Rによる多変量解析

今野良彦ゼミ

20516030 河原崎由佳　　20516049 佐藤有香里

20516073 長谷川香奈　　20516083 柳田麻美

20516087 山本裕美子　　20516088 吉田美穂

【多変量データ】

- 採集されるデータのほとんどは多変量データ
- 変数を同時に調査しなければ、データの構造や特徴をとらえることはできない

【多変量解析】

- 無意味に見えるデータの中から意味のあるデータを探り当てる
- 必要な計算量が膨大なので適切なソフトを用いなければならない

【確率プロット】

柳田麻美　山本裕美子

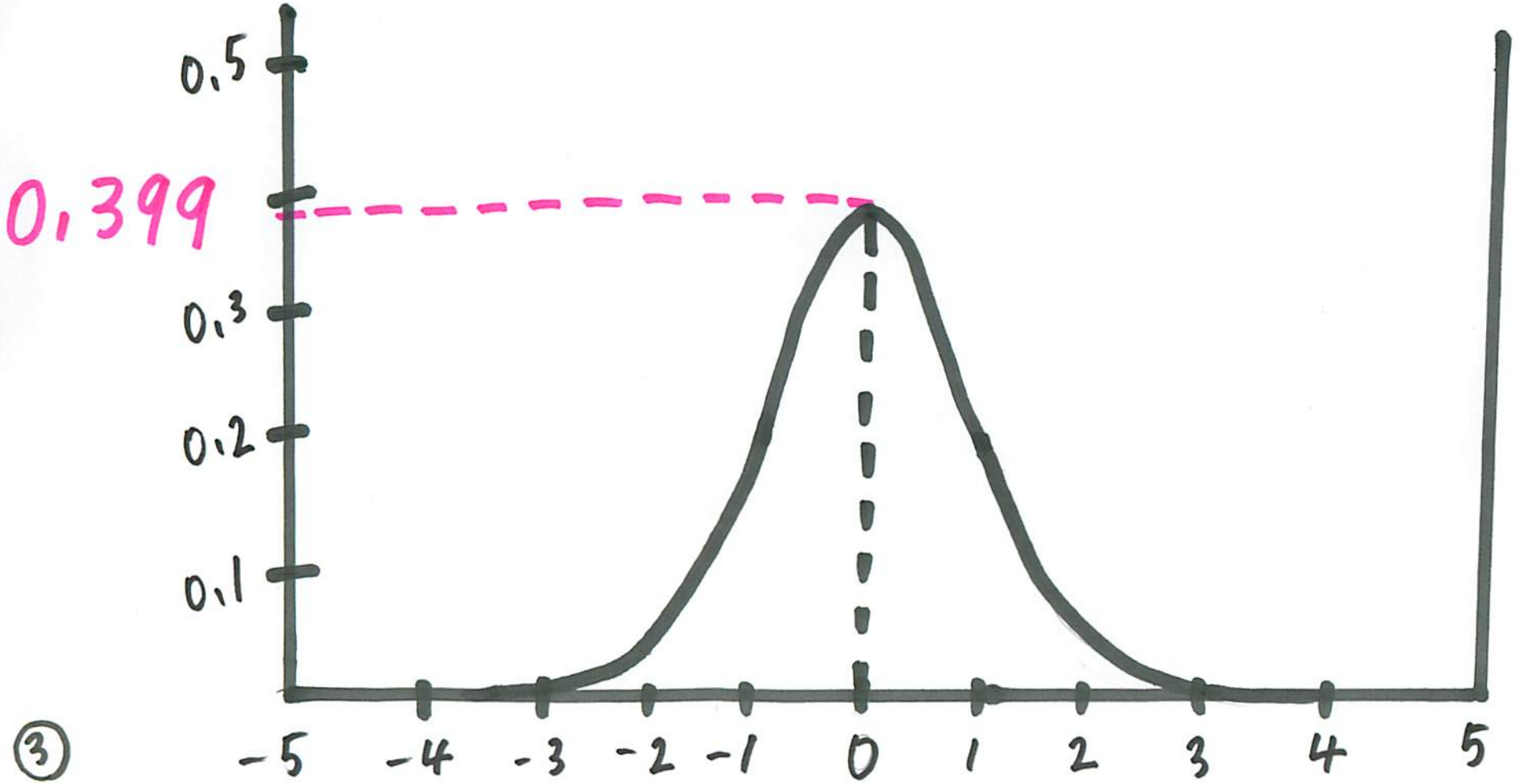
吉田美穂　佐藤有香里

【主成分分析】

河原崎由佳　長谷川香奈

標準正規分布の確率密度関数

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$$



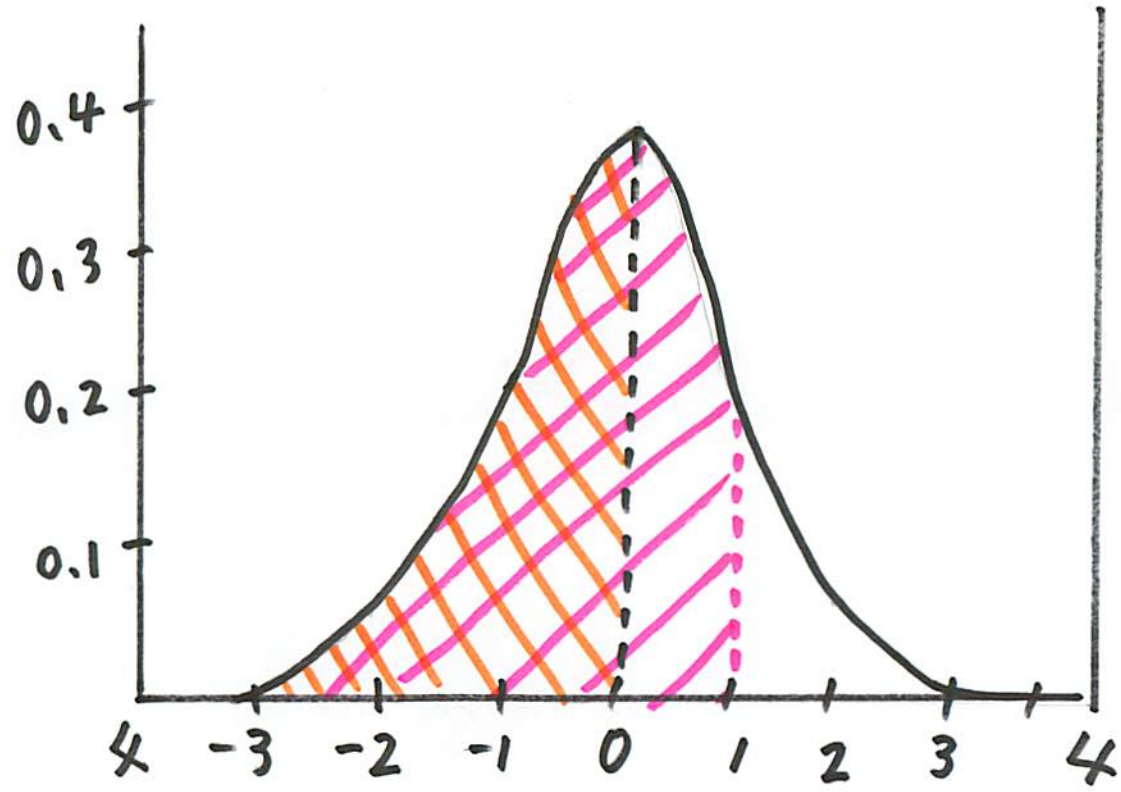
累積分布関数

$$\Phi(x)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du$$

$x=0$ のとき

$$\Phi(x)=\int_{-\infty}^{0}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u^2}du$$

$$=\frac{1}{2}$$



主成分分析

主成分分析とは･･･

- 多変量解析の手法の一つ
- データを互いに無相関なデータに書き換える手法

主成分分析の流れ･･･

①データを直交するベクトルを用いて無相関に変換
②新しいデータの大きさを規準化する
③新しいデータを情報量の多い順に並べる
④データの低次元での分析を行う

主成分分析の原理

多変量データ$\mathbf{x}=\begin{pmatrix}x_1\\x_2\\\vdots\\x_q\end{pmatrix}$ 係数$\mathbf{a_i}=\begin{pmatrix}a_{i1}\\a_{i2}\\\vdots\\a_{iq}\end{pmatrix},(i=1,2,\dots q)$

この２つの変数の線形結合

$$y_1=a_{11}x_1+a_{12}x_2+\cdots+a_{1q}x_q$$

$$y_2=a_{21}x_1+a_{22}x_2+\cdots+a_{2q}x_q$$

$$\vdots$$

$$y_q=a_{q1}x_1+a_{q2}x_2+\cdots+a_{qq}x_q$$ が主成分

係数ベクトルa_i決定の条件　$(i = 1, 2, \ldots, q)$

・各主成分を無相関にする

$$^t a_i a_j = 0 \ (i \neq j)$$

・各主成分を規準化する

$$^t a_i a_i = 1$$

・第1主成分y_1の分散が最も大きくなるようにする

・第2主成分y_2以降は同様に分散の大きい順に並べる

※特定の条件下での最大化にはラグランジュの
未定乗数法が用いられる

米国都市の大気汚染に関するデータ

	x_1	x_2	x_3	x_4	x_5	x_6
国名	Temp	Manuf	Pop	Wind	Precip	Days
Phoenix	70.3	213	582	6.0	7.05	36
Miami	75.5	207	335	9.0	59.80	128
Chicago	50.6	3344	3369	10.4	34.44	122
Buffalo	47.1	391	463	12.4	36.11	166
Philadelphia	54.6	1692	1950	9.6	39.93	115
Columbus	51.5	266	540	8.6	37.01	134
	⋮	⋮	⋮	⋮	⋮	⋮
Houston	68.9	721	1233	10.8	48.19	103
Richmond	57.8	197	299	7.6	42.59	115

主成分分析の結果

	主成分1	主成分2	主成分3	主成分4	主成分5	主成分6
分散	2.196	1.500	1.395	0.760	0.114	0.035
分散の割合	0.366	0.250	0.232	0.127	0.019	0.006
分散の累積	0.366	0.616	0.848	0.975	0.994	1.000

分散の合計＝6

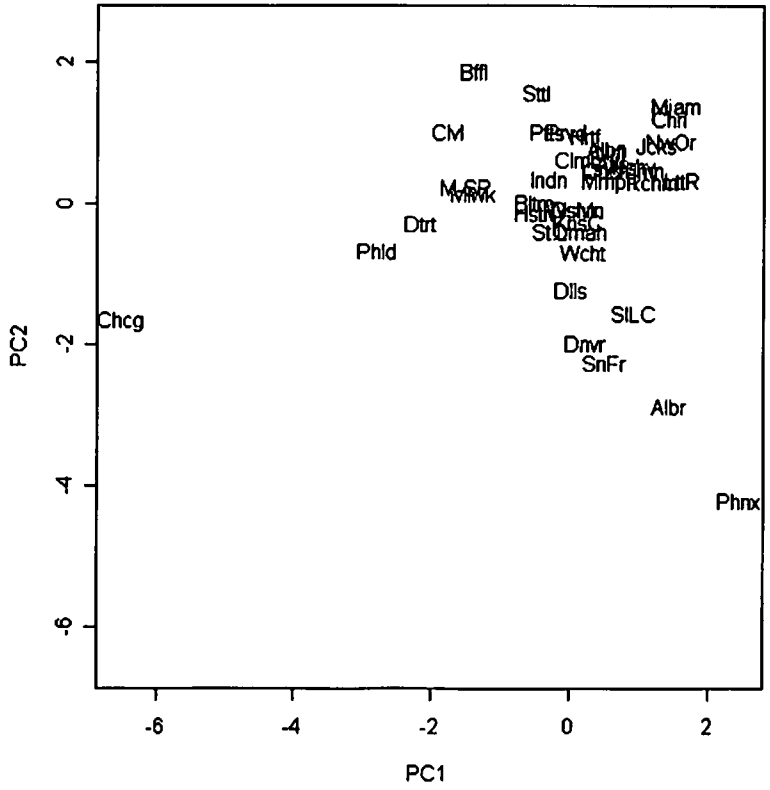
		Temp	Manuf	Pop	Wind	Precip	Days
a_1	主成分1	-0.330	-0.612	-0.578	-0.354	0	-0.238
a_2	主成分2	0.128	-0.168	-0.222	0.131	0.623	0.708

つまり、主成分は以下のとおりになる。

$$y_1 = (-0.330)\cdot x_1 + (-0.612)\cdot x_2 + \cdots + (-0.238)\cdot x_6$$

$$y_2 = 0.128\cdot x_1 + (-0.168)\cdot x_2 + \cdots + 0.708\cdot x_6$$

第一主成分と第二主成分の相関図



シカゴ（Chcg）

$$y_{11.1} = \begin{pmatrix} -0.330 \\ -0.612 \\ -0.578 \\ -0.354 \\ 0 \\ -0.238 \end{pmatrix}^{T} \begin{pmatrix} 0.714 \\ 5.113 \\ 4.767 \\ 0.669 \\ -0.197 \\ 0.305 \end{pmatrix} = -6.43$$

フェニックス（Phnx）

$$y_{1.1} = \begin{pmatrix} -0.330 \\ -0.612 \\ -0.578 \\ -0.354 \\ 0 \\ -0.238 \end{pmatrix}^{T} \begin{pmatrix} -2.011 \\ -0.444 \\ -0.046 \\ -2.411 \\ -2.522 \\ -2.939 \end{pmatrix} = 2.515$$