

```
#####
##第七回「主成分分析」
#####

#####
#ボストンハウジングデータ
x <- read.table("housing_table.data", header=T)
names(x)      #14変数
summary(x)
plot(x)        #散布図のプロット

x <- scale(x)   #標準化
colMeans(x)    #全変数の平均0
diag(var(x))   #全変数の分散1

Sigma = cov(x) #分散共分散行列を計算.
Sigma[1:5, 1:5]

res <- eigen(Sigma) #分散共分散行列を固有値固有ベクトル変換
res

norm(Sigma - res$vectors %*% diag(res$values) %*% t(res$vectors)) #確認
norm(Sigma %*% res$vectors - res$vectors %*% diag(res$values) ) #確認
tmp <- res$vectors %*% t(res$vectors); tmp[1:5, 1:5] #直交性の確認
tmp <- t(res$vectors) %*% res$vectors; tmp[1:5, 1:5] #直交性の確認

#主成分分析
Lam <- res$values
V <- res$vectors
rownames(V) <- colnames(x) #変数に名前をつける.
colnames(V) <- paste("PC", seq(1, length(res$values)))

y = x %*% V #主成分スコアの計算
plot(y[, 1], y[, 2], type='n') #第一, 第二主成分スコアをプロット
text(y[, 1], y[, 2], seq(length=nrow(y)), cex=0.5) #サンプル番号を表示

biplot(y[, 1:2], V[, 1:2], cex=0.5) #第一, 第二主成分スコア
biplot(y[, 1:3], V[, 1:3], cex=0.5) #第一, 第三主成分スコア

round(cov(y), 10) #yは無相関

# Rの組み込み関数を使ってみる.
PCA <- princomp(x, cor=TRUE) #データの標準化をするにはcor=TRUE
biplot(PCA)
biplot(PCA, scale = 0) #scale=0とすると先と同じプロットを得る.

#####
# 統計で見る都道府県のすがた【自然環境】
kenkan <- read.csv("ken-kankyo-kakou.csv", header=TRUE)
PCA.nature <- princomp(kenkan, cor=TRUE)
biplot(PCA.nature) #都道府県の自然環境から見た特徴が見えてくる.
```