

How to use ClustersAnalysis Package

Contents

Introduction	1
Short Descriptions of datasets	1
Iris :	1
Infert :	2
Esoph :	2
Import ClustersAnalysis from Github (using devtools)	2

Introduction

This is a demonstration of using the R package Clusters. You will see how to analyze classes according to one or more variables. The group variable must be of the type factor or character and the exploratory variables can be quantitative or qualitative. In this demonstration we are going to use natives dataset from R such as “iris”, “infert” or “esoph”.

Short Descriptions of datasets

Iris : The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other.

```
summary(iris)
```

```
##   Sepal.Length   Sepal.Width   Petal.Length   Petal.Width
## Min.    :4.300   Min.    :2.000   Min.    :1.000   Min.    :0.100
## 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
## Median :5.800   Median :3.000   Median :4.350   Median :1.300
## Mean    :5.843   Mean    :3.057   Mean    :3.758   Mean    :1.199
## 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
## Max.    :7.900   Max.    :4.400   Max.    :6.900   Max.    :2.500
##      Species
## setosa    :50
## versicolor:50
## virginica :50
##
##
##
```

Infert : This is a matched case-control study dating from before the availability of conditional logistic regression.

```
summary(infert)
```

```
##      education      age      parity      induced
## 0-5yrs : 12   Min.    :21.00   Min.    :1.000   Min.    :0.0000
## 6-11yrs:120   1st Qu.:28.00   1st Qu.:1.000   1st Qu.:0.0000
## 12+ yrs:116   Median :31.00   Median :2.000   Median :0.0000
##              Mean    :31.50   Mean    :2.093   Mean    :0.5726
##              3rd Qu.:35.25   3rd Qu.:3.000   3rd Qu.:1.0000
##              Max.    :44.00   Max.    :6.000   Max.    :2.0000
##      case      spontaneous      stratum      pooled.stratum
## Min.    :0.0000   Min.    :0.0000   Min.    : 1.00   Min.    : 1.00
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:21.00   1st Qu.:19.00
## Median :0.0000   Median :0.0000   Median :42.00   Median :36.00
## Mean    :0.3347   Mean    :0.5766   Mean    :41.87   Mean    :33.58
## 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:62.25   3rd Qu.:48.25
## Max.    :1.0000   Max.    :2.0000   Max.    :83.00   Max.    :63.00
```

Esoph :

Data from a case-control study of (o)esophageal cancer in Ille-et-Vilaine, France. This is a data frame with records for 88 age/alcohol/tobacco combinations.

```
summary(esoph)
```

```
##      agegp      alcgp      tobgp      ncases      ncontrols
## 25-34:15   0-39g/day:23   0-9g/day:24   Min.    : 0.000   Min.    : 1.00
## 35-44:15   40-79      :23   10-19      :24   1st Qu.: 0.000   1st Qu.: 3.00
## 45-54:16   80-119      :21   20-29      :20   Median : 1.000   Median : 6.00
## 55-64:16   120+        :21   30+        :20   Mean    : 2.273   Mean    :11.08
## 65-74:15                                     3rd Qu.: 4.000   3rd Qu.:14.00
## 75+      :11                                     Max.    :17.000   Max.    :60.00
```

Import ClustersAnalysis from Github (using devtools)

```
#Use the below line to install devtools if necessary
install.packages("devtools")
library(devtools)

install.packages("devtools")
devtools::install_github("clepadellec/ClustersAnalysis")

load package
library(ClustersAnalysis)

ClustersAnalysis::u_plot_size_effect(Univariate_object(esoph,1)),2)

ClustersAnalysis::u_sil_pca_plot(Univariate_object(iris,5),interact = FALSE)
```

