

Analyse en Composantes Principales

Haeji Yun

2023-03-25

Dans cette étude, nous allons analyser la consommation journalière de protéine des pays européens. Les individus sont les pays et les variables sont les différentes sources de protéine qui donnent l'information sur la quantité consommée exprimée en même unité pour toutes les sources.

1. Analyse Préliminaire

Après la manipulation de données, nous retrouvons avec 25 observations et 9 variables sans valeurs manquantes.

Nous observons un grand écart entre la troisième et la quatrième quartile pour la viande rouge, la poisson, et les céréales, qui nous suggère qu'il y a quelques pays qui consomment une quantité remarquablement élevée de ces trois aliments par rapport aux autres pays.

Nous constatons également que, dans la globalité, les pays consomment plus de protéine issue des céréales et du lait par rapport aux autres sources alimentaires.

```
##      viandr      viandb.      oeuf      lait
## Min.   : 4.400   Min.    : 1.400   Min.   :0.500   Min.    : 4.90
## 1st Qu.: 7.800   1st Qu.: 4.900   1st Qu.:2.700   1st Qu.:11.10
## Median : 9.500   Median : 7.800   Median :2.900   Median :17.60
## Mean   : 9.828   Mean    : 7.896   Mean    :2.936   Mean    :17.11
## 3rd Qu.:10.600   3rd Qu.:10.800   3rd Qu.:3.700   3rd Qu.:23.30
## Max.    :18.000   Max.    :14.000   Max.    :4.700   Max.    :33.70
##      poisson      cereals      feculent      oleagine
## Min.   : 0.200   Min.    :18.60   Min.    :0.600   Min.    :0.700
## 1st Qu.: 2.100   1st Qu.:24.30   1st Qu.:3.100   1st Qu.:1.500
## Median : 3.400   Median :28.00   Median :4.700   Median :2.400
## Mean   : 4.284   Mean    :32.25   Mean    :4.276   Mean    :3.072
## 3rd Qu.: 5.800   3rd Qu.:40.10   3rd Qu.:5.700   3rd Qu.:4.700
## Max.    :14.200   Max.    :56.70   Max.    :6.500   Max.    :7.800
##      fruitleg
## Min.   :1.400
## 1st Qu.:2.900
## Median :3.800
## Mean   :4.136
## 3rd Qu.:4.900
## Max.    :7.900
```

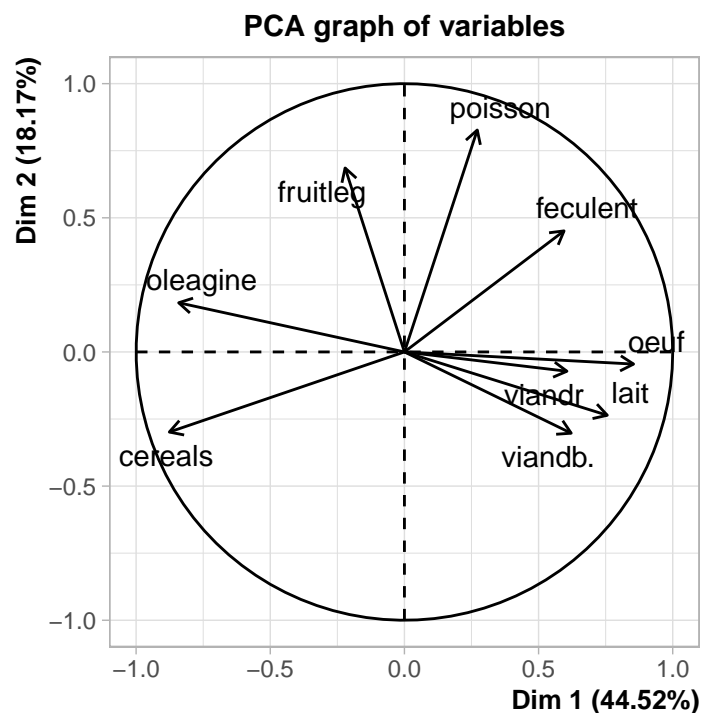
2. ACP normé

Analyse du premier plan (axe 1 et 2)

Le premier axe est positivement corrélé avec la viande, le lait, l'oeuf et le féculent. Il est négativement corrélé avec les céréales et les oléagineux. Cet axe oppose donc les pays qui consomment le plus de viande, d'oeuf, de lait et de féculent par rapport à la moyenne et les pays qui consomment le plus d'oléagineux et de céréales par rapport à la moyenne.

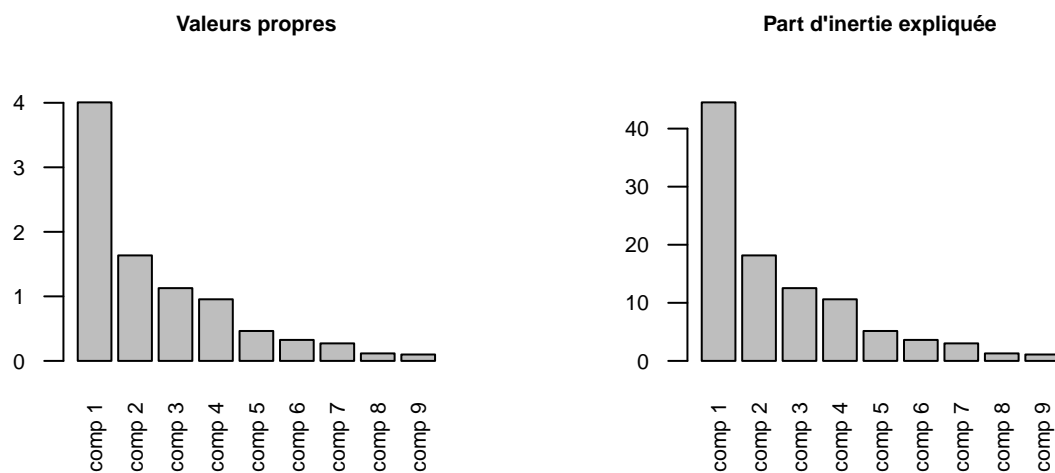
Quant au deuxième axe, il y a une corrélation positive avec la poisson, et les fruits & légumes. Donc cet axe oppose les pays qui consomment le plus de ces deux catégories de protéine que la moyenne et les pays qui en consomment le moins par rapport à la moyenne.

La part d'inertie portée par les deux axes est à l'ordre de 62,69% qui n'est pas peu mais pas suffisant pour expliquer toute la variabilité de données.



En effet, nous observons qu'il y a d'autres axes qui portent une part d'inertie qui peut être informative. Certes, les axes 1 et 2 portent les plus grandes parts d'inertie de 44,52% et 18,17% respectivement, les axes 3 et 4 présentent également une part d'inertie supérieure à 10%. De ce fait, nous ne pourrions pas expliquer suffisamment de la variabilité des observations avec seuls 2 ou 3 axes.

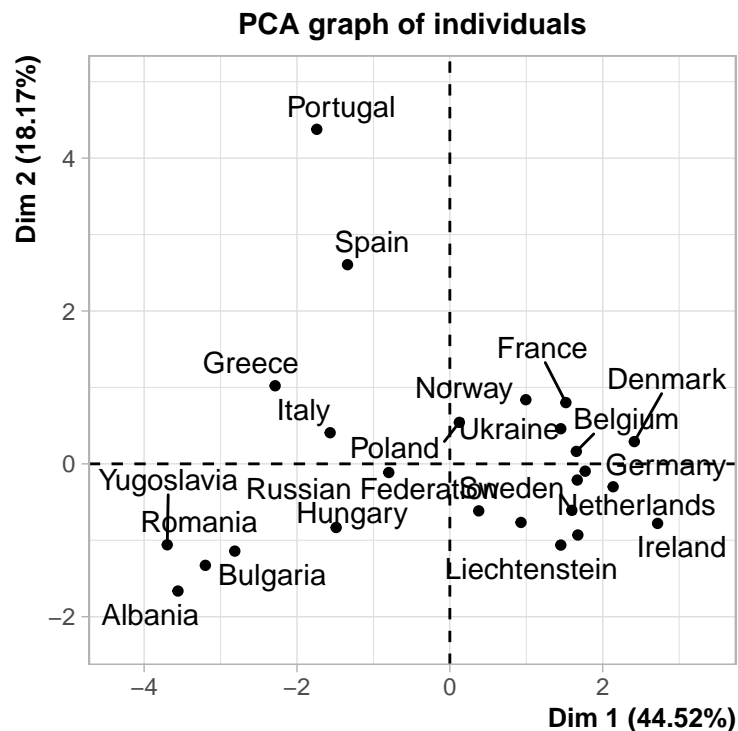
##	eigenvalue	percentage of variance	cumulative percentage of variance
## comp 1	4.0064376	44.515973	44.51597
## comp 2	1.6349994	18.166661	62.68263
## comp 3	1.1279195	12.532439	75.21507
## comp 4	0.9546640	10.607377	85.82245
## comp 5	0.4638384	5.153760	90.97621
## comp 6	0.3251310	3.612566	94.58878
## comp 7	0.2716063	3.017848	97.60662
## comp 8	0.1162919	1.292132	98.89876
## comp 9	0.0991119	1.101243	100.00000



Sur l'axe 1, le Danemark, l'Allemagne et l'Irlande sont situés plus à droite. Ils consommeraient le plus de viande, d'oeuf, de lait et de féculent et le moins de céréales et d'oléagineux par rapport à la moyenne. La Yougoslavie, la Bulgarie, et l'Albanie sont situés plus à gauche de l'axe. Ils consommeraient le plus de céréales et d'oléagineux et le moins de viande, de lait, d'oeuf et de féculent par rapport à la moyenne.

Sur l'axe 2, le Portugal et l'Espagne qui se trouvent en haut seraient les pays qui consomment le plus de poisson et de fruits & légumes par rapport à la moyenne. La Bulgarie et l'Albanie qui se trouvent en bas de l'axe 2, consommeraient le moins de poisson et de fruits & légumes.

Nous voyons déjà que la Bulgarie et l'Albanie présentent une caractéristique extrême puisque ces deux pays présentent à la fois la coordonnée la moins élevée sur les deux axes.



Analyse de l'axe 1

Nous avons supposé une corrélation positive de l'axe 1 avec la viande, le lait, l'oeuf et le féculent et une corrélation négative avec les céréales et les oléagineux.

Pour vérifier cela, nous pouvons comparer la moyenne globale avec la moyenne du groupe des individus plus remarquables sur l'axe.

Apparemment, les 5 individus ayant les coordonnées les plus élevées sur l'axe 1 ont une consommation moyenne de viande, d'oeuf, de lait, et de féculent plus élevée que la moyenne globale et une consommation moyenne d'oléagineux et de fruits & légumes moins élevée que la moyenne globale. Le contraire est observé chez les 5 individus ayant les coordonnées les moins élevées sur l'axe.

Cela confirme bien notre supposition sur l'axe 1.

```
## [1] "Moyenne des 5 individus ayant les coordonnées les plus élevées de l'axe 1"
```

```
##   viandr viandb.   oeuf   lait poisson cereals feculent oleagine
##   12.56   10.52   4.16   22.72    4.46   22.24    5.02    1.80
## fruitleg
##    3.22
```

```
## [1] "Moyenne des 5 individus ayant les coordonnées les moins élevées de l'axe 1"
```

```
##   viandr viandb.   oeuf   lait poisson cereals feculent oleagine
##    7.74    4.34   1.52   11.08    1.78   49.24    2.00    5.60
## fruitleg
##    3.68
```

```
## [1] "Moyenne globale"
```

```
##   viandr viandb.   oeuf   lait poisson cereals feculent oleagine
##   9.828   7.896   2.936  17.112    4.284   32.248    4.276    3.072
## fruitleg
##    4.136
```

Nous pouvons aussi la vérifier avec la description automatique de l'axe par la fonction `dimdesc()` qui explique bien la corrélation positive de l'axe avec l'oeuf, le lait, la viande, et le féculent et sa corrélation négative avec les oléagineux et les céréales.

```
## $quanti
##      correlation      p.value
## oeuf      0.8540434 5.631033e-08
## lait      0.7560622 1.234504e-05
## viandb.    0.6216121 9.101673e-04
## viandr     0.6057056 1.333313e-03
## feculent   0.5949735 1.706062e-03
## oleagine  -0.8413450 1.376380e-07
## cereals   -0.8761911 9.505468e-09
##
## attr(,"class")
## [1] "condes" "list"
```

Analyse de l'axe 2

Pour l'axe 2, nous avons supposé une corrélation positive avec la poisson et les fruits & légumes.

Comme pour l'axe 1, nous pouvons comparer la moyenne globale avec la moyenne du groupe des individus plus remarquables sur l'axe.

Apparemment, les 5 individus ayant les coordonnées les plus élevées sur l'axe 2 ont une consommation moyenne de poisson et de fruits & légumes plus élevée que la moyenne globale. Les 5 individus ayant les coordonnées les moins élevées sur l'axe présentent une caractéristique opposée.

Cela confirme bien notre supposition sur l'axe 2.

```
## [1] "Moyenne des 5 individus ayant les coordonnées les plus élevées de l'axe 2"
```

```
##   viandr viandb.   oeuf   lait poisson cereals feculent oleagine
##   10.18   4.94    2.60   14.78   8.50   29.80    4.64    4.48
## fruitleg
##    6.16
```

```
## [1] "Moyenne des 5 individus ayant les coordonnées les moins élevées de l'axe 2"
```

```
##   viandr viandb.   oeuf   lait poisson cereals feculent oleagine
##    7.48   6.54    1.82   11.54    1.02   46.50    2.28    4.30
## fruitleg
##    3.24
```

```
## [1] "Moyenne globale"
```

```
##   viandr viandb.   oeuf   lait poisson cereals feculent oleagine
##   9.828   7.896    2.936   17.112   4.284   32.248    4.276    3.072
## fruitleg
##    4.136
```

Nous voyons bien la corrélation positive de l'axe avec la poisson et les fruits & légumes par la description automatique de l'axe donnée par la fonction `dimdesc()`.

```
## $quanti
##      correlation      p.value
## poisson    0.8270696 3.441990e-07
## fruitleg    0.6856106 1.552354e-04
## feculent    0.4511479 2.360026e-02
##
## attr(,"class")
## [1] "condes" "list"
```

3. ACP NON NORMÉE

Lorsque les données sont exprimées en même unité comme dans notre cas, nous pouvons effectuer une ACP non normée, c'est-à-dire sur les données non standardisées. En effectuant une ACP non normée, nous prenons en considération la variabilité liée au volume de différentes variables.

Nous avons vu dans l'analyse préliminaire que les pays consomment plus de protéine issue du lait et des céréales par rapport aux autres sources de protéine. Donc nous pouvons supposer que le lait et les céréales vont recevoir un poids plus important en ACP non normée.

Analyse du premier plan (axe 1 et 2)

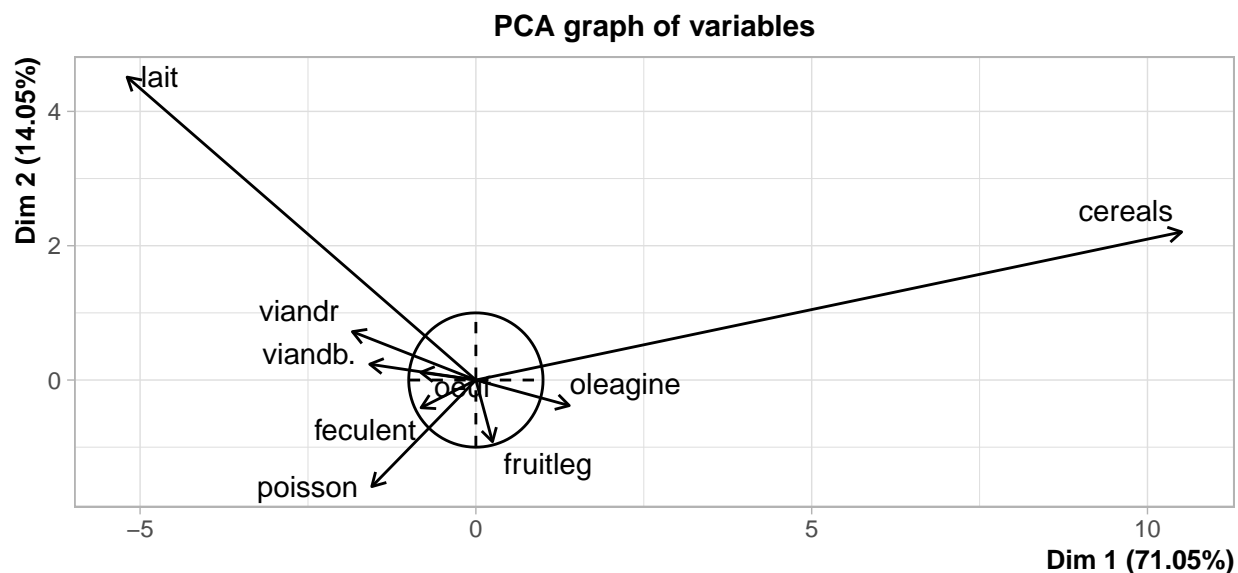
Effectivement, sur le premier plan, nous voyons que les céréales et le lait ont visiblement un impact plus important. La poisson a également un impact important par rapport aux autres variables.

Nous observons que les signes de corrélation de chaque variable avec les axes changent mais les oppositions des axes restent similaires.

C'est-à-dire, l'axe 1 oppose toujours les pays qui consomment le plus de viande, d'oeuf, et de lait et le moins de céréales et d'oléagineux par rapport à la moyenne contre les pays qui consomment le moins de viande, d'oeuf, et de lait et le plus de céréales et d'oléagineux.

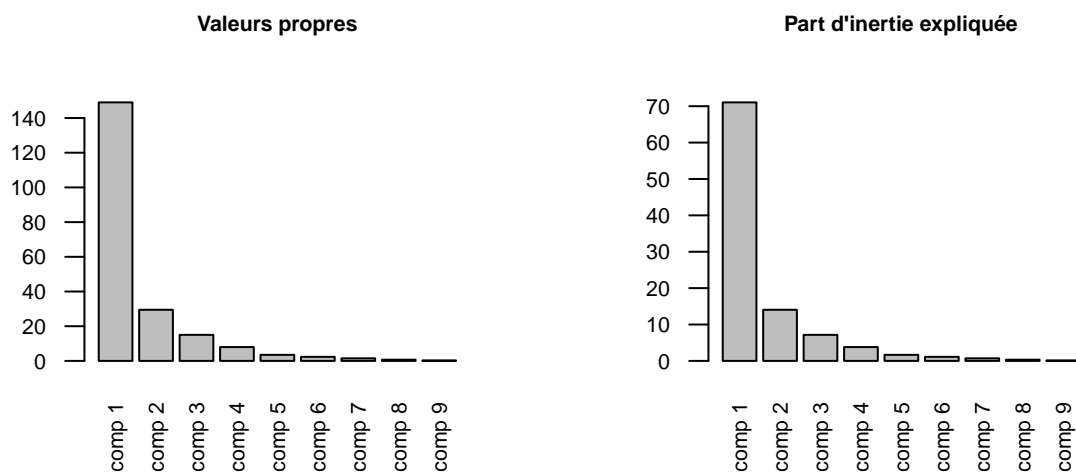
Quant à l'axe 2, il oppose les pays qui consomment le plus de poisson et de fruits & légumes et le moins de lait par rapport à la moyenne contre les pays qui consomment le moins de poisson et de fruits & légumes et le plus de lait par rapport à la moyenne.

La part d'inertie portée par les deux axes est égale à 85,1%. En prenant en compte la variabilité liée au volume des variables, nous arrivons à expliquer plus de variance avec le premier plan.



En effet, nous observons que presque toute la partie de variance, 92,26% est expliquée par les 3 premières valeurs propres et les autres valeurs propres sont presque comme du bruit.

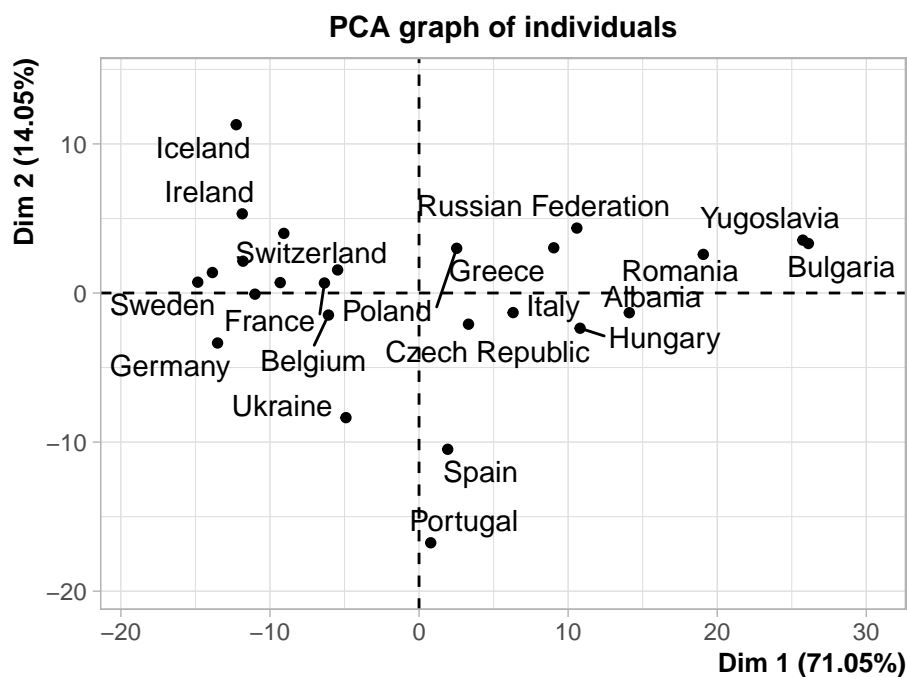
##	eigenvalue	percentage of variance	cumulative percentage of variance
## comp 1	149.0246349	71.0532519	71.05325
## comp 2	29.4707873	14.0513364	85.10459
## comp 3	15.0177185	7.1602774	92.26487
## comp 4	7.9842531	3.8068011	96.07167
## comp 5	3.4847247	1.6614771	97.73314
## comp 6	2.3316395	1.1116992	98.84484
## comp 7	1.4945262	0.7125731	99.55742
## comp 8	0.6862067	0.3271756	99.88459
## comp 9	0.2420531	0.1154082	100.00000



Sur le plan, nous observons que les pays sont plus éloignés les uns aux autres. L'intégration de la variabilité liée au volume différencie plus les pays entre eux.

Sur l'axe 1, la Suède se trouve à l'extrémité alors qu'elle était au milieu dans le cas d'ACP normée. Cela suggère qu'elle consomme plus grande quantité de viande, d'oeuf, de lait et de féculant par rapport à la moyenne. L'Allemagne, le Danemark, et l'Irlande restent toujours les pays qui en consomment le plus par rapport à la moyenne. La Bulgarie et la Yougoslavie restent toujours les pays qui en consomment le moins par rapport à la moyenne.

Sur l'axe 2, l'Islande se trouve à l'extrémité alors qu'elle était au milieu dans le cas d'ACP normée. Nous pouvons supposer que l'Islande consomme la plus grande quantité du lait par rapport à la moyenne. Le Portugal et l'Espagne restent toujours les pays qui consomment plus de poisson et de fruits & légumes par rapport à la moyenne.



Analyse de l'axe 1

En effectuant une ACP non normée, nous avons détecté un nouvel individu remarquable sur l'axe 1, la Suède. Cela suggère qu'elle consomme plus de viande, d'oeuf, et de lait par rapport à la moyenne.

Nous pouvons le vérifier en comparant la consommation suédoise à la moyenne globale. Certes, nous observons que la consommation de viande, d'oeuf et de lait de Suède est plus élevée que la moyenne.

```
##          viandr viandb. oeuf lait poisson cereals feculent oleagine fruitleg
## Sweden      9.9      7.8  3.5 24.7      7.5      19.5      3.7      1.4      2
```

```
## [1] "Moyenne Globale"
```

```
##  viandr viandb.      oeuf      lait poisson cereals feculent oleagine
##    9.828   7.896   2.936   17.112    4.284   32.248    4.276    3.072
## fruitleg
##     4.136
```

Nous avons également détecté que les céréales et les oléagineux avaient plus d'impact par rapport à l'ACP normée. La Bulgarie et la Yougoslavie étant les deux pays qui consomment visiblement le plus de céréales et d'oléagineux par rapport à la moyenne, nous pouvons comparer leur consommation moyenne par rapport à la moyenne globale.

Nous observons bien que ces deux pays consomment visiblement plus de céréales et moins de viande, d'oeuf, et de lait par rapport à la moyenne globale.

```
##  viandr viandb.      oeuf      lait poisson cereals feculent oleagine
##    6.10   5.50   1.40    8.90    0.90   56.30    2.05    4.70
## fruitleg
##     3.70
```

```
## [1] "Moyenne Globale"
```

```
##  viandr viandb.      oeuf      lait poisson cereals feculent oleagine
##    9.828   7.896   2.936   17.112    4.284   32.248    4.276    3.072
## fruitleg
##     4.136
```

Analyse de l'axe 2

En ACP non normée, le lait est une nouvelle variable qui impacte manifestement l'axe 2. Et c'est l'Islande qui est l'individu remarquable lié à la consommation du lait, suggérant qu'elle consomme le plus de lait en quantité par rapport à la moyenne.

Nous pouvons le vérifier en comparant la consommation d'Islande à la moyenne globale. Nous voyons bien que la consommation de lait d'Islande est plus élevée, presque double de la moyenne globale.

```
##          viandr viandb. oeuf lait poisson cereals feculent oleagine fruitleg
## Iceland      9.5      4.9  2.7 33.7      5.8      26.3      5.1      1      1.4
```

```
## [1] "Moyenne Globale"
```



```
##   viandr viandb.   oeuf   lait poisson cereals feculent oleagine
##   9.828   7.896   2.936  17.112   4.284   32.248   4.276   3.072
## fruitleg
##   4.136
```

À l'opposé des pays qui consomment le plus de lait, se trouvent les pays qui consomment le plus de poisson. Comme en ACP normée, l'Espagne et le Portugal sont les pays qui en consomment en plus par rapport à la moyenne.

En comparant la consommation moyenne de Espagne & Portugal avec la moyenne globale, nous pouvons vérifier que l'Espagne et le Portugal consomment plus de lait en moyenne par rapport à la moyenne globale. Ils consomment plus que le double de la moyenne globale.

```
## [1] "Moyenne de Espagne & Portugal"
```

```
##   viandr viandb.   oeuf   lait poisson cereals feculent oleagine
##   6.65   3.55   2.10   6.75   10.60   28.10   5.80   5.30
## fruitleg
##   7.55
```

```
## [1] "Moyenne Globale"
```

```
##   viandr viandb.   oeuf   lait poisson cereals feculent oleagine
##   9.828   7.896   2.936  17.112   4.284   32.248   4.276   3.072
## fruitleg
##   4.136
```

4. Conclusion

L'analyse nous a permis de caractériser la consommation de protéine des pays :

- Les pays de l'Est tels que la Yougoslavie, la Bulgarie, la Roumanie, l'Hongrie, et l'Albanie privilégient la consommation de céréales et d'oléagineux sur la viande, le féculent, le lait, et l'oeuf par rapport à la moyenne des autres pays.
- Les pays situés au Nord-Ouest comme la Suède, le Danemark, l'Irlande et l'Allemagne consomment plus de viande, d'oeuf, de lait et de féculent et moins de céréales et d'oléagineux par rapport à la moyenne.
- Les pays du Sud comme l'Espagne et le Portugal sont de grands consommateurs de poisson et de fruits & légumes par rapport à la moyenne.
- L'Islande consomment beaucoup de lait par rapport à la moyenne.

Donc nous constatons que, dans la globalité, le comportement des pays européens en consommation de protéine est fortement lié à leur situation géographique. De façon intéressante, à part quelques exceptions, la représentation des individus sur le premier plan de l'ACP non normée ressemble beaucoup à leur position géographique réelle sur la carte de l'Europe.

En passant de l'ACP normée à l'ACP non normée, la variabilité liée à la quantité de chaque variable est réintégrée dans l'analyse. En effet, l'ACP normée aurait gommé cette variabilité et agrandi le biais.

Étant donné que l'objectif de l'ACP est de réduire le nombre de variables explicatives en gardant plus d'information sur la variabilité de données, l'ACP non normée se révèle plus satisfaisant et interprétable. Il semble qu'une part significative de la variabilité est expliquée par la différence de volume de chaque variable.