データの分析と知識発見

第2回 講義の構成

- 関数を用いた基本統計量の計算
- 行列について
- Rを用いた行列の計算

Rでの関数の利用

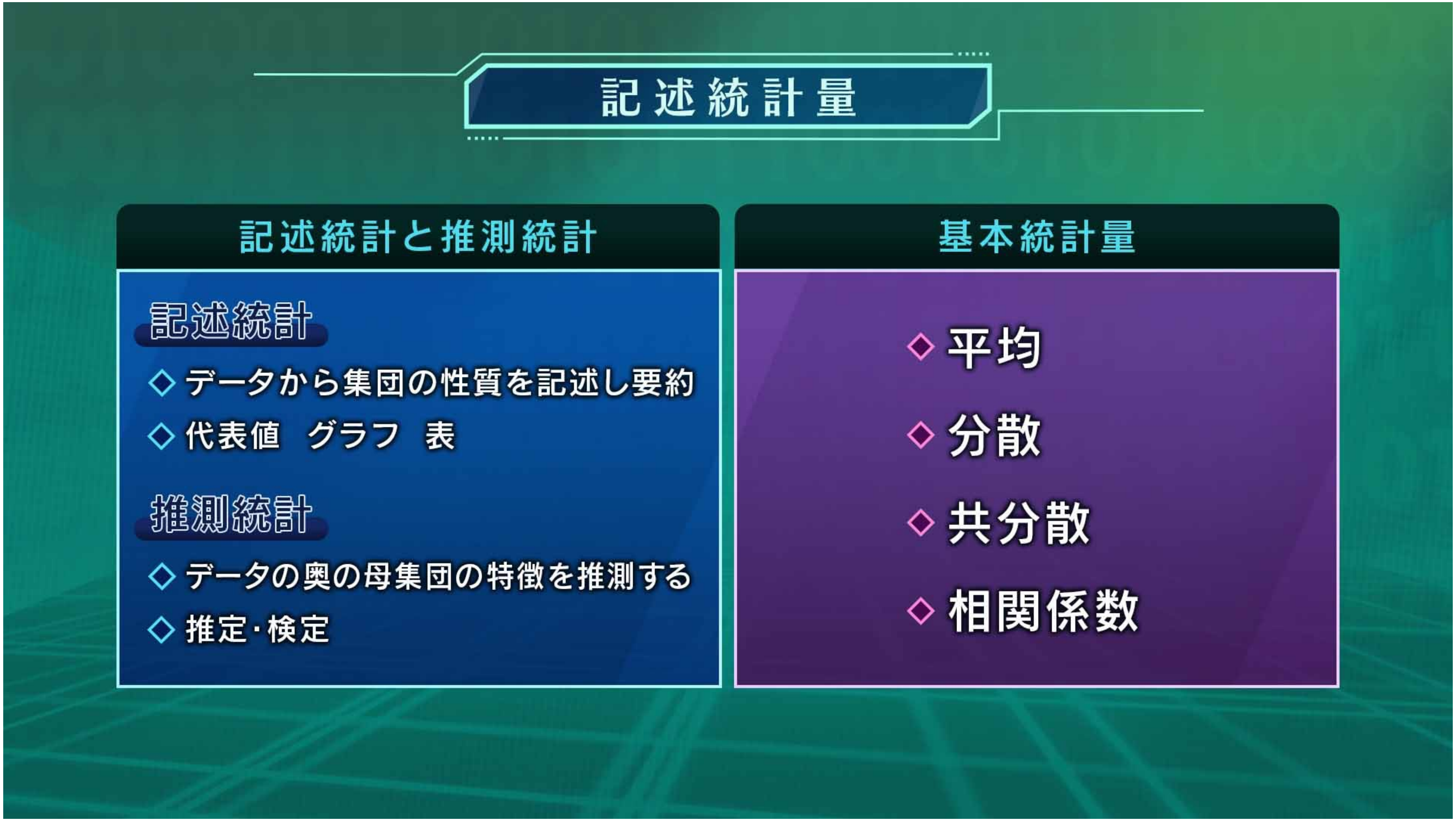
```
x1 <- c(118, 119, 121, 122, 170 )
x2 <- c(128, 129, 130, 131, 132 )
x1
x1[3]

mean(x1)

var(x1)

cor(x1,x2)
```

c()　結合する

3番目の要素にアクセスするにはx1[3]

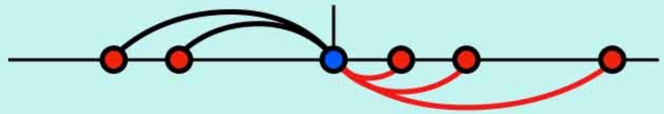
R における関数の例

統計量	説　明
sum	合計，例 sum(a,b,c)
mean	平均値，中央値は median
max	最大値，最小値は min
var	分散，行列を与えると分散共分散行列を計算する
sd	標準偏差，それぞれの列の標準偏差を求める
cor	相関，行列を与えると相関行列を求める


記述統計量
記述統計と推測統計
記述統計
◇ データから集団の性質を記述し要約
◇ 代表値　グラフ　表
推測統計
◇ データの奥の母集団の特徴を推測する
◇ 推定・検定
基本統計量
◇ 平均
◇ 分散
◇ 共分散
◇ 相関係数

基本統計量

平均：重心
標準偏差：ばらつきの大きさ



平均 $$\mu_x = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

分散 $$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

不偏分散 $$\sigma_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

標準偏差 $$\sigma_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

共分散と相関係数

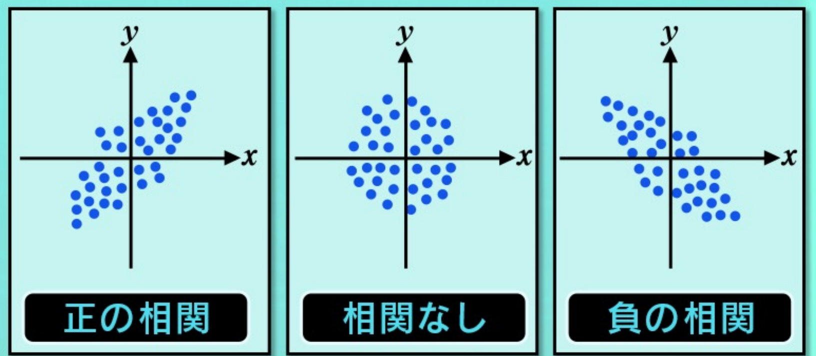
（不偏）共分散

$$Cov(x, y) = \sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

相関係数 ：関係の強さ

$$Cor(x, y) = \rho_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y}$$

$$-1 \leq \rho_{xy} \leq 1$$



行　列

$(x_1, x_2, \cdots, x_n)$ から $(y_1, y_2, \cdots, y_m)$ への変換が

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \qquad\vdots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{cases}$$

と表されるとするとき　これを

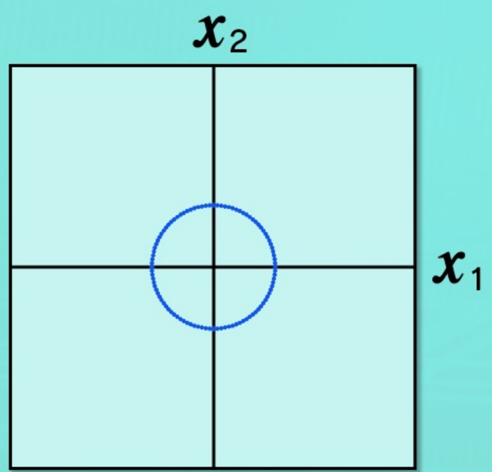
$$\underbrace{\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}}_{\text{ベクトル}} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}}_{\text{行列}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\text{ベクトル}} = A\boldsymbol{x}$$

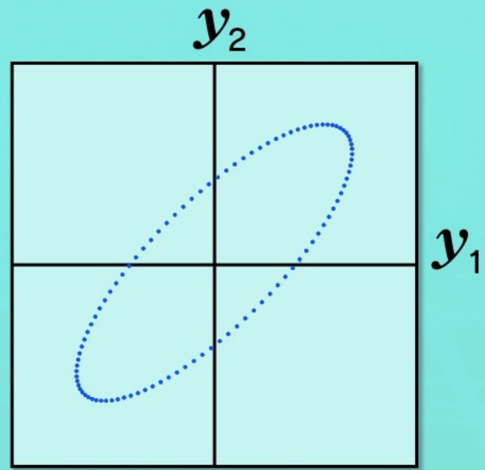
と表す

行列による変換の例（1）

$$\begin{cases} y_1 = 2x_1 + x_2 \\ y_2 = x_1 + 2x_2 \end{cases} \qquad A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

- ◇ 全体から全体へ変化
- ◇ 違う点は違う場所に変換



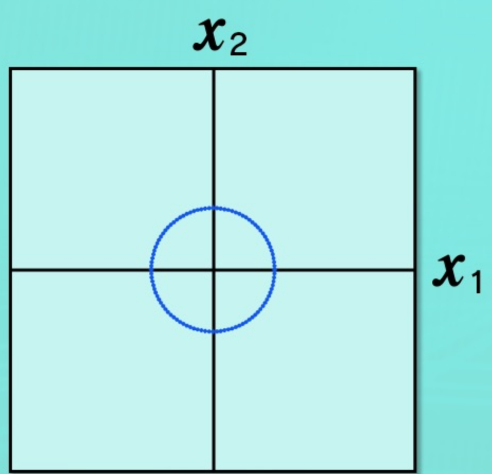


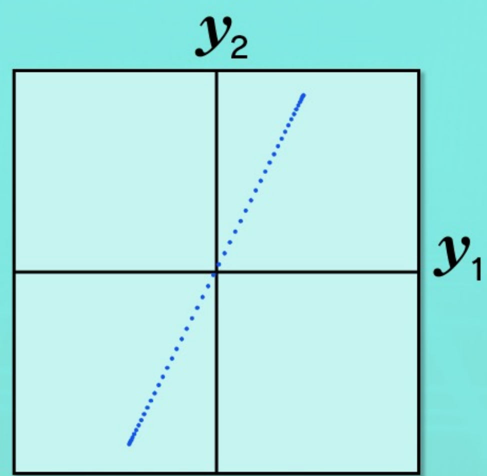
変換先から元の点へ
戻すことができる

行列による変換の例(2)

$$
\begin{cases}
y_1 = x_1 + x_2 \\
y_2 = 2x_1 + 2x_2
\end{cases}
\qquad
A = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}
$$

- ◇ 全体から直線へ変化
- ◇ 違う点が同じ点に
  変換されることもある





変換先から元の点へ
戻すことはできない

行列の特徴を調べる

ゼロ行列と単位行列

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

ゼロ行列 :変換するとすべての点が0

$$0 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

単位行列 :変換しても もとの位置のまま

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

行列の積

$$\underset{l \times n}{\begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{l1} & \cdots & c_{ln} \end{pmatrix}} = \underset{l \times m}{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \ddots & \ddots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{pmatrix}} \underset{m \times n}{\begin{pmatrix} b_{11} & \cdots & b_{1n} \\ b_{21} & \cdots & b_{2n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}}$$

ij成分：m 個の積の和

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj}$$

正方行列（行と列数が同じ行列）であっても AB と BA が同じとは限らない

逆行列

割り算 $a \neq 0$ のとき

$b \div a = \dfrac{b}{a}$ は $b \times \dfrac{1}{a} = \dfrac{b}{a}$ となる

ちなみに $a \times \dfrac{1}{a} = 1$

$\dfrac{1}{a} = a^{-1}$ 逆数

$B \div A$ ではなく

$AA^{-1} = I$ である A^{-1} を用いて

BA^{-1} 逆行列 を計算する

ここで $AA^{-1} = A^{-1}A = I$

ただし BA^{-1} と $A^{-1}B$ が等しいとは限らない

固有値・固有ベクトル

A：n 行 n 列の正方行列（逆行列を持つ場合を考える）

$$
\begin{cases}
A\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1 \\
A\boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_2 \\
\quad\vdots \\
A\boldsymbol{x}_n = \lambda_n \boldsymbol{x}_n
\end{cases}
$$

λ_i：固有値
$\boldsymbol{x}_i$：固有ベクトル
n 組存在する

n 組存在して それらが互いに独立となる

$$
P = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{pmatrix}
$$

が逆行列を持つ

固有値と対角化（1）

$$AP=\begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{pmatrix}=(A\boldsymbol{x}_1,\ A\boldsymbol{x}_2,\ \cdots,\ A\boldsymbol{x}_n)$$

$$=(\lambda_1\boldsymbol{x}_1,\ \lambda_1\boldsymbol{x}_2,\ \cdots,\ \lambda_n\boldsymbol{x}_n)=\begin{pmatrix} \lambda_1 x_{11} & \lambda_2 x_{21} & \cdots & \lambda_n x_{n1} \\ \lambda_1 x_{12} & \lambda_2 x_{22} & \cdots & \lambda_n x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1 x_{1n} & \lambda_2 x_{2n} & \cdots & \lambda_n x_{nn} \end{pmatrix}$$

$$=\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}=P\begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

固有値と対角化(2)

$$\begin{cases} A\boldsymbol{x}_1 = \lambda_1 \boldsymbol{x}_1 \\ A\boldsymbol{x}_2 = \lambda_2 \boldsymbol{x}_2 \\ \qquad \vdots \\ A\boldsymbol{x}_n = \lambda_n \boldsymbol{x}_n \end{cases}$$

のとき $P = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{pmatrix}$

とすると

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

A の特徴を表す

が成り立つ

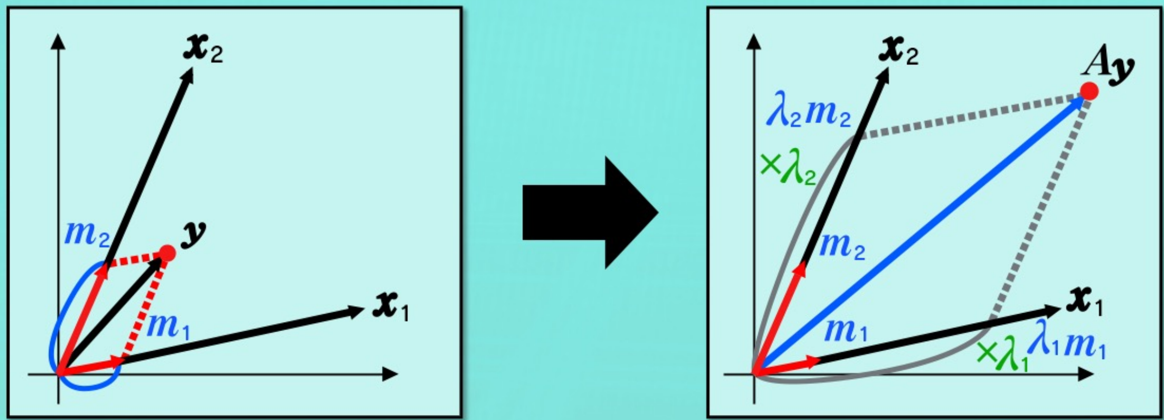
$$(P^{-1}AP)^n = P^{-1}AP \cdot P^{-1}AP \cdot \cdots \cdot P^{-1}AP = P^{-1}A^nP$$

固有ベクトルと変換との関係

$$\boldsymbol{y} = m_1\boldsymbol{x}_1 + m_2\boldsymbol{x}_2 + \cdots + m_n\boldsymbol{x}_n$$

$$A\boldsymbol{y} = \lambda_1 m_1\boldsymbol{x}_1 + \lambda_2 m_2\boldsymbol{x}_2 + \cdots + \lambda_n m_n\boldsymbol{x}_n$$

$$\boldsymbol{y} = P\begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{pmatrix} \qquad A\boldsymbol{y} = P\begin{pmatrix} \lambda_1 m_1 \\ \lambda_2 m_2 \\ \vdots \\ \lambda_n m_n \end{pmatrix}$$



Rと行列

- 行列を作るには matrix() という関数を使う
- 行：row　列：column
- 行数と列数を指定する (ncol= 2 など)
- 引数はベクトル
- c() でカンマで区切って与えてもよい
- 何も指定しないと1列目から入力
- byrow = TRUE とすると1行目から

C(1,2,3,4,5,6)

$$\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

行列の計算

```
A <- matrix( c( 1, 2, 3, 4), ncol= 2, nrow=2)
B <- matrix( c( 1, 2, 3, 4), ncol= 2, byrow=T )
A %*% B
C <- solve(A)
A %*% C
C %*% A

D <- eigen(A)
D
P <- D$vectors

solve(P) %*% A %*% P
```

行：row　列：column
matrix（ベクトル，行数と列数を指定）

何も指定しないと 列目ごと
byrow = TRUE とすると 行ごとに入力

solve() で逆行列

eigen()　　　固有値 固有ベクトル
$values　：固有値
$vectors　：固有ベクトル