

## 1 3. 多変量解析 2 (主成分分析)

### 主成分分析

2

複数の項目が持つ共通の特徴を集約し合成変数を作る

$$z_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m$$

$$z_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m$$

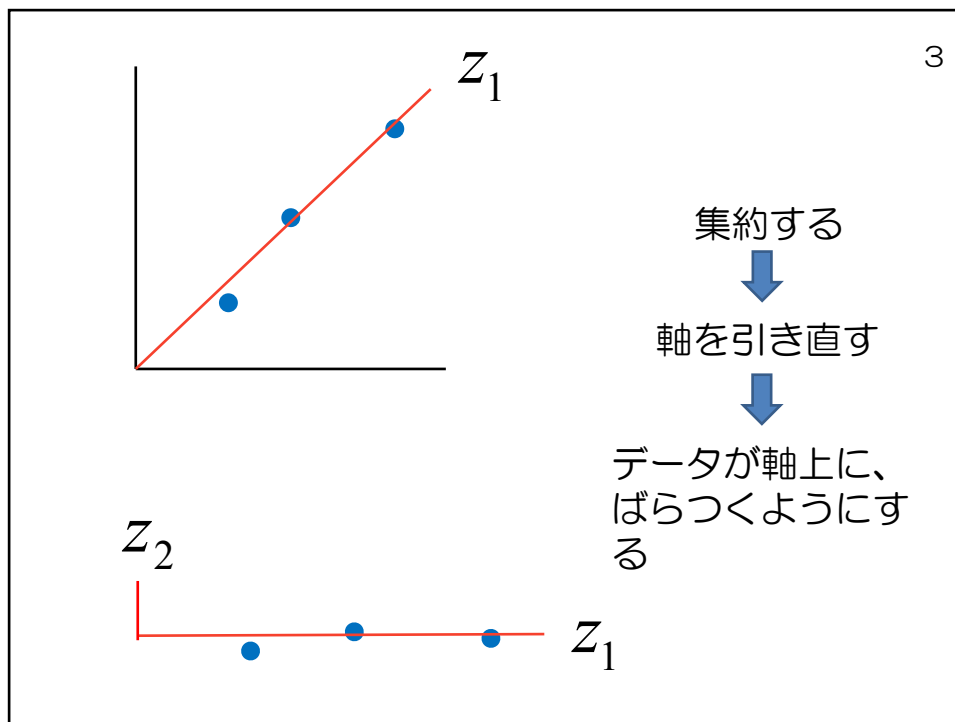
⋮

$$z_n = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m$$

$$\sum_{i=1}^m a_{im}^2 = 1$$

$z_i$  : 第*i*主成分

$x_i$  : 基準化された特性値



4

主成分分析のためのRの関数

- princomp()
- prcomp()
- principal()  
psych/パッケージ
- pca()  
青木先生の関数

5

英語	国語	社会	数学	理科	英語	国語	社会	数学	理科
60	65	76	54	45	85	73	77	85	91
70	68	67	66	56	90	82	84	93	94
43	56	54	23	34	77	63	75	74	84
82	78	78	45	69	69	58	62	76	67
89	89	93	78	82	56	63	45	72	82
67	78	92	56	45	59	73	67	54	62
87	92	70	69	72	79	87	65	83	85
72	58	74	78	68	68	66	74	80	78
67	82	77	67	62	82	72	81	89	93
58	67	84	78	83	72	78	83	95	89

6

```
eigo <- c(60, 70, 43, 82, 89, 67, 87, 72, 67, 58, 85, 90, 77, 69, 56, 59, 79, 68, 82, 72)
kokugo <- c(65, 68, 56, 78, 89, 78, 92, 58, 82, 67, 73, 82, 63, 58, 63, 73, 87, 66, 72, 78)
syakai <- c(76, 67, 54, 78, 93, 92, 70, 74, 77, 84, 77, 84, 75, 62, 45, 67, 65, 74, 81, 83)
sugaku <- c(54, 66, 23, 45, 78, 56, 69, 78, 67, 78, 87, 93, 74, 76, 72, 54, 83, 80, 89, 95)
rika <- c(45, 56, 34, 69, 82, 45, 72, 68, 62, 83, 91, 94, 84, 67, 82, 62, 85, 78, 93, 89)
kyoka <- cbind(eigo, kokugo, syakai, sugaku, rika)
colnames(kyoka) <- c("英語", "国語", "社会", "数学", "理科")
rownames(kyoka) <- 1:20
```

	英語	国語	社会	数学	理科
1	60	65	76	54	45
2	70	68	67	66	56
3	43	56	54	23	34
4	82	78	78	45	69
5	89	89	93	78	82
6	67	78	92	56	45

## 主成分分析

7

```
source("all.R", encoding="euc-jp")
a <- pca(kyoka)
print(a, npca=5)
```

主成分負荷量

## 主成分分析の結果

	PC1	PC2	PC3	PC4		
英語	-0.890	0.137	-0.161	0.401	-0.000	
国語	-0.681	0.539	-0.431	-0.245	-0.000	
社会	-0.627	0.523	0.572	-0.041	0.000	
数学	-0.818	-0.488	0.150	-0.139	-0.000	
理科	-0.821	-0.509	-0.054	-0.061	0.000	
固有値	Eigenvalue	2.992	1.081	0.564	0.246	0.000
寄与率	Contribution	0.598	0.216	0.113	0.049	0.000
累積寄与率	Cum. contrib.	0.598	0.815	0.927	0.976	1.000

## 主成分負荷量

8

主成分と変数との相関係数

## 主成分得点

主成分を軸とした場合の変数の座標

## 固有値

データの分散共分散行列（あるいは相関係数行列）の固有値

主成分得点の標準偏差の2乗

固有値 1 以上

9

print(a)

主成分分析の結果

主成分負荷量

	PC1	PC2	Contribution
英語	-0.890	0.137	0.811
国語	-0.681	0.539	0.754
社会	-0.627	0.523	0.668
数学	-0.818	-0.488	0.907
理科	-0.821	-0.509	0.933
Eigenvalue	2.992	1.081	
Contribution	0.598	0.216	
Cum. contrib.	0.598	0.815	

固有値

寄与率

累積寄与率

詳細

10

summary(a)

```

$mean
  英語  国語  社会  数学  理科
71.60 72.40 73.90 70.85 72.05

$variance
  英語  国語  社会  数学  理科
155.94 113.52 138.62 309.19 297.52

$standard.deviation
  英語  国語  社会  数学  理科
12.487 10.654 11.774 17.584 17.249

```

\$r 相関係数					
	英語	国語	社会	数学	理科
英語	1.00000	0.65201	0.51878	0.59152	0.63394
国語	0.65201	1.00000	0.47193	0.26723	0.31892
社会	0.51878	0.47193	1.00000	0.33576	0.23379
数学	0.59152	0.26723	0.33576	1.00000	0.86421
理科	0.63394	0.31892	0.23379	0.86421	1.00000

\$factor. loadings 主成分負荷量					
	PC1	PC2	PC3	PC4	PC5
英語	-0.89028	0.13696	-0.161263	0.400707	-0.045573
国語	-0.68111	0.53859	-0.430674	-0.245450	-0.017021
社会	-0.62735	0.52346	0.572245	-0.041279	0.057047
数学	-0.81754	-0.48835	0.149789	-0.138789	-0.226820
理科	-0.82069	-0.50922	-0.054288	-0.061170	0.245905

\$eval 固有値					
	PC1	PC2	PC3	PC4	PC5
	2.99197	1.08064	0.56433	0.24552	0.11754

\$contribution 寄与率					
	PC1	PC2	PC3	PC4	PC5
	59.8394	21.6128	11.2867	4.9104	2.3508

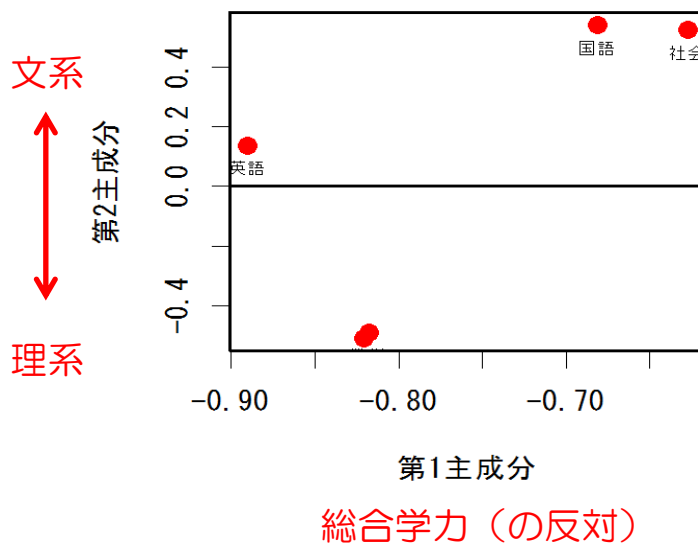
  

\$cum. contribution 累積寄与率					
	PC1	PC2	PC3	PC4	PC5
	59.839	81.452	92.739	97.649	100.000

\$\$fs	主成分得点				
	PC1	PC2	PC3	PC4	PC5
1	1.932830	0.847408	0.672752	0.041021	-0.311066
2	1.039278	0.060974	-0.174331	0.350787	-0.559063
3	4.853628	0.419238	-0.304299	0.087929	0.326023
4	0.017248	1.369304	-0.508062	0.838983	0.786858
5	-2.446876	1.368759	0.084960	0.035688	0.156042
6	0.583051	2.219032	0.916957	-0.262836	-0.294875
7	-1.218730	1.025608	-1.633851	0.147186	-0.249166
8	0.443060	-0.787704	0.895151	0.625634	-0.382871
9	0.122333	0.963554	-0.244666	-0.649355	-0.230854
10	-0.045548	-0.488473	1.244537	-0.916614	0.511953
11	-1.667552	-0.683815	0.042755	0.436114	0.080767
12	-2.691615	-0.125381	-0.020719	0.168111	-0.018997
13	-0.330803	-0.796793	0.482000	0.659968	0.390144
14	1.032558	-1.262780	0.132531	0.553450	-0.489564

主成分負荷量の散布図

plot(a, which="loadings", cex=1.5, col="red")



主成分得点の散布図

15

```
plot(a, which="scores", cex=1.5, col="red")
text(a$fs, labels=rownames(kyoka), pos=3)
```

