階層的クラスター分析

|  |
| --- |
| 記述統計的なコマンドは心理系のところに多そうだから"psych"を読み込むlibrary(psych)library(stats)library(proxy)#データを読み込む。environmentのimport data setを使う。次に以下のようにdatとdata frameに入れる。dat<-ptdat#後のことも考えて標準化する。この場合、データが数値データであれば、(data-平均)/標準偏差 だから以下のコマンドで標準化できるndata<-scale(dat)#コンピュータの計算能力を超えるので、数字を丸めるcdata<-100\*(ndata)rdata<-round(cdata)rdata#後で見たくなりそうだから、総当たり相関行列も作る。cordat<-cor(rdata)write.table(cordat,"ptcor.csv",sep=",")#マハラノビス距離の計算mdist<-proxy::dist(rdata, method="mahalanobis")edc<-hclust(mdist,method="ward.D")edcplot(edc,hang=-1,main="mahalanobis-ward")#念のためユークリッド距離によってデンドログラムをつくる。クラスターの結合はWard法を用いる。データが標準化してあっても、項目間の相関性があれば、マハラノビス距離の結果と違うはず。edist<-dist(rdata)edc<-hclust(edist,method="ward.D")plot(edc,hang=-1,main="euclid-ward")#データー分布を視覚化してみたいので、古典的な多次元尺度法で２次元プロットする。mcmd<-cmdscale(mdist)plot(mcmd) text(mcmd)#クラスターの仕分け(mclster<-cutree(edc,k=6))(eclster<-cutree(edc,k=6))#データの書き出しwrite.table(mclster,"clustermah.csv",sep=",")write.table(eclster,"clustereuc.csv",sep=",")#非階層的クラスタ分析kmeans(mdist, 6, iter.max = 200, nstart = 1, algorithm = c("Hartigan-Wong"))mclsep<-kmeans(mdist, 6, iter.max = 100, nstart = 1, algorithm = c("Hartigan-Wong"))#データの書き出しwrite.table(mclsep$cluster,"categorymah.csv",sep=",") |