

```
1   ###階層クラスター分析
2   library(MASS)
3   library(psych)
4   library(dplyr)
5   data("Cars93")
6   colnames(Cars93)
7   ###欠測値処理(リストワイズ削除したデータを作成する)
8   anyNA(Cars93) #TRUE;NAあり
9   sum(is.na(Cars93))
10  na <- subset(Cars93 , complete.cases(Cars93) == F) #NA,Itemsを確認
11  #このデータはNAがある.内容を確認する
12  #NA.omitして良いか.OKなら以下の作業をする
13  #na.omit <- na.omit(Cars93) #ある場合NAをomitしたファイル作成
14  #nrow(na.omit) ##リストワイズ後のサンプル数
15  #write.csv(na.omit,"data/Cars93.na.omit.csv") #削除後を保存
16
17  ##基本統計量の計算
18  describe(Cars93) ###NA;omitして計算
19  ##階層的クラスター分析;距離行列を分析
20  ##Distance
21  c.dist <- dist(Cars93[,c("Price","MPG.city","EngineSize","RPM",
22                          "Length","Turn.circle","Weight")])
23  #c.dist2 <- dist(c.dat[,c(4,6,11,13,18,21,24)])
24  #method:single,complet,average,centroid,median,ward.D2,mcquitty
25  ward.h <- hclust(c.dist , method = "ward.D2") #ward法
26  plot(ward.h , hang = -1 , main = "Ward" ,
27       labels = Cars93$Model) #labels:列name
28  ward.h$order      #クラスタリング後のサンプルの順番
29  rect.hclust(ward.h , k = 3 , border = "red") #3CLに切る
30  rect.hclust(ward.h , k = 5 , border = "blue") #5CLに切る
31
32  comp.h <- hclust(c.dist , method = "complet") #最遠隣法
33  plot(comp.h , hang = -1 , main = "C-Link",      #デンドログラム
34       labels = Cars93$Model)
35  rect.hclust(comp.h , k = 3 , border = "red") #3CLに切る
36  rect.hclust(comp.h , k = 5 , border = "blue") #5CLに切る
37
38
39  ##cutoffreeを使ってクラスターを保存する
40  ward.h <- hclust(c.dist , method = "ward.D2")
41  plot(ward.h , hang = -1 , main = "Ward")
42  rect.hclust(ward.h , k = 3 , border = "red") #3CLに切る
43  cluster <- cutree(ward.h, k=3) # K=でいくつのクラスターに分けるか指定
44  cluster <- factor(cluster)    # 因子の型に変更
45  table(cluster)                # それぞれのクラスターの人数確認
46
47
```