

# Rによる主成分分析

---

- Rによる主成分分析
  - 手順①(データセットを読み込む)
  - 手順②(`princomp(<データセット>)`を実行)
  - 各主成分得点の寄与率の確認
  - スクリーンプロット
  - 重み付け係数の確認
  - 主成分得点の確認
  - 元の変数と主成分得点の相関係数の確認

# 仮想データセット

|    | A  | B    | C      | D      | E    | F      |
|----|----|------|--------|--------|------|--------|
| 1  | id | eigo | sugaku | kokugo | rika | syakai |
| 2  | 1  | 77   | 73     | 87     | 58   | 71     |
| 3  | 2  | 78   | 80     | 91     | 64   | 72     |
| 4  | 3  | 65   | 74     | 78     | 46   | 64     |
| 5  | 4  | 71   | 98     | 76     | 65   | 58     |
| 6  | 5  | 63   | 82     | 81     | 49   | 67     |
| 7  | 6  | 72   | 85     | 80     | 55   | 66     |
| 8  | 7  | 65   | 77     | 85     | 53   | 62     |
| 9  | 8  | 73   | 58     | 77     | 40   | 60     |
| 10 | 9  | 74   | 79     | 82     | 45   | 63     |
| 11 | 10 | 69   | 71     | 93     | 47   | 75     |
| 12 | 11 | 70   | 68     | 74     | 50   | 65     |
| 13 | 12 | 79   | 100    | 89     | 61   | 74     |
| 14 | 13 | 52   | 48     | 68     | 43   | 53     |
| 15 | 14 | 67   | 53     | 71     | 39   | 57     |
| 16 | 15 | 65   | 76     | 84     | 52   | 61     |
| 17 | 16 | 76   | 60     | 86     | 41   | 73     |
| 18 | 17 | 62   | 56     | 83     | 44   | 68     |
| 19 | 18 | 75   | 72     | 73     | 51   | 69     |
| 20 | 19 | 69   | 81     | 94     | 48   | 77     |
| 21 | 20 | 69   | 64     | 70     | 42   | 59     |

# Rによる主成分分析

- 事前準備  
とくになし
- 手順
  - ① データセットを読み込む
  - ② `princomp(<データセット>)`を実行  
たったこれだけ！

※各主成分得点の寄与率や累積寄与率をみるには、②の実行結果を代入したオブジェクトを、`summary()`で見る

※スクリープロットをみるには、②の実行結果を代入したオブジェクトを、`screeplot()`で見る


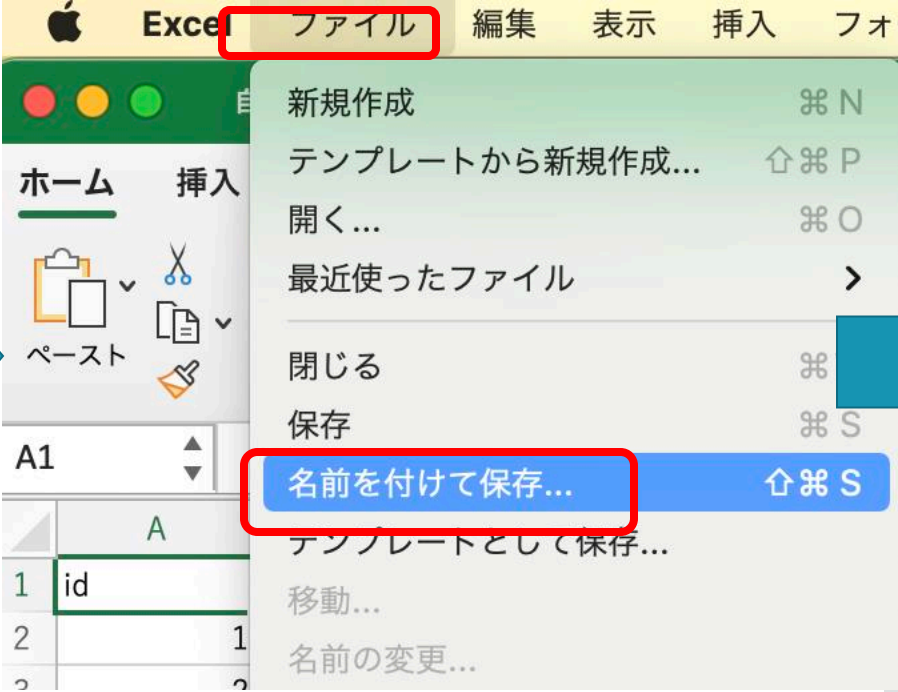
※重み付け係数をみるには、②の実行結果を代入したオブジェクトに、`$loadings`を付けて見る

※主成分得点をみるには、②の実行結果を代入したオブジェクトに、`$scores`を付けて見る

# 手順① (データセットを読み込む)

EXCELで先ほどの仮想データセットをつかって、  
**csv形式で**、ワーキングディレクトリ(作業用フォルダ)に保存しましょう

|    | A  | B    | C      | D      | E    | F      |
|----|----|------|--------|--------|------|--------|
| 1  | id | eigo | sugaku | kokugo | rika | syakai |
| 2  | 1  | 77   | 73     | 87     | 58   | 71     |
| 3  | 2  | 78   | 80     | 91     | 64   | 72     |
| 4  | 3  | 65   | 74     | 78     | 46   | 64     |
| 5  | 4  | 71   | 98     | 76     | 65   | 58     |
| 6  | 5  | 63   | 82     | 81     | 49   | 67     |
| 7  | 6  | 72   | 85     | 80     | 55   | 66     |
| 8  | 7  | 65   | 77     | 85     | 53   | 62     |
| 9  | 8  | 73   | 58     | 77     | 40   | 60     |
| 10 | 9  | 74   | 79     | 82     | 45   | 63     |
| 11 | 10 | 69   | 71     | 93     | 47   | 75     |
| 12 | 11 | 70   | 68     | 74     | 50   | 65     |
| 13 | 12 | 79   | 100    | 89     | 61   | 74     |
| 14 | 13 | 52   | 48     | 68     | 43   | 53     |
| 15 | 14 | 67   | 53     | 71     | 39   | 57     |
| 16 | 15 | 65   | 76     | 84     | 52   | 61     |
| 17 | 16 | 76   | 60     | 86     | 41   | 73     |
| 18 | 17 | 62   | 56     | 83     | 44   | 68     |
| 19 | 18 | 75   | 72     | 73     | 51   | 69     |
| 20 | 19 | 69   | 81     | 94     | 48   | 77     |
| 21 | 20 | 69   | 64     | 70     | 42   | 59     |



ワーキングディレクトリに保存する

※注意:環境によっては、日本語は文字化けします

あとは、下記を実行すればOKでしたね!

```
dat <- read.csv("testScore.csv", header=T, row.names=1, na.strings=".")
```

1行目は列ラベルとして読み込むよう指定

1列目は行ラベルとして読み込むよう指定

半角ピリオドを欠損値として読み込むよう指定

row.names=1は指定しなくても今回は動きます。

今回も欠損値はありませんが、クセにしておくの良いです

## 手順② (princomp(<データセット>))を実行

```
> dat <- read.csv("testScore.csv", header=T, row.names=1, na.strings=".")
> dat
```

```
      eigo sugaku kokugo rika syakai
1      77     73     87   58     71
2      78     80     91   64     72
3      65     74     78   46     64
4      71     98     76   65     58
5      63     82     81   49     67
```

...(後略)

たったこれだけ!

```
> princomp(dat)
Call:
princomp(x = dat)
```

Standard deviations:

```
      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5
15.975419   8.346407   5.148363   4.057466   2.572413
```

```
5 variables and 20 observations.
```

各主成分得点の標準偏差  
(標本分散の正の平方根)

| H                       | I      | J      | K     | L     | M     |
|-------------------------|--------|--------|-------|-------|-------|
|                         | 第1主成分  | 第2主成分  | 第3主成分 | 第4主成分 | 第5主成分 |
|                         | 8.20   | 6.92   | 4.60  | -7.19 | 0.48  |
|                         | 17.87  | 6.51   | 3.15  | -9.68 | 1.48  |
|                         | -2.81  | -3.79  | -3.13 | 3.76  | -1.36 |
|                         | 23.68  | -19.86 | 0.66  | -1.95 | 1.08  |
|                         | 5.91   | -4.09  | -7.12 | 4.40  | -2.98 |
|                         | 12.35  | -5.45  | 1.24  | 1.07  | -1.04 |
|                         | 4.01   | -3.14  | -6.16 | -1.97 | 3.34  |
|                         | -17.41 | 2.16   | 6.34  | 2.27  | 4.33  |
|                         | 3.87   | -1.25  | 2.52  | 6.94  | 4.19  |
|                         | 2.90   | 14.19  | -5.31 | 0.67  | -0.06 |
|                         | -5.79  | -3.15  | 4.04  | -1.65 | -3.36 |
|                         | 32.88  | 0.13   | 2.02  | 2.35  | -1.02 |
|                         | -33.37 | -9.96  | -7.25 | -5.52 | -1.44 |
|                         | -25.66 | -2.91  | 4.09  | 1.06  | 1.72  |
|                         | 2.29   | -3.81  | -5.65 | -1.47 | 3.66  |
|                         | -9.20  | 15.52  | 5.11  | 2.14  | 0.27  |
|                         | -16.49 | 7.96   | -5.53 | -3.09 | -1.75 |
|                         | -0.39  | -1.78  | 8.46  | -0.10 | -5.80 |
|                         | 12.10  | 12.22  | -7.03 | 4.27  | -1.33 |
|                         | -14.96 | -6.43  | 4.95  | 3.69  | -0.40 |
| 分散                      | 255.21 | 69.66  | 26.51 | 16.46 | 6.62  |
| 元の情報量を表現<br>できた割合 (寄与率) | 68.15% | 18.60% | 7.08% | 4.40% | 1.77% |



※各主成分得点の寄与率や累積寄与率をみるには、  
②の実行結果を代入したオブジェクトを、summary()で見る

寄与率

累積  
寄与率

```
> result <- princomp(dat)
> summary(result)
Importance of components:
              Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
Standard deviation 15.9754188  8.3464072  5.14836328  4.05746603  2.57241271
Proportion of Variance 0.6815476 0.1860333 0.07078317 0.04396443 0.01767148
Cumulative Proportion 0.6815476 0.8675809 0.93836409 0.98232852 1.00000000
```

第1主成分得点と第2主成分得点の  
2つ合わせて、元の5変数の情報(分散)  
の約86.76%を表現できている  
(0.6815476...+0.1860333...≐0.8675809)

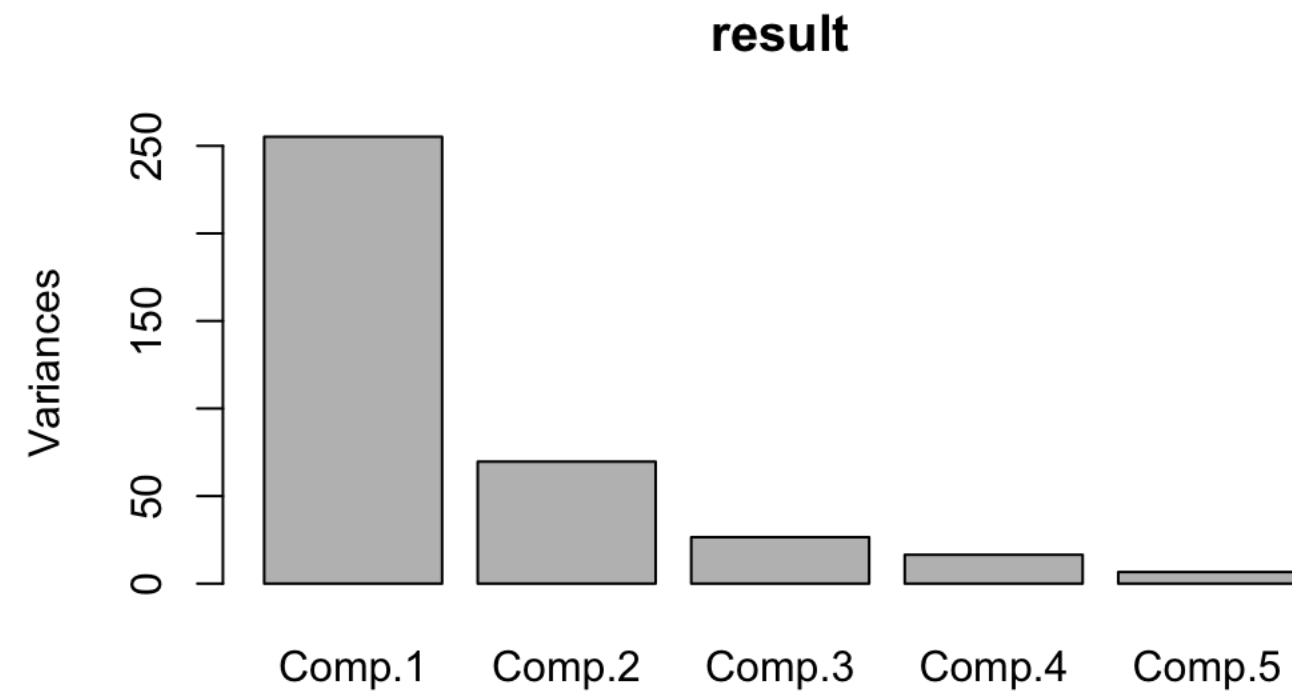
論文等での報告の仕方の例：  
5教科の成績について主成分分析を適用したところ、第1主成分の  
寄与率が約68.15%であった。

| H                       | I      | J      | K     | L     | M     |
|-------------------------|--------|--------|-------|-------|-------|
|                         | 第1主成分  | 第2主成分  | 第3主成分 | 第4主成分 | 第5主成分 |
|                         | 8.20   | 6.92   | 4.60  | -7.19 | 0.48  |
|                         | 17.87  | 6.51   | 3.15  | -9.68 | 1.48  |
|                         | -2.81  | -3.79  | -3.13 | 3.76  | -1.36 |
|                         | 23.68  | -19.86 | 0.66  | -1.95 | 1.08  |
|                         | 5.91   | -4.09  | -7.12 | 4.40  | -2.98 |
|                         | 12.35  | -5.45  | 1.24  | 1.07  | -1.04 |
|                         | 4.01   | -3.14  | -6.16 | -1.97 | 3.34  |
|                         | -17.41 | 2.16   | 6.34  | 2.27  | 4.33  |
|                         | 3.87   | -1.25  | 2.52  | 6.94  | 4.19  |
|                         | 2.90   | 14.19  | -5.31 | 0.67  | -0.06 |
|                         | -5.79  | -3.15  | 4.04  | -1.65 | -3.36 |
|                         | 32.88  | 0.13   | 2.02  | 2.35  | -1.02 |
|                         | -33.37 | -9.96  | -7.25 | -5.52 | -1.44 |
|                         | -25.66 | -2.91  | 4.09  | 1.06  | 1.72  |
|                         | 2.29   | -3.81  | -5.65 | -1.47 | 3.66  |
|                         | -9.20  | 15.52  | 5.11  | 2.14  | 0.27  |
|                         | -16.49 | 7.96   | -5.53 | -3.09 | -1.75 |
|                         | -0.39  | -1.78  | 8.46  | -0.10 | -5.80 |
|                         | 12.10  | 12.22  | -7.03 | 4.27  | -1.33 |
|                         | -14.96 | -6.43  | 4.95  | 3.69  | -0.40 |
| 分散                      | 255.21 | 69.66  | 26.51 | 16.46 | 6.62  |
| 元の情報量を表現<br>できた割合 (寄与率) | 68.15% | 18.60% | 7.08% | 4.40% | 1.77% |

※スクリープロットをみるには、  
②の実行結果を代入したオブジェクトを、`screeplot()`で見る

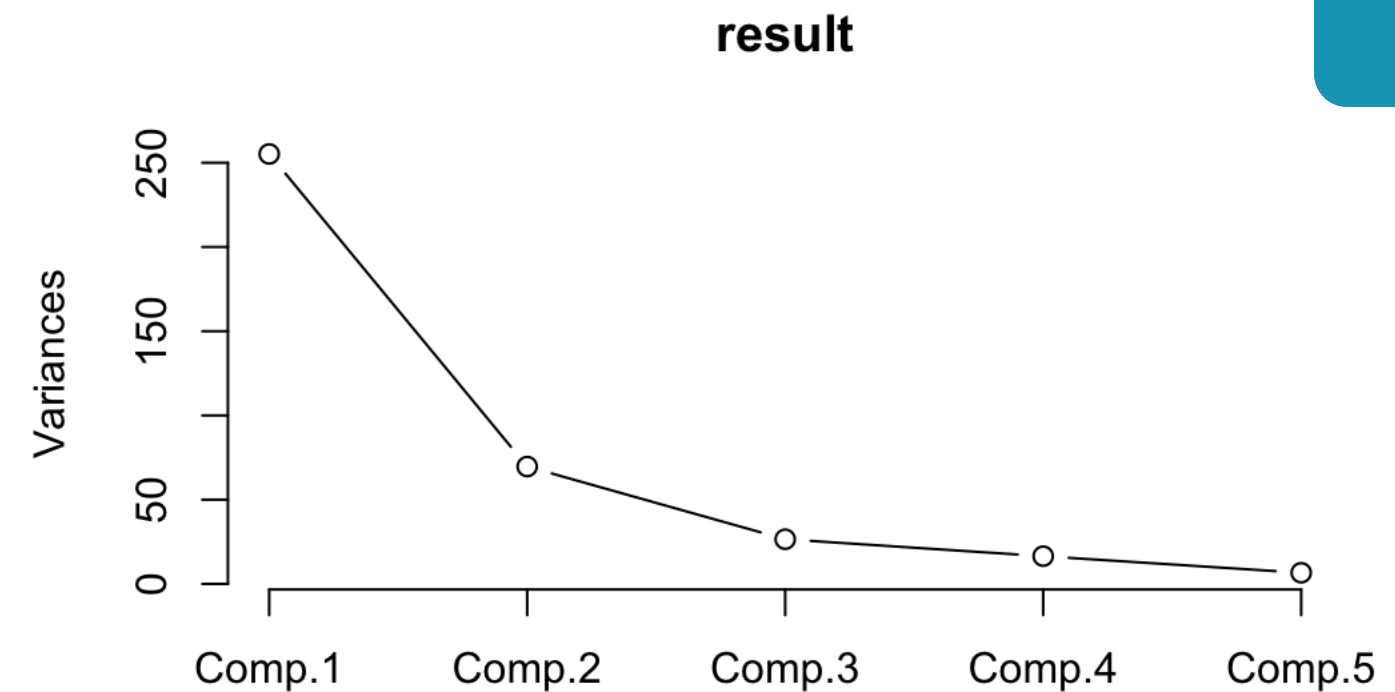
- 棒グラフ

```
> result <- princomp(dat)
> screeplot(result)
```



- 折れ線グラフ

```
> screeplot(result, type="l")
```



グラフの種類の指定  
(折れ線グラフ)

(参考)

resultという表タイトルを消したい場合は、  
さらに、`main=NULL`という引数を加えれば良い。  
(`main=""`でもOK)



※重み付け係数を見るには、  
②の実行結果を代入したオブジェクトに、\$loadingsを付けて見る

(参考)  
デフォルトでは、0.1以下の係数は非表示。  
全表示したい場合は print(<結果を代入したオブジェクト>, cutoff=0)

```
> result$loadings

Loadings:
      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
eigo    0.235  0.246  0.897      0.276
sugaku  0.806 -0.360 -0.137  0.449
kokugo  0.289  0.630 -0.419 -0.113  0.575
rika    0.403 -0.211      -0.883 -0.111
syakai  0.222  0.606      -0.762

      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
SS loadings      1.0   1.0   1.0   1.0   1.0
Proportion Var   0.2   0.2   0.2   0.2   0.2
Cumulative Var   0.2   0.4   0.6   0.8   1.0
```

重み付け係数の二乗和を示している。  
主成分分析においては常に1なので、  
あまり意味のない情報(気にしなくて良い)。

```
> print(result$loadings, cutoff=0)

Loadings:
      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
eigo    0.235  0.246  0.897  0.062  0.276
sugaku  0.806 -0.360 -0.137  0.449 -0.021
kokugo  0.289  0.630 -0.419 -0.113  0.575
rika    0.403 -0.211  0.047 -0.883 -0.111
syakai  0.222  0.606  0.006  0.053 -0.762

      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
SS loadings      1.0   1.0   1.0   1.0   1.0
Proportion Var   0.2   0.2   0.2   0.2   0.2
Cumulative Var   0.2   0.4   0.6   0.8   1.0
```

※主成分得点を見るには、  
②の実行結果を代入したオブジェクトに、\$scoresを付けて見る

(参考)  
桁数を調整したい場合は round(<結果を代入したオブジェクト>, digits=2)

```
> result <- princomp(dat)
> result$scores
```

|    | Comp.1      | Comp.2      | Comp.3     | Comp.4      | Comp.5      |
|----|-------------|-------------|------------|-------------|-------------|
| 1  | 8.2023838   | 6.9207514   | 4.5991465  | -7.18546167 | 0.47811287  |
| 2  | 17.8739603  | 6.5088126   | 3.1526074  | -9.67617483 | 1.47550351  |
| 3  | -2.8082148  | -3.7883555  | -3.1343547 | 3.76349711  | -1.36444220 |
| 4  | 23.6813668  | -19.8586475 | 0.6609076  | -1.95174559 | 1.08091462  |
| 5  | 5.9103788   | -4.0861151  | -7.1185941 | 4.40446501  | -2.98075348 |
| 6  | 12.3505066  | -5.4490204  | 1.2357778  | 1.07015104  | -1.04321314 |
| 7  | 4.0084304   | -3.1437313  | -6.1563259 | -1.96672354 | 3.34194385  |
| 8  | -17.4125738 | 2.1576447   | 6.3392835  | 2.26757293  | 4.32503667  |
| 9  | 3.8717675   | -1.2462143  | 2.5246992  | 6.93878203  | 4.18605597  |
| 10 | 2.9038602   | 14.1934832  | -5.3060676 | 0.66525455  | -0.05668379 |
| 11 | -5.7906768  | -3.1509103  | 4.0368633  | -1.64692587 | -3.36432709 |
| 12 | 32.8837494  | 0.1310715   | 2.0224536  | 2.34644951  | -1.01646428 |
| 13 | -33.3661234 | -9.9641788  | -7.2543510 | -5.51518621 | -1.43581500 |
| 14 | -25.6595442 | -2.9099752  | 4.0902387  | 1.05606534  | 1.72104865  |
| 15 | 2.2883385   | -3.8094683  | -5.6538244 | -1.47303223 | 3.66106601  |
| 16 | -9.1981937  | 15.5215268  | 5.1100194  | 2.13893163  | 0.27497793  |
| 17 | -16.4885053 | 7.9576494   | -5.5274058 | -3.09232370 | -1.75437414 |
| 18 | -0.3879502  | -1.7755769  | 8.4614213  | -0.09982322 | -5.80333578 |
| 19 | 12.0986229  | 12.2219982  | -7.0332189 | 4.26593453  | -1.33011918 |
| 20 | -14.9615830 | -6.4307441  | 4.9507242  | 3.69029321  | -0.39513201 |

```
> round(result$scores, digits=2)
```

|    | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|----|--------|--------|--------|--------|--------|
| 1  | 8.20   | 6.92   | 4.60   | -7.19  | 0.48   |
| 2  | 17.87  | 6.51   | 3.15   | -9.68  | 1.48   |
| 3  | -2.81  | -3.79  | -3.13  | 3.76   | -1.36  |
| 4  | 23.68  | -19.86 | 0.66   | -1.95  | 1.08   |
| 5  | 5.91   | -4.09  | -7.12  | 4.40   | -2.98  |
| 6  | 12.35  | -5.45  | 1.24   | 1.07   | -1.04  |
| 7  | 4.01   | -3.14  | -6.16  | -1.97  | 3.34   |
| 8  | -17.41 | 2.16   | 6.34   | 2.27   | 4.33   |
| 9  | 3.87   | -1.25  | 2.52   | 6.94   | 4.19   |
| 10 | 2.90   | 14.19  | -5.31  | 0.67   | -0.06  |
| 11 | -5.79  | -3.15  | 4.04   | -1.65  | -3.36  |
| 12 | 32.88  | 0.13   | 2.02   | 2.35   | -1.02  |
| 13 | -33.37 | -9.96  | -7.25  | -5.52  | -1.44  |
| 14 | -25.66 | -2.91  | 4.09   | 1.06   | 1.72   |
| 15 | 2.29   | -3.81  | -5.65  | -1.47  | 3.66   |
| 16 | -9.20  | 15.52  | 5.11   | 2.14   | 0.27   |
| 17 | -16.49 | 7.96   | -5.53  | -3.09  | -1.75  |
| 18 | -0.39  | -1.78  | 8.46   | -0.10  | -5.80  |
| 19 | 12.10  | 12.22  | -7.03  | 4.27   | -1.33  |
| 20 | -14.96 | -6.43  | 4.95   | 3.69   | -0.40  |

\*後で紹介しているprintを使っても桁数を調整できますが、printの場合は、有効数字2桁まで表示するようになるので、id=10の人のComp.5の-0.05668..という値を有効数字2桁まで表示するためにComp.5が小数第3位まで表示されるようになります。



# (参考)主成分得点の確かめ算

## ●行列の積を使った確かめ算

scale(<データセット>, scale=F)  
で中心化得点になる

|    | A  | B      |     |
|----|----|--------|-----|
| 1  | id | 第1主成分  | ... |
| 2  | 1  | 8.20   |     |
| 3  | 2  | 17.87  |     |
| 4  | 3  | -2.81  |     |
| 5  | 4  | 23.68  |     |
| 6  | 5  | 5.91   |     |
| 7  | 6  | 12.35  |     |
| 8  | 7  | 4.01   |     |
| 9  | 8  | -17.41 |     |
| 10 | 9  | 3.87   |     |
| 11 | 10 | 2.90   |     |
| 12 | 11 | -5.79  |     |
| 13 | 12 | 22.88  |     |

第1主成分得点  $\hat{=}$

$$0.24 \times \text{英語}_c + 0.81 \times \text{数学}_c + \\ 0.29 \times \text{国語}_c + 0.40 \times \text{理科}_c + \\ 0.22 \times \text{社会}_c$$

\* $c$  は中心化得点  
(主成分得点の平均=0になるように、  
データから当該変数の平均値をひいている)

例)

id=1の人の第1主成分得点  $\hat{=}$

$$0.24 \times (77 - 69.55) + 0.81 \times (73 - 72.75) + \\ 0.29 \times (87 - 81.10) + 0.40 \times (58 - 49.65) + \\ 0.22 \times (71 - 65.70) \\ \hat{=} 8.20$$

## ●確かめ算

> # (参考) 確かに、主成分得点が重み付け係数から算出した値と一致するか確認

> # 第1主成分得点

```
> result$loadings["eigo", "Comp.1"] * (dat$eigo - mean(dat$eigo)) +  
+ result$loadings["sugaku", "Comp.1"] * (dat$sugaku - mean(dat$sugaku)) +  
+ result$loadings["kokugo", "Comp.1"] * (dat$kokugo - mean(dat$kokugo)) +  
+ result$loadings["rika", "Comp.1"] * (dat$rika - mean(dat$rika)) +  
+ result$loadings["syakai", "Comp.1"] * (dat$syakai - mean(dat$syakai))
```

```
[1] 8.2023838 17.8739603 -2.8082148 23.6813668 5.9103788 12.3505066 4.0084304  
[8] -17.4125738 3.8717675 2.9038602 -5.7906768 32.8837494 -33.3661234 -25.6595442  
[15] 2.2883385 -9.1981937 -16.4885053 -0.3879502 12.0986229 -14.9615830
```

```
> t(t(result$loadings) %*% t(scale(dat, scale=F)))  
      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5  
1    8.2023838  6.9207514  4.5991465 -7.18546167  0.47811287  
2   17.8739603  6.5088126  3.1526074 -9.67617483  1.47550351  
3   -2.8082148 -3.7883555 -3.1343547  3.76349711 -1.36444220  
4   23.6813668 -19.8586475  0.6609076 -1.95174559  1.08091462  
5    5.9103788 -4.0861151 -7.1185941  4.40446501 -2.98075348  
6   12.3505066 -5.4490204  1.2357778  1.07015104 -1.04321314  
7    4.0084304 -3.1437313 -6.1563259 -1.96672354  3.34194385  
8  -17.4125738  2.1576447  6.3392835  2.26757293  4.32503667  
9    3.8717675 -1.2462143  2.5246992  6.93878203  4.18605597  
10   2.9038602 14.1934832 -5.3060676  0.66525455 -0.05668379  
11  -5.7906768 -3.1509103  4.0368633 -1.64692587 -3.36432709  
12  32.8837494  0.1310715  2.0224536  2.34644951 -1.01646428  
13 -33.3661234 -9.9641788 -7.2543510 -5.51518621 -1.43581500  
14 -25.6595442 -2.9099752  4.0902387  1.05606534  1.72104865  
15   2.2883385 -3.8094683 -5.6538244 -1.47303223  3.66106601  
16  -9.1981937 15.5215268  5.1100194  2.13893163  0.27497793  
17 -16.4885053  7.9576494 -5.5274058 -3.09232370 -1.75437414  
18  -0.3879502 -1.7755769  8.4614213 -0.09982322 -5.80333578  
19  12.0986229 12.2219982 -7.0332189  4.26593453 -1.33011918  
20 -14.9615830 -6.4307441  4.9507242  3.69029321 -0.39513201
```



# 元の変数と主成分得点の相関係数の確認

```
> if (!require('psych')) install.packages('psych'); library('psych')
> dat.ex <- cbind(dat, result$scores)
> corr.test(dat.ex)
```

Call:corr.test(x = dat.ex)

Correlation matrix

|        | eigo | sugaku | kokugo | rika  | syakai | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|--------|------|--------|--------|-------|--------|--------|--------|--------|--------|--------|
| eigo   | 1.00 | 0.46   | 0.40   | 0.43  | 0.55   | 0.59   | 0.32   | 0.73   | 0.04   | 0.11   |
| sugaku | 0.46 | 1.00   | 0.44   | 0.80  | 0.36   | 0.96   | -0.23  | -0.05  | 0.14   | 0.00   |
| kokugo | 0.40 | 0.44   | 1.00   | 0.37  | 0.82   | 0.62   | 0.70   | -0.29  | -0.06  | 0.20   |
| rika   | 0.43 | 0.80   | 0.37   | 1.00  | 0.28   | 0.85   | -0.23  | 0.03   | -0.47  | -0.04  |
| syakai | 0.55 | 0.36   | 0.82   | 0.28  | 1.00   | 0.55   | 0.78   | 0.00   | 0.03   | -0.30  |
| Comp.1 | 0.59 | 0.96   | 0.62   | 0.85  | 0.55   | 1.00   | 0.00   | 0.00   | 0.00   | 0.00   |
| Comp.2 | 0.32 | -0.23  | 0.70   | -0.23 | 0.78   | 0.00   | 1.00   | 0.00   | 0.00   | 0.00   |
| Comp.3 | 0.73 | -0.05  | -0.29  | 0.03  | 0.00   | 0.00   | 0.00   | 1.00   | 0.00   | 0.00   |
| Comp.4 | 0.04 | 0.14   | -0.06  | -0.47 | 0.03   | 0.00   | 0.00   | 0.00   | 1.00   | 0.00   |
| Comp.5 | 0.11 | 0.00   | 0.20   | -0.04 | -0.30  | 0.00   | 0.00   | 0.00   | 0.00   | 1.00   |

Sample Size

[1] 20

Probability values (Entries above the diagonal are adjusted for multiple tests.)

|        | eigo | sugaku | kokugo | rika | syakai | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|--------|------|--------|--------|------|--------|--------|--------|--------|--------|--------|
| eigo   | 0.00 | 1.00   | 1.00   | 1.00 | 0.44   | 0.22   | 1.00   | 0.01   | 1      | 1      |
| sugaku | 0.04 | 0.00   | 1.00   | 0.00 | 1.00   | 0.00   | 1.00   | 1.00   | 1      | 1      |
| kokugo | 0.08 | 0.05   | 0.00   | 1.00 | 0.00   | 0.14   | 0.02   | 1.00   | 1      | 1      |
| rika   | 0.06 | 0.00   | 0.11   | 0.00 | 1.00   | 0.00   | 1.00   | 1.00   | 1      | 1      |
| syakai | 0.01 | 0.12   | 0.00   | 0.23 | 0.00   | 0.44   | 0.00   | 1.00   | 1      | 1      |
| Comp.1 | 0.01 | 0.00   | 0.00   | 0.00 | 0.01   | 0.00   | 1.00   | 1.00   | 1      | 1      |
| Comp.2 | 0.16 | 0.34   | 0.00   | 0.33 | 0.00   | 1.00   | 0.00   | 1.00   | 1      | 1      |
| Comp.3 | 0.00 | 0.83   | 0.22   | 0.89 | 0.98   | 1.00   | 1.00   | 0.00   | 1      | 1      |
| Comp.4 | 0.87 | 0.57   | 0.80   | 0.04 | 0.89   | 1.00   | 1.00   | 1.00   | 0      | 1      |
| Comp.5 | 0.64 | 0.99   | 0.40   | 0.87 | 0.20   | 1.00   | 1.00   | 1.00   | 1      | 0      |

データセットの相関行列を求める(psychパッケージの関数)

cbind(<データセット1>, <データセット2>)でデータセットを横方向に結合する(※行の数は同じである必要がある)

(参考) 小数第3位まで見たい場合は print(<結果を代入したオブジェクト>, digits=3)

```
> print(corr.test(dat.ex), digits=3)
Call:corr.test(x = dat.ex)
Correlation matrix
      eigo sugaku kokugo  rika syakai Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
eigo  1.000  0.465  0.404  0.428  0.548  0.593  0.324  0.728  0.039  0.112
sugaku 0.465  1.000  0.442  0.804  0.357  0.963 -0.225 -0.053  0.136 -0.004
kokugo 0.404  0.442  1.000  0.373  0.823  0.617  0.703 -0.288 -0.061  0.198
rika   0.428  0.804  0.373  1.000  0.280  0.849 -0.232  0.032 -0.473 -0.038
syakai 0.548  0.357  0.823  0.280  1.000  0.547  0.780  0.005  0.033 -0.302
Comp.1 0.593  0.963  0.617  0.849  0.547  1.000  0.000  0.000  0.000  0.000
Comp.2 0.324 -0.225  0.703 -0.232  0.780  0.000  1.000  0.000  0.000  0.000
Comp.3 0.728 -0.053 -0.288  0.032  0.005  0.000  0.000  1.000  0.000  0.000
Comp.4 0.039  0.136 -0.061 -0.473  0.033  0.000  0.000  0.000  1.000  0.000
Comp.5 0.112 -0.004  0.198 -0.038 -0.302  0.000  0.000  0.000  0.000  1.000
Sample Size
[1] 20
Probability values (Entries above the diagonal are adjusted for multiple tests.)
      eigo sugaku kokugo  rika syakai Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
eigo  0.000  1.000  1.000  1.000  0.444  0.218  1.000  0.011  1  1
sugaku 0.039  0.000  1.000  0.001  1.000  0.000  1.000  1.000  1  1
kokugo 0.077  0.051  0.000  1.000  0.000  0.142  0.022  1.000  1  1
rika   0.060  0.000  0.105  0.000  1.000  0.000  1.000  1.000  1  1
syakai 0.012  0.122  0.000  0.233  0.000  0.444  0.002  1.000  1  1
Comp.1 0.006  0.000  0.004  0.000  0.013  0.000  1.000  1.000  1  1
Comp.2 0.163  0.340  0.001  0.325  0.000  1.000  0.000  1.000  1  1
Comp.3 0.000  0.825  0.219  0.894  0.985  1.000  1.000  0.000  1  1
Comp.4 0.869  0.567  0.797  0.035  0.889  1.000  1.000  1.000  0  1
Comp.5 0.639  0.986  0.404  0.874  0.196  1.000  1.000  1.000  1  0
```

To see confidence intervals of the correlations, print with the short=FALSE option

下半分は、相関係数の検定結果(調整済みp値)

To see confidence intervals of the correlations, print with the short=FALSE option



# 標準得点を用いた主成分分析

scale(<データセット>)で  
標準得点(z得点)にした後は同様

```
> dat.std <- data.frame(scale(dat))
> result.std <- princomp(dat.std)
> dat.std.ex <- cbind(dat.std, result.std$scores)
> print(corr.test(dat.std.ex), digits=3)
Call:corr.test(x = dat.std.ex)
Correlation matrix
```

|        | eigo  | sugaku | kokugo | rika   | syakai | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|--------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| eigo   | 1.000 | 0.465  | 0.404  | 0.428  | 0.548  | 0.730  | 0.052  | 0.675  | 0.006  | 0.089  |
| sugaku | 0.465 | 1.000  | 0.442  | 0.804  | 0.357  | 0.796  | -0.497 | -0.129 | -0.318 | -0.027 |
| kokugo | 0.404 | 0.442  | 1.000  | 0.373  | 0.823  | 0.796  | 0.444  | -0.330 | 0.022  | 0.244  |
| rika   | 0.428 | 0.804  | 0.373  | 1.000  | 0.280  | 0.745  | -0.581 | -0.124 | 0.300  | -0.038 |
| syakai | 0.548 | 0.357  | 0.823  | 0.280  | 1.000  | 0.785  | 0.558  | -0.045 | 0.011  | -0.267 |
| Comp.1 | 0.730 | 0.796  | 0.796  | 0.745  | 0.785  | 1.000  | 0.000  | 0.000  | 0.000  | 0.000  |
| Comp.2 | 0.052 | -0.497 | 0.444  | -0.581 | 0.558  | 0.000  | 1.000  | 0.000  | 0.000  | 0.000  |
| Comp.3 | 0.675 | -0.129 | -0.330 | -0.124 | -0.045 | 0.000  | 0.000  | 1.000  | 0.000  | 0.000  |
| Comp.4 | 0.006 | -0.318 | 0.022  | 0.300  | 0.011  | 0.000  | 0.000  | 0.000  | 1.000  | 0.000  |
| Comp.5 | 0.089 | -0.027 | 0.244  | -0.038 | -0.267 | 0.000  | 0.000  | 0.000  | 0.000  | 1.000  |

```
Sample Size
[1] 20
Probability values (Entries above the diagonal are adjusted for multiple tests.)
```

|        | eigo  | sugaku | kokugo | rika  | syakai | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|--------|-------|--------|--------|-------|--------|--------|--------|--------|--------|--------|
| eigo   | 0.000 | 1.000  | 1.000  | 1.000 | 0.432  | 0.010  | 1.000  | 0.041  | 1      | 1      |
| sugaku | 0.039 | 0.000  | 1.000  | 0.001 | 1.000  | 0.001  | 0.875  | 1.000  | 1      | 1      |
| kokugo | 0.077 | 0.051  | 0.000  | 1.000 | 0.000  | 0.001  | 1.000  | 1.000  | 1      | 1      |
| rika   | 0.060 | 0.000  | 0.105  | 0.000 | 1.000  | 0.007  | 0.265  | 1.000  | 1      | 1      |
| syakai | 0.012 | 0.122  | 0.000  | 0.233 | 0.000  | 0.002  | 0.383  | 1.000  | 1      | 1      |
| Comp.1 | 0.000 | 0.000  | 0.000  | 0.000 | 0.000  | 0.000  | 1.000  | 1.000  | 1      | 1      |
| Comp.2 | 0.828 | 0.026  | 0.050  | 0.007 | 0.011  | 1.000  | 0.000  | 1.000  | 1      | 1      |
| Comp.3 | 0.001 | 0.588  | 0.156  | 0.601 | 0.851  | 1.000  | 1.000  | 0.000  | 1      | 1      |
| Comp.4 | 0.981 | 0.171  | 0.928  | 0.199 | 0.963  | 1.000  | 1.000  | 1.000  | 0      | 1      |
| Comp.5 | 0.709 | 0.910  | 0.300  | 0.873 | 0.256  | 1.000  | 1.000  | 1.000  | 1      | 0      |

To see confidence intervals of the correlations, print with the short=FALSE option



# 課題

- 前回のアンケート調査データのうち、ラーメンに関する5項目に対して主成分分析を適用してください。算出された第1主成分得点についてその意味を解釈し、第1主成分得点の平均値が東日本出身者と西日本出身者で異なるか否か、下記の文章に続ける形で、報告してください。なお、文中の〇〇は、適切な言葉や数値に置き換えてください。また、結論付けるのに必要な平均値の差の検定結果については、t値（またはF値）、p値等の各種統計量を、論文と同じように報告してください。さらに、両群での第1主成分得点の分布を比較するための箱ひげ図もあわせて報告してください。

本研究では、ラーメンが好きな程度は、東日本出身者か西日本出身者であるかで異なるのか否か検討する。

Web調査の結果、大学生〇〇名が回答した。5種類のラーメン（塩、担々麺、醤油、豚骨、味噌）が好きな程度を問う5項目について主成分分析を適用したところ、第1主成分の寄与率が約〇〇%であった。第1主成分の重み付け係数は、塩：〇〇、担々麺：〇〇、醤油：〇〇、豚骨：〇〇、味噌：〇〇であったため、第1主成分は〇〇を表していると解釈した。第1主成分得点の平均値が、東日本出身者と西日本出身者で異なるか否か検討するため、第1主成分得点について、出身地（東日本or西日本）を対応の〇〇要因とする〇〇を行ったところ、・・・

t検定か分散分析  
（どちらでも良い）

「ある」か「ない」のどちらか

# 課題の補足

- princomp関数には、主成分分析を適用したい変数のみが含まれているデータセットを読み込ませる必要があります(今回の場合は、ラーメンに関する5項目のみ)。データセットから特定の列だけ抽出するには、以下のようになります。

1～4列目だけ抽出する場合

```
> dat
  eigo sugaku kokugo rika syakai
1   77    73    87   58    71
2   78    80    91   64    72
3   65    74    78   46    64
4   71    98    76   65    58
5   63    82    81   49    67
```

```
> dat[1:4]
  eigo sugaku kokugo rika
1   77    73    87   58
2   78    80    91   64
3   65    74    78   46
4   71    98    76   65
5   63    82    81   49
```

1列目と5列目だけ抽出する場合

```
> dat[c(1, 5)]
  eigo syakai
1   77     71
2   78     72
3   65     64
4   71     58
5   63     67
```

3列目だけ取り除いて抽出する場合

```
> dat[-3]
  eigo sugaku rika syakai
1   77    73   58    71
2   78    80   64    72
3   65    74   46    64
4   71    98   65    58
5   63    82   49    67
```

- 標準得点を用いた主成分分析にするかどうかは任意です。
- 平均値の差の検定は、t検定と分散分析のいずれで検討してもOKです。(ぜひ、これまでの内容を思い出しながらチャレンジしよう!)

## チャレンジ問題(任意)

---

- 先の課題について、第2主成分得点についても同様に解釈をして、東日本出身者と西日本出身者で平均値が異なるか否か検討してください。