

# TD3 : Consommation de protéines en Europe

## 1 Consommation de protéines en Europe

On considère des données qui concernent la consommation de protéines dans différents pays d'Europe en 1973.

Les pays sont : Albanie (al), Autriche (at), Belgique (be), Bulgarie (bg), Suisse (ch), Tchécoslovaquie (cz), République Démocratique d'Allemagne (dd), République Fédérale d'Allemagne (de), Danemark (dk), Espagne (es), Finlande (fi), France (fr), Grèce (gr), Hongrie (hu), Irlande (ie), Italie (it), Pays bas (nl), Norvège (no), Pologne (pl), Portugal (pt), Roumanie (ro), Russie (ru), Suède (se), Royaume Uni (uk), Yougoslavie (yu).

Les sources de protéines sont : viande rouge (VROU), viande blanche (VBLA), œufs (OEUF), lait (LAIT), poisson (POISS), céréales (CERE), amidon (AMID), légumes secs, noix et graines (NOIX), fruits et légumes (FRLEG). Ces données ont été collectées en pleine guerre froide ; on ajoute une variable EST qui vaut 1 pour les pays du « bloc de l'Est ».

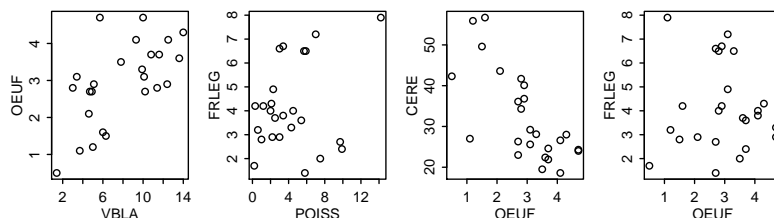
On donne ci-dessous le tableau de données brutes, la matrice des corrélations et la représentation des 4 paires de variables (VBLA, OEUF), (POISS, FRLEG), (OEUF, CERE) et (OEUF, FRLEG).

	VROU	VBLA	OEUF	LAIT	POISS	CERE	AMID	NOIX	FRLEG	EST
al	10.1	1.4	0.5	8.9	0.2	42.3	0.6	5.5	1.7	1
at	8.9	14.0	4.3	19.9	2.1	28.0	3.6	1.3	4.3	0
be	13.5	9.3	4.1	17.5	4.5	26.6	5.7	2.1	4.0	0
bg	7.8	6.0	1.6	8.3	1.2	56.7	1.1	3.7	4.2	1
ch	13.1	10.1	3.1	23.8	2.3	25.6	2.8	2.4	4.9	0
cz	9.7	11.4	2.8	12.5	2.0	34.3	5.0	1.1	4.0	1
dd	8.4	11.6	3.7	11.1	5.4	24.6	6.5	0.8	3.6	1
de	11.4	12.5	4.1	18.8	3.4	18.6	5.2	1.5	3.8	0
dk	10.6	10.8	3.7	25.0	9.9	21.9	4.8	0.7	2.4	0
es	7.1	3.4	3.1	8.6	7.0	29.2	5.7	5.9	7.2	0
fi	9.5	4.9	2.7	33.7	5.8	26.3	5.1	1.0	1.4	0
fr	18.0	9.9	3.3	19.5	5.7	28.1	4.8	2.4	6.5	0
gr	10.2	3.0	2.8	17.6	5.9	41.7	2.2	7.8	6.5	0
hu	5.3	12.4	2.9	9.7	0.3	40.1	4.0	5.4	4.2	1
ie	13.9	10.0	4.7	25.8	2.2	24.0	6.2	1.6	2.9	0
it	9.0	5.1	2.9	13.7	3.4	36.8	2.1	4.3	6.7	0
nl	9.5	13.6	3.6	23.4	2.5	22.4	4.2	1.8	3.7	0
no	9.4	4.7	2.7	23.3	9.7	23.0	4.6	1.6	2.7	0
pl	6.9	10.2	2.7	19.3	3.0	36.1	5.9	2.0	6.6	1
pt	6.2	3.7	1.1	4.9	14.2	27.0	5.9	4.7	7.9	0
ro	6.2	6.3	1.5	11.1	1.0	49.6	3.1	5.3	2.8	1
ru	9.3	4.6	2.1	16.6	3.0	43.6	6.4	3.4	2.9	1
se	9.9	7.8	3.5	24.7	7.5	19.5	3.7	1.4	2.0	0
uk	17.4	5.7	4.7	20.6	4.3	24.3	4.7	3.4	3.3	0
yu	4.4	5.0	1.2	9.5	0.6	55.9	3.0	5.7	3.2	1

### Corrélations

	VROU	VBLA	OEUF	LAIT	POISS	CERE	AMID	NOIX	FRLEG
VROU	1.00	0.15	0.59	0.50	0.06	-0.50	0.14	-0.35	-0.07
VBLA	0.15	1.00	0.62	0.28	-0.23	-0.41	0.31	-0.63	-0.06
OEUF	0.59	0.62	1.00	0.58	0.07	-0.71	0.45	-0.56	-0.05
LAIT	0.50	0.28	0.58	1.00	0.14	-0.59	0.22	-0.62	-0.41
POISS	0.06	-0.23	0.07	0.14	1.00	-0.52	0.40	-0.15	0.27
CERE	-0.50	-0.41	-0.71	-0.59	-0.52	1.00	-0.53	0.65	0.05
AMID	0.14	0.31	0.45	0.22	0.40	-0.53	1.00	-0.47	0.08
NOIX	-0.35	-0.63	-0.56	-0.62	-0.15	0.65	-0.47	1.00	0.37
FRLEG	-0.07	-0.06	-0.05	-0.41	0.27	0.05	0.08	0.37	1.00

### Représentation de paires



### 1.1 Premier regard sur les données

**Question 1** Quelles sont les variables qui sont particulièrement corrélées ou décorrélées ?

**Question 2** Pour chacun des couples (VBLA, OEUF), (POISS, FRLEG), (OEUF, CERE) et (OEUF, FRLEG), commentez la répartition des valeurs, identifiez les éventuels individus « anormaux » et expliquez le lien avec les corrélations mesurées.

### 1.2 Analyse en composante principale

On met pour l'instant de côté la variable EST. On obtient les données suivantes par ACP sur les données centrées-réduites sur les variables restantes : valeurs propres, corrélations avec les 5 premiers axes, valeurs test pour les 5 premiers axes, coordonnées des individus sur les 5 premiers axes, et enfin que la qualité de leur représentation par les 5 premiers axes principaux (en %).

Valeurs propres

	Axis1	Axis2	Axis3	Axis4	Axis5		Axis1	Axis2	Axis3	Axis4	Axis5
[1]	4.01	1.63	1.13	0.95	0.46	0.33	0.27	0.12	0.10		

Corrélations

	Comp1	Comp2	Comp3	Comp4	Comp5
VROU	-0.61	0.07	0.32	0.63	0.22
VBLA	-0.62	0.30	-0.66	-0.04	-0.20
OEUF	-0.85	0.05	-0.19	0.31	0.05
LAIT	-0.76	0.24	0.41	0.00	-0.14
POISS	-0.27	-0.83	0.34	-0.21	-0.20
CERE	0.88	0.30	-0.10	-0.01	0.16
AMID	-0.59	-0.45	-0.26	-0.33	0.50
NOIX	0.84	-0.18	0.06	0.32	0.10
FRLEG	0.22	-0.69	-0.43	0.45	-0.16

Valeurs test pour variable EST

	Axis1	Axis2	Axis3	Axis4	Axis5
EST	2.77	1.84	-1.64	-1.79	1.54

al	3.56	1.66	1.80	0.23	0.02	al	61.2	13.4	15.6	0.3	0.0
at	-1.45	1.06	-1.37	0.17	-0.95	at	33.8	18.1	29.9	0.5	14.6
be	-1.66	-0.16	-0.22	0.53	0.77	be	70.9	0.7	1.3	7.3	15.4
bg	3.20	1.33	-0.15	0.22	-0.49	bg	74.1	12.8	0.2	0.3	1.8
ch	-0.93	0.77	0.16	1.19	-0.85	ch	20.6	14.0	0.6	33.9	17.1
cz	-0.38	0.62	-1.22	-0.47	0.26	cz	4.6	12.1	47.5	7.1	2.2
dd	-1.45	-0.46	-1.33	-1.16	0.43	dd	32.6	3.3	27.3	20.8	2.9
de	-2.14	0.30	-0.82	0.11	-0.07	de	79.6	1.6	11.7	0.2	0.1
dk	-2.41	-0.29	0.77	-0.99	-0.77	dk	66.3	1.0	6.7	11.1	6.7
es	1.34	-2.61	-0.53	0.37	0.53	es	17.2	65.3	2.7	1.3	2.7
fi	-1.60	0.61	2.09	-1.44	0.04	fi	23.7	3.4	40.8	19.5	0.0
fr	-1.52	-0.80	0.00	2.00	0.26	fr	25.4	7.1	0.0	43.9	0.7
gr	2.29	-1.02	0.90	1.83	-0.41	gr	42.7	8.5	6.6	27.4	1.4
hu	1.49	0.83	-1.95	-0.22	-0.04	hu	27.6	8.7	47.6	0.6	0.0
ie	-2.72	0.78	0.02	0.44	1.04	ie	77.6	6.4	0.0	2.1	11.3
it	1.57	-0.41	-0.13	1.25	-0.82	it	44.6	3.0	0.3	28.3	12.2
nl	-1.68	0.93	-0.78	-0.13	-0.78	nl	53.0	16.4	11.6	0.3	11.4
no	-0.99	-0.84	1.74	-1.16	-0.42	no	15.7	11.2	48.1	21.4	2.9
pl	-0.12	-0.54	-1.51	-0.47	-0.02	pl	0.3	6.0	46.2	4.5	0.0
pt	1.74	-4.38	-0.04	-0.91	-0.39	pt	12.7	80.5	0.0	3.5	0.6
ro	2.81	1.14	-0.07	-0.63	0.32	ro	80.1	13.2	0.1	4.0	1.1
ru	0.80	0.11	0.38	-0.95	1.70	ru	12.8	0.3	2.9	18.0	58.3
se	-1.67	0.21	1.31	-0.75	-0.84	se	45.4	0.7	27.9	9.2	11.4
uk	-1.77	0.10	1.18	1.77	1.11	uk	33.3	0.1	14.7	33.3	13.0
yu	3.70	1.06	-0.21	-0.84	0.39	yu	86.1	7.1	0.3	4.4	0.9

**Question 3** Commentez la répartition de l'inertie : combien d'axes principaux voudrait-t-on retenir ? quelle alors est la qualité globale de représentation ?

**Question 4** Quelles sont les sources de protéines qui déterminent les axes que l'on retient ? Précisez les critères utilisés. Y a-t-il un effet de taille ?

**Question 5** Quels sont les pays d'Europe qui déterminent les axes que l'on retient ? Précisez les critères utilisés.

**Question 6** Comment peut-on interpréter les axes à partir des deux questions précédentes ?

**Question 7** Quels sont les cinq pays dont la qualité de représentation est mauvaise sur l'espace propre retenu ? Précisez les critères utilisés.

**Question 8** Que peut-on dire du Portugal ? Calculer sa contribution à l'axe qui nous intéresse et donner des pistes sur la manière dont on aurait pu le traiter. Peut-on dire qu'il n'a pas sa place dans l'analyse ?

### 1.3 Variable supplémentaire : le bloc de l'est

Comme il a été dit en introduction, ces variables ont été mesurées en pleine guerre froide. Il paraît donc intéressant de regarder ce que l'on peut dire du bloc des pays de l'est (variable EST). Pour cela on utilise les valeurs test données plus haut.

**Question 9** Avez-vous toutes les données nécessaires dans cet énoncé pour le calcul des valeurs test ? Détaillez.

**Question 10** Les conditions d'utilisation des valeurs test sont-elles réunies ? En laissant de côté les problèmes éventuels, expliquez ce que les valeurs test nous apprennent.

## 2 Qualité de représentation et part d'inertie expliquée

On réalise une ACP sur des données centrées-réduites avec  $p$  variables et  $n$  individus. On note  $p_i$  le poids de l'individu  $i$  et  $c_{ik}$  la coordonnée de l'individu  $i$  sur l'axe principal  $k$ . On rappelle que la  $k$ -ème valeur propre  $\lambda_k$  vérifie  $\lambda_k = \text{var } \mathbf{c}_k$ . Les questions 11 et 12 sont indépendantes.

**Question 11** Rappeler la formule définissant la part d'inertie expliquée par les  $q$  premiers axes. Montrer qu'elle peut s'écrire

$$\frac{\sum_{k=1}^q \sum_{i=1}^n p_i c_{ik}^2}{\sum_{k=1}^p \sum_{i=1}^n p_i c_{ik}^2}$$

**Question 12** Montrer que pour toutes variables réelles  $x_1, \dots, x_n$  et  $y_1, \dots, y_n$  on a la propriété

$$\text{Si pour tout } i, \frac{x_i}{y_i} = K \text{ alors } \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} = K.$$

**Question 13** On suppose que tous les individus ont la même qualité de représentation  $Q$  par le sous-espace formé des  $q$  premiers axes principaux :

$$\frac{\sum_{k=1}^q c_{ik}^2}{\sum_{k=1}^p c_{ik}^2} = Q, \quad i = 1, \dots, n.$$

Déduire des deux questions précédentes que la part d'inertie expliquée par les  $q$  premiers axes est aussi égale à  $Q$ .