第5回：多変量解析における行列の基礎

Code

1.　行列

数を長方形に並べたものを**行列**（matrix）といいます。特に縦に m 個、横に n 個だけ数を並べた行列をサイズ (m,n) の行列といいます。例えば、

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

は、サイズ $(2,3)$ 行列です。また行の数と列の数が同じ行列を**正方行列**といいます。また、m 行 1 列と数を縦に一列並べたものを**列ベクトル**、1 行 n 列と数を横に一行並べたものを**行ベクトル**といいます。

多変量解析には行列がたくさん登場します。みなさんがよく扱っているデータフレームは、まさに行列の代表例です。データフレームを**データ行列**とよんだりします。

Hide

```
# データの読み込み
dat <- read.csv("./data/scores.csv", fileEncoding = "utf-8")
dat
```

国語 <int>	数学 <int>	英語 <int>
90	30	155
95	35	115
100	50	200
90	45	150
85	20	120

5 rows

またデータの分散・共分散を一つにまとめて表現する**分散共分散行列**や、相関係数を一つにまとめて表現する**相関行列**も重要な役割を果たします。例えば先ほどのデータの分散共分散行列は、次のようなものです。

$$\begin{pmatrix} \text{国語の分散} & \text{国語と数学の共分散} & \text{国語と英語の共分散} \\ \text{数学と国語の共分散} & \text{数学の分散} & \text{数学と英語の共分散} \\ \text{英語と国語の共分散} & \text{英語と数学の共分散} & \text{英語の分散} \end{pmatrix}$$

R 言語では `cov` 関数を使って、分散共分散行列を計算することができます。

Hide

```
# 分散共分散行列
cov(dat)
```

```
      国語   数学    英語
国語  32.50  53.75  123.75
数学  53.75 142.50  296.25
英語 123.75 296.25 1157.50
```

この結果から「国語の分散は32.50」や「英語と数学の共分散は296.25」などがわかります。また相関行列とは、次のようなものです。

$$\begin{pmatrix} 1 & \text{国語と数学の相関係数} & \text{国語と英語の相関係数} \\ \text{数学と国語の相関係数} & 1 & \text{数学と英語の相関係数} \\ \text{英語と国語の相関係数} & \text{英語と数学の相関係数} & 1 \end{pmatrix}$$

R 言語では cor 関数を使って、分散共分散行列を計算することができます。

Hide

```
# 相関行列
cor(dat)
```

```
        国語      数学      英語
国語 1.0000000 0.7898222 0.6380328
数学 0.7898222 1.0000000 0.7294414
英語 0.6380328 0.7294414 1.0000000
```

この結果から「国語と数学の相関係数は0.79」などがわかります。

「行列を使わなくても何とかなるではないか」と思うかもしれませんが、行列を使うことで様々な数値をまとめて扱えるようになり、様々な統計量の計算に役立つわけです。

2. 行列・ベクトルの演算

行列・ベクトルの足し算、引き算、掛け算を説明します。

2.1 行列の足し算・引き算

同じサイズの行列どうしでは、各要素ごとに足し算・引き算をすることで、行列の足し算・引き算ができます。例を示します。

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 11 & 22 \\ 33 & 44 \end{pmatrix} = \begin{pmatrix} 12 & 24 \\ 36 & 48 \end{pmatrix}$$

$$\begin{pmatrix} 11 & 22 \\ 33 & 44 \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 10 & 20 \\ 30 & 40 \end{pmatrix}$$

2.2 行列の掛け算

行列の掛け算の説明は容易でないので、例を使って説明します。

$$\begin{aligned} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} &= \begin{pmatrix} 1 \times 5 + 2 \times 7 & 1 \times 6 + 2 \times 8 \\ 3 \times 5 + 4 \times 7 & 3 \times 6 + 4 \times 8 \end{pmatrix} \\ &= \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix} \end{aligned}$$

このように行列の掛け算は、足し算や引き算と異なり、サイズ (l, m) の行列とサイズ (m, n) の行列を掛けることで、サイズ (l, n) の行列を得ることができます。

2.3 演習

行列の足し算・引き算・掛け算の練習をやってみましょう。

問題 : 以下の行列の計算を行ってください。ただし、計算ができないものもあります。できないものについては、理由を述べてください。

1. $$\begin{pmatrix} 3 & 1 & -2 \\ 2 & -3 & 2 \end{pmatrix} + \begin{pmatrix} -1 & -5 & 2 \\ 2 & -3 & -1 \end{pmatrix}$$

2. $$\begin{pmatrix} 1 & 5 & -1 \\ -1 & 1 & 2 \end{pmatrix} + \begin{pmatrix} -1 & -5 \\ 2 & 2 \\ 3 & 3 \end{pmatrix}$$

3. $$\begin{pmatrix} 3 & 1 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} -1 & -5 \\ 2 & -3 \end{pmatrix}$$

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

5. $$\begin{pmatrix} 1 & 5 & -1 \end{pmatrix} \begin{pmatrix} -1 & -5 & 1 \end{pmatrix}$$

解答 :

1. $$\begin{pmatrix} 3 & 1 & -2 \\ 2 & -3 & 2 \end{pmatrix} + \begin{pmatrix} -1 & -5 & 2 \\ 2 & -3 & -1 \end{pmatrix} = \begin{pmatrix} 2 & -4 & 0 \\ 4 & -6 & 1 \end{pmatrix}$$

2. 足し算は同じサイズの行列どうしでしか計算できません。ところがこの問題の式は、左の行列のサイズが $(2, 3)$、右の行列のサイズが $(3, 2)$ になっています。つまり、足し算ができません。

3. $$\begin{aligned} \begin{pmatrix} 3 & 1 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} -1 & -5 \\ 2 & -3 \end{pmatrix} &= \begin{pmatrix} 3 \times (-1) + 1 \times 2 & 3 \times (-5) + 1 \times (-3) \\ 2 \times (-1) + (-3) \times 2 & 2 \times (-5) + (-3) \times (-3) \end{pmatrix} \\ &= \begin{pmatrix} -1 & -18 \\ -8 & -1 \end{pmatrix} \end{aligned}$$

4. $$\begin{pmatrix} -1 & -5 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 2 & -3 \end{pmatrix} = \begin{pmatrix} -13 & 14 \\ 0 & 11 \end{pmatrix}$$

5. 行列の掛け算では、サイズ (l, m) の行列とサイズ (m, n) の行列を掛けることで、サイズ (l, n) を得ることができるのでした。この問題の式では左の行列のサイズが $(1, 3)$、右の行列のサイズが $(1, 3)$ であることから、行列の掛け算を行うことができません。

2.4 単位行列と逆行列

正方行列には**単位行列**といって、数の掛け算でいうところの 1 の役割を果たす行列があります。例えば、サイズ $(2, 2)$ の行列の場合

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$
$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

が成り立ちます。同様にして、対角成分が 1 でその他の成分が 0 であるようなサイズ (n, n) の正方行列

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

は、サイズ (n, n) のどんな正方行列に対しても $AI_n = A$ かつ $I_nA = A$ をみたします。正方行列 I_n を**単位行列**といいます。

またサイズ (n, n) の正方行列 A に対して、掛け算すると単位行列 I_n になるような正方行列 B を行列 A の**逆行列**といいます。逆行列はどんな正方行列にも存在するとは限りません。$(2, 2)$ の正方行列の場合、次のような公式が知られています。

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \Rightarrow A^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

この公式を証明しておきましょう。$AA^{-1} = I_2$ と $A^{-1}A = I_2$ を確認する必要がありますが、$AA^{-1} = I_2$ のみ示します。

$$\begin{aligned} AA^{-1} &= \frac{1}{ad - bc}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \\ &= \frac{1}{ad - bc}\begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

問題 : 以下の行列に逆行列があれば求めてください。逆行列がない場合は理由を説明してください。

1. $\begin{pmatrix} 3 & 1 & -2 \\ 2 & -3 & 2 \end{pmatrix}$
2. $\begin{pmatrix} 3 & 1 \\ -8 & -3 \end{pmatrix}$
3. $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

解答 :

1. 正方行列でないので、逆行列は存在しません。
2. サイズ $(2, 2)$ の正方行列に対する逆行列の公式を用います。$A^{-1} = \begin{pmatrix} 3 & 1 \\ -8 & -3 \end{pmatrix}$
3. サイズ $(2, 2)$ の正方行列に対する逆行列の公式を用います。$A^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

3. 行列と連立一次方程式

連立一次方程式は行列を用いて表現することができます。この事実は主成分分析の仕組みを理解するときに役に立ちます。例えば、連立一次方程式

$$\begin{cases} 2x & + & 4y & = 10 \\ x & + & 3y & = 6 \end{cases}$$

は行列を用いると次のように表現することができます。

$$\begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 10 \\ 6 \end{pmatrix}$$

なお、逆行列を用いてこの連立一次方程式の解を求めることもできます。$A = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix}$ の逆行列は

$$A^{-1} = \frac{1}{2} \begin{pmatrix} 3 & -4 \\ -1 & 2 \end{pmatrix}$$

なので、以下のように解を得ることができます。

$$\begin{aligned} \begin{pmatrix} x \\ y \end{pmatrix} &= A^{-1} \begin{pmatrix} 10 \\ 6 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 3 & -4 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 10 \\ 6 \end{pmatrix} \\ &= \begin{pmatrix} 3 \\ 1 \end{pmatrix} \end{aligned}$$

問題 : 以下の連立一次方程式を行列を用いて表してください。

1. $\begin{cases} x & + & 3y & = 4 \\ 2x & + & 3y & = 5 \end{cases}$
2. $\begin{cases} -x & + & 2y & = 3 \\ 2x & + & y & = 4 \end{cases}$

解答 :

1. $\begin{pmatrix} 1 & 3 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 \\ 5 \end{pmatrix}$
2. $\begin{pmatrix} -1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$

4. 固有値問題

主成分分析の仕組みを理解するときに**固有値問題**が現れるので、ここで紹介します。固有値問題とは、正方行列 A に対して $Av = \lambda v$ をみたすような 0 でないベクトル v と値 λ の組 (v, λ) を求める問題のことです。ベクトル v のことを**固有ベクトル**、値 λ のことを**固有値**といいます。この問題は線形代数で解き方が知られているのですが、ここでは R 言語の関数 `eigen` を用いて解く方法を紹介します。

例題 : 行列 $A = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix}$ の固有値と固有ベクトルを求めてください。

解答 :

Hide

```
A <- matrix(c(3, 1,
              1, 2), nrow = 2, byrow = TRUE)
eigen(A)
```

```
eigen() decomposition
$values
[1] 3.618034 1.381966

$vectors
           [,1]       [,2]
[1,] -0.8506508  0.5257311
[2,] -0.5257311 -0.8506508
```

この結果から固有値 3.618 の固有ベクトル $\begin{pmatrix}-0.851\\-0.526\end{pmatrix}$ と固有値 0.526 の固有ベクトル $\begin{pmatrix}0.526\\-0.851\end{pmatrix}$ の二つがあることがわかります。■

問題：行列 $A = \begin{pmatrix}4 & 1 & 5\\1 & 9 & 3\\5 & 3 & 1\end{pmatrix}$ の固有値と固有ベクトルを求めてください。

解答：

Hide

```
A <- matrix(c(4, 1, 5,
              1, 9, 3,
              5, 3, 1), nrow = 3, byrow = TRUE)
eigen(A)
```

```
eigen() decomposition
$values
[1] 11.179454  5.828534 -3.007988

$vectors
           [,1]       [,2]       [,3]
[1,] -0.4167586  0.7170824  0.5586636
[2,] -0.7958462 -0.5848058  0.1569428
[3,] -0.4392506  0.3792031 -0.8144102
```

この結果から固有値 11.179 の固有ベクトル $\begin{pmatrix}-0.417\\-0.796\\-0.439\end{pmatrix}$、固有値 5.829 の固有ベクトル $\begin{pmatrix}0.717\\-0.585\\0.379\end{pmatrix}$、固有値 -3.008 の固有ベクトル \begin{pmatrix}0.559\\0.157\\-0.814\end{pmatrix の三つがあることがわかります。■