

Exercise 17

Explorative methods

We study data which measure protein consumption in twenty-five European countries for nine food groups. We investigate whether meat consumption is related to that of other foods and whether there are groupings of countries.

Country:	Country name
RdMeat:	Red meat
WhMeat:	White meat
Eggs:	Eggs
Milk:	Milk
Fish:	Fish
Cereal:	Cereals
Starch:	Starchy foods
Nuts:	Pulses, nuts, and oil-seeds
FrVeg:	Fruits and vegetables

For each of the food groups per country the average amount of protein per head per day is given in units of 0.1 gram.

(Reference: Weber, A. (1973) *Agrarpolitik im Spannungsfeld der internationalen Ernährungspolitik*, Institut für Agrarpolitik und Marktlehre, Kiel. Also found in: Gabriel, K.R. (1981) Biplot display of multivariate matrices for inspection of data and diagnosis. In *Interpreting Multivariate Data* (Ed. V. Barnett), New York: John Wiley & Sons, 147-173. Hand, D.J., et al. (1994) *A Handbook of Small Data Sets*, London: Chapman & Hall, 297-298. Downloaded from *The Data and Story Library*, see <http://lib.stat.cmu.edu/DASL/>)

Download the table protein.sav from: http://www.let.rug.nl/~heeringa/statistics/stat03_2013/ and load the table in SPSS.

1. Test whether the food group variables are roughly normally distributed by means of scatter plots and Shapiro-Wilk tests.
2. Perform a Principal Component Analysis. Simultaneously with performing this analysis a scree plot should be created. Use varimax rotation. Save the factor scores as variables, and choose the Anderson-Rubin method. See to it that loadings of 0.4 or higher only are displayed in the tables with the loadings.
3. Look at the table with the communalities. Which variable is the most unique? Which one is the least unique?
4. Look at the table Total Variance explained. How many factors are meaningful? What criterium is used?
5. Look at the scree plot. How many factors (or components) are meaningful?
6. Compare the Component Matrix with the Rotated Component Matrix. Has rotation made the results clearer? Try to interpret the three factors using the Rotated Component Matrix.
7. Create a scatter plot on the basis of the factor scores, for each pair of factor scores (Factor 1 vs.

Factor 2, Factor 1 vs. Factor 3, Factor 2 vs. Factor 3). Try to interpret them.

8. Perform a metric PROXSCAL Multidimensional Scaling. Create proximities from data. Use Euclidean distance as a measure. Use Ratio as Proximity Transformations. Scale to two dimensions. How much variance in the original distances is explained by the two dimensions?

9. Perform a non-metric PROXSCAL Multidimensional Scaling. Create proximities from data. Use Euclidean distance as a measure. Use Ordinal as Proximity Transformations. Scale to two dimensions. How much variance in the original distances is explained by the two dimensions?

10. Compare the two plots. Which model does explain the data better?

11. Perform hierarchical cluster analysis. The cases should be labeled by Country. See to it that a dendrogram will be generated. Use Between-groups linkage, and use Euclidean distance as interval measure.

12. Look at the dendrogram. What groupings do you find?

13. Repeat the same analysis, but now use Ward's method. Compare the dendrogram with the previous one. Which one looks most plausible?