

The data set (and description) can be downloaded here:

<http://cran.r-project.org/web/packages/rrcov/>

### Description:

#### Hemophilia Data

##### Description

The hemophilia data set contains two measured variables on 75 women, belonging to two groups:  $n_1=30$  of them are non-carriers (normal group) and  $n_2=45$  are known hemophilia A carriers (obligatory carriers).

##### Usage

```
data(hemophilia)
```

##### Format

A data frame with 75 observations on the following 3 variables.

AHFActivity

AHF activity

AHFantigen

AHF antigen

gr

group - normal or obligatory carrier

##### Details

Originally analyzed in the context of discriminant analysis by Hebemma and Hermans (1974). The objective is to find a procedure for detecting potential hemophilia A carriers on the basis of two measured variables:

$x_1 = \log_{10}(\text{AHV activity})$  and  $x_2 = \log_{10}(\text{AHV-like antigen})$ . The first group of  $n_1=30$  women consists of known non-carriers (normal group) and the second group of  $n_2=45$  women is selected from known hemophilia A carriers (obligatory carriers). This data set was also analyzed by Johnson and Wichern (1998) as well as, in the context of robust Linear Discriminant Analysis by Hawkins and McLachlan (1997) and Hubert and Van Driessen (2004).

##### Source

Habemma, J.D.F, Hermans, J. and van den Broek, K. (1974) Stepwise Discriminant Analysis Program Using Density Estimation in Proceedings in Computational statistics, COMPSTAT`1974 (Physica Verlag, Heideberg, 1974, pp 101-110).

##### References

Johnson, R.A. and Wichern, D. W. Applied Multivariate Statistical Analysis (Prentice Hall, International Editions, 2002, fifth edition)

Hawkins, D. M. and McLachlan, G.J. (1997) High-Breakdown Linear Discriminant Analysis J. Amer. Statist. Assoc. 92 136-143.

Hubert, M., Van Driessen, K. (2004) Fast and robust discriminant analysis, Computational Statistics and Data Analysis, 45 301-320.

### Descriptive statistics:

Dataset= hemophilia : n= 75 , d= 2

Class1: n= 30

#### Covariance matrix:

```
      [,1] [,2]
[1,] 0.0209 0.0155
[2,] 0.0155 0.0179
```

#### Correlation matrix:

```
      [,1] [,2]
[1,] 1.0000 0.8017
[2,] 0.8017 1.0000
```

Median: -0.1269 -0.0681  
 Mean: -0.1349 -0.0779  
 MCD-estimated:  
 MDC-0.975-Mean: -0.1292 -0.0603  
 MDC-0.750-Mean: -0.1292 -0.0603  
 MDC-0.500-Mean: -0.1292 -0.0603

Class2: n= 45

Covariance matrix:  
 [,1] [,2]  
 [1,] 0.0238 0.0154  
 [2,] 0.0154 0.0240

Correlation matrix:  
 [,1] [,2]  
 [1,] 1.0000 0.6431  
 [2,] 0.6431 1.0000

Median: -0.3049 -0.0018  
 Mean: -0.3079 -0.006  
 MCD-estimated:  
 MDC-0.975-Mean: -0.3079 -0.006  
 MDC-0.750-Mean: -0.3079 -0.006  
 MDC-0.500-Mean: -0.3079 -0.006

Measures:  
 Mah.Dist: 2.1388  
 Mah.Dist-MCD-0.975: 2.0689  
 Mah.Dist-MCD-0.750: 2.0689  
 Mah.Dist-MCD-0.500: 2.0689

All the MCD estimates have been obtained after a slight perturbation of the data set

