「陽明文庫所蔵史料における料紙構成物の変化：統計ソフトウェアRを用いた可視化の試み」関連データ

渋谷綾子・尾上陽介・高島晶彦・天野真志・野村朋弘・山田太造・後藤真

　このファイルは、「陽明文庫所蔵史料における料紙構成物の変化：統計ソフトウェアRを用いた可視化の試み」に使用したRマークダウンのコードである。  
　図5は、現生標本（イネ、アワ、キビ、ヒエ）と陽明文庫社蔵史料で確認された料紙のデンプン粒（イネ、トロロアオイ、種不明）について粒径の比較・検討を行い、それぞれの特徴を可視化した。デンプン粒の粒径範囲は標本によって左右されるが（藤本1994、註50)論文）、現生標本は渋谷（2010、註29)論文）で計測したデータ（任意で20個抽出）にもとづくものである。陽明文庫所蔵史料の料紙のデンプン粒は，調査史料90点の撮影箇所における計測結果を用いており，イネ329個（函番号11：89個，函番号47：223個，函番号132：17個），トロロアオイ111個（函番号11：49個，函番号47：42個，函番号47：20個），種不明3個（函番号11のみ）である。  
　図6～8は、構成物のデンプン粒、鉱物、細胞組織、繊維に対する主成分分析の結果を示し、図9は料紙面積と構成物（合計）の相関分析の結果である。さらに、各料紙構成物に対する因子分析のコードを示す。これらの因子分析の結果は、本文中で説明している。

# パッケージの読み込み  
library(ggplot2)  
library(readr)  
library(tidyverse)  
library(knitr)  
library(rmarkdown)  
library(revealjs)  
library(scales)  
library(reshape2)

# 現生デンプン粒標本と料紙のデンプン粒の比較

## 陽明文庫所蔵史料の料紙に含有されたデンプン粒の特徴

# 図5作成のためのCSVファイルの読み取り  
starch <- read\_csv("yomei-starch.csv")  
  
head(starch)　 # データフレームの上6行を表示

# A tibble: 6 x 2  
 デンプン粒の種類 粒径範囲  
 <chr> <dbl>  
1 現生アワ 10   
2 現生アワ 11.5   
3 現生アワ 7.69  
4 現生アワ 8.46  
5 現生アワ 11.5   
6 現生アワ 7.69

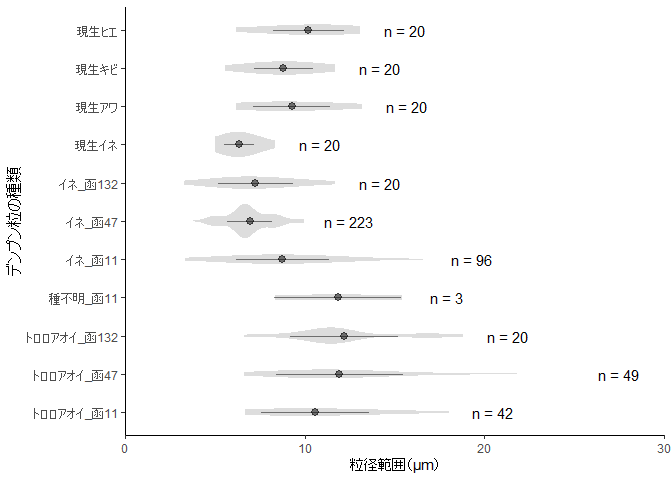
names(starch) # starchに含まれるすべての変数名

[1] "デンプン粒の種類" "粒径範囲"

dim(starch)　 # starchに含まれる観測数と変数の数を表示させる

[1] 533 2

n\_fun <- function(x){  
return(data.frame(y = max(x)+2.5, label = paste0("n = ",length(x))))  
}  
  
ggplot(starch, aes(x = デンプン粒の種類, y = 粒径範囲)) +  
 geom\_violin(trim=T,fill="#999999",linetype="blank",alpha=I(1/3),adjust=2.5)+ # バイオリンプロット作成  
 stat\_summary(geom="pointrange",fun = mean, fun.min = function(x) mean(x)-sd(x),   
 fun.max = function(x) mean(x)+sd(x), size=.5,alpha=.5)+ # 平均値±標準偏差をプロット  
 stat\_summary(fun.data = n\_fun, geom = "text",colour="black",size=4)+ # 各グループのデータ数を最大値の位置に追加  
 scale\_y\_continuous(breaks = c(0,10,20,30), limits = c(0,30), expand = c(0,0))+ # 数値軸の目盛りを指定  
 scale\_x\_discrete(limit=c("トロロアオイ\_函11","トロロアオイ\_函47","トロロアオイ\_函132","種不明\_函11","イネ\_函11","イネ\_函47","イネ\_函132","現生イネ","現生アワ","現生キビ","現生ヒエ")) + # 文字軸の順番を指定  
 coord\_flip() +　# 90度横向きにする  
 labs(x = "デンプン粒の種類", y = "粒径範囲（μm）") + # ラベルの指定  
 theme\_classic()



ggsave(file = "fig5.png", dpi = 300, width=5,height=6) # ファイルの保存

# 陽明文庫所蔵史料の料紙構成物に対する主成分分析

tbs3 <- read\_csv("yomei-compo.csv") # CSVファイルの読み取り  
tbs3 # 読み込んだデータ

# A tibble: 89 x 7  
 番号 紙素材 デンプン粒 鉱物 細胞組織 繊維 ほか  
 <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
 1 11-391 コウゾ 20 3 97 1 28  
 2 11-392 コウゾ 0 0 62 0 0  
 3 11-393 コウゾ 5 1 119 0 0  
 4 11-394 コウゾ 0 0 74 0 10  
 5 11-396 コウゾ 0 1 108 0 0  
 6 11-397 コウゾ 0 1 58 1 0  
 7 11-398 コウゾ 0 0 91 0 0  
 8 11-399 コウゾ 0 0 48 0 10  
 9 11-400 コウゾ 81 4 95 0 0  
10 11-401 コウゾ 0 0 19 0 5  
# ... with 79 more rows

# 構成物の種類を実数型に変換  
tbs4 <-  
 tbs3 %>%  
 filter(紙素材 %in% "コウゾ") %>%　　　 # コウゾだけを選択  
 select(デンプン粒,鉱物,細胞組織,繊維,ほか) %>%  
 mutate(  
 デンプン粒 = as.numeric(デンプン粒), # デンプン粒を実数に変換  
 鉱物 = as.numeric(鉱物),　　　　　　 # 鉱物を実数に変換  
 細胞組織 = as.numeric(細胞組織),　　 # 細胞組織を実数に変換  
 繊維 = as.numeric(繊維),　　　　　　 # 繊維を実数に変換  
 ほか = as.numeric(ほか))　　　　　　 # ほか（他の物質）を実数に変換  
  
# 主成分分析を行うパッケージFactoMineRを読み込み，主成分分析を実行  
library(FactoMineR)  
# 主成分分析を実行  
res.pca <-   
 PCA(tbs4,graph = FALSE)  
  
# 多変量解析の可視化に特化したfactoextraパッケージ  
library(factoextra)   
# 各主成分の寄与率を描画  
fviz\_screeplot(res.pca)



ggsave(file = "fig6.png", width = 6, height = 4, dpi = 300) # ファイルの保存  
  
# 主成分分析の概要を表示  
summary(res.pca)

Call:  
PCA(X = tbs4, graph = FALSE)   
  
  
Eigenvalues  
 Dim.1 Dim.2 Dim.3 Dim.4 Dim.5  
Variance 1.377 1.266 0.930 0.838 0.589  
% of var. 27.532 25.325 18.609 16.760 11.774  
Cumulative % of var. 27.532 52.857 71.466 88.226 100.000  
  
Individuals (the 10 first)  
 Dist Dim.1 ctr cos2 Dim.2 ctr cos2 Dim.3 ctr  
1 | 3.623 | 2.731 6.374 0.568 | 1.492 2.069 0.170 | 1.497 2.835  
2 | 0.903 | -0.761 0.495 0.710 | -0.239 0.053 0.070 | -0.400 0.202  
3 | 1.461 | -0.634 0.344 0.188 | 1.145 1.218 0.614 | 0.638 0.514  
4 | 0.895 | -0.384 0.126 0.184 | -0.326 0.099 0.133 | 0.149 0.028  
5 | 1.298 | -0.735 0.462 0.321 | 0.912 0.773 0.494 | 0.515 0.335  
6 | 1.294 | 0.302 0.078 0.055 | -0.203 0.038 0.025 | 0.137 0.024  
7 | 1.036 | -0.965 0.796 0.867 | 0.048 0.002 0.002 | 0.155 0.030  
8 | 1.015 | -0.201 0.035 0.039 | -0.583 0.316 0.330 | -0.348 0.153  
9 | 6.146 | 3.290 9.250 0.287 | 4.873 22.063 0.629 | -1.045 1.380  
10 | 1.419 | -0.228 0.045 0.026 | -0.767 0.547 0.292 | -1.063 1.428  
 cos2   
1 0.171 |  
2 0.196 |  
3 0.190 |  
4 0.028 |  
5 0.157 |  
6 0.011 |  
7 0.022 |  
8 0.118 |  
9 0.029 |  
10 0.561 |  
  
Variables  
 Dim.1 ctr cos2 Dim.2 ctr cos2 Dim.3 ctr cos2  
デンプン粒 | 0.650 30.655 0.422 | 0.432 14.755 0.187 | -0.262 7.365 0.069  
鉱物 | 0.396 11.373 0.157 | 0.756 45.131 0.571 | 0.032 0.112 0.001  
細胞組織 | -0.345 8.627 0.119 | 0.466 17.117 0.217 | 0.771 63.929 0.595  
繊維 | 0.541 21.281 0.293 | -0.470 17.409 0.220 | 0.375 15.128 0.141  
ほか | 0.622 28.064 0.386 | -0.266 5.588 0.071 | 0.354 13.465 0.125  
   
デンプン粒 |  
鉱物 |  
細胞組織 |  
繊維 |  
ほか |

res.pca$eig %>%  
 kable()

|  |  |  |  |
| --- | --- | --- | --- |
|  | eigenvalue | percentage of variance | cumulative percentage of variance |
| comp 1 | 1.3766096 | 27.53219 | 27.53219 |
| comp 2 | 1.2662437 | 25.32487 | 52.85707 |
| comp 3 | 0.9304291 | 18.60858 | 71.46565 |
| comp 4 | 0.8380248 | 16.76050 | 88.22614 |
| comp 5 | 0.5886928 | 11.77386 | 100.00000 |

# eigenvaluesは主成分の分散，percentage of variancevは寄与率，cumulative percentage of varianceが累積寄与率を示す  
# スクリープロットを作成するfviz\_screeplot()は，自動的にpercentage of varianceをy値に出力する

第1主成分が27％超，第2主成分も合わせると80％近い。

## 主成分に対する各変数の寄与率を出図

fviz\_contrib(res.pca,   
 choice = "var", # 変数ごとの寄与率(ctr)  
 axes = 1, # 主成分1を指定（変更すると各主成分が指定できる）  
 top = 10)　 # 表示する変数の数を指定



ggsave(file = "fig7.png", width = 6, height = 4, dpi = 300) # ファイルの保存  
  
res.pca$var$contrib %>%  
 kable() # y軸に指定されている"var"でres.pcaオブジェクトの要素であるres.pca$varを引数に指定

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Dim.1 | Dim.2 | Dim.3 | Dim.4 | Dim.5 |
| デンプン粒 | 30.654628 | 14.755364 | 7.3649657 | 17.899973 | 29.325069 |
| 鉱物 | 11.372538 | 45.130749 | 0.1119077 | 6.230577 | 37.154228 |
| 細胞組織 | 8.627449 | 17.117302 | 63.9292627 | 3.562390 | 6.763597 |
| 繊維 | 21.281049 | 17.408648 | 15.1283692 | 29.670742 | 16.511191 |
| ほか | 28.064335 | 5.587937 | 13.4654946 | 42.636318 | 10.245915 |

第1主成分はデンプン粒，ほか（塵や墨などの物質）が高い寄与率を占めることから，第1主成分は「填料」と要約できる。

## 主成分得点の散布図を出力

# 主成分1と2を表示  
fviz\_pca\_biplot(res.pca) # 主成分1と2を表示，axes = C(○,○))で別の主成分を表示可能



ggsave(file = "fig8.png", width = 6, height = 4, dpi = 300) # ファイルの保存

デンプン粒と鉱物は，同じ意味を持つ変数，すなわち填料である。細胞組織の断片，繊維とほか（他の物質）は異なる変数を示すため，素材由来の構成物だけの含有ではない。

# 料紙面積と構成物の相関分析（無相関検定）

帰無仮説H₀：母相関は0である「調査史料では料紙面積と構成物に相関がない」  
対立仮説H₁：母相関は0ではない「調査史料では料紙面積と構成物に相関がある」

tbs5 <- read\_csv("yomei-square.csv") # CSVファイルの読み取り  
tbs5 # 読み込んだデータ

# A tibble: 89 x 4  
 番号 紙素材 料紙面積 構成物合計  
 <chr> <chr> <dbl> <dbl>  
 1 11-391 コウゾ 2612. 149  
 2 11-392 コウゾ 2650. 62  
 3 11-393 コウゾ 2015. 125  
 4 11-394 コウゾ 1991. 84  
 5 11-396 コウゾ 2020. 109  
 6 11-397 コウゾ 2141. 60  
 7 11-398 コウゾ 2036. 91  
 8 11-399 コウゾ 1944. 58  
 9 11-400 コウゾ 2056. 180  
10 11-401 コウゾ 2116. 24  
# ... with 79 more rows

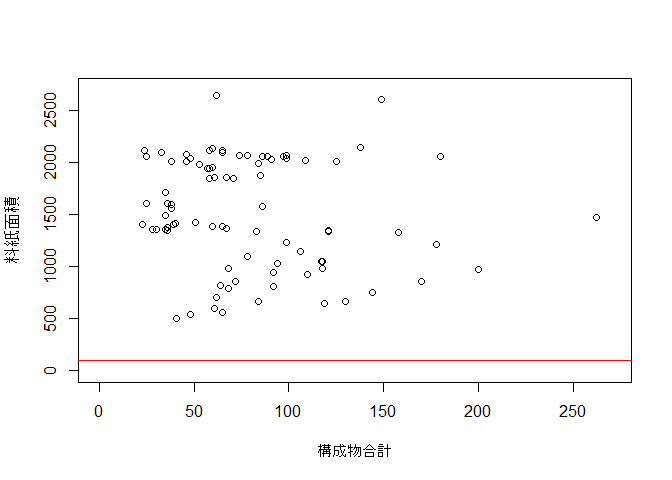
# 構成物の種類を実数型に変換  
tbs6 <-  
 tbs5 %>%  
 filter(紙素材 %in% "コウゾ") %>%　　　 # コウゾだけを選択  
 mutate(  
 面積 = as.numeric(料紙面積), 　　　　# 料紙面積を実数に変換  
 構成物合計 = as.numeric(構成物合計)) # 構成物合計を実数に変換  
  
# 料紙面積と構成物合計の相関計数と無相関検定  
attach(tbs6)  
  
cor(構成物合計,料紙面積, method="spearman") 　　 # スピアマンの相関係数

[1] -0.186486

cor.test(構成物合計,料紙面積, method="pearson") # 無相関かどうかの検定

Pearson's product-moment correlation  
  
data: 構成物合計 and 料紙面積  
t = -1.3249, df = 83, p-value = 0.1888  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 -0.34641471 0.07139816  
sample estimates:  
 cor   
-0.1439158

plot(構成物合計,料紙面積, xlim=c(0,270), ylim=c(0,2700)) # xlimとylimで範囲を指定  
# 回帰直線を入れる場合は以下を追加  
abline(lm(構成物合計~料紙面積), col="red") # 回帰直線を入れる



ggsave(file = "fig9.png", width = 6, height = 4, dpi = 300) # ファイルの保存

相関係数が-0.186486であり，対立仮説「調査史料では料紙面積と構成物に相関がある」は棄却される（ほとんど相関はない）。

# 料紙構成物の因子分析

料紙構成物に共通して影響する因子を仮定，この因子から変数間の相関関係を考える。

tbs9<- read\_csv("yomei-compo.csv") # CSVファイルの読み取り  
tbs9 # 読み込んだデータ

# A tibble: 89 x 7  
 番号 紙素材 デンプン粒 鉱物 細胞組織 繊維 ほか  
 <chr> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
 1 11-391 コウゾ 20 3 97 1 28  
 2 11-392 コウゾ 0 0 62 0 0  
 3 11-393 コウゾ 5 1 119 0 0  
 4 11-394 コウゾ 0 0 74 0 10  
 5 11-396 コウゾ 0 1 108 0 0  
 6 11-397 コウゾ 0 1 58 1 0  
 7 11-398 コウゾ 0 0 91 0 0  
 8 11-399 コウゾ 0 0 48 0 10  
 9 11-400 コウゾ 81 4 95 0 0  
10 11-401 コウゾ 0 0 19 0 5  
# ... with 79 more rows

# 構成物の種類を実数型に変換  
tbs10 <-  
 tbs9 %>%  
 filter(紙素材 %in% "コウゾ") %>%　　　 # コウゾだけを選択  
 select(デンプン粒,鉱物,細胞組織,繊維,ほか) %>%  
 mutate(  
 デンプン粒 = as.numeric(デンプン粒), # デンプン粒を実数に変換  
 鉱物 = as.numeric(鉱物),　　　　　　 # 鉱物を実数に変換  
 細胞組織 = as.numeric(細胞組織),　　 # 細胞組織を実数に変換  
 繊維 = as.numeric(繊維),　　　　　　 # 繊維を実数に変換  
 ほか = as.numeric(ほか))　　　　　　 # ほか（他の物質）を実数に変換  
  
# 構成物間の相関係数を出す  
相関行列 <- cor(tbs10)  
相関行列

デンプン粒 鉱物 細胞組織 繊維 ほか  
デンプン粒 1.00000000 0.29252650 -0.07469807 0.11404069 0.06666722  
鉱物 0.29252650 1.00000000 0.10766770 -0.09679362 0.07799736  
細胞組織 -0.07469807 0.10766770 1.00000000 -0.09181340 -0.11933025  
繊維 0.11404069 -0.09679362 -0.09181340 1.00000000 0.21947530  
ほか 0.06666722 0.07799736 -0.11933025 0.21947530 1.00000000

# 構成物の因子分析を実施  
eigen(相関行列) # 固有値を算出し因子数を決定（2に決定）

eigen() decomposition  
$values  
[1] 1.3766096 1.2662437 0.9304291 0.8380248 0.5886928  
  
$vectors  
 [,1] [,2] [,3] [,4] [,5]  
[1,] -0.5536662 0.3841271 0.27138470 0.4230836 0.5415263  
[2,] -0.3372319 0.6717942 -0.03345261 -0.2496112 -0.6095427  
[3,] 0.2937252 0.4137306 -0.79955777 0.1887429 0.2600692  
[4,] -0.4613139 -0.4172367 -0.38895204 0.5447086 -0.4063397  
[5,] -0.5297578 -0.2363882 -0.36695360 -0.6529649 0.3200924

構成物因子分析 <- factanal(tbs10,factors=2) # 因子分析を実施  
構成物因子分析

Call:  
factanal(x = tbs10, factors = 2)  
  
Uniquenesses:  
デンプン粒 鉱物 細胞組織 繊維 ほか   
 0.859 0.005 0.934 0.716 0.821   
  
Loadings:  
 Factor1 Factor2  
デンプン粒 0.340 0.159   
鉱物 0.970 -0.234   
細胞組織 -0.252   
繊維 0.532   
ほか 0.174 0.385   
  
 Factor1 Factor2  
SS loadings 1.089 0.575  
Proportion Var 0.218 0.115  
Cumulative Var 0.218 0.333  
  
Test of the hypothesis that 2 factors are sufficient.  
The chi square statistic is 1 on 1 degree of freedom.  
The p-value is 0.317

共通性 <- 1-構成物因子分析$uniquenesses # 共通性を表示  
共通性

デンプン粒 鉱物 細胞組織 繊維 ほか   
0.14105652 0.99500000 0.06593406 0.28368629 0.17861696

# 因子負荷が小さい値でも0である場合を除いて表示，因子ごとに因子負荷の高い項目をまとめて並べる  
print(構成物因子分析,cutoff=0,sort=TRUE)

Call:  
factanal(x = tbs10, factors = 2)  
  
Uniquenesses:  
デンプン粒 鉱物 細胞組織 繊維 ほか   
 0.859 0.005 0.934 0.716 0.821   
  
Loadings:  
 Factor1 Factor2  
鉱物 0.970 -0.234   
繊維 0.029 0.532   
デンプン粒 0.340 0.159   
細胞組織 0.050 -0.252   
ほか 0.174 0.385   
  
 Factor1 Factor2  
SS loadings 1.089 0.575  
Proportion Var 0.218 0.115  
Cumulative Var 0.218 0.333  
  
Test of the hypothesis that 2 factors are sufficient.  
The chi square statistic is 1 on 1 degree of freedom.  
The p-value is 0.317

# 構成物の因子分析を実施，プロマックス回転で因子間に相関をもたせる  
構成物因子分析斜交 <- factanal(tbs10,factors=2,rotation="promax")  
print(構成物因子分析斜交,cutoff=0,sort=TRUE)

Call:  
factanal(x = tbs10, factors = 2, rotation = "promax")  
  
Uniquenesses:  
デンプン粒 鉱物 細胞組織 繊維 ほか   
 0.859 0.005 0.934 0.716 0.821   
  
Loadings:  
 Factor1 Factor2  
鉱物 1.008 -0.112   
繊維 -0.118 0.537   
デンプン粒 0.287 0.203   
細胞組織 0.118 -0.246   
ほか 0.063 0.409   
  
 Factor1 Factor2  
SS loadings 1.130 0.570  
Proportion Var 0.226 0.114  
Cumulative Var 0.226 0.340  
  
Factor Correlations:  
 Factor1 Factor2  
Factor1 1.000 -0.148  
Factor2 -0.148 1.000  
  
Test of the hypothesis that 2 factors are sufficient.  
The chi square statistic is 1 on 1 degree of freedom.  
The p-value is 0.317