



# 第9回データ分析勉強会

---

Rではじめる多変量解析

---

[鈴木さんぽう, CC BY-NC-SA 4.0]

2019年02月19日

# 本日の内容

---

## 1. 多変量解析入門

- ・ 多変量解析とは？
- ・ 予測や制御を行う
  - 重回帰分析
- ・ 集約や探索を行う
  - 主成分分析、因子分析
  - クラスター分析(別資料)

## 2. 演習

- ・ 演習I & 演習II, 演習III & 発表

# 多变量解析入門

# 多変量解析とは？

---

多変量解析(multivariate analysis)とは文字通り多数の変量(変数)を解析するための方法です。例えば、散布図行列では2変量間の関係は見えますが変量全体の関係はとらえにくいものです。変量全体の関係をとらえてモデルリングしたい場合やデータマイニングのように多数の情報から傾向を調べたい探索を行う場合などに多変量解析を用います。

主な多変量解析の代表的な手法としては以下のようなものがあります。

- ・ 重回帰分析(予測や制御に使える)
  - 過去のデータが持つ構造が近い将来も使える(≒プロセスが安定している)
- ・ 主成分分析(情報の縮約=次元縮約)
- ・ 因子分析(主成分分析とは似て非なるもの)
- ・ クラスター分析(データの分類に便利)
- ・ コンジョイント分析(上記の分析とは多少趣が異なる)

# 多変量解析の手法

---

多変量解析には前述のように様々な手法があり、目的に応じて使い分ける必要があります。

## 1. 予測や制御

線形単回帰のように**目的変数と説明変数の関係**を統計モデルとし、それに基づく予測を行いたい場合に用います

- ・ 連続データ:重回帰分析、分散分析
- ・ カテゴリカルデータ:判別分析、ロジスティック回帰分析

## 2. 集約や探索

データが持つ情報を集約したり**変量の関係**を知りたい場合に用います。近年、データマイニングなどにも使われています。

- ・ 主成分分析、因子分析
- ・ クラスター分析(データを分類したい場合)

予測や制御を行う

# 変量(変数)間の関係

---

予測や制御を行う前には変量間の関係を把握しておくことが大切です。変量間の関係は以下のように分類できます。[出典: 統計的因果推論 岩崎学 朝倉書店]

- ・ **相関**関係 (Correlation)
  - 関係性があることしか言えない(双方向の関係)
  - ゆえに予測や制御には使えない
- ・ **回帰**関係 (Regression)
  - 因果関係と同様な一方的な関係(必ずしも因果関係ではない)
  - 予測に用いることができる
    - 回帰関係と因果関係の違いがよく分からない...
- ・ **因果**関係 (Causality)
  - 原因から結果へ一方的な関係(回帰関係を因果関係としているものもある)
  - 予測や制御に用いることができる

# 単回帰分析(復習)

---

(線形) 単回帰分析はひとつの説明変数と目的変数の関係を最小二乗法を用いて下記のような回帰式を求める分析です。最小二乗法を用いるため、ほとんどのデータで回帰分析ができてしまいしますので、単回帰分析では精度(決定係数)が重要になります。

$$y = a + bx$$

## 外挿

説明変数がない範囲(実際にデータが観察されていない範囲)は外挿と呼ばれ、回帰式が成り立つ関係があるのか分からないので注意が必要です。

## 切片はどう考える?

例えば説明変数が規模の場合、切片(説明変数 = 0の時の目的変数の値)をどう考えるのが良いでしょうか?

# (線形) 単回帰分析の結果を説明してください

---

```
##  
## Call:  
## lm(formula = mpg ~ hp, data = mtcars)  
##  
## Residuals:  
##      Min       1Q   Median      3Q      Max  
## -5.7121 -2.1122 -0.8854  1.5819  8.2360  
##  
## Coefficients:  
##                 Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 30.09886   1.63392 18.421 < 2e-16 ***  
## hp          -0.06823   0.01012 -6.742 1.79e-07 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 3.863 on 30 degrees of freedom  
## Multiple R-squared:  0.6024, Adjusted R-squared:  0.5892  
## F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```

# 重回帰分析

---

(線形)重回帰分析では一般的に下記の回帰式( $n = 1, 2, \dots$ )を用います。

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$$

この回帰式を用いると以下のようなことができるようになります。

- ・ 説明変数(与えられた値)から目的変数を予測する
- ・ 目的変数の値を大きく(または小さく)するには説明変数(これから与える値)をどうかするか(制御)【因果関係がある場合】

線形モデル以外にも、交互作用モデル、2次式モデル、共分散分析モデル、一般線形化モデルなど様々なモデルを作ることができます。

$$y = b_0 + b_1x_1 + \dots + b_nx_n + b_{12}x_1x_2 + \dots + b_{(n-1)n}x_{n-1}x_n$$

# 重回帰式をどう評価するか

---

重回帰式の解釈や使い方は単回帰に比べると難しいと言われています。目的により解釈が変わってきますので注意してください。

## 予測で使う場合(回帰関係)

- ・個々の偏回帰係数の解釈はあまり重要でない
- ・偏回帰係数はより良い予測モデルの選択のために用いられる

## 制御で使う場合(因果関係)

- ・個々の偏回帰係数の解釈(値)が重要である
- ・偏回帰係数は制御に用いるために選択される
  - 0に近いような係数は目的変数への影響が小さいので除外するなど

# 重回帰式の評価

(線形)重回帰モデルを評価するには線形単回帰モデルと同様に決定係数(自由度調整済み)とp値に着目します。

## 自由度調整済み決定係数

```
lm(mpg ~ cyl + disp + hp + drat + wt, data = mtcars) %>% broom::glance()
```

|   | r.squared<br><dbl> | adj.r.squared<br><dbl> | sigma<br><dbl> | statistic<br><dbl> | p.value<br><dbl> | df<br><int> |
|---|--------------------|------------------------|----------------|--------------------|------------------|-------------|
| 1 | 0.8513152          | 0.8227219              | 2.53761        | 29.77331           | 5.618479e-10     | 6           |

1 row | 1-7 of 12 columns

**自由度調整済み決定係数**(adj.r.squared)は説明変数が目的変数の変動の何パーセントを説明できているかを意味する値です。この例では約82%を説明できていることが分かります。予測に使う場合にはどの程度説明できているかは重要です。**p値**(p.value)は重回帰式自体の評価結果です。

# 重回帰式の評価(つづき)

重回帰式 자체の評価を行ったら次に偏回帰係数のp値(p.value)を確認します。

## 偏回帰係数とp値

```
lm(mpg ~ cyl + disp + hp + drat + wt, data = mtcars) %>% broom::tidy()
```

| term        | estimate | std.error | statistic | p.value |
|-------------|----------|-----------|-----------|---------|
| <chr>       | <dbl>    | <dbl>     | <dbl>     | <dbl>   |
| (Intercept) | 36.008   | 7.571     | 4.756     | 0.000   |
| cyl         | -1.107   | 0.716     | -1.547    | 0.134   |
| disp        | 0.012    | 0.012     | 1.039     | 0.308   |
| hp          | -0.024   | 0.013     | -1.809    | 0.082   |
| drat        | 0.952    | 1.391     | 0.685     | 0.500   |
| wt          | -3.673   | 1.059     | -3.469    | 0.002   |

6 rows

# 偏回帰係数の解釈

---

偏回帰係数の解釈は意外と難しくどの偏回帰係数を選ぶかは前述の目的(下記に再掲)により異なってきます。

## 予測で使う場合(回帰関係)

- ・個々の偏回帰係数の解釈はあまり重要でない
- ・偏回帰係数はより良い予測モデルの選択のために用いられる

## 制御で使う場合(因果関係)

- ・個々の偏回帰係数の解釈(値)が重要である
- ・偏回帰係数は制御に用いるために選択される
  - 0に近いような係数は目的変数への影響が小さいので除外するなど

# 変数選択

---

説明変数の選択にはAIC(赤池情報量基準)を元にしたステップワイズ変数選択を用いて最も適していると考えらえる変量を選択する方法があります。

```
step(lm(mpg ~ cyl + disp + hp + drat + wt, data = mtcars)) %>%  
  broom:::tidy()
```

このモデル式では自由度調整済み決定係数が0.8263であり、先程のモデルよりは多くの変動を説明できているので予測に使うにはより適していると考えられます。

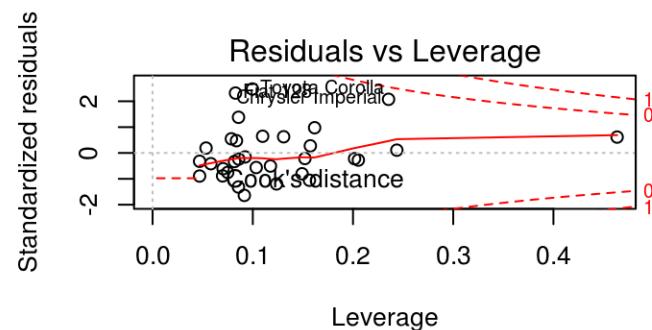
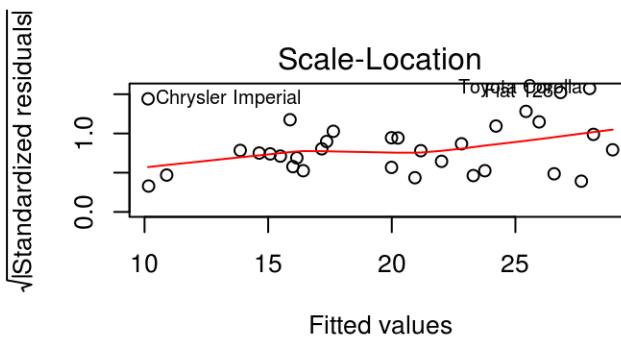
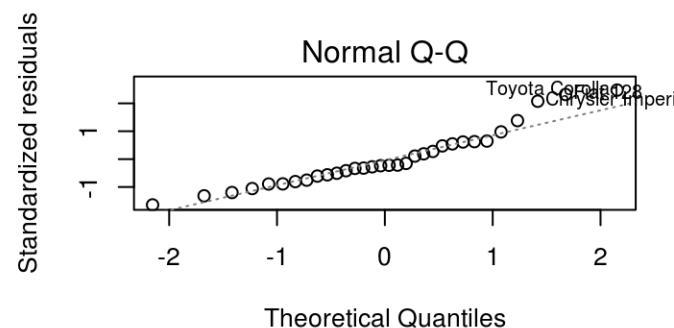
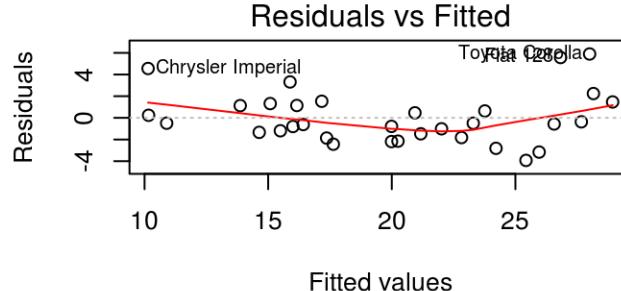
# 変数選択は意外と難しい

---

- ・ 説明変数の選択は必要なのか?
  - 説明変数が決まっている
  - 分析の目的が予測(使えるものは使ってしまう)
- ・ 説明変数の選択が必要
  - 説明変数をスクリーニングしたい(分析の初期段階)
  - 予備調査の結果を用いて本調査を行う
  - 分析の目的が制御(少ない方が制御しやすい)
- ・ ただし、残差分析は常にすること

# 残差分析

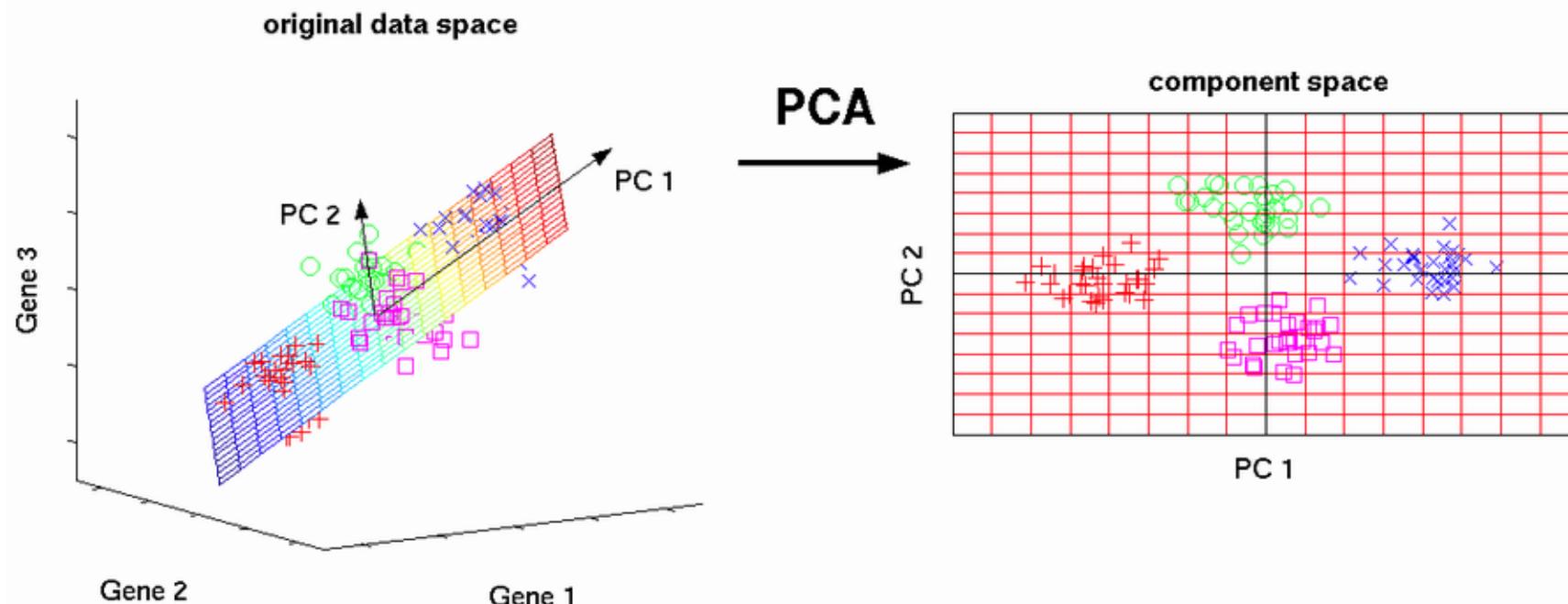
```
lm(mpg ~ cyl + hp + wt, data = mtcars) %>% plot()
```



集約や探索を行う

# 主成分分析とは

主成分分析は次元縮約を使って任意数の変量(変数)から「共通因子」を探索し「可視化」できる分析方法です。乱暴にいえばn次元空間の変量(変数)の関係を2次元空間の主成分の関係へ写像する(次元縮約する)方法です。



『主成分分析について具体的な過程抜きでざっくり』より引用

# 次元縮約とは

---

主成分分析や因子分析の目的である次元縮約とは多数ある変量をより少数の次元に要約することで変量の関係を把握しやすくするための手法です。次元縮約することで下表のようなことができるようになります。

|   | 主成分分析   | 因子分析    |
|---|---------|---------|
| 1 | 共通因子の探索 | 共通因子の探索 |
| 2 |         | 独自要因の推定 |
| 3 | 重みつき合計  |         |
| 4 | 可視化     |         |

# 主成分分析の分析手順

---

主成分分析は以下の手順で行います。

- ・ n次元空間の中で最も分散が大きくなる第一主成分軸を探す
- ・ 第一主成分軸に直行し最も分散が大きくなる第二主成分軸を探す
- ・ 第一、第二主成分軸に直行し最も分散が大きくなる第三主成分軸を探す
  - 以降、同様に第n主成分軸までを探す
- ・ 各主成分軸上での個々の変量(変数)の座標値(主成分得点)を求める
- ・ 以下の指標を使用して主成分数を決める
  - 累積分散説明率(寄与率)(7~8割以上ならまあいいかな)
  - 値が「1」を超える(共分散の)固有値の数
  - 降順に並べた(共分散の)固有値の折れ線の「屈曲点 - 1」
    - 上記の値が一致しない場合は分析者の判断による

# 分散総和

---

主成分分析は各変量の分散を分散の総和を変えずに再分配する分析です。再分配する先を主成分(軸)と呼びます。以下のように総和に変化はありません。

## 変量の分散総和と各変量の分散

## 主成分得点の分散総和と主成分得点の分散

# 分散説明率(寄与率)

---

主成分数を決めるためにまず分散説明率を確認します。分散説明率とは分散の総和に対する各主成分得点の分散の比のことです。先程の例では下表の通りになります。

## 主成分得点の分散総和と主成分得点の分散

| 主成分 | 分散                 | 寄与率 |
|-----|--------------------|-----|
| 1   | 1.2                | 40% |
| 2   | 0.8                | 27% |
| 3   | 0.4                | 13% |
| 4   | 0.2                | 7%  |
| 5   | 0.1                | 3%  |
| 6   | 0.05               | 2%  |
| 7   | 0.02               | 1%  |
| 8   | 0.01               | 1%  |
| 9   | 0.005              | 1%  |
| 10  | 0.002              | 1%  |
| 11  | 0.001              | 1%  |
| 12  | 0.0005             | 1%  |
| 13  | 0.0002             | 1%  |
| 14  | 0.0001             | 1%  |
| 15  | 0.00005            | 1%  |
| 16  | 0.00002            | 1%  |
| 17  | 0.00001            | 1%  |
| 18  | 0.000005           | 1%  |
| 19  | 0.000002           | 1%  |
| 20  | 0.000001           | 1%  |
| 21  | 0.0000005          | 1%  |
| 22  | 0.0000002          | 1%  |
| 23  | 0.0000001          | 1%  |
| 24  | 0.00000005         | 1%  |
| 25  | 0.00000002         | 1%  |
| 26  | 0.00000001         | 1%  |
| 27  | 0.000000005        | 1%  |
| 28  | 0.000000002        | 1%  |
| 29  | 0.000000001        | 1%  |
| 30  | 0.0000000005       | 1%  |
| 31  | 0.0000000002       | 1%  |
| 32  | 0.0000000001       | 1%  |
| 33  | 0.00000000005      | 1%  |
| 34  | 0.00000000002      | 1%  |
| 35  | 0.00000000001      | 1%  |
| 36  | 0.000000000005     | 1%  |
| 37  | 0.000000000002     | 1%  |
| 38  | 0.000000000001     | 1%  |
| 39  | 0.0000000000005    | 1%  |
| 40  | 0.0000000000002    | 1%  |
| 41  | 0.0000000000001    | 1%  |
| 42  | 0.00000000000005   | 1%  |
| 43  | 0.00000000000002   | 1%  |
| 44  | 0.00000000000001   | 1%  |
| 45  | 0.000000000000005  | 1%  |
| 46  | 0.000000000000002  | 1%  |
| 47  | 0.000000000000001  | 1%  |
| 48  | 0.0000000000000005 | 1%  |
| 49  | 0.0000000000000002 | 1%  |
| 50  | 0.0000000000000001 | 1%  |

## 主成分得点の分散総和と分散説明率(寄与率)

| 主成分 | 分散                 | 寄与率 |
|-----|--------------------|-----|
| 1   | 1.2                | 40% |
| 2   | 0.8                | 27% |
| 3   | 0.4                | 13% |
| 4   | 0.2                | 7%  |
| 5   | 0.1                | 3%  |
| 6   | 0.05               | 2%  |
| 7   | 0.02               | 1%  |
| 8   | 0.01               | 1%  |
| 9   | 0.005              | 1%  |
| 10  | 0.002              | 1%  |
| 11  | 0.001              | 1%  |
| 12  | 0.0005             | 1%  |
| 13  | 0.0002             | 1%  |
| 14  | 0.0001             | 1%  |
| 15  | 0.00005            | 1%  |
| 16  | 0.00002            | 1%  |
| 17  | 0.00001            | 1%  |
| 18  | 0.000005           | 1%  |
| 19  | 0.000002           | 1%  |
| 20  | 0.000001           | 1%  |
| 21  | 0.0000005          | 1%  |
| 22  | 0.0000002          | 1%  |
| 23  | 0.0000001          | 1%  |
| 24  | 0.00000005         | 1%  |
| 25  | 0.00000002         | 1%  |
| 26  | 0.00000001         | 1%  |
| 27  | 0.000000005        | 1%  |
| 28  | 0.000000002        | 1%  |
| 29  | 0.000000001        | 1%  |
| 30  | 0.0000000005       | 1%  |
| 31  | 0.0000000002       | 1%  |
| 32  | 0.0000000001       | 1%  |
| 33  | 0.00000000005      | 1%  |
| 34  | 0.00000000002      | 1%  |
| 35  | 0.00000000001      | 1%  |
| 36  | 0.000000000005     | 1%  |
| 37  | 0.000000000002     | 1%  |
| 38  | 0.000000000001     | 1%  |
| 39  | 0.0000000000005    | 1%  |
| 40  | 0.0000000000002    | 1%  |
| 41  | 0.0000000000001    | 1%  |
| 42  | 0.00000000000005   | 1%  |
| 43  | 0.00000000000002   | 1%  |
| 44  | 0.00000000000001   | 1%  |
| 45  | 0.000000000000005  | 1%  |
| 46  | 0.000000000000002  | 1%  |
| 47  | 0.000000000000001  | 1%  |
| 48  | 0.0000000000000005 | 1%  |
| 49  | 0.0000000000000002 | 1%  |
| 50  | 0.0000000000000001 | 1%  |

この場合、第一、第二主成分得点で全体の98%が説明できることが分かります。

# (共分散行列の)固有値

---

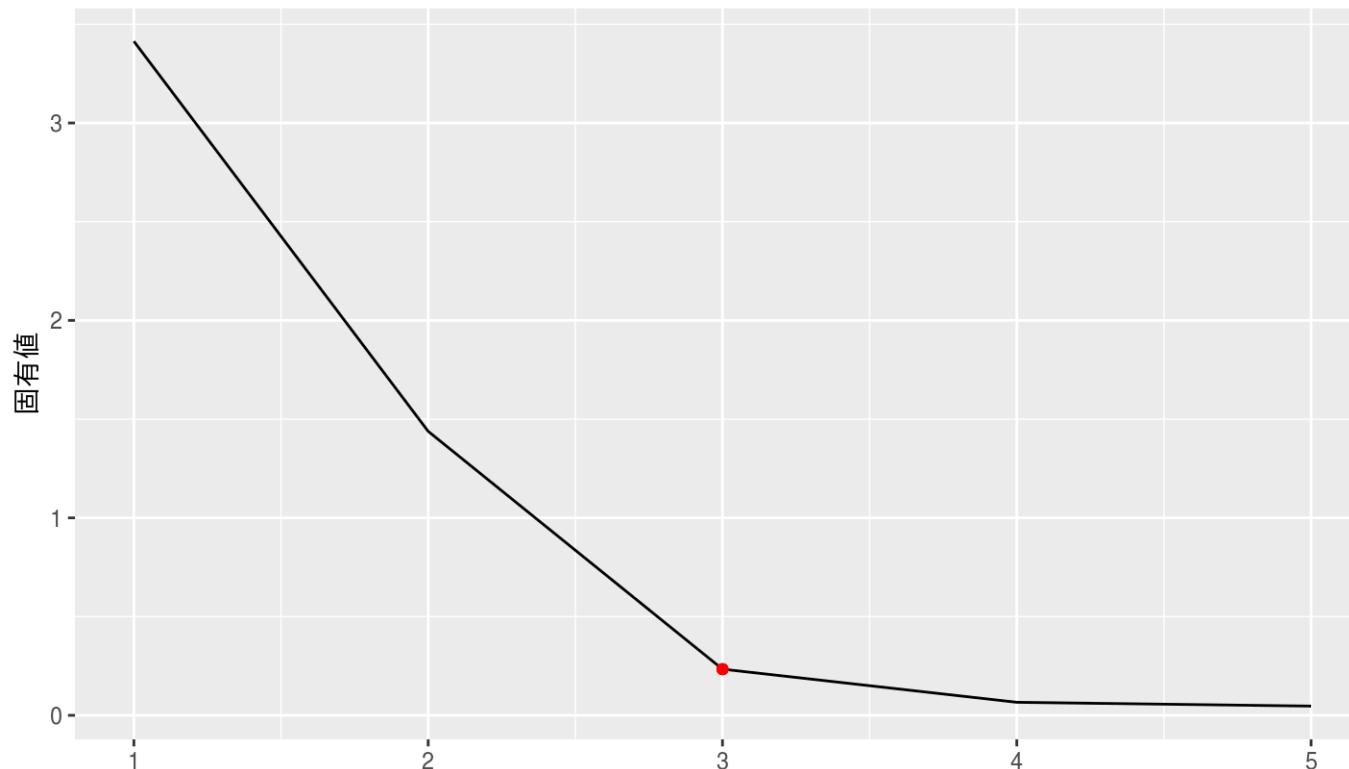
分散共分散行列(相関係数行列)の固有値は主成分得点の標準偏差の二乗(主成分得点の分散)に等しく、固有値が1を超えているものを主成分として採用します。

## 固有値

# 固有値の屈曲点

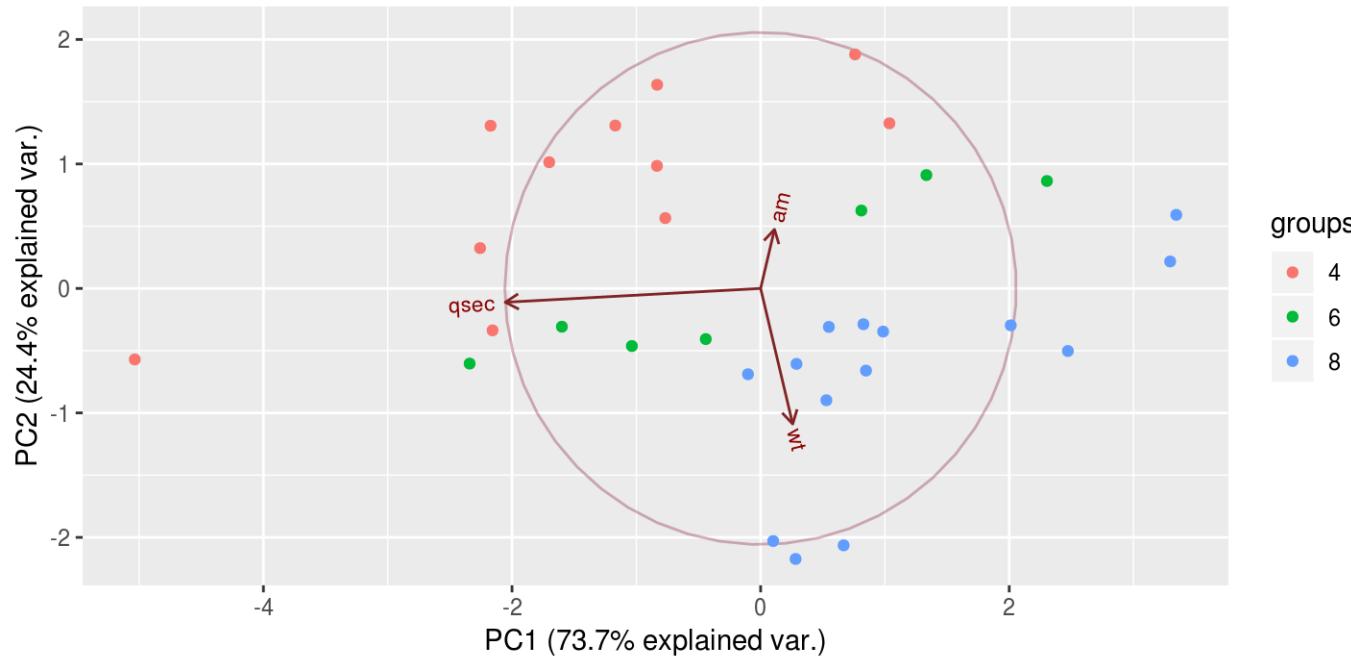
---

固有値を降順に並べて折れ線グラフを描き、屈曲点(変曲点)から1を引いたものが主成分数です。下記の場合は、3番目が屈曲点とみなせますので、主成分数は2となります。



# 主成分分析の可視化

主成分得点を用いて可視化することでデータの関係を把握しやすくなります。二次元の可視化にはggbiplot関数が便利です。



# 単純構造化(より理解しやすくする)

---

以下の条件をすべて満たせば主成分負荷量を任意に変換(回転と呼ばれることが多い)することができます。

- ・ 全ての主成分得点の**分散総和が「1」**
- ・ 第k成分得点と第l成分得点( $\neq k$ )は**無相関**
- ・ **誤差の値を変えない**

回転させるのは主成分負荷量に「0」が多くなるようするためで、主成分負荷量が「0」(または、0に近い値)になれば、変量に対応する主成分は影響がない(少ない)ということであり、より、合理的な説明がしやすくなるからです。

代表的な回転方法としてはバリマックス回転法があります。

# Rで解くには

---

Rを利用して主成分分析を行うにはprcomp関数を用います。数値データ以外は受け付けませんので注意してください。なお、prcomp関数の出力は**主成分得点**である点に注意してください（主成分負荷量ではありません）。

```
mtcars %>% dplyr::select(wt, qsec, am) %>%  
  prcomp(scale. = FALSE)
```

```
## Standard deviations (1, .., p=3):  
## [1] 1.8007420 1.0360726 0.2888232  
##  
## Rotation (n x k) = (3 x 3):  
##          PC1        PC2        PC3  
## wt    0.12448262 -0.91286076 -0.38883069  
## qsec -0.99076363 -0.09312021 -0.09857007  
## am    0.05377276  0.39750956 -0.91602109
```

# 分散説明率(寄与率)を求める

---

分散説明率(寄与率)はprcomp関数の結果をサマライズするだけで求めることができます。

```
mtcars %>% dplyr::select(wt, qsec, am) %>%  
  prcomp(scale. = FALSE) %>% summary()
```

```
## Importance of components:  
##                 PC1    PC2    PC3  
## Standard deviation 1.801 1.036 0.28882  
## Proportion of Variance 0.737 0.244 0.01896  
## Cumulative Proportion 0.737 0.981 1.00000
```

Proportion of Varianceが分散説明率で、Cumulative Proportionが累積分散説明率になります。

# prcomp関数の返り値詳細

---

prcomp関数の返り値は下記のようなリスト型になっています。属性(attr)を見ると分かりますが与えたデータフレームの順に結果が返されます。

```
mtcars %>% dplyr::select(wt, qsec, am) %>%  
  prcomp(scale. = FALSE, retx = FALSE) %>% str()  
  
## List of 4  
## $ sdev    : num [1:3] 1.801 1.036 0.289  
## $ rotation: num [1:3, 1:3] 0.1245 -0.9908 0.0538 -0.9129 -0.0931 ...  
## ..- attr(*, "dimnames")=List of 2  
##   ..$ : chr [1:3] "wt" "qsec" "am"  
##   ..$ : chr [1:3] "PC1" "PC2" "PC3"  
## $ center   : Named num [1:3] 3.217 17.849 0.406  
## ..- attr(*, "names")= chr [1:3] "wt" "qsec" "am"  
## $ scale    : logi FALSE  
## - attr(*, "class")= chr "prcomp"
```

# 固有値を求める

---

固有値は前述の通り主成分得点の標準偏差(\$sdev)の二乗(=分散)に等しいので二乗することで求めることができます。

```
result <- mtcars %>% dplyr::select(wt, qsec, am) %>%  
  # prcomp(scale. = TRUE)  
  prcomp(scale. = FALSE)
```

```
t(result$sdev ^ 2) %>% as.data.frame() %>%  
  dplyr::rename(PC1 = V1, PC2 = V2, PC3 = V3)
```

## 参考資料

# 主成分負荷量を求める

---

主成分負荷量は因子分析との関連から出てくることが多い主成分得点と変量(変数)との相関係数です。主成分得点(\$rotation)と固有値の平方根(\$sdev)の積で求めることができます。転置処理(t関数)が必要な点に注意してください。

```
result <- mtcars %>% dplyr::select(wt, qsec, am) %>%  
  prcomp(scale. = FALSE)  
t(result$rotation) * result$sdev) %>% as.data.frame()
```

# prcomp関数のscale.オプション

---

今回の説明では手動で計算した各変量の分散と主成分得点の分散が一致するようにprcomp関数のscale.オプションは常にFALSE(デフォルト値)に設定していますが、一般的にはscale. = TRUEを設定し各変量が単位分散を持つようにスケーリング<sup>1</sup>されることが推奨されています。

scale.

a logical value indicating whether the variables should be scaled to have unit variance before the analysis takes place. **The default is FALSE** for consistency with S, **but in general scaling is advisable**. Alternatively, a vector of length equal the number of columns of x can be supplied. The value is passed to base::scale function.

<sup>1</sup> 各変量が異なる単位を持つ場合、標準化により無単位化(無次元化)すること

# 主成分分析のまとめ

---

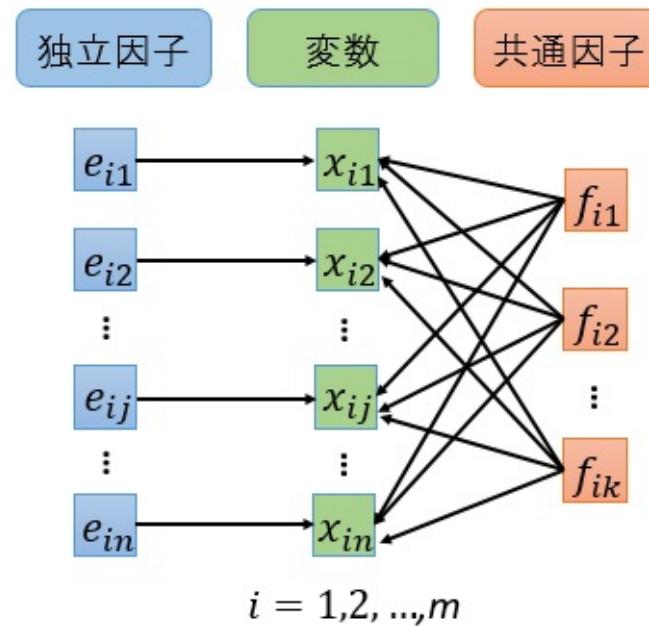
- ・ 多数の変量を少数の主成分に縮約する(まとめる)のが主成分分析
  - 今回の例では3変量が2変量に縮約された
  - 各変量の分散の総和と主成分の分散の総和は同じであることが条件
  - 分散の総和に占める割合(分散説明率)により主成分数が決まる
- ・ 主成分の解釈は分析者が行う
  - 主成分が何を意味しているかは分析結果には示されない
  - 分析結果を可視化することで解釈しやすくなる

このような特徴から変量の多いアンケート分析などに使われることが多い主成分分析ですが、プロダクトメトリクスやチケットデータの分析にも使えるのではないかと考えています。ただし、チケットデータはカテゴリカルデータが多いので適用には工夫が必要だと思われます。

# 因子分析

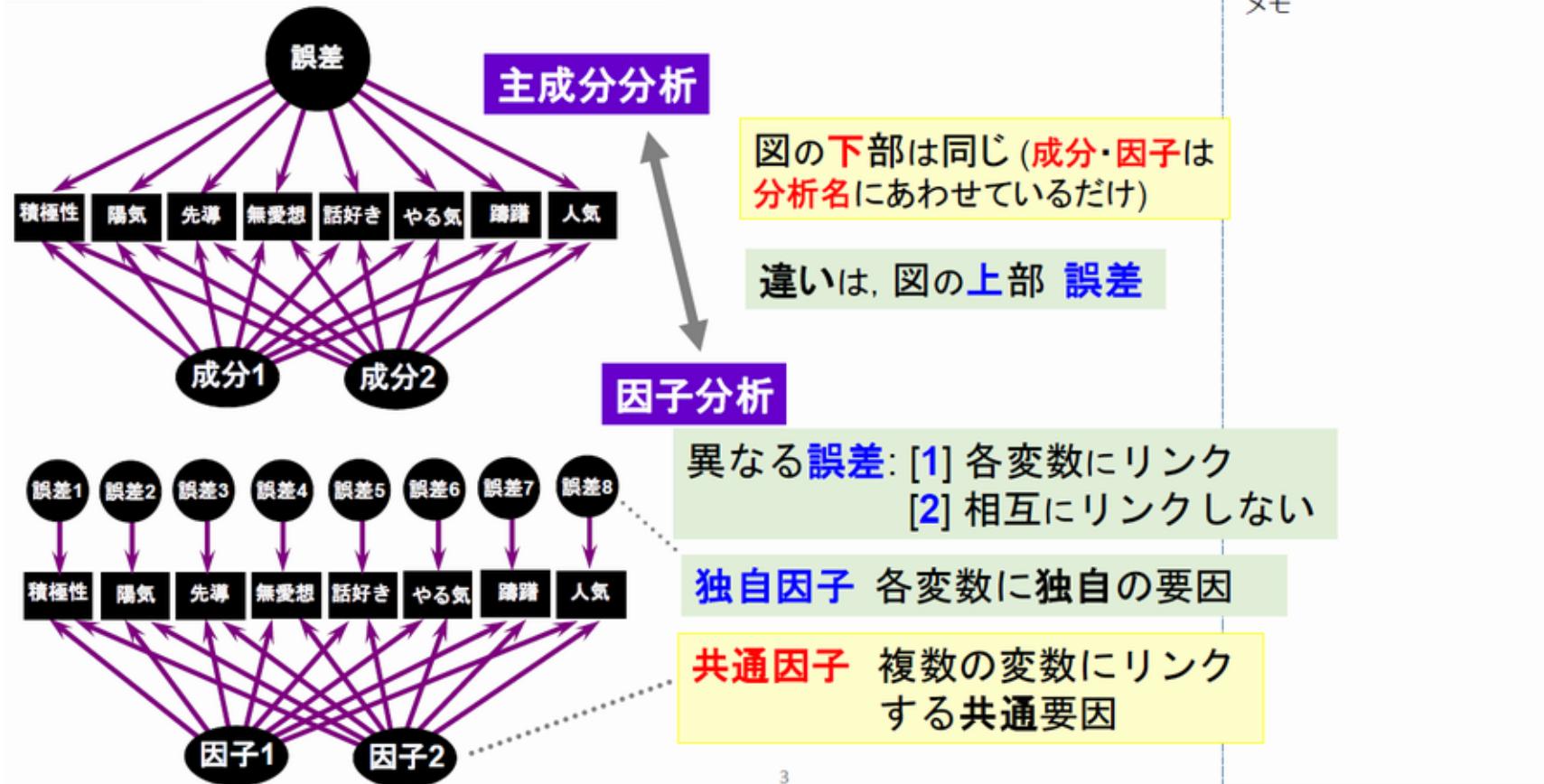
# 因子分析とは

因子分析とは主成分分析と同様に変量を次元縮約することで共通要因を探査し**独自要因を推定する**ための分析方法です。主成分分析と異なるのは、誤差は各変量に固有のもの（各変量に独自の要因）として考える点です。なお、因子分析における共通因子とは主成分分析における主成分に同じです。



# 主成分分析と因子分析

## 2. 主成分分析と因子分析の類似と相違



gacco 統計学III 第5回テキストより

# Rで解くには

---

Rで因子分析を行うには最尤法を用いたfactanal関数がありますが、今回は説明を省略します。「復習のためのリソース」などを参考にしてください。

# 演習

# 演習問題について

---

演習問題は以下の三問ですが、全てを解く必要はありません。また、今回の演習問題は計算をする問題ではなく各分析手法の結果を見て何が言えるかを話し合ってまとめてください。最後に発表していただきます。

| 演習    | 概要                         | 分析手法  |
|-------|----------------------------|-------|
| 演習I   | バグの多いモジュールはどのような傾向があるかを考える | 主成分分析 |
| 演習II  | 成績上位の投手はどのような傾向があるかを考える    | 主成分分析 |
| 演習III | (独立)因子がなにを表しているかを考える       | 因子分析  |

---

演習I

# プロダクト・メトリクス

---

あるオブジェクト指向プログラムのプロダクト・メトリクス(ソースコード・メトリクス)を用いてバグが発見されたメソッド(モジュール)の特徴、バグが発見されていないメソッド(モジュール)の特徴を整理してみてください。なお、各メトリクスの意味は以下の通りです。

TMLOC: 総行数(総メソッド行数)

WMC: 重み付きメソッド数 - サイクロマチック複雑度の合計

CBO: クラス間の結合度 - 値が小さいほど結合度が弱い(作りがシンプル)

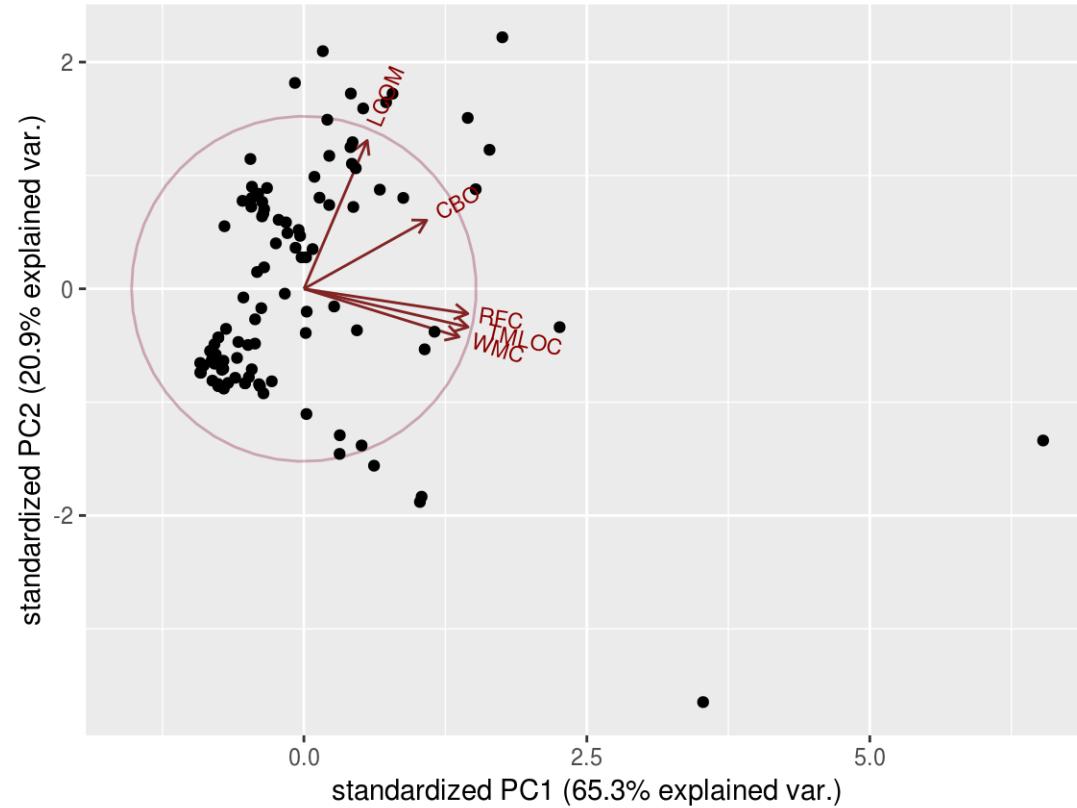
LCOM: クラスの凝集度の欠如 - 値が小さいほど凝集度が高い

RFC: 応答処理の多さ - 値が大きいほどメソッド呼び出しの種類が多い

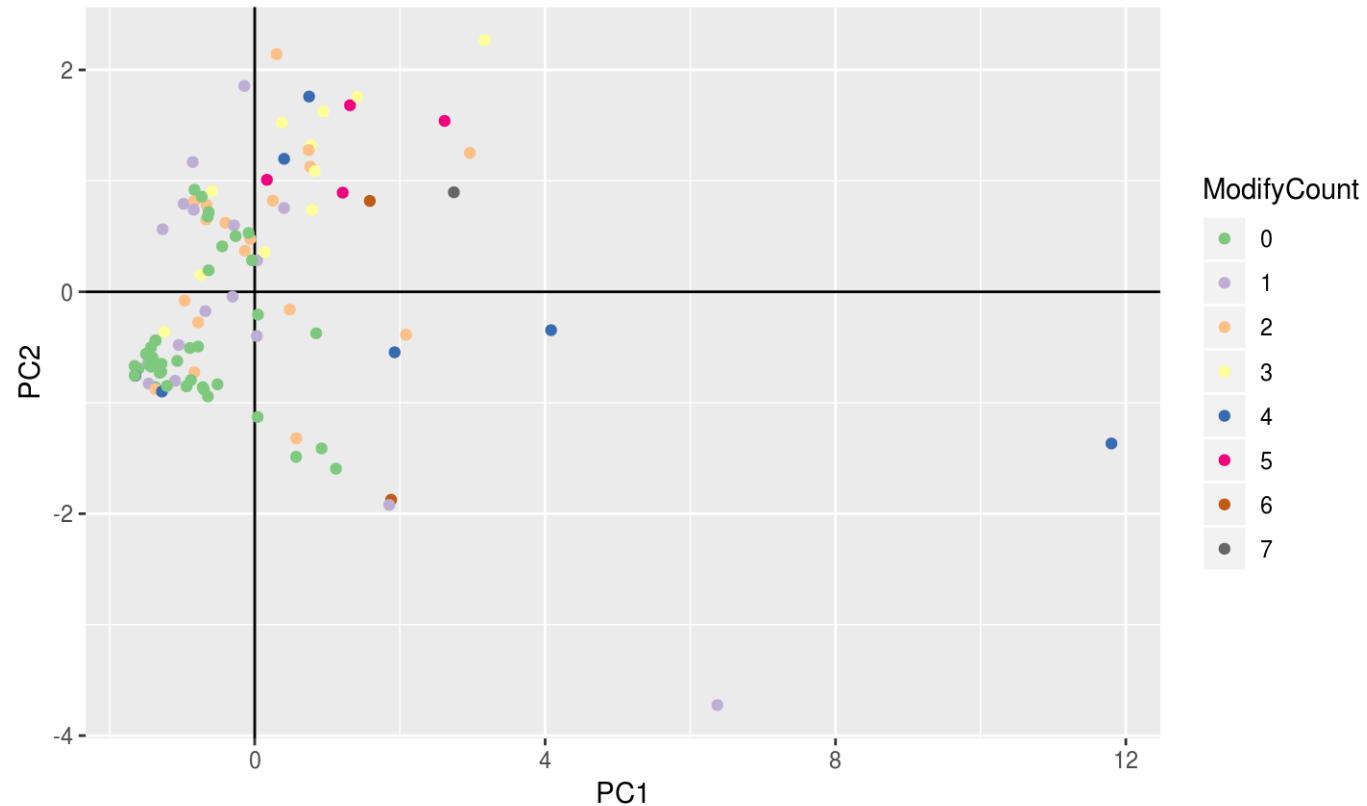
データなど出典: デート本

# 主成分プロット

---



# 主成分プロット(色分け)



# 演習 II

# 対象データ

---

対象となるデータはセリーグの投手成績(最終結果)です。成績には、総合、セーブ、ホールドポイントの三種類があります。次ページ以降の主成分分析結果を見て各投手がどのような特徴を持っているか分析してください。

## セーブ

「セーブとは、リードしているチームの救援投手が試合終了までリードを守りきることで付く投手記録。最多のセーブを記録した投手に最多セーブ投手のタイトルが与えられる。」([Wikipedia](#) より引用し修正)

## ホールドポイント

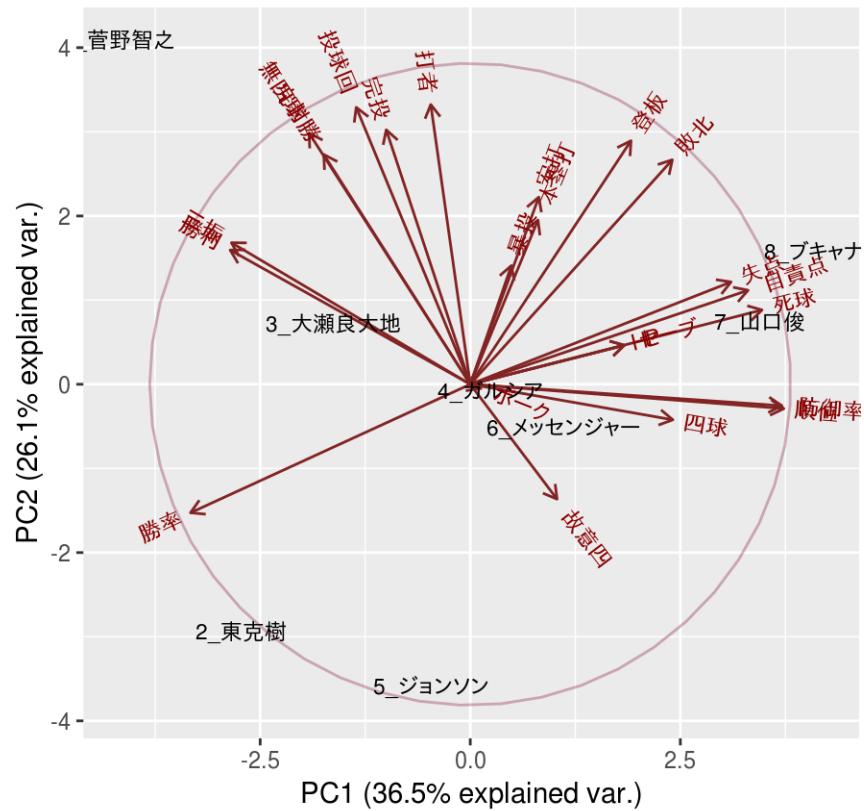
「ホールドポイントとは、ホールドと救援勝利の数を合計した数字のこと。2005年にセ・パ交流戦が始まったことに伴い、最優秀中継ぎ投手の新しい選考基準として考案された記録である。」([Wikipedia](#) より引用し修正)

# 総合成績

# 主成分分析表(総合成績)

---

# 主成分分析結果(総合成績)

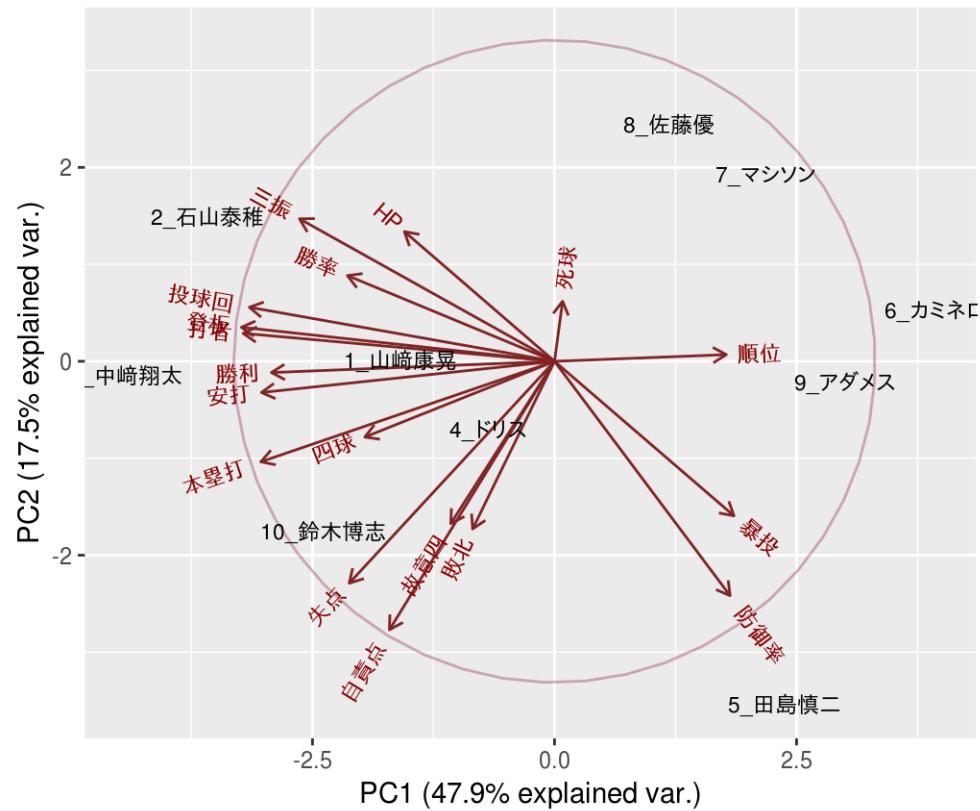


セーブ上位

# 主成分分析表(セーブ上位)

---

# 主成分分析結果(セーブ上位)

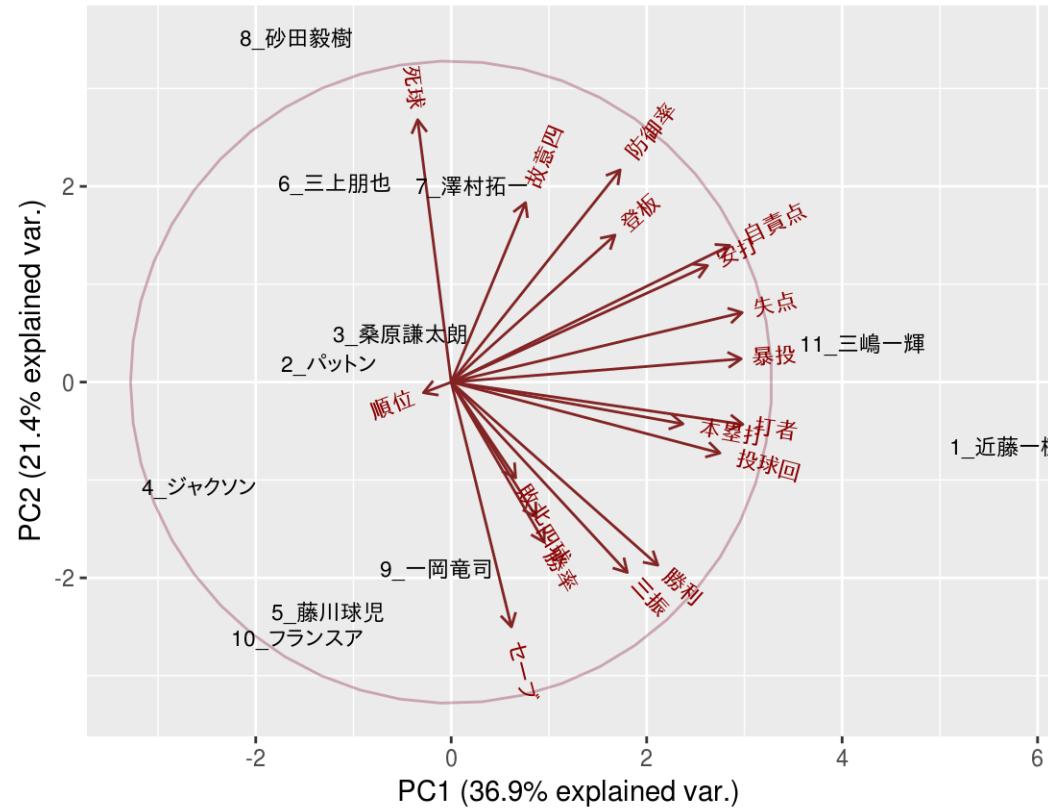


ホールドポイント上位

# 主成分分析表(ホールドポイント上位)

---

# 主成分分析結果(ホールドポイント上位)



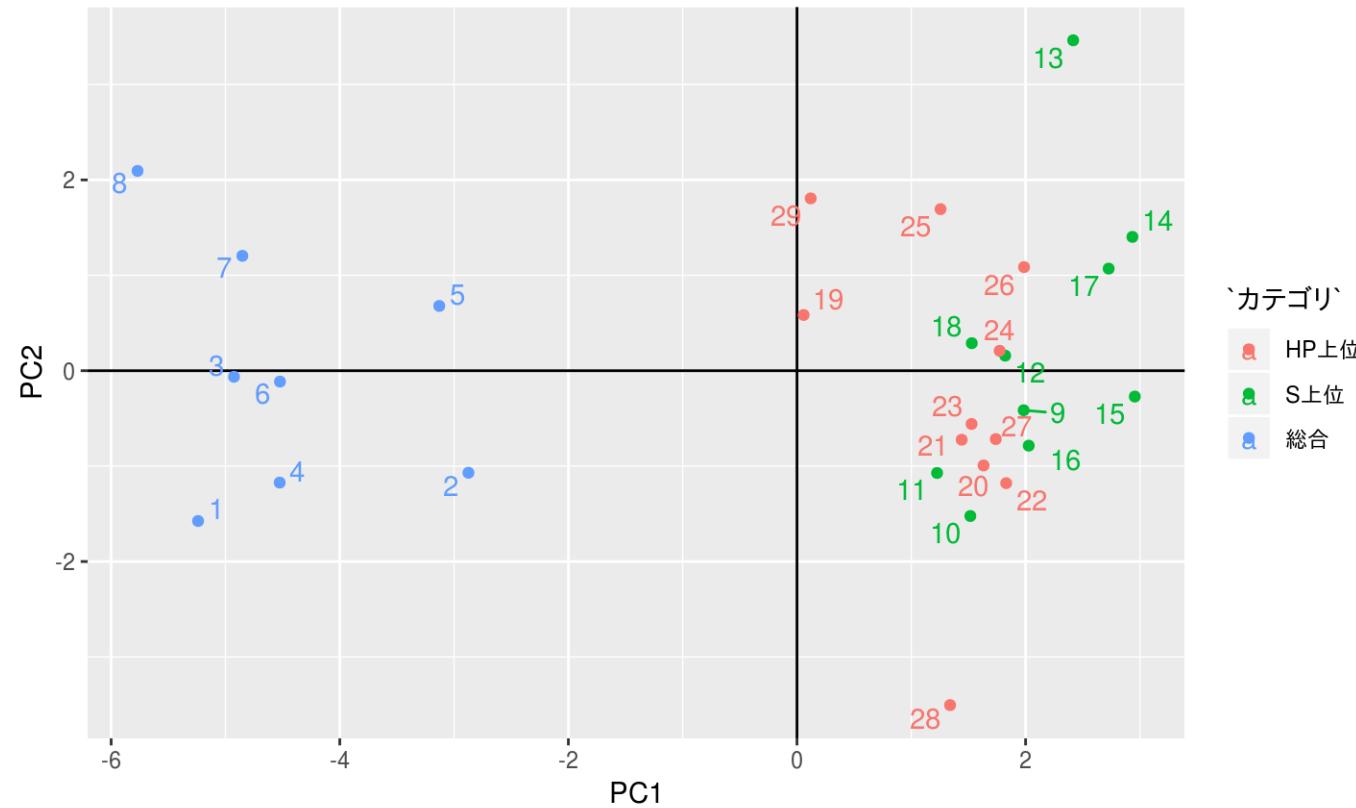
# 総合、セーブ上位、ホールドポイント上位 比較

# 主成分分析表

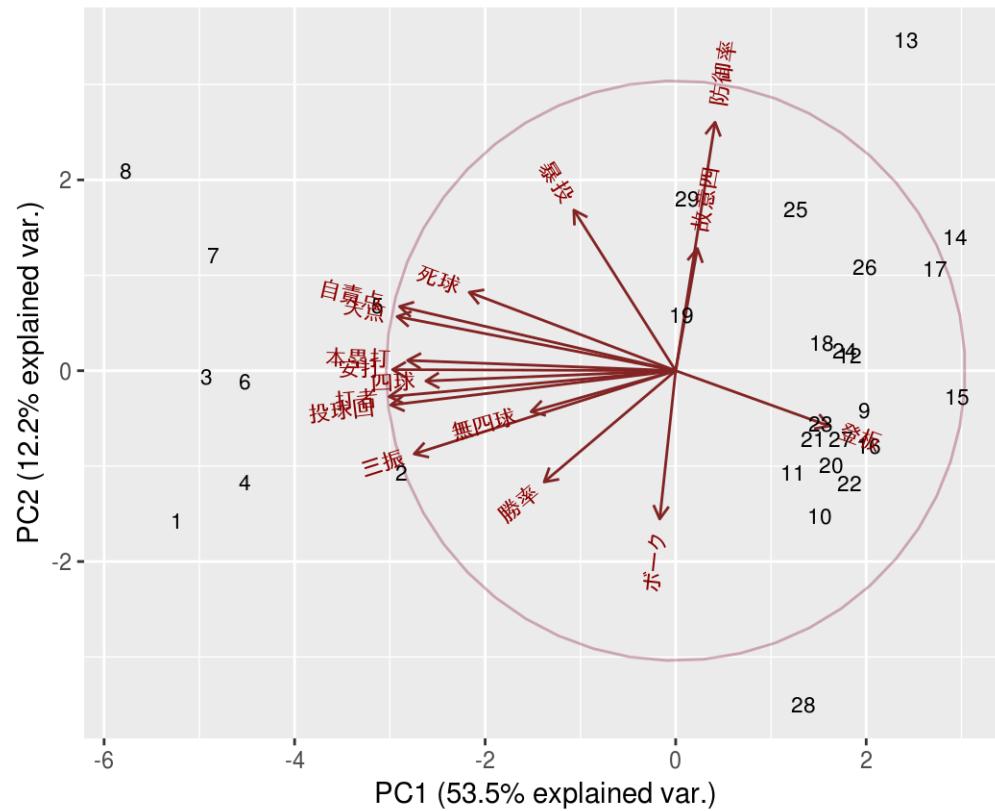
---

全てのカテゴリで比較するために「セーブ、HP、勝利、敗北、完投、完封勝」を除いた主成分分析結果です。

# 主成分プロット



# 主成分プロット



# 演習III

# 因子分析

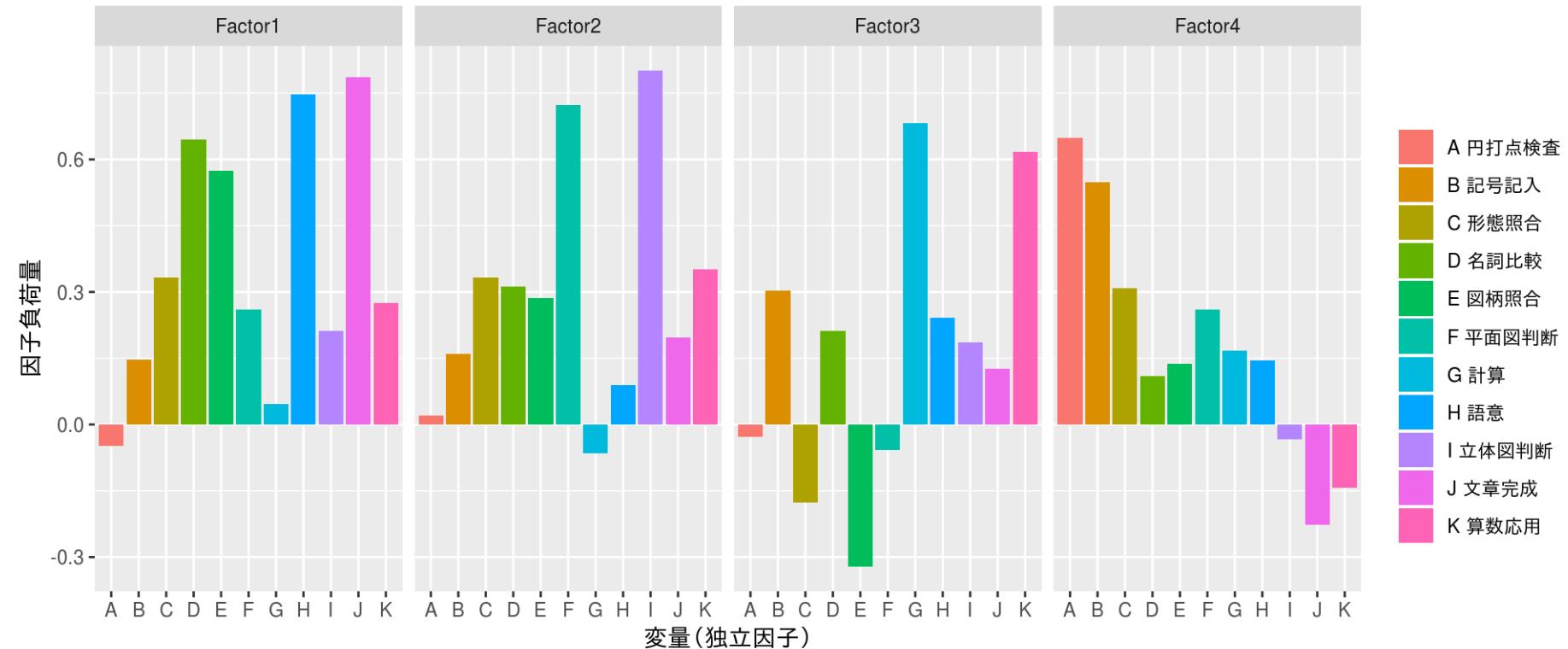
---

厚生労働省編一般職業適性検査 の筆記試験結果を因子分析したグラフが次ページにあります。各独自因子がどのような適性(能力)なのか分析してください。

| 内容                       | 内容                         |
|--------------------------|----------------------------|
| A 円打点(円の中に点を打つ)          | G 計算(加減乗除の計算を行う)           |
| B 記号記入(記号を記入する)          | H 語意(同意語かまたは反意語を見つけだす)     |
| C 形態照合(形と大きさの同じ図形をさがします) | I 立体図判断(展開図で表された立体形をさがします) |
| D 名詞比較(文字・数字の違いを見つける)    | J 文章完成(文章を完成する)            |
| E 図柄照合(同じ図形を見つけだす)       | K 算数応用(応用問題を解く)            |
| F 平面図判断(置き方をかえた図形を見つけだす) |                            |

# 因子分析結果

GATB



[データなど出典](#)

最後に

# 復習のためのリソース

---

- [gacco \(統計学IIIの開講予定は未定\)](#)
- [Rで主成分分析](#)
- [主成分分析とは](#)
- [\(Rを用いた\) 主成分分析 \(PDF\)](#)
- [Rと主成分分析 \(同志社大学\)](#)
- [Rと因子分析 \(同志社大学\)](#)

# Enjoy !

---

CC BY-NC-SA 4.0, Sampo Suzuki