

Statistiques appliquées

Avec SPSS

1. Méthodologie de recherche et traitement des données en sciences humaines
2. Utilisation de SPSS

Méthodologie de recherche et traitement des données en sciences humaines

- I. Démarche de la recherche
- II. Types de variables et échelles de mesure
- III. Le questionnaire
- IV. Du questionnaire à la base de données

I. Démarche de la recherche

Au départ de toute recherche : un questionnement

Les 4 étapes de la recherche :

1. Circonscrire l'objet de recherche
2. Construire l'objet de recherche
3. Construire les instruments d'observations
4. Analyser l'objet d'observations

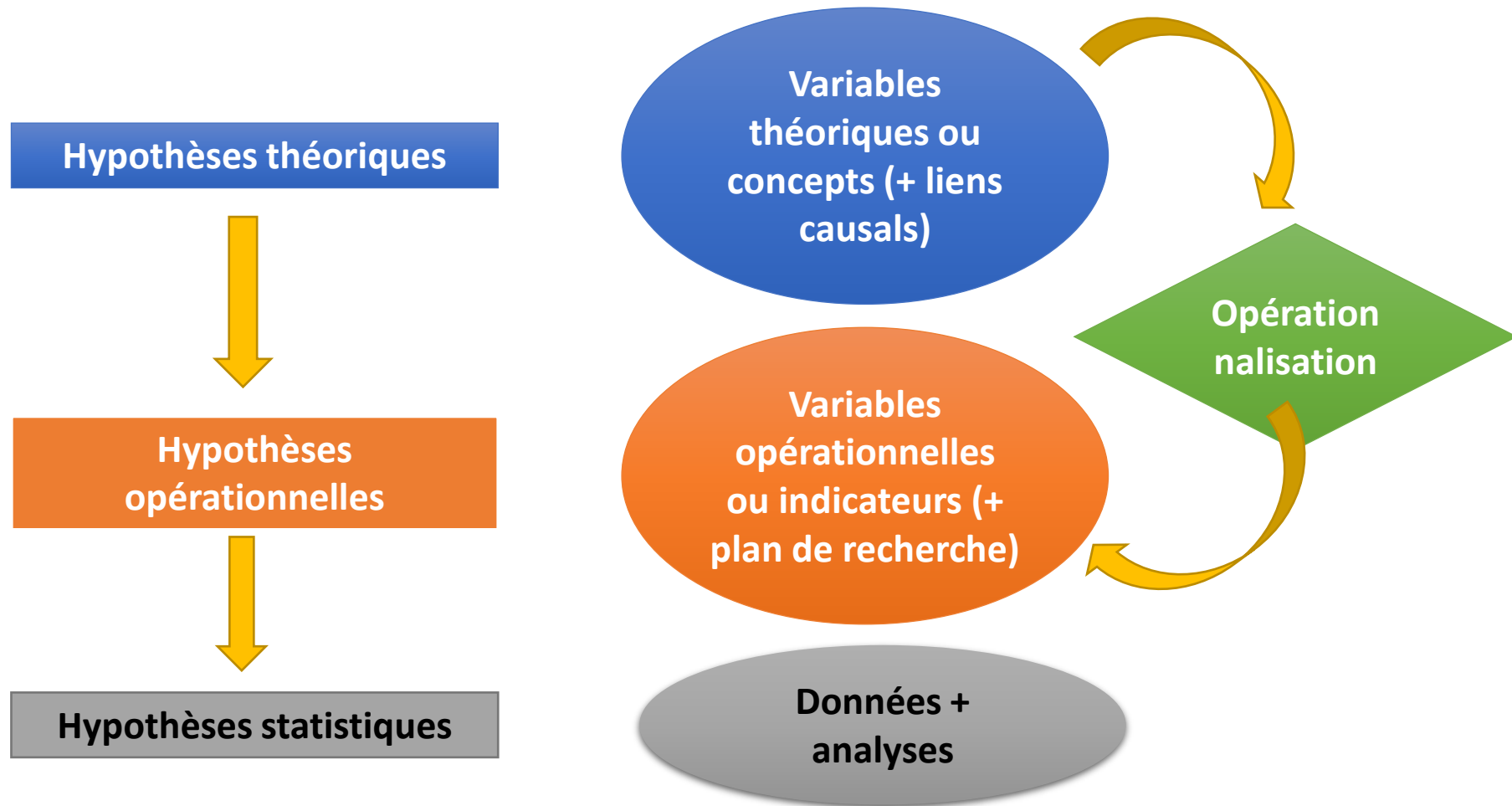
I.1. Circonscrire l'objet de recherche

- 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.

I.2. 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
- Pour cela il faut déterminer :
 1. les dimensions des concepts
 2. les composantes des dimensions
 3. les indicateurs des composantes

Opérationnalisation...



I.3. 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 ?

Comment observer ?

- Questionnaire
- Entretien
- Tests
- Observation
- Mesures psychophysiologiques
- Analyse du texte
- ...

1.4. Analyser l'objet d'observation

Pourquoi analyser les données récoltées ?

- Pour tester notre hypothèse de départ → besoin d'un logiciel statistique.

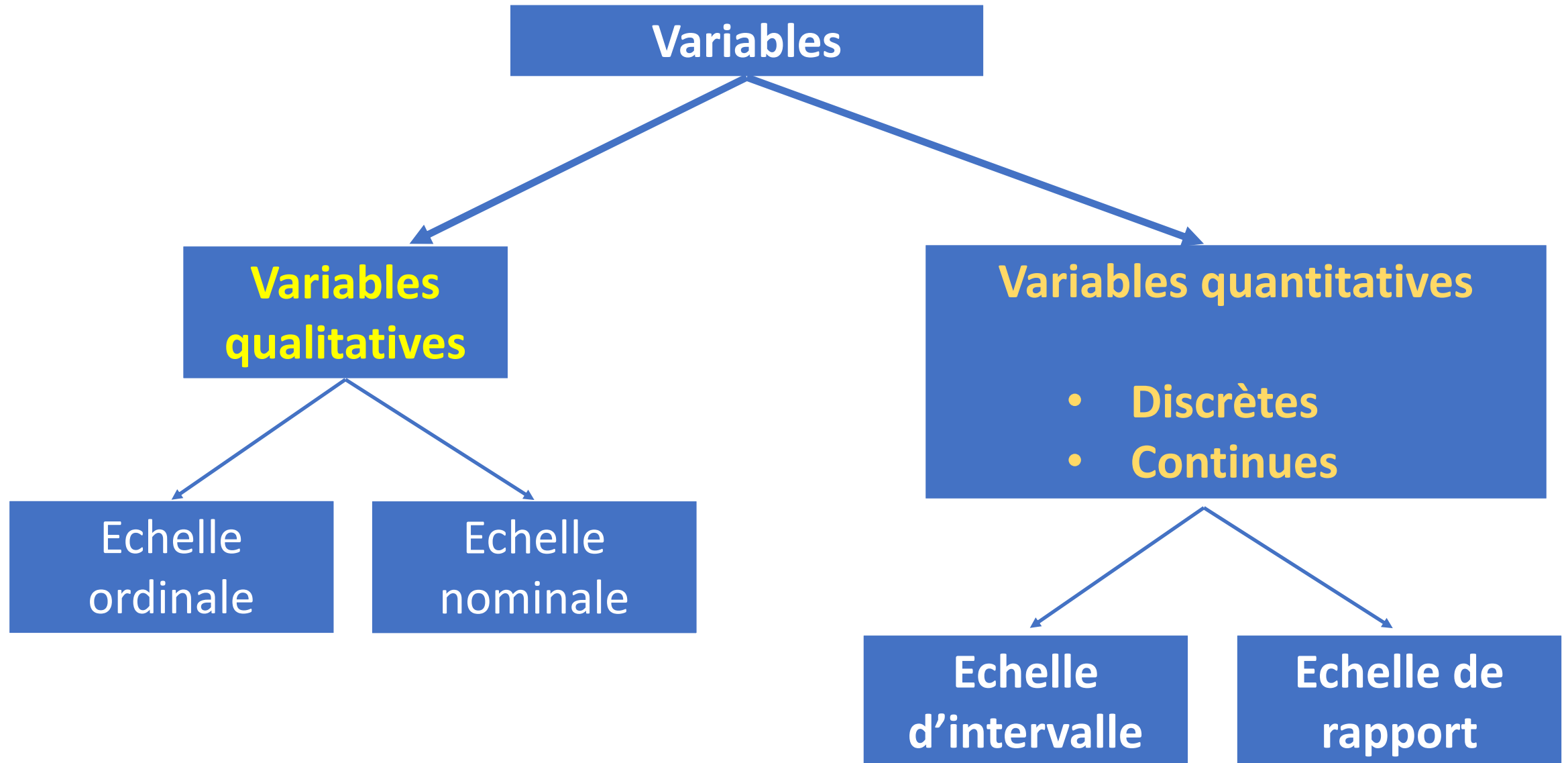
1.4.1. Analyser l'objet d'observation : 3 opérations

- 1. Préparation des données (décrire et agréger).
- 2. Analyse des relations entre les variables.
- 3. Comparaison des résultats observés avec les résultats attendus et interprétation de l'écart.

1.4.2. Analyse statistique

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

II. Types de variables et échelles de mesure



II.1. Variables qualitatives

Les modalités sont : Catégories (mots, lettres), non numériques

- sexe, étude, profession, nationalité, niveau scolaire, diplôme, etc.

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

(échelle de likert : Une échelle de Likert est un outil psychométrique permettant de mesurer une attitude chez des individus.) :

Exemple: degré d'approbation

Pas du tout d'accord	Peu d'accord	Plutôt d'accord	Tout à fait d'accord
---------------------------------	---------------------	----------------------------	---------------------------------

II.1.1. 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.

II.1.2. 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.

II.2. 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

- II.2.1. Variables quantitatives : Niveaux de mesure

- II.2.1.1. L'échelle d'intervalles (ou continue)

Pour représenter la différence et l'ordre entre les sujets. Contrairement à l'échelle ordinaire les distances entre les valeurs ont un sens.

- II.2.1.2. L'échelle de rapport

Pour représenter la différence et l'ordre entre les sujets (d'une manière relative).

III. Le questionnaire :

III.1. Exemples :

1. Echelle d'expressivité : Amour, Joie, Peur, Colère, Tristesse, Honte
2. Extraversion (amical, sociable, vif,....)
4. Intellect (malin, intelligent, cultivé,)
5. Agréabilité (chaleureux, serviable...)
7. Echelle d'ouverture
 - Ouverture aux sentiments,
 - Ouverture aux idées,
 - Ouverture aux actions et
 - Ouverture aux valeurs
8. Données sociodémographiques : Age, Sexe, Taille, Poids, Nationalité, Etude,
9. Etc.

III.2. Types de questions

- III2.1. Questions fermées :

	Pas du tout d'accord	Un peu d'accord	Moyenne ment d'accord	Plutôt d'accord	Tout à fait d'accord
Je considère que le taux d'analphabétisation est plus élevé dans les zones rurales	1	2	3	4	5

- III2.2. Questions ouvertes :

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

- III2.3. Questions à réponse unique :

Sexe: Femme

Homme

- III2.4. Questions à réponses multiples :

Quelles loisirs préférez-vous? (plusieurs réponses sont possibles)

- sport
- rencontre des amis
- activités culturelles
- jouer de la musique
- télévision
- surfer sur internet
- bricolage
- autre (précisez):.....

IV. Du questionnaire à la base de données

- Une fois le questionnaire rempli, on dispose de :
 1. N individus (objets, unités statistiques) qui ont répondu au questionnaire
 2. 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 (ou tableau) et le fichier des données.

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

IV.1. Codage

- Exemple de tableau de codage pour un questionnaire

Variable (ou question)	Valeur
Numéro d'identification	
Expressivité	1 = jamais ou presque jamais 2 = parfois 3 = souvent 4 = presque toujours ou toujours
Satisfaction de....	1 = pas du tout 5 = tout à fait
Loisirs	0 = ne préfère pas ce loisir 1 = préfère ce loisir
Sexe	1 = femme 2 = homme
Age	

Variable	Codage
Région ou pays (nationalité)	1 = Adrar 2 = Chlef Ou 1 = Algérie 2 = Maroc
Domaine d'études	1 = mathématiques, 2 = psychologie, 3 = sociologie, 4 = droit, 5 = langues, 7 = ingénierie, 8 = économie, 9 = informatique
Fourchette de revenue	1 = de 15 à 20 milles dinars 2 = de 21 à 40 milles dinars 3 = de 41 à 80 milles dinars 6 = plus de 150 milles dinars
Taille en cm	
Poids en kg	
Code de sécurité social	

IV.2. 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.

Exemples : restreindre l'échelle

Passage de 4 modalités à 2:

Ancienne variable				
	Pas satisfait	Plutôt pas satisfait	Plutôt satisfait	Satisfait
Etes-vous satisfait de la manière dont votre employé gère son entreprise?	1	2	3	4

Nouvelle variable		
	Insatisfait	Satisfait
Etes-vous satisfait de la manière dont votre employé gère son entreprise?	1	2

Age: variable quantitative allant de 15 à 40 ans → variable recodée en 5 modalités:

- 15 à 20
- 21 à 25
- 26 à 30
- 31 à 35
- 36 à 40

De « Wilaya » a « Région »:

- 1 – Adrar
- 2 – Chlef
- 3 – Aghouat



....

....

58 - El Meniaa

- 1 - Ouest de l'Algérie
- 2 - Est de l'Algérie
- 3 - Nord de l'Algérie
- 4 - Sud de l'Algérie

AXE 2

Utilisation de SPSS:

- 1. Les fenêtres et leurs contenus**
- 2. Manipulations des données d'un fichier**
- 3. Manipulation des résultats**

SPSS (Statistical Package for the Social Sciences) (SPSS Inc)

En 2009, la compagnie décide de changer le nom de ses produits en PASW, pour Predictive Analytics Software et est rachetée par IBM

Introduction

- Le logiciel SPSS adapté à la mise en œuvre de techniques d'analyse des données statistiques. Il favorise la gestion des données dans un environnement graphique convivial associant menu descriptif et boîtes de dialogue.
Il permet de traiter les données avec efficacité et d'effectuer des analyses sur de grandes bases de données. Il offre plusieurs possibilités pour organiser et synthétiser les informations statistiques.
- L'exploitation et le traitement statistique des données d'enquête de terrain , notamment pour d'organisation, la restructuration, le traitement et l'exploitation des informations

1 . Les fenêtres et leurs contenus

- Le logiciel SPSS est composé de trois fenêtres :
 1. la fenêtre des données (ou fenêtre de l'éditeur),
 2. La fenêtre des résultats (souvent appelée viewer) et
 3. La fenêtre des syntaxes.

recidivism.sav [Ensemble_de_données1] - IBM SPSS Statistics Editeur de données

Fichier Edition Affichage Données Transformer Analyse Marketing direct Graphes Utilitaires Fenêtre Aide

📁 🖨️ 📄 ↶ ↷ 📊 ⬇️ 📑 🔍 📈 ⚖️ 📊 🔄 🗑️ ABC

	Nom	Type	Largeur	Décimales	Etiquette	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	age	Numérique	4	0	Age en années	Aucun	Aucun	4	☰ Droite	📏 Echelle	↘ Entrée
2	agecat									📏 Echelle	↘ Entrée
3	marital									📏 Echelle	↘ Entrée
4	social									📊 Ordinales	↘ Entrée
5	ed									📊 Ordinales	↘ Entrée
6	employ									📏 Echelle	↘ Entrée
7	gender									📏 Echelle	↘ Entrée
8	crime1									📊 Ordinales	↘ Entrée
9	violent1									📏 Echelle	↘ Entrée
10	bail									📏 Echelle	↘ Entrée
11	rehab									📏 Echelle	↘ Entrée
12	arrest2									📏 Echelle	↘ Entrée
13	time									📏 Echelle	↘ Entrée
14	crime2									📊 Ordinales	↘ Entrée
15	violent2									📏 Echelle	↘ Entrée
16	convict2									📏 Echelle	↘ Entrée
17											
18											
19											
20											
21											
22											
23											
24											
25											

*Résultats1 [Document1] - IBM SPSS Statistics Viewer

Fichier Edition Affichage Données Transforme Insérer Format Analyse Marketing dire Graphes Utilitaires Fenêtre Aide

📁 🖨️ 📄 🔍 📄 ↶ ↷ 📊 ⬇️ 📑 🔍 📈 ⚖️ 📊 🔄 🗑️ ABC

← → + - 📄 📄 📄

Résultats

- Log
- Explorer
 - Titre
 - Remarques
 - Ensemble de don
 - Récapitulatif du tr
 - Descriptives
 - Age en années
 - Titre
 - Diagramme t
 - Boîte à mous

Explorer

[Ensemble_de_données1] E:\Cours SPSS\data files\recidiv

Récapitulatif du traitement des observations

	Observations				
	Valide		Manquante		Tot
	N	Pourcent	N	Pourcent	N
Age en années	1000	100.0%	0	0.0%	1000

Descriptives

		Statistique
Age en années	Moyenne	28.07
	Intervalle de confiance à	Borne inférieure
		27.57

Le processeur IBM SPSS Statistics est prêt

Affichage des données Affichage des variables

Le processeur IBM SPSS Statistics est prêt

1.1 La fenêtre des données

- La fenêtre des données permet d'entrer les données. Cette fenêtre présente deux onglets (en bas à gauche) : affichage des variables, et affichage des données.

2.2 La fenêtre des résultats

- Elle affiche les résultats des commandes exécutées par SPSS. C'est dans cette fenêtre que nous observerons les sorties de tableaux, de graphiques, etc. Cette fenêtre s'ouvre automatiquement dès que vous lancez une procédure qui doit afficher des résultats.

1.3 La fenêtre des syntaxes

- Elle permet de saisir les commandes SPSS. En effet, SPSS présente une configuration Windows avec menu déroulant (on peut donc travailler avec la souris) et une possibilité de saisir les programmes en langage SPSS.

2. Manipulations des données d'un fichier

Caractéristiques d'une variable

- On ne peut bien manipuler les données sans connaître parfaitement les caractéristiques des variables.
- Dans SPSS, dans la fenêtre des données et dans l'affichage des variables, il y a les caractéristiques suivantes: **Nom**, **Type**, **Largeur**, **Décimales**, **Etiquette**, **Valeur**, **Manquante**, **Colonnes**, **Align**, **Mesure** et **Rôle**.

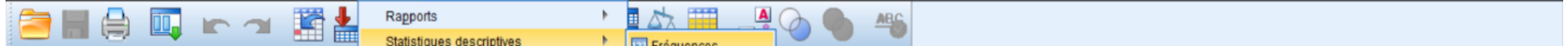
	Nom	Type	Largeur	Décimales	Etiquette	Valeurs	Manquant	Colonnes	Align	Mesure	Rôle
1	age	Numérique	4	0	Age en an...	Aucun	Aucun	4	Droite	Echelle	Entrée
2	agecat	Numérique	4	0	Catégorie ...	{1, 18-21}...	Aucun	6	Droite	Echelle	Entrée
3	marital	Numérique	4	0	Statut mari...	{0, Non ma...	Aucun	7	Droite	Echelle	Entrée
4	social	Numérique	4	0	Statut social	{1, Classe ...	Aucun	6	Droite	Ordinales	Entrée
5	ed	Numérique	4	0	Niveau d'éd...	{1, Inf au B...	Aucun	4	Droite	Ordinales	Entrée
6	employ	Numérique	4	0	Employé	{0, Non}...	Aucun	6	Droite	Echelle	Entrée

3. Manipulation des résultats sur SPSS

• On ouvre le fichier recidivism.sav : Ce fichier de données d'hypothèse concerne les mesures qu'entreprend une agence administrative d'application de la loi pour interpréter les taux de récidive dans la juridiction. Chaque observation correspond à un récidiviste et enregistre les informations démographiques qui lui sont propres, certains détails sur le premier délit commis, ainsi que le temps écoulé jusqu'à la seconde arrestation si elle s'est produite dans les deux années suivant la première.

• Exploration des données :

• Variable : Age en années



2 : age 19

	age	agecat	marital	social
1	22	2	0	1
2	19	1	0	1
3	28	2	0	3
4	36	3	1	3
5	27	2	1	2
6	29	2	1	2
7	24	2	0	1
8	28	2	1	1
9	36	3	1	1
10	25	2	0	3
11	40	4	0	1
12	21	1	0	2
13	18	1	1	2
14	20	1	0	2
15	41	4	0	2
16	19	1	0	2
17	20	1	1	1
18	20	1	0	3
19	39	3	0	2
20	23	2	0	2
21	47	4	1	2
22	36	3	0	2
23	28	2	0	3
24	35	3	0	1
25	24	2	0	1
26	27	2	0	1
27	34	3	1	1

- Rapports
- Statistiques descriptives
- Tableaux
- Comparer les moyennes
- Modèle linéaire général
- Modèles linéaires généralisés
- Modèles Mixtes
- Corrélation
- Régression
- Log Linéaire
- Réseaux neuronaux
- Classification
- Réduction des dimensions
- Echelle
- Tests non paramétriques
- Prévisions
- Survie
- Réponses multiples
- Analyse des valeurs manquantes
- Imputation multiple
- Echantillons complexes
- Simulation...
- Contrôle de qualité
- Courbe ROC...

	arrest2	time	crime2	violent2	convict2	var	var	var	var
1	1	74,00	4	0	0				
2	1	239,00	3	0	1				
3	1	208,00	1	0	0				
4	0	9999,00	-1	-1	-1				
5	0	9999,00	-1	-1	-1				
6	1	287,00	3	0	1				
7	0	528,00	1	1	1				
8	0	9999,00	-1	-1	-1				
9	0	9999,00	-1	-1	-1				
10	0	431,00	2	0	1				
11	0	9999,00	-1	-1	-1				
12	0	177,00	3	0	1				
13	0	446,00	1	0	1				
14	0	264,00	3	0	1				
15	1	262,00	4	1	1				
16	0	538,00	1	0	1				
17	0	9999,00	-1	-1	-1				
18	0	512,00	3	0	0				
19	0	155,00	1	0	1				
20	0	167,00	2	0	1				
21	0	9999,00	-1	-1	-1				
22	0	9999,00	-1	-1	-1				
23	0	9999,00	-1	-1	-1				
24	0	9999,00	-1	-1	-1				
25	0	708,00	3	1	1				
26	0	9999,00	-1	-1	-1				
27	0	9999,00	-1	-1	-1				

Visible : 16 variables sur 16

Vue de données Vue des variables

Active Windows
Accédez aux paramètres pour activer Windows.



	age	agecat	marital	social	ed	employ	gender	crime1	violent1	bail	rehab	arrest2	time	crime2	violent2	convict2	var	var	var	var
1	22	2	0	1	1	1	0	1	0	1	0	1	74,00	4	0	0				
2	19	1	0	1	2	0	1	2	0	0	1	1	239,00	3	0	1				
3	28	2	0	3	1	0	0	1	0	1	1	1	208,00	1	0	0				
4	36	3	1	3	2															
5	27	2	1	2	2															
6	29	2	1	2	2															
7	24	2	0	1	2															
8	28	2	1	1	1															
9	36	3	1	1	1															
10	25	2	0	3	1															
11	40	4	0	1	2															
12	21	1	0	2	3															
13	18	1	1	2	1															
14	20	1	0	2	2															
15	41	4	0	2	2															
16	19	1	0	2	1															
17	20	1	1	1	2															
18	20	1	0	3	1															
19	39	3	0	2	1															
20	23	2	0	2	2															
21	47	4	1	2	3															
22	36	3	0	2	1															
23	28	2	0	3	2	0	0	2	0	1	1	0	9999,00	-1	-1	-1				
24	35	3	0	1	2	1	1	1	0	1	1	0	9999,00	-1	-1	-1				
25	24	2	0	1	1	0	1	4	0	1	0	1	708,00	3	1	1				
26	27	2	0	1	1	0	0	3	0	1	1	0	9999,00	-1	-1	-1				
27	34	3	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
28	26	2	1	1	2	0	0	2	0	1	0	1	692,00	1	0	1				

Fréquences

Variable(s) :

- Age en années [age]

Afficher les tables de fréquences

OK Coller Réinitialiser Annuler Aide

Statistiques...
 Graphiques...
 Format...
 Style...
 Bootstrap...

I.1. Fréquences :

Statistiques		
Age en années		
N	Valide	80
	Manquant	0

Fréquence	Fréquence relative	Fréquence cumulée
$\sum_{i=1}^k n_i = n$	$f_i = \frac{n_i}{n}$	$F_1 = f_1$ $F_2 = f_1 + f_2$ $F_3 = f_1 + f_2 + f_3$ $F_k = f_1 + \dots + f_k = 1$ (ou 100%)
n_i : le nombre d'éléments ayant la même modalité ou valeur n : le nombre totale d'éléments Le pourcentage = fréquence * 100		

		Age en années			
		Fréquence	Pourcentage	Pourcentage valide	Pourcentage cumulé
Valide	18	5	6,3	6,3	6,3
	19	6	7,5	7,5	13,8
	20	5	6,3	6,3	20,0
	21	5	6,3	6,3	26,3
	22	4	5,0	5,0	31,3
	23	3	3,8	3,8	35,0
	24	4	5,0	5,0	40,0
	25	3	3,8	3,8	43,8
	26	3	3,8	3,8	47,5
	27	3	3,8	3,8	51,2
	28	5	6,3	6,3	57,5
	29	4	5,0	5,0	62,5
	30	2	2,5	2,5	65,0
	32	4	5,0	5,0	70,0
	34	1	1,3	1,3	71,3
	35	2	2,5	2,5	73,8
	36	5	6,3	6,3	80,0
	37	3	3,8	3,8	83,8
	39	3	3,8	3,8	87,5
	40	2	2,5	2,5	90,0
41	2	2,5	2,5	92,5	
42	1	1,3	1,3	93,8	
43	2	2,5	2,5	96,3	
44	1	1,3	1,3	97,5	
47	2	2,5	2,5	100,0	
Total		80	100,0	100,0	

I.2. Descriptives :

The screenshot displays the IBM SPSS Statistics interface. The 'Analyse' menu is open, and 'Descriptives' is selected. The data table in the background shows the following columns: age, agecat, marital, social, arrest2, time, crime2, violent2, convict2, and four empty 'var' columns. The 'arrest2' column is highlighted in yellow.

	age	agecat	marital	social	arrest2	time	crime2	violent2	convict2	var	var	var	var
1	22	2	0	1	1	74,00	4	0	0				
2	19	1	0	1	1	239,00	3	0	1				
3	28	2	0	3	1	208,00	1	0	0				
4	36	3	1	3	0	9999,00	-1	-1	-1				
5	27	2	1	2	0	9999,00	-1	-1	-1				
6	29	2	1	2	1	287,00	3	0	1				
7	24	2	0	1	0	528,00	1	1	1				
8	28	2	1	1	0	9999,00	-1	-1	-1				
9	36	3	1	1	0	9999,00	-1	-1	-1				
10	25	2	0	3	0	431,00	2	0	1				
11	40	4	0	1	0	9999,00	-1	-1	-1				
12	21	1	0	2	0	177,00	3	0	1				
13	18	1	1	2	0	446,00	1	0	1				
14	20	1	0	2	0	264,00	3	0	1				
15	41	4	0	2	1	262,00	4	1	1				
16	19	1	0	2	0	538,00	1	0	1				
17	20	1	1	1	0	9999,00	-1	-1	-1				
18	20	1	0	3	0	512,00	3	0	0				
19	39	3	0	2	0	155,00	1	0	1				
20	23	2	0	2	0	167,00	2	0	1				
21	47	4	1	2	0	9999,00	-1	-1	-1				
22	36	3	0	2	0	9999,00	-1	-1	-1				
23	28	2	0	3	2	9999,00	-1	-1	-1				
24	35	3	0	1	2	9999,00	-1	-1	-1				
25	24	2	0	1	1	708,00	3	1	1				
26	27	2	0	1	1	9999,00	-1	-1	-1				
27	34	3	1	1	2	9999,00	-1	-1	-1				



3 : arrest2 1 Visible : 16 variables sur 16

	age	agecat	marital	social	ed	employ	gender	crime1	violent1	bail	rehab	arrest2	time	crime2	violent2	convict2	var	var	var	var
1	22	2	0	1	1	1	0	1	0	1	0	1	74,00	4	0	0				
2	19	1	0	1	2	0	1	2	0	0	1	1	239,00	3	0	1				
3	28	2	0	3	1	0	0	1	0	1	1	1	999,00	4	0	0				
4	36	3	1	3	2	0	0	0	0	0	0	0	999,00	1	0	0				
5	27	2	1	2	2	0	0	0	0	0	0	0	999,00	1	0	0				
6	29	2	1	2	2	0	0	0	0	0	0	0	999,00	1	0	0				
7	24	2	0	1	2	0	0	0	0	0	0	0	999,00	1	0	0				
8	28	2	1	1	1	0	0	0	0	0	0	0	999,00	1	0	0				
9	36	3	1	1	1	0	0	0	0	0	0	0	999,00	1	0	0				
10	25	2	0	3	1	0	0	0	0	0	0	0	999,00	1	0	0				
11	40	4	0	1	2	0	0	0	0	0	0	0	999,00	1	0	0				
12	21	1	0	2	3	0	0	0	0	0	0	0	999,00	1	0	0				
13	18	1	1	2	1	0	0	0	0	0	0	0	999,00	1	0	0				
14	20	1	0	2	2	0	0	0	0	0	0	0	999,00	1	0	0				
15	41	4	0	2	2	0	0	0	0	0	0	0	999,00	1	0	0				
16	19	1	0	2	1	0	0	0	0	0	0	0	999,00	1	0	0				
17	20	1	1	1	2	0	0	0	0	0	0	0	999,00	1	0	0				
18	20	1	0	3	1	0	0	0	0	0	0	0	999,00	1	0	0				
19	39	3	0	2	1	0	0	0	0	0	0	0	999,00	1	0	0				
20	23	2	0	2	2	0	0	0	0	0	0	0	999,00	1	0	0				
21	47	4	1	2	3	0	0	0	0	0	0	0	999,00	1	0	0				
22	36	3	0	2	1	0	0	0	0	0	0	0	999,00	1	0	0				
23	28	2	0	3	2	0	0	2	0	1	1	0	999,00	-1	-1	-1				
24	35	3	0	1	2	1	1	1	0	1	1	0	999,00	-1	-1	-1				
25	24	2	0	1	1	0	1	4	0	1	0	1	708,00	3	1	1				
26	27	2	0	1	1	0	0	3	0	1	1	0	999,00	-1	-1	-1				
27	34	3	1	1	2	0	0	2	0	0	0	0	999,00	-1	-1	-1				
28	26	2	1	1	2	0	0	2	0	1	0	1	500,00	4	0	1				

Descriptives

Variable(s) :

- Age en années [age]

- Catégorie d'âge [agecat]
- Statut marital [marital]
- Statut social [social]
- Niveau d'éducation [ed]
- Employé [employ]
- Sexe [gender]
- Sévérité de la première agression [crime1]
- Première agression violente [violent1]
- Caution demandée [bail]
- Réhabilitation [rehab]
- Seconde arrestation [arrest2]
- Durée avant seconde arrestation [time]
- Sévérité de la deuxième agression [crime2]
- Deuxième agression violente [violent2]
- Deuxième condamnation [convict2]

Enregistrer des valeurs standardisées dans des variables

Options Style... Bootstrap...

OK Coller Réinitialiser Annuler Aide

Vue de données Vue des variables

Activer Windows
Accédez aux paramètres pour activer Windows.

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
Age en années	80	18	47	28,50	8,146
N valide (liste)	80				

I.3. Explorer

recidivism_réduit.sav [Jeu_de_données6] - IBM SPSS Statistics Editeur de données

Fichier Edition Affichage Données Transformer **Analyse** Marketing direct Graphiques Utilitaires Fenêtre Aide

3 : arrest2 1

Statistiques descriptives

- Fréquences
- Descriptives
- Explorer**
- Tableaux croisés
- Analyse TURF
- Rapport
- Tracés P-P...
- Tracé Q-Q...

	age	agecat	marital	social	arrest2	time	crime2	violent2	convict2	var	var	var	var
1	22	2	0	1	1	74,00	4	0	0				
2	19	1	0	1	1	239,00	3	0	1				
3	28	2	0	3	1	208,00	1	0	0				
4	36	3	1	3	0	9999,00	-1	-1	-1				
5	27	2	1	2	0	9999,00	-1	-1	-1				
6	29	2	1	2	1	287,00	3	0	1				
7	24	2	0	1	0	528,00	1	1	1				
8	28	2	1	1	0	9999,00	-1	-1	-1				
9	36	3	1	1	0	9999,00	-1	-1	-1				
10	25	2	0	3	0	9999,00	-1	-1	-1				
11	40	4	0	1	0	431,00	2	0	1				
12	21	1	0	2	0	9999,00	-1	-1	-1				
13	18	1	1	2	0	177,00	3	0	1				
14	20	1	0	2	0	446,00	1	0	1				
15	41	4	0	2	0	264,00	3	0	1				
16	19	1	0	2	1	262,00	4	1	1				
17	20	1	1	1	0	538,00	1	0	1				
18	20	1	0	3	0	9999,00	-1	-1	-1				
19	39	3	0	2	0	512,00	3	0	0				
20	23	2	0	2	0	155,00	1	0	1				
21	47	4	1	2	0	167,00	2	0	1				
22	36	3	0	2	0	9999,00	-1	-1	-1				
23	28	2	0	3	0	9999,00	-1	-1	-1				
24	35	3	0	1	0	9999,00	-1	-1	-1				
25	24	2	0	1	0	708,00	3	1	1				
26	27	2	0	1	0	9999,00	-1	-1	-1				
27	34	3	1	1	0	9999,00	-1	-1	-1				

Visible : 16 variables sur 16

Vue de données | Vue des variables

Explorer

Le processeur IBM SPSS Statistics est prêt | Unicode:OFF



3 : arrest2 1 Visible : 16 variables sur 16

	age	agecat	marital	social	ed	employ	gender	crime1	violent1	bail	rehab	arrest2	time	crime2	violent2	convict2	var	var	var	var
1	22	2	0	1	1	1	0	1	0	1	0	1	74,00	4	0	0				
2	19	1	0	1	2	0	1	2	0	0	1	1	239,00	3	0	1				
3	28	2	0	3	1	0	0	1	0	1	1	1	208,00	1	0	0				
4	36	3	1	3	2	1	0	1	0	1	0	0	9999,00	-1	-1	-1				
5	27	2	1	2	2	0	0	1	0	1	1	0	9999,00	-1	-1	-1				
6	29	2	1	2	2	1	0													
7	24	2	0	1	2	1	0													
8	28	2	1	1	1	1	0													
9	36	3	1	1	1	1	0													
10	25	2	0	3	1	1	0													
11	40	4	0	1	2	1	1													
12	21	1	0	2	3	0	0													
13	18	1	1	2	1	0	0													
14	20	1	0	2	2	1	0													
15	41	4	0	2	2	1	0													
16	19	1	0	2	1	1	0													
17	20	1	1	1	2	1	0													
18	20	1	0	3	1	1	0													
19	39	3	0	2	1	1	0													
20	23	2	0	2	2	0	0													
21	47	4	1	2	3	1	0	1	0	1	0	0	9999,00	-1	-1	-1				
22	36	3	0	2	1	0	1	2	0	1	0	0	9999,00	-1	-1	-1				
23	28	2	0	3	2	0	0	2	0	1	1	0	9999,00	-1	-1	-1				
24	35	3	0	1	2	1	1	1	0	1	1	0	9999,00	-1	-1	-1				
25	24	2	0	1	1	0	1	4	0	1	0	1	708,00	3	1	1				
26	27	2	0	1	1	0	0	3	0	1	1	0	9999,00	-1	-1	-1				
27	34	3	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
28	26	2	1	1	2	0	0	2	0	0	0	1	502,00	1	0	1				

Explorer

Liste Variables dépendantes :
 Age en années [age] [Statistiques...]

Liste des facteurs :
 [Tracés...]

Etiqueter les observations par :
 [Options]

Affichage
 Les deux Statistiques Tracés

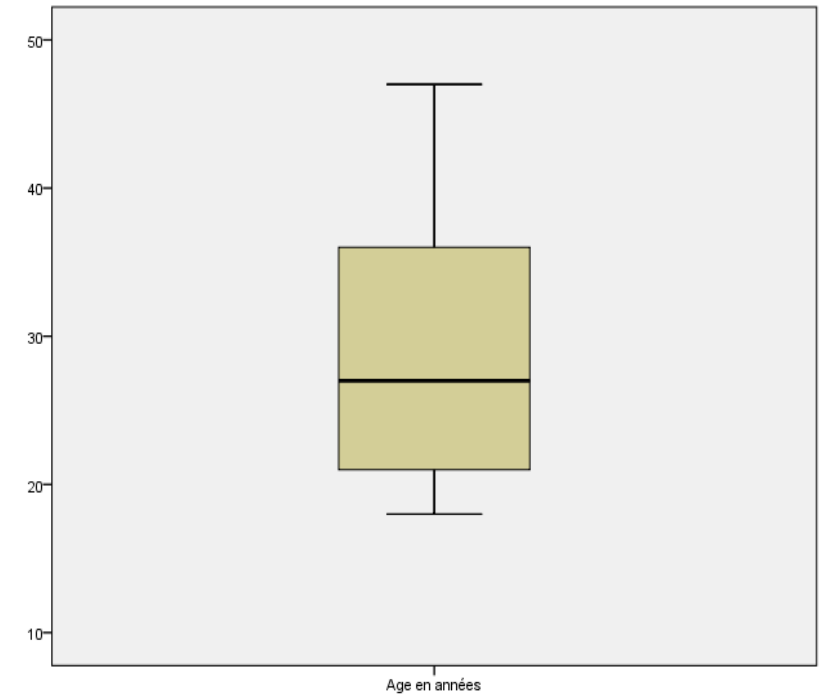
[Bootstrap...]

[OK] [Coller] [Réinitialiser] [Annuler] [Aide]

Récapitulatif de traitement des observations						
	Observations					
	Valide		Manquant		Total	
	N	Pourcentage	N	Pourcentage	N	Pourcentage
Age en années	80	100,0%	0	0,0%	80	100,0%

Descriptives				
		Statistiques	Erreur standard	
Age en années	Moyenne	28,50	,911	
	Intervalle de confiance à 95 % pour la moyenne	Borne inférieure	26,69	
		Borne supérieure	30,31	
	Moyenne tronquée à 5 %	28,15		
	Médiane	27,00		
	Variance	66,354		
	Ecart type	8,146		
	Minimum	18		
	Maximum	47		
	Plage	29		
	Plage interquartile	15		
	Asymétrie	,527	,269	
	Kurtosis	-,840	,532	

Valeurs extrêmes				
			Numéro de l'observation	Valeur
Age en années	Supérieure	1	21	47
		2	33	47
		3	79	44
		4	46	43
		5	61	43
	Inférieure	1	74	18
		2	68	18
		3	63	18
		4	49	18
		5	13	18



Boîte à moustache

- Une boîte à moustaches est un autre graphique utile pour la visualisation de la distribution des données. Elle contient plusieurs mesures statistiques.

Remarque : Cet exemple utilise Employee data.

A partir des menus, sélectionnez :

- **Graphiques > Sélecteur de modèles de représentations graphiques...**
 1. Dans l'onglet de base, sélectionnez sexe et salaire. (Utilisez la combinaison Ctrl+clic pour sélectionner plusieurs champs/variables.)
 2. Sélectionnez Boîte à moustaches.
 3. Cliquez sur OK.

Fichier Edition Affichage Données Transformer Analyse Marketing direct Graphiques Utilitaires Fenêtre Aide

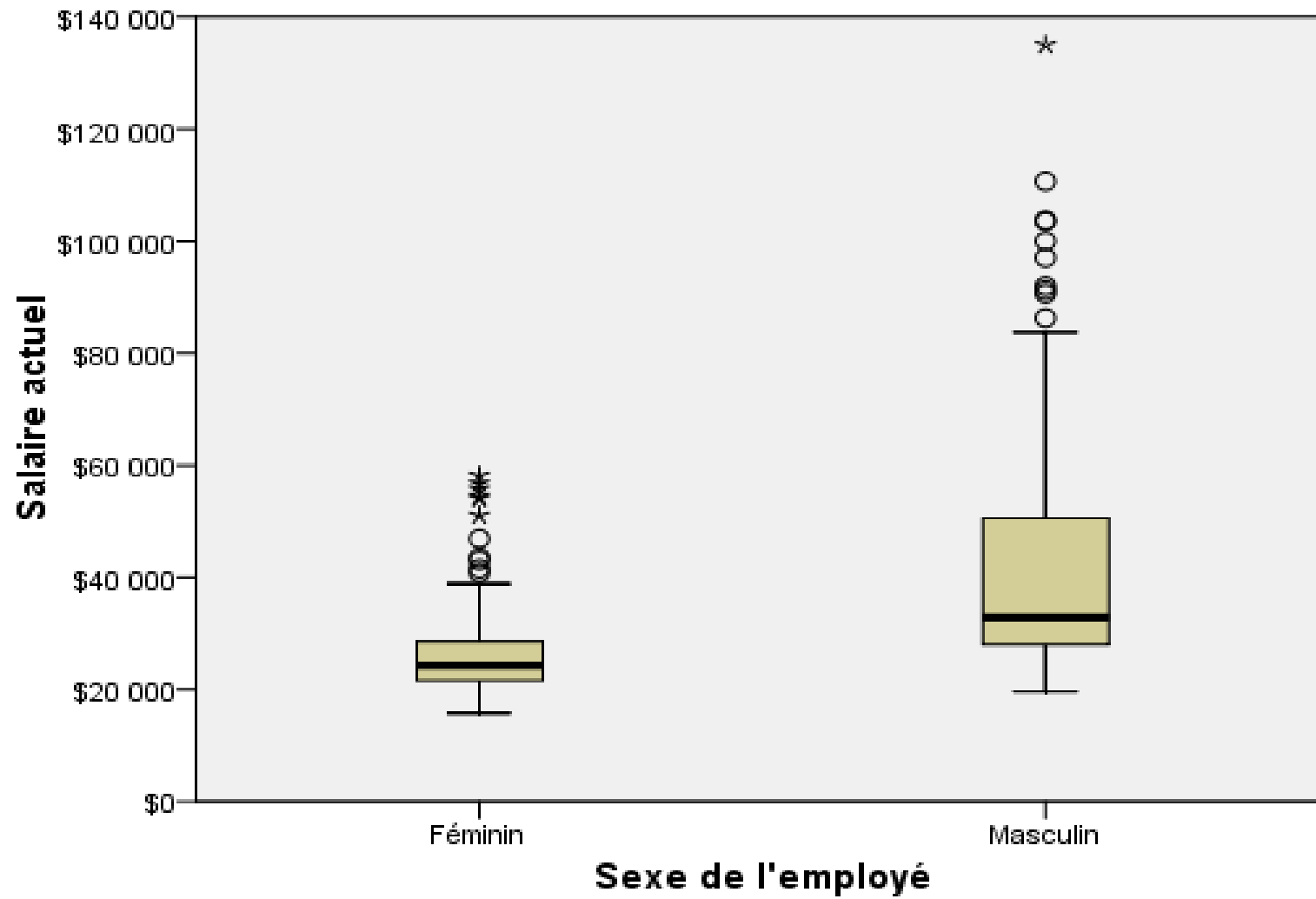
Générateur de graphiques...
Sélecteur de modèles de représentations graphiques...
Comparer les sous-groupes
Diagrammes de variables de régression
Boîtes de dialogue ancienne version

Visible : 10 variables sur 10

	id	sexe	datenais	educ	catemp														
1	1	m	02/03/1952	15															
2	2	m	05/23/1958	16															
3	3	f	07/26/1929	12	1	\$21,450	\$12,000	98	381	0									
4	4	f	04/15/1947	8	1	\$21,900	\$13,200	98	190	0									
5	5	m	02/09/1955	15	1	\$45,000	\$21,000	98	138	0									
6	6	m	08/22/1958	15	1	\$32,100	\$13,500	98	67	0									
7	7	m	04/26/1956	15	1	\$36,000	\$18,750	98	114	0									
8	8	f	05/06/1966	12	1	\$21,900	\$9,750	98	0	0									
9	9	f	01/23/1946	15	1	\$27,900	\$12,750	98	115	0									
10	10	f	02/13/1946	12	1	\$24,000	\$13,500	98	244	0									
11	11	f	02/07/1950	16	1	\$30,300	\$16,500	98	143	0									
12	12	m	01/11/1966	8	1	\$28,350	\$12,000	98	26	1									
13	13	m	07/17/1960	15	1	\$27,750	\$14,250	98	34	1									
14	14	f	02/26/1949	15	1	\$35,100	\$16,800	98	137	1									
15	15	m	08/29/1962	12	1	\$27,300	\$13,500	97	66	0									
16	16	m	11/17/1964	12	1	\$40,800	\$15,000	97	24	0									
17	17	m	07/18/1962	15	1	\$46,000	\$14,250	97	48	0									
18	18	m	03/20/1956	16	3	\$103,750	\$27,510	97	70	0									
19	19	m	08/19/1962	12	1	\$42,300	\$14,250	97	103	0									
20	20	f	01/23/1940	12	1	\$26,250	\$11,550	97	48	0									
21	21	f	02/19/1963	16	1	\$38,850	\$15,000	97	17	0									
22	22	m	09/24/1940	12	1	\$21,750	\$12,750	97	315	1									
23	23	f	03/15/1965	15	1	\$24,000	\$11,100	97	75	1									
24	24	f	03/27/1933	12	1	\$16,950	\$9,000	97	124	1									
25	25	f	07/01/1942	15	1	\$21,150	\$9,000	97	171	1									
26	26	m	11/08/1966	15	1	\$31,050	\$12,600	96	14	0									
27	27	m	03/19/1954	19	3	\$60,375	\$27,480	96	96	0									
28	28	m	04/14/1962	15	1	\$33,550	\$14,250	96	12	0									

Activier Windows
Accédez aux paramètres pour activer Windows.

Sélecteur de modèles de représentations graphiques... Le processeur IBM SPSS Statistics est prêt Unicode:OFF

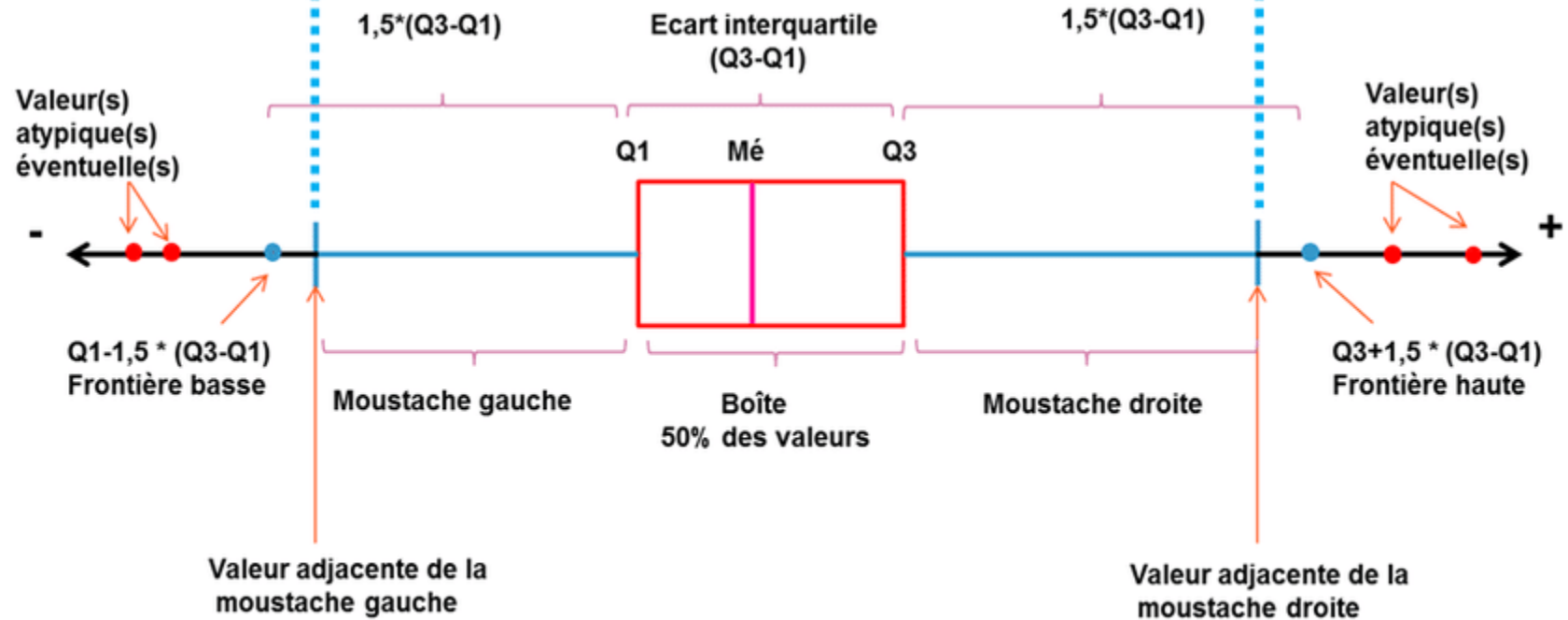
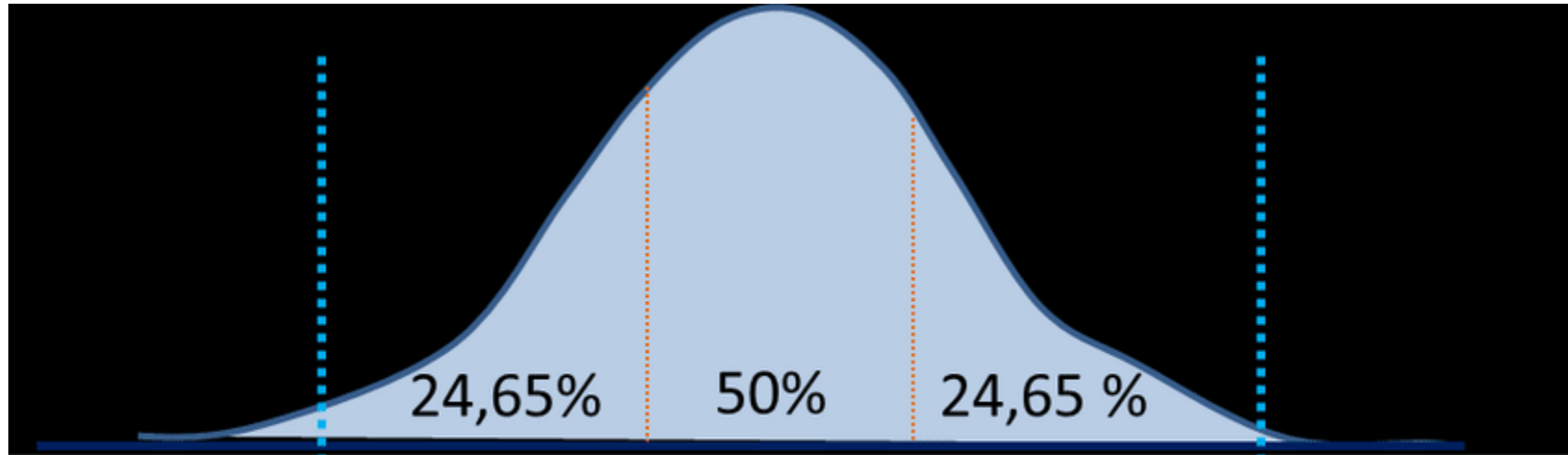


Différentes parties de la boîte à moustaches :

- La ligne sombre au milieu des boîtes est la médiane du salaire. La moitié des observations/lignes a une valeur supérieure à la médiane, et la moitié a une valeur inférieure.
- Comme la moyenne, la médiane est une mesure de la tendance centrale.
- Contrairement à la moyenne, elle est moins influencée par les observations/lignes avec des valeurs extrêmes.
- Dans cet exemple, la médiane est inférieure à la moyenne.
- Le fait que la moyenne et la médiane soient différentes indique qu'il existe quelques observations/lignes avec des valeurs extrêmes qui élèvent la moyenne. **Autrement dit, il existe quelques employés qui gagnent des salaires élevés.**

- Le bas de la boîte indique le 25e percentile. Vingt-cinq pour cent des observations/lignes ont des valeurs au-dessous du 25e percentile.
- Le haut de la boîte indique le 75e percentile. Vingt-cinq pour cent des observations/lignes ont des valeurs au-dessus du 75e percentile.
- Cela signifie que 50 % des observations/lignes sont situées dans la boîte.
- La boîte est beaucoup plus petite pour les femmes que pour les hommes. Ceci indique que le salaire varie moins pour les femmes que pour les hommes.
- Le haut et le bas de la boîte sont souvent appelés charnières.
- Les barres en T qui partent des boîtes sont appelées limites internes ou moustaches. Elles s'étendent jusqu'à 1,5 fois la hauteur de la zone.
- Si les données sont distribuées normalement, environ 95 % des données doivent être situées entre les limites internes.

- Dans cet exemple, les limites internes sont moins étendues pour les femmes que pour les hommes, ce qui constitue une autre indication que le salaire varie moins pour les femmes que pour les hommes.
- Les points sont des valeurs extrêmes. Il s'agit de valeurs qui n'entrent pas dans les limites internes.
- Les valeurs extrêmes sont des valeurs aberrantes. Les astérisques ou les étoiles sont des valeurs extrêmes et éloignées.
- Elles représentent des observations/lignes qui ont des valeurs égales à plus de trois fois la hauteur des boîtes.
- Il existe plusieurs valeurs extrêmes pour les femmes et les hommes.
- Il faut noter que la moyenne est supérieure à la médiane. La moyenne est plus élevée du fait de ces valeurs extrêmes.



1.4. Tableaux croisés

- La technique des tableaux croisés est une technique de base qui permet d'examiner la relation entre deux variables catégorielles. Par exemple, en utilisant la variable Tranche d'âge comme variable de ligne et la variable Sexe comme variable de colonne, vous pouvez créer un tableau croisé bidimensionnel présentant le nombre d'hommes et de femmes dans chaque tranche d'âge.
 1. Ouvrez le générateur de tableaux (menu Analyse, Tableaux, Tableaux personnalisés).
 2. Cliquez sur Restaurer pour supprimer les sélections précédentes dans le générateur de tableaux.
 3. Dans le générateur de tableaux, faites glisser la variable Tranche d'âge de la liste des variables vers la zone Lignes du panneau de canevas.
 4. Faites glisser la variable Sexe de la liste des variables vers la zone Colonnes du panneau de canevas. (Vous pouvez être amené à parcourir la liste pour accéder à cette variable.)
 5. Cliquez sur OK pour créer le tableau.

Fichier Édition Affichage Données Transformer Analyser Marketing direct Graphiques Utilitaires Fenêtre Aide

Rapports

Statistiques descriptives

Tableaux

Comparer les moyennes

Modèle linéaire général

Modèles linéaires généralisés

Modèles Mixtes

Corrélation

Régression

Log Linéaire

Réseaux neuronaux

Classification

Réduction des dimensions

Echelle

Tests non paramétriques

Prévisions

Survie

Réponses multiples

Analyse des valeurs manquantes

Imputation multiple

Echantillons complexes

Simulation...

Contrôle de qualité

Courbe ROC...

Align

Mesure

Rôle

Droite Echelle Entrée

Gauche Nominale Entrée

Droite Echelle Entrée

Droite Ordinales Entrée

Droite Nominale Entrée

Droite Echelle Entrée

Droite Echelle Entrée

Droite Echelle Entrée

Droite Echelle Entrée

Droite Echelle Entrée

Droite Echelle Entrée

Droite Nominale Entrée

un 8 Droite Echelle Entrée

8 Droite Nominale Entrée

Fréquences

Descriptives

Explorer

Tableaux croisés

Analyse TURF

Rapport

Tracés P-P...

Tracé Q-Q...

Activer Windows

Accédez aux paramètres pour activer Windows.

Vue de données

Vue des variables

Tableaux croisés

Le processeur IBM SPSS Statistics est prêt

Unicode:OFF

recidivism_reduit.sav [Jeu_de_donnees6] - IBM SPSS Statistics Editeur de données

Fichier Edition Affichage Données Transformer Analyse Marketing direct Graphiques Utilitaires Fenêtre Aide

3 : arrest2 1 Visible : 16 variables sur 16

	age	agecat	marital	social	ed	employ	gender	crime1	violent1	bail	rehab	arrest2	time	crime2	violent2	convict2	var	var	var	var
1	22	2	0	1	1	1	0	1	0	1	0	1	74,00	4	0	0				
2	19	1	0	1	2	0	1	2	0	0	1	1	239,00	3	0	1				
3	28	2	0	3	1	0	0	1	0	1	1	1	208,00	1	0	0				
4	36	3	1	3	2	0	0	0	0	0	0	0								
5	27	2	1	2	2	0	0	0	0	0	0	0								
6	29	2	1	2	2	0	0	0	0	0	0	0								
7	24	2	0	1	2	0	0	0	0	0	0	0								
8	28	2	1	1	1	0	0	0	0	0	0	0								
9	36	3	1	1	1	0	0	0	0	0	0	0								
10	25	2	0	3	1	0	0	0	0	0	0	0								
11	40	4	0	1	2	0	0	0	0	0	0	0								
12	21	1	0	2	3	0	0	0	0	0	0	0								
13	18	1	1	2	1	0	0	0	0	0	0	0								
14	20	1	0	2	2	0	0	0	0	0	0	0								
15	41	4	0	2	2	0	0	0	0	0	0	0								
16	19	1	0	2	1	0	0	0	0	0	0	0								
17	20	1	1	1	2	0	0	0	0	0	0	0								
18	20	1	0	3	1	0	0	0	0	0	0	0								
19	39	3	0	2	1	0	0	0	0	0	0	0								
20	23	2	0	2	2	0	0	0	0	0	0	0								
21	47	4	1	2	3	0	0	0	0	0	0	0								
22	36	3	0	2	1	0	0	2	0	1	1	0	9999,00	-1	-1	-1				
23	28	2	0	3	2	0	0	2	0	1	1	0	9999,00	-1	-1	-1				
24	35	3	0	1	2	1	1	1	0	1	1	0	9999,00	-1	-1	-1				
25	24	2	0	1	1	0	1	4	0	1	0	1	708,00	3	1	1				
26	27	2	0	1	1	0	0	3	0	1	1	0	9999,00	-1	-1	-1				
27	34	3	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
28	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
29	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
30	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
31	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
32	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
33	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
34	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
35	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
36	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
37	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
38	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
39	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
40	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
41	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
42	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
43	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
44	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
45	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
46	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
47	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
48	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
49	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				
50	26	2	1	1	2	0	0	2	0	0	0	0	9999,00	-1	-1	-1				

Tableaux croisés

Ligne(s):
 Catégorie d'âge [agecat]
 Age en années [age]

Colonne(s):
 Sexe [gender]

Couche 1 de 1

Précédent Suivant

Afficher les variables de couche dans les couches du tableau

Afficher les graphiques à barres en cluster

Supprimer les tableaux

OK Coller Réinitialiser Annuler Aide

Activer Windows

Accédez aux paramètres pour activer Windows.

Le processeur IBM SPSS Statistics est prêt Unicode:OFF

Tableau croisé Age en années * Sexe

Effectif		Sexe		Total
		Homme	Femme	
Age en années	18	4	1	5
	19	4	2	6
	20	5	0	5
	21	4	1	5
	22	3	1	4
	23	3	0	3
	24	3	1	4
	25	3	0	3
	26	3	0	3
	27	3	0	3
	28	5	0	5
	29	4	0	4
	30	2	0	2
	32	3	1	4
	34	1	0	1
	35	1	1	2
	36	4	1	5
	37	3	0	3
	39	3	0	3
	40	1	1	2
	41	2	0	2
	42	1	0	1
	43	2	0	2
	44	1	0	1
	47	2	0	2
	Total	70	10	80