

## S-plus 版 BECON-EEM の R への移植方法

S-plus 版のソースコード:

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

BECON-EEMのコードは、PDFファイルの176～187ページ。

- ① メモ帳など R で使うエディタ（行番号が表示されるものが望ましい）に該当部分のコードをコメント行なども含めてすべてコピーペーストする。総行数は 767。
- ② ページのフッターのページ番号(176-186)とヘッダーの" ROBUST MULTIVARIATE OUTLIER DETECTION AND IMPUTATION"が並んでページ毎に入っているので取り除く。作業後の総行数は、11 ページ×2 行減って 745。
- ③ 以下の場所の二乗の記号「^」が文字化けしていれば修正する。[xx]は行番号を示す。

```
[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
```

※ ファイルをたとえば C ドライブの R というディレクトリに"BEM.r"という名前で保存し、R の source 関数で以下のように読み込むと、その際の文法チェックでこれらの箇所がエラー表示される。

```
source("c:/R/BEM2.r")
```

- ④ この他、R と S-plus での相違により、以下の三か所を修正する。

```
[44] # pos <- compare(cumsum(w),sum(w)/2)
```

↓ (書き換え)

```
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] #
```

↓ (直後に追加)

```
oka.kakeru <- function(x,w) x*w # メモリ使用抑制用
```

```
[64] # return((t(sweep(x,2,mean))%*%diag(w)%*%sweep(x,2,mean))/(sum(w)-1))
```

↓ (書き換え)

```
xs <- apply(sweep(x,2,mean), 2, oka.kakeru, w=w)
return((t(xs)%*%sweep(x,2,mean))/(sum(w)-1))
```