

# 「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・高島晶彦・天野真志・野村朋弘・山田太造・尾上陽介

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)
library(ggfortify)
```

## 現生デンプン粒標本と料紙のデンプン粒の比較

### 料紙のデンプン粒の特徴

```
# 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
```

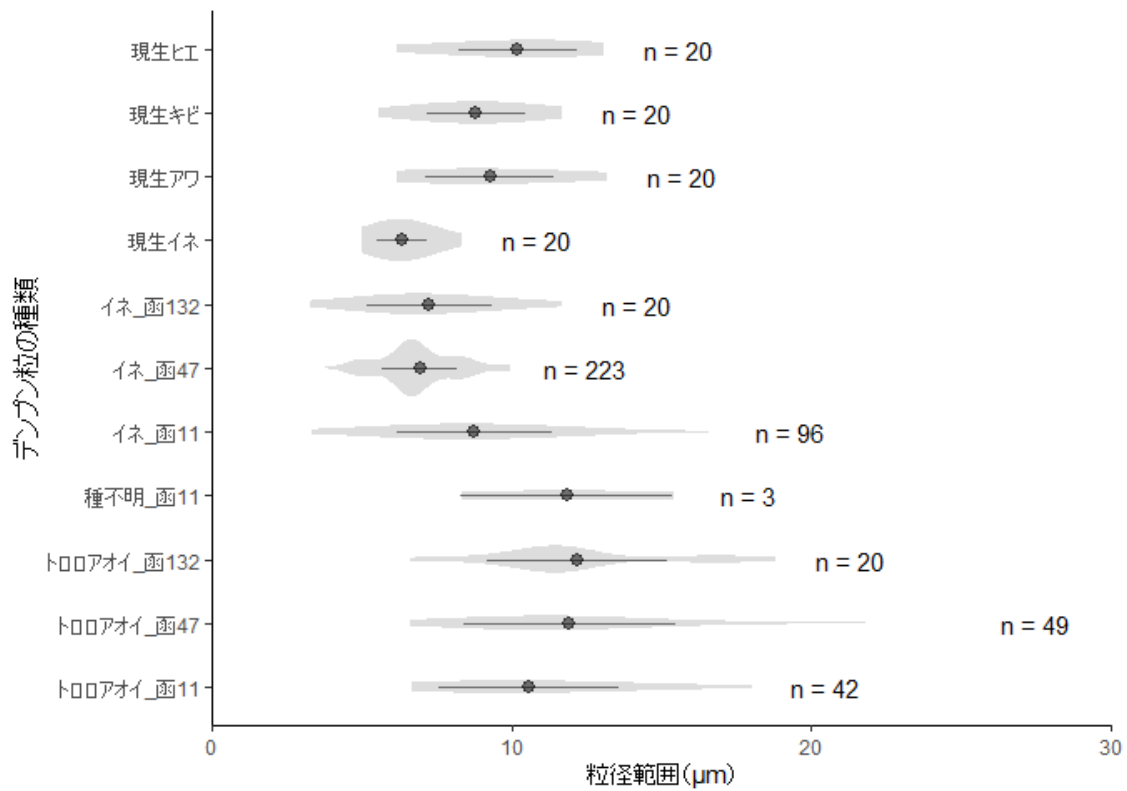
```
[1] "デンプン粒の種類" "粒径範囲"
```

```
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
```

## 料紙構成物に対する主成分分析

```
tbs1 <- read_csv("yomei-compo.csv") # CSV ファイルの読み取り
tbs1 # 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
tbs2 <-
  tbs1 %>%
    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(tbs2, 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_screplot(res.pca)

```



2 202		0.903		-0.761	0.495	0.710		-0.239	0.053	0.070		-0.400	0.
3 514		1.461		-0.634	0.344	0.188		1.145	1.218	0.614		0.638	0.
4 028		0.895		-0.384	0.126	0.184		-0.326	0.099	0.133		0.149	0.
5 335		1.298		-0.735	0.462	0.321		0.912	0.773	0.494		0.515	0.
6 024		1.294		0.302	0.078	0.055		-0.203	0.038	0.025		0.137	0.
7 030		1.036		-0.965	0.796	0.867		0.048	0.002	0.002		0.155	0.
8 153		1.015		-0.201	0.035	0.039		-0.583	0.316	0.330		-0.348	0.
9 380		6.146		3.290	9.250	0.287		4.873	22.063	0.629		-1.045	1.
10 428		1.419		-0.228	0.045	0.026		-0.767	0.547	0.292		-1.063	1.

	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 0.069	0.650	30.655	0.422	0.432	14.755	0.187	-0.262	7.365
鉤物 0.001	0.396	11.373	0.157	0.756	45.131	0.571	0.032	0.112
細胞組織 0.595	-0.345	8.627	0.119	0.466	17.117	0.217	0.771	63.929
繊維 0.141	0.541	21.281	0.293	-0.470	17.409	0.220	0.375	15.128
ほか 0.125	0.622	28.064	0.386	-0.266	5.588	0.071	0.354	13.465
デンプン粒								
鉤物								

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

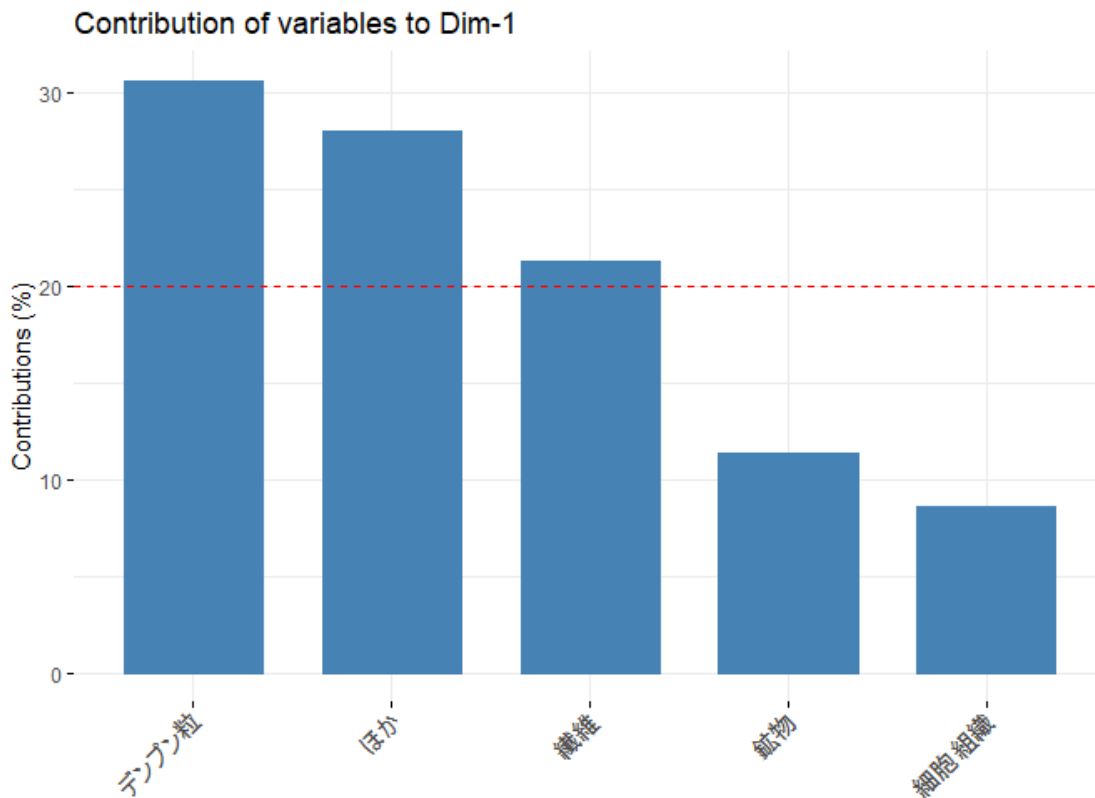
# *eigenvalues* は主成分の分散, *percentage of variance* は寄与率, *cumulative percentage of variance* が累積寄与率を示す

# スクリーン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 = 6, 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

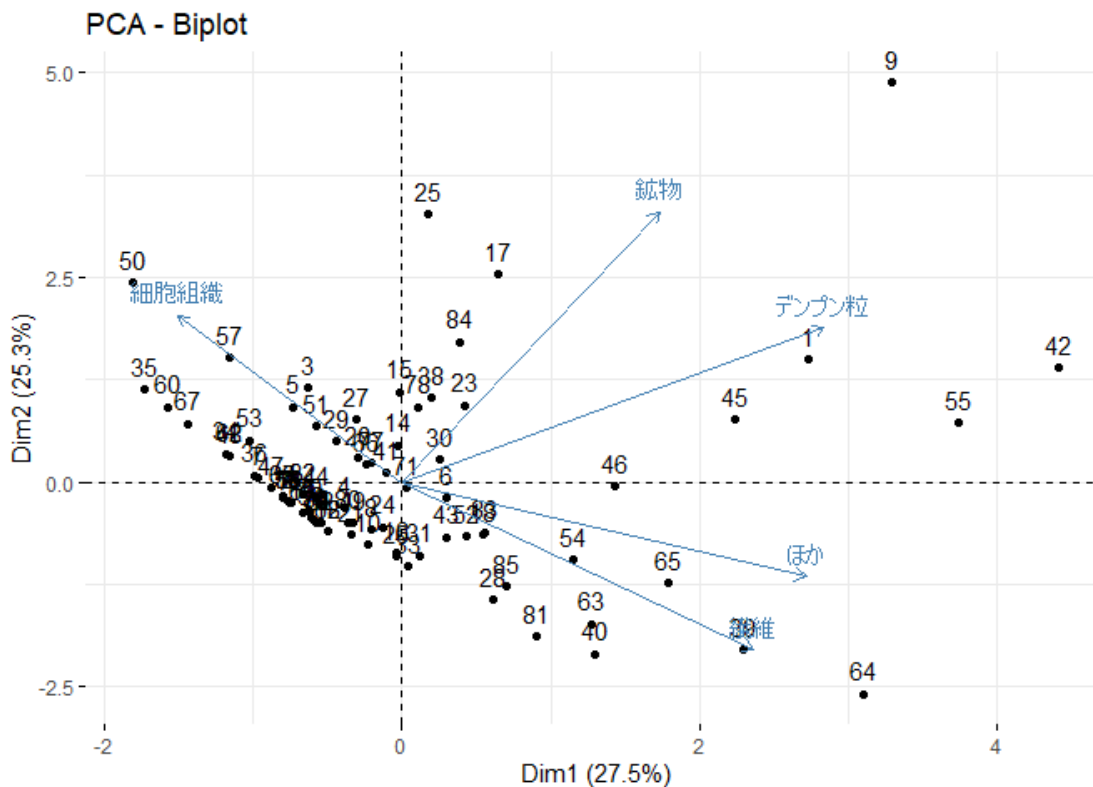
第 1 主成分はデンプン粒、ほか（塵や墨などの物質）が高い寄与率を占めることから、第 1 主成分は「填料とたの物質の混合」と要約できる。



## 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(0,0)) で別の主成分を表示可能
```



```
ggsave(file = "fig8.png", width = 6, height = 6, 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  $H_0$  : 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  $H_1$  : 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」

```
tbs3 <- read_csv("yomei-square.csv") # CSV ファイルの読み取り  
tbs3 # 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
tbs4 <-
  tbs3 %>%
  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(tbs4)

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

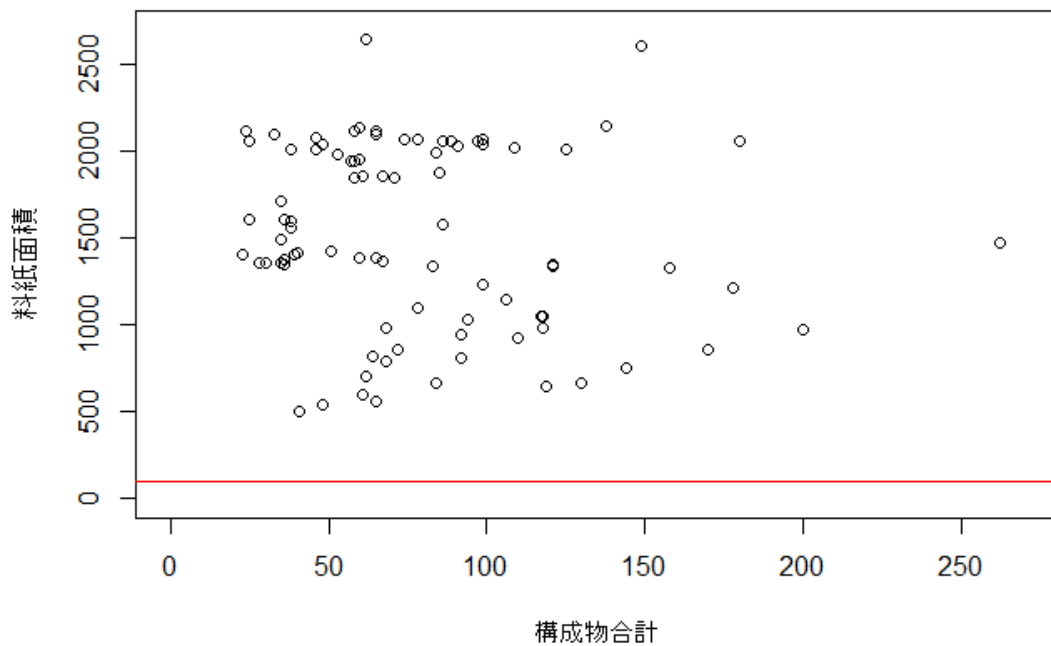
cor.test(構成物合計,料紙面積, method="pearson")  # 無相関かどうかの検定

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→fig10 として保存
```



相関係数が-0.186486 であり、t 値=-1.3249、df (自由度) =83、p-value (p 値) =0.1888 という数値が得られた。今回の p 値は有意水準 0.05 (5%) より大きい。つまり、2 つの変数「料紙面積」と「構成物の量」の間には有意な相関があるとはいえず、相関係数については、意味のある相関係数であるとはいえない。

## 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。

```
# 因子分析を行うパッケージを読み込む
library(psych)
library(GPArotation)
```

```
tbs5<- read_csv("yomei-compo.csv") # CSV ファイルの読み取り
tbs5 # 読み込んだデータ
```

```
# A tibble: 89 x 7
```

	番号	紙素材	デンプン粒	鉍物	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
```

```
tbs6 <-
  tbs5 %>%
  filter(紙素材 %in% "コウゾ") %>% # コウゾだけを選択
  select(デンプン粒, 鉍物, 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
    鉍物 = as.numeric(鉍物), # 鉍物を実数に変換
    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
```

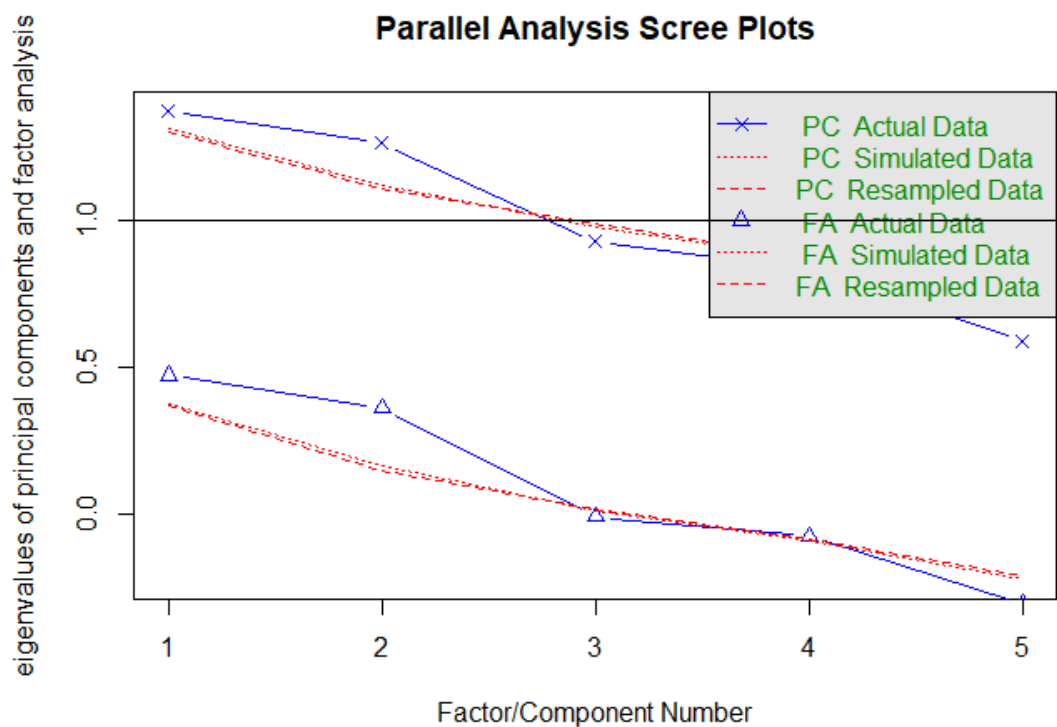
```
相関行列 <- cor(tbs6)
```

```
相関行列
```

	デンプン粒	鉍物	細胞組織	繊維	ほか
デンプン粒	1.00000000	0.29252650	-0.07469807	0.11404069	0.06666722
鉍物	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

```
# 因子数を決める
```

```
fa.parallel(tbs6, SMC=TRUE) # スクリーンプロットを表示
```



Parallel analysis suggests that the number of factors = 0 and the number of components = 0

```
vss(tbs6, n.obs=N, rotate="varimax")
```

Warning in fac(r = r, nfactors = nfactors, n.obs = n.obs, rotate = rotate, : An ultra-Heywood case was detected. Examine the results carefully



```
4 1.7e-18 3.2e-11 NA NA
5 1.7e-18 3.2e-11 NA NA
```

# 結果として、平行分析では3因子、MAP法では1因子、適合度基準（BIC）では2因子が良い。ここでは3因子で決める。

# 因子分析を行う

```
fa.result1 <- fa(tbs6, nfactored=3, fm="ML")
print(fa.result1, sort=T, cut=0.3) # 因子負荷が0.3以下の値を非表示
```

Factor Analysis using method = ml

Call: fa(r = tbs6, nfactored = 3, fm = "ML")

Standardized loadings (pattern matrix) based upon correlation matrix

	item	ML1	ML2	ML3	h2	u2	com
鉤物	2	0.99			0.995	0.005	1.0
デンプン粒	1		0.85		0.721	0.279	1.0
ほか	5			0.68	0.465	0.535	1.0
繊維	4			0.32	0.145	0.855	2.0
細胞組織	3				0.061	0.939	2.7

	ML1	ML2	ML3
SS loadings	1.03	0.75	0.60
Proportion Var	0.21	0.15	0.12
Cumulative Var	0.21	0.36	0.48
Proportion Explained	0.43	0.31	0.25
Cumulative Proportion	0.43	0.75	1.00

With factor correlations of

	ML1	ML2	ML3
ML1	1.00	0.33	0.08
ML2	0.33	1.00	0.14
ML3	0.08	0.14	1.00

Mean item complexity = 1.6

Test of the hypothesis that 3 factors are sufficient.

The degrees of freedom for the null model are 10 and the objective function was 0.22 with Chi Square of 18.17

The degrees of freedom for the model are -2 and the objective function was 0

The root mean square of the residuals (RMSR) is 0

The df corrected root mean square of the residuals is NA

The harmonic number of observations is 85 with the empirical chi square 0 with prob < NA

```
The total number of observations was 85 with Likelihood Chi Square = 0
with prob < NA
```

```
Tucker Lewis Index of factoring reliability = 2.294
```

```
Fit based upon off diagonal values = 1
```

```
Measures of factor score adequacy
```

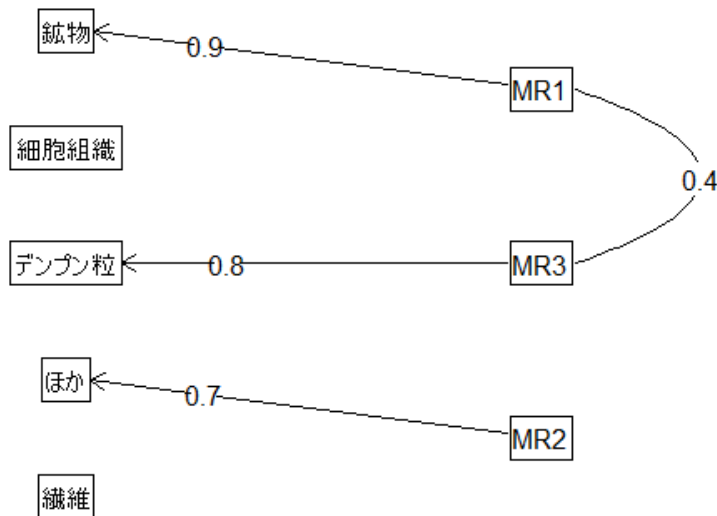
	ML1	ML2	ML3
Correlation of (regression) scores with factors	1.00	0.86	0.71
Multiple R square of scores with factors	0.99	0.73	0.51
Minimum correlation of possible factor scores	0.99	0.47	0.02

```
# 因子負荷の可視化
```

```
fa.result1 = fa(tbs6, nfactors=3, fm="minres", rotate="oblimin", use="complete.obs")
```

```
fa.diagram(fa.result1)
```

### Factor Analysis



```
# 因子負荷量の表示
```

```
unclass(fa.result1$loadings)
```

	MR1	MR3	MR2
デンプン粒	0.01654569	0.79113716	-0.009188201
鉍物	0.92299697	0.01169930	0.008554870
細胞組織	0.18572833	-0.14466103	-0.163285306



```
繊維      -0.20071818  0.18314100  0.295960324
ほか      0.02345439 -0.01816758  0.737600749
```

```
# 描画
biplot(fa.result1$scores,fa.result1$loading,cex=1)
```

```
# 描画
biplot(fa.result1$scores, fa.result1$loading, cex=1)
```

```
biplot(fa.result1$scores,fa.result1$loading,cex=1)
```

