主成分分析

基本的な考え方

村田 昇

講義の内容

- **第 1 日 : 主成分分析の考え方**
- 第 2 日 : 分析の評価と視覚化

主成分分析の考え方

主成分分析

- 多数の変量のもつ情報の分析・視覚化
  - 変量を効率的に縮約して少数の特徴量を構成する
  - 特徴量に関与する変量間の関係を明らかにする
- **PCA** (Principal Component Analysis)
  - 構成する特徴量 : **主成分** (princial component)

主成分分析の例

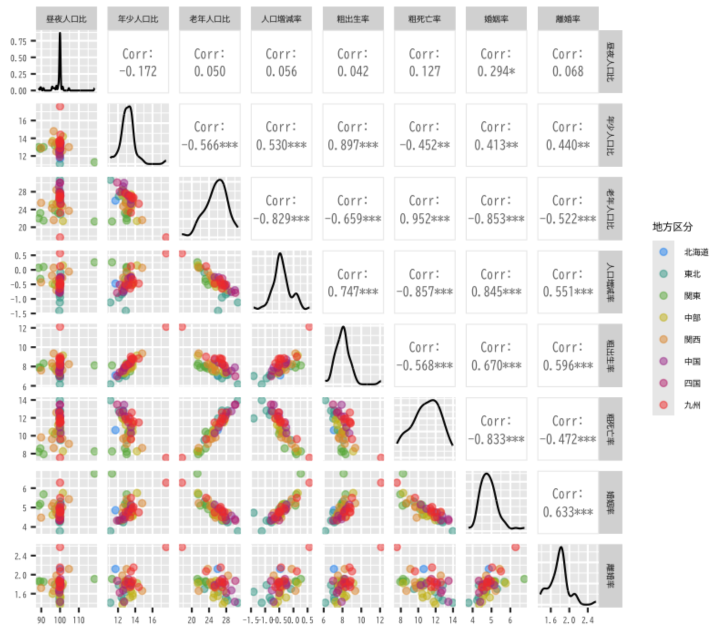

Figure 1: 人口関連データの散布図

Table 1: 都道府県別の人口関連データ

都道府県名	地方区分	昼夜人口比	年少人口比	老年人口比	人口増減率	粗出生率	粗死亡率	婚姻率	離婚率
北海道	北海道	100.0	11.7	26.0	-0.47	7.09	10.63	4.86	2.12
青森県	東北	100.0	12.1	27.0	-0.95	6.79	12.81	4.33	1.78
岩手県	東北	99.7	12.4	27.9	-0.84	7.12	12.33	4.32	1.52
宮城県	東北	100.2	13.0	22.9	-0.09	8.05	9.51	5.30	1.70
秋田県	東北	99.9	11.1	30.7	-1.12	6.16	13.98	3.78	1.41
山形県	東北	99.8	12.6	28.3	-0.78	7.13	12.81	4.24	1.46
福島県	東北	99.6	12.9	26.1	-1.41	7.02	11.94	4.73	1.64
茨城県	関東	97.2	13.2	23.8	-0.51	7.78	10.20	4.92	1.79
栃木県	関東	99.1	13.2	23.2	-0.40	8.02	10.43	5.13	1.85
群馬県	関東	99.9	13.4	24.9	-0.45	7.49	10.63	4.64	1.77
埼玉県	関東	88.6	13.0	22.0	0.07	7.90	8.20	5.10	1.86
千葉県	関東	89.5	12.8	23.2	-0.31	7.89	8.59	5.19	1.86
東京都	関東	118.4	11.3	21.3	0.26	8.12	8.25	6.75	1.91
神奈川県	関東	91.2	13.0	21.5	0.10	8.32	7.94	5.68	1.85
新潟県	中部	100.0	12.5	27.2	-0.64	7.45	11.97	4.35	1.37
富山県	中部	99.8	12.7	27.6	-0.55	7.28	11.79	4.50	1.43
石川県	中部	100.2	13.4	25.0	-0.26	8.21	10.51	4.91	1.52
福井県	中部	100.1	13.7	26.0	-0.50	8.40	11.01	4.55	1.55
山梨県	中部	99.0	12.9	25.6	-0.58	7.44	11.21	4.60	1.87
長野県	中部	99.9	13.5	27.4	-0.47	7.81	11.48	4.67	1.66
岐阜県	中部	96.0	13.7	25.2	-0.48	8.00	10.45	4.62	1.60
静岡県	中部	99.9	13.4	24.9	-0.37	8.25	10.23	5.17	1.84
愛知県	中部	101.5	14.2	21.4	0.15	9.14	8.26	5.75	1.82
三重県	関西	98.1	13.5	25.3	-0.38	8.00	10.44	4.89	1.76
滋賀県	関西	96.6	14.8	21.6	0.07	9.35	8.64	5.22	1.66
京都府	関西	101.2	12.6	24.7	-0.27	7.66	9.68	5.02	1.77
大阪府	関西	104.7	13.0	23.7	-0.06	8.24	9.09	5.43	2.12
兵庫県	関西	95.7	13.5	24.3	-0.20	8.34	9.63	5.07	1.84
奈良県	関西	89.9	12.9	25.5	-0.43	7.60	9.82	4.48	1.72
和歌山県	関西	98.1	12.5	28.4	-0.70	7.51	12.59	4.72	1.98
鳥取県	中国	100.0	13.2	27.2	-0.51	8.20	12.15	4.74	1.83
島根県	中国	100.0	12.7	30.0	-0.70	7.90	13.46	4.40	1.43
岡山県	中国	99.9	13.5	26.2	-0.26	8.41	10.94	4.94	1.82
広島県	中国	100.3	13.5	25.3	-0.25	8.72	10.28	5.15	1.78
山口県	中国	99.5	12.6	29.2	-0.76	7.55	12.74	4.58	1.67
徳島県	四国	99.7	12.2	28.0	-0.51	7.40	12.60	4.34	1.62
香川県	四国	100.2	13.2	27.1	-0.30	8.25	11.50	4.84	1.91
愛媛県	四国	100.1	12.8	27.8	-0.56	7.87	12.17	4.51	1.79
高知県	四国	99.9	11.9	30.1	-0.79	7.00	13.49	4.33	1.87
福岡県	九州	100.1	13.5	23.3	0.12	9.01	9.63	5.50	2.07
佐賀県	九州	100.2	14.4	25.3	-0.47	8.83	11.48	4.75	1.74
長崎県	九州	99.8	13.4	27.0	-0.64	8.33	11.92	4.50	1.74
熊本県	九州	99.6	13.7	26.5	-0.33	8.85	11.38	4.96	1.87
大分県	九州	100.0	12.9	27.6	-0.50	8.14	11.86	4.77	1.85
宮崎県	九州	100.0	13.8	26.7	-0.44	8.75	11.59	5.03	2.15
鹿児島県	九州	99.9	13.6	27.0	-0.53	8.78	12.59	4.78	1.84
沖縄県	九州	100.0	17.6	17.7	0.57	12.12	7.54	6.28	2.58

分析の枠組み

- $x_1, \dots, x_p$: **変数**
- $z_1, \dots, z_d$: **特徴量** ($d \le p$)
- 変数と特徴量の関係 (線形結合)

$$z_k = a_{1k}x_1 + \cdots + a_{pk}x_p \quad (k = 1, \dots, d)$$

  - 特徴量は定数倍の任意性があるので以下を仮定

$$\|\boldsymbol{a}_k\|^2 = \sum_{j=1}^{p} a_{jk}^2 = 1$$

主成分分析の用語

- 特徴量 z_k

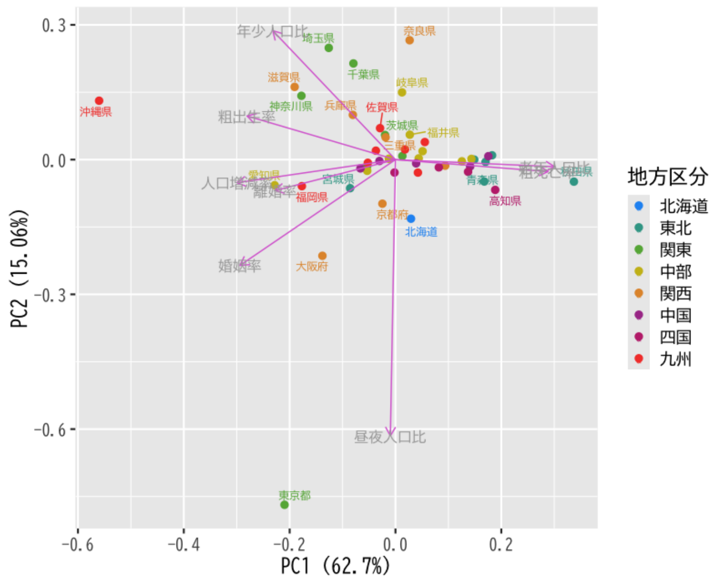

Figure 2: 主成分得点のバイプロット

  - 第 k **主成分得点** (principal component score)
  - 第 k **主成分**
- 係数ベクトル $\boldsymbol{a}_k$
  - 第 k **主成分負荷量** (principal component loading)
  - 第 k **主成分方向** (principal component direction)

分析の目的

- 目的

  主成分得点 $z_1, \dots, z_d$ が変数 $x_1, \dots, x_p$ の情報を効率よく反映するように主成分負荷量 $\boldsymbol{a}_1, \dots, \boldsymbol{a}_d$ を観測データから決定する
- 分析の方針 (以下は同値)
  - データの情報を最も保持する変量の **線形結合を構成**
  - データの情報を最も反映する **座標軸を探索**
- **教師なし学習** の代表的手法の 1 つ
  - 特徴抽出 : 情報処理に重要な特性を変数に凝集
  - 次元縮約 : 入力をできるだけ少ない変数で表現

第 1 主成分の計算

記号の準備

- 変数 : $x_1, \dots, x_p$ (p 次元)
- 観測データ : n 個の $(x_1, \dots, x_p)$ の組

$$\{(x_{i1}, \dots, x_{ip})\}_{i=1}^n$$

- ベクトル表現
  - $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top$: i 番目の観測データ (p 次元空間内の 1 点)
  - $\boldsymbol{a} = (a_1, \ldots, a_p)^\top$: 長さ 1 の p 次元ベクトル

係数ベクトルによる射影

- データ $\boldsymbol{x}_i$ の $\boldsymbol{a}$ 方向成分の長さ

$$\boldsymbol{a}^\top \boldsymbol{x}_i \quad (\text{スカラー})$$

- 方向ベクトル $\boldsymbol{a}$ をもつ直線上への点 $\boldsymbol{x}_i$ の直交射影

$$(\boldsymbol{a}^\top \boldsymbol{x}_i)\,\boldsymbol{a} \quad (\text{スカラー} \times \text{ベクトル})$$

幾何学的描像

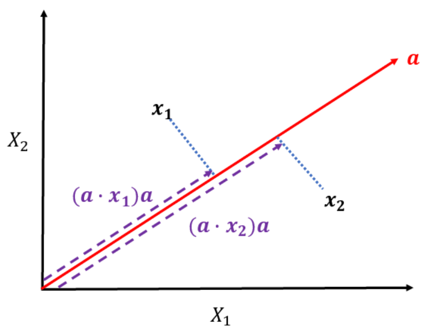

Figure 3: 観測データの直交射影 ($p = 2, n = 2$ の場合)

ベクトル a の選択の指針

- 射影による特徴量の構成

  ベクトル $\boldsymbol{a}$ を **うまく** 選んで観測データ $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n$ の情報を最も保持する 1 変量データ $z_1, \cdots, z_n$ を構成

$$z_1 = \boldsymbol{a}^\top \boldsymbol{x}_1, z_2 = \boldsymbol{a}^\top \boldsymbol{x}_2, \ldots, z_n = \boldsymbol{a}^\top \boldsymbol{x}_n$$

- 特徴量のばらつきの最大化

  観測データの **ばらつき**を最も反映するベクトル $\boldsymbol{a}$ を選択

$$\arg\max_{\boldsymbol{a}} \sum_{i=1}^{n} (\boldsymbol{a}^\top \boldsymbol{x}_i - \boldsymbol{a}^\top \bar{\boldsymbol{x}})^2, \quad \bar{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i,$$

ベクトル a の最適化

- 最適化問題

  制約条件 $\|\boldsymbol{a}\| = 1$ の下で以下の関数を最大化せよ

$$f(\boldsymbol{a}) = \sum_{i=1}^{n} (\boldsymbol{a}^\top \boldsymbol{x}_i - \boldsymbol{a}^\top \bar{\boldsymbol{x}})^2$$

- この最大化問題は必ず解をもつ
  - $f(\boldsymbol{a})$ は連続関数
  - 集合 $\{\boldsymbol{a} \in \mathbb{R}^p : \|\boldsymbol{a}\| = 1\}$ はコンパクト (有界閉集合)

演習

問題

- 以下の問に答えなさい
  - 評価関数 $f(\boldsymbol{a})$ を以下の中心化したデータ行列で表しなさい

$$X = \begin{pmatrix} \boldsymbol{x}_1^\top - \bar{\boldsymbol{x}}^\top \\ \vdots \\ \boldsymbol{x}_n^\top - \bar{\boldsymbol{x}}^\top \end{pmatrix} = \begin{pmatrix} x_{11} - \bar{x}_1 & \cdots & x_{1p} - \bar{x}_p \\ \vdots & & \vdots \\ x_{n1} - \bar{x}_1 & \cdots & x_{np} - \bar{x}_p \end{pmatrix}$$

  - 上の結果を用いて次の最適化問題の解の条件を求めなさい

$$\text{maximize} \quad f(\boldsymbol{a}) \quad \text{s.t.} \quad \boldsymbol{a}^\top \boldsymbol{a} = 1$$

解答例

- 定義どおりに計算する

$$\begin{aligned} f(\boldsymbol{a}) &= \sum_{i=1}^{n} (\boldsymbol{a}^\top \boldsymbol{x}_i - \boldsymbol{a}^\top \bar{\boldsymbol{x}})^2 \\ &= \sum_{i=1}^{n} (\boldsymbol{a}^\top \boldsymbol{x}_i - \boldsymbol{a}^\top \bar{\boldsymbol{x}})(\boldsymbol{x}_i^\top \boldsymbol{a} - \bar{\boldsymbol{x}} \boldsymbol{a}^\top) \\ &= \boldsymbol{a}^\top X^\top X \boldsymbol{a} \end{aligned}$$

  - 回帰分析の Gram 行列を参照

- 制約付き最適化なので未定係数法を用いればよい

$$L(\boldsymbol{a}, \lambda) = f(\boldsymbol{a}) + \lambda(1 - \boldsymbol{a}^\top \boldsymbol{a})$$

  の鞍点

$$\frac{\partial}{\partial \boldsymbol{a}} L(\boldsymbol{a}, \lambda) = 0$$

  を求めればよいので

$$\begin{aligned} 2X^\top X \boldsymbol{a} - 2\lambda \boldsymbol{a} &= 0 \\ X^\top X \boldsymbol{a} &= \lambda \boldsymbol{a} \quad \text{(固有値問題)} \end{aligned}$$

第 1 主成分の解

ベクトル $\boldsymbol{a}$ の解

- 最適化問題

$$\text{maximize} \quad f(\boldsymbol{a}) = \boldsymbol{a}^\top X^\top X \boldsymbol{a} \quad \text{s.t.} \quad \boldsymbol{a}^\top \boldsymbol{a} = 1$$

- 固有値問題
  $f(\boldsymbol{a})$ の極大値を与える $\boldsymbol{a}$ は $X^\top X$ の固有ベクトルとなる
  $$X^\top X \boldsymbol{a} = \lambda \boldsymbol{a}$$

第 1 主成分

- 固有ベクトル $\boldsymbol{a}$ に対する $f(\boldsymbol{a})$ は行列 $X^\top X$ の固有値

$$f(\boldsymbol{a}) = \boldsymbol{a}^\top X^\top X \boldsymbol{a} = \boldsymbol{a}^\top \lambda \boldsymbol{a} = \lambda$$

- 求める $\boldsymbol{a}$ は行列 $X^\top X$ の最大固有ベクトル (長さ 1)
- **第 1 主成分負荷量** : 最大 (第一) 固有ベクトル $\boldsymbol{a}$
- **第 1 主成分得点**

$$z_{i1} = a_1 x_{i1} + \cdots + a_p x_{ip} = \boldsymbol{a}^\top \boldsymbol{x}_i, \quad (i = 1, \ldots, n)$$

Gram 行列の性質

Gram 行列の固有値

- $X^\top X$ は半正定値行列
- $X^\top X$ の固有値は 0 以上の実数
  - 固有値を重複を許して降順に並べる

$$\lambda_1 \geq \cdots \geq \lambda_p \quad (\geq 0)$$

  - 固有値 λ_k に対する固有ベクトルを $\boldsymbol{a}_k$(長さ 1) とする

$$\|\boldsymbol{a}_k\| = 1, \quad (k = 1, \ldots, p)$$

Gram 行列のスペクトル分解

- $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_p$ は **互いに直交** するようとることができる

$$j \neq k \quad \Rightarrow \quad \boldsymbol{a}_j^\top \boldsymbol{a}_k = 0$$

- 行列 $X^\top X$ (半正定値行列) のスペクトル分解

$$\begin{aligned} X^\top X &= \lambda_1 \boldsymbol{a}_1 \boldsymbol{a}_1^\top + \lambda_2 \boldsymbol{a}_2 \boldsymbol{a}_2^\top + \cdots + \lambda_p \boldsymbol{a}_p \boldsymbol{a}_p^\top \\ &= \sum_{k=1}^{p} \lambda_k \boldsymbol{a}_k \boldsymbol{a}_k^\top \end{aligned}$$

  - 固有値と固有ベクトルによる行列の表現

演習

問題

- 以下の問に答えなさい
  - Gram 行列のスペクトル分解において λ_j と $\boldsymbol{a}_j$ が固有値・固有ベクトルとなることを確かめなさい

$$X^\top X = \sum_{k=1}^{p} \lambda_k \boldsymbol{a}_k \boldsymbol{a}_k^\top$$

  - 以下の行列を用いて Gram 行列のスペクトル分解を書き直しなさい

$$A = \begin{pmatrix} \boldsymbol{a}_1^\top \\ \vdots \\ \boldsymbol{a}_p^\top \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & & 0 \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

解答例

- 固有ベクトルの直交性に注意する

$$\begin{aligned} X^\top X \boldsymbol{a}_j &= \sum_{k=1}^{p} \lambda_k \boldsymbol{a}_k \boldsymbol{a}_k^\top \boldsymbol{a}_j && (\text{直交性}) \\ &= \lambda_j \boldsymbol{a}_j \boldsymbol{a}_j^\top \boldsymbol{a}_j && (\text{単位ベクトル}) \\ &= \lambda_j \boldsymbol{a}_j \end{aligned}$$

- 転置に注意して計算する

$$X^\top X = A^\top \Lambda A$$

第 2 主成分以降の計算

第 2 主成分の考え方

- 第 1 主成分
  - 主成分負荷量 : ベクトル $\boldsymbol{a}_1$
  - 主成分得点 : $\boldsymbol{a}_1^\top \boldsymbol{x}_i$ $(i = 1, \ldots, n)$
- 第 1 主成分負荷量に関してデータが有する情報

$$(\boldsymbol{a}_1^\top \boldsymbol{x}_i)\, \boldsymbol{a}_1 \quad (i = 1, \ldots, n)$$

- 第 1 主成分を取り除いた観測データ (分析対象)

$$\tilde{\boldsymbol{x}}_i = \boldsymbol{x}_i - (\boldsymbol{a}_1^\top \boldsymbol{x}_i)\, \boldsymbol{a}_1 \quad (i = 1, \ldots, n)$$

第 2 主成分の最適化

- 最適化問題

  制約条件 $\|\boldsymbol{a}\| = 1$ の下で以下の関数を最大化せよ

$$\tilde{f}(\boldsymbol{a}) = \sum_{i=1}^{n}(\boldsymbol{a}^{\top}\tilde{\boldsymbol{x}}_i - \boldsymbol{a}^{\top}\bar{\tilde{\boldsymbol{x}}})^2 \quad \text{ただし} \quad \bar{\tilde{\boldsymbol{x}}} = \frac{1}{n}\sum_{i=1}^{n}\tilde{\boldsymbol{x}}_i$$

演習

問題

- 以下の問に答えなさい
  - 以下の中心化したデータ行列を X と $\boldsymbol{a}_1$ で表しなさい

$$\tilde{X} = \begin{pmatrix} \tilde{\boldsymbol{x}}_1^{\top} - \bar{\tilde{\boldsymbol{x}}}^{\top} \\ \vdots \\ \tilde{\boldsymbol{x}}_n^{\top} - \bar{\tilde{\boldsymbol{x}}}^{\top} \end{pmatrix}$$

  - 上の結果を用いて次の最適化問題の解を求めなさい

$$\text{maximize} \quad \tilde{f}(\boldsymbol{a}) \quad \text{s.t.} \quad \boldsymbol{a}^{\top}\boldsymbol{a} = 1$$

解答例

- 定義どおりに計算する

$$\tilde{X} = \begin{pmatrix} \tilde{\boldsymbol{x}}_1^{\top} - \bar{\tilde{\boldsymbol{x}}}^{\top} \\ \vdots \\ \tilde{\boldsymbol{x}}_n^{\top} - \bar{\tilde{\boldsymbol{x}}}^{\top} \end{pmatrix} = X - X\boldsymbol{a}_1\boldsymbol{a}_1^{\top}$$

- Gram 行列 $\tilde{X}^{\top}\tilde{X}$ を計算する

$$\begin{aligned}
\tilde{X}^{\top}\tilde{X} &= (X - X\boldsymbol{a}_1\boldsymbol{a}_1^{\top})^{\top}(X - X\boldsymbol{a}_1\boldsymbol{a}_1^{\top}) \\
&= X^{\top}X - X^{\top}X\boldsymbol{a}_1\boldsymbol{a}_1^{\top} - \boldsymbol{a}_1\boldsymbol{a}_1^{\top}X^{\top}X + \boldsymbol{a}_1\boldsymbol{a}_1^{\top}X^{\top}X\boldsymbol{a}_1\boldsymbol{a}_1^{\top} \\
&= X^{\top}X - \lambda_1\boldsymbol{a}_1\boldsymbol{a}_1^{\top} \\
&= \sum_{k=2}^{p}\lambda_k\boldsymbol{a}_k\boldsymbol{a}_k^{\top}
\end{aligned}$$

  元の Gram 行列 $X^{\top}X$ の固有ベクトル $\boldsymbol{a}_1$ の固有値が 0 となっていると考えることができる

第 2 主成分以降の解

第 2 主成分

- Gram 行列 $\tilde{X}^{\top}\tilde{X}$ の固有ベクトル $\boldsymbol{a}_1$ の固有値は 0

$$\tilde{X}^{\top}\tilde{X}\boldsymbol{a}_1 = 0$$

- Gram 行列 $\tilde{X}^{\top}\tilde{X}$ の最大固有値は λ_2
- 解は第 2 固有値 λ_2 に対応する固有ベクトル $\boldsymbol{a}_2$

- 以下同様に第 k 主成分負荷量は $X^{\top}X$ の第 k 固有値 λ_k に対応する固有ベクトル $\boldsymbol{a}_k$

解析の事例

データセットについて

- 総務省統計局より取得した都道府県別の社会生活統計指標 (自然環境・経済基盤) の一部
  - 総務省 https://www.e-stat.go.jp/SG1/estat/List.do?bid=000001083999&cycode=0
  - 整理したものを `japan_social.csv` として配布
    * 都道府県名
    * 地方区分
    * 森林面積割合 (%) 2014 年
    * 就業者 1 人当たり農業産出額 (販売農家)（万円) 2014 年
    * 全国総人口に占める人口割合 (%) 2015 年
    * 土地生産性（耕地面積 1 ヘクタール当たり）(万円) 2014 年
    * 商業年間商品販売額［卸売業＋小売業］（事業所当たり）(百万円) 2013 年

社会生活統計指標の分析

- 変数間の関係を見る

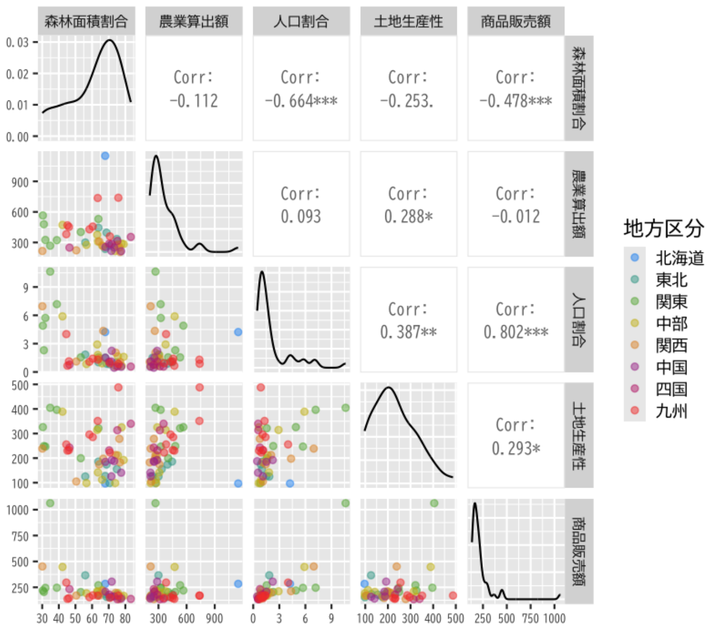

Figure 4: データの散布図

- 変数のばらつきに大きな違いがある

- より詳細な分布を確認する

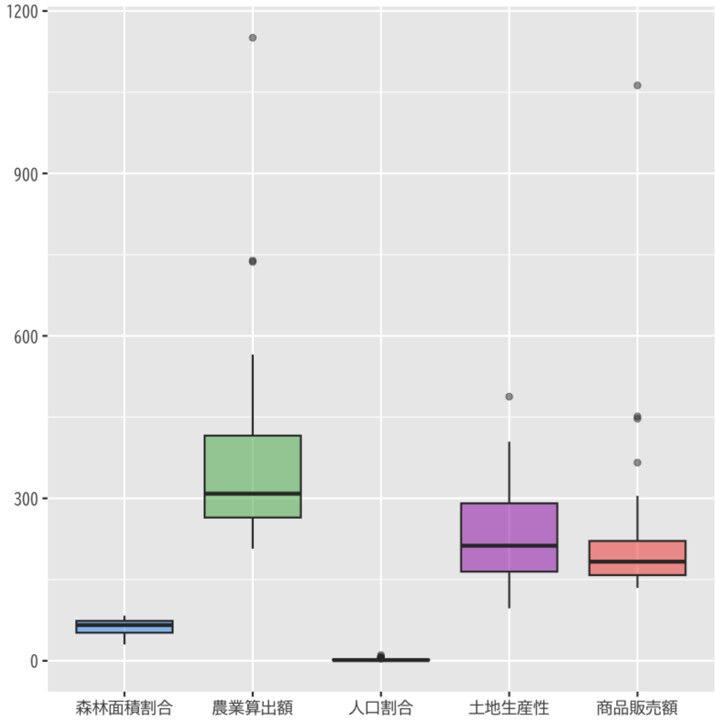

Figure 5: 変数別の箱ひげ図

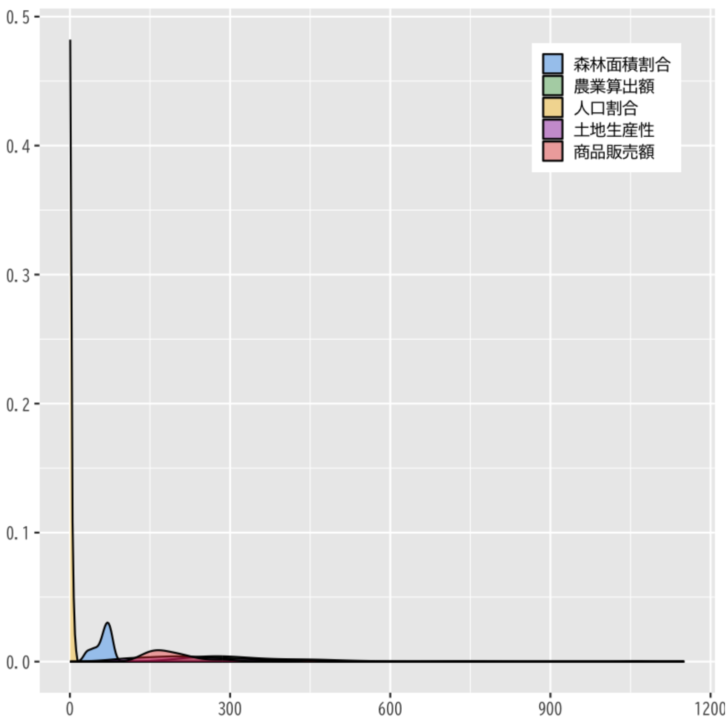

Figure 6: 変数別の確率密度

- データのばらつきを揃える

- 変数のばらつきを確認する

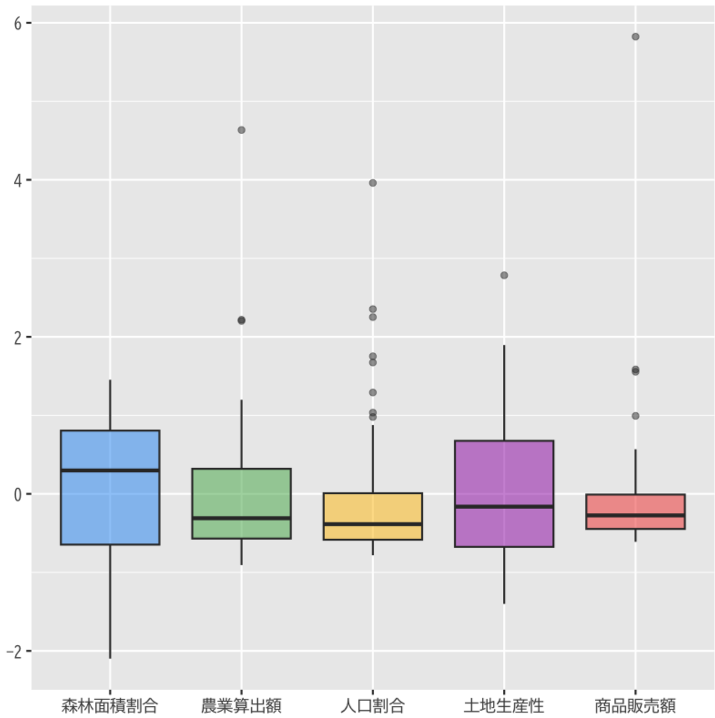

Figure 7: 標準化したデータの箱ひげ図

- より詳細な分布を確認する

- 主成分負荷量を計算する (標準化後)

- 主成分方向から読み取れること
  - 第 1 : 人の多さに関する成分 (正の向きほど人が多い)
  - 第 2 : 農業生産力に関する成分 (正の向きほど高い)

- 各主成分における負荷量の違いを確認する

- 主成分得点を表示する

- データの標準化を行わない場合

- ばらつきの大きな変数に主成分方向が偏る
- 主成分負荷量を比較する

- 主成分得点を比較する

次回の予定

- 第 1 日 : 主成分分析の考え方
- **第 2 日 : 分析の評価と視覚化**

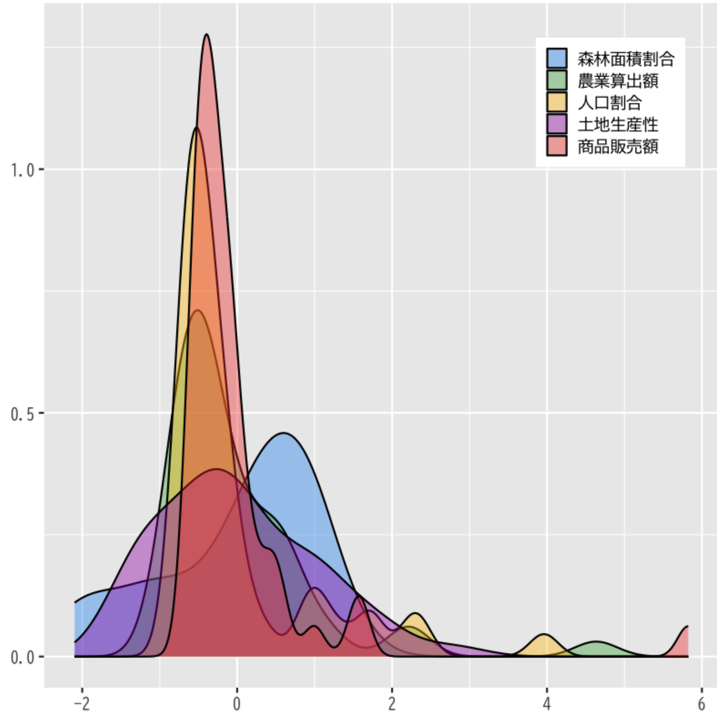

Figure 8: 標準化したデータの確率密度

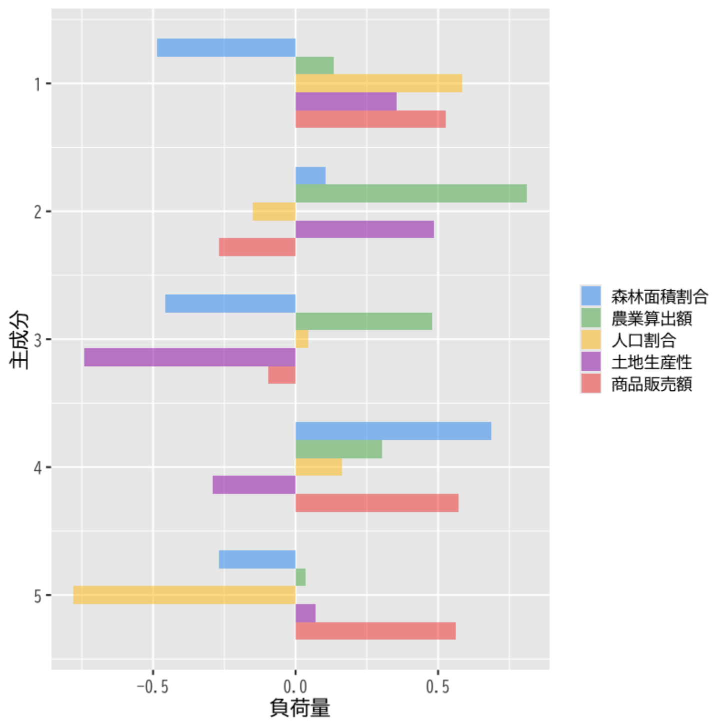

Figure 9: 主成分負荷量の視覚化

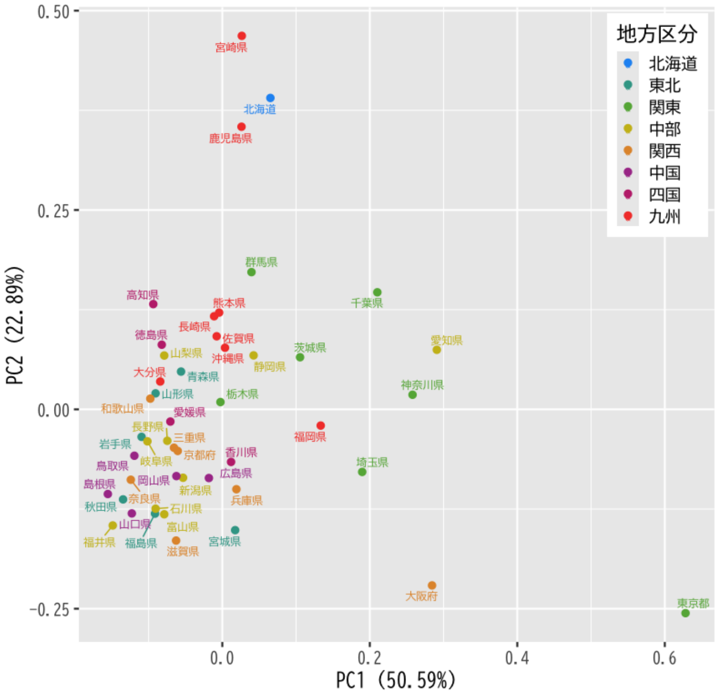

Figure 10: 主成分得点の散布図

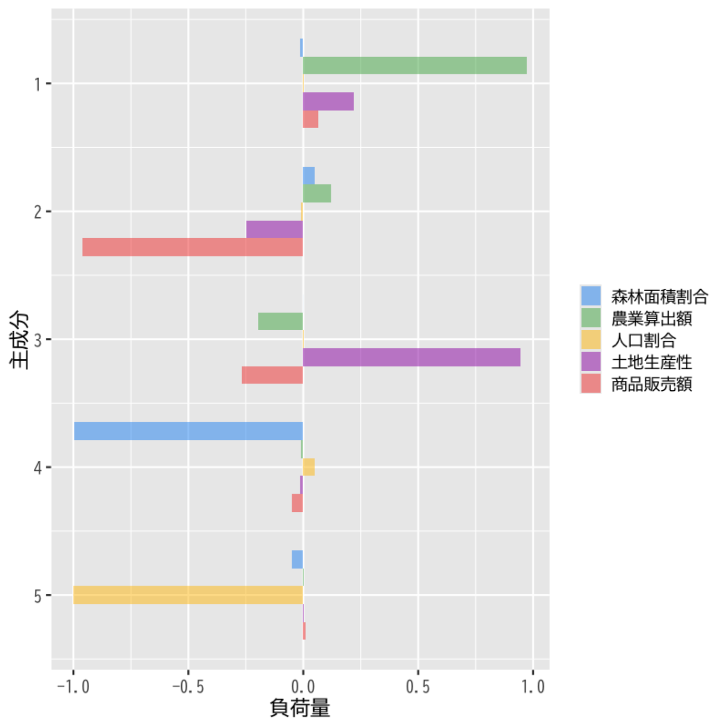

Figure 11: 標準化しない場合の主成分負荷量

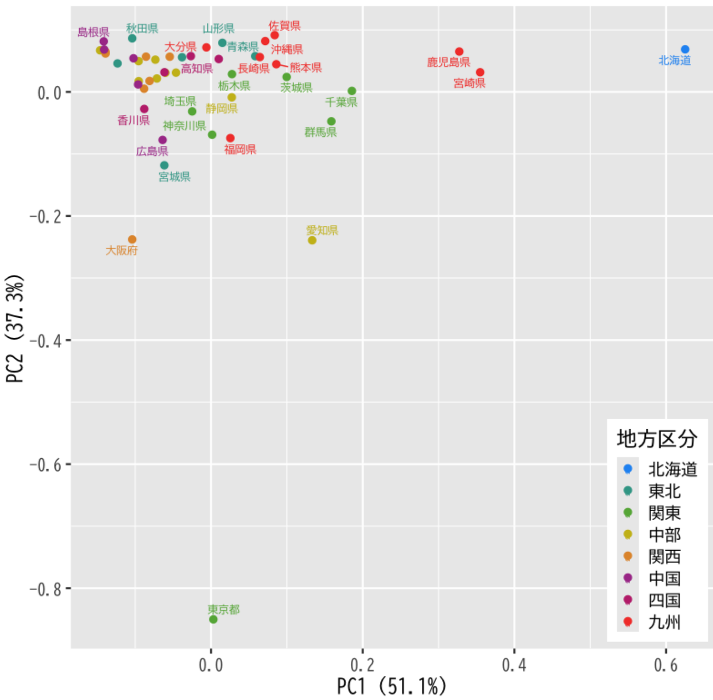

Figure 12: 標準化しない場合の主成分得点の散布図

Table 2: 都道府県別の自然環境・経済基盤データ

都道府県名	地方区分	森林面積割合	農業算出額	人口割合	土地生産性	商品販売額
北海道	北海道	67.9	1150.6	4.23	96.8	283.3
青森県	東北	63.8	444.7	1.03	186.0	183.0
岩手県	東北	74.9	334.3	1.01	155.2	179.4
宮城県	東北	55.9	299.9	1.84	125.3	365.9
秋田県	東北	70.5	268.7	0.81	98.5	153.3
山形県	東北	68.7	396.3	0.88	174.1	157.5
福島県	東北	67.9	236.4	1.51	127.1	184.5
茨城県	関東	31.0	479.0	2.30	249.1	204.9
栃木県	関東	53.2	402.6	1.55	199.6	204.3
群馬県	関東	63.8	530.6	1.55	321.6	270.0
埼玉県	関東	31.9	324.7	5.72	247.0	244.7
千葉県	関東	30.4	565.5	4.90	326.1	219.7
東京都	関東	34.8	268.5	10.63	404.7	1062.6
神奈川県	関東	38.8	322.8	7.18	396.4	246.1
新潟県	中部	63.5	308.6	1.81	141.9	205.5
富山県	中部	56.6	276.1	0.84	98.5	192.4
石川県	中部	66.0	271.3	0.91	112.0	222.9
福井県	中部	73.9	216.1	0.62	98.5	167.3
山梨県	中部	77.8	287.4	0.66	325.3	156.2
長野県	中部	75.5	280.0	1.65	211.3	194.4
岐阜県	中部	79.0	283.7	1.60	192.1	167.9
静岡県	中部	63.1	375.8	2.91	314.5	211.4
愛知県	中部	42.2	472.3	5.89	388.9	446.9
三重県	関西	64.3	310.6	1.43	174.3	170.1
滋賀県	関西	50.5	222.8	1.11	104.9	170.7
京都府	関西	74.2	267.8	2.05	212.5	196.7
大阪府	関西	30.1	216.3	6.96	238.8	451.2
兵庫県	関西	66.7	261.2	4.35	197.7	212.5
奈良県	関西	76.8	207.0	1.07	182.7	147.0
和歌山県	関西	76.4	251.1	0.76	278.4	136.4
鳥取県	中国	73.3	249.9	0.45	187.6	162.2
島根県	中国	77.5	214.1	0.55	140.8	141.1
岡山県	中国	68.0	254.8	1.51	184.9	207.8
広島県	中国	71.8	286.2	2.24	192.2	304.6
山口県	中国	71.6	216.9	1.11	125.8	158.9
徳島県	四国	75.2	315.4	0.59	313.5	134.5
香川県	四国	46.4	249.5	0.77	242.9	232.9
愛媛県	四国	70.3	288.5	1.09	231.6	179.4
高知県	四国	83.3	354.2	0.57	339.9	137.9
福岡県	九州	44.5	381.0	4.01	255.6	295.7
佐賀県	九州	45.2	468.7	0.66	230.3	137.9
長崎県	九州	58.4	428.9	1.08	296.0	154.0
熊本県	九州	60.4	456.6	1.41	285.5	172.5
大分県	九州	70.7	360.1	0.92	222.8	148.3
宮崎県	九州	75.8	739.1	0.87	487.7	170.6
鹿児島県	九州	63.4	736.5	1.30	351.2	169.4
沖縄県	九州	46.1	452.4	1.13	232.8	145.4

Table 3: 標準化したデータ

都道府県名	地方区分	森林面積割合	農業算出額	人口割合	土地生産性	商品販売額
北海道	北海道	0.4250	4.6300	0.9790	-1.4000	0.42100
青森県	東北	0.1510	0.4890	-0.5120	-0.4460	-0.27400
岩手県	東北	0.8920	-0.1590	-0.5210	-0.7760	-0.29900
宮城県	東北	-0.3760	-0.3610	-0.1340	-1.1000	0.99300
秋田県	東北	0.5990	-0.5440	-0.6140	-1.3800	-0.48000
山形県	東北	0.4790	0.2050	-0.5810	-0.5740	-0.45100
福島県	東北	0.4250	-0.7340	-0.2880	-1.0800	-0.26400
茨城県	関東	-2.0400	0.6910	0.0801	0.2290	-0.12300
栃木県	関東	-0.5560	0.2420	-0.2690	-0.3010	-0.12700
群馬県	関東	0.1510	0.9940	-0.2690	1.0100	0.32900
埼玉県	関東	-1.9800	-0.2150	1.6700	0.2070	0.15300
千葉県	関東	-2.0800	1.2000	1.2900	1.0500	-0.02000
東京都	関東	-1.7800	-0.5460	3.9600	1.9000	5.82000
神奈川県	関東	-1.5200	-0.2270	2.3500	1.8100	0.16300
新潟県	中部	0.1310	-0.3100	-0.1480	-0.9180	-0.11800
富山県	中部	-0.3290	-0.5010	-0.6000	-1.3800	-0.20900
石川県	中部	0.2980	-0.5290	-0.5670	-1.2400	0.00214
福井県	中部	0.8260	-0.8530	-0.7030	-1.3800	-0.38300
山梨県	中部	1.0900	-0.4350	-0.6840	1.0500	-0.46000
長野県	中部	0.9330	-0.4780	-0.2230	-0.1750	-0.19500
岐阜県	中部	1.1700	-0.4560	-0.2460	-0.3810	-0.37900
静岡県	中部	0.1050	0.0846	0.3640	0.9300	-0.07760
愛知県	中部	-1.2900	0.6510	1.7500	1.7300	1.56000
三重県	関西	0.1850	-0.2980	-0.3250	-0.5720	-0.36400
滋賀県	関西	-0.7370	-0.8140	-0.4740	-1.3100	-0.36000
京都府	関西	0.8460	-0.5500	-0.0364	-0.1630	-0.17900
大阪府	関西	-2.1000	-0.8520	2.2500	0.1190	1.58000
兵庫県	関西	0.3450	-0.5880	1.0400	-0.3210	-0.07000
奈良県	関西	1.0200	-0.9070	-0.4930	-0.4820	-0.52400
和歌山県	関西	0.9930	-0.6480	-0.6370	0.5430	-0.59800
鳥取県	中国	0.7860	-0.6550	-0.7820	-0.4290	-0.41900
島根県	中国	1.0700	-0.8650	-0.7350	-0.9300	-0.56500
岡山県	中国	0.4320	-0.6260	-0.2880	-0.4580	-0.10300
広島県	中国	0.6860	-0.4420	0.0521	-0.3800	0.56900
山口県	中国	0.6720	-0.8490	-0.4740	-1.0900	-0.44200
徳島県	四国	0.9130	-0.2700	-0.7170	0.9190	-0.61100
香川県	四国	-1.0100	-0.6570	-0.6330	0.1630	0.07150
愛媛県	四国	0.5850	-0.4280	-0.4840	0.0420	-0.29900
高知県	四国	1.4500	-0.0422	-0.7260	1.2000	-0.58700
福岡県	九州	-1.1400	0.1150	0.8770	0.2990	0.50700
佐賀県	九州	-1.0900	0.6300	-0.6840	0.0281	-0.58700
長崎県	九州	-0.2090	0.3960	-0.4880	0.7320	-0.47600
熊本県	九州	-0.0756	0.5590	-0.3350	0.6190	-0.34700
大分県	九州	0.6120	-0.0076	-0.5630	-0.0522	-0.51500
宮崎県	九州	0.9530	2.2200	-0.5860	2.7800	-0.36000
鹿児島県	九州	0.1250	2.2000	-0.3860	1.3200	-0.36900
沖縄県	九州	-1.0300	0.5340	-0.4650	0.0548	-0.53500

Table 4: 標準化した場合の主成分負荷量

column	PC1	PC2	PC3	PC4	PC5
森林面積割合	-0.4871	0.1046	-0.4575	0.6860	-0.2682
農業算出額	0.1339	0.8115	0.4791	0.3045	0.0348
人口割合	0.5851	-0.1511	0.0447	0.1641	-0.7784
土地生産性	0.3548	0.4851	-0.7417	-0.2897	0.0689
商品販売額	0.5258	-0.2689	-0.0952	0.5708	0.5624

Table 5: 標準化しない場合の主成分負荷量

column	PC1	PC2	PC3	PC4	PC5
森林面積割合	-0.0142	0.0482	-0.0004	-0.9975	-0.0495
農業算出額	0.9729	0.1208	-0.1971	-0.0080	0.0004
人口割合	0.0022	-0.0116	0.0000	0.0489	-0.9987
土地生産性	0.2217	-0.2467	0.9433	-0.0155	0.0026
商品販売額	0.0647	-0.9602	-0.2672	-0.0476	0.0090