主成分分析

評価と視覚化

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**

主成分分析の復習

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
  - 変量の間の関係を明らかにする
- 分析の方針
  - **データの情報を保持する = データを区別することができる**
  - データの情報を最大限保持する変量の線形結合を構成
  - データの情報を最大限反映する座標 (方向) を探索

分析の考え方

- 1 変量の特徴量 $\boldsymbol{a}^\top\boldsymbol{x}_1,\dots,\boldsymbol{a}^\top\boldsymbol{x}_n$ を構成
  - 観測データ $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ のもつ情報を最大限保持するベクトル $\boldsymbol{a}$ を **適切に** 選択
  - $\boldsymbol{a}^\top\boldsymbol{x}_1,\dots,\boldsymbol{a}^\top\boldsymbol{x}_n$ の変動 (ばらつき) が最も大きい方向を選択
- **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

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

行列による表現

- 中心化したデータ行列

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

- 評価関数 $f(\boldsymbol{a})$ は行列 $X^\top X$ の二次形式

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

固有値問題

- 最適化問題

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

- 解の条件

  $f(\boldsymbol{a})$ の極大値を与える $\boldsymbol{a}$ は $X^\top X$ の固有ベクトルである

  $$X^\top X\boldsymbol{a} = \lambda\boldsymbol{a}$$

  - 未定係数法を用いている

主成分負荷量と主成分得点

- $\boldsymbol{a}$: **主成分負荷量** (principal component loading)
- $\boldsymbol{a}^\top\boldsymbol{x}_i$: **主成分得点** (principal component score)
- 第 1 主成分負荷量

  $X^\top X$ の第 1(最大) 固有値 λ_1 に対応する固有ベクトル $\boldsymbol{a}_1$
- 第 k 主成分負荷量

  $X^\top X$ の第 k 固有値 λ_k に対応する固有ベクトル $\boldsymbol{a}_k$

演習

問題

- 以下の問に答えなさい
  - ベクトル $\boldsymbol{a}$ を $X^\top X$ の単位固有ベクトルとするとき

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

    の値を求めよ
  - 行列 X を中心化したデータ行列，ベクトル $\boldsymbol{a}_k$ を第 k 主成分負荷量とするとき，第 k 主成分得点の平均まわりの平方和

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

    を X と $\boldsymbol{a}_k$ で表せ

解答例

- 固有値・固有ベクトルの性質を利用する

  $X^\top X$ の固有値・固有ベクトルを $\lambda_k, \boldsymbol{a}_k$ とする．$\boldsymbol{a} = \boldsymbol{a}_k$ とすれば

  $$\begin{aligned} f(\boldsymbol{a}_k) &= \boldsymbol{a}_k^\top X^\top X\boldsymbol{a}_k \\ &= \boldsymbol{a}_k^\top\lambda_k\boldsymbol{a}_k && \text{(固有ベクトル)} \\ &= \lambda_k && \text{(単位ベクトル)} \end{aligned}$$

  - 定義に従い計算すればよい (前回の復習)

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

寄与率

寄与率の考え方

- 回帰分析で考察した寄与率の一般形

$$
(\text{寄与率}) = \frac{(\text{その方法で説明できる変動})}{(\text{データ全体の変動})}
$$

- 主成分分析での定義 (proportion of variance)

$$
(\text{寄与率}) = \frac{(\text{主成分の変動})}{(\text{全体の変動})}
$$

Gram 行列のスペクトル分解

- 行列 $X^\top X$ (半正定値行列) のスペクトル分解

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

  - 固有値と固有ベクトルによる行列の表現
- 主成分の変動の評価

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

  - 固有ベクトル (単位ベクトル) の直交性を利用

寄与率の計算

- 主成分と全体の変動

$$
\begin{aligned}
(\text{主成分の変動}) &= \sum_{i=1}^{n} (\boldsymbol{a}_k^\top \boldsymbol{x}_i - \boldsymbol{a}_k^\top \bar{\boldsymbol{x}})^2 = \boldsymbol{a}_k^\top X^\top X \boldsymbol{a}_k = \lambda_k \\
(\text{全体の変動}) &= \sum_{i=1}^{n} \|\boldsymbol{x}_i - \bar{\boldsymbol{x}}\|^2 = \sum_{l=1}^{p} \boldsymbol{a}_l^\top X^\top X \boldsymbol{a}_l = \sum_{l=1}^{p} \lambda_l
\end{aligned}
$$

- 固有値による寄与率の表現

$$
(\text{寄与率}) = \frac{\lambda_k}{\sum_{l=1}^{p} \lambda_l}
$$

累積寄与率

- **累積寄与率** (cumulative proportion) :

  第 k 主成分までの変動の累計

$$(\text{累積寄与率}) = \frac{\sum_{l=1}^{k} \lambda_l}{\sum_{l=1}^{p} \lambda_l}$$

  - 累積寄与率はいくつの主成分を用いるべきかの基準
  - 一般に累積寄与率が 80%程度までの主成分を用いる

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

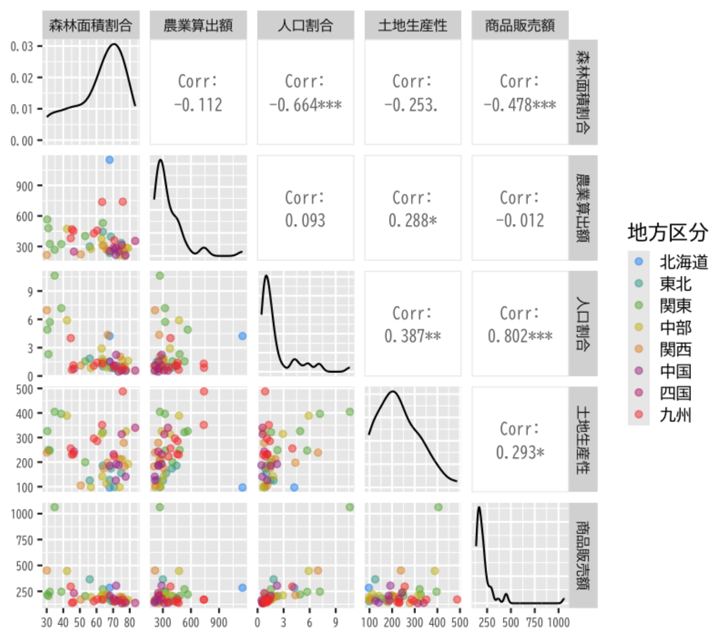


Figure 1: データの散布図

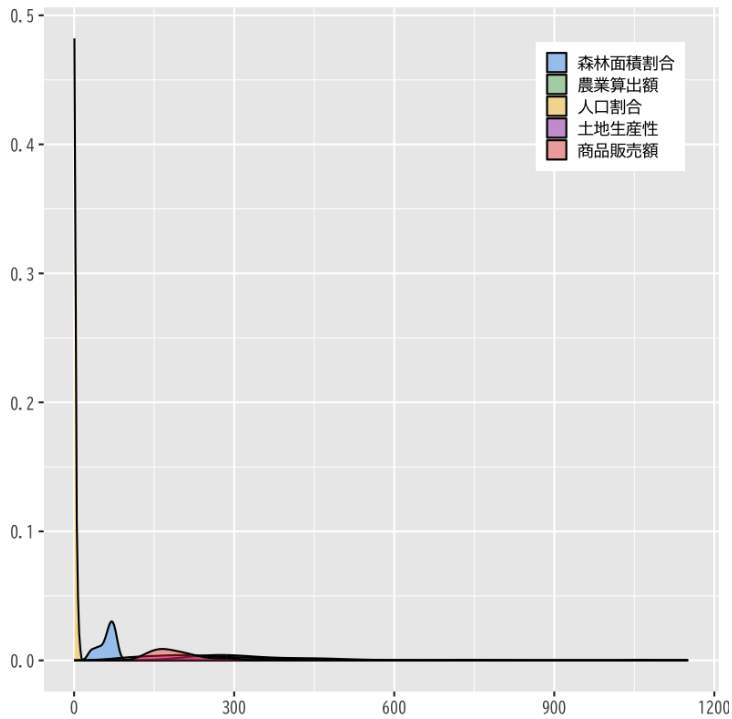


Figure 2: 変数別の確率密度

- 変数のばらつきに大きな違いがある

- 標準化なしで分析した場合
  - 第 1 : 分散が大きい農業算出額が支配的
  - 第 2 : 次に分散が大きな商品販売額が支配的

- 寄与率を確認する

- 寄与率を視覚化する

- 第 1,2 主成分得点の表示

- 第 3,2 主成分得点の表示

- データのばらつきを揃える

- 変数のばらつきを確認する

- 標準化して分析した場合
  - 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)

- 寄与率を確認する

- 寄与率を視覚化する

- 第 1,2 主成分得点の表示

- 第 3,2 主成分得点の表示

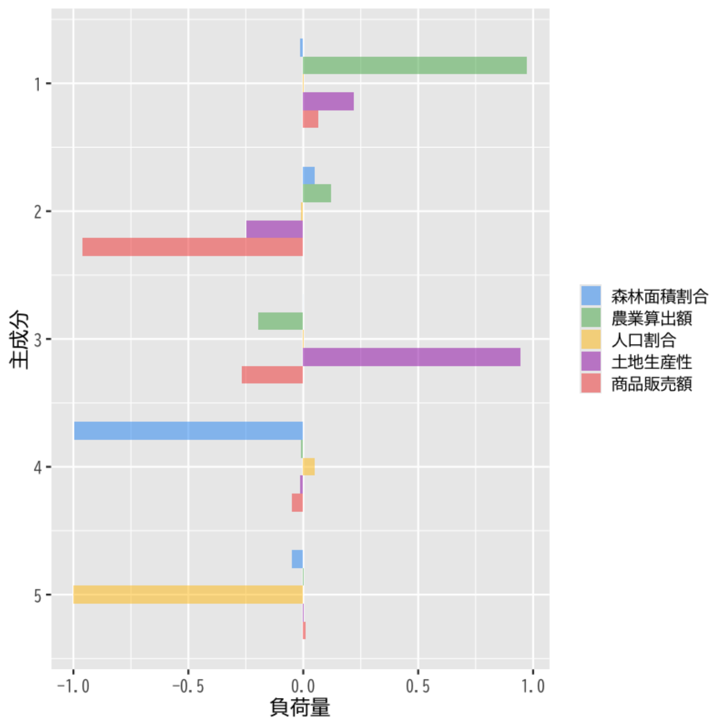


Figure 3: 標準化しない場合の主成分負荷量

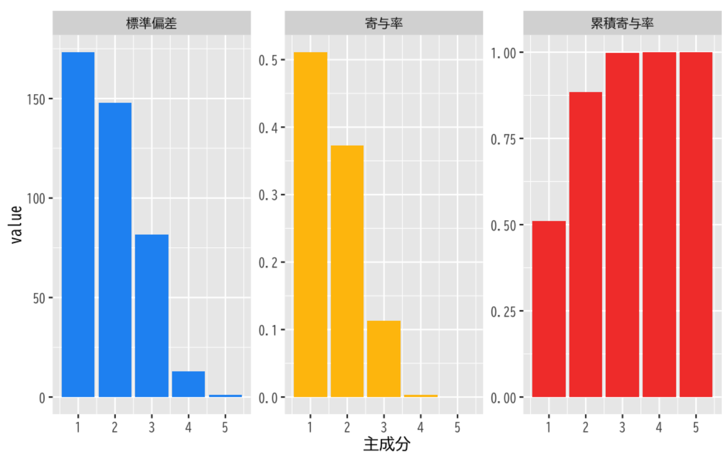


Figure 4: 寄与率と累積寄与率

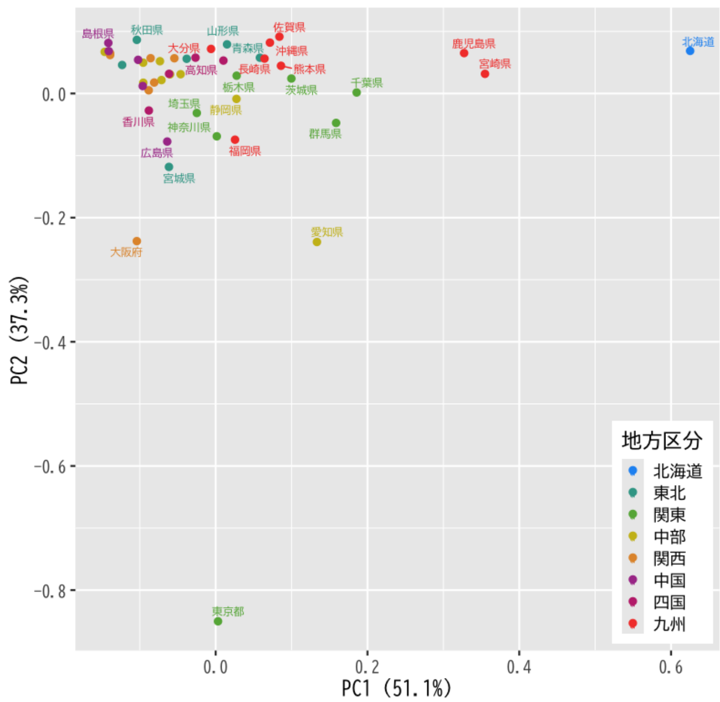


Figure 5: 主成分得点の散布図

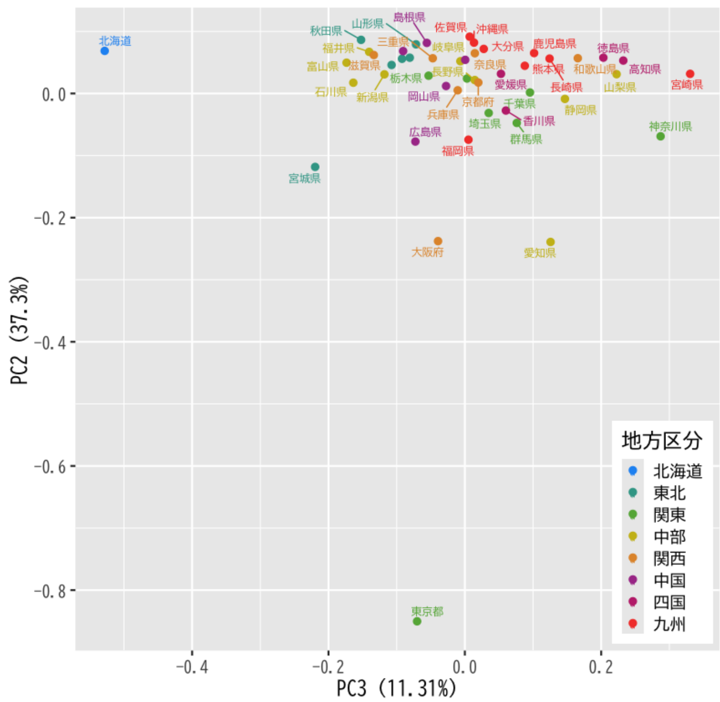


Figure 6: 主成分得点の散布図

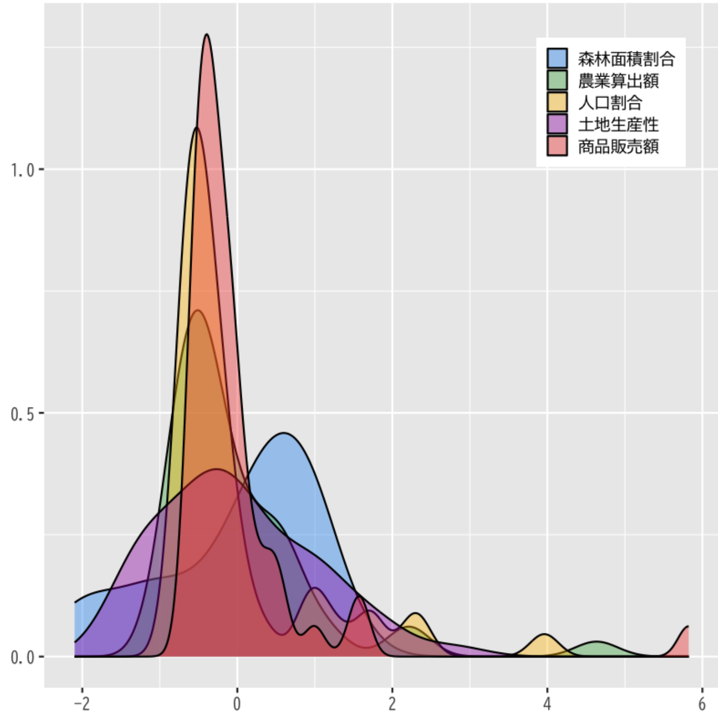


Figure 7: 標準化したデータの確率密度

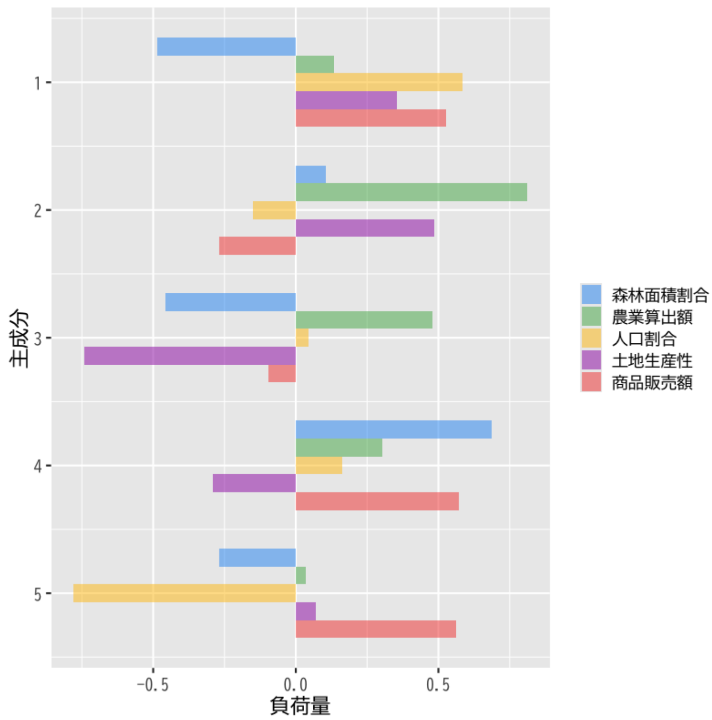


Figure 8: 標準化しない場合の主成分負荷量

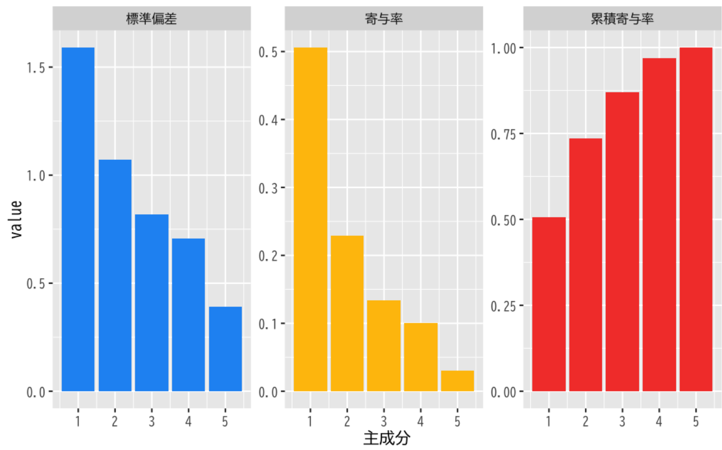


Figure 9: 寄与率と累積寄与率

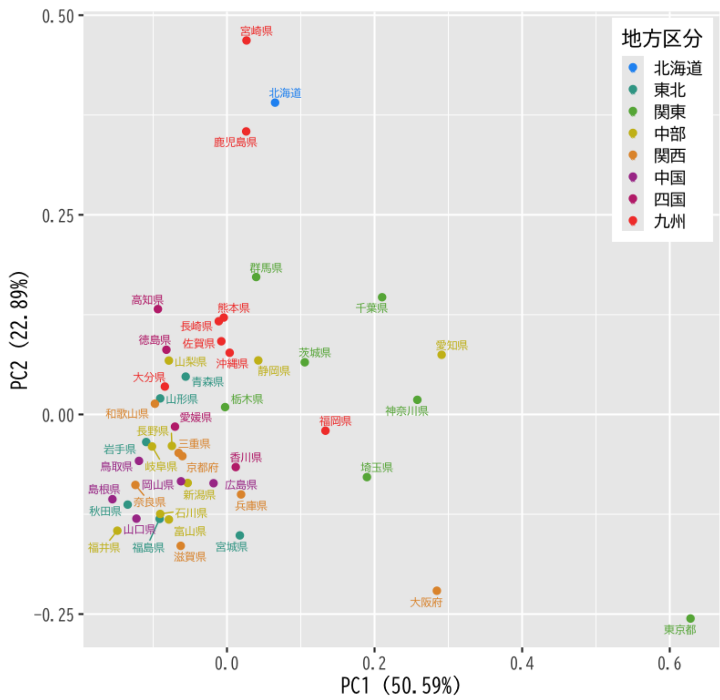


Figure 10: 主成分得点の散布図

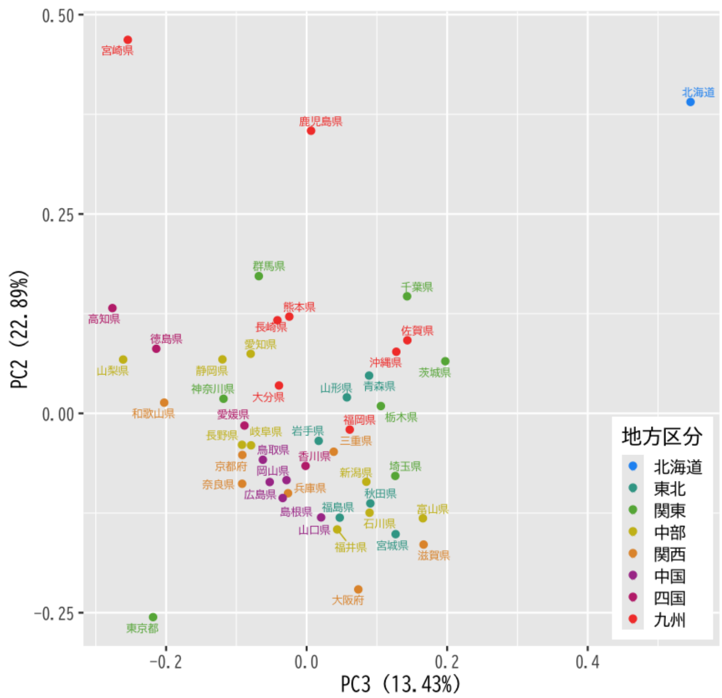


Figure 11: 主成分得点の散布図

演習

問題

- 以下の問に答えなさい
  - 標準化条件を満たす線形変換 $x'_{ij} = a_j(x_{ij} - b_j)$ を求めよ

$$\frac{1}{n}\sum_{i=1}^{n} x'_{ij} = 0, \quad \frac{1}{n-1}\sum_{i=1}^{n}(x'_{ij})^2 = 1$$

  - 標準化されたデータ行列を

$$X' = \begin{pmatrix} \boldsymbol{x}_1'^{\top} \\ \vdots \\ \boldsymbol{x}_n'^{\top} \end{pmatrix} = \begin{pmatrix} x'_{11} & \cdots & x'_{1p} \\ \vdots & & \vdots \\ x'_{n1} & \cdots & x'_{np} \end{pmatrix}$$

  と書くとき，$X'^{\top}X'$ の対角成分を求めよ

解答例

- 標本平均の定義どおりに計算すればよい

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} x'_{ij} &= \frac{1}{n}\sum_{i=1}^{n}\left(a_j(x_{ij} - b_j)\right) \\ &= a_j\left(\frac{1}{n}\sum_{i=1}^{n} x_{ij} - b_j\right) \\ &= 0 \end{aligned}$$

したがって

$$b_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} = \bar{x}_j \quad \text{(元の変数の標本平均)}$$

- 不偏分散も同様に計算すればよい

$$\begin{aligned}\frac{1}{n-1}\sum_{i=1}^{n}(x'_{ij})^2 &= a_j^2 \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2 \\ &= 1\end{aligned}$$

したがって

$$a_j = \left(\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2\right)^{-1/2} \quad \text{(標準偏差の逆数)}$$

- 不偏分散での標準化であることに注意する

$$(X'^{\top}X')_{jj} = \sum_{i=1}^{n}(x'_{ij})^2 = n-1$$

主成分負荷量

主成分負荷量と主成分得点

- 負荷量 (得点係数) の大きさ: 変数の貢献度
- **問題点**
  - 変数のスケールによって係数の大きさは変化する
  - 変数の標準化 (平均 0，分散 1) がいつも妥当とは限らない
- スケールによらない変数と主成分の関係
  - **相関係数** を考えればよい

相関係数

- $\boldsymbol{e}_j$: 第 j 成分は 1，それ以外は 0 のベクトル
- $X\boldsymbol{e}_j$: 第 j 変数ベクトル
- $X\boldsymbol{a}_k$: 第 k 主成分得点ベクトル
- 主成分と変数の相関係数:

$$\begin{aligned}\mathrm{Cor}(X\boldsymbol{a}_k, X\boldsymbol{e}_j) &= \frac{\boldsymbol{a}_k^{\top}X^{\top}X\boldsymbol{e}_j}{\sqrt{\boldsymbol{a}_k^{\top}X^{\top}X\boldsymbol{a}_k}\sqrt{\boldsymbol{e}_j^{\top}X^{\top}X\boldsymbol{e}_j}} \\ &= \frac{\lambda_k\boldsymbol{a}_k^{\top}\boldsymbol{e}_j}{\sqrt{\lambda_k}\sqrt{(X^{\top}X)_{jj}}} = \frac{\sqrt{\lambda_k}(\boldsymbol{a}_k)_j}{\sqrt{(X^{\top}X)_{jj}}}\end{aligned}$$

相関係数による評価

- 標準化されたデータの場合
  - $X^\top X$ の対角成分は全て $n-1$ ($(X^\top X)_{jj} = n-1$)
  - 第 k 主成分に対する相関係数ベクトル

$$\boldsymbol{r}_k = \sqrt{\lambda_k/(n-1)} \cdot \boldsymbol{a}_k, \quad (\boldsymbol{r}_k)_j = \sqrt{\lambda_k/(n-1)} \cdot (\boldsymbol{a}_k)_j$$

  - **主成分負荷量の比較**
    - 同じ主成分 (k を固定) への各変数の影響は固有ベクトルの成分比
    - 同じ変数 (j を固定) の各主成分への影響は固有値の平方根で重みづけ
- 標準化されていない場合
  - 変数の分散の影響を考慮する必要がある

データ行列の分解表現

特異値分解

- 階数 r の $n \times p$ 型行列 X の分解

$$X = U\Sigma V^\top$$

  - U は $n \times n$ 型直交行列, V は $p \times p$ 型直交行列
  - Σ は $n \times p$ 型行列

$$\Sigma = \begin{pmatrix} D & O_{r,p-r} \\ O_{n-r,r} & O_{n-r,m-r} \end{pmatrix}$$

    - $O_{s,t}$ は $s \times t$ 型零行列
    - D は $\sigma_1 \geq \sigma_2 \geq \sigma_r > 0$ を対角成分とする $r \times r$ 型対角行列

特異値

- 行列 Σ の成分表示

$$\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & O_{r,p-r} \\ & & \sigma_r & \\ & & & \\ & O_{n-r,r} & & O_{n-r,m-r} \end{pmatrix}$$

- D の対角成分: X の **特異値** (singular value)

特異値分解による Gram 行列の表現

- Gram 行列の展開

$$\begin{aligned} X^\top X &= (U\Sigma V^\top)^\top (U\Sigma V^\top) \\ &= V\Sigma^\top U^\top U\Sigma V^\top \\ &= V\Sigma^\top \Sigma V^\top \end{aligned}$$

- 行列 $\Sigma^\top\Sigma$ は対角行列

$$\Sigma^\top\Sigma = \begin{pmatrix} \sigma_1^2 & & & & & \\ & \ddots & & & & \\ & & \sigma_r^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

演習

問題

- 行列 X の特異値分解を $U\Sigma V^\top$ とし，行列 U の第 k 列ベクトルを $\boldsymbol{u}_k$，行列 V の第 k 列ベクトルを $\boldsymbol{v}_k$ とするとき，以下の問に答えなさい
  - 行列 U, V の列ベクトルを用いて X を展開しなさい
  - Gram 行列 $X^\top X$ の固有値を特異値で表しなさい
  - 行列 X の主成分負荷量を求めなさい
  - それぞれの負荷量に対応する主成分得点を求めなさい

解答例

- Σ が対角成分しか持たないことに注意すると以下のように展開される

$$X = U\Sigma V^\top = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^\top$$

- 先週の演習問題と特異値分解を比較する

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

  より

$$\lambda_k = \begin{cases} \sigma_k^2, & k \leq r \\ 0, & k > r \end{cases}$$

- 転置に気をつけて同様に比較すればよい

$$A = \begin{pmatrix} \boldsymbol{a}_1^\top \\ \vdots \\ \boldsymbol{a}_p^\top \end{pmatrix}$$

  と定義されているので主成分負荷量 (固有ベクトル) は行列 V の列ベクトル

$$\boldsymbol{a}_k = \boldsymbol{v}_k$$

- 主成分得点の定義どおり計算する

$$X\boldsymbol{a}_k = U\Sigma V^\top \boldsymbol{v}_k = \sigma_k \boldsymbol{u}_k$$

  - ただし $k > r$ のとき $\sigma_k = 0$ とする
  - V と U は大きさが異なるので注意する

バイプロット

データ行列の分解

- 行列 U の第 k 列ベクトル $\boldsymbol{u}_k$
- 行列 V の第 k 列ベクトル $\boldsymbol{v}_k$
- データ行列の特異値分解: (Σ の非零値に注意)

$$X = U\Sigma V^{\top} = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\top}$$

データ行列の近似表現

- 第 k 主成分と第 l 主成分を用いた行列 X の近似 X'

$$X \simeq X' = \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^{\top} + \sigma_l \boldsymbol{u}_l \boldsymbol{v}_l^{\top}$$

- 行列の積による表現

$$\begin{aligned} X' =& GH^{\top}, (0 \le s \le 1) \\ & G = \begin{pmatrix} \sigma_k^{1-s}\boldsymbol{u}_k & \sigma_l^{1-s}\boldsymbol{u}_l \end{pmatrix}, \quad H = \begin{pmatrix} \sigma_k^{s}\boldsymbol{v}_k & \sigma_l^{s}\boldsymbol{v}_l \end{pmatrix} \end{aligned}$$

バイプロット

- 関連がある 2 枚の散布図を 1 つの画面に表示する図を一般に**バイプロット** (biplot) と呼ぶ
- 行列 G, H の各行を 2 次元座標と見なす

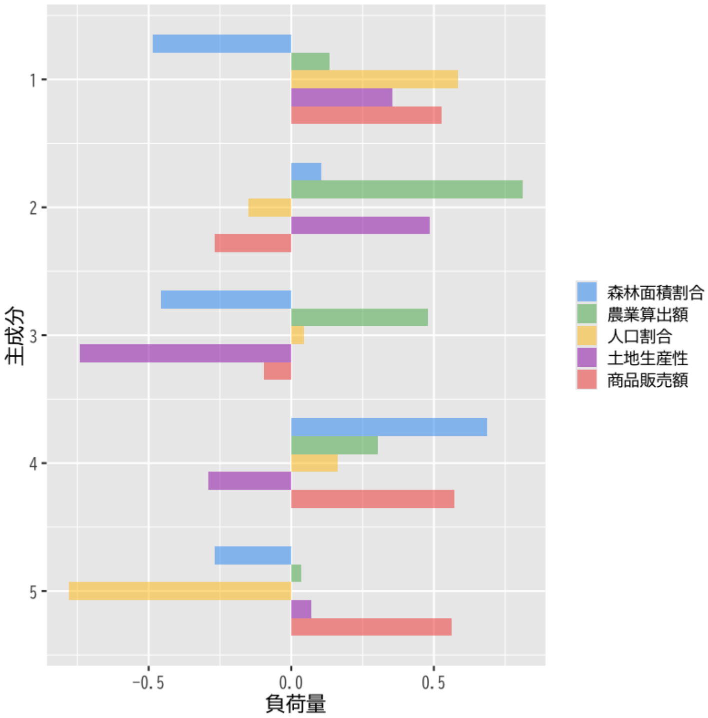
$$X' = GH^{\top}$$

  - 行列 G の各行は各データの 2 次元座標
  - 行列 H の各行は各変量の 2 次元座標
- パラメタ s は 0, 1 または 1/2 が主に用いられる
- X の変動を最大限保持する近似は $k = 1, l = 2$

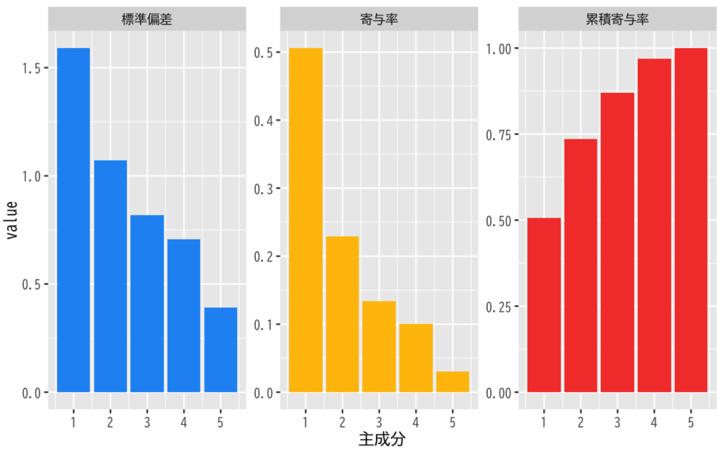
解析の事例

バイプロット

- 主成分負荷量 (標準化あり)



- 寄与率



- 第 1,2 主成分によるバイプロット

- 第 3,2 主成分によるバイプロット

- 中心部の拡大

次回の予定

- **第 1 日 : 判別分析の考え方**
- 第 2 日 : 分析の評価

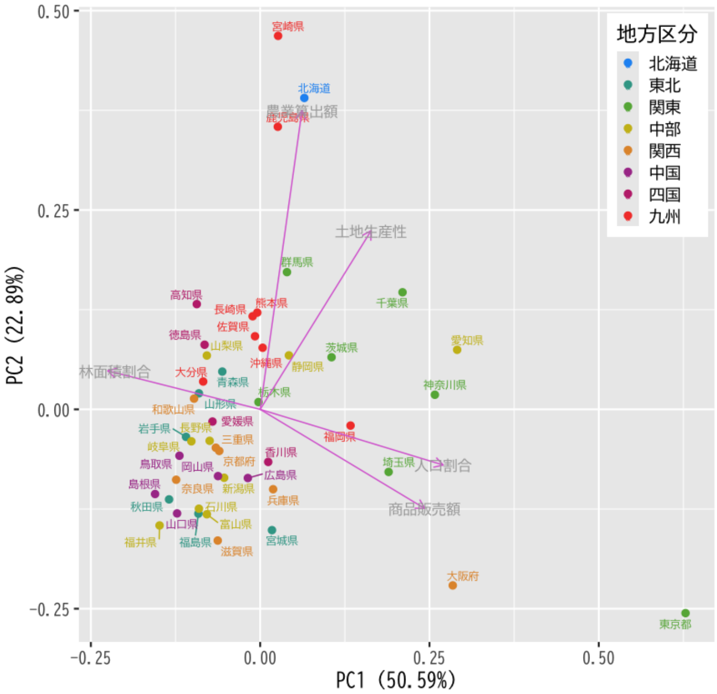


Figure 12: バイプロット (第主成分得点と負荷量の同時散布図)

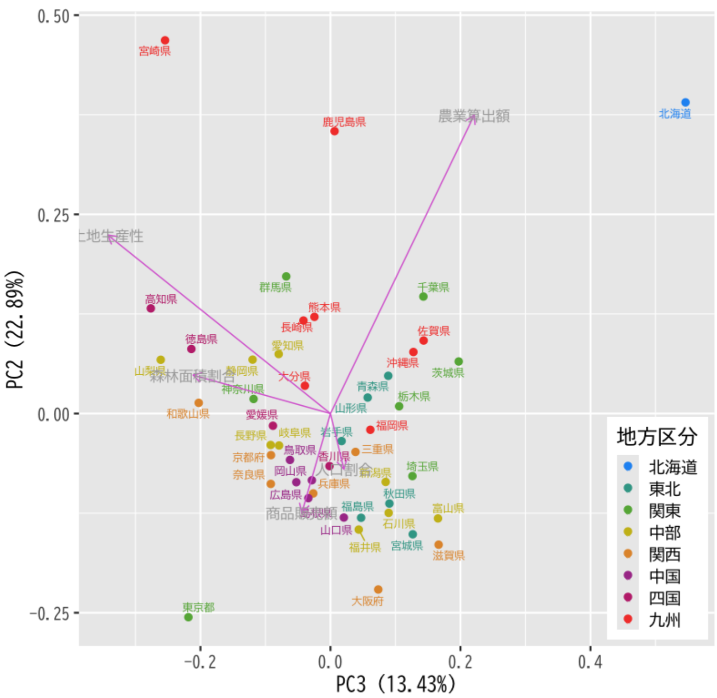


Figure 13: 第 3 主成分 (横軸) と第 2 主成分 (縦軸) のバイプロット

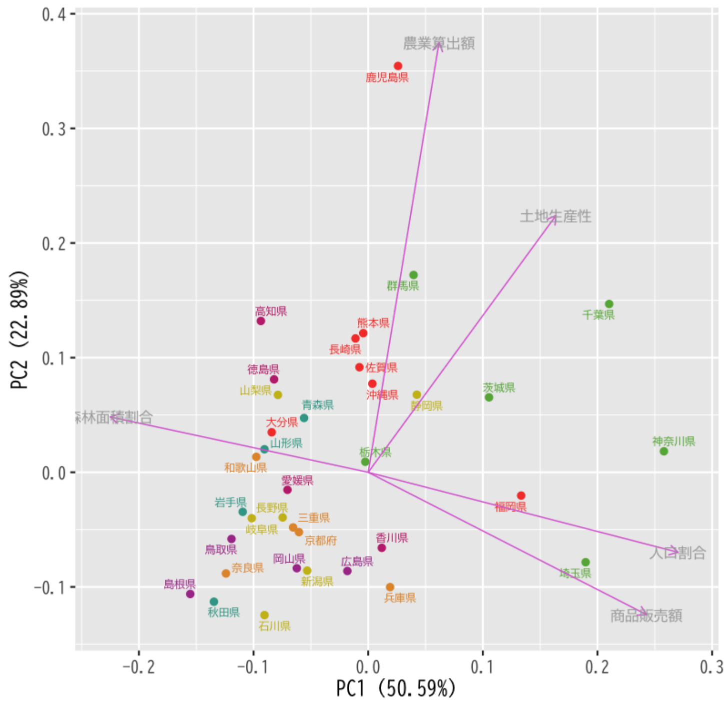


Figure 14: 第 1,2 主成分によるバイプロットの拡大図

Table 1: 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

主成分	標準偏差	寄与率	累積寄与率
1	173.275	0.511	0.511
2	148.037	0.373	0.884
3	81.523	0.113	0.997
4	12.972	0.003	1.000
5	1.052	0.000	1.000

Table 2: 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 3: 寄与率

主成分	標準偏差	寄与率	累積寄与率
1	1.590	0.506	0.506
2	1.070	0.229	0.735
3	0.820	0.134	0.869
4	0.708	0.100	0.969
5	0.392	0.031	1.000

Table 4: 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: 寄与率 (再掲)

主成分	標準偏差	寄与率	累積寄与率
1	1.590	0.506	0.506
2	1.070	0.229	0.735
3	0.820	0.134	0.869
4	0.708	0.100	0.969
5	0.392	0.031	1.000