#######################################################################################

#

# 階層型クラスター分析スクリプト エストレーラ2021年3月号の「研究ノート」用

#

# SSDSE-2020c をCSVファイルでダウンロードしたものから、前処理を行い「標準化データA」に

# 変換し、ward法クラスター分析を行うスクリプト

# 加えて、SSDSE-2020C.csvファイルから、12中分類項目による同様な分析も行う。

# 最終修正版 2020/12/28

#######################################################################################

library(dplyr)

######################

# SSDSE-2020C.csv から小分類項目だけを抜き出し「標準化データA」に変換する前処理

ssdse\_2020c <- read.csv("ssdse-2020C.csv")

ssdse\_2020c\_small <- select(ssdse\_2020c, -X2017\_2019, -City, -LA03, -LB01, -LB02, -LB03,

-LB04, -LB05, -LB06, -LB07, -LB08, -LB09, -LB10, -LB11, -LB12)

ssdse\_2020c\_small\_pref <- filter(ssdse\_2020c\_small, Prefecture != "全国")

colnames(ssdse\_2020c\_small\_pref) <- ssdse\_2020c\_small\_pref[1,]

ssdse\_2020c\_edited\_pref <- filter(ssdse\_2020c\_small\_pref, 都道府県!="都道府県")

rownames(ssdse\_2020c\_edited\_pref) <- c(ssdse\_2020c\_edited\_pref$都道府県)

ssdse <- select(ssdse\_2020c\_edited\_pref, -都道府県)

### 全て文字型となってしまっているので、いったんCSVファイルに""抜きで書き出して改めて読み込む

write.csv(ssdse, file="ssdse\_2020c\_edited.csv", quote = FALSE)

ssdse <- read.csv("ssdse\_2020c\_edited.csv", row.name = 1)

### 「標準化データA」を作成する

ssdse.p <- (ssdse %>% select(米:学校給食) %>% sweep(.,1,c(as.numeric(ssdse[,1])),"/")\*100)

ssdse.std <- scale(ssdse.p) # 「標準化データA」 (ssdse.std)

######################

# ward法クラスター分析（Murtagh&Legendre(2014)による適正な初期測度を用いる）

ssdse.dist <- (1/2)\* dist(ssdse.std)

methname <- "ward.D2" # method を指定

cluster\_num.p <- 12 # クラスター数を指定

ssdse.hclust <- hclust(ssdse.dist, method=methname)

plot(ssdse.hclust)

rect.hclust(ssdse.hclust, k=cluster\_num.p)

######################

# SSDSE-2020C.csv から中項目分類だけを抜き出し「標準化データ」に変換する前処理

ssdse\_2020c\_middle <- select(ssdse\_2020c, Prefecture, LB00, LB01, LB02, LB03, LB04, LB05, LB06, LB07,

LB08, LB09, LB10, LB11, LB12)

ssdse\_2020c\_middle\_pref <- filter(ssdse\_2020c\_middle, Prefecture != "全国")

colnames(ssdse\_2020c\_middle\_pref) <- ssdse\_2020c\_middle\_pref[1,]

ssdse\_2020c\_middle\_edited\_pref <- filter(ssdse\_2020c\_middle\_pref, 都道府県!="都道府県")

rownames(ssdse\_2020c\_middle\_edited\_pref) <- c(ssdse\_2020c\_middle\_edited\_pref$都道府県)

ssdse\_middle <- select(ssdse\_2020c\_middle\_edited\_pref, -都道府県)

### 全て文字型となってしまっているので、いったんCSVファイルに""抜きで書き出して改めて読み込む

write.csv(ssdse\_middle, file="ssdse\_2020c\_middle\_edited.csv", quote = FALSE)

ssdse\_middle <- read.csv("ssdse\_2020c\_middle\_edited.csv", row.name = 1)

### 「標準化データ」を作成する

ssdse\_middle.p <- (ssdse\_middle %>% select(X01.穀類:X12.外食) %>%

sweep(.,1,c(as.numeric(ssdse\_middle[,1])),"/")\*100)

ssdse\_middle.std <- scale(ssdse\_middle.p) # 「標準化データ」 (ssdse\_middle.std)

######################

# ward法クラスター分析（Murtagh&Legendre(2014)による適正な初期測度を用いる）

ssdse\_middle.dist <- (1/2)\* dist(ssdse\_middle.std)

ssdse\_middle.hclust <- hclust(ssdse\_middle.dist, method=methname)

plot(ssdse\_middle.hclust)

rect.hclust(ssdse\_middle.hclust, k=cluster\_num.p)