

CHAPITRE 8 : ANALYSE EN COMPOSANTE PRINCIPALE

8-1- Les objectifs de l'ACP :

L'Analyse en Composantes principales (ACP) fait partie des méthodes d'analyses descriptives multivariées. Le but de cette analyse est de résumer le maximum d'informations possibles pour :

- A. Faciliter l'interprétation d'un plus grand nombre de données initiales.
- B. Donner plus de sens aux données réduites.

L'ACP permet donc de réduire des tableaux de grandes tailles en un petit nombre de variables (2 ou 3 généralement) tout en conservant le maximum d'information.

On utilise l'ACP pour faire apparaître :

- Les liaisons entre variables : les systèmes de relation qui existent entre elles. C'est-à-dire : *leurs associations ou leurs oppositions*.
- La répartition des individus les uns par rapport aux autres, en relation avec les variables traitées. Autrement dit, les individus qui représentent des caractéristiques communes ou antagonistes.

L'Analyse en Composantes Principales (ACP) consiste à transformer des variables liées entre elles dites "corrélées" en nouvelles variables décorrélées les unes avec autres.

Ces nouvelles variables sont nommées "*COMPOSANTES PRINCIPALES*".

Elle permet aux praticiens de réduire l'information en un nombre de composantes plus limité que le nombre initial de variables.

8-2- Types de tableaux pouvant être traités par une ACP :

L'analyse en composantes principales peut être appliquée sur :

- Les tableaux de mesures : variables quantitatives ;
- Les tableaux de notes : Variables qualitatives ordinales ;
- Les tableaux de rangs : classement des individus de 1 à n, du meilleur au mauvais, du plus rapide au plus lent, etc.

8-3- Le tableau de donnée initial :

On suppose que l'on dispose des observations de "**P**" variables quantitatives sur "**n**" individus. Les valeurs sont "rangées" dans un tableau à **n** lignes et **p** colonnes ; on note **X** la matrice associée à ce tableau :

$$X = \begin{bmatrix} \dots & \dots & \dots \\ \dots & X_i^j & \dots \\ \dots & \dots & \dots \end{bmatrix}$$

Où x_i^j est la valeur prise par la variable j sur l'individu i .

Une variable j sera alors identifiée au vecteur $x^j = [x_1^j, \dots, x_n^j]^t$ et un individu i sera identifié au vecteur $e_i = [x_i^1, \dots, x_i^p]$.

On peut interpréter géométriquement les lignes et les colonnes du tableau **X** par des points dans deux espaces différents : l'espace des variables et l'espace des individus.

- L'espace des individus : Les n lignes peuvent être considérés comme n points de l'espace des individus à p dimensions. Deux points sont très proches si les p coordonnées de ces deux points sont très proches (mêmes valeurs pour les différentes variables).
- L'espace des variables : Les p colonnes peuvent être considérés comme p points dans un espace à n dimensions. Cet espace est appelé l'espace des variables. Si les valeurs prises par deux variables sont très voisines pour l'ensemble des individus, ces variables seront très proches (ce qui peut signifier que les variables mesurent la même chose ou encore qu'elles soient liées par une relation particulière).

Les données sont les mesures effectuées sur n unités $\{u_1, u_2, \dots, u_i, \dots, u_n\}$. Les p variables qui représentent ces mesures sont $\{v_1, v_2, \dots, v_j, \dots, v_p\}$.

8-4- Individus et variables supplémentaires

a- Individus supplémentaires :

Afin de faciliter l'interprétation des résultats, on peut introduire dans le tableau de données de départ des données que l'on appelle individus supplémentaires.

Les unités statistiques supplémentaires sont des unités statistiques sur lesquelles on dispose des observations sur l'ensemble des variables mais dont on ne tient pas compte dans le calcul des paramètres statistiques.

L'intérêt des données supplémentaires est de caractériser les graphiques des groupes d'unités statistiques supplémentaires.

b- Variables supplémentaires :

Ce sont des variables n'ayant pas de rapport direct avec l'analyse mais que l'on souhaite voir représentées dans les graphiques.

Certains auteurs utilisent les termes de variables actives pour les variables de départ et passives pour les variables supplémentaires.

8-5- Transformation des données initiales :

a- Données centrées :

On appelle données centrées le tableau de données :

$$\tilde{X} = (X_i^j - \bar{X}^j)$$

Dans ce tableau, la somme des valeurs d'une même colonne est nulle.

b- Données centrées réduites :

On appelle données centrées réduites le tableau de données :

$$\hat{X} = \frac{(X_i^j - \bar{X}^j)}{\delta_j}$$

Dans ce tableau, la somme des valeurs d'une même colonne est nulle et la somme des carrés des valeurs d'une même colonne vaut 1.

c- Matrice de variances/covariances :

On appelle matrice de variances/covariances de X la matrice centrée au carrée.

La diagonale de cette matrice contient les variances de chaque variable étudiée et les autres valeurs sont des covariances.

d- Matrice de corrélations :

On appelle matrice de corrélations R de X la matrice centrée réduite au carrée.

La diagonale de cette matrice contient des 1 et les autres valeurs représentent les coefficients de corrélation entre les variables deux à deux. Ainsi, R résume la structure des dépendances linéaires entre les p variables.

e- La diagonalisation de la matrice de corrélation :

On diagonalise la matrice de corrélation pour obtenir les valeurs propres en calculant le déterminant de la matrice suivante : $|\text{cor}(X) - \lambda I| = 0$

La résolution de polynôme caractéristique issu de calcul de déterminant donne les valeurs propres λ_i .

Chaque valeur propre nous renseigne sur la quantité d'information située sur chaque axe factoriel.

8-6- La construction des espaces factoriels :

Définir l'espace factoriel revient à :

- Définir q nouvelles variables comme axes du repère du nuage de points-individus : les composantes principales.
- Définir q nouveaux individus comme axes du repère du nuage de points-variables.

a- L'espace des individus :

L'analyse du nuage de point utilise la notion fondamentale de distance. On munit l'espace des individus de la distance euclidienne classique.

Le principe de la méthode est d'obtenir une représentation approchée du nuage des n individus dans un sous-espace de dimension faible.

Le choix de l'espace de projection s'effectue selon le critère suivant qui revient à déformer le moins possible les distances en projection. Le sous-espace de dimension k recherché est tel que la moyenne des carrés des distances entre projections soit le plus grand possible, en d'autres termes il faut que l'inertie du nuage projeté sur le sous-espace H_k soit maximale. Autrement dit, L' inertie est la quantité réelle qui mesure la dispersion des individus dans le nouvel espace à p dimensions.

b- L'espace des variables :

Changement d'origine : $g = 0$ (centrage des variables). La recherche des sous-espaces H_k se fait de proche en proche pour $k=1$ à p .

K : les valeurs propres et p : le nombre de variables initiales.

La détermination de H_1 revient à chercher une droite passant par l'origine qui s'ajuste le mieux au nuage de points-individus (maximisant l'inertie expliquée). La détermination de cette droite, revient à déterminer un vecteur unitaire u_1 porté par cette droite avec $d(0, u_1)=1$.

Une fois u_1 déterminé, on peut démontrer que le sous-espace H_1 s'ajustant au mieux au nuage de points contenant nécessairement u_1 . Pour déterminer le sous-espace H_2 , on recherche u_2 tel que u_2 **perpendiculaire** à u_1 et tel que la droite portée par u_2 , passant par 0, ait une inertie maximale. On peut démontrer que le sous-espace H_3 contient nécessairement u_1 et u_2 ...

Les vecteurs $u_1, u_2, \dots, u_p$ peuvent s'obtenir à partir de la matrice d'inertie C (covariance ou corrélation) entre les variables du tableau.

Cette matrice est telle qu'il existe p vecteurs et λ constantes qui vérifient l'équation matricielle suivante : $C.v = \lambda$.

Les p vecteurs sont les vecteurs propres et les constantes λ associées sont les valeurs propres. Ces vecteurs sont orthogonaux deux à deux et unitaires (de longueur égale à 1). Ils peuvent être rangés par ordre décroissant des valeurs propres associées : le premier vecteur propre v_1 est associé à la valeur propre la plus élevée λ_1 .

Ces vecteurs sont les vecteurs u_1 à u_p recherchés.

Les droites engendrées par ces vecteurs propres sont appelées respectivement le 1^{er}, 2^{ème}, et $p^{\text{ième}}$ axe principal d'inertie du nuage.

8-7- Etude de cas : Consommation du gaz naturel dans le secteur résidentiel

L'objet de cette étude est de faire une analyse descriptive des principaux déterminants de la consommation du gaz naturel dans le secteur résidentiel, et ce durant la période allant de 1990 jusqu'à 2012.

a- Statistiques descriptives (analyse univariée) :

Tableau 8-1 : Statistiques descriptives

Variable	Min	Max	Moyenne	Ecart-type	CV
REV	104029,27	1258164,11	542778,83	315852,18	58,19
PE	3,46	31,84	21,75	10,45	48,08
PG	0,07	0,74	0,52	0,25	47,10
PGO	11041,67	16309,52	13545,87	2207,27	16,29
COEL	314520,07	1412034,00	679604,16	290660,20	42,77
COGN	797054,71	5350950,00	2065608,83	1203476,62	58,26
COGO	1349,03	125205,08	69002,89	48673,14	70,54

Avec : **Min** : Minimum ; **Max** : Maximum ; **Moy** : Moyenne ; **CV** : Coefficient de variation ;
 REV : Revenu ; PE : Prix d'électricité ; PG : Prix du gaz naturel ; PGO : prix du gazole ;
 COEL : consommation d'électricité ; COGN : Consommation du gaz naturel et COGO :
 Consommation du gazole.

Les résultats de l'analyse univariée permettent d'apprécier l'importance voire le poids de chaque variable dans notre analyse.

On en déduit alors à partir de ces résultats de statistiques descriptive que la variable consommation du gaz naturel a le poids le plus important dans notre base de données, parce qu'elle détienne la moyenne la plus élevée et on trouve en deuxième et troisième position les variables consommation d'électricité et revenu des ménages.

Quant à la représentativité des moyennes des variables, on déduit que les variables : PE, REV, COEL et COGN sont *moyennement représentés* du fait de leur coefficient de variation¹ qui est moyen, c'est à dire tournent autour de 50%.

Cependant, nous avons la variable COGO dont la moyenne est non représentative, avec un coefficient de variation de 70,54%. Enfin la variable ayant la moyenne la plus représentative dans notre base de données est le prix du gazole (PGO), elle se caractérise avec une moyenne élevée et un coefficient de variation faible (16,29%).

b- *Matrice de corrélation (analyse bivariée) :*

Elle nous indique le niveau des relations entre les variables prises deux à deux.

Tableau 8-2 : Matrice de corrélation (résultat ACP).

Variables	REV	PE	PG	PGO	COEL	COGN	COGO
REV	1	0,836	0,815	0,841	0,980	0,962	-0,929
PE	0,836	1	0,998	0,753	0,843	0,682	-0,862
PG	0,815	0,998	1	0,725	0,824	0,654	-0,839
PGO	0,841	0,753	0,725	1	0,846	0,826	-0,933
COEL	0,980	0,843	0,824	0,846	1	0,931	-0,921
COGN	0,962	0,682	0,654	0,826	0,931	1	-0,875
COGO	-0,929	-0,862	-0,839	-0,933	-0,921	-0,875	1

¹ Cet indicateur permet de mesurer la représentativité de l'indicateur moyenne arithmétique.

Le résultat de la matrice de corrélation indique qu'il existe de fortes corrélations positives entre toutes les variables prise deux à deux, exception faite pour la variable COGO qui a une corrélation négative avec l'ensemble des variables.

Plus précisément, la forte corrélation positive (corrélation supérieur 0.9) se manifeste pour les couples de variables suivants : (PG, PE), (REV, COEL), (COGN, COEL). Quant à la corrélation négative elle se manifeste pour les couples des variables suivants : (COGO, REV), (COGO, PE), (COGO, PGO).

c- Applicabilité de l'ACP (Test de Bartlett) :

Le test d Bartlett sert à comparer la matrice de corrélation avec la matrice identité, pour voir si les deux matrices se ressemblent.

Tableau 8-3 : Test de Bartlett (Résultat ACP) :

Khi ² (Valeur observée)	363,015
Khi ² (Valeur critique)	32,671
DDL	21
p-value	< 0,0001
Alpha	0,05

Interprétation du test :

H₀: Il n'y a pas de corrélation significativement différente de 0 entre les variables.

H₁: Au moins l'une des corrélations entre les variables est significativement différente de zéro.

Etant donné que la p-value calculée est inférieure au risque d'erreur alpha=0,05, on doit rejeter l'hypothèse nulle H₀, et retenir l'hypothèse alternative H₁.

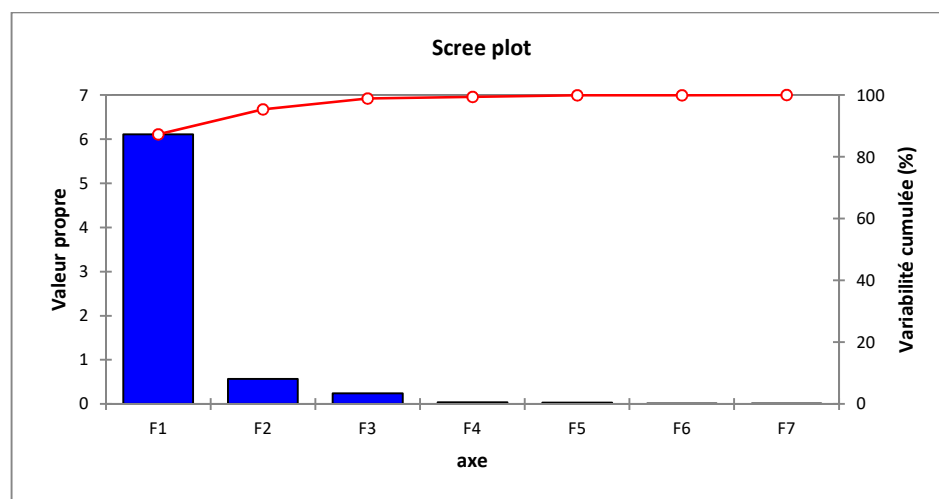
d- Valeurs propres :

Les valeurs représentent la masse d'information expliquée par chaque axe (nouvelle variable extraite à partir de la base de données de départ).

Tableau 8-4 : Valeurs propres (Résultat ACP).

	F1	F2	F3	F4	F5	F6	F7
Valeur propre	6,114	0,564	0,244	0,041	0,031	0,006	0,001
Variabilité (%)	87,348	8,052	3,479	0,580	0,449	0,083	0,010
% cumulé	87,348	95,400	98,879	99,459	99,907	99,990	100,000

Figure 8-1 : Valeurs propres.

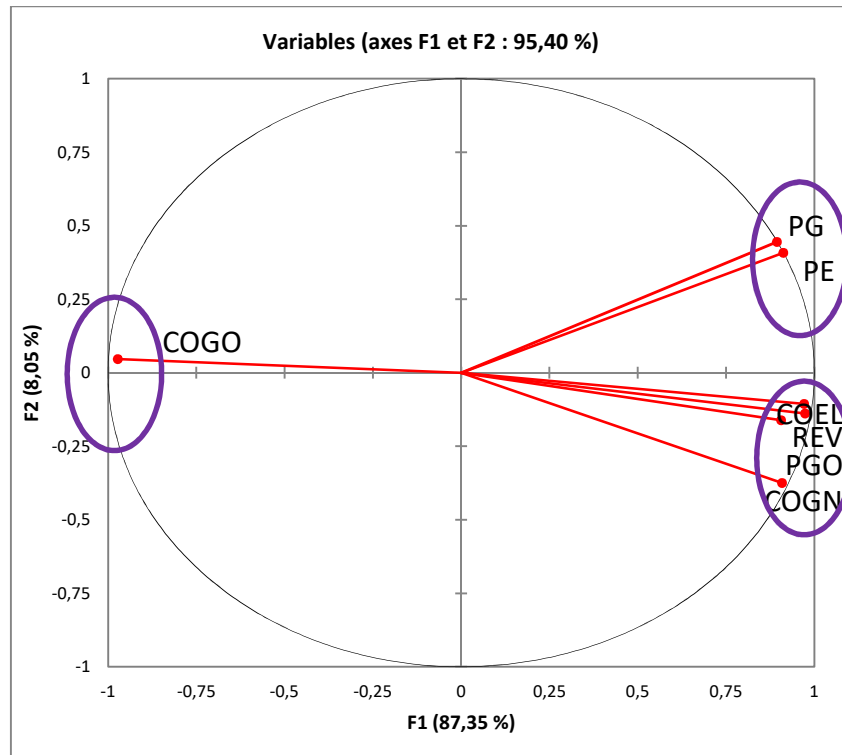


A partir du graphique, on peut déduire le nombre d'axes à prendre en considération pour l'analyse. Le nombre d'axes représente le nombre de points jusqu'au point d'inflexion.

Dans notre cas, le nombre de points est arrêté à « 02 », soit l'équivalent de deux axes, ce résultat est confirmé par le tableau des valeurs propres qui correspond à la colonne F2 et du cumule de la variabilité, soit un peu plus de 95% de l'information qui est prise en charge par les deux premiers axes.

e- L'analyse multivariée des variables :

Figure 8-2 : Analyse multivariée des variables.



Comme le montre le graphique ci-dessus, on distingue trois groupes de variables fortement corrélées entre elles et opposent en même temps les autres groupes de variables.

- Groupe (1) : composé des variables suivantes : COEL, PGO, REV et COGN.
- Groupe (2) : contient les variables suivantes : PG et PE.
- Groupe (3) : représenté uniquement par la variable COGO.

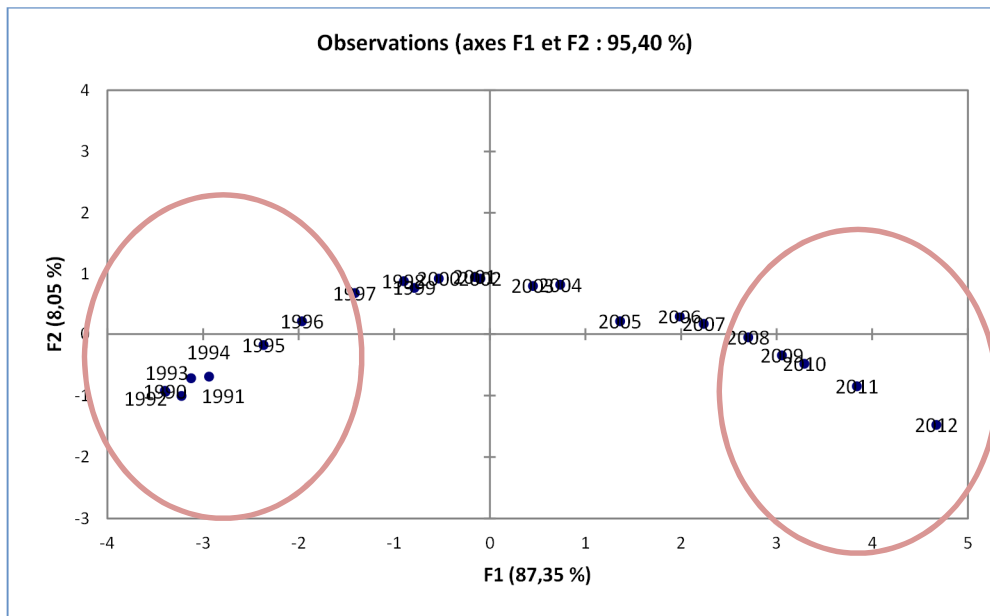
Interprétation : l'interprétation se fait par rapport aux deux axes (F1 et F2).

Par rapport à l'axe 1 (F1) : les variables PGO, REV, COEL, PG, PE et COGN sont corrélés positivement entre elles et négativement avec la variable COGO.

Par rapport à l'axe 2 (F2) : les variables COEL, PGO, REV et COGN sont corrélés négativement avec les variables PG et PE.

f- Analyse multivariée des individus :

Figure 8-3 : analyse multivariée des individus.



L'analyse multivariée des individus, consiste à calquer le graphique des valeurs propres sur celui des individus afin d'en tirer une interprétation optimale.

De ce fait, on peut procéder à l'interprétation suivante :

L'analyse multivariée des individus dégage une courbe, celle-ci est caractérisée par une forte

La consommation du gasoil (COGO) durant la période (1991-1997) et une faible consommation des autres produits énergétiques (COGN, COEL), ainsi que de faibles valeurs pour le revenu/ménage et les prix des produits énergétiques étaient relativement faible dans cette période.

Contrairement, durant la période 2008-2012 où on assiste clairement à un changement du comportement du consommateur. En effet, on observe un effet de substitution du gasoil par l'électricité et le gaz. Egalement, cette période se caractérise par l'augmentation du revenu/ménage ainsi que l'augmentation des prix des produits énergétiques.

Tous ces changements peuvent être liés à certains facteurs dont les principales sont les suivants :

1. Augmentation des salaires, ce qui entraîne une augmentation du revenu des ménages.

2. Compte tenu des subventions sur les prix des produits énergétiques, la consommation sur ces derniers continuent à accroître.
3. Pour des motifs d'ordre environnementaux, la consommation du gasoil a été remplacée par la consommation de l'électricité et du gaz.
4. Les ménages utilisent des équipements plus adaptés au développement de la société, ce qui favorise l'utilisation de certains équipements à usage électrique.

EXERCICE D'APPLICATION DU CHAPITRE 8 :

Exercice 1 :

Une étude portant sur les avantages fondamentaux recherchés par les consommateurs lors de l'achat d'un dentifrice.

Tableau de données de l'enquête

N°	V1	V2	V3	V4	V5	V6
1	7	3	6	4	2	4
2	1	3	2	4	5	4
3	6	2	7	4	1	3
4	4	5	4	6	2	5
5	1	2	2	3	6	2
6	6	3	6	4	2	4
7	5	3	6	3	4	3
8	6	4	7	4	1	4
9	3	4	2	3	6	3
10	2	6	2	6	7	6
11	6	4	7	3	2	3
12	2	3	1	4	5	4
13	7	2	6	4	1	3
14	4	6	4	5	3	6
15	1	3	2	2	6	4
16	6	4	6	3	3	4
17	5	3	6	3	3	4
18	7	3	7	4	1	4
19	2	4	3	3	6	3
20	3	5	3	6	4	6
21	1	3	2	3	5	3
22	5	4	5	4	2	4
23	2	2	1	5	4	4
24	4	6	4	6	4	7
25	6	5	4	2	1	4
26	3	5	4	6	4	7
27	4	4	7	2	2	5
28	3	7	2	6	4	3
29	4	6	3	7	2	7
30	2	3	2	4	7	2

Le sondage est réalisé dans un centre commercial, avec un échantillon de 30 personnes, qui ont donné leur avis sur les affirmations suivantes, sur une échelle de 1 à 7 (1 : en total désaccord ; 7 : entièrement d'accord).

V₁ : Il est important d'utiliser un dentifrice qui prévient la formation des caries

V₂ : Un dentifrice doit rendre les dents brillantes ;

V₃ : Un dentifrice doit renforcer les gencives ;

V₄ : Un dentifrice doit rafraîchir l'haleine ;

V₅ : La prévention contre les caries n'est pas un avantage important du dentifrice ;

V₆ : Un dentifrice doit, avant tout, donner de belles dents.

Question :

Appliquez la méthode d'Analyse en Composante Principe "ACP" sur le tableau ci-dessus, et interprétez les résultats obtenus.

Exercice 2 :

Lors d'une étude sur la relation entre le comportement au sein du foyer et comportement lié aux achats, on a obtenu des données sur les rapports de styles de vie suivants sur une échelle en sept points (1= n'est pas d'accord, 7= est d'accord) :

V₁ : Je préfère passer une soirée tranquille à la maison plutôt que d'aller à la fête ;

V₂ : Je préfère toujours les prix, même sur les petits articles ;

V₃ : Les magazines sont plus intéressants que les films ;

V₄ : Je n'achète pas les produits affichés sur les panneaux publicitaires ;

V₅ : Je suis casanier (qui aime rester à la maison) ;

V₆ : Je garde et utilise les coupons de réduction ;

V₇ : Les entreprises gaspillent beaucoup d'argent en publicité.

Les données obtenues à partir d'un échantillon de 25 individus sont regroupées ci-après :

N°	V1	V2	V3	V4	V5	V6	V7
1	6	2	7	6	5	3	5
2	5	7	5	6	6	6	4
3	5	3	4	5	6	6	7
4	3	2	2	5	1	3	2
5	4	2	3	2	2	1	3
6	2	6	2	4	3	7	5
7	1	3	3	6	2	5	7
8	3	5	1	4	2	5	6
9	7	3	6	3	5	2	4
10	6	3	3	4	4	6	5
11	6	6	2	6	4	4	7
12	3	2	2	7	6	1	6
13	5	7	6	2	2	6	1
14	6	3	5	5	7	2	3
15	3	2	4	3	2	6	5
16	2	7	5	1	4	5	2
17	3	2	2	7	2	4	6
18	6	4	5	4	7	3	3
19	7	2	6	2	5	2	1
20	5	6	6	3	4	5	3
21	2	3	3	2	1	2	6
22	3	4	2	1	4	3	6
23	2	6	3	2	1	5	3
24	6	5	7	4	5	7	2
25	7	6	5	4	6	5	3

Faites une Analyse en Composante Principale "ACP" sur ces données ?

Exercice 3 :

Afin d'étudier les caractéristiques de 24 marques de voiture, nous disposons de la base de données suivantes.

<i>Numéro</i>	<i>Modèle</i>	<i>Cylindrée</i>	<i>Puissance</i>	<i>Vitesse</i>	<i>Poids</i>	<i>Longueur</i>	<i>Largeur</i>
1	HONDA Civic	1396	90	174	850	369	166
2	RENAULT 19	1721	92	180	965	415	169
3	FIAT Tipo	1580	83	170	970	395	170
4	PEUGEOT 405	1769	90	180	1080	440	169
5	RENAULT 21	2068	88	180	1135	446	170
6	CITROEN BX	1769	90	182	1060	424	168
7	BMW 530i	2986	188	226	1510	472	175
8	ROVER 827i	2675	177	222	1365	469	175
9	RENAULT 25	2548	182	226	1350	471	180
10	OPEL Omega	1998	122	190	1255	473	177
11	PEUGEOT 405	1905	125	194	1120	439	171
12	FORD Sierra	1993	115	185	1190	451	172
13	BMW 325iX	2494	171	208	1300	432	164
14	AUDI 90 Quattro	1994	160	214	1220	439	169
15	FORD Scorpio	2933	150	200	1345	466	176
16	RENAULT Espace	1995	120	177	1265	436	177
17	NISSAN Vanette	1952	87	144	1430	436	169
18	VW Caravelle	2109	112	149	1320	457	184
19	FORD Fiesta	1117	50	135	810	371	162
20	FIAT Uno	1116	58	145	780	364	155
21	PEUGEOT 205	1580	80	159	880	370	156
22	PEUGEOT 205	1294	103	189	805	370	157
23	SEAT Ibiza SX I	1461	100	181	925	363	161
24	CITROEN AX Sport	1294	95	184	730	350	160

Moyenne et écart-type des colonnes

	Moyenne	Ecart-type
Cylindrée	1906,125	516,794
Puissance	113,667	37,968
Vitesse	183,083	24,685
Poids	1110,833	225,442
Longueur	421,583	40,470
Largeur	168,833	7,493

Matrice de corrélation :

	Cylindrée	Puissance	Vitesse	Poids	Longueur	Largeur
Cylindrée	1	0,861	0,693	0,905	0,864	0,709
Puissance	0,861	1	0,894	0,746	0,689	0,552
Vitesse	0,693	0,894	1	0,491	0,532	0,363
Poids	0,905	0,746	0,491	1	0,917	0,791
Longueur	0,864	0,689	0,532	0,917	1	0,864

Largeur	0,709	0,552	0,363	0,791	0,864	1
---------	-------	-------	-------	-------	-------	---

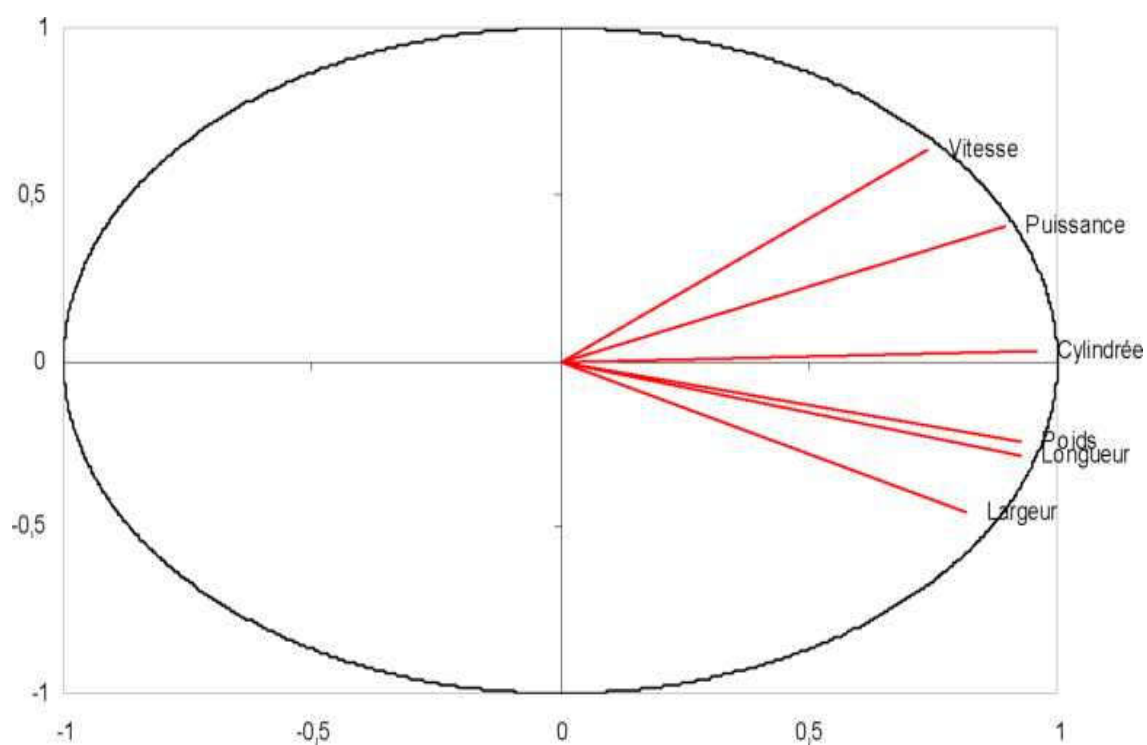
Test de sphéricité de Bartlett :

Khi ² (valeur observée)	178,583
Khi ² (valeur critique)	24,996
ddl	15
p-value unilatérale	< 0.0001
Alpha	0,05

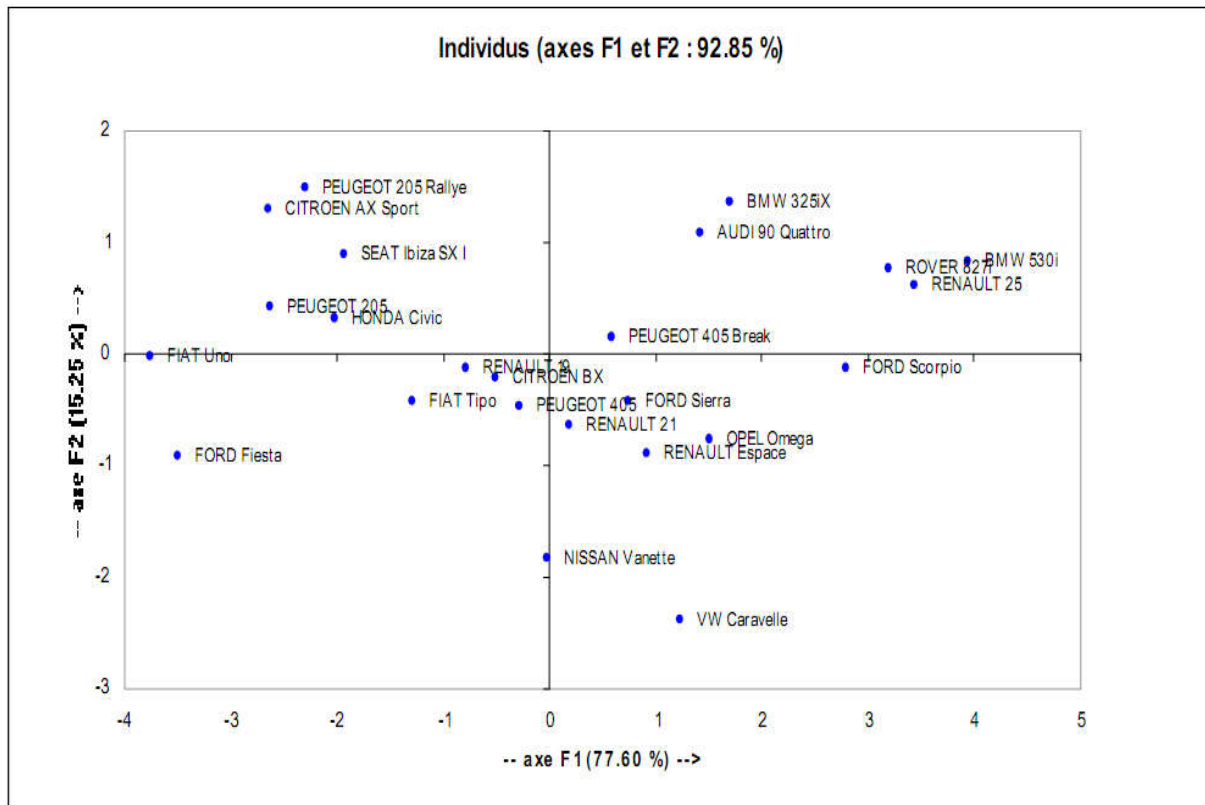
Valeurs propres

	F1	F2	F3	F4	F5	F6
Valeur propre	4,656	0,915	0,240	0,103	0,065	0,021
% variance	77,600	15,254	4,007	1,712	1,078	0,349
% cumulé	77,600	92,854	96,861	98,573	99,651	100,00

Représentation des variables (axes F1 et F2 : 92.85 %)



Position des individus :



Questions :

1. Commenter les tableaux ci-dessus ?
2. L'ACP utilisée est-elle normée ?
3. Quel est le pourcentage d'inertie du 1^{er} plan factoriel ?
4. Quelle est la qualité de la représentation du 1^{er} plan factoriel ? Avons-nous besoin de l'axe 3 ?
5. Quels sont les individus et les variables qui contribuent le plus à la construction du 1^{er} et du 2^{ème} axe factoriel ?
6. Renault 25 et Fiat Tipo Sont-ils bien représentés sur l'Axe1 ? et Sur l'axe 2 ?

CONCLUSION DE LA PARTIE 3 :

Les méthodes d'analyses de données multivariées sont nombreuses et permettent de synthétiser l'information portant sur des bases de données de grandes tailles en quelques résultats très facile à lire et à expliquer. Elles sont très élaborées et elles font appelle souvent à des représentations géométriques sur des plans (espaces vectoriels de deux dimensions).

L'analyse en composante principale « ACP » est la méthode de base dans la mesure où les autres méthodes sont des prolongements de cette dernière.

Par ailleurs, l'ACP présente certaines limites, notamment son application sur des données quantitatives ou qualitatives ordinaires. Afin de pallier à cette limite, d'autre méthodes ont été développés, tels que l'analyse factorielle des correspondances, l'analyse discriminante, l'analyse des correspondances multiples...

CONCLUSION GENERALE

Le présent support de cours est le fruit de presque une dizaine d'années d'enseignement aux profils des étudiants universitaires (à l'Université de Mouloud Mammeri de Tizi Ouzou, à l'EHEC (Ex-INC), à l'ENSM) ainsi que pour le compte des personnels des entreprises économique (publiques et privés) et d'administration publiques.

Par ailleurs, le recours aux méthodes d'analyse de données est de nos jours très indispensable pour donner un sens à l'action prise. En effet, les méthodes d'analyse de données univariées permettent de décrire une seule variable statistique (présentation tabulaires, graphiques et synthèses numériques) et les méthodes d'analyse de données bivariées, quant à elles, permettent de traiter le système de relation qui existe entre deux variables, deux à deux mutuellement.

La méthode d'Analyse en Composante Principale (ACP) est une méthode d'analyse de données multivariée qui a pour but de faire une analyse descriptive d'une base de données comportant plusieurs variables (quantitatives ou qualitatives ordinales) et qui sont mesurées sur plusieurs individus. Cependant, la présence de bases de données avec des variables quantitatives et qualitatives et l'essor de l'outil informatique (logiciels spécialisés dans le traitement de données statistiques tels que : SPSS, Xlstat et Stata) nous amènent à traiter dans la prochaine édition de ce cours les autres méthodes d'analyse de données multivariées, tels que : l'AFC, l'ACM, la CAH, la régression multiple, la régression logistique...).

ANNEXES

ANNEXES 1 : TABLES STATISTIQUES

ANNEXE 2 : APERÇU THEORIQUE SUR LES TESTS DE CONFORMITE « Z »

ANNEXE 3 : APERÇU THEORIQUE SUR LES TESTS D'HOMOGENEITE «TEST t»

ANNEXES 1 : TABLES STATISTIQUES

Annexe 1-1 : Table de Khi-deux.

$$P(\chi_v^2 \geq \chi_{v,\alpha}^2) = \alpha$$

1 - α	0,001	0,005	0,01	0,025	0,05	0,1	0,5	0,9	0,95	0,975	0,99	0,995	0,999
α	0,999	0,995	0,99	0,975	0,95	0,9	0,5	0,1	0,05	0,025	0,01	0,005	0,001
$v = \text{ddl}$													
1	0,00	0,00	0,00	0,00	0,00	0,02	0,45	2,71	3,84	5,02	6,63	7,88	10,83
2	0,00	0,01	0,02	0,05	0,10	0,21	1,39	4,61	5,99	7,38	9,21	10,60	13,82
3	0,02	0,07	0,11	0,22	0,35	0,58	2,37	6,25	7,81	9,35	11,34	12,84	16,27
4	0,09	0,21	0,30	0,48	0,71	1,06	3,36	7,78	9,49	11,14	13,28	14,86	18,47
5	0,21	0,41	0,55	0,83	1,15	1,61	4,35	9,24	11,07	12,83	15,09	16,75	20,51
6	0,38	0,68	0,87	1,24	1,64	2,20	5,35	10,64	12,59	14,45	16,81	18,55	22,48
7	0,60	0,99	1,24	1,69	2,17	2,83	6,35	12,02	14,07	16,01	18,48	20,28	24,32
8	0,86	1,34	1,65	2,18	2,73	3,49	7,34	13,36	15,51	17,53	20,09	21,95	26,12
9	1,15	1,73	2,09	2,70	3,33	4,17	8,34	14,68	16,92	19,02	21,67	23,59	27,88
10	1,48	2,16	2,56	3,25	3,94	4,87	9,34	15,99	18,31	20,48	23,21	25,19	29,59
11	1,83	2,60	3,05	3,82	4,57	5,58	10,34	17,28	19,68	21,92	24,73	26,76	31,26
12	2,21	3,07	3,57	4,40	5,23	6,30	11,34	18,55	21,03	23,34	26,22	28,30	32,91
13	2,62	3,57	4,11	5,01	5,89	7,04	12,34	19,81	22,36	24,74	27,69	29,82	34,53
14	3,04	4,07	4,66	5,63	6,57	7,79	13,34	21,06	23,68	26,12	29,14	31,32	36,12
15	3,48	4,60	5,23	6,26	7,26	8,55	14,34	22,31	25,00	27,49	30,58	32,80	37,70
16	3,94	5,14	5,81	6,91	7,96	9,31	15,34	23,54	26,30	28,85	32,00	34,27	39,25
17	4,42	5,70	6,41	7,56	8,67	10,09	16,34	24,77	27,59	30,19	33,41	35,72	40,79
18	4,90	6,26	7,01	8,23	9,39	10,86	17,34	25,99	28,87	31,53	34,81	37,16	42,31
19	5,41	6,84	7,63	8,91	10,12	11,65	18,34	27,20	30,14	32,85	36,19	38,58	43,82
20	5,92	7,43	8,26	9,59	10,85	12,44	19,34	28,41	31,41	34,17	37,57	40,00	45,31
21	6,45	8,03	8,90	10,28	11,59	13,24	20,34	29,62	32,67	35,48	38,93	41,40	46,80
22	6,98	8,64	9,54	10,98	12,34	14,04	21,34	30,81	33,92	36,78	40,29	42,80	48,27
23	7,53	9,26	10,20	11,69	13,09	14,85	22,34	32,01	35,17	38,08	41,64	44,18	49,73
24	8,08	9,89	10,86	12,40	13,85	15,66	23,34	33,20	36,42	39,36	42,98	45,56	51,18
25	8,65	10,52	11,52	13,12	14,61	16,47	24,34	34,38	37,65	40,65	44,31	46,93	52,62
26	9,22	11,16	12,20	13,84	15,38	17,29	25,34	35,56	38,89	41,92	45,64	48,29	54,05
27	9,80	11,81	12,88	14,57	16,15	18,11	26,34	36,74	40,11	43,19	46,96	49,65	55,48
28	10,39	12,46	13,56	15,31	16,93	18,94	27,34	37,92	41,34	44,46	48,28	50,99	56,89
29	10,99	13,12	14,26	16,05	17,71	19,77	28,34	39,09	42,56	45,72	49,59	52,34	58,30
30	11,59	13,79	14,95	16,79	18,49	20,60	29,34	40,26	43,77	46,98	50,89	53,67	59,70

Pour $v > 30$, La loi du χ^2 peut être approximée par la loi normale $N(v, \sqrt{v})$

Annexe 1-2 : Table de Student

(Valeurs de *T* ayant la probabilité *P* d'être dépassée en valeur absolue)

DL / P	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,929
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	6,869
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	5,408
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,263	2,821	3,250	4,781
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707
27	0,137	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,690
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,674
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,649
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,656
40	0,126	0,255	0,388	0,529	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,551
80	0,126	0,254	0,387	0,527	0,679	0,848	1,046	1,296	1,671	2,000	2,390	2,660	3,460
120	0,126	0,254	0,386	0,526	0,677	0,845	1,041	1,289	1,658	1,980	2,358	2,617	3,373
Infini	0,126	0,253	0,385	0,524	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,291

Annexe 1-3 : Table de Fisher.

$$P(F_{\nu_1, \nu_2} < f_{\nu_1, \nu_2, \alpha}) = \alpha$$

$$\alpha = 0,95$$

		v ₁																	
		1	2	3	4	5	6	7	8	9	10	15	20	30	50	100	200	500	•
v ₂	1	161	200	216	225	230	234	237	239	241	242	246	248	250	252	253	254	254	254
	2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
	3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,66	8,62	8,58	8,55	8,54	8,53	8,53
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,80	5,75	5,70	5,66	5,65	5,64	5,63
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,56	4,50	4,44	4,41	4,39	4,37	4,37
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87	3,81	3,75	3,71	3,69	3,68	3,67
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,44	3,38	3,32	3,27	3,25	3,24	3,23
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15	3,08	3,02	2,97	2,95	2,94	2,93
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94	2,86	2,80	2,76	2,73	2,72	2,71
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77	2,70	2,64	2,59	2,56	2,55	2,54
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,72	2,65	2,57	2,51	2,46	2,43	2,42	2,40
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,62	2,54	2,47	2,40	2,35	2,32	2,31	2,30
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,53	2,46	2,38	2,31	2,26	2,23	2,22	2,21
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,46	2,39	2,31	2,24	2,19	2,16	2,14	2,13
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33	2,25	2,18	2,12	2,10	2,08	2,07
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,35	2,28	2,19	2,12	2,07	2,04	2,02	2,01
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,31	2,23	2,15	2,08	2,02	1,99	1,97	1,96
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,27	2,19	2,11	2,04	1,98	1,95	1,93	1,92
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,23	2,16	2,07	2,00	1,94	1,91	1,89	1,88
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,12	2,04	1,97	1,91	1,88	1,86	1,84
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,15	2,07	1,98	1,91	1,85	1,82	1,80	1,78
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,11	2,03	1,94	1,86	1,80	1,77	1,75	1,73
	26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,07	1,99	1,90	1,82	1,76	1,73	1,71	1,69
	28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,04	1,96	1,87	1,79	1,73	1,69	1,67	1,65
	30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,93	1,84	1,76	1,70	1,66	1,64	1,62
	40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	1,92	1,84	1,74	1,66	1,59	1,55	1,53	1,51
	50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,87	1,78	1,69	1,60	1,52	1,48	1,46	1,44
	60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,84	1,75	1,65	1,56	1,48	1,44	1,41	1,39
	80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,79	1,70	1,60	1,51	1,43	1,38	1,35	1,32
	100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,77	1,68	1,57	1,48	1,39	1,34	1,31	1,28
	200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,72	1,62	1,52	1,41	1,32	1,26	1,22	1,19
	500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,69	1,59	1,48	1,38	1,28	1,21	1,16	1,11
	•	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,67	1,57	1,46	1,35	1,24	1,17	1,11	1,00

ANNEXE 2 :

APERÇU THEORIQUE SUR LES TESTS DE CONFORMITE « Z »

Annexe 2-1- Test de conformité d'une moyenne par rapport à une norme :

1-1- La variance de la PM est connue :

Hypothèses	Statistique du test	Décision
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$Z = \frac{\mu - \mu_0}{\rho/\sqrt{n}}$	RH_0 si $Z \geq \mu_{\alpha/2}$ $\bar{R}H_0$ si $Z < \mu_{\alpha/2}$
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$Z = \frac{\mu - \mu_0}{\rho/\sqrt{n}}$	RH_0 si $Z \geq \mu_{\alpha}$ $\bar{R}H_0$ si $Z < \mu_{\alpha}$
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$Z = \frac{\mu - \mu_0}{\rho/\sqrt{n}}$	RH_0 si $Z < \mu_{\alpha}$ $\bar{R}H_0$ si $Z \geq \mu_{\alpha}$

μ_{α} : Valeur critique lue sur la table de la loi normale centrée réduite au risque d'erreur α ou $(\alpha/2)\%$.

1-2- La variance de la PM est inconnue :

Hypothèses	Statistique du test	Décision
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$Z = \frac{\mu - \mu_0}{\delta/\sqrt{n-1}}$	RH_0 si $Z \geq t_{\alpha/2 ; n-1}$ $\bar{R}H_0$ si $Z < t_{\alpha/2}$
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$Z = \frac{\mu - \mu_0}{\delta/\sqrt{n-1}}$	RH_0 si $Z \geq t_{\alpha/2}$ $\bar{R}H_0$ si $Z < t_{\alpha/2}$
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$Z = \frac{\mu - \mu_0}{\delta/\sqrt{n-1}}$	RH_0 si $Z < t_{\alpha/2}$ $\bar{R}H_0$ si $Z \geq t_{\alpha/2}$

δ : représente l'écart type de l'échantillon.

$t_{\alpha/2}$: valeur critique lu sur la table de Student au risque d'erreur $\alpha/2$ et un degré de liberté $(n-1)$;

Remarque : On met toujours les résultats de la statistique calculée « Z » en valeur absolue.

Annexe 2-2- Test de conformité d'une proportion :

Hypothèses	Statistique du test	Décision
$H_0 : P = f$ $H_1 : P \neq f$	$Z = \frac{P-f}{\sqrt{f(1-f)/n}}$	RH_0 si $Z \geq U_{\alpha/2}$ $\bar{R}H_0$ si $Z < U_{\alpha/2}$
$H_0 : p = f$ $H_1 : P > f$	$Z = \frac{P-f}{\sqrt{f(1-f)/n}}$	RH_0 si $Z \geq U_{\alpha}$ $\bar{R}H_0$ si $Z < U_{\alpha}$
$H_0 : p = f$ $H_1 : P < f$	$Z = \frac{P-f}{\sqrt{f(1-f)/n}}$	RH_0 si $Z < U_{\alpha}$ $\bar{R}H_0$ si $Z \geq U_{\alpha}$

Annexe 2-3- Test de conformité d'une variance :

Hypothèses	Statistique du test	Décision
$H_0 : \sigma_e^2 = \sigma_0^2$ $H_1 : \sigma_e^2 \neq \sigma_0^2$	$\chi^2 = \frac{(n-1)\sigma_e^2}{\sigma_0^2}$	RH_0 si $\chi^2 \geq U_{\alpha/2}$ $\bar{R}H_0$ si $\chi^2 < U_{\alpha/2}$
$H_0 : \sigma_e^2 = \sigma_0^2$ $H_1 : \sigma_e^2 > \sigma_0^2$	$\chi^2 = \frac{(n-1)\sigma_e^2}{\sigma_0^2}$	RH_0 si $\chi^2 \geq U_{\alpha}$ $\bar{R}H_0$ si $\chi^2 < U_{\alpha}$
$H_0 : \sigma_e^2 = \sigma_0^2$ $H_1 : \sigma_e^2 < \sigma_0^2$	$\chi^2 = \frac{(n-1)\sigma_e^2}{\sigma_0^2}$	RH_0 si $\chi^2 < U_{\alpha}$ $\bar{R}H_0$ si $\chi^2 \geq U_{\alpha}$

ANNEXE 3 :

APERÇU THEORIQUE SUR LES TESTS D'HOMOGENEITE « TEST "T" »

Annexe 3-1 : Test d'homogénéité de deux populations (Test de Student 't'):

Soient deux échantillons indépendants de tailles n_1 et n_2 , et chacun des deux suit une loi normale : $N(\mu_1, \sigma_1)$, $N(\mu_2, \sigma_2)$.

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

$$\text{Statistique de test : } T = \left| \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \right|$$

Règle de décision : Si n_1 et n_2 sont inférieurs à 30, on compare la statistique t avec la statistique lue sur la table de Student. Et si sont supérieurs à 30, on compare la statistique t avec la statistique lue sur la table de la loi normale centrée réduite.

Si le caractère étudié est un caractère qualitatif, on utilise la statistique Z suivante :

$$t = \left| \frac{P_1 - P_2}{\sqrt{Pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \quad \text{Où : } P = \frac{n_1 * P_1 + n_2 * P_2}{n_1 + n_2}$$

REFERENCES BIBLIOGRAPHIQUES :

- ✓ Bourbonnés, R (2005), Manuel d'Econométrie, Edition Dunod.
- ✓ Cadoret, I ; Benjamin, C ; Martin, F ; Herrard, N et Tanguy, S (2009), Econométrie appliquée : Méthodes, Application et corrigés, Edition De Boeck.
- ✓ David, R. Anderson, (2015), Statistiques pour l'économie et la gestion, Ed De Boeck ;
- ✓ Escofier, P (1998), Analyses Factorielles Simples et Multiples, Objectifs, Méthodes et Interprétation, Dunod.
- ✓ Fenneteau, H et Bialès, C (1993), Analyse statistique des données : Applications et cas pour le Marketing, Ed Ellipses, Paris.
- ✓ Godelieve, M-S et Rafael, C (2013), Analyser les données en sciences sociales ; Ed P.I.E Peter Lang, Bruxelles.
- ✓ Goldfard, B et Pardoux, C (2011), Introduction à la méthode statistique; ed Dunod, 6eme édition, Paris.
- ✓ Hahn, C et Macé, S ; Méthodes statistiques appliquées au management, 2e Ed PEARSON.
- ✓ Hamrouni, A (2017), Analyse des données : Traitement statistiques des enquetes avec SPSS; Editions universitaires européennes; France.
- ✓ HUBLIER, J et RAIMBOURG, P (1996), Statistiques pour l'économie, Ed Bréal.
- ✓ Morineau, L-P (1995) : Statistique Exploratoire Multidimensionnelle, Dunod.
- ✓ Malhotra, N, Décaudin, J-M et Bouguerra, A. (2004), Études marketing avec SPSS. 4^e édition, Pearson Éducation.
- ✓ Py, B (2007) ; Statistique descriptive : Nouvelle méthode pour bien comprendre et réussir ; 5e éd Economica.
- ✓ Saporta, G (1990), Probabilités, Analyse des Données et Statistiques, Technip.
- ✓ Tenenhaus, M (1995) : Méthodes Scientifiques de Gestion, Dunod.
- ✓ Tuffery, S (2005), Data mining et statistique décisionnnelle, Editions Technip.