1. Determine the population principal component Y and Y for the covariance matrix

also calculate the proportion of the total variance explained by the first principal component.

1. Let X1,X2. …,, X5 denote observed weekly rates of return for JP Morgan, Citybank, Wells Fargo, Royal Dutch Shell and ExxonMobil respectively.



1. Compute the eigenvalues and eigenvectors for this correlation matrix
2. How many PCs to be retained?
3. What percentage of total variance is explained by the first PC?

Install packages

Psych, MASS and parallel

1. This exercise aims to develop a Consumer Price Index (CPI) of US based on the estimated retail food prices by cites. For each of 23 cities, the prices of five kinds of food are measured. The developed CPI should reflect most information given by the five prices, of Bread Burger Milk Oranges Tomatoes, but contain a lower number of variables;

The data, which are reported by US Bureau of Labour Statistics (1973), and is given as a text file with file “LabourStat1973.csv”.

* 1. Load the dataset into R
  2. Compute the principal components from the covariance matrix (mean-correlated data).

1. How many PCs to be retained
2. What percentage of total variance is explained by the first PC?
3. Obtain scree plot and discuss.
4. The data for this exercise come from 200 specimens of a certain type of crabs which is avialabe the MASS package. The crabs come in two colours (blue and orange). In the experiment 100 of each type were collected, 50 males and 50 females, and for each of the 200 crabs, five quantities were measured: The carapace/shell length (CL), carapace/shell width (CW), size of frontal lobe (FL), rear width (RW), and body depth (BD).
   1. Install MASS package and load the dataset.
   2. Compute the principal components from the covariance matrix (mean-correlated data).
   3. How many PCs to be retained
   4. What percentage of total variance is explained by the first PCA?
   5. Obtain scree plot and discuss.
5. The data ” MAGIC Gamma Telescope Data Set”are MC generated to simulate registration of high energy gamma particles in a ground-based atmospheric Cherenkov gamma telescope using the imaging technique. Cherenkov gamma telescope observes high energy gamma rays, taking advantage of the radiation emitted by charged particles produced inside the electromagnetic showers initiated by the gammas, and developing in the atmosphere. This Cherenkov radiation (of visible to UV wavelengths) leaks through the atmosphere and gets recorded in the detector, allowing reconstruction of the shower parameters. The available information consists of pulses left by the incoming Cherenkov photons on the photomultiplier tubes, arranged in a plane, the camera. Depending on the energy of the primary gamma, a total of few hundreds to some 10000 Cherenkov photons get collected, in patterns (called the shower image), allowing to discriminate statistically those caused by primary gammas (signal) from the images of hadronic showers initiated by cosmic rays in the upper atmosphere (background).   
     
   The data set was generated by a Monte Carlo program, Corsika, described in:   
   D. Heck et al., CORSIKA, A Monte Carlo code to simulate extensive air showers,   
   The variables are

fLength: continuous # major axis of ellipse [mm]   
fWidth: continuous # minor axis of ellipse [mm]   
fSize: continuous # 10-log of sum of content of all pixels [in #phot]   
fConc: continuous # ratio of sum of two highest pixels over fSize [ratio]   
fConc1: continuous # ratio of highest pixel over fSize [ratio]   
fAsym: continuous # distance from highest pixel to center, projected onto major axis [mm]   
fM3Long: continuous # 3rd root of third moment along major axis [mm]   
fM3Trans: continuous # 3rd root of third moment along minor axis [mm]   
fAlpha: continuous # angle of major axis with vector to origin [deg]   
fDist: continuous # distance from origin to center of ellipse [mm]   
class: g,h # gamma (signal), hadron (background)

* 1. Load the dataset “MTD.csv”
  2. Compute the principal components from the covariance matrix (mean-correlated data).
  3. How many PCs to be retained
  4. What percentage of total variance is explained by the first PCA?
  5. Obtain scree plot and discuss.