

主成分負荷量と結果の解釈

重み・主成分係数を用いた代謝物の選び方

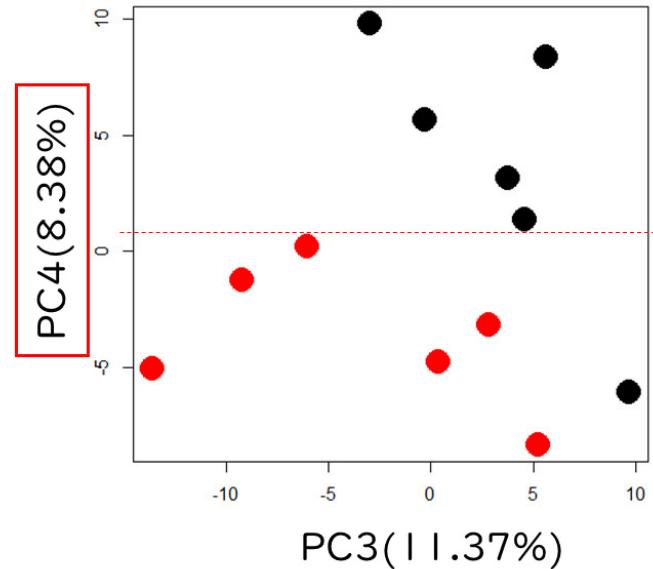
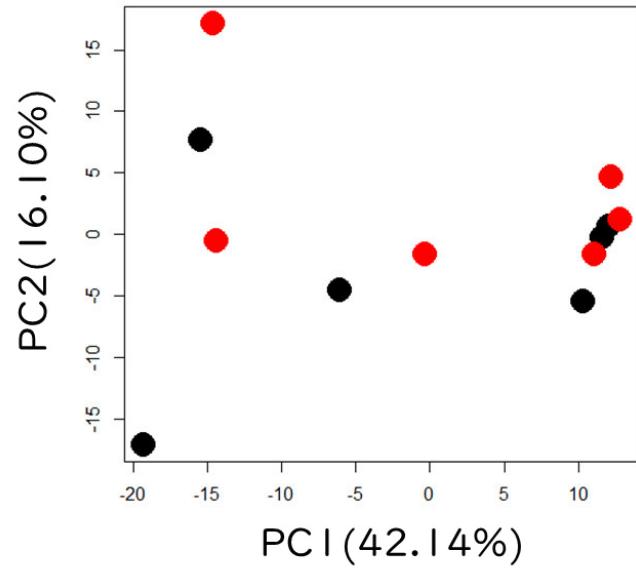
主成分スコア=(変数1)×w₁+(変数2)×w₂+…+(変数P)×w_p

$$\begin{array}{c} \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \\ \text{↑} = x_1 w_1 + x_2 w_2 + \dots + x_p w_p \end{array}$$

主成分スコアは、各変数のデータを重みwを係数として足し合わせたものであり、wの値が大きい変数が主成分スコアと関連が強く、wの値が小さい変数が主成分スコアと関連が弱い

例. 係数w₁が0であれば、主成分スコアと変数1のデータは関連が無い

Rでの主成分係数(固有ベクトル)の取得



- 主成分係数(PC4)の取得

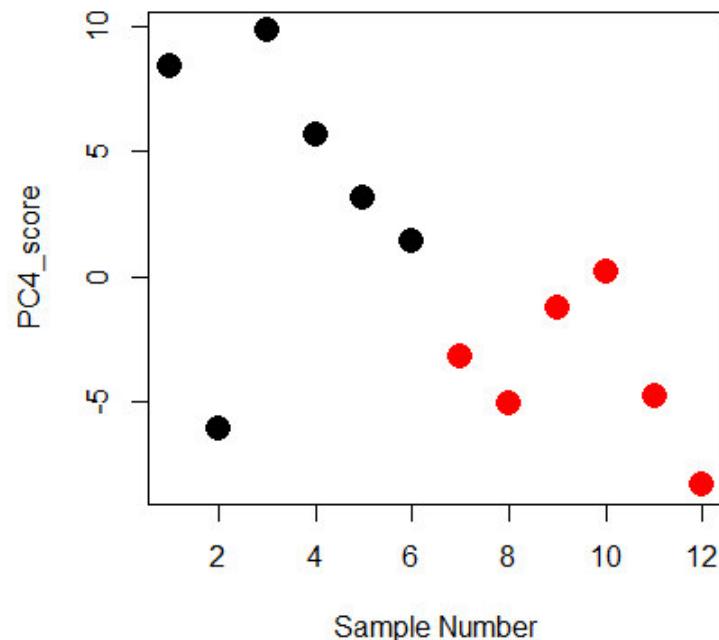
```
v <- pca$rotation[,4]  
names(v) <- X0[,1]
```

第1主成分スコアと固有ベクトルの結果

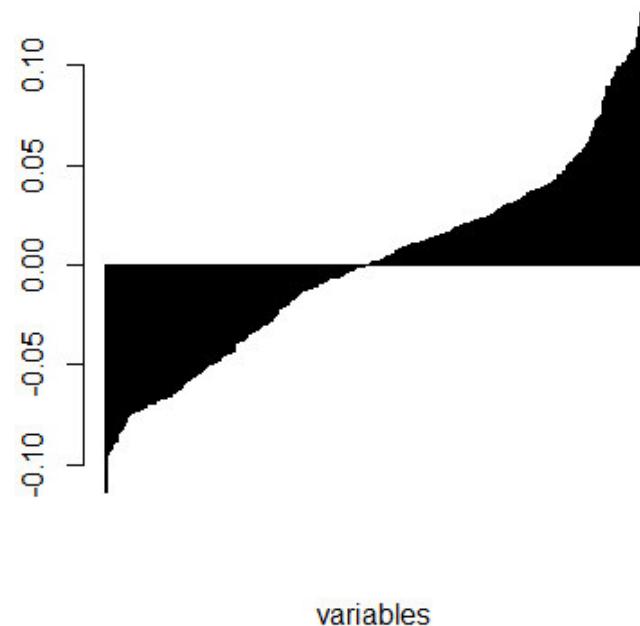
- PC4のスコアと主成分係数(固有ベクトル)

```
PC4_score <- PC_score[,4] # 第4主成分  
plot(PC4_score,col=class, pch=16,cex=2, xlab="Sample Number")  
barplot(sort(v), xlab="variables", xaxt="n")
```

スコアプロット



主成分係数wのプロット



結果をcsvファイルに出力

- 主成分係数(PC4)をcsvファイルに出力

```
write.csv(v, file="C:/R/weight_PC4.csv")
```

The screenshot shows a Microsoft Excel spreadsheet titled "weight_PC4.csv - Excel". The ribbon menu is visible at the top. The data is organized into two columns: "変数名" (Variable Name) and "PC4の主成分係数" (PC4 Principal Component Coefficient). The first column contains numerical values, and the second column contains their ratios. A green box highlights the range from A1 to H1.

変数名	PC4の主成分係数
1	
2 200.1/2926	-0.063572219
3 205/2791	-0.007743185
4 206/2791	-0.038441572
5 207.1/2719	-0.00384172
6 219.1/2524	-0.04998422

主成分係数(PC4)の値で並び替え

データタブを開く

weight_PC4.csv - Excel

ファイル ホーム 挿入 ページレイアウト 数式 データ 校閲 表示 開発 アドイン ヘルプ チーム

外部データの取り込み 新しいケリ ケリの表示 テーブルから 最近使ったソース 取得と変換

すべて更新 接続 プロパティ リンクの編集

接続 データの種類

株式 地理

クリア 再適用 詳細設定

並べ替え フィルター 並べ替えとフィルター

A1

	A	B
1		
2	384.2/3993	0.127934218
3	491.2/3398	0.125986297
4	449.1/3290	0.119687939
5	322.1/3392	0.119519714
6	439.3/4056	0.113390845
7	438.3/4056	0.112686375
8	301.2/3389	0.108496889
9	354.2/3618	0.107604041
10	300.2/3392	0.107492645

1行目のA列とB列のセルを選択した状態で、
フィルタを実行

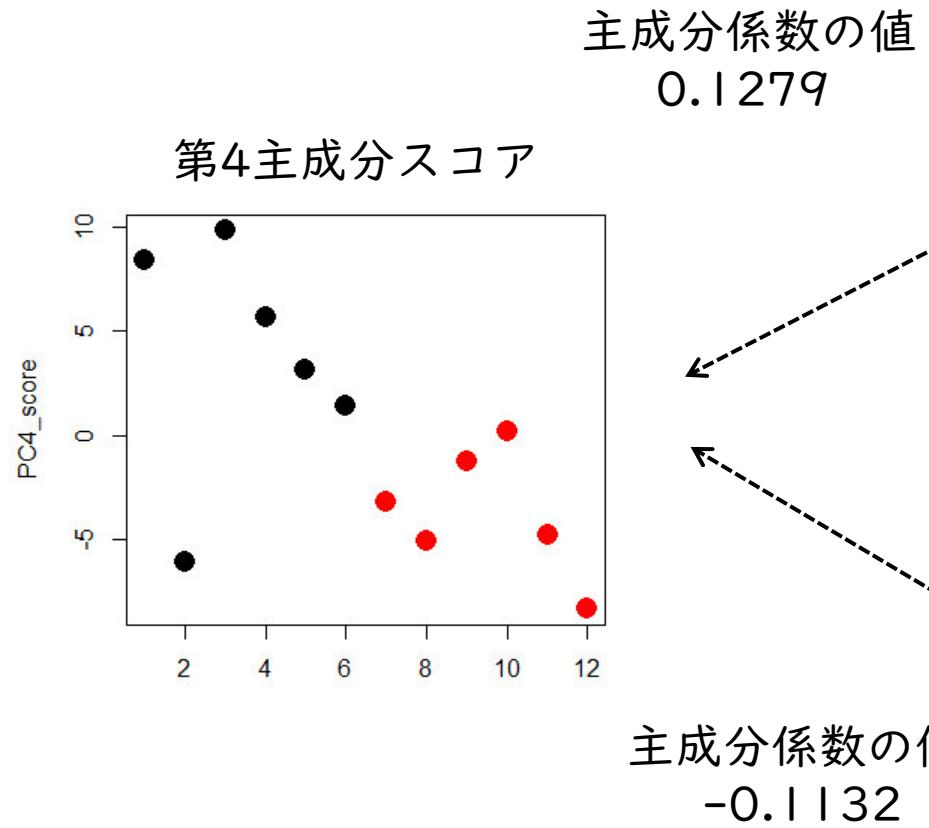
降順(または昇順)で並び替え

PC4と関連のある変数(ピーク)を抽出できる

主成分係数の値が大きい2つの変数

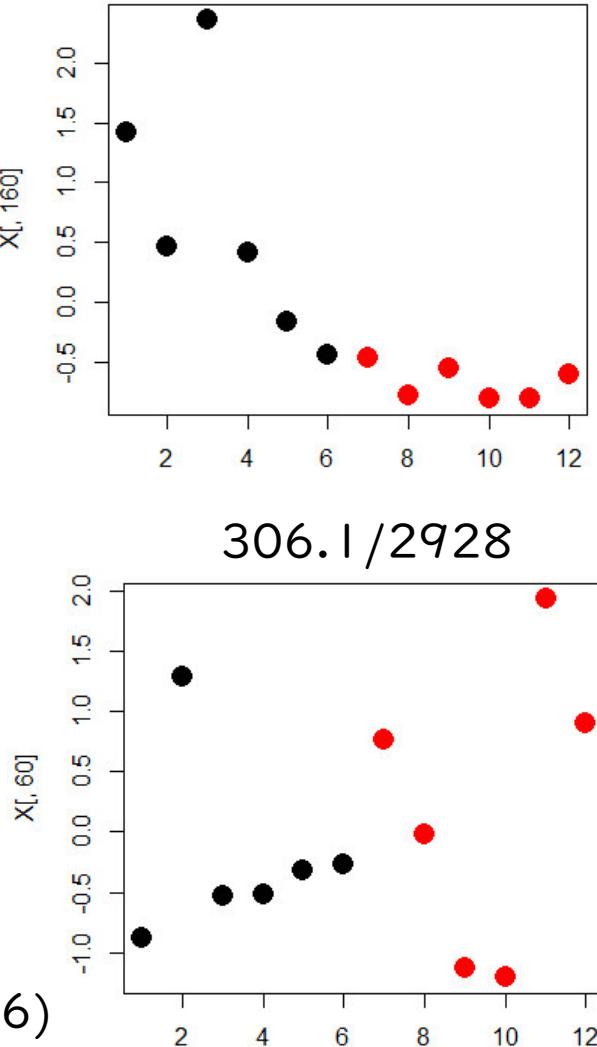
```
plot(X[,160],col=class, cex=2, pch=16)
```

384.2/3993



```
plot(X[,60],col=class, cex=2, pch=16)
```

306.1/2928



主成分係数(PC4)の値が大きな上位10変数

正に値が大きな変数上位10個

ピーク	主成分係数
384.2/3993	0.1279
491.2/3398	0.1260
449.1/3290	0.1197
322.1/3392	0.1195
439.3/4056	0.1134
438.3/4056	0.1127
301.2/3389	0.1085
354.2/3618	0.1076
300.2/3392	0.1075
410.3/3937	0.1064

負に値が大きな変数上位10個

ピーク	主成分係数
306.1/2928	-0.1132
414.2/3060	-0.0943
246.1/2517	-0.0939
591.3/3003	-0.0926
288.1/2798	-0.0923
590.3/3003	-0.0916
546.3/3015	-0.0888
532.3/3729	-0.0884
547.2/3015	-0.0881
576.3/2860	-0.0844

主成分係数の値が大きな上位いくつまでが重要な変数、
もしくは主成分係数の値がいくつまでを
重要な変数と考えれば良いか？

演習内容

- csvファイル(fasting.csv)の読み込み
- 主成分分析を計算
- 主成分スコア
- 主成分係数(固有ベクトル)から代謝物を選ぶ
- MetaboAnalystの結果と比べてみる

読み込むファイルに関する注意

変数

fasting.csv

サンプル

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Sample	fasting	gamma-G	gamma-G	gamma-B	beta-Ala	Xanthosin	Xanthopte	Xanthine	Val	UTP	Urocanic	Uridine	Uric acid	Urea	Uracil	UMP
1 L1-1	0	0.00213	0.00003	0.00225	0.00698	0.00204	0.00012	0.00834	0.02848	0.00002	0.00002	0.00929	0.00183	0.14336	0.00148	0.00639
2 L1-2	0	0.00212	0.00004	0.0022	0.01274	0.00138	0.0001	0.00506	0.04297	0.00002	0.00002	0.0107	0.00359	0.20963	0.0017	0.00657
3 L1-3	0	0.00115	0.00004	0.00193	0.00995	0.00053	0.00009	0.00427	0.03943	0.00004	0.00002	0.00681	0.0022	0.2026	0.00114	0.00975
4 L1-4	0	0.00112	0	0.00154	0.00762	0.00076	0.00008	0.00605	0.04423	0.00003	0.00002	0.00663	0.00225	0.15643	0.00108	0.00683
5 L1-5	0	0.00076	0	0.00143	0.00626	0.00093	0.00005	0.00547	0.04475	0.00006	0	0.00704	0.00311	0.13072	0.00118	0.00913
7 L2-1	12	0.0005	0.00007	0.00153	0.01778	0.00099	0	0.00903	0.03792	0.00002	0.00018	0.00643	0.0024	0.14257	0.00113	0.00709
8 L2-2	12	0.00046	0.00007	0.0011	0.01317	0.00058	0.00003	0.00824	0.03474	0.00001	0.00004	0.00649	0.00242	0.10548	0.00111	0.00741
9 L2-3	12	0.00076	0.0001	0.0011	0.01042	0.00043	0.00004	0.00566	0.03678	0	0.00003	0.00672	0.00224	0.11282	0.0012	0.00999
10 L2-4	12	0.00037	0.00005	0.00093	0.01232	0.00051	0.00004	0.00785	0.02622	0.00002	0.00005	0.00611	0.00351	0.13681	0.00109	0.00819
11 L2-5	12	0.00022	0.00004	0.00113	0.01408	0.00035	0	0.00731	0.03029	0.00002	0.00004	0.00524	0.00258	0.13154	0.00101	0.00827

サンプル

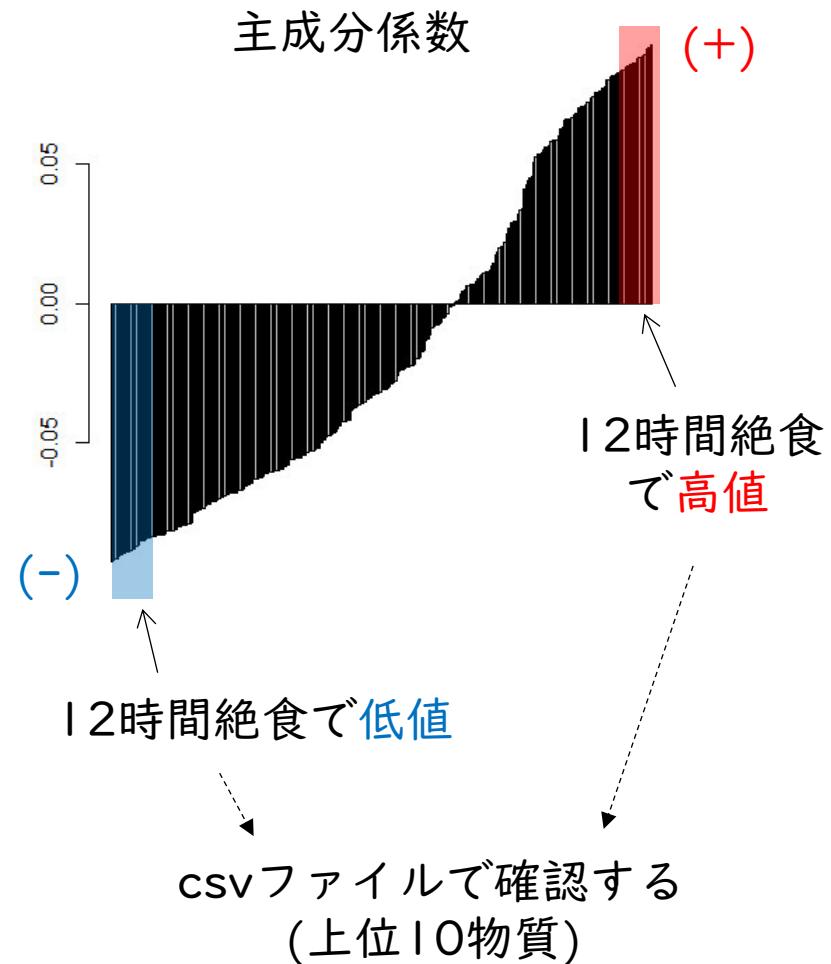
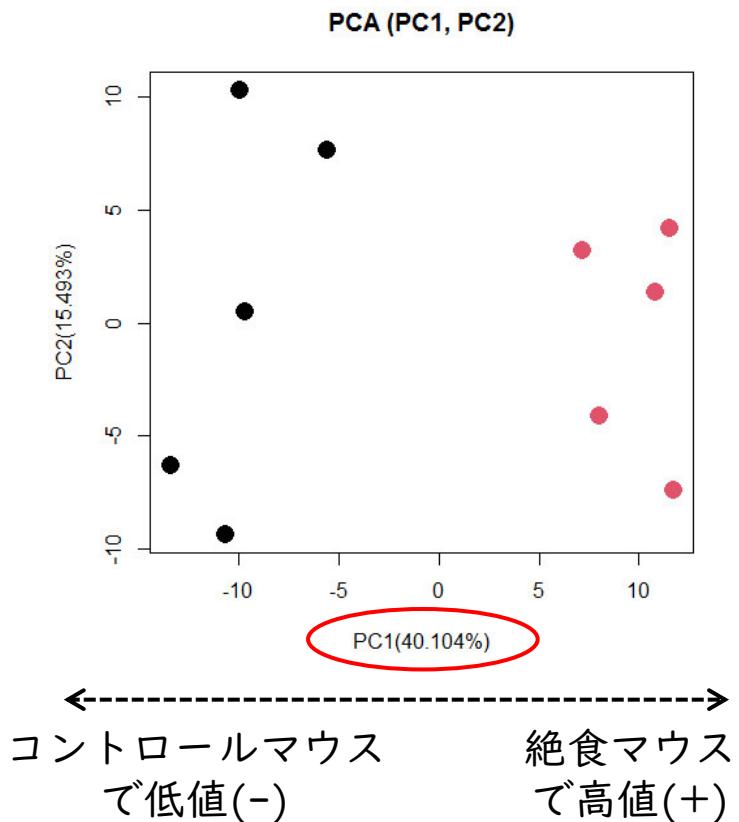
mouse_data_original.csv

変数

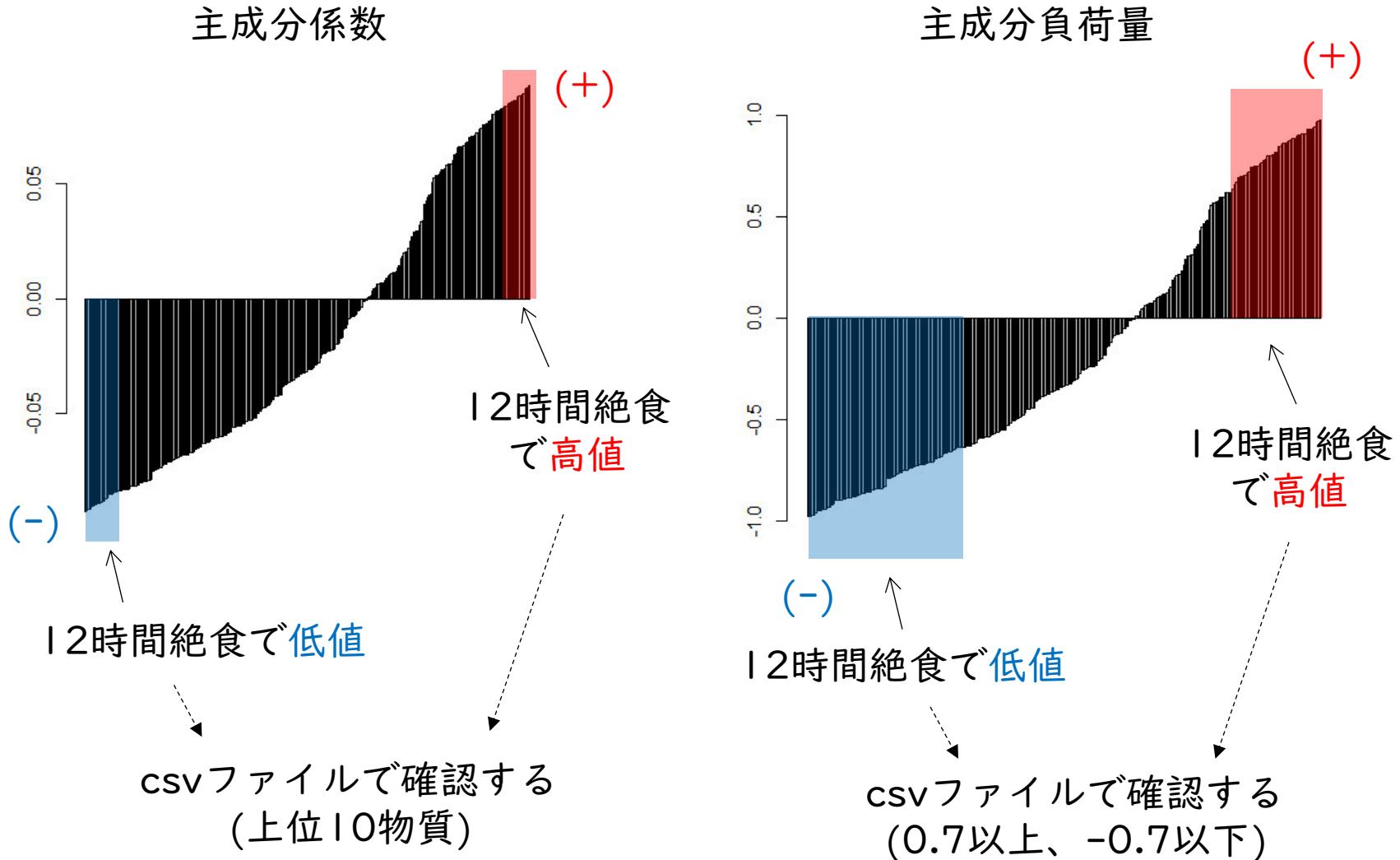
A	B	C	D	E	F	G	H	I	J	K	L	M
Label	ko15	ko16	ko18	ko19	ko21	ko22	wt15	wt16	wt18	wt19	wt21	wt22
3 200.1/2926	147887.5	451600.7	65290.38	56540.93	85146.33	162012.4	175177.1	82619.48	51951.61	69198.22	153273.5	98144.28
4 205/2791	1778569	1567038	1482796	1039130	1223132	1072038	1950287	1466781	1572679	1275313	1356014	1231442
5 206/2791	237993.6	269714	201393.4	150107.3	176989.7	156797	276541.8	222366.2	211717.7	186850.9	188285.9	172348.8
6 207.1/2719	380873	460629.7	351750.1	219288	286848.6	235022.6	417169.6	324892.5	277990.7	220972.4	236728.2	
7 219.1/2524	235544.9	173623.4	82364.59	79480.4	185792.4	174458.8	244584.5	161184.1	72029.38	75096.99	238194.4	173830
8 231/2516	117649.8	48960.63	226609.1	286232.2	435094.5	62168.71	465898	61234.44	96841.46	240261.2	201316.2	179437.7
9 233/3023	399145.3	356951.3	410550.7	198416.5	363381.7	317805.23	397107.8	271252.1	334459.9	181901.3	456900.5	294270.6
10 234/3024	76880.87	99526.27	97493.76	53461.71	88227.79	81072.23	65215.84	55952.44	73781.01	45211.66	83693.39	57516.2
11 235.1/2695	171995.2	128945.2	155442.5	115286.3	37769.45	7663.88	199981.5	30028.6	156968.3	52596.48	14641.59	27713.29
12 236.1/2524	252282	2060319	71763.79	73602.47	186661	198804.3	253791.1	187225.7	79389.63	90012.64	256263.3	206487
13 240.2/3682	112440.6	90466.98	193768.7	170641.5	81223.74	146653	43974.38	270872.9	176425.2	187063	75235.94	86068.18
14 241.1/3879	1465989	1318747	1215369	632037.4	579968.2	561964.9	1468103	1594705	1006929	805533.2	731777.6	522916.1
15 242.1/3679	280767.6	248792.4	224467.7	109019.4	103841.9	101092.9	280280.2	299453.2	188328.2	139748.3	139968.5	95347.85
16 244.1/2832	612169.9	256316.5	90539.86	33269.29	54610.78	115505.4	627835	56883.71	168524.5	27260.37	110170.7	1352930
17 246.1/2517	27932.12	51557.09	38018.76	41747.89	56145.39	47901.55	105508.8	70508.82	46090.83	45630.55	77205.36	486427.55
18 249.1/3668	1435001	1228148	1193347	641881.8	536808.3	574005.7	1297986	1566269	1076654	747969.8	688496.3	485427.6
19 250.1/3668	347794.8	238893.4	248245.6	126623.8	107844.1	118627	281745.8	336058.8	215534.9	149872.8	139317.3	99814.5
20 254.1/3231	78911.6	564680.8	8937	127186.8	21055.46	24185.93	54645.64	124200.9	108479	30096	66315.96	27554.78
21 255.2/3679	1420403	1187752	1264222	673816.2	581177.9	651560.3	1211727	1550071	1204087	875296.3	759075.9	640885.7
22 256.2/3679	307708.4	263317.7	275104.9	138168.4	121284.4	128540.4	258452.2	338390.9	213618.1	180384.5	164675.7	91822.7

主成分分析の結果

マウスの肝臓のメタボローム解析
2群比較: ●コントロール、●絶食12時間後



主成分係数と主成分負荷量の比較



一般的な主成分負荷量の定義

杉山高一、多変量データ解析入門、朝倉書店(2001) 35ページ

『主成分と変数の相関係数を、主成分の因子負荷量(factor loading)という』

塩谷實、多変量解析概論、朝倉書店(2001) 114ページ

『共分散行列を用いた主成分分析の場合、第*i*主成分 U_i と*j*番目の成分変数 X_j の間の相関係数を X_j の U_i への因子負荷量という』

(その後に、相関係数行列についても同じであることが書かれている)

中村永友、Rで学ぶデータサイエンス 多次元データ解析法、共立出版(2009) 102ページ

『得られた主成分と本来の変数の相関係数を主成分負荷量あるいは因子負荷量という』

一般的に、主成分分析の主成分負荷量といえば、主成分スコアと各変数との相関係数として定義される

主成分分析のローディングの2つの定義

- 定義その1：主成分スコアと各変数の相関係数
 - 各変数を平均0、分散1にスケーリングするときに利用されることが多い
 - 二乗して足した値は1にならない
- 定義その2：相関係数行列の固有ベクトル(主成分係数)
 - 各変数を平均0、分散1にスケーリングしないときに利用されることが多い
 - 二乗して足した値が1になる

定義その1 の主成分負荷量

『主成分スコアと各変数の相関係数』の使用を推奨

ソフトウェアがどちらを使用しているか
確認する必要がある

主成分負荷量の計算方法

主成分スコア t_i と各変数 X_i との相関係数 r_i

$$r_i = \text{corr}(t_i, X_i) = \frac{\sqrt{\lambda} w_i}{\sqrt{\frac{1}{n-1} X_i' X_i}}$$

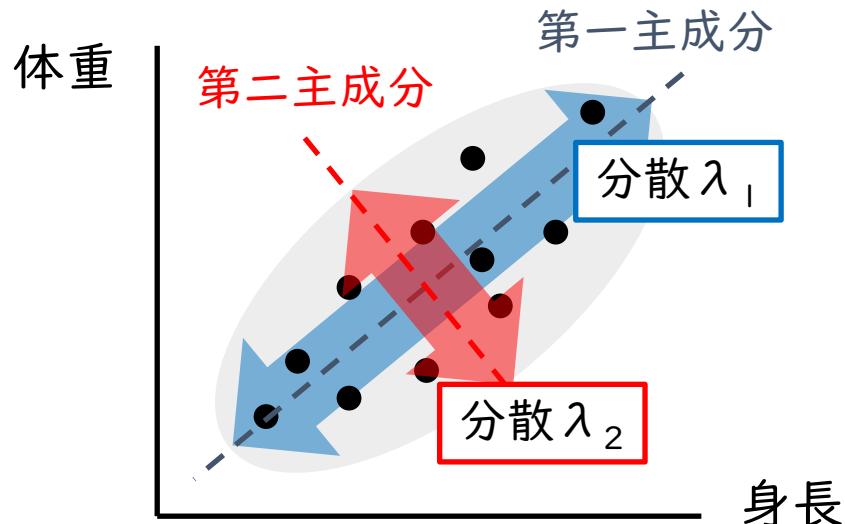
各変数をautoscaling(平均0分散1)した時は、次のように書ける

$$r_i = \sqrt{\lambda} w_i \quad (\text{途中の式は省略})$$

主成分負荷量は、固有ベクトル w に固有値 λ の平方根を掛けて計算できる

また固有ベクトル w は、主成分スコアと各変数の相関係数に比例する、とも表現できる

[参考] 寄与率と主成分スコアの分散の関係



第一主成分の寄与率(%) :
 $100 \times \lambda_1 / (\lambda_1 + \lambda_2)$

$$\overline{Var(\mathbf{t})} = \frac{1}{n-1} \frac{\mathbf{w}' \mathbf{X}' \mathbf{X} \mathbf{w}}{\|\mathbf{w}\|^2} = \frac{\mathbf{w}' \lambda \mathbf{w}}{\|\mathbf{w}\|^2} = \lambda$$

$$\frac{1}{n-1} \mathbf{X}' \mathbf{X} \mathbf{w} = \lambda \mathbf{w}$$

固有値 λ は主成分スコアの分散を表す

第1主成分(体の大きさ)

第2主成分(スタイル)

最大固有値に対応する固有ベクトル

$\| \mathbf{w} \|$
 主成分スコアの分散が最大の
 固有ベクトル(PC1)

主成分係数(PC4)の値が大きな上位10変数

正に値が大きな変数上位10個

ピーク	主成分係数
384.2/3993	0.1279
491.2/3398	0.1260
449.1/3290	0.1197
322.1/3392	0.1195
439.3/4056	0.1134
438.3/4056	0.1127
301.2/3389	0.1085
354.2/3618	0.1076
300.2/3392	0.1075
410.3/3937	0.1064

·
·
·

負に値が大きな変数上位10個

ピーク	主成分係数
306.1/2928	-0.1132
414.2/3060	-0.0943
246.1/2517	-0.0939
591.3/3003	-0.0926
288.1/2798	-0.0923
590.3/3003	-0.0916
546.3/3015	-0.0888
532.3/3729	-0.0884
547.2/3015	-0.0881
576.3/2860	-0.0844

·
·
·

主成分係数の値が大きな上位いくつまでが重要な変数、
もしくは主成分係数の値がいくつまでを
重要な変数と考えれば良いか？

主成分負荷量が正または負に大きな上位の変数

正に値が大きな変数上位10個

ピーク 主成分負荷量

384.2/3993	0.7489
491.2/3398	0.7375
449.1/3290	0.7007
322.1/3392	0.6997
439.3/4056	0.6638
438.3/4056	0.6597
301.2/3389	0.6352
354.2/3618	0.6299
300.2/3392	0.6293
410.3/3937	0.6228

⋮

0.7以上が3変数

0.4以上が38変数

負に値が大きな変数上位10個

ピーク 主成分負荷量

306.1/2928	-0.6628
414.2/3060	-0.5523
246.1/2517	-0.5498
591.3/3003	-0.5419
288.1/2798	-0.5402
590.3/3003	-0.5365
546.3/3015	-0.5200
532.3/3729	-0.5177
547.2/3015	-0.5156
576.3/2860	-0.4944

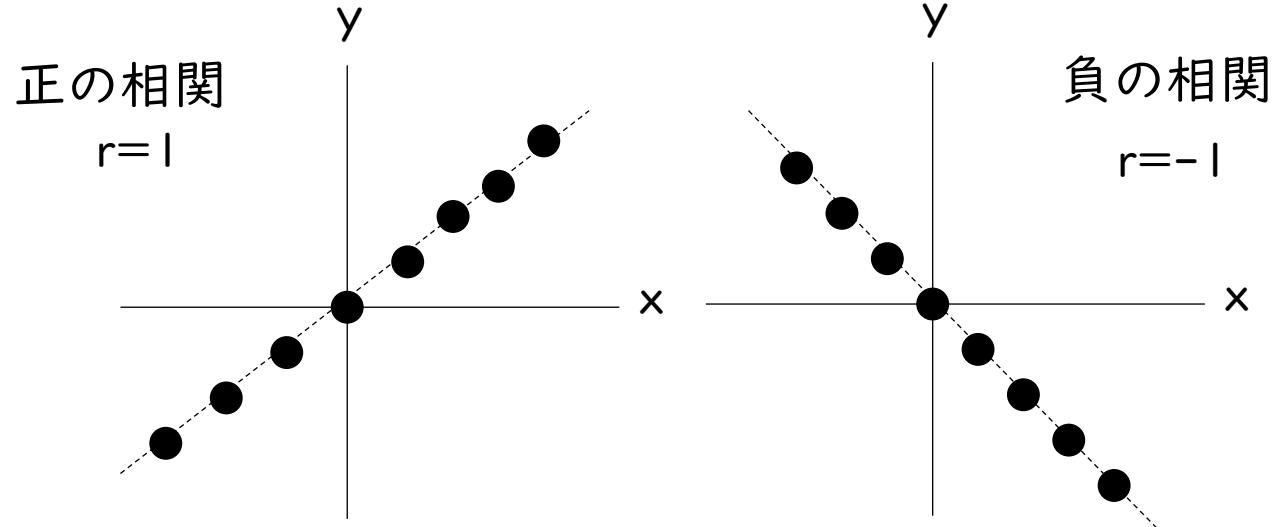
⋮

-0.7以下はなし

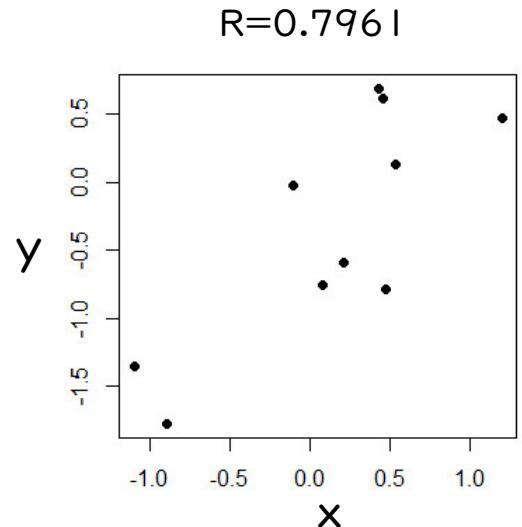
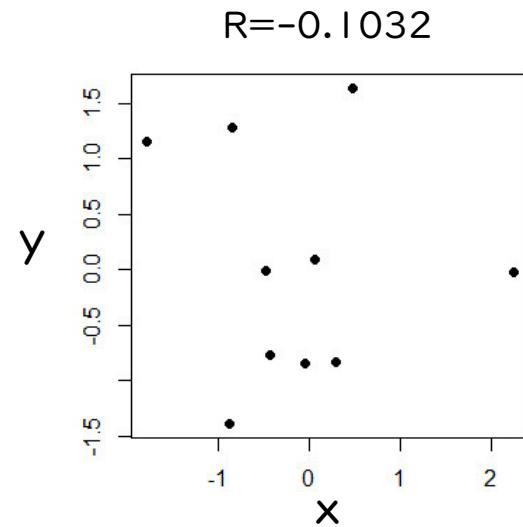
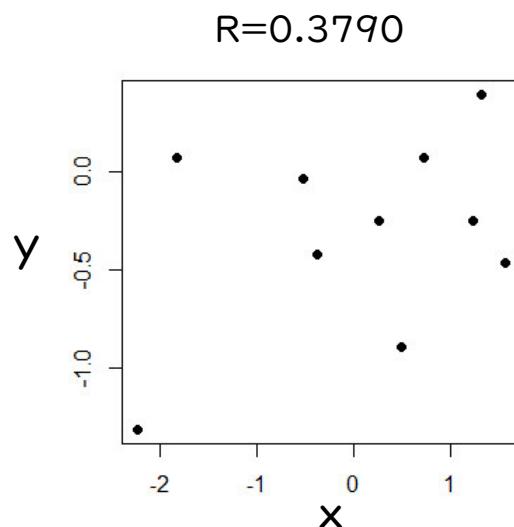
-0.4以下が38変数

上位10個の変数を選ぶのが一概に良いとは言えない

相関係数とは



実際に計算した例



主成分負荷量の検定により有意な代謝物を選ぶ

- 主成分負荷量の定義

- 主成分スコアと各変数の相関係数(教科書的な定義)
- 固有ベクトルが相関係数と比例するのは、データをautoscalingした時に限り、主成分負荷量として使用する際には変換が必要

- ピアソンの相関係数の仮説検定

$$t = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1 - r_{xy}^2}} \quad n : \text{サンプルサイズ}$$
$$r_{xy} = \text{corr}(x, y)$$

は自由度n-2のt分布に従う

主成分負荷量の統計的仮説検定を行える

METHODOLOGY ARTICLE

Open Access

Statistical hypothesis testing of factor loading in principal component analysis and its application to metabolite set enrichment analysis

Hiroyuki Yamamoto*, Tamaki Fujimori, Hajime Sato, Gen Ishikawa, Kenjiro Kami and Yoshiaki Ohashi

Abstract

Background: Principal component analysis (PCA) has been widely used to visualize high-dimensional metabolomic data in a two- or three-dimensional subspace. In metabolomics, some metabolites (e.g., the top 10 metabolites) have been subjectively selected when using factor loading in PCA, and biological inferences are made for these metabolites. However, this approach may lead to biased biological inferences because these metabolites are not objectively selected with statistical criteria.

Results: We propose a statistical procedure that selects metabolites with statistical hypothesis testing of the factor loading in PCA and makes biological inferences about these significant metabolites with a metabolite set

主成分負荷量の統計的仮説検定を用いて有意な代謝物を選び、代謝パスウェイデータベースを参照(エンリッチメント解析)し、統計的に有意な代謝パスウェイを選ぶ方法を提案

データの準備

- CSVファイルの読み込み(実行済み)

```
file <- "C:/R/mouse_data_original.csv"  
X0 <- read.csv(file, skip=1) # 1行目をスキップ
```

- データの準備

```
X <- X0[,-1] # 1列目の群情報の文字列を除く  
class <- c(1,1,1,1,1,1,2,2,2,2,2,2) # 群情報  
X <- t(X) # データの転置  
X <- scale(X) # スケーリング
```

- 主成分分析の計算

```
pca <- prcomp(X, scale=TRUE)
```

主成分分析(Ⅰ)

- 主成分スコア(PC4)

```
PC_score <- pca$x
```

```
PC4_score <- PC_score[,4]
```

```
write.csv(PC4_score, file="C:/R/PC4_score.csv")
```

- 主成分係数、固有ベクトル(PC4)

```
v <- pca$rotation[,4]
```

```
names(v) <- X0[,1]
```

```
write.csv(v, file="C:/R/weight_PC4.csv")
```

主成分負荷量のp-valueの計算

- ライブラリの読み込み

```
library(loadings)
```

- 主成分負荷量の計算と仮説検定

```
pca <- pca_loading(pca)
```

```
PC4_loading <- pca$loading$R[,4] # 主成分負荷量
```

```
p_PC4 <- pca$loading$p.value[,4] # p値
```

- 主成分負荷量の出力

```
names(PC4_loading) <- X0[,1]
```

```
write.csv(PC4_loading, file="C:/R/loading_PC4.csv")
```

Excel(csvファイル)での確認手順

- 主成分負荷量の解析結果(csvファイル)

weight_PC4.csv

PC4の主成分係数(B列)

B列の二乗和を計算
=SUMSQ(B2:B最終行)

A	B	C	D	E	F	G	H	I
1	x							
2	200.1/2926	-0.063572219						
3	205/2791	-0.007743185						
4	206/2791	-0.038441572						
5	207.1/2719	-0.00384172						
6	219.1/2524	-0.04998422						
	⋮							

PC4の係数の二乗和が1になっていることから、
主成分係数、固有ベクトルであることが確認できる

固有ベクトルから主成分負荷量(相関係数) への変換方法

主成分負荷量(相関係数)
= 主成分スコアの分散の平方根 × 主成分係数

PC4_score.csv

	A	B
2	R1 zaMt 1	-10.3932
3	R1 zaMt 2	-9.74371
4	R1 zaMt 3	-9.2806
5	R1 zaMt 4	-9.36079
6	R1 zaMt 5	-8.98823
7	R1 zaMt 6	-7.52343
8	R2 zaMt 0	-11.1691
9	R2 zaMt 1	-10.8171
10	R2 zaMt 2	-10.6018
11	R2 zaMt 3	-10.0279
12	R2 zaMt 4	-9.59622
13	R2 zaMt 5	-9.49064
14	R2 zaMt 6	-9.54685
15	R3_1 zaMt 0	-11.9542
16	R3_1 zaMt 1	-11.2057
17	R3_1 zaMt 2	-11.3681
18	R3_1 zaMt 3	-10.2499
19	R3_1 zaMt 4	-9.50741
20	R3_1 zaMt 5	-9.94565
21	R3_1 zaMt 6	-8.98555

PC4_score.csvは事前に出力しておく

```
write.csv(PC4_score, file="C:/R/PC4_score.csv")
```

読み込むファイルに関する注意

変数

fasting.csv

サンプル

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1 Sample	fasting	gamma-G	gamma-G	gamma-B	beta-Ala	Xanthosin	Xanthopte	Xanthine	Val	UTP	Urocanic	Uridine	Uric acid	Urea	Uracil	UMP
2 L1-1	0	0.00213	0.00003	0.00225	0.00698	0.00204	0.00012	0.00834	0.02848	0.00002	0.00002	0.00929	0.00183	0.14336	0.00148	0.00639
3 L1-2	0	0.00212	0.00004	0.0022	0.01274	0.00138	0.0001	0.00506	0.04297	0.00002	0.00002	0.0107	0.00359	0.20963	0.0017	0.00657
4 L1-3	0	0.00115	0.00004	0.00193	0.00995	0.00053	0.00009	0.00427	0.03943	0.00004	0.00002	0.00681	0.0022	0.2026	0.00114	0.00975
5 L1-4	0	0.00112	0	0.00154	0.00762	0.00076	0.00008	0.00605	0.04423	0.00003	0.00002	0.00663	0.00225	0.15643	0.00108	0.00683
6 L1-5	0	0.00076	0	0.00143	0.00626	0.00093	0.00005	0.00547	0.04475	0.00006	0	0.00704	0.00311	0.13072	0.00118	0.00913
7 L2-1	12	0.0005	0.00007	0.00153	0.01778	0.00099	0	0.00903	0.03792	0.00002	0.00018	0.00643	0.0024	0.14257	0.00113	0.00709
8 L2-2	12	0.00046	0.00007	0.0011	0.01317	0.00058	0.00003	0.00824	0.03474	0.00001	0.00004	0.00649	0.00242	0.10548	0.00111	0.00741
9 L2-3	12	0.00076	0.0001	0.0011	0.01042	0.00043	0.00004	0.00566	0.03678	0	0.00003	0.00672	0.00224	0.11282	0.0012	0.00999
10 L2-4	12	0.00037	0.00005	0.00093	0.01232	0.00051	0.00004	0.00785	0.02622	0.00002	0.00005	0.00611	0.00351	0.13681	0.00109	0.00819
11 L2-5	12	0.00022	0.00004	0.00113	0.01408	0.00035	0	0.00731	0.03029	0.00002	0.00004	0.00524	0.00258	0.13154	0.00101	0.00827
12																

mouse_data_original.csv

変数

A	B	C	D	E	F	G	H	I	J	K	L	M
1	ko15	ko16	ko18	ko19	ko21	ko22	wt15	wt16	wt18	wt19	wt21	wt22
2 Label	KO	KO	KO	KO	KO	KO	WT	WT	WT	WT	WT	WT
3 200.1/2926	147887.5	451600.7	65290.38	56540.93	85146.33	162012.4	175177.1	82619.48	51951.61	69198.22	153273.5	98144.28
4 205/2791	1778569	1567038	1482796	1039130	1223132	1072038	1950287	1466781	1572679	1275313	1356014	1231442
5 206/2791	237993.6	269714	201393.4	150107.3	176989.7	156797	276541.8	222366.2	211717.7	186850.9	188285.9	172348.8
6 207.1/2719	380873	460629.7	351750.1	219288	286848.6	235022.6	417169.6	324892.5	277990.7	220972.4	236728.2	
7 219.1/2524	235544.9	173623.4	82364.59	79480.4	185792.4	174458.8	244584.5	161184.1	72029.38	75096.99	238194.4	173830
8 231/2516	117649.8	48960.63	226609.1	286232.2	435094.5	62168.71	465898	61234.44	96841.46	240261.2	201316.2	179437.7
9 233/3023	399145.3	356951.3	410550.7	198416.5	363381.7	317805.23	397107.8	271252.1	334459.9	181901.3	456900.5	294270.6
10 234/3024	76880.87	99526.27	97493.76	53461.71	88227.79	81072.23	65215.84	55952.44	73781.01	45211.66	83693.39	57516.2
11 235.1/2695	171995.2	128945.2	155442.5	115286.3	37769.45	7663.88	199981.5	30028.6	156968.3	52596.48	14641.59	27713.29
12 236.1/2524	252282	2060319	71763.79	73602.47	186661	198804.3	253791.1	187225.7	73989.63	90012.64	256263.3	206487
13 240.2/3682	112440.6	90466.98	193768.7	170641.5	81223.74	146653	43974.38	270872.9	176425.2	187063	75235.94	86068.18
14 241.1/3879	1465989	1318747	1215369	632037.4	579968.2	561964.9	1468103	1594705	1006929	805533.2	731777.6	522916.1
15 242.1/3679	280767.6	248792.4	224467.7	109019.4	103841.9	101092.9	280280.2	299453.2	188328.2	139748.3	139968.5	95347.85
16 244.1/2832	612169.9	256316.5	90539.86	33269.29	54610.78	115505.4	627835	56883.71	168524.5	27260.37	110170.7	1352930
17 246.1/2517	27932.12	51557.09	38018.76	41747.89	56145.39	47901.55	105508.8	70508.82	46090.83	45630.55	77205.36	48624.55
18 249.1/3668	1435001	1228148	1193347	641881.8	536808.3	574005.7	1297986	1566269	1076654	747969.8	688496.3	485427.6
19 250.1/3668	347794.8	238893.4	248245.6	126623.8	107844.1	118627	281745.8	336058.8	215534.9	149872.8	139317.3	99814.5
20 254.1/3231	78911.6	1564680.8	8937	127186.8	21055.46	24185.93	54645.64	124200.9	108479	30096	66315.96	27554.78
21 255.2/3679	1420403	1187752	1264222	673816.2	581177.9	651560.3	1211727	1550071	1204087	875296.3	759075.9	640885.7
22 256.2/3679	307708.4	263317.7	275104.9	183168.4	121284.4	128540.4	258452.2	338390.9	213618.1	180384.5	164675.7	91822.7

主成分スコアと各変数の相関係数(1)

主成分スコア \mathbf{t} と各変数 \mathbf{X}_i との相関係数 r_i

$$r_i = \text{corr}(\mathbf{t}, \mathbf{X}_i) = \frac{\text{cov}(\mathbf{t}, \mathbf{X}_i)}{\sqrt{\text{var}(\mathbf{t})} \sqrt{\text{var}(\mathbf{X}_i)}} = \frac{\frac{1}{n-1} \mathbf{t}' \mathbf{X}_i}{\sqrt{\frac{1}{n-1} \mathbf{t}' \mathbf{t}} \sqrt{\frac{1}{n-1} \mathbf{X}_i' \mathbf{X}_i}}$$

$\mathbf{t} = \mathbf{X}\mathbf{w}$ を代入 ただし、 \mathbf{t} と \mathbf{X}_i の平均は0

$$r_i = \frac{\frac{1}{n-1} \mathbf{w}' \mathbf{X}' \mathbf{X}_i}{\sqrt{\frac{1}{n-1} \mathbf{w}' \mathbf{X}' \mathbf{X} \mathbf{w}}} \sqrt{\frac{1}{n-1} \mathbf{X}_i' \mathbf{X}_i} = \frac{\frac{1}{n-1} \mathbf{w}' \mathbf{X}' \mathbf{X} \mathbf{c}}{\sqrt{\frac{1}{n-1} \mathbf{w}' \mathbf{X}' \mathbf{X} \mathbf{w}}} \sqrt{\frac{1}{n-1} \mathbf{X}_i' \mathbf{X}_i}$$

ただし、 \mathbf{c} は i 番目の要素が1で、それ以外が0の縦ベクトル
 $\mathbf{c} = [0, 0, 0, \dots, 0, \underset{i\text{番目の要素}}{1}, 0, 0]'$

i番目の要素

主成分スコアと各変数の相関係数(2)

$$\frac{1}{n-1} \mathbf{X}' \mathbf{X} \mathbf{w} = \lambda \mathbf{w} \text{ を転置した } \boxed{\frac{1}{n-1} \mathbf{w}' \mathbf{X}' \mathbf{X}} = \lambda \mathbf{w}' \text{ を代入して}$$

$$r_i = \frac{\frac{1}{n-1} \mathbf{w}' \mathbf{X}' \mathbf{X} \mathbf{c}}{\sqrt{\frac{1}{n-1} \mathbf{w}' \mathbf{X}' \mathbf{X} \mathbf{w}} \sqrt{\frac{1}{n-1} \mathbf{X}_i' \mathbf{X}_i}} = \frac{\lambda \mathbf{w}' \mathbf{c}}{\sqrt{\lambda \mathbf{w}' \mathbf{w}} \sqrt{\frac{1}{n-1} \mathbf{X}_i' \mathbf{X}_i}}$$

$\mathbf{w}' \mathbf{w} = 1$ と $\mathbf{w}' \mathbf{c} = w_i$ を代入して

$$r_i = \frac{\lambda w_i}{\sqrt{\lambda} \sqrt{\frac{1}{n-1} \mathbf{X}_i' \mathbf{X}_i}} = \frac{\sqrt{\lambda} w_i}{\sqrt{\frac{1}{n-1} \mathbf{X}_i' \mathbf{X}_i}}$$

各変数をautoscaling(平均0 分散1)した時は、次のように書ける

$$r_i = \sqrt{\lambda} w_i \quad \boxed{\mathbf{w} \propto \text{corr}(\mathbf{t}, \mathbf{x}_i)}$$

固有ベクトルは、主成分スコアと各変数の相関係数に比例する