

Projet ingénierie

Pour réaliser ce devoir, j'ai décidé d'utiliser une base de données avec des informations sur les habitudes alimentaires de plusieurs individus afin de comprendre les habitudes alimentaires des individus et à évaluer la qualité de leur alimentation en fonction de ces différentes variables (10). Elle peut également être utilisée pour élaborer des stratégies d'amélioration de la qualité de l'alimentation des individus en fonction de leurs habitudes alimentaires actuelles. Les colonnes représentent différentes variables. Ces variables sont :

- "calories/j" : nombre de calories consommées par jour avec comme
Moyenne : 2020
Médiane : 1975
Écart-type : 286
Valeur maximum : 2500
Valeur minimum : 1500
⇒ On peut voir que les individus consomment en moyenne 2020 calories par jour, avec une variance importante de 286. De plus, 50% des individus consomment moins de 1975 calories par jour, et que les 50% restants consomment plus de 1975 calories par jour.
- "fruit/j" : nombre de portions de fruits consommées par jour ;
Moyenne : 2,4
Médiane : 2
Écart-type : 1,5
Valeur maximum : 7
Valeur minimum : 0
⇒ On peut voir que les individus consomment en moyenne 2,4 soit 3 fruits par jour, avec une variance importante de 1,5. De plus, 50% des individus consomment moins de 2 fruits par jour, et que les 50% restants consomment plus de 2 fruits par jour.
- "légumes/j" : nombre de portions de légumes consommées par jour
Moyenne : 3,45
Médiane : 3
Écart-type : 1,6
Valeur maximum : 8
Valeur minimum : 1
⇒ On peut voir que les individus consomment en moyenne 3,45 soit 4 légumes par jour, avec une variance importante de 1,6. De plus, 50% des individus consomment moins de 3 légumes par jour, et que les 50% restants consomment plus de 3 légumes par jour.
- "plat prep maison" : nombre de repas préparés à la maison consommés par jour ;
Moyenne : 1,85
Médiane : 2
Écart-type : 0,7
Valeur maximum : 3
Valeur minimum : 1
⇒ On peut voir que les individus consomment en moyenne 1,85, soit 2 plats préparés à la maison par jour, avec une variance importante de 0,7. De plus, 50% des individus consomment moins de 2 plats préparés à la maison par jour, et que les 50% restants consomment plus de 2 plats préparés à la maison par jour.
- "repas resto" : nombre de repas pris à l'extérieur (restaurant) consommés par mois ;
Moyenne : 2,05
Médiane : 2
Écart-type : 1,7

Valeur maximum : 9

Valeur minimum : 0

⇒ On peut voir que les individus consomment en moyenne 2,05 soit 2 repas au restaurant par mois, avec une variance importante de 1,7. De plus, 50% des individus consomment moins de 2 repas au restaurant par mois, et que les 50% restants consomment plus de 2 repas au restaurant par mois.

- "snack/j" : nombre de snacks consommés par jour

Moyenne : 2,4

Médiane : 2

Écart-type : 2,05

Valeur maximum : 10

Valeur minimum : 1

⇒ On peut voir que les individus consomment en moyenne 2,4 soit 2 snacks par jour, avec une variance importante de 2,05. De plus, 50% des individus consomment moins de 2 snacks par jour, et que les 50% restants consomment plus de 2 snacks par jour.

- "bois.sucres/j" : nombre de boissons sucrées consommées par jour

Moyenne : 1,85

Médiane : 2

Écart-type : 1,2

Valeur maximum : 5

Valeur minimum : 0

⇒ On peut voir que les individus consomment en moyenne 1,85 soit 2 boissons sucrées par jour, avec une variance importante de 1,2. De plus, 50% des individus consomment moins de 2 boissons sucrées par jour, et que les 50% restants consomment plus de 2 boissons sucrées par jour.

- "alcools/s" : nombre de boissons alcoolisées consommées par semaine ;

Moyenne : 0,85

Médiane : 0,5

Écart-type : 1

Valeur maximum : 3

Valeur minimum : 0

⇒ On peut voir que les individus consomment en moyenne 0,85 soit 1 boisson alcoolisée par semaine, avec une variance importante de 1. De plus, 50% des individus consomment moins de 0,5 boisson alcoolisée par semaine (soit la moitié de leurs boisson), et que les 50% restants consomment plus de 0,5 boisson alcoolisée par semaine.

- "repas veg/m" : nombre de repas végétariens consommés par mois ;

Moyenne : 1,5

Médiane : 2

Écart-type : 1

Valeur maximum : 3

Valeur minimum : 0

⇒ On peut voir que les individus consomment en moyenne 1,5 soit 2 repas végétariens par mois, avec une variance importante de 1. De plus, 50% des individus consomment moins de 2 repas végétariens par mois, et que les 50% restants consomment plus de 2 repas végétariens par mois.

- "rps sns glut/s" : nombre de repas sans gluten consommés par semaine

Moyenne : 0,75

Médiane : 1

Écart-type : 0,8

Valeur maximum : 2

Valeur minimum : 0

⇒ On peut voir que les individus consomment en moyenne 0,75 soit 1 repas sans gluten par semaine, avec une variance importante de 0,8. De plus, 50% des individus consomment moins de 1 repas sans gluten par semaine, et que les 50% restants consomment plus de 1 repas sans gluten par semaine.

Il y a 20 individus dans cette base de données et chaque individu a des habitudes alimentaires différentes. Certains individus consomment plus de fruits et légumes que d'autres, tandis que d'autres consomment plus de plats préparés à la maison ou de repas au restaurant. Certains individus consomment également plus d'alcool et de sucres que d'autres, tandis que d'autres ont une alimentation plus équilibrée. Pour réaliser cette base de données, j'ai demandé à mes proches de répondre à mon sondage :

Répondez à ces questions : (par des chiffres)
nombre de calories consommées par jour
nombre de portions de fruits consommées par jour
nombre de portions de légumes consommées par jour
nombre de repas préparés à la maison consommés par jour
nombre de repas pris à l'extérieur (restaurant) consommés par jour
nombre de snacks consommés par jour
nombre de boissons sucrées consommées par jour
nombre de boissons alcoolisées consommées par semaine
nombre de repas végétariens consommés par mois
nombre de repas sans gluten consommés par semaine

A partir de leurs réponses, j'ai pu créer ma base de données sur Excel :

A	B	C	D	E	F	G	H	I	J	K
individu	calories/j	fruit/j	legumes/j	plat prep mais	repas resto	snack/j	bois.sucres/j	alcools/s	repas veg/m	rps sns glut/s
1	2000	2	3	1	2	1	2	2	3	1
2	2500	3	4	2	1	2	3	1	2	2
3	1800	1	2	3	3	1	1	0	1	0
4	2200	2	4	2	2	3	2	3	2	1
5	1900	1	3	1	2	2	1	0	2	2
6	2400	3	4	2	1	1	2	1	1	0
7	1800	2	2	3	3	2	1	0	2	1
8	2000	1	3	1	2	1	2	1	1	2
9	2500	4	4	2	1	3	3	2	3	0
10	1900	2	3	1	2	2	1	0	2	1
11	1700	2	7	1	1	10	5	0	0	0
12	1800	2	3	2	1	2	0	0	0	0
13	1950	2	4	3	2	5	1	0	1	0
14	2000	0	1	1	9	1	1	0	2	0
15	1500	4	8	2	0	1	0	0	0	0
16	2300	7	3	3	1	1	3	0	0	0
17	1850	3	3	2	2	2	1	1	2	1
18	2100	2	2	1	2	1	3	3	2	1
19	2500	4	3	2	2	3	2	2	3	2
20	1700	1	3	2	2	4	3	1	2	1

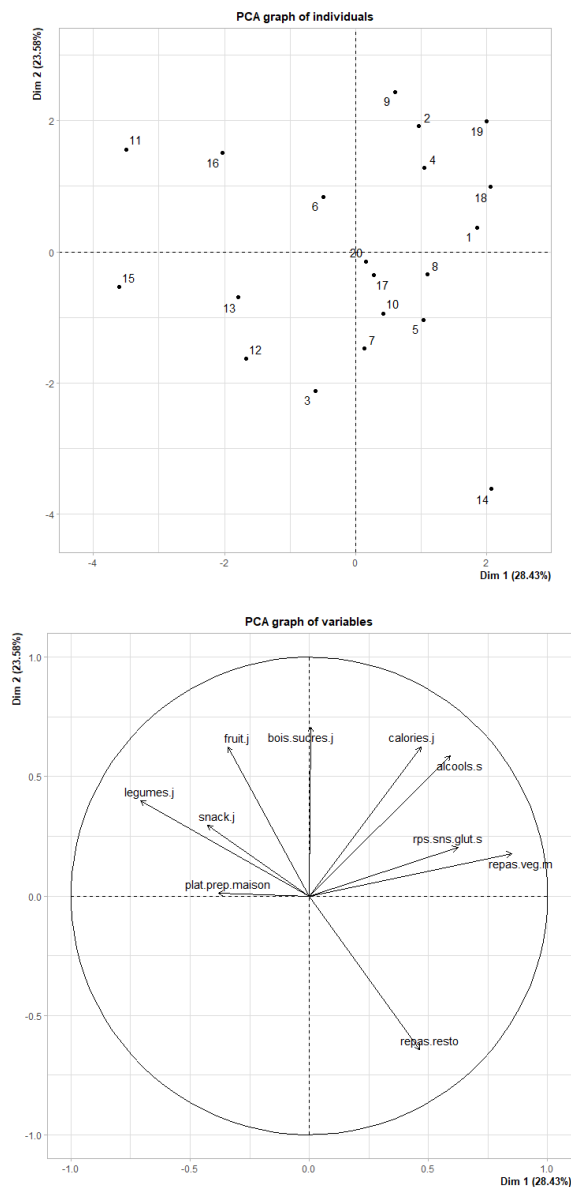
L'ACP est une méthode d'analyse multivariée qui permet de réduire la dimensionnalité d'un ensemble de données en créant de nouvelles variables, appelées composantes principales, qui représentent une grande partie de la variance totale des données. Ces nouvelles variables sont des combinaisons linéaires des variables d'origine.

Avant de réaliser l'ACP, il est nécessaire de standardiser les données afin que chaque variable ait une moyenne de zéro et une variance d'un. Cela permet d'éviter que les variables ayant des échelles de mesure différentes ne soient surreprésentées dans l'analyse.

Après avoir standardisé les données, l'ACP consiste à calculer les vecteurs propres et les valeurs propres de la matrice de covariance des données. Les vecteurs propres représentent les axes de la

nouvelle base de données et les valeurs propres représentent la proportion de variance expliquée par chaque composante principale.

Ensuite, les données peuvent être projetées sur ces nouveaux axes pour obtenir les coordonnées des individus dans l'espace des composantes principales. Cette projection permet de visualiser les relations entre les individus et les variables de la base de données. Grâce à l'application R, et la méthode de l'ACP, j'ai pu obtenir les graphiques suivant :



Pour l'axe 1 :

- En termes d'individus : les plus fortes contributions sont celles des individus 15, 1, 8 et 18
- En termes de variables : les plus fortes contributions sont celles des légumes/j, plat.prep/j, repas sans gluten/s et des repas végétariens.
- Sur l'axe 1, on a donc une opposition entre les individus qui consomment beaucoup de légumes et plat préparés à la maison (individus N°15) et les individus qui consomment beaucoup de repas sans gluten et repas végétariens (1, 8, 18). Ce qui signifie que les individus avec des scores élevés sur cet axe consomment davantage de légumes/j, plat.prep/j, repas sans gluten/s et des repas végétariens que ceux avec des scores plus faibles.

Pour l'axe 2 :

- En termes d'individus : les plus fortes contributions sont celles des individus 9 et 14
- En termes de variables : les plus fortes contributions sont celles des fruits, boissons sucrées, calories et boissons alcoolisées. On voit aussi une opposition entre les fruits et les repas pris au restaurant.

Sur l'axe 2, les individus avec des scores élevés consomment davantage de fruits, boissons sucrées, calories et boissons alcoolisées.

Donc, la première composante principale suggère que les individus qui ont des valeurs élevées sur cette composante consomment plus de légumes et plats préparés par jour, mais consomment moins d'aliments riches en glutamate monosodique. Tandis que la deuxième composante principale représente la distinction entre les personnes qui mangent des fruits et ceux qui vont au restaurant.

En conclusion, l'ACP permet de visualiser les différentes habitudes alimentaires de chaque individu et met en évidence l'opposition qui existe entre eux. Cette base de données Excel, m'a permis d'analyser plusieurs habitudes alimentaires tels que manger dehors/à la maison ou avoir une habitude alimentaire saine/pas saine. Grâce aux graphiques obtenus par la méthode de l'ACP, nous avons pu visualiser les relations entre les variables et les composantes principales et également identifié les variables qui contribuent le plus à chaque composante principale, ce qui peut nous aider à identifier les variables importantes dans notre analyse. Nous pouvons conclure que l'ACP est une méthode utile pour l'analyse de données multidimensionnelles. En réduisant la dimensionnalité des données, nous pouvons mieux comprendre les relations entre les variables et obtenir des insights importants pour la prise de décision.