2013/09/01

How to port BECON-EEM for S-Plus to R

I. Get the source codes for S-plus.

The paper including the codes are as follows:

Béguin, Cédric,, Beat Hulliger (2003). Robust multivariate outlier detection and imputation with incomplete survey data, EUREDIT Deliverable D4/5.2.1/2 Part C, pp.176-187. [http://www.cs.york.ac.uk/euredit/results/Results/Robust/Part%20C.zip]

Download the zip file, unzip it, and copy the codes for BECON-EEM(BEM) algorithm at pages 176 to 187 including comments and paste them on a text editor (the one which shows the line numbers are desirable). The number of total lines are 767 at this stage.

II. Remove the page header and fotter.

A page footer (page number) and a header ("ROBUST MULTIVARIATE OUTLIER DETECTION AND IMPUTATION") are included in each page. Remove them. The number of lines become 745, after 11 pages x 2 lines are removed.

III. Some characters might be corrupted probably depending on environment. Check it by saving the codes, as "BEM.r" for example, and loading it to R by 'source("c:/R/BEM2.r").

In case of Japanese Windows environment, the character "^" is corrupted and the following corrections are needed. The figures in the belackets are line numbers.

[86] return(sum(w*(x-mean)^2)/(sum(w)-1))
[296] apply(sweep(data,2,EM.mean)^2,1,sum,na.rm=T)*p/(p-apply(is.na(data),1,sum))
[492] test <- (c.np+c.hr)^2*chi.sq

IV. Modify the codes as follows:

```
[44]  # pos <- compare(cumsum(w),sum(w)/2)
                          ⇩        (replace)
      pos <- rep(0, n)
      pos[c(cumsum(w) > sum(w)/2)] <- 1
      pos[c(cumsum(w) < sum(w)/2)] <- -1
```

```
[56]  ############ Weighted variance/covariance matrix ############
[57]  #
[58]  # Does not accept missing values
[59]  #
                          ⇩        (insert the following line right after the comments above)
      oka.kakeru <- function(x,w) x*w                  # restrain memory usage
```

```
[64]  # return((t(sweep(x,2,mean))%*%diag(w)%*%sweep(x,2,mean))/(sum(w)-1))
                          ⇩        (replace)
      xs <- apply(sweep(x,2,mean), 2, oka.kakeru, w=w)
      return((t(xs)%*%sweep(x,2,mean))/(sum(w)-1))
```

Porting: OKAMOTO, Masato Documentation: WADA, Kazumi