

【データサイエンス基礎講座】:2014年12月度

データサイエンス基礎講座(超初級・実践編)  
2014年11月26日～12月17日<全5回>演習用資料

# R入門

## インストールから活用まで

- ・主催:株式会社インプレス
- ・企画/製作:フューチャーブリッジパートナーズ株式会社



# 第2限目

# 主成分分析2

- Rでのデータ入力

- koku = c(4,2,5,5,5,3,3,3,2)

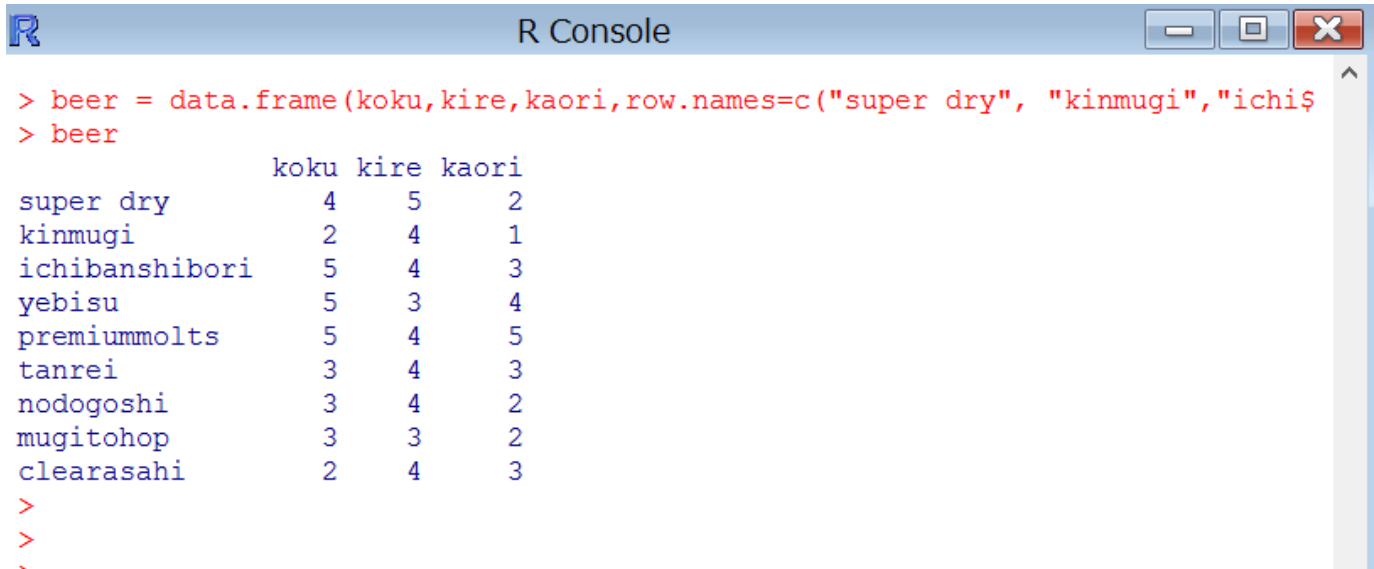
- kire = c(5,4,4,3,4,4,4,3,4)

- kaori = c(2,1,3,4,5,3,2,2,3)

- beer =

- data.frame(koku,kire,kaori,row.names=c("super  
dry", "kinmugi", "ichiban shibori", "yebisu",  
"premiummolts", "tanrei", "nodogoshi",  
"mugitohop", "clearasahi"))

- beer

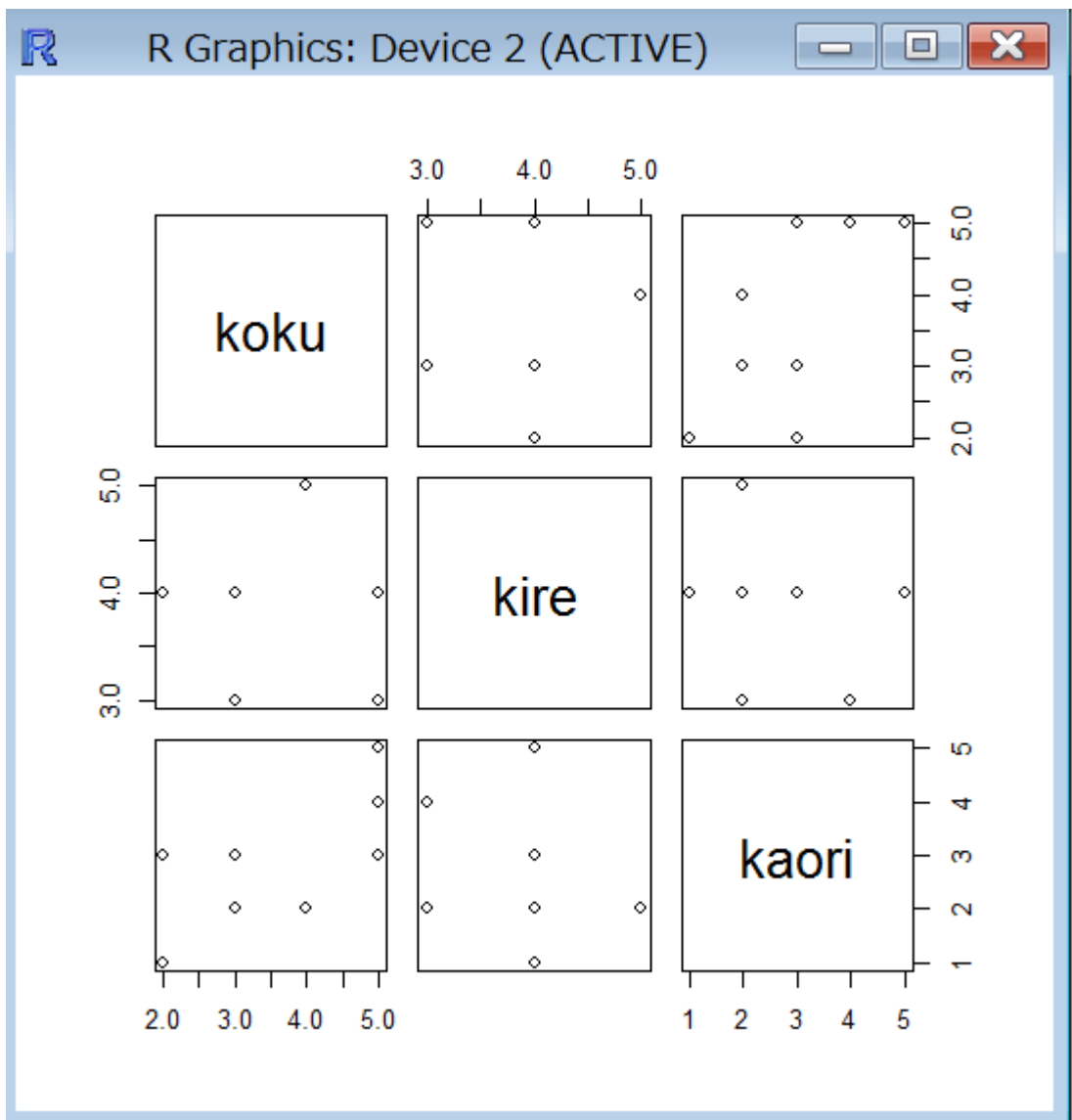


```
R Console
> beer = data.frame(koku,kire,kaori,row.names=c("super dry", "kinmugi", "ichi$
> beer
      koku kire kaori
super dry      4     5     2
kinmugi        2     4     1
ichibanshibori 5     4     3
yebisu         5     3     4
premiummolts   5     4     5
tanrei         3     4     3
nodogoshi      3     4     2
mugitohop      3     3     2
clearasahi     2     4     3
>
>
`
```

# 主成分分析3

- 散布図を描画

```
R Console  
> pairs(beer)  
> |
```



# 重回帰分析

- MASSパッケージをインストール
  - `install.packages("MASS")`
  - `library(MASS)`
- X1～X4を説明変数として、重回帰分析
  - `result = lm(Y ~ X1+X2+X3+X4, data=aictest)`
- AICに基づく変数選択
  - `result2 = stepAIC(result)`

```
R Console
> result2 = stepAIC(result)
Start: AIC=29.39
Y ~ X1 + X2 + X3 + X4

      Df Sum of Sq  RSS   AIC
- X4   1     3.233 72.763 27.846
<none>      69.531 29.392
- X1   1    41.797 111.328 32.099
- X2   1    43.966 113.497 32.292
- X3   1   165.279 234.810 39.562

Step: AIC=27.85
Y ~ X1 + X2 + X3

      Df Sum of Sq  RSS   AIC
<none>      72.763 27.846
- X1   1    39.909 112.673 30.219
- X2   1    84.007 156.770 33.522
- X3   1   168.982 241.746 37.853
>
>
>
> |
```

X4がない場合のAIC  
X4がある場合のAIC

現状のAICが一番  
小さいのでX1,X2,X3  
が最適解