Introduction au traitement des données avec SPSS

1. Logiciels de traitement de données
2. Découverte de SPSS
3. Entrer les données à partir d'un questionnaire
4. Préparation des données
5. Représentations graphiques
6. Mesures descriptives

1. Logiciels de traitement de données

Quelques logiciels de traitement des données car ils sont nombreux :

- **Excel**
- **StatBox et Question**
- **Sphinx**
- **Minitab**
- **SAS** (Système d'Analyse Statistique)
- **SPSS (Statistical Package for the Social Science**

2. Découverte de SPSS

Qu'est ce que SPSS

SPSS signifie **Statistical Package for the Social Science**

Logiciel spécialisé de traitement statistique des données dont l'objectif est d'offrir un logiciel intégré pour réaliser la totalité des tests statistiques. Il comprend plusieurs modules :

- Système de base
- Modèles de régression (regression models)
- Modèles avancés (advanced models)
- Tableaux (tables)
- Tests exacts (exact tests)
- Catégories (categories)
- Tendances (trends)
- Autres modules spécialisés

2. Découverte de SPSS

Comment démarrer SPSS

Pour lancer SPSS, 2 méthodes peuvent être utilisée :

- Faites un **double clic sur l'icône SPSS** apparaissant sur le bureau ou ;
- Cliquez sur **Démarrer**, puis **Programmes** et *IBM SPSS Statistics*.

Types de fenêtre dans SPSS

Une session typique SPSS a toujours 3 fenêtres :

- L'éditeur de données/ Data Editor
- La fenêtre des résultats/ Viewer
- La fenêtre de syntaxe/ Syntax Editor

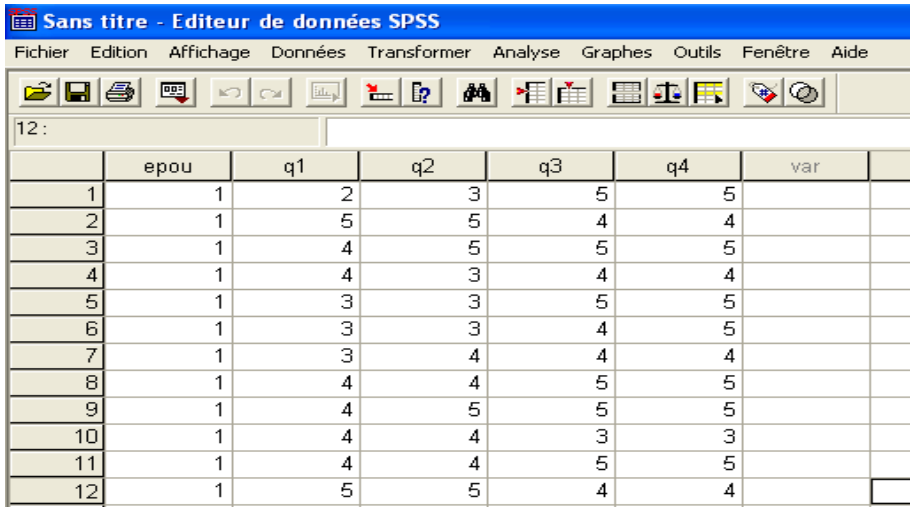
INTRODUCTION AU TRAITEMENT DES DONNÉES

2. Découverte de SPSS

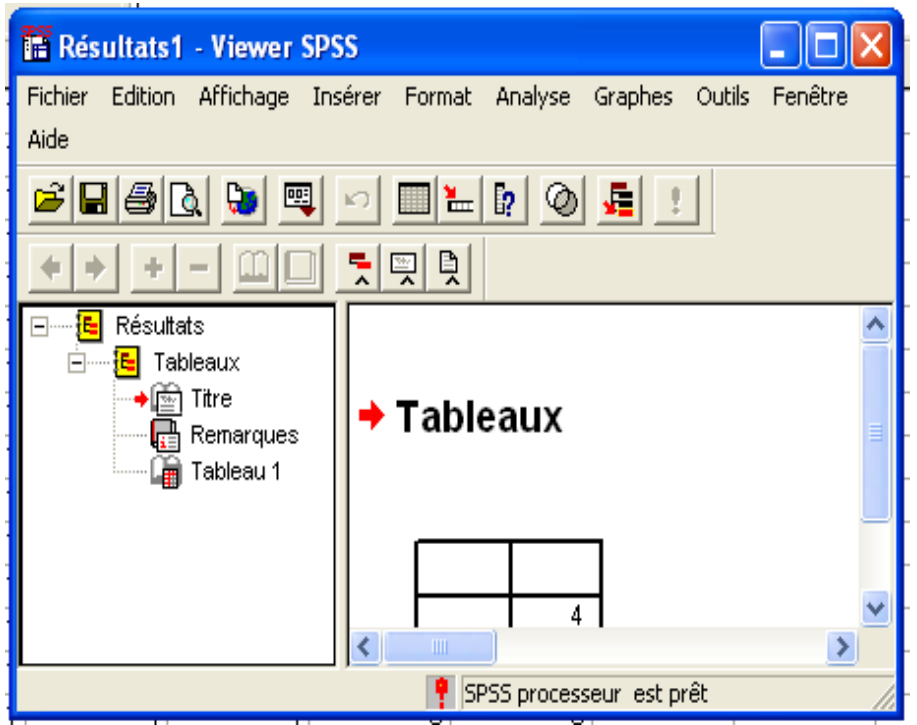
Types de fenêtre dans SPSS

- **L'éditeur de données**: cette fenêtre permet créer de nouveaux fichiers de données ou modifier des fichiers de données existants. ***Un fichier de données à l'extension «.sav.»***

- **La fenêtre des résultats/ Viewer:** s'ouvre automatiquement la 1ère fois que vous exécutez une procédure qui génère des résultats (tableaux et diagrammes) : ***c'est un fichier d'extension «.spo.»***



	epou	q1	q2	q3	q4	var
1	1	2	3	5	5	
2	1	5	5	4	4	
3	1	4	5	5	5	
4	1	4	3	4	4	
5	1	3	3	5	5	
6	1	3	3	4	5	
7	1	3	4	4	4	
8	1	4	4	5	5	
9	1	4	5	5	5	
10	1	4	4	3	3	
11	1	4	4	5	5	
12	1	5	5	4	4	

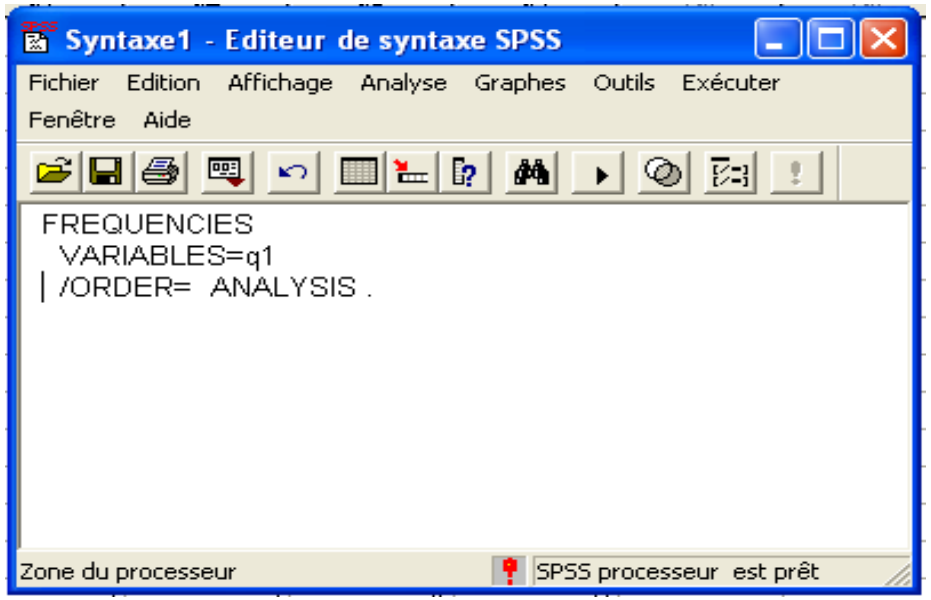


2. Découverte de SPSS

Types de fenêtre dans SPSS

La fenêtre de syntaxe: permet d'écrire les commandes d'analyses statistiques; *c'est un fichier* ***«.sps.»*.**

Lorsqu'une commande est complète, on peut l'exécuter en allant dans le menu "Run : Current" (ou encore en tapant Ctrl-R).



2. Découverte de SPSS

Barre des menus et Barre des boutons

La barre des boutons est uniquement un raccourci de la barre des menus

La barre des menus contient :

- ✓ **FICHIER/ FILE** : permet la gestion des fichiers (ex : ouvrir un nouveau fichier, fermer, enregistrer, etc.)
- ✓ **EDITION/ EDIT** : permet d'effectuer les opérations de traitement de texte (ex : copier, couper, coller, sélectionner, etc.)
- ✓ **AFFICHAGE/ VIEW** : permet de définir les options de l'écran (ex : barres d'outils)
- ✓ **DONNÉES/ DATA** : traite de tout ce qui est lié à la gestion de la barre de données (ex : définir ou insérer une variable, trier les données, etc.)ti

2. Découverte de SPSS

Barre des menus et Barre des boutons

La barre des menus contient *(suite)*:

- ✓ **TRANSFORMER / TRANSFORM** : présente les différentes opérations de transformation possibles sur les variables de la barre de données (ex : recodification, catégorisation, création d'indices, etc.)
- ✓ **ANALYSE/ ANALYZE :** permet d'accéder à toutes les analyses statistiques que SPSS rend possibles (ex : analyses descriptives, corrélations, etc.)
- ✓ **GRAPHES/ GRAPHS** : présente tous les types de graphiques que SPSS permet de créer (ex : histogrammes, boîtes à moustaches, courbes, etc.)
- ✓ **OUTILITAIRES/ UTILITIES** : comprend les utilitaires du programme (ex : informations sur les fichiers, informations sur les variables, etc.)
- ✓ **FENÊTRE/ WINDOWS** : permet la gestion des fenêtres
- ✓ **AIDE/ HELP** : propose des rubriques d'aide à l'utilisation de SPSS

3. Entrer les données à partir du questionnaire

Saisie des données

Avec SPSS, on peut ajouter les données de deux façons différentes:

- *1ère façon* : **saisir directement** dans l'écran ***AFFICHAGE/VIEW***
- *2ème façon* ***: importer les données*** d'un autre logiciel, par exemple Excel ou Access, etc.

Encoder le questionnaire

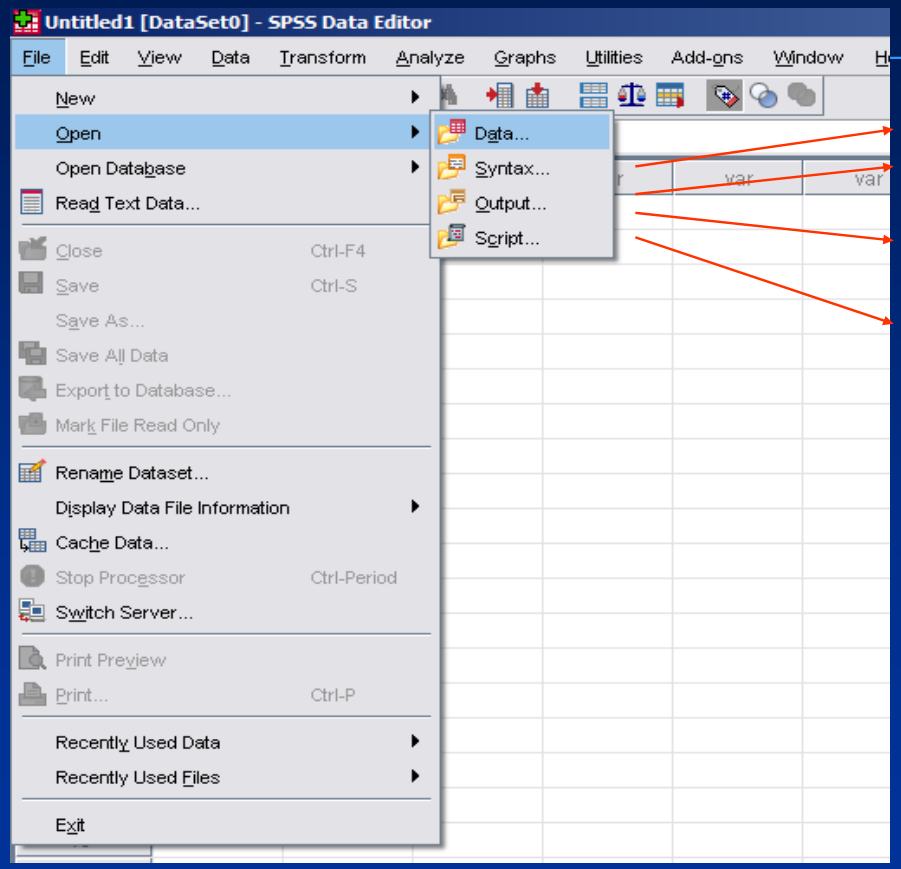
Il est recommandé de résumer les informations les plus importantes sur les variables rassemblées dans un « tableau de codage ». Ce tableau de codage à deux utilités à deux moments bien précis :

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

INTRODUCTION AU TRAITEMENT DES DONNÉES

Colonne	Nom SPSS	Nom de variable	Label	Values
Col1	V1	Date	Date de la mesure	
Col2	V2	Identifiant	Identifiant de l'athlète	
Col3	V3	Sexe	Sexe de l'athlète	1=Homme 2=Femme
Col4	V4	Vitamine	Vitamine prise par l'athlète	1=VitamineA 2=VitamineB 3=Vitamine C
Col5	V5	Absence	Nombre de jours de repos	
Col6	V6	Recup1	Nombre de seconde pour récupérer après le marathon 1	
Col7	V7	Recup2	Nombre de seconde pour récupérer après le marathon 2	
Col8	V8	Recup3	Nombre de seconde pour récupérer après le marathon 3	
Col9	V9	Arret1	Marathon 1 réalisé avec ou sans arrêt	1=Sans arrêt 2=Avec arrêts
Col10	V10	Arret2	Marathon 2 réalisé avec ou sans arrêt	0=Sans arrêt 1=Avec arrêts
Col11	V10	Fausse_Date	Date inventée	

INTRODUCTION AU TRAITEMENT DES DONNÉES

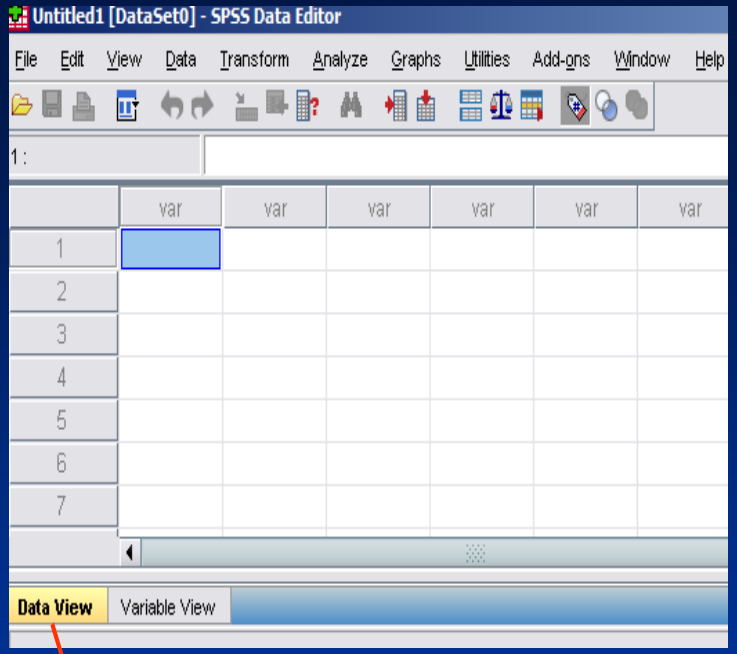


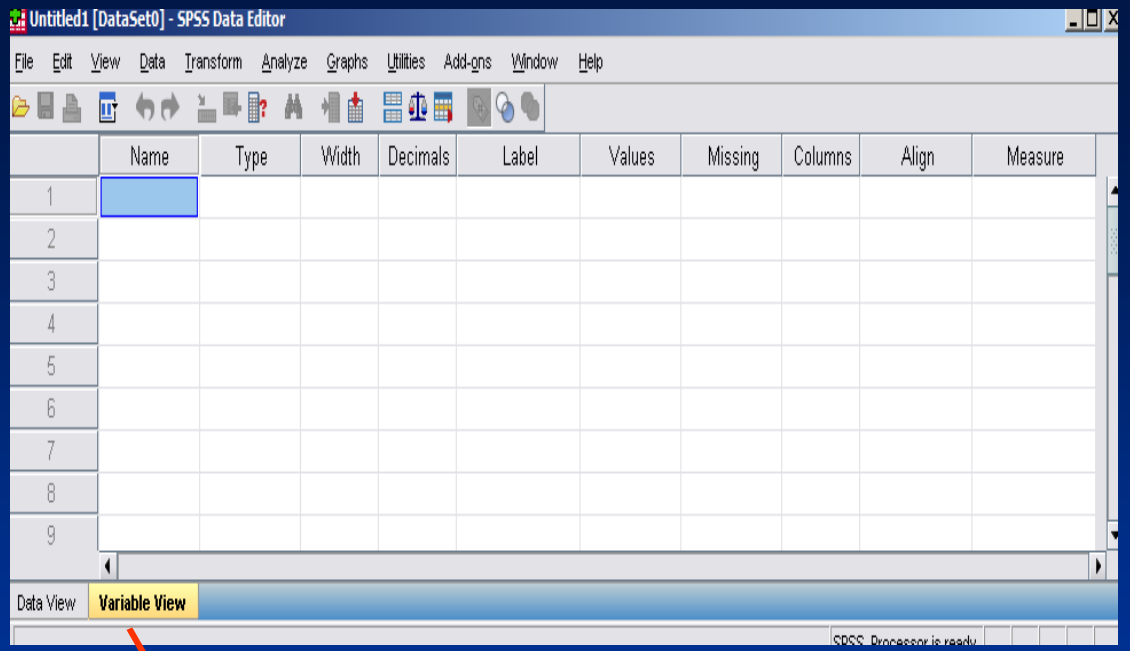
Plusieurs types de fichiers

- Data : Fichier de données
- Syntax : Fichier de syntaxe incluant le code de commandes SPSS
- Output : Fichier incluant les résultats des analyses
- Script : Fichier incluant du langage de programmation objet

=> Ces différents fichiers peuvent être sauvés et réutilisés par la suite

Le fichier de données

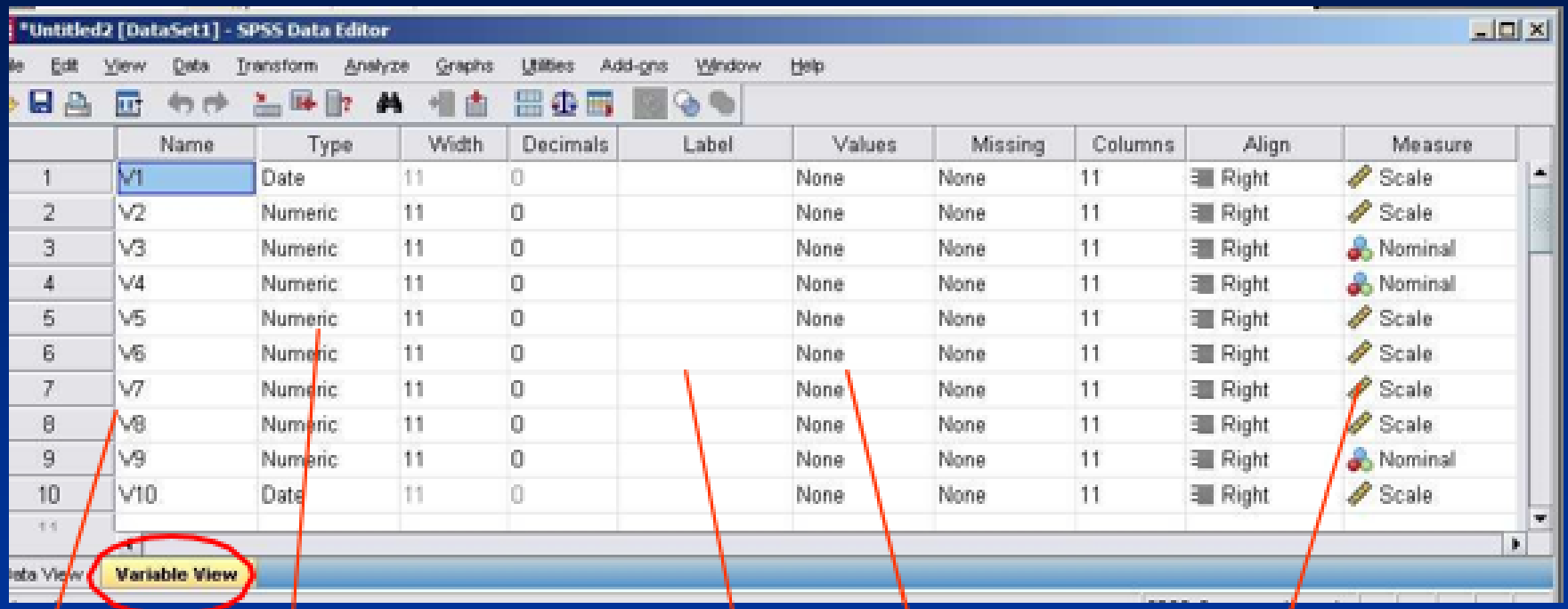




- Data View : Visualisation des données
  - ➢ permet de modifier les données

- Variable View : Visualisation des variables
  - ➢ permet de modifier les caractéristiques des variables

Définir les variables



Changer le **nom** des variables

Définir le **type** : Eviter les variables « string » (chaîne de caractères) car ça limite certaines analyses

Donner un **label** : nom complet des variables

Indiquer la signification de chaque **valeur**

Indiquer le type de **mesure** : échelle, ordinale, nominale

⇒ A vous d'essayer avec les informations reçues (aller voir ce qu'il est possible de faire dans chaque menu: changer l'affichage des dates, définir les missing…)

4. Préparation des données

Transformer les données

Le logiciel SPSS permet certaines procédures de transformation :

- **Créer une nouvelle variable à partir d'une formule de calcul**, faisant intervenir un ou plusieurs paramètres (calculer des scores d'échelle, des sous échelle ; centrer et réduire une variable, etc.)
- Changer la présentation des données d'une variable, en regroupant certaines valeurs d'une ou des variables cela s'appelle « **Recodage** »

D'autres procédures de transformation sont disponibles également sous SPSS

4. Préparation des données

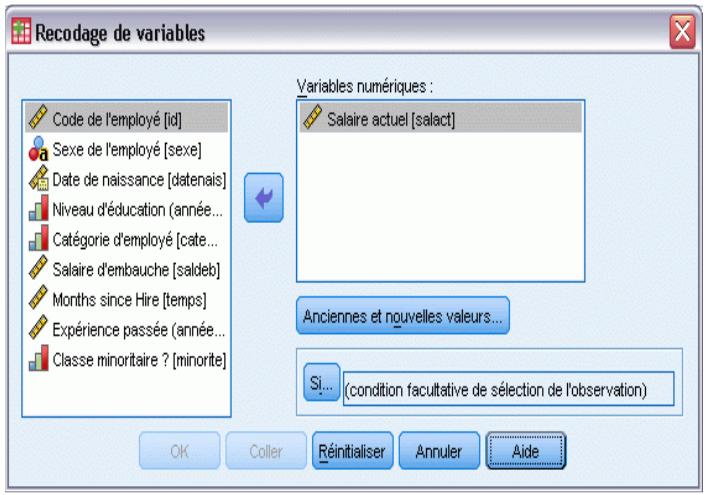
Recoder les variables

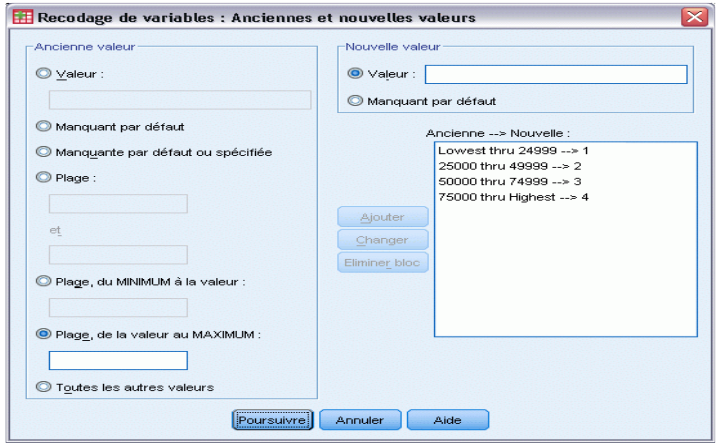
Pour recoder les valeurs d'une variable il faut :

✓ Sélectionner **Transformer > Recoder des variables > Dans une variable différente**

Sélectionner les variables que vous désirez recoder. *Si vous sélectionnez plusieurs variables, elles doivent être du même type (numérique ou alphanumérique)* ;

✓ Cliquer sur **Anciennes et nouvelles valeurs** et spécifier comment recoder les valeurs.





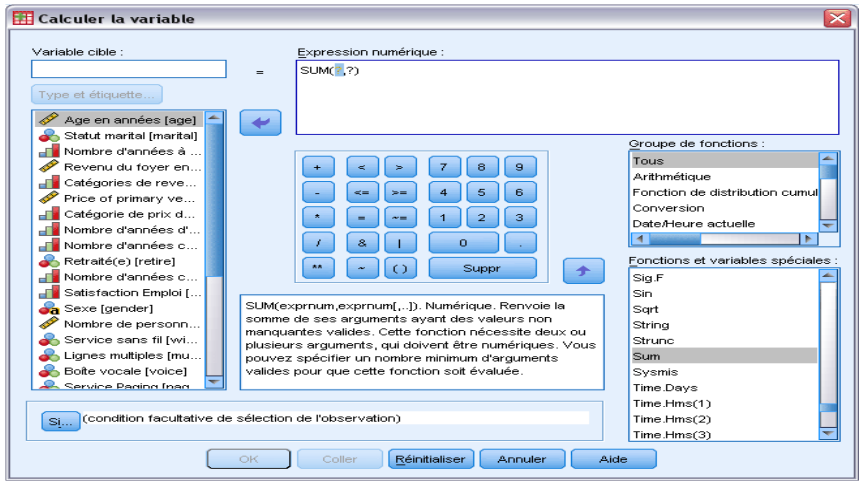
INTRODUCTION AU TRAITEMENT DES DONNÉES

4. Préparation des données

Construire les indicateurs

Pour construire une nouvelle variable à partir de plusieurs variables de départ :

Transformer > Calculer la variable



Transformer les données en utilisant la page syntaxe

Transformer > Calculer la variable et appuyer sur le bouton «**Coller**» au lieu de «**OK**».

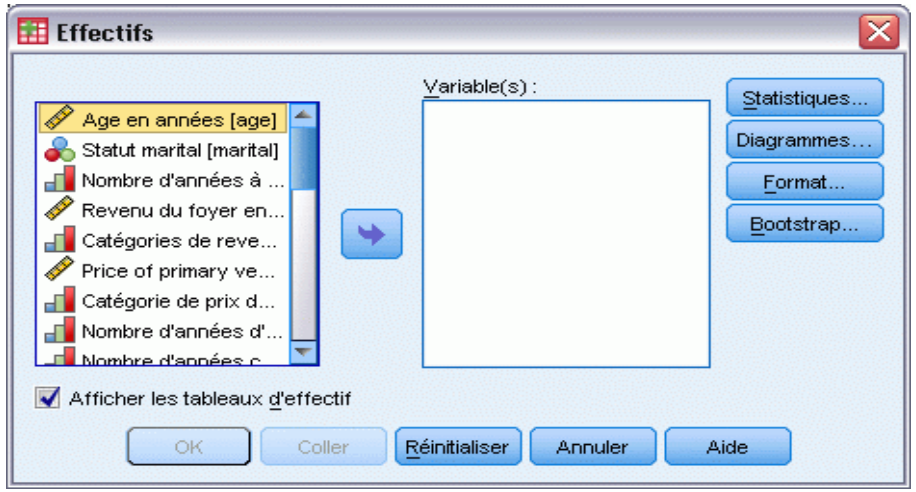
La commande exécutée s'inscrira dans la page de syntaxe.

Pour exécuter les commandes, on les sélectionne et on envoie la syntaxe en appuyant sur le bouton ▶

5. Représentations graphiques

Fréquence

Les tableaux de fréquences indiquent pour une variable donnée, toutes les valeurs prises par cette variable, le nombre de fois que chaque valeur apparaît et la proportion qu'elle représente par rapport à l'ensemble des autres valeurs de la variable.



Analyse > Statistiques descriptives > Effectifs

Choisissez les variables à analyser et faites-les glisser dans la liste **Variable(s)** à droite > **OK**

5. Représentations graphiques

Graphiques pour les variables nominales et ordinales (fréquences)

Sélectionner : **Graphes > Générateur de diagrammes > Galerie** (s'il n'est pas sélectionné) ;

- Cliquer sur le diagramme dont vous avez besoin et faites le glisser dans la zone étendue au-dessus de la galerie ;
- Renseigner les axes à partir des variables qui sont à gauche ; sélectionner et faire glisser dans le cadre de l'axe réservé.

Graphiques pour les variables métrique

Graphes > Générateur de diagrammes > Galerie (s'il n'est pas sélectionné) > cliquer et faire glisser **Histogramme** l'espace réservé > renseigner les axes

Ou Graphes > Boîtes de dialogue ancienne version > Histogramme

6. Mesures descriptives

Mesures descriptives

Analyse > Statistiques descriptives > Effectifs

- ✓ ***Afficher les tableaux d'effectifs*** **:** tableaux de distribution de fréquences
- ✓ *Le bouton* ***Statistiques*** : permet d'ajouter des statistiques de Fractiles, de Tendances centrale, de dispersion et de distribution
- ✓ *Le bouton* ***Diagrammes*** : permet d'ajouter un diagramme au tableau de fréquences

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!

6. Mesures descriptives

Comparer les groupes

L'on peut analyser séparément des sous-groupes de l'échantillon afin de les comparer.

Données > Scinder un fichier > Comparer les groupes

Les options :

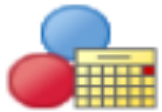
✓ ***« Comparer les groupes » : donne un tableau commun pour les sous-groupes.***

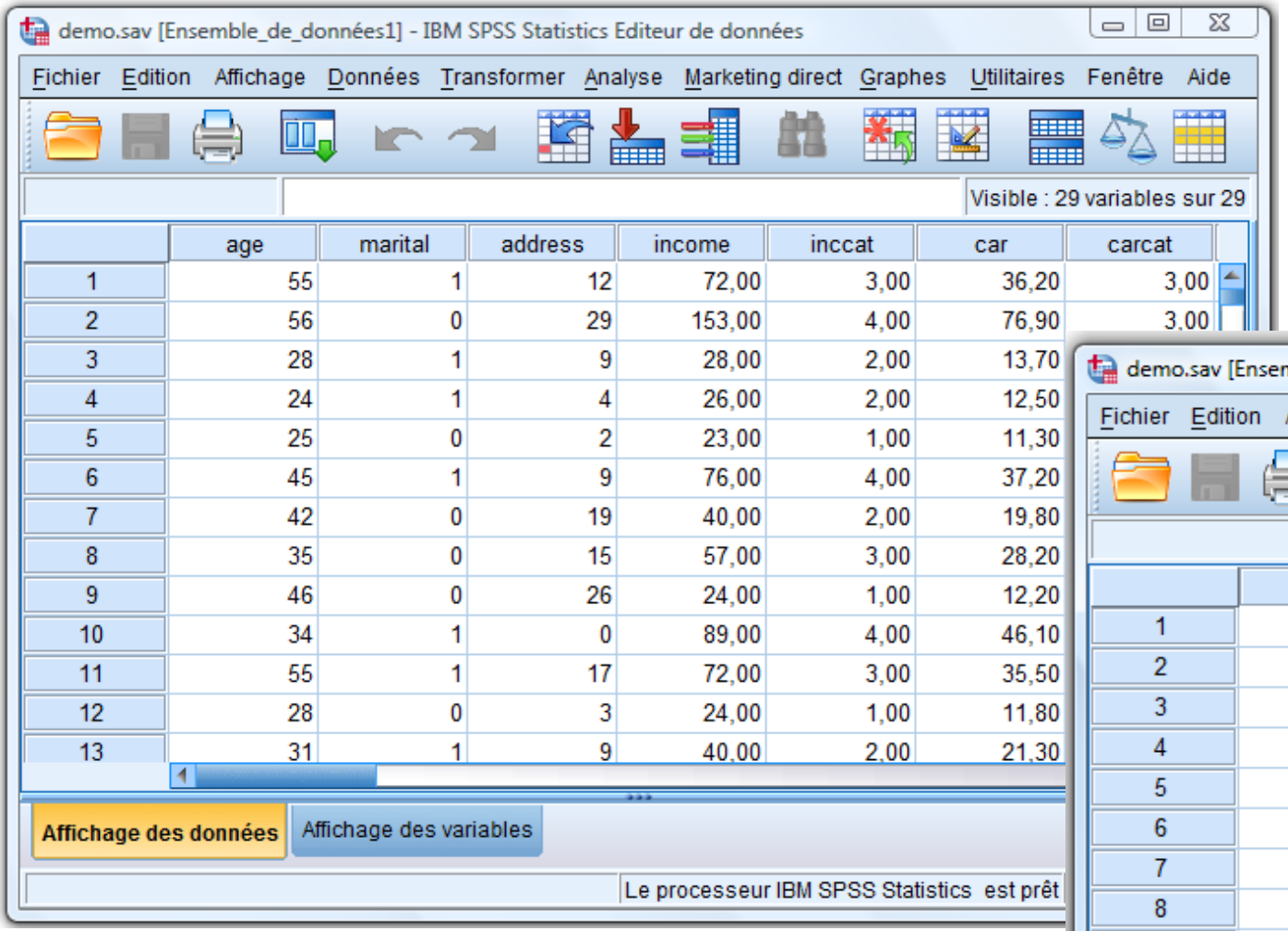
✓ ***« Séparer les résultats par groupe » : donne des tableaux séparés pour les sous-groupes.***

Pour la désactiver, **Données > Scinder un fichier > Analyser toutes les observations, ne pas créer de groupes**.

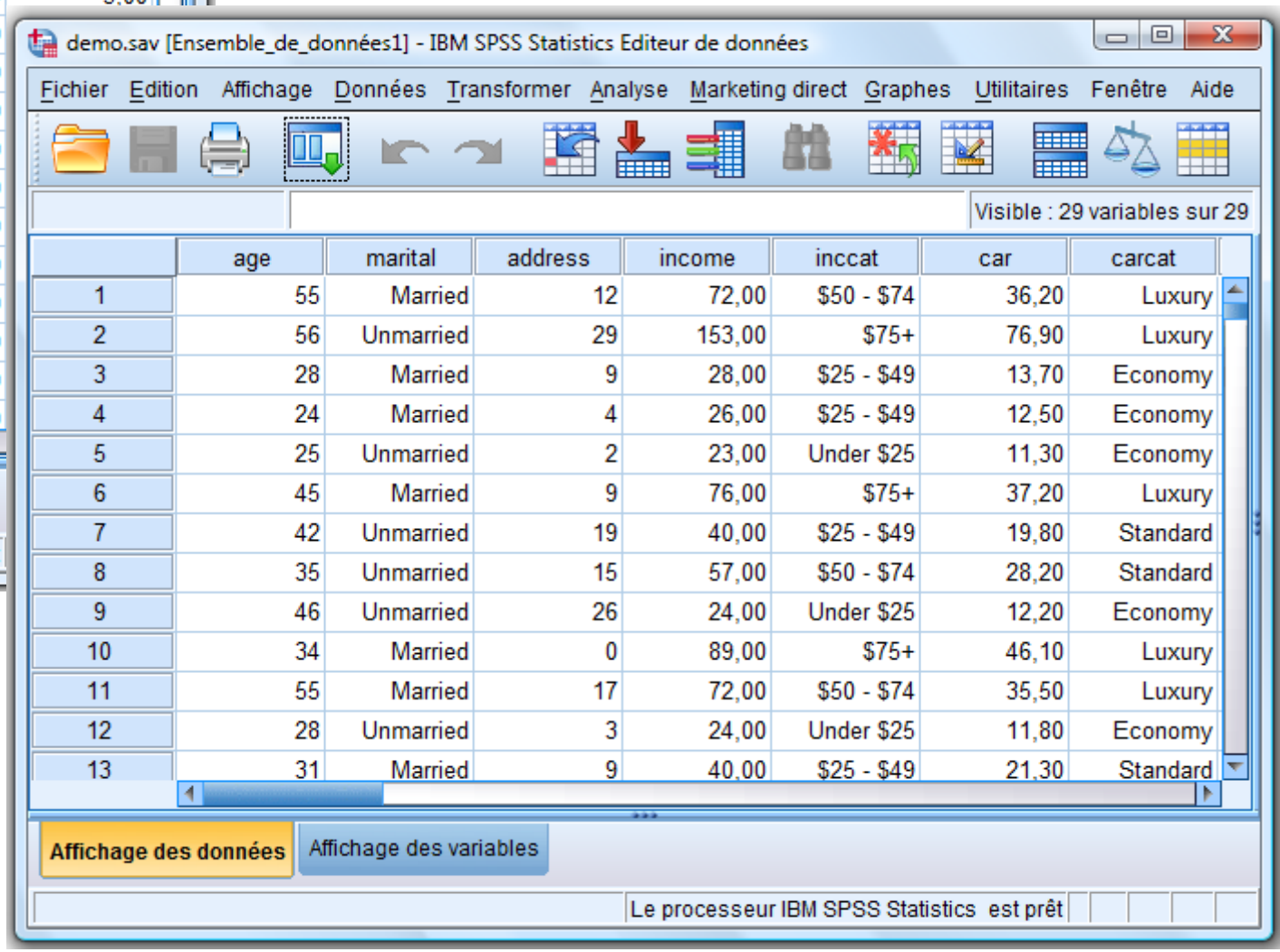
Type de variable :

- **Variable qualitative : valeurs non numériques, dites modalités, (profession)**
- **Variable dichotomique : variable qualitative ne** prenant que deux modalités (sexe)
- **Variable quantitative discrète : valeurs numériques** isolées (nombre d'enfants)
- **Variable quantitative continue : valeurs numériques**
  sur un intervalle continu (salaire, poids, durée)

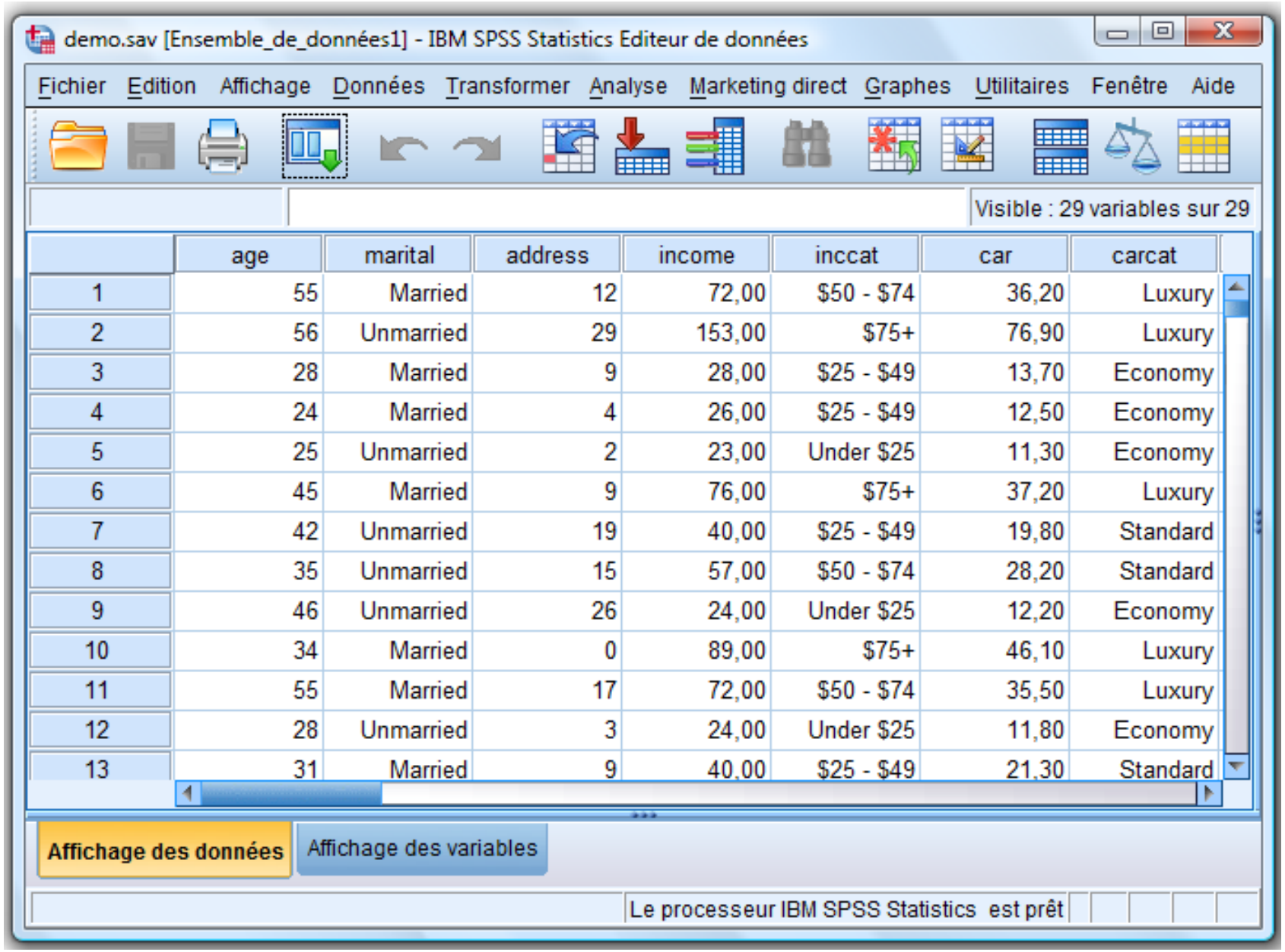
Niveau de mesure	Le type de données			
	Numérique	Chaîne	Date	Heure
Echelle (continue).		n/a		
Ordinales				
Nominales				



	age	marital	address	income	inccat	car	carcat
1	55	1	12	72,00	3,00	36,20	3,00
2	56	0	29	153,00	4,00	76,90	3,00
3	28	1	9	28,00	2,00	13,70	
4	24	1	4	26,00	2,00	12,50	
5	25	0	2	23,00	1,00	11,30	
6	45	1	9	76,00	4,00	37,20	
7	42	0	19	40,00	2,00	19,80	
8	35	0	15	57,00	3,00	28,20	
9	46	0	26	24,00	1,00	12,20	
10	34	1	0	89,00	4,00	46,10	
11	55	1	17	72,00	3,00	35,50	
12	28	0	3	24,00	1,00	11,80	
13	31	1	9	40,00	2,00	21,30	



	age	marital	address	income	inccat	car	carcat
1	55	Married	12	72,00	$50 - $74	36,20	Luxury
2	56	Unmarried	29	153,00	$75+	76,90	Luxury
3	28	Married	9	28,00	$25 - $49	13,70	Economy
4	24	Married	4	26,00	$25 - $49	12,50	Economy
5	25	Unmarried	2	23,00	Under $25	11,30	Economy
6	45	Married	9	76,00	$75+	37,20	Luxury
7	42	Unmarried	19	40,00	$25 - $49	19,80	Standard
8	35	Unmarried	15	57,00	$50 - $74	28,20	Standard
9	46	Unmarried	26	24,00	Under $25	12,20	Economy
10	34	Married	0	89,00	$75+	46,10	Luxury
11	55	Married	17	72,00	$50 - $74	35,50	Luxury
12	28	Unmarried	3	24,00	Under $25	11,80	Economy
13	31	Married	9	40,00	$25 - $49	21,30	Standard



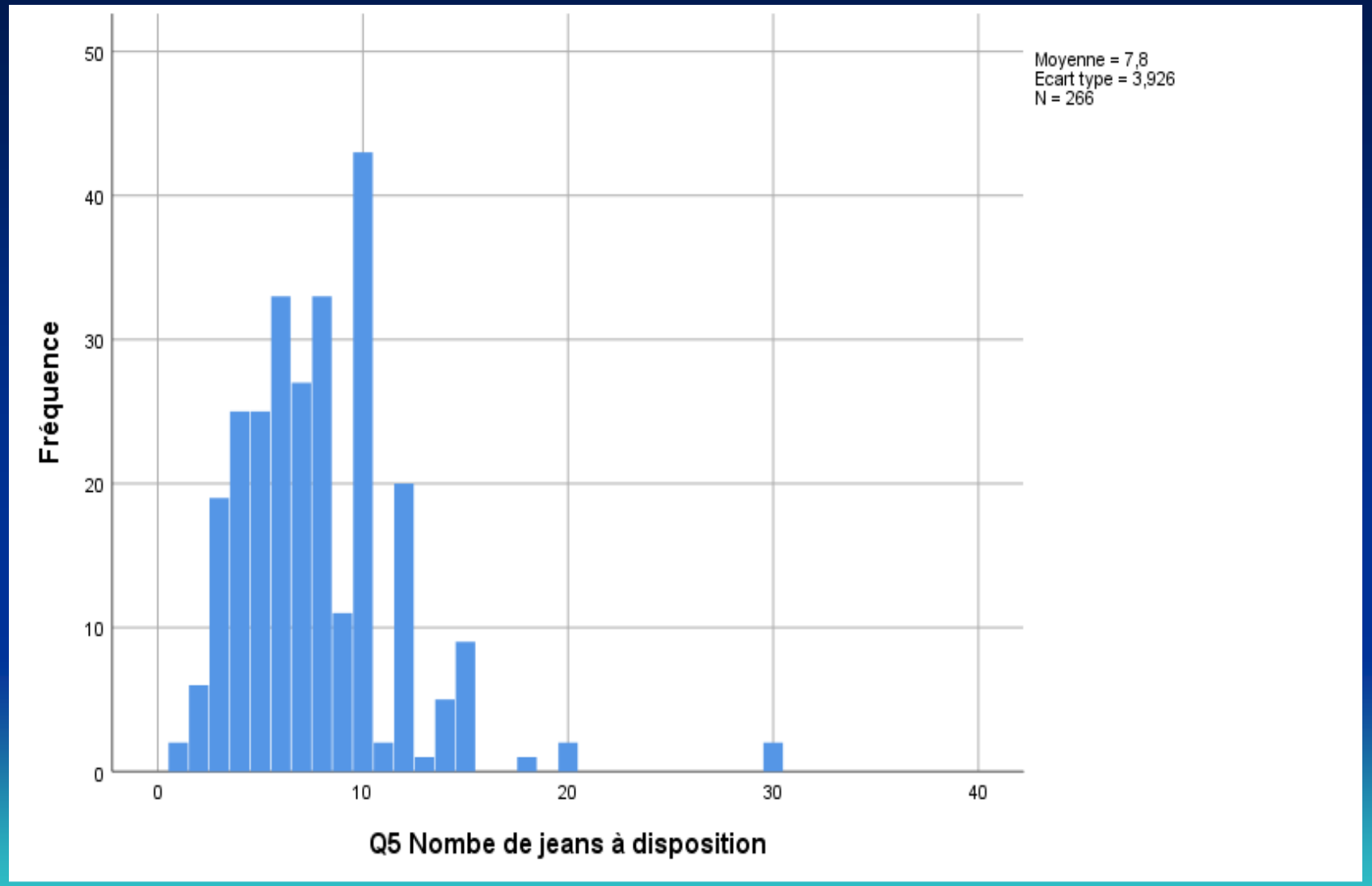
	age	marital	address	income	inccat	car	carcat
1	55	Married	12	72,00	$50 - $74	36,20	Luxury
2	56	Unmarried	29	153,00	$75+	76,90	Luxury
3	28	Married	9	28,00	$25 - $49	13,70	Economy
4	24	Married	4	26,00	$25 - $49	12,50	Economy
5	25	Unmarried	2	23,00	Under $25	11,30	Economy
6	45	Married	9	76,00	$75+	37,20	Luxury
7	42	Unmarried	19	40,00	$25 - $49	19,80	Standard
8	35	Unmarried	15	57,00	$50 - $74	28,20	Standard
9	46	Unmarried	26	24,00	Under $25	12,20	Economy
10	34	Married	0	89,00	$75+	46,10	Luxury
11	55	Married	17	72,00	$50 - $74	35,50	Luxury
12	28	Unmarried	3	24,00	Under $25	11,80	Economy
13	31	Married	9	40,00	$25 - $49	21,30	Standard

Analyse d’une variable Quantitative

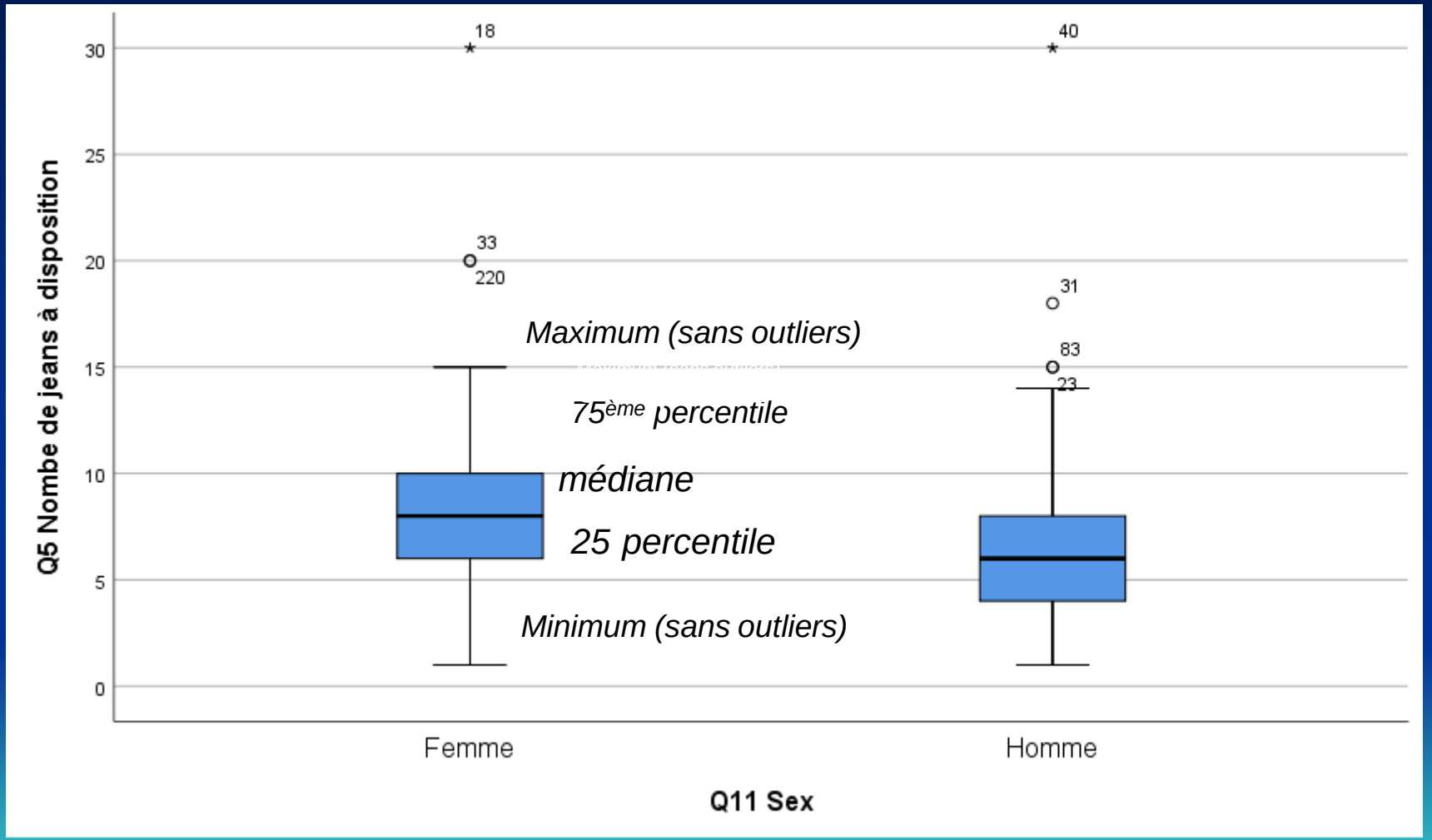
Visualisation graphique

Statistique descriptive

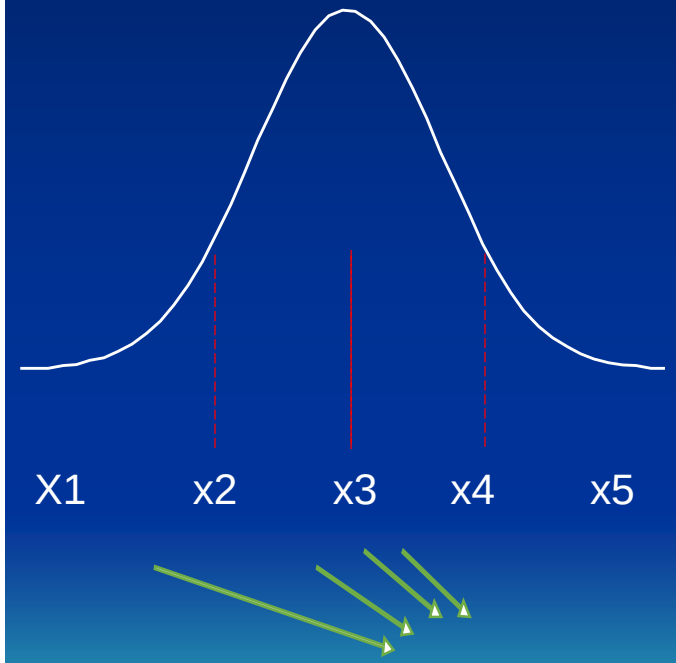
Histogramme

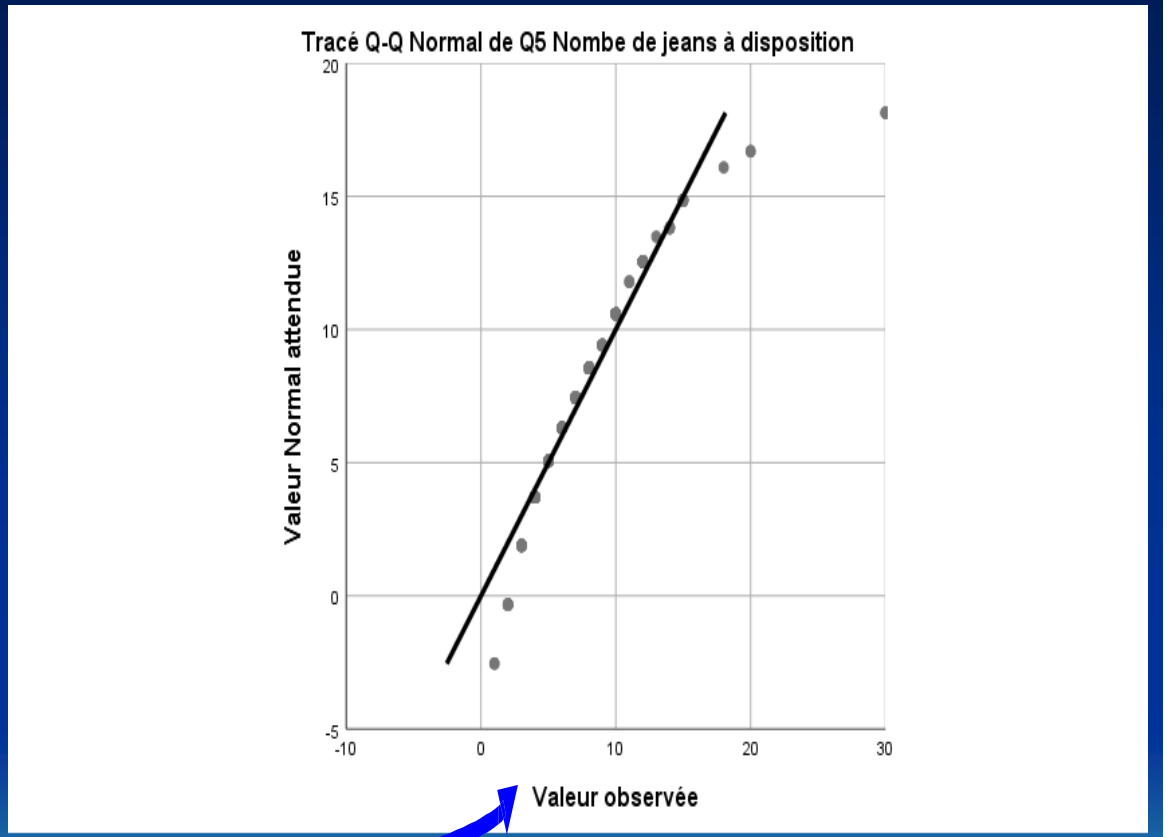


Boite à moustache

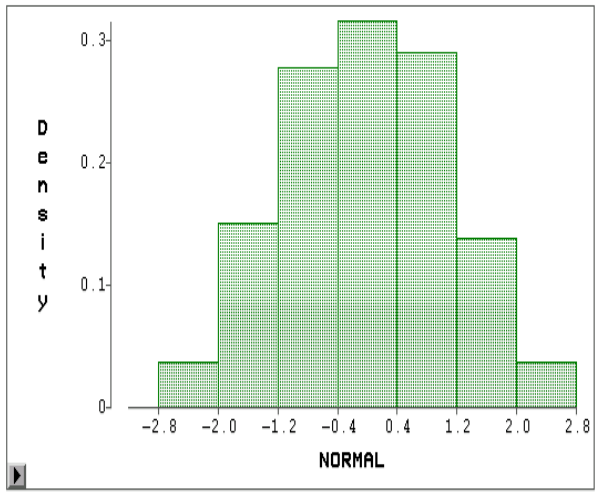


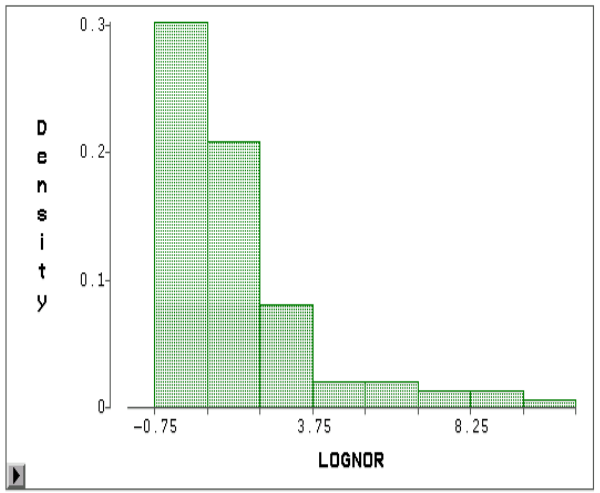
QQ-plot

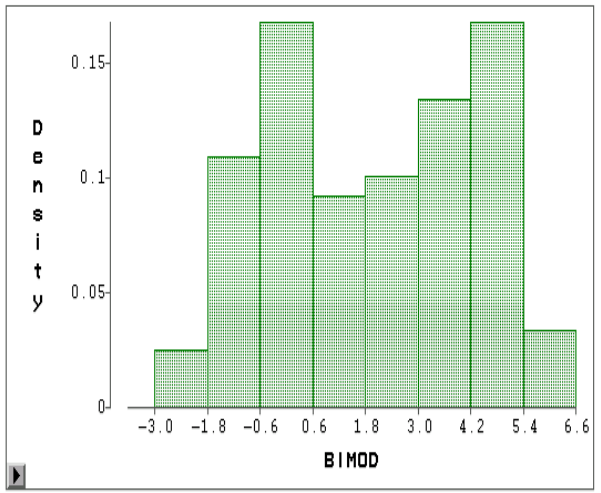

X1
x2
x3
x4
x5
observations

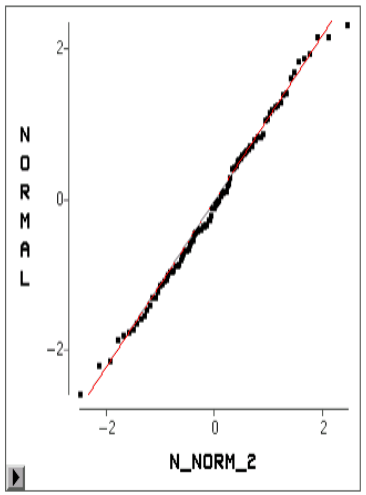

Tracé Q-Q Normal de Q5 Nombe de jeans à disposition
Valeur Normal attendue
Valeur observée
20
15
10
5
0
-5
-10
0
10
20
30

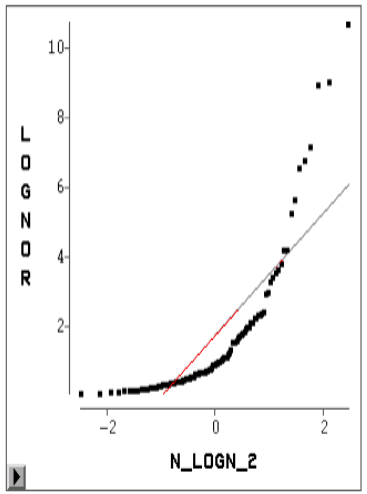
Quelques exemples de vérification de la normalité par QQplot

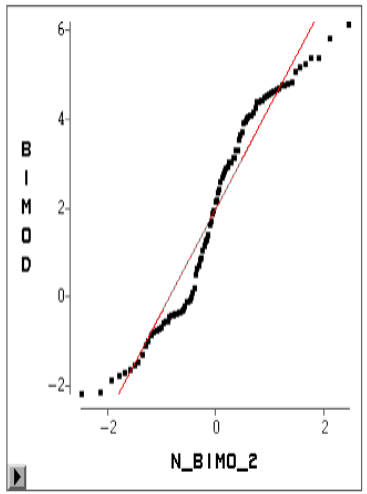












ENCGT
المدرسة الوطنية للتجارة والتسيير
ENCG TANGER

Notes

Histogramme

- *Graphiques → Boites de dialogues → Histogramme*

Boite à moustache

- *Graphiques → Boites de dialogues (→ Interactive)→*

ou Graphiques → Génrateur de graphiques → Boite à moustache

QQ-plot (Pour la normalité)

- *Analyse → Statistiques descriptives→ Q-Q*

Tableaux: Statistiques descriptives

SPSS : Analyse → Reports → Récapitulatif des observations

Analyse → Statistiques descriptives→ Explorer

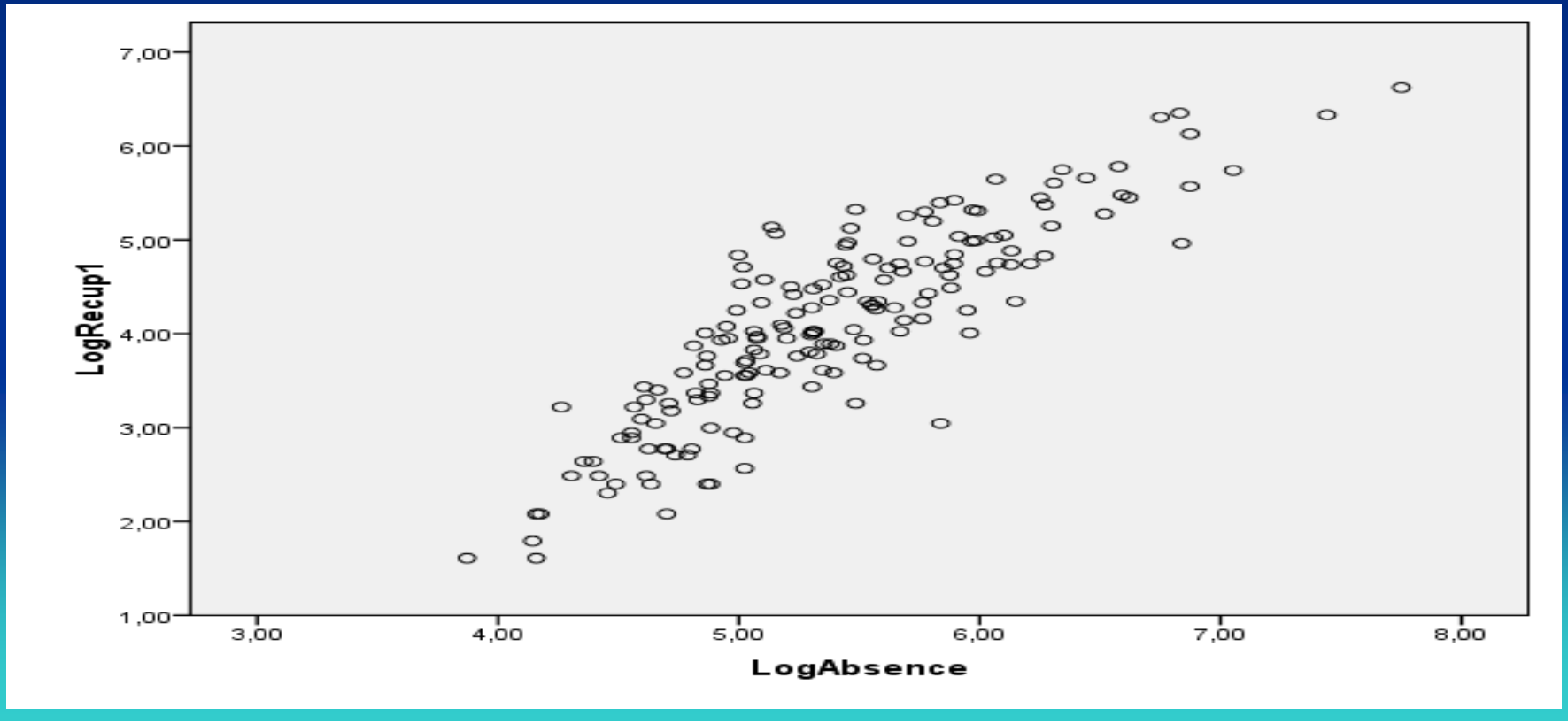
Analyse → Statistiques descriptives→ Fréquences

Analyse avec plusieurs variables quantitatives

Visualisation graphique

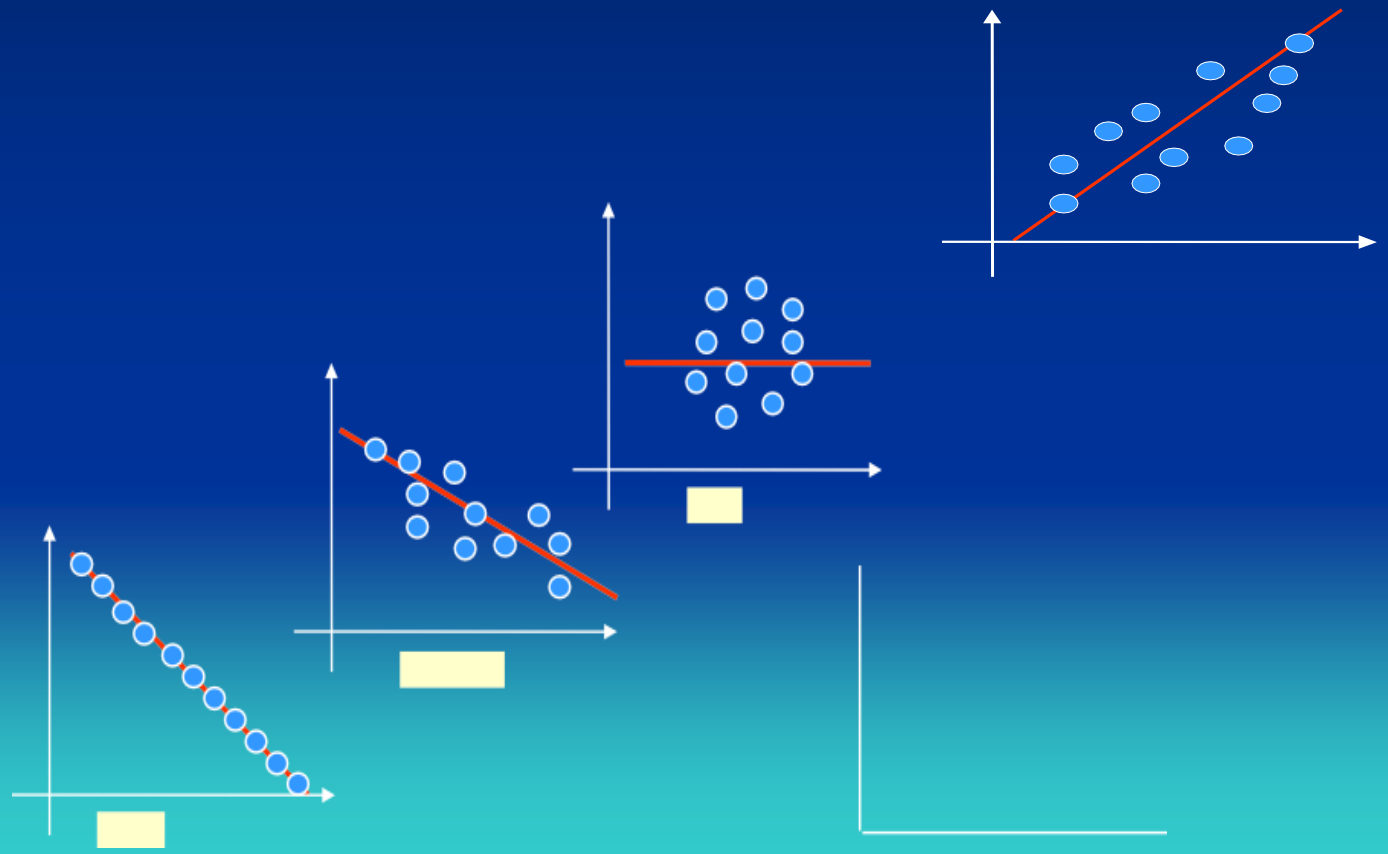
Visualiser le lien entre nombre de jean à disposition et prix.

Graphiques → Boites de dialogues → Dispersion



Statistique descriptive et inférence

Coefficient de corrélation de Pearson



Coefficient de corrélation et test d'hypothèse sur le coefficient

- Il existe plusieurs coefficients de corrélation dans SPSS :

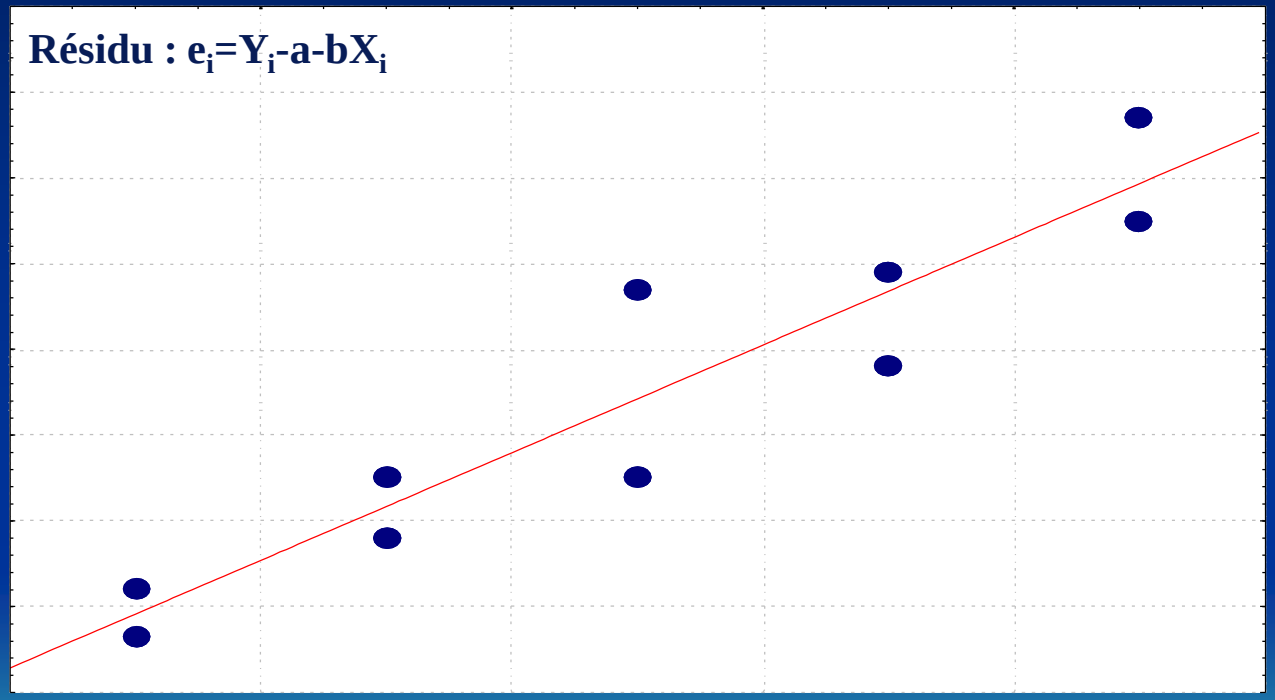
  - Pearson: utilisé quand on a deux variables continues
  - Spearman : quantitative non normales ou les variables qualitatives ordinales
  - Kendall tau-b (basé sur le nombre de concordances et discordances des rangs) : pour des variables ordinales

- Il existe un test d'hypothèse pour tester si le coefficient est égal versus différent de 0 (= versus > 0):

  H_0: $\rho=0$ contre H_1: $\rho\neq0$

Modélisation : Régression linéaire

La régression linéaire simple : $Y = \alpha + \beta X + \varepsilon$



Modélisation : Régression linéaire

- Comment juger si le modèle est valide ?

  → En analysant les résidus et les points influents

  Les hypothèses suivantes doivent toujours être vérifiées

$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

Modèle linéaire

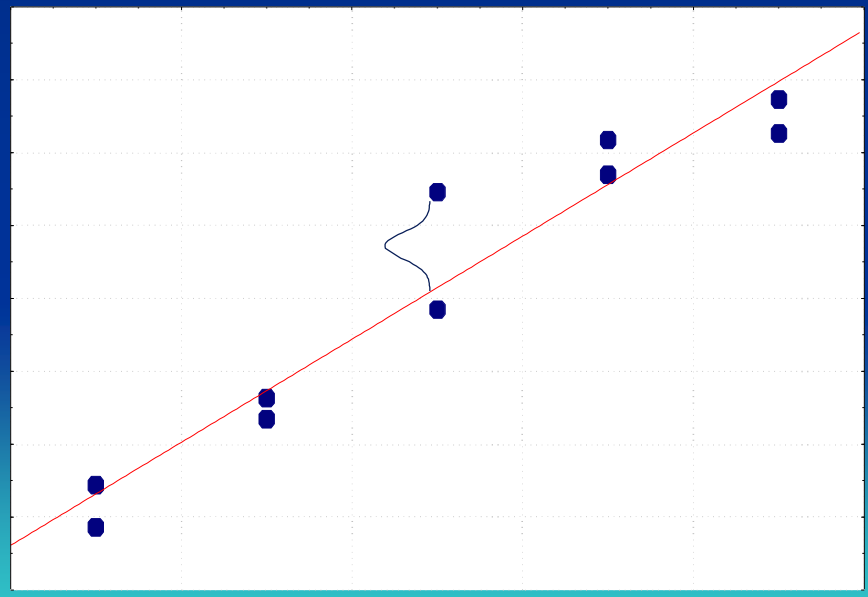
Termes d'erreur

$\varepsilon_i \sim iN(0,\sigma^2)$

Indépendance observations

Normalité de la distribution

Homogénéité des variances



L’ Analyse des données (ADD)

Rappel:

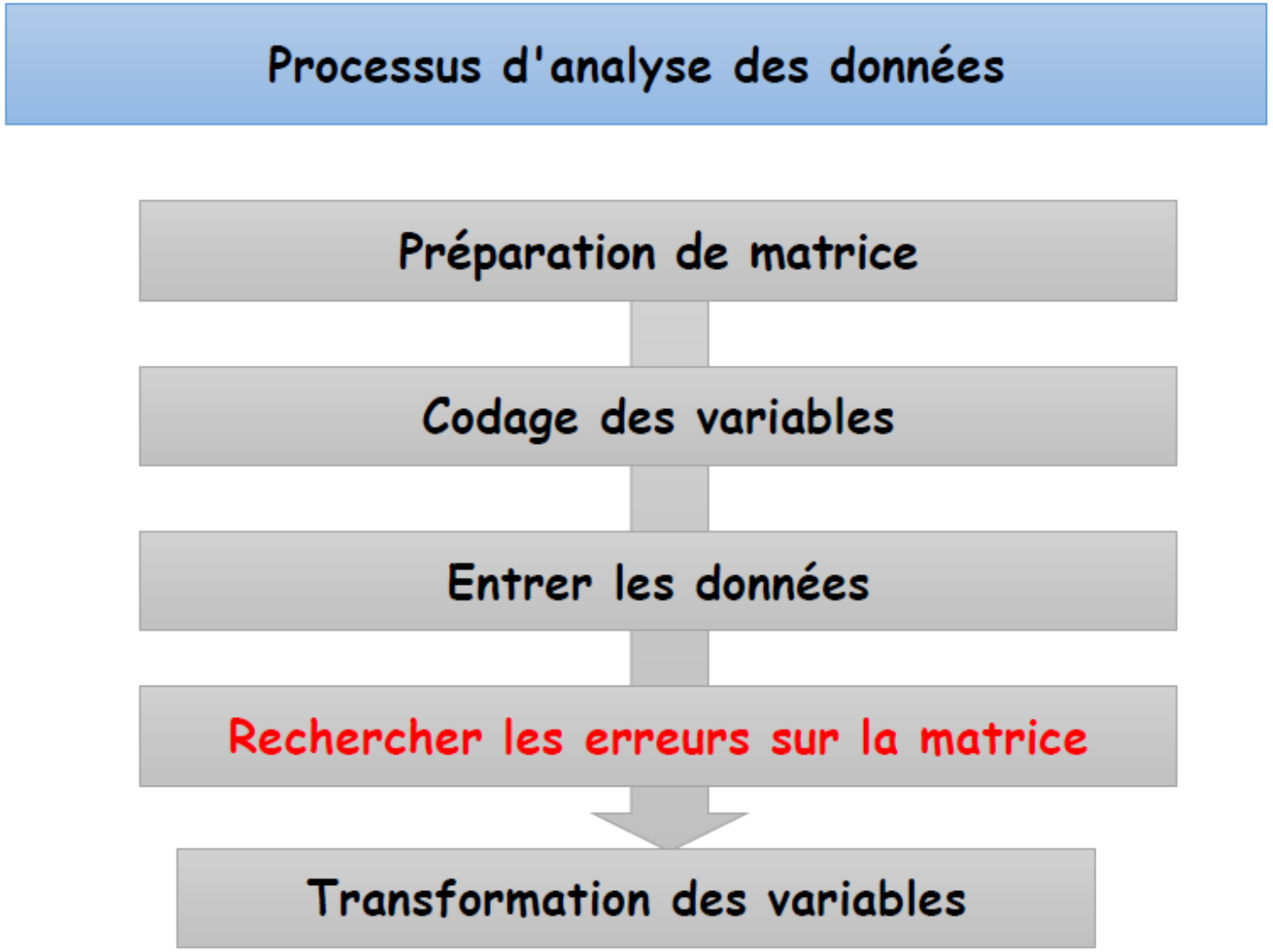
L’ Analyse des données (ADD) : l’ensemble de méthodes **descriptives** ayant pour objectif de **résumer et visualiser l’information** contenue dans un grand tableau de données

Objectifs:

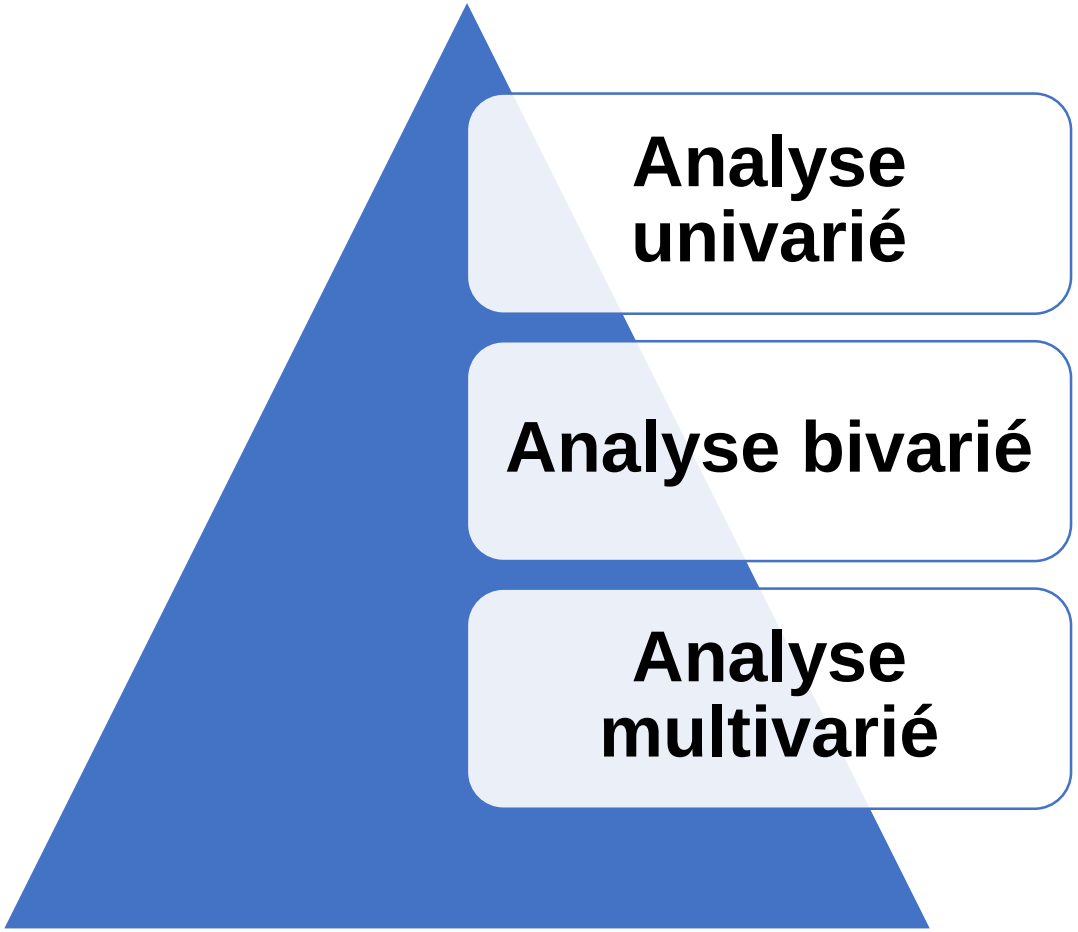
- Répondre aux problèmes posés par des tableaux de grandes dimensions
- Résumer les informations contenues dans un grand tableau sous forme d’une matrice
- Organiser et visualiser les informations

Outils : SPSS, EVIEWS..

Rappel


Processus d'analyse des données
Préparation de matrice
Codage des variables
Entrer les données
Rechercher les erreurs sur la matrice
Transformation des variables

Les types d’analyse des données



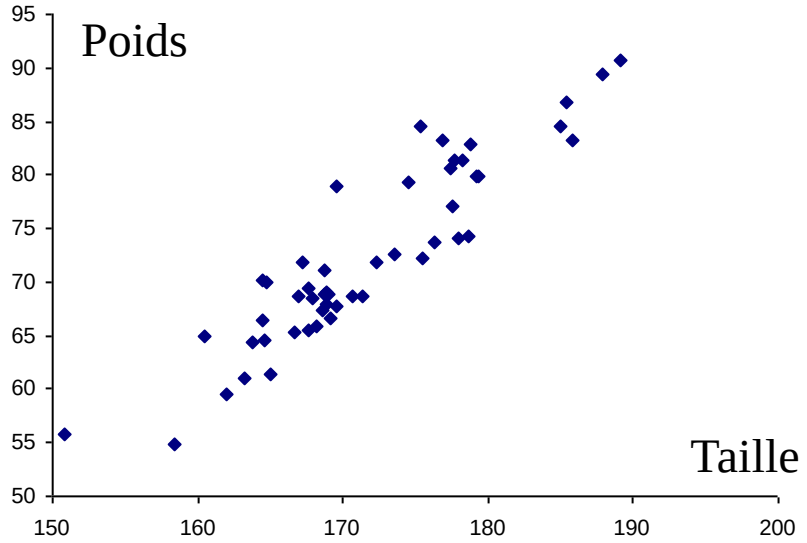
Les **tests statistiques** dans l'analyse des données les plus couramment utilisés :

1. Analyse descriptive (les paramètres de tendance et les paramètres de dispersion).
2. Test de corrélation
3. Analyse de régression (régression linéaire simple et multiple)
4. Test de khi II
5. ANOVA
6. ACP
7. ……

ETUDE DE 2 VARIABLES QUANTITATIVES

(1) MESURE DE LA LIAISON ENTRE 2 VARIABLES QUANTITATIVES

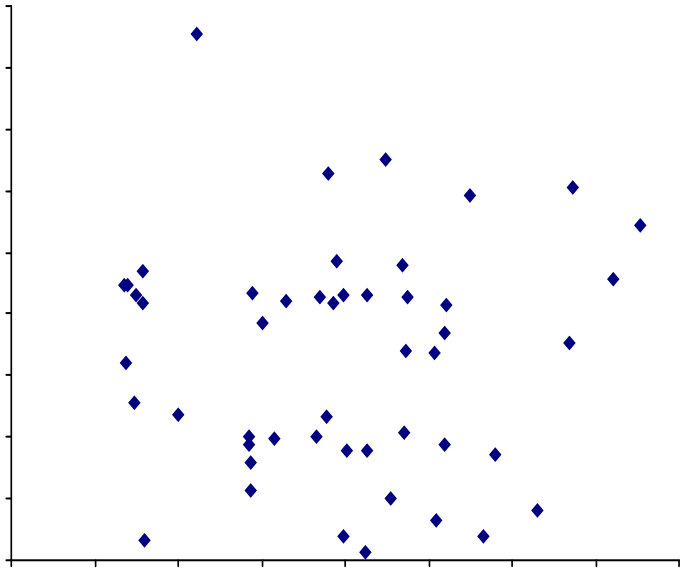
Nom	Taille x_i (cm)	Poids y_i (kg)
Pierre	175	73
Arantxa	168	56
…..	…..	…..
Martin	185	87



La connaissance de la taille x apporte une certaine information sur le poids y

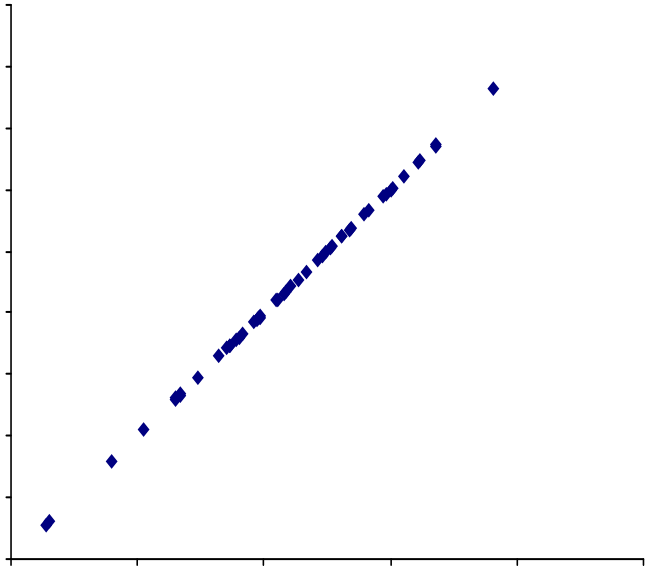
Il existe une relation de dépendance entre x et y

(2) MESURE DE LA LIAISON ENTRE 2 VARIABLES QUANTITATIVES



La connaissance de x n'apporte aucune certaine information sur y

x et y sont indépendantes



La connaissance de x permet de connaître exactement la valeur de y

Il existe une relation fonctionnelle entre x et y

(3) MESURE DE LA LIAISON ENTRE 2 VARIABLES QUANTITATIVES

Covariance : $$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Propriétés :

$Cov(x,y) > 0 \Leftrightarrow$ x et y varient dans le même sens

$Cov(x,y) < 0 \Leftrightarrow$ x et y varient en sens contraire

$Cov(x,y) = Cov(y,x)$

$Cov(x,x) = V(x)$

$Cov(a\,x + b\,y\,,\,z) = a\,Cov(x,z) + b\,Cov(y,z)$

(4) MESURE DE LA LIAISON ENTRE 2 VARIABLES QUANTITATIVES

Corrélation linéaire: $\rho = \dfrac{\mathrm{cov}(x,y)}{\sigma(x)\,\sigma(y)}$

Propriétés :

$-1 \le \rho \le 1$

$y = a\,x + b \Leftrightarrow \begin{cases} \rho = 1 & \text{si} \quad a > 0 \\ \rho = -1 & \text{si} \quad a < 0 \end{cases}$

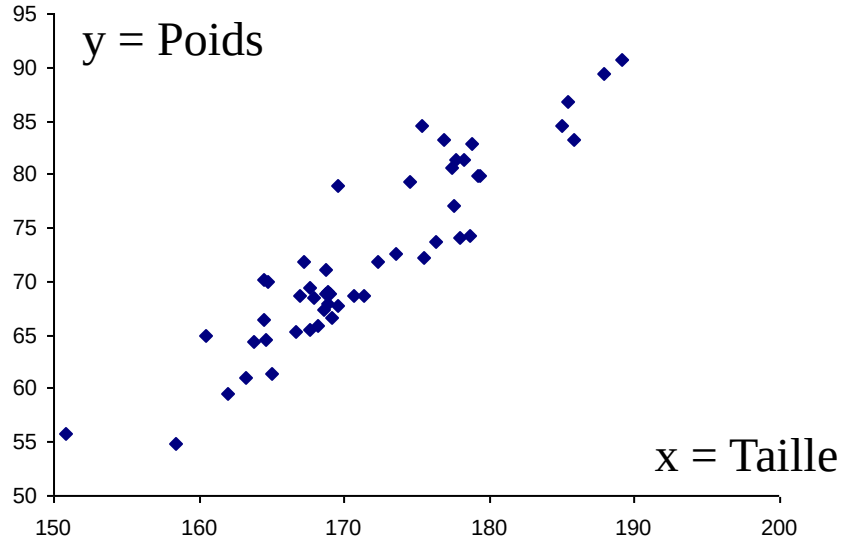
$|\rho| = 1 \Leftrightarrow$ Il existe une relation fonctionnelle entre x et y

$\rho = 0 \Leftrightarrow$ x et y sont indépendantes

$0 < |\rho| < 1 \Leftrightarrow$ Il existe une dépendance linéaire d'autant plus forte que $|\rho|$ est grand

Ne pas confondre causalité et corrélation

(1) AJUSTEMENT LINEAIRE

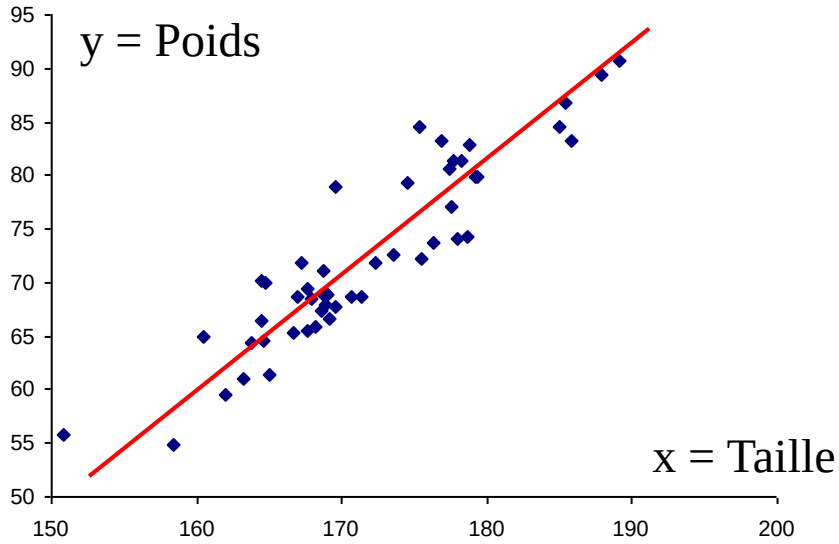


Est-il possible de trouver une fonction numérique f telle que y = f (x) ?

Si une telle fonction existe, on dit que f est un modèle du phénomène étudié.

x est la variable explicative.
y est la variable expliquée.

(2) AJUSTEMENT LINEAIRE



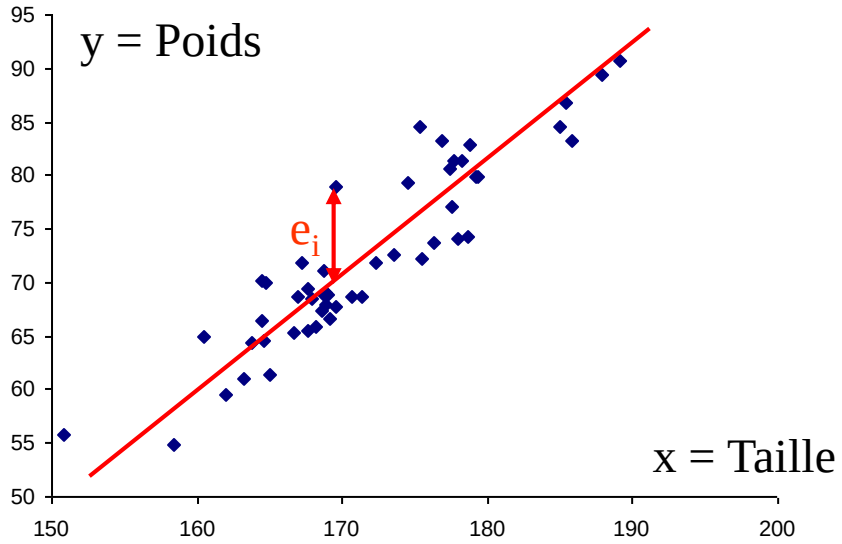
On désire trouver la droite qui passe « au mieux » à l'intérieur du nuage de points

(3) AJUSTEMENT LINEAIRE

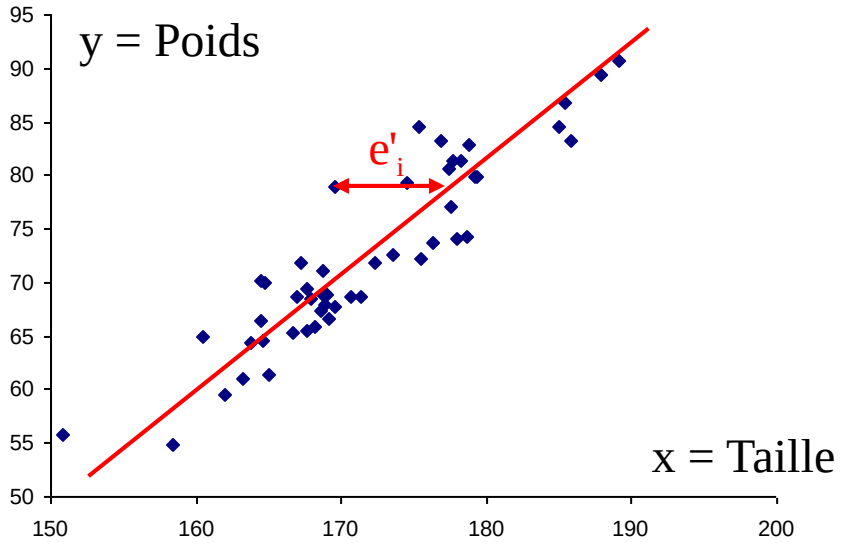
« au mieux »

Minimiser $S = \sum_{i=1}^{n} e_i^2$

Minimiser $S' = \sum_{i=1}^{n} e'^2_i$



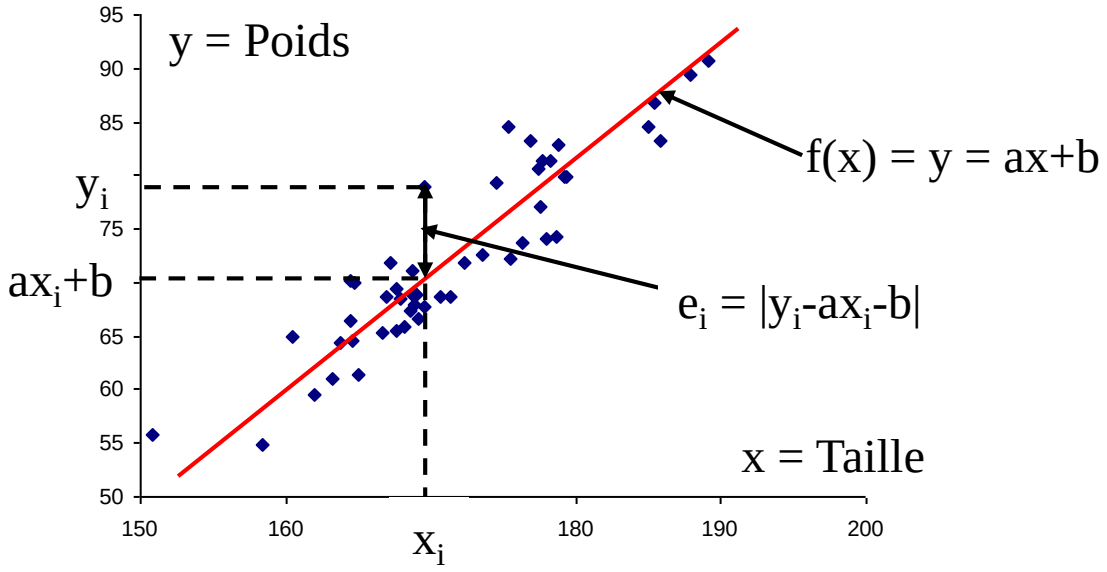
Droite de régression de y en x



Droite de régression de x en y

(4) AJUSTEMENT LINEAIRE
REGRESSION LINEAIRE DE Y EN X

Droite de régression linéaire de y en x
$y = f(x) = ax + b$



La droite de régression linéaire de y en x, notée $D_{y/x}$, minimise $S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2$

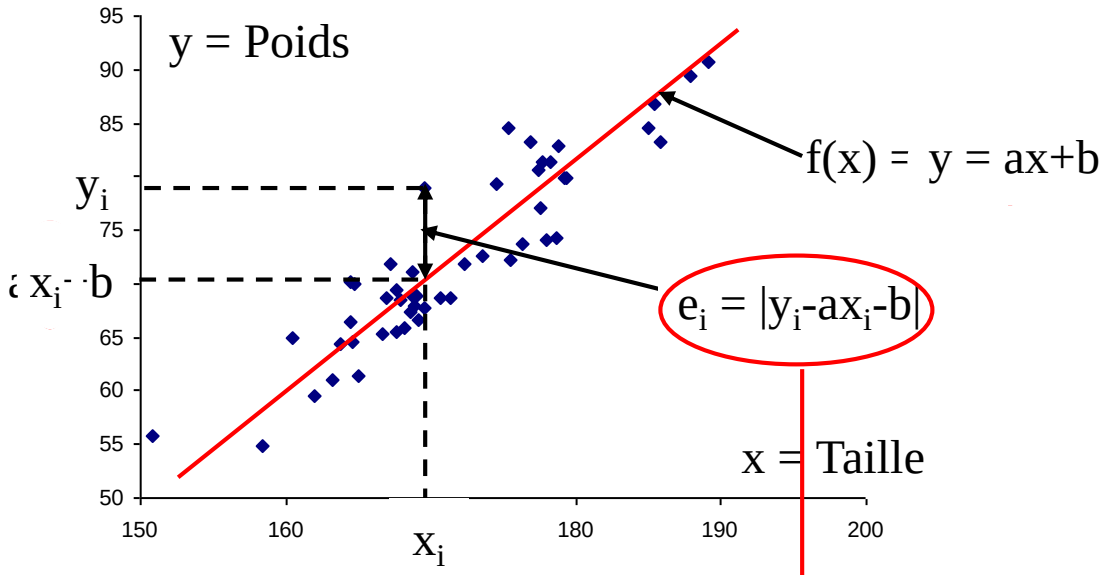
$$a = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{Cov(x,y)}{V(x)}$$

$$b = \bar{y} - a\bar{x}$$

$D_{y/x}$ passe par le point moyen $(\bar{x}, \bar{y})$

(5) AJUSTEMENT LINEAIRE
REGRESSION LINEAIRE DE Y EN X

Droite de régression linéaire de y en x
$y = f(x) = ax + b$



$y = a x + b$ définit un modèle affine

$\hat{y}_i = a x_i + b$ = valeur de y_i prévue par le modèle

$r_i = y_i - \hat{y}_i$ = résidu de la ième observation

$e_i = |r_i| = |y_i - a x_i - b|$ = erreur due au modèle

Rappel
Concepts statistiques fondamentaux
-Analyse de donnée-

Bilan Tests Statistiques

X1 \ X2	Quantitative	Qualitative
Quantitative	- Corrélation Pearson - Corrélation Spearman* - R^2	- *Test de Student* - *Test U de Mann-Whitney** - ANOVA - Test de Fisher - Test de Wilcoxon*
Qualitative	- *Test de Student* - *Test U de Mann-Whitney** - ANOVA - Test de Fisher - Test de Wilcoxon*	- Chi-2 - T de Tschuprow - V de Cramer

Bilan Tests Statistiques

Analyses Statistique

Paramétrique

Usages

1. Tester le nombre de proportion
2. Tester la différence de proportion
3. Comparer les groupes

Variables

1. Petit échantillon (n<30) : T-test
2. Grand échantillon (n>30): Z-test, F-test, ANOVA

Non Paramétrique

1. Test de signe
2. Test de rang
3. Test U de Mann-Whitney
4. Test de Wilcoxon
5. etc…

QUANTITATIVE VS QUANTITATIVE

https://www.linkedin.com/in/ernanda-chouacheu/

	Pearson	Spearman	Kendall
Principe	Mesure la **relation linéaire** entre deux variables quantitatives continues.	Évalue la **relation monotone** entre deux variables en se basant sur leurs **rangs**.	Mesure la **force de concordance** entre deux variables en comparant les **paires de données**.
Nature des données	Quantitatives continues	Ordinale ou continues	ordinale ou continues
Type de relation	Linéaire	Monotone	Monotone
Distribution requise	Normale	Aucune	Aucune
Échantillon optimal	Grand	Grand	Petit
Sensibilité aux outliers	Très sensible	Moyennement sensible	Peu sensible

QUANTITATIVE VS QUANTITATIVE

TYPES DE TEST DE STUDENT

	Test de student à un échantillon	Test de student à deux échantillons	Test de student apparié
Autres appellations	*Test de Student*	➢ *Test de student par groupes indépendants* ➢ *Test de Student pour échantillons indépendants*	➢ *Test de student par groupes appariés* ➢ *Test de Student pour échantillons dépendants*
Nombre de variables	*Une*	*Deux*	*Deux*
Type de variable	➢ *Mesures continues*	➢ *Mesures continues* ➢ *Catégorielle ou nominale pour définir les groupes*	➢ *Mesures continues* ➢ *Catégorielle ou nominale pour définir les paires dans les groupes*
Objet du test	*Décider si la moyenne de population est égale à la valeur spécifique ou pas*	*Décider si les moyennes de population pour deux groupes différents sont égales ou pas.*	*Décider si la différence entre des mesures appariées pour la population est nulle ou pas.*
Exemple:	*La taille moyenne d'un groupe de personnes est égale à 1,68m ou pas ?*	*Les tailles moyennes pour deux groupes de personnes sont identiques ou pas ?*	*La différence des tailles moyennes pour un groupe de personnes avant et après deux années est nulle ou pas ?*
Estimation de moyenne de la population	*Moyennes de l'échantillon*	*Moyenne de l'échantillon pour chaque groupe*	*Moyenne de l'échantillon des différences dans les mesures appariées.*
Écart-type de la population	*Utilisez l'écart-type de l'échantillon*	*Utilisez les écarts-types de l'échantillon pour chaque*	*Utiliser l'écart-type de l'échantillon des différences dans les mesures appariées*
Dlb (Degrés de liberté)	*Nombre d'observations dans l'échantillon moins 1, ou : n-1*	*Somme des observations dans chaque, chaque échantillon moins 2, ou: n1 + n2 – 2*	*Nombre d'observations dans l'échantillon moins 1, ou: n-1*

QUANTITATIVE VS QUANTITATIVE

Vous souhaitez comparer plus de deux groupes, optez pour l'ANOVA. Le test de Student est plus approprié lorsque vous comparez seulement deux variables.

Test ANOVA

Type de test Anova		
	Anova à un facteur	**Anova à deux facteurs**
Description	*Étudie la relation entre une variable explicative et un variable dépendante.*	*Étudie l'effet de deux ou plusieurs variables sur la variable dépendante.*
Exemple	*Le niveau d'étude (variable indépendante) d'un individu influence t-il son revenu?*	*Le niveau d'étude (première variable indépendante) et le sexe (deuxième variable indépendante) d'un indivudu influencent-ils son revenu?*

Vous travaillez avec deux groupes de données, mais vous doutez que vos données suivent une distribution normale ? Pas de panique, le test de Mann-Whitney et le test de Wilcoxon sont là pour vous !

Les tests Wilcoxon & Mann-Whitney

Test 📊	Type d'échantillons 🔄	Ce qu'il compare 🎯	Exemple 📌
Wilcoxon (signed-rank test)	**Appariés** (mêmes individus, avant/après)	Différences entre **paires de mesures**	Comparer les performances d'étudiants **avant et après** une formation 🎓
Mann-Whitney (U test)	**Indépendants** (deux groupes distincts)	Différence entre **distributions**	Comparer les niveaux de stress entre **deux groupes d'employés** 🧑‍💼

TD 1

TD 1

Exercice n°1 :

Le tableau suivant présente le chiffre d'affaire mensuel de la société X et le nombre des projets réalisés par mois

chiffre d'affaire (Y)	10	14	24	30	38	44
Nombre de projets (X)	2	3	6	8	10	12

1°) Calculer et interpréter la Moyenne, la variance et la covariance.

2°) Calculer et interpréter le coefficient de corrélation.

Correction EX1

Nombre Projet (xi)	Chiffre d'affaire (millier dh) (yi)	xi.yi	xi^2	yi^2
2	10	20	4	100
3	14	42	9	196
6	24	144	36	576
8	30	240	64	900
10	38	380	100	1444
12	44	528	144	1936
41	**160**	**1354**	**357**	**5152**

Correction EX1

2. Calcule de la covariance

- **Moyenne** $(\overline{X}) = \frac{\sum xi}{\sum ni} = \frac{41}{6} = 6{,}83 \approx$ **7 projets**

Interprétation : la moyenne mensuelle des projets réalisés par l'entreprise ALPHA est de 7 projets

- **Moyenne** $(\overline{Y}) = \frac{\sum yi}{\sum ni} = \frac{160}{6} = 26{,}67 =$ **26 670 Dhs**

Interprétation : le Chiffre d'affaire moyen mensuel réalisé par l'entreprise ALPHA est de 26670 dhs.

- **Covariance (Cov)** $= \frac{\sum xiyi}{\sum ni} - (\overline{X}\overline{Y}) \Rightarrow \frac{1354}{6} - (6{,}83 \times 26{,}67) =$ **43,51**

Interprétation : la covariance donne un chiffre positif qui montre qu'il existe une relation positive entre le nombre de projet réalisés et l'évolution du chiffre d'affaire.

Correction EX1

3. Calcule du Coefficient de corrélation

- **Variance de X:** $\sigma^2_x = \frac{\sum xi^2}{\sum ni} - \bar{X}^2 \Rightarrow \frac{357}{6} - (6{,}83)^2 = 12{,}85$
- **Variance de Y :** $\sigma^2_y = \frac{\sum yi^2}{\sum ni} - \bar{Y}^2 \Rightarrow \frac{5152}{6} - (26{,}67)^2 = 147{,}37$
- **Coefficient de corrélation (r)** $= \frac{Covariance}{\sqrt{\sigma_x{}^2 \sigma_y{}^2}} \Rightarrow \frac{43{,}51}{\sqrt{12{,}85 x 147{,}37}} = 100\%$

Interprétation : le test de corrélation montre qu'il existe une relation positive très forte de 100% entre les deux variables étudiées, cela veut dire que toute augmentation du nombre de projet réalisé implique une augmentation du chiffre d'affaire de l'entreprise, et vice versa.

EX 2

1. Décrivez le type d’échelle associé à chacune des questions du tableau.

1. Quelle est votre année de naissance ?
2. Quel est votre niveau d’étude ?
3. Quel est votre statut marital ?
4. En incluant les enfants de moins de 18 ans, quelle est la taille de votre foyer ?
5. Quels sont approximativement les revenus de votre foyer ?
6. Quel est votre sexe ?
7. Possédez-vous une carte de fidélité de l’enseigne ?

Solution EX 2

Quelle est votre année de naissance ?	Numérique
Quel est votre niveau d'étude ?	Nominale (échelle)
Quel est votre statut marital ?	Nominale (échelle)
En incluant les enfants de moins de 18 ans, quelle est la taille de votre foyer ?	Numérique
Quels sont approximativement les revenus de votre foyer ?	Nominale (échelle)
Quel est votre sexe ?	Nominale
Possédez-vous une carte de fidélité de l'enseigne ?	Nominale

EX 2

1. Fréquentez-vous ce point de vente au moins toutes les deux semaines ?
2. Quel montant moyen dépensez-vous par mois dans ce type de point de vente ?
3. Seriez-vous prêt à faire vos achats dans ce (nouveau) point de vente ?
4. À combien estimez-vous le prix moyen d’une paire de chaussures dans ce point de vente ?
5. Vous décririez-vous comme un auditeur régulier de radio ?
6. Quel type de programme de radio écoutez-vous le plus souvent ?

Solution EX 2

Fréquentez-vous ce point de vente au moins toutes les deux semaines?	Nominale
Quel montant moyen dépensez-vous par mois dans ce type de point de vente?	Numérique
Seriez-vous prêt à faire vos achats dans ce (nouveau) point de vente?	Échelle métrique
À combien estimez-vous le prix moyen d'une paire de chaussures dans ce point de vente?	Numérique
Vous décririez-vous comme un auditeur régulier de radio?	Nominale
Quel type de programme de radio écoutez-vous le plus souvent?	Nominale (échelle)

EX 2

Fréquentez-vous ce point de vente au moins toutes les deux semaines?	Nominale
Quel montant moyen dépensez-vous par mois dans ce type de point de vente?	Numérique
Seriez-vous prêt à faire vos achats dans ce (nouveau) point de vente?	Échelle métrique
À combien estimez-vous le prix moyen d'une paire de chaussures dans ce point de vente?	Numérique
Vous décririez-vous comme un auditeur régulier de radio?	Nominale
Quel type de programme de radio écoutez-vous le plus souvent?	Nominale (échelle)
Regardez-vous régulièrement le journal télévisé?	Nominale
Quel journal TV regardez-vous le plus fréquemment?	Nominale (échelle)
Lisez-vous la presse quotidienne?	Nominale
Quelle rubrique de presse quotidienne lisez-vous le plus souvent?	Nominale (échelle)
Êtes-vous abonné à un titre de presse magazine?	Nominale
La décoration de la boutique est importante à mes yeux.	Échelle métrique
Je préfère un point de vente à moins de 30 minutes de chez moi.	Échelle métrique
Je préfère être conseillé(e) par des vendeurs(euses).	Échelle métrique

EX 2

2. Donnez trois exemples de tests que vous pourriez mettre en oeuvre à partir de ces variables.

- ***Exemple1:*** **un tri croisé** entre **le montant moyen dépensé** dans le point de vente et le niveau d'études, afin de mettre en évidence un impact de la CSP sur les achats ;
- ***Exemple2:*** **une analyse typologique** afin de classer les individus de l'enquête en fonction de leur profil de réponse ;
- ***Exemple3:*** **ANOVA a deux facteurs** dont l'objet serait d'expliquer le **montant moyen dépensé** par une série de variables explicatives (**niveau d'études, statut marital**)

EX3

1. Quel(s) test(s) recommanderiez-vous à un chargé d'étude souhaitant comparer l'intention d'achat d'un produit avant et après son exposition dans un film publicitaire ?

Réponse:

Dans ce cas de figure, le chargé d'étude doit **comparer la moyenne des réponses de deux échantillons à deux périodes distinctes**, en d'autres termes **avant et après l'exposition du produit dans un message publicitaire**.

Il s'agit donc d'une mesure sur **échantillons appariés.**

EX 4

Pour chacune des questions de recherche suivantes, trouvez le type d'analyse de variance approprié en spécifiant le nombre de facteurs avec leurs niveaux.

1. L'intention d'achat des consommateurs varie-t-elle en fonction de la couleur du packaging (rouge, vert ou bleu) ?
2. La CSP (5 catégories) a-t-elle un effet sur la qualité du service perçu ?
3. L'attitude vis-à-vis de la marque d'un produit de luxe varie-t-elle en fonction du pays d'origine de la marque (France, Espagne, Italie, États-Unis) et de son réseau de distribution (très sélectif ou non sélectif) ?

EX 4

Pour chacune des questions de recherche suivantes, trouvez le type d'analyse de variance approprié en spécifiant le nombre de facteurs avec leurs niveaux.

1. L'intention d'achat des consommateurs varie-t-elle en fonction de la couleur du packaging (rouge, vert ou bleu) ?

Réponse : ANOVA à un facteur, la couleur du packaging ayant trois niveaux (rouge, vert, bleu).

2. La CSP (5 catégories) a-t-elle un effet sur la qualité du service perçu ?

Réponse : ANOVA à un facteur, la CSP ayant cinq niveaux.

3. L'attitude vis-à-vis de la marque d'un produit de luxe varie-t-elle en fonction du pays d'origine de la marque (France, Espagne, Italie, États-Unis) et de son réseau de distribution (très sélectif ou non sélectif) ?

Réponse : ANOVA à deux facteurs, le pays d'origine de la marque et le réseau de distribution, lesquels ayant respectivement quatre niveaux (France, Espagne, Italie, États-Unis) et deux niveaux (très sélectif, non sélectif).

TD 2 : ACP

Objectifs pédagogiques de ces exercices

- Comprendre comment l'ACP permet de **réduire la dimensionnalité des données** tout en conservant **l'essentiel de l'information**.
- Apprendre à standardiser, analyser et visualiser les résultats d'une ACP.
- Savoir interpréter les axes factoriels et en tirer des insights stratégiques.
- Développer une capacité à appliquer l'ACP à différents contextes (marketing, éducation, économie, santé, etc.).

Les types de mesure

Mesure nominale : Sexe; Situation matrimoniale

- Méthode : AFC

Mesure ordinale

- Méthode : AFC

Mesure métrique

- Méthode : ACP

Exemple : Les critères importants dans l'évaluation d'un club de sport

Dans une enquête sur les attentes des clients vis-à-vis de leur salle de sport, on interroge les individus sur une vingtaine de critères.

L'analyse factorielle sert à regrouper les attentes en trois ou quatre points plus simples.

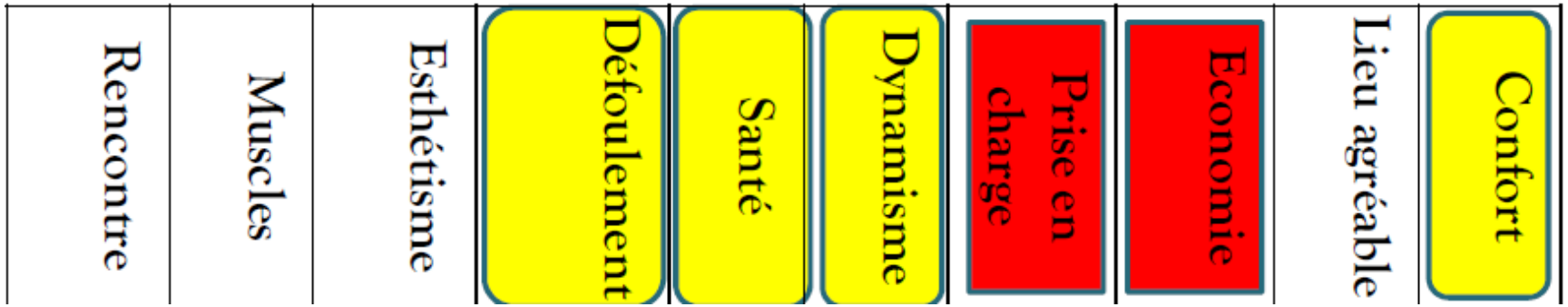
Elle agrège les variables en facteurs ou combinaisons de variables.

Exemple : Les critères importants dans l'évaluation d'un club de sport

	Rencontre	Muscles	esthétisme	Défoulement	Santé	Dynamisme	Prise en charge	Economie	Lieu agréable	confort
1	4	1	4	2	3	2	3	1	2	1
2	1	2	3	4	2	5	2	4	2	1
3	3	1	4	2	5	4	2	5	2	1
4	1	2	1	3	2	2	1	1	2	3
5	3	1	2	1	2	2	2	3	3	1
6	1	2	4	3	4	4	4	4	3	2

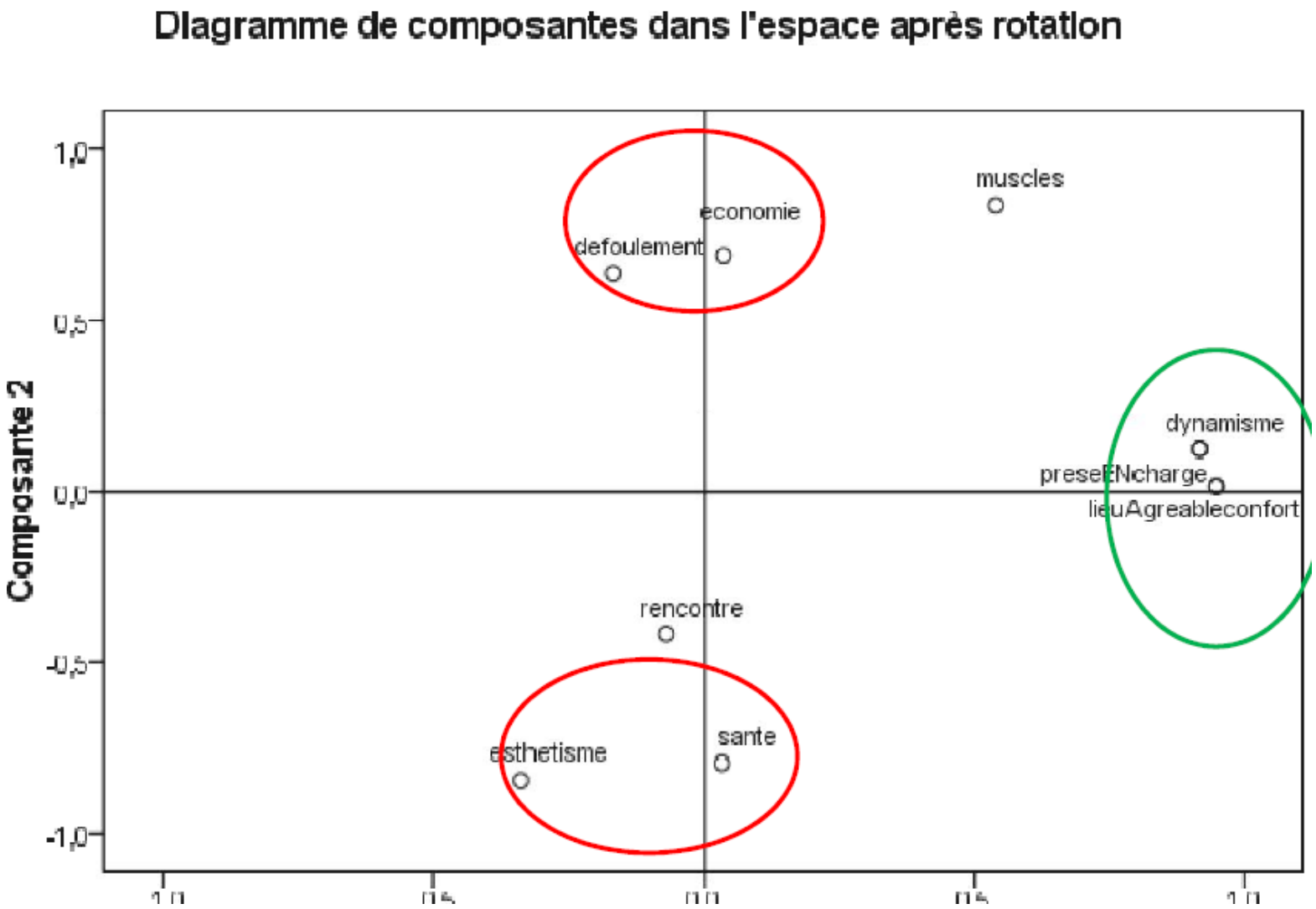
ACP

Exemple : Les critères importants dans l'évaluation d'un club de sport



	Facteur 1 (FORME)	Facteur 2 (Economie)	Facteur 3
1			
2			
3			
4			
5			
6			

ACP



À titre d'exemple, **le confort, les aspects défoulement, dynamisme et santé** représentent peut-être en fait la même chose: **être en forme (F1)**

ENCGT
المدرسة الوطنية للتجارة والتسيير
ENCG TANGER
École Nationale de Commerce et de Gestion

L’adéquation des données

- **La** « ***Measure of Sampling Adequacy*** » (MSA) **ou** *Kaiser- Meyer-Olkin* **(KMO)** teste si les corrélations partielles entre les variables ne sont pas trop faibles.

Des valeurs de KMO comprises entre 0,3 et 0,7 représentent des solutions factorielles tout juste acceptables. Il est préférable que le KMO dépasse le seuil de 0,7.

- **Le test de Sphéricité de Bartlett est assez peu utile.**

L’extraction des facteurs

Il est souvent conseillé d'imposer un pourcentage de variance expliquée égal à 60%*, mais ce seuil doit être adapté aux objectifs poursuivis. Ce critère, qui a pour objectif d'éviter une forte déformation de l'information, peut parfois être celui à privilégier.

La rotation des facteurs

- **Varimax**: (méthode la plus courante) rotation orthogonale qui minimise le nombre de variables ayant de fortes corrélations sur chaque facteur. Simplifie l'interprétation des facteurs .

Exercice ACP :

- Une enquête portant sur les perceptions de différentes marques de voitures a été réalisée auprès des consommateurs. Les individus ont évalué **10 marques d'après 15 critères**, notés sur des échelles de Likert de 1 à 9. Les variables perceptuelles sont les suivantes : **Notoriété, Ergonomie, Finition, Pre tige, Qualité, Familial, Confort, Economique, Nouveauté, Image, Qualité-prix, Innovation, Robutesse, Sportif, Spacieux**.
- Les résultats de l'enquête sont répertoriés **dans le fichier « C » (déjà envoyé).** Sur ces données, une analyse factorielle exploratoire peut permettre d'identifier les perceptions des consommateurs, mais aussi de r présenter les marques en fonction de ces perceptions dans ce que nous nommons carte perceptuelle, ou mapping perceptuel.

1. Réaliser une analyse factorielle sur ces données de l'étude.

2. Interpréter l'analyse factorielle. Quelle conclusion tirez-vous de cette analyse ?

*FICHIER C ACP .sav [Jeu_de_données1] - IBM SPSS Statistics Editeur de données

Fichier Edition Affichage Données Transformer Analyse Graphiques Utilitaires Extensions Fenêtre Aide

15 : Confort

Visible : 16 variables sur 16

	Modèle	Notoriété	Nouveauté	QualitéPrix	Robustesse	Spacieux	Ergonomie	Prestige	Familial	Economique	Image
1	Série 1 (BMW)	5,6	3,6	4,1	3,2	4,2	4,6	5,4	3,5	3,6	5,3
2	C3 (Citroën)	4,0	5,0	4,9	4,0	3,9	4,9	3,5	3,6	3,7	4,2
3	147 (AlfaRomeo)	4,6	4,0	3,8	2,4	5,3	3,5	5,6	3,4	3,6	5,0
4	Focus (Ford)	5,6	4,3	6,2	3,7	3,5	4,9	5,3	2,9	3,2	5,5
5	Megane (Renault)	4,0	3,9	4,9	4,0	3,6	4,6	2,8	4,3	4,9	3,7
6	A3 Sportback (Audi)	5,2	3,4	5,1	3,3	3,9	5,0	4,7	3,9	5,0	5,6
7	Classe A (Mercedes)	5,3	3,4	4,3	2,5	5,8	3,8	5,4	1,9	4,3	5,3
8	C4 (Citroën)	3,9	5,4	5,7	4,3	3,3	4,7	3,8	4,3	3,1	4,4
9	Golf (VolksWagen)	5,7	3,3	4,1	3,5	4,3	4,1	6,4	2,8	4,3	5,9
10	307 (Peugeot)	3,9				3,6	4,6	3,3	3,9	4,6	3,9
11		.									
12		.									
13		.									
14		.									
15		.									
16		.									
17											
18											
19											
20											
21											
22											

Menu Analyse :
- Rapports
- Statistiques descriptives
- Statistiques de Bayes
- Tableaux
- Comparer les moyennes
- Modèle linéaire général
- Modèles linéaires généralisés
- Modèles Mixtes
- Corrélation
- Régression
- Log Linéaire
- Réseaux neuronaux
- Classifier
- Réduction des dimensions
  - Analyse factorielle
  - Analyse des correspondances...
  - Codage optimal
- Echelle
- Tests non paramétriques
- Prévisions
- Survie
- Réponses multiples
- Analyse des valeurs manquantes
- Imputation multiple
- Echantillons complexes
- Simulation...
- Contrôle de qualité
- Courbe ROC...
- Modélisation spatio-temporelle...
- Marketing direct

Vue de données | Vue des variables

Analyse factorielle

Le processeur IBM SPSS Statistics est prêt

Unicode:ON

Taper ici pour rechercher

19°C Eclaircies

08:34
24/03/2025

*FICHIER C ACP .sav [Jeu_de_données1] - IBM SPSS Statistics Editeur de données

Fichier Edition Affichage Données Transformer Analyse Graphiques Utilitaires Extensions Fenêtre Aide

15 : Confort

Visible : 16 variables sur 16

	Modèle	Notoriété	Finition	Qualité	Confort	Nouveauté	QualitéPrix	Robustesse	Spacieux	Ergonomie	Prestige	Familial	Economique	Image
1	Série 1 (BMW)	5,6	6,3	2,9	1,6	3,6	4,1	3,2	4,2	4,6	5,4	3,5	3,6	5,3
2	C3 (Citroën)	4,0	3,6	4,2	4,2	5,0	4,9	4,0	3,9	4,9	3,5	3,6	3,7	4,2
3	147 (AlfaRomeo)	4,6	5,2							3,5	5,6	3,4	3,6	5,0
4	Focus (Ford)	5,6	4,2							4,9	5,3	2,9	3,2	5,5
5	Megane (Renault)	4,0	3,5							4,6	2,8	4,3	4,9	3,7
6	A3 Sportback (Audi)	5,2	5,4							5,0	4,7	3,9	5,0	5,6
7	Classe A (Mercedes)	5,3	4,8							3,8	5,4	1,9	4,3	5,3
8	C4 (Citroën)	3,9	2,8							4,7	3,8	4,3	3,1	4,4
9	Golf (VolksWagen)	5,7	5,0							4,1	6,4	2,8	4,3	5,9
10	307 (Peugeot)	3,9	3,3							4,6	3,3	3,9	4,6	3,9
11		.	.							.	.	.	.	.
12		.	.							.	.	.	.	.
13		.	.							.	.	.	.	.
14		.	.							.	.	.	.	.
15		.	.							.	.	.	.	.
16		.	.							.	.	.	.	.
17														
18														
19														
20														
21														
22														

Analyse factorielle

Modèle
Notoriété
Finition
Qualité
Confort
Nouveauté
Qualité-Prix [Qual...
Robustesse
Spacieux
Ergonomie
Prestige

Variables :

Variable de filtrage :

Valeur...

Descriptives
Extraction...
Rotation...
Scores...
Options

OK Coller Réinitialiser Annuler Aide

Vue de données Vue des variables

Le processeur IBM SPSS Statistics est prêt Unicode:ON

Taper ici pour rechercher

19°C Eclaircies 08:35 24/03/2025

*FICHIER C ACP .sav [Jeu_de_données1] - IBM SPSS Statistics Editeur de données

Fichier Edition Affichage Données Transformer Analyse Graphiques Utilitaires Extensions Fenêtre Aide

1 : Modèle | Série 1 (BMW) | Visible : 16 variables sur 16

	Modèle	Notoriété	Finition	Qualité	Confort	Nouveauté	QualitéPrix	Robustesse	Spacieux	Ergonomie	Prestige	Familial	Economique	Image
1	Série 1 (BMW)	5,6	6,3	2,9	1,6	3,6	4,1	3,2	4,2	4,6	5,4	3,5	3,6	5,3
2	C3 (Citroën)	4,0	3,6	4,2	4,2	5,0	4,9	4,0	3,9	4,9	3,5	3,6	3,7	4,2
3	147 (AlfaRomeo)	4,6	5,2							3,5	5,6	3,4	3,6	5,0
4	Focus (Ford)	5,6	4,2							4,9	5,3	2,9	3,2	5,5
5	Megane (Renault)	4,0	3,5							4,6	2,8	4,3	4,9	3,7
6	A3 Sportback (Audi)	5,2	5,4							5,0	4,7	3,9	5,0	5,6
7	Classe A (Mercedes)	5,3	4,8							3,8	5,4	1,9	4,3	5,3
8	C4 (Citroën)	3,9	2,8							4,7	3,8	4,3	3,1	4,4
9	Golf (VolksWagen)	5,7	5,0							4,1	6,4	2,8	4,3	5,9
10	307 (Peugeot)	3,9	3,3							4,6	3,3	3,9	4,6	3,9
11		.	.							.	.	.	.	.
12		.	.							.	.	.	.	.
13		.	.							.	.	.	.	.
14		.	.							.	.	.	.	.
15		.	.							.	.	.	.	.
16		.	.							.	.	.	.	.
17														
18														
19														
20														
21														
22														

Analyse factorielle

Modèle

Variables :
- Notoriété
- Finition
- Qualité
- Confort
- Nouveauté
- Qualité-Prix [Qual...
- Robustesse

Variable de filtrage :

Valeur...

Descriptives... Extraction... Rotation... Scores... Options

OK Coller Réinitialiser Annuler Aide

Vue de données | Vue des variables

Le processeur IBM SPSS Statistics est prêt | Unicode:ON

Taper ici pour rechercher

19°C Eclaircies 08:37 24/03/2025

Correction EX1

Sous SPSS

Correction EX1

Variance totale expliquée									
	Valeurs propres initiales			Sommes extraites du carré des chargements			Sommes de rotation du carré des chargements		
Composante	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé	Total	% de la variance	% cumulé
1	7,745	51,634	51,634	7,745	51,634	51,634	6,948	46,323	46,323
2	2,795	18,635	70,270	2,795	18,635	70,270	3,592	23,946	70,270
3	2,062	13,750	84,019						
4	1,276	8,510	92,529						
5	,442	2,948	95,477						
6	,388	2,585	98,062						
7	,201	1,343	99,405						
8	,069	,459	99,864						
9	,020	,136	100,000						
10	4,895E-16	3,263E-15	100,000						
11	3,902E-16	2,601E-15	100,000						
12	8,282E-17	5,521E-16	100,000						
13	-1,209E-16	-8,062E-16	100,000						
14	-2,130E-16	-1,420E-15	100,000						
15	-2,878E-16	-1,919E-15	100,000						
Méthode d'extraction : Analyse en composantes principales.									

On conseille en général d'arrêter l'extraction de facteurs lorsque 60 % de variance cumulée a été extraite (Hair et al. 1998).

Correction EX1

Le tableau de la variance totale expliquée présente **les deux dimensions** qui résument l'information. La première dimension permet d'expliquer **46,32 % de la variance du phénomène**, en ajoutant le deuxième nous arrivons à expliquer plus de 70 % de la variance totale.

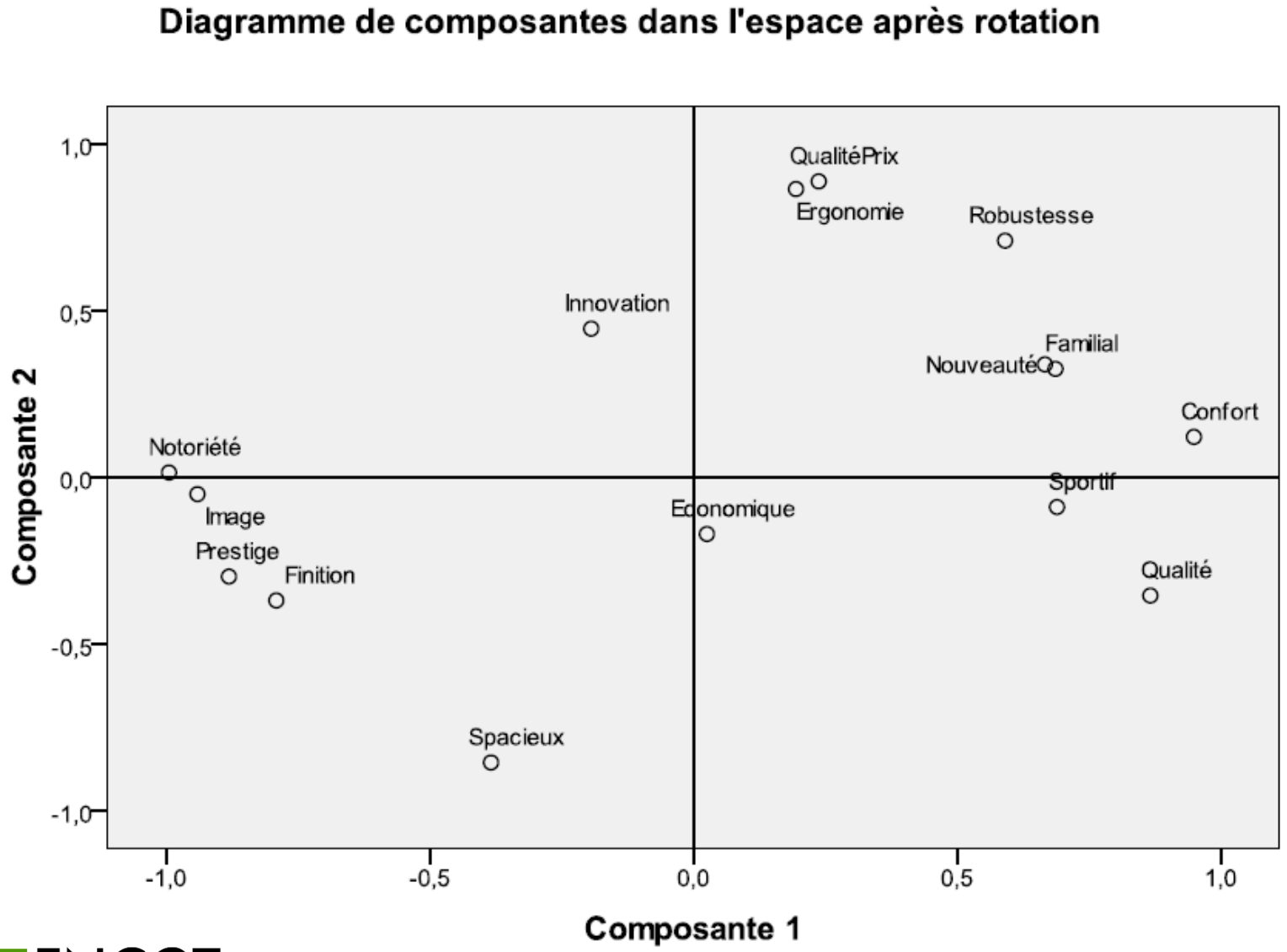
Cette variance cumulée indique que la réduction des variables à deux composantes permet de conserver l'essentiel du phénomène mesuré par les quinze variables perceptuelles initiales. Notre représentation du phénomène est donc de qualité.

Correction EX1

Qualités de représentation		
	Initiales	Extraction
Notoriété	1,000	,989
Finition	1,000	,761
Qualité	1,000	,878
Confort	1,000	,916
Nouveauté	1,000	,560
Qualité-Prix	1,000	,844
Robustesse	1,000	,854
Spacieux	1,000	,878
Ergonomie	1,000	,785
Prestige	1,000	,864
Familial	1,000	,578
Economique	1,000	,029
Image	1,000	,886
Innovation	1,000	,236
Sportif	1,000	,484
Méthode d'extraction : Analyse en composantes principales.		

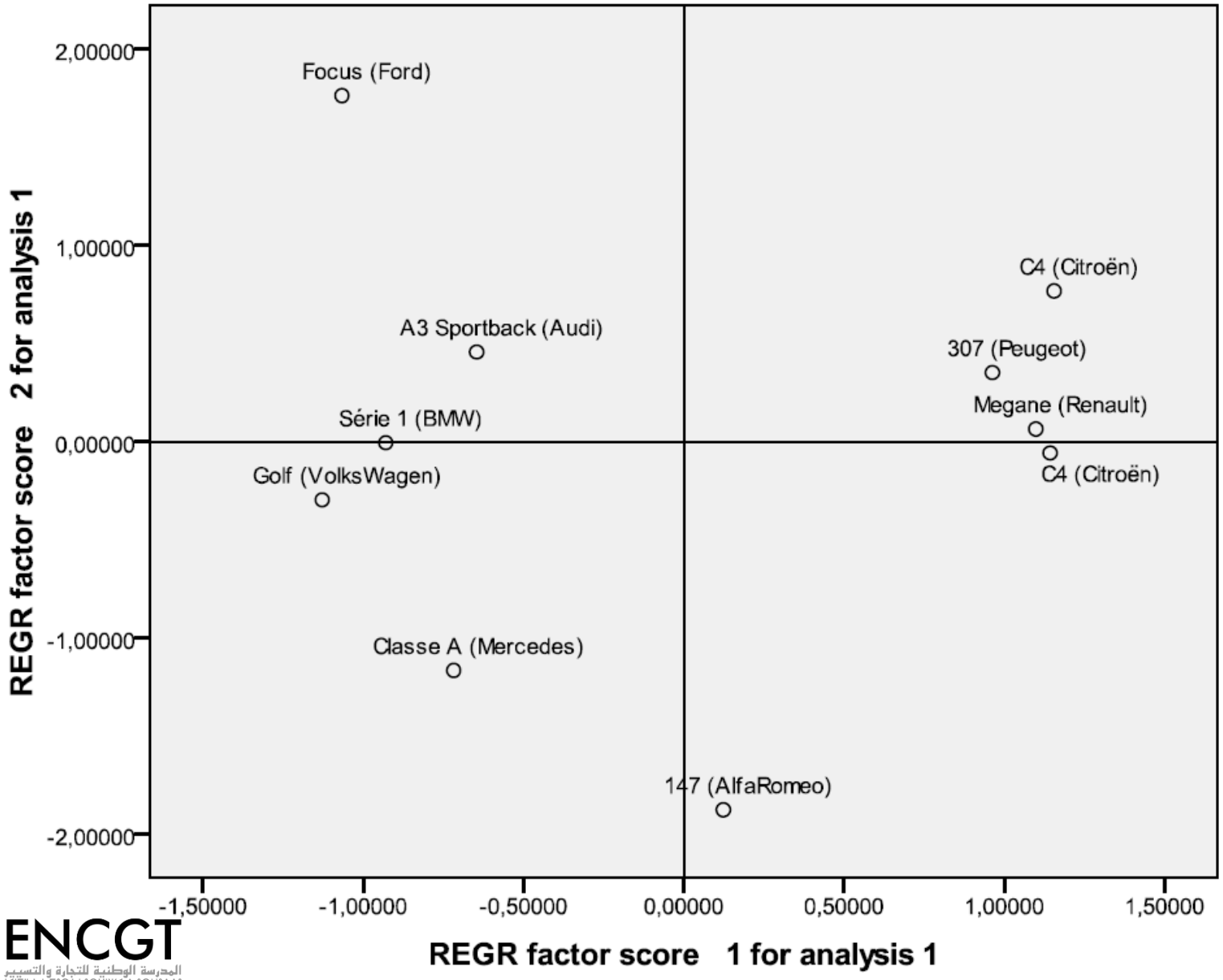
La **qualité de la représentation** permet de vérifier si les variables initiales sont bien prises en compte par les composantes extraites. Ici, la qualité de représentation ou communalité de la ***variable notoriété*** est de **0,989**. Ce qui signifie que 98,9 % de la variance de la variable est prise en compte par les composantes extraites. Dans cet exemple, les variables ***Economique et innovation*** ne sont pas bien représentées: leur communalité est inférieure à 0,5.

Correction EX1



La **première composante** relève de l'opposition entre l'image perçue (**image, notoriété, prestige**) à gauche de l'axe et le caractère familial, rassurant du véhicule (**confort, familial et qualit**é) à droite de l'axe. La seconde composante relève du rapport qualité-prix perçu.

Pour commander le graphique sous SPSS, sélectionnez le menu :
Graphes > Boîtes de dialogue héritées > Dispersion/Points, puis cliquez sur **Définir**



Les voitures sportives à l'image de prestige de la partie gauche du graphique s'opposent aux voitures plus familiales de la partie droite du graphique. En outre, le rapport qualité-prix de l'Alpha Romeo 147 est jugé médiocre contrairement à celui de la Ford Focus.

Matrice des composantes[a]

	Composante	
	1	2
Notoriété	-,905	
Confort	,949	
Image	-,940	
Prestige	-,880	
Qualité	,867	-,354
Finition	-,791	-,369
Sportif	,690	
Familial	,687	,326
Nouveauté	,667	,338
Qualité-Prix		,887
Ergonomie		,864
Spacieux	-,384	-,855
Robustesse	,592	,710
Innovation		,446
Economique		

Méthode d'extraction : Analyse en composantes principales.

a. 2 composantes extraites.

La qualité et le confort sont ainsi **positivement reliés à la dimension 1**, de même que, dans une moindre mesure, **le caractère familial et sportif. Notoriété, image, finition et prestige** sont en revanche **négativement corrélés à cet axe**.

Le rapport **qualité-prix et l'ergonomie** du modèle **sont positivement reliés à la dimension 2**.

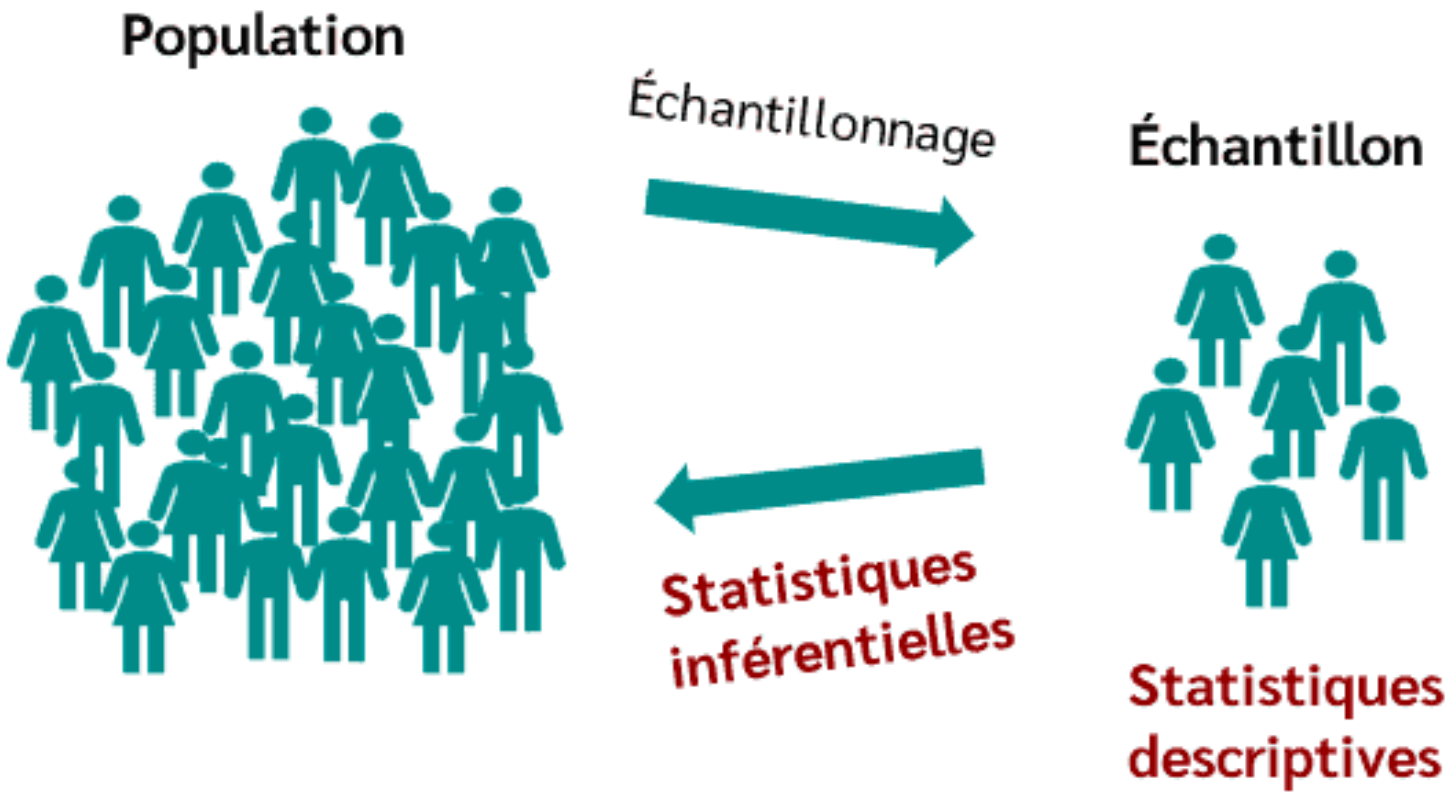
Exercice 2 : Analyse des avis sur les hôtels (A faire)

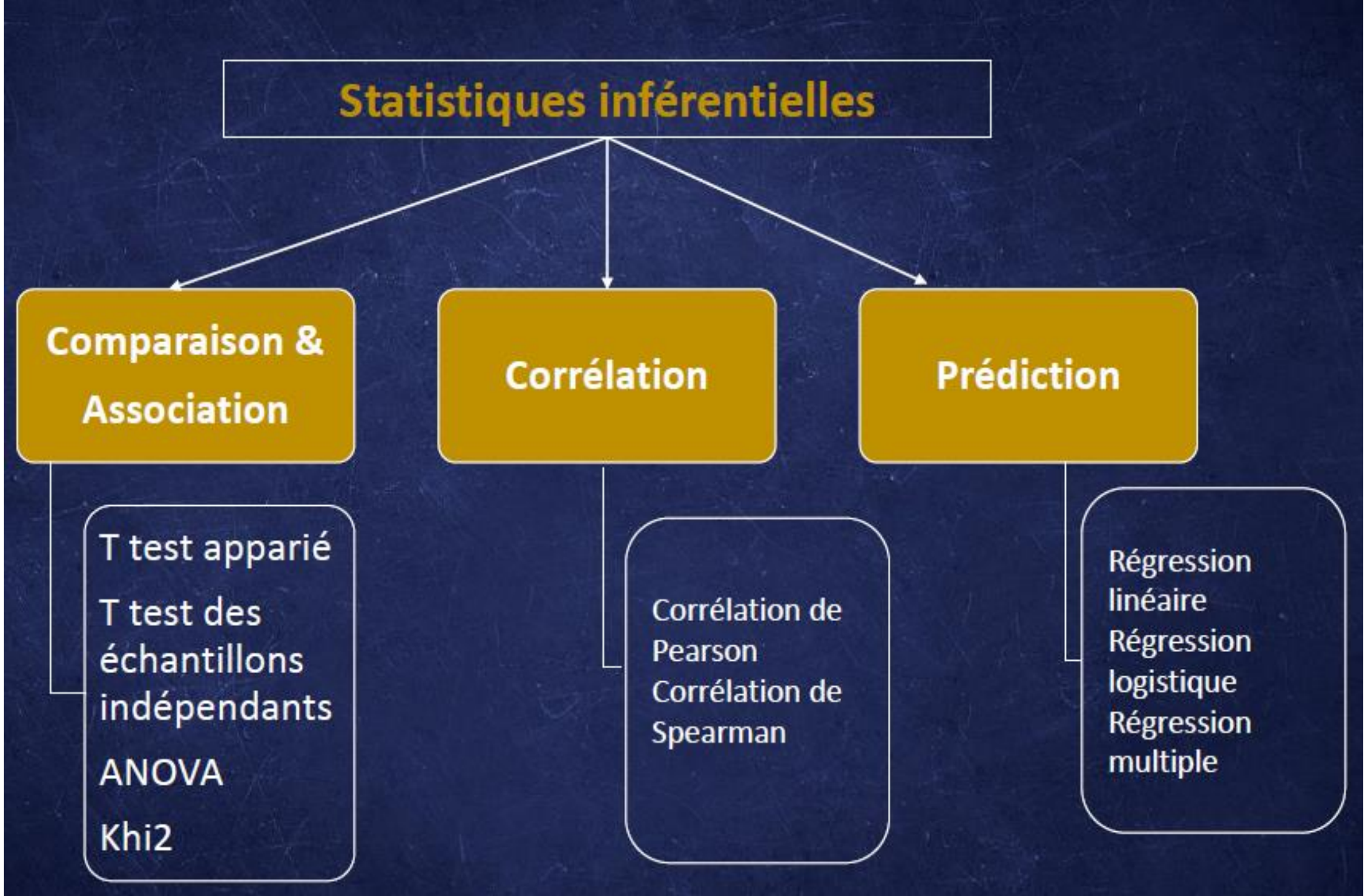
- Une plateforme de réservation d'hôtels a collecté des avis de clients sur **15 hôtels** en fonction de **12 critères** (Propreté, Confort, Localisation, Prix, Service, Wifi, Petit-déjeuner, Bruit, Vue, Équipement, Personnel, Recommandation).

1. Effectuer une **ACP** pour voir quels critères sont les plus déterminants dans la perception des hôtels.
2. Représenter les hôtels sur la carte factorielle et analyser les similarités

Rappel

Statistiques descriptives et statistiques inférentielles



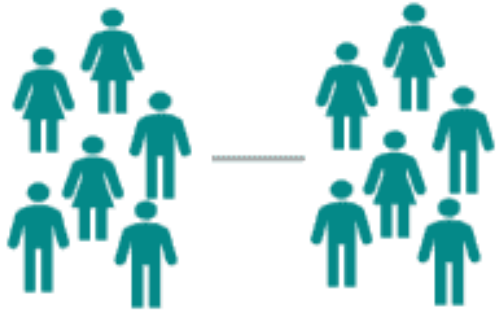

Statistiques inférentielles
Comparaison & Association
Corrélation
Prédiction
T test apparié
T test des échantillons indépendants
ANOVA
Khi2
Corrélation de Pearson
Corrélation de Spearman
Régression linéaire
Régression logistique
Régression multiple

Test t pour un échantillon

Test t pour un échantillon

Y a-t-il une différence entre un groupe et la population ?

Test t pour échantillons indépendants



Y a-t-il une différence entre deux groupes ?

Test t pour échantillons appariés

Y a-t-il une différence au sein d'un groupe entre deux moments dans le temps ?

Exemple de test t à un échantillon

Nous examinons si un didacticiel de statistiques en ligne nouvellement introduit à l'ENCG a un effet sur les résultats des étudiants aux examens.

La note moyenne à l'examen de statistiques est de 28 points depuis des années. Ce semestre, un nouveau cours de statistiques en ligne a été introduit. La direction du cours aimerait maintenant savoir si la réussite des études a changé depuis l'introduction du tutoriel de statistiques : **le cours de statistiques en ligne a-t-il un effet positif sur les résultats aux examens** ?

La population considérée est l'ensemble des étudiants qui ont passé l'examen de statistique depuis l'introduction du nouveau didacticiel de statistique. La valeur de référence à comparer est 28.

Exemple de test t à un échantillon

Étudiant	Note
1	28
2	29
3	35
4	37
5	32
6	26
7	37
8	39
9	22
10	29
11	36
12	38

H0 : La valeur moyenne de l'échantillon et la valeur prédéfinie ne diffèrent pas de manière significative.

→ Le didacticiel de statistique en ligne n'a pas d'effet significatif sur les résultats de l'examen

Exemple de test t à un échantillon

Statistiques

	n	Valeur moyenne	Écart-type	Erreur standard de la valeur moyenne
Score	12	32.33	5.47	1.58

Test t à un échantillon (valeur du test = 28)

	t	ddl	p
Score	2.75	11	0.02

Intervalle de confiance à 95% de la différence

	Différence de valeur moyenne	Inférieure	Supérieure
Score	4.33	0.86	7.81

la valeur p (bilatérale) est égale à 0,02, , cela signifie que la probabilité qu'un échantillon présentant une différence moyenne de 4,33 ou plus soit tiré de la population est de 2 %.
Le seuil de signification a été fixé à 5 %, ce qui est supérieur à 2 %.
→C'est pourquoi on suppose qu'il existe une différence significative entre l'échantillon et la population.

Le test du Khi-deux

Exemple de création d'un tableau de contingence

Sexe	Avec parapluie
femme	oui
homme	oui
femme	oui
femme	oui
homme	oui
homme	non
femme	non
homme	non
femme	non
femme	non
homme	non

Dans cet exemple, on suppose qu'un jour de pluie, un étudiant compte combien de personnes "avec" et combien de personnes "sans" parapluie viennent au cours de statistiques. En outre, il note le sexe des étudiants.

Exemple de création d'un tableau de contingence

Le résultat peut maintenant être automatiquement affiché dans un tableau de contingence.

Le tableau croisé contient les fréquences absolues des combinaisons de caractéristiques respectives.

		Avec parapluie		
		oui	non	Total
Sexe	femme	5	7	12
	homme	5	5	10
	Total	10	12	22

Test de signification d'un tableau croisé

Un tableau croisé peut être utilisé pour examiner s'il existe une relation entre les deux variables.

Cependant, étant donné qu'un tableau croisé est une statistique descriptive, une affirmation ne peut être faite que sur l'échantillon. Si une affirmation doit être faite à l'échelle de la population, le test **du chi-deux** est nécessaire

Exemple : Le test du Khi-deux

Supposons que nous voulions étudier s'il existe un lien entre **le sexe et le niveau d'éducation le plus élevé**. Pour ce faire, nous créons un questionnaire dans lequel les participants cochent leur sexe et leur niveau d'études le plus élevé.

Le résultat de l'enquête est ensuite affiché dans un tableau de contingence.

	Femelle	Masculin
Sans diplôme	6	7
Collège	13	16
Baccalauréat	16	15
Maîtrise	8	11
Total	**43**	**49**

Existe-t-il une relation entre le sexe et le niveau d'éducation le plus élevé ?

⬇

Test du khi-deux

Exemple : Le test du Khi-deux

Hypothèses:

H0 : Il n’y a pas de relation entre **le sexe et le niveau d'éducation le plus élevé.**

H1 : Il existe une corrélation entre **le sexe et le niveau d'études le plus élevé**

Les logiciels statistiques, dont SPSS, donnent une signification ou p-value, s'interprétant comme le niveau risque de se tromper en rejetant H_0 *.*
Ainsi, si elle est inférieure à 5 %, on rejette l'hypothèse d'indépendance entre les deux variables, qui sont alors significativement associées.

Résultats du test *khi-deux*

Tests du khi-carré

	Valeur	ddl	Signification asymptotique (bilatérale)	Sig. exacte (bilatérale)	Sig. exacte (unilatérale)
khi-carré de Pearson	,494[a]	1	,482		
Correction pour continuité[b]	,337	1	,562		
Rapport de vraisemblance	,497	1	,481		
Test exact de Fisher				,541	,282
Association linéaire par linéaire	,493	1	,483		
N d'observations valides	436				

a. 0 cellules (0,0%) ont un effectif théorique inférieur à 5. L'effectif théorique minimum est de 35,87.

b. Calculée uniquement pour une table 2x2

Mesures symétriques

		Valeur	Signification approximative
Nominal par Nominal	Phi	-,034	,482
	V de Cramer	,034	,482
N d'observations valides		436	

A interpréter ?

Coefficient de Cramer V

Le V de Cramer : Mesure **la force de l'association** entre deux variables **qualitatives.** Il est basé sur le Chi-carré et est ajusté pour le nombre de catégories des variables.

Une fois le test du $\chi 2$ réaliser, il est crucial de mesurer l'intensité de la liaison entre les variables.

Il varie entre 0 et 1.

si ***V* > 0.60 : Association forte**.

		V	$\geq$	0,70	relation très forte
0,50	$\leq$	V	$\leq$	0,69	relation forte
0,30	$\leq$	V	$\leq$	0,49	relation modérée
0,10	$\leq$	V	$\leq$	0,29	relation faible
0,01	$\leq$	V	$\leq$	0,09	relation très faible
		V	$=$	0,00	relation nulle

Analyse de la variance (ANOVA)

Analyse de la variance (ANOVA)

- L'analyse de la variance (ANOVA) permet de déterminer s'il existe des différences statistiquement significatives entre plusieurs échantillons (plus de deux).
    - Elle compare les moyennes et les variances des différents groupes pour identifier des variations entre eux.

Exemples d’utilisation :

- Comparaison des performances moyennes de différentes usines d’une entreprise.
- Evaluation de l’effet de plusieurs traitements médicaux sur un groupe de patients.
- Analyse de la satisfaction des clients selon différentes régions géographiques.

Hypothèses de l'analyse de la variance

$H0 : \mu 1 = \mu 2 = \cdots = \mu p = \mu$

- ✓ La moyenne de la variable dépendante est la même pour tous les groupes.
- ✓ Le facteur étudié n'a aucune influence sur la variable dépendante.

$H1 : \exists j \; ; \mu j \neq \mu$

- ✓ Il existe au moins un groupe avec une moyenne significativement différente des autres.

Regle de décision:
Si la **p-valeur** < 0.05 ⇒ on rejette HO.
Si la **p-valeur** > 0.05 ⇒ on accepte H0.

Remarque : Si H0 est rejetée, des **tests post-hoc** peuvent être nécessaires pour identifier précisément les groupes qui diffèrent

ANOVA

ANOVA à un facteur VS **ANOVA à deux facteurs**

ANOVA à un facteur	**ANOVA à deux facteurs**
Le lieu de résidence d'une personne (variable indépendante) influence-t-il son salaire ?	Le lieu de résidence (1ère variable indépendante) et le sexe (2e variable indépendante) d'une personne influencent-ils son salaire ?

Exemple Analyse de la variance à un facteur :

Vous voulez vérifier s'il y a une différence dans la consommation de café entre les étudiants de différentes matières. Pour ce faire, vous interrogez 10 étudiants de chaque filière

Cas	Consommation de café	Sujet
1	21	Mathématiques
2	23	Mathématiques
3	17	Mathématiques
4	11	Mathématiques
5	9	Mathématiques
6	27	Mathématiques
7	22	Mathématiques
8	12	Mathématiques
9	20	Mathématiques
10	4	Mathématiques
11	18	Économie
12	22	Économie
13	19	Économie
14	26	Économie
15	13	Économie
16	24	Économie
17	23	Économie
18	17	Économie
19	21	Économie
20	15	Économie
21	17	Psychologie
22	16	Psychologie
23	23	Psychologie
24	7	Psychologie
25	26	Psychologie
26	9	Psychologie
27	25	Psychologie
28	21	Psychologie
29	14	Psychologie
30	20	Psychologie

Nous voulons vérifier s'il y a une différence dans la **consommation de café** entre les étudiants de **différentes matières**.

Pour ce faire, vous interrogez 10 étudiants de chaque filière

Cas	Consommation de café	Sujet
1	21	Mathématiques
2	23	Mathématiques
3	17	Mathématiques
4	11	Mathématiques
5	9	Mathématiques
6	27	Mathématiques
7	22	Mathématiques
8	12	Mathématiques
9	20	Mathématiques
10	4	Mathématiques
11	18	Économie
12	22	Économie
13	19	Économie
14	26	Économie
15	13	Économie
16	24	Économie
17	23	Économie
18	17	Économie
19	21	Économie
20	15	Économie
21	17	Psychologie
22	16	Psychologie
23	23	Psychologie
24	7	Psychologie
25	26	Psychologie
26	9	Psychologie
27	25	Psychologie
28	21	Psychologie
29	14	Psychologie
30	20	Psychologie

	n	Moyenne	SD
Math	10	16.6	7.291
Economie	10	19.8	4.131
Psychologie	10	17.8	6.443
Total	30	18.067	5.938

	Somme des carrés	ddl	Carrés moyen	F	p
Entre les groupes	52.267	2	26.133	0.702	0.505
Au sein des groupes	1005.6	27	37.244		
Total	1057.867	29			

A interpréter ?

Analyse de la variance à un facteur : SPSS

A suivre