

数量化分析

1. 数量化 1 類の分析ができる
2. 数量化 2 類ができる
3. 数量化 4 類の分析ができる
4. コレスポンデンス分析ができる

【1】数量化 1 類は重回帰分析

数量化 1 類は重回帰分析です。

数量化〇〇とカテゴライズするからかえってわかりにくい！

数量化 1 類と重回帰分析の違いは説明変数が計数値か計量値かどうかの違い本質は同じ

重回帰分析を復習してから、説明変数を計量値から計数値に変えた場合の重回帰分析をします。それが数量化 1 類の分析です。

【2】重回帰分析の解き方（復習）

(1) 重回帰分析の復習ができる関連記事

【関連記事】【まとめ】重回帰分析がよくわかる

<https://qcplanets.com/method/multi-regression/mr-summary/>

(2) 重回帰分析の復習ポイント

重回帰分析から数量化 1 類へ変化していく際に、比較するために必要な変数を復習しましょう。

1. 平方和 S
2. 分散分析
3. 回帰直線(切片と傾き)
4. 寄与率 S_R

この記事では説明変数は 2 つとし、回帰直線 $y = a + bx_1 + cx_2$ を考えます。

●平方和は

- ・ $S_y = \sum_{i=1}^n (y_i - \bar{y})^2$
- ・ $S_{11} = \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2$
- ・ $S_{22} = \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2$
- ・ $S_{12} = \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)$
- ・ $S_{1y} = \sum_{i=1}^n (x_{1i} - \bar{x}_1)(y_i - \bar{y})$
- ・ $S_{2y} = \sum_{i=1}^n (x_{2i} - \bar{x}_2)(y_i - \bar{y})$

●回帰直線(切片と傾き)は

◎傾き

- ・ $S_{11}b + S_{12}c = S_{1y}$
- ・ $S_{12}b + S_{22}c = S_{2y}$

◎切片

$$\bar{y} = a + b\bar{x}_1 + c\bar{x}_2$$

●寄与率 S_R は

$$S_R = bS_{1y} + cS_{2y}$$

【3】数量化 1 類と重回帰分析を比較

(1) データを用意

数量化 1 類と重回帰分析を比較するために、次の 3 つのデータを用意します。

①データ(type1)

No	x1	x2	y
1	3	1	3
2	2	4	4
3	4	2	4
4	4	5	7
5	5	4	7
6	6	2	5
合計	24	18	30
平均	4	3	5

次に、説明変数 x_1, x_2 において、

●0～3⇒0

●4～6⇒1

という基準を設けてダミー変数化して重回帰分析します。

データ(type2)は、説明変数 x_1 のみ

データ(type3)は、説明変数 x_1, x_2 両方
とします。

②データ(type2)

No	x1	x2	y
1	0	1	3
2	0	4	4
3	1	2	4
4	1	5	7
5	1	4	7
6	1	2	5
合計	4	18	30
平均	0.67	3	5

③データ(type3)

No	x1	x2	y
1	0	0	3
2	0	1	4
3	1	0	4
4	1	1	7
5	1	1	7
6	1	0	5
合計	4	3	30
平均	0.67	0.5	5

(2) 分析結果を比較

平方和、分散分析、回帰直線、寄与率を比較しますが、

数量化1類も重回帰分析も同じ解き方です。

(3) 平方和

平方和	データ (type1)	データ (type2)	データ (type3)
S_{11}	10	68	68
S_{1y}	6	3	3
S_{12}	-1	-1	50
S_{22}	12	12	39
S_{2y}	10	10	3
S_{yy}	14	14	14

上表のマーカ一部ですが、説明変数 x_i の値が変化したところの平方和が変化していますね。

(4) 分散分析

-	データ(type1)		データ(type2)		データ(type3)	
	平方和S	自由度φ	平方和S	自由度φ	平方和S	自由度φ
回帰R	13.042	2	12.089	2	12.75	2
残差e	0.958	3	1.911	3	1.25	3
計T	14	5	14	5	14	5

データ type1 から 3 にかけて、回帰平方和に若干の差が出ていますが、総平方和は不変であることがわかりますね。

(5) 回帰直線

-	定数項	x_1 の係数	x_2 の係数	式
データ (type1)	-0.429	0.689	0.891	$y = -0.429 + 0.689x_1 + 0.891x_2$
データ (type2)	1.778	1.733	0.689	$y = 1.778 + 1.733x_1 + 0.689x_2$
データ (type3)	2.5	2.25	2	$y = 2.5 + 2.25x_1 + 2x_2$

それぞれのケースで若干値が変わっていますね。

(6) 寄与率

-	寄与率R
データ (type1)	0.932
データ (type2)	0.863
データ (type3)	0.911

数量化 1 類の分析ができましたね!

以上、「数量化 1 類の分析ができる」を解説しました。

数量化 2 類ができる

【1】数量化 2 類は判別分析である

数量化 2 類は判別分析です！

ただし、説明変数が計数値（質的データ）である点だけが判別分析と異なります。

判別分析は 2 つ解析方法があります。

1. 線形判別関数による判別
2. マハラビノス距離による判別

https://qcplanets.com/wp-content/uploads/2025/11/qc_planets_discriminant_analysis_text_3ed0etU84.pdf
で復習しましょう。

【2】データ事例

以下のデータを使って、線形判別関数とマハラビノス距離の両手法で判別分析をします。

判別分析の場合は下左表のとおりですが、今回は数量化 2 類なので、説明変数 x_1, x_2 を黄色マーカーのようにあえて変えてみます。

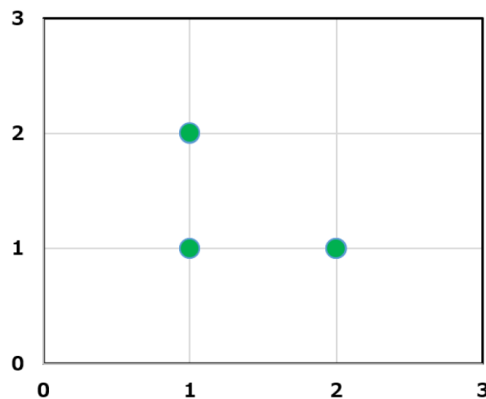
<判別分析>

群	No	x_1	x_2	平均 $\bar{x}_1$	平均 $\bar{x}_2$
1	1	4	8	4	10
	2	6	10		
	3	2	12		
2	4	10	16	7.5	13.5
	5	5	10		
	6	8	12		
	7	7	16		
合計	42	84	全平均	6	12

<数量化 2 類>

群	No	x_1	x_2	平均 $\bar{x}_1$	平均 $\bar{x}_2$
1	1	1	1	1.333	1.333
	2	2	1		
	3	1	2		
2	4	2	1	1.5	1.25
	5	1	1		
	6	2	1		
	7	1	2		
合計	10	9	全平均	1.429	1.286

グラフは下図のとおりです。



【3】線形判別関数で判別分析

線形判別関数から判別分析を、関連記事で解いています。途中経過は関連記事でご確認ください。

【関連記事】線形判別関数が計算できる(2次元)

https://qcplanets.com/method/multivariate/ldf_cal2-1/

(1) 変動を計算

実際に、全変動 S_T と群間変動 S_B を計算して、ラグランジュの未定乗数法を使って固有方程式を作ります。解き方の流れは関連記事に書いています。変動を求めるための表を作ります。

群	NO	x	y	Z	群平均
1	1	1	1	$1a_1+1a_2$	$1.33a_1+1.33a_2$
	2	2	1	$2a_1+1a_2$	-
	3	1	2	$1a_1+2a_2$	-
2	4	2	1	$2a_1+1a_2$	$1.5a_1+1.25a_2$
	5	1	1	$1a_1+1a_2$	-
	6	2	1	$2a_1+1a_2$	-
	7	1	2	$1a_1+2a_2$	-
-	-	-	-	全体の平均	$1.429a_1+1.286a_2$

$$\begin{aligned}
& \text{全変動 } S_T \\
&= \sum_{i=1}^7 (Z - \bar{Z})^2 \\
&= ((1a_1 + 1a_2) - (1.429a_1 + 1.286a_2))^2 + \dots \\
&= 1.714a_1^2 - 1.714a_1a_2 + 1.429a_2^2
\end{aligned}$$

$$\begin{aligned}
& \text{群間変動 } S_B \\
&= \sum_{i=1}^3 (\bar{Z}_1 - \bar{Z})^2 + \sum_{i=1}^4 (\bar{Z}_2 - \bar{Z})^2 \\
&= 0.0486a_1^2 - 0.0476a_1a_2 + 0.011a_2^2
\end{aligned}$$

(2) ラグランジュの未定乗数法から固有方程式を作る

ラグランジュの未定乗数法から固有方程式を作ります。ラグランジュの未定乗数法から

関数 F を

$$\begin{aligned}
F(a_1, a_2) &= 1.714a_1^2 - 1.714a_1a_2 + 1.429a_2^2 - \\
&\lambda(0.0486a_1^2 - 0.0476a_1a_2 + 0.011a_2^2)
\end{aligned}$$

を定義して、偏微分=0 の式を立てます。その後、固有方程式と流れますね。

係数が小数で複雑なので、文字式で簡単に表記します。

$$\begin{aligned}
\bullet \frac{\partial F(a_1, a_2)}{\partial a_1} &= sa_1 + ta_2 - \lambda(ua_1 + va_2) = 0 \\
\bullet \frac{\partial F(a_1, a_2)}{\partial a_2} &= (ea_1 + fa_2) - \lambda(ga_1 + ha_2) = 0
\end{aligned}$$

行列表記しましょう。

$$\begin{pmatrix} s - u\lambda & t - v\lambda \\ e - g\lambda & f - h\lambda \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

計算すると $\lambda(6.857\lambda - 0.190) = 0$

よって、固有値 λ は $\lambda = 0, 0.027$ となります。

(3) 固有ベクトルを算出

固有ベクトルから (a, b) を計算します。固有値 0 のときは、固有ベクトルの成分がすべて 0 になるので、固有値が 0.027 についてのみ計算します。

● $\lambda = 0.027$ のとき、

$$\begin{pmatrix} 0.09542 & -0.04762 \\ -0.04762 & 0.02381 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

より、

● $2a = b$

となります。

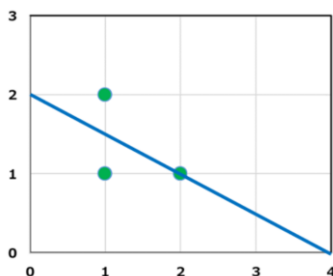
(4) 線形判別関数を作る

係数比は $2a = b$ であり、

7つのデータの平均を通る関数を作ればよいので、

$$Z = x_1 + 2x_2 + 4$$

となります。グラフを描くとわかりやすいですね。



【4】マハラビノス距離で判別分析

(1) マハラビノス距離の求め方(復習)

マハラビノス距離については関連記事で導出方法や具体的な算出事例を紹介しています。ご確認ください。

【関連記事】マハラビノス距離が計算できる

https://qcplanets.com/method/multivariate/mahalanobis-distance_cal/

マハラビノス距離（2次元）の場合、関連記事から

$$D_M^2 = (x_1 - \bar{x}_1, x_2 - \bar{x}_2) \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}^{-1} \begin{pmatrix} x_1 - \bar{x}_1 \\ x_2 - \bar{x}_2 \end{pmatrix}$$

と書けます。

実際は

● $S_{11} \Rightarrow \sigma_x^2$

● $S_{22} \Rightarrow \sigma_y^2$

● $S_{12} \Rightarrow \sigma_{xy}^2$

から計算します。

必要な数値をデータ表から計算すると、下表にまとめられます。

値	1群	2群
σ_x^2	0.333	0.333
σ_y^2	0.333	0.25
σ_{xy}^2	-0.167	-0.167
$\bar{x}$	1.333	1.5
$\bar{y}$	1.333	1.25
a	4	4.5
b	4	6
c	4	6

ここで、マハラビノス距離を展開すると楕円の方程式になるので、

$$a(x - \bar{x})^2 + b(x - \bar{x})(y - \bar{y}) + c(y - \bar{y})^2 = D$$

の係数 a, b, c を上表に載せています。計算して確認ください。

なお（右辺）の D は距離です。

(2) マハラビノス距離から分析

マハラビノス距離を計算すると、

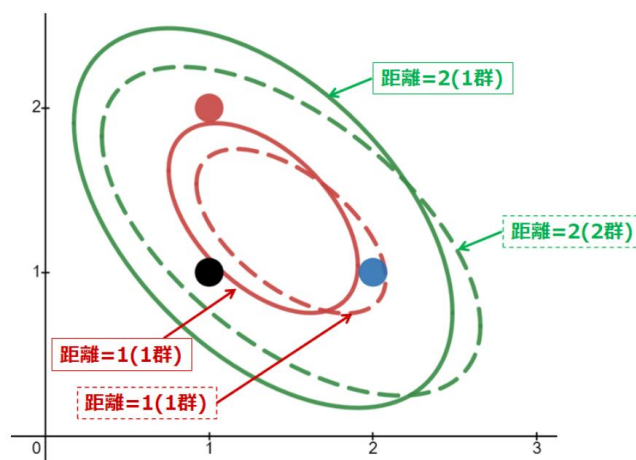
●1群: $4(x - 1.333)^2 + 4(x - 1.333)(y - 1.333) + 4(y - 1.333)^2 = D$

●2群: $4.5(x - 1.5)^2 + 6(x - 1.5)(y - 1.25) + 6(y - 1.25)^2 = D$

となります。

★グラフ表示

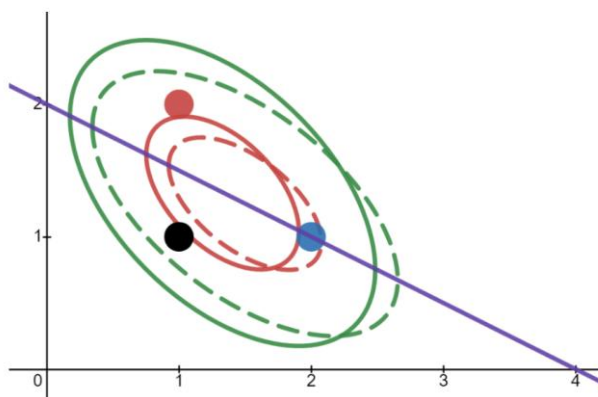
楕円を図示します。



1 群、2 群の違いがあるのかが、ちょっとわかりにくいですね。

【5】線形判別関数とマハラビノス距離の分析結果を比較

関連記事で求めた線形判別関数とマハラビノス距離の結果を1つのグラフに表示します。



どうでしょうか？

線形判別関数で作った直線の方がデータを2つに分けることがはっきりわかりますね。

このようにして、数量化Ⅱ類を分析しますが、

数量化2類は判別分析です！説明変数が質的か量的かが、違うだけで本質は同じです。

以上、「数量化2類ができる」を解説しました。

【1】数量化 4 類とは

数量化 4 類とは

- 互いの親近性から関係を数量化するもの
- 数量化 3 類に親近度の重みづけをしたもの
- 数量化 3 類の解法を応用したもの

なので、数量化 3 類の解法がベースとなります。関連記事で事前に確認しましょう。まったく同じ解法で本記事を解説します。

【関連記事】数量化 3 類の分析ができる

<https://qcplanets.com/method/multivariate/theory-class3/>

なお、分析手法を分類します。

目的	説明変数		
	目的変数	量的データ	質的データ
予測	量的データ	重回帰分析 正準相関分析	数量化1類 コンジョイント分析
	質的データ	判別分析 ロジスティック回帰分析	数量化2類
要約	なし	主成分分析 因子分析 クラスター分析 多次元尺度法 数量化IV類	数量化3類 コレスポネンス分析

【2】数量化 4 類の解き方

(1) 解法手順

では、数量化 4 類の解法を解説します！ 次のステップで解いていきます。

1. データ表を用意
2. 距離の平方和が最大になる条件を求める
3. ラグランジュの未定乗数法を使う
4. 結果的に固有方程式になる
5. 最大の固有値解からデータの関係性を求める

よく見ると、

主成分分析と同じ解法！
数量化 3 類と同じ解法！

なので、注意したいのは、

固有値を計算することより、条件式がラグランジュの未定乗数法を用いた結果
固有方程式になるという意識が大事でしたね！

数量化 4 類は親近度の重みづけが数量化 3 類の解法から追加されます。

では、実データを使いながら解説します。

【3】 解法 1.データ表を用意

(1) データ表を用意

あるアンケートを取ったら、下表のようになったとしましょう。これを数量化 4 類で分析しましょう。

親近度	S_1	S_2	S_3
S_1	–	1	2
S_2	0	–	1
S_3	3	0	–

ここで、表の意味を説明します。

S_i と S_j の親近度は

● S_i にとっての S_j の親近度 T_{ij} と

● S_j にとっての S_i の親近度 T_{ji} が

それぞれあるので、

$T_{ij} \neq T_{ji}$

も十分起こりえます。この点に注意しましょう。

その中で、 S_i と S_j の距離を

x_i と x_j を使って表現し、その距離の平方和Qを考えます。

Q の式が親近度の重みづけが入る点が数量化 3 類の解法と違って数量化 4 類の解法なのです。

(2) データの平均と分散を 0,1 と標準化する

解析しやすくするために、

●平均 0

●分散 1

とします。

★平均

$$\bullet \bar{x} = \frac{x_1 + x_2 + x_3}{3} = 0$$

★分散

$$\bullet V = \sum_{i=1}^3 \frac{(x_i - \bar{x})^2}{3} = \sum_{i=1}^3 \frac{x_i^2}{3} \\ = \frac{1}{3}(x_1^2 + x_2^2 + x_3^2) = 1$$

【4】 解法 2. 距離の平方和が最大になる条件を求める

(1) 距離の平方和Qを計算

距離の平方和Qは

$$Q = 1 \times (x_1 - x_2)^2 + 2 \times (x_1 - x_3)^2 + 0 \times (x_2 - x_1)^2 \\ + 1 \times (x_2 - x_3)^2 + 3 \times (x_3 - x_1)^2 + 0 \times (x_3 - x_2)^2 \\ = (x_1 - x_2)^2 + 2(x_1 - x_3)^2 + (x_2 - x_3)^2 + 3(x_3 - x_1)^2$$

$$Q = (x_1 - x_2)^2 + 2(x_1 - x_3)^2 + (x_2 - x_3)^2 + 3(x_3 - x_1)^2$$

【5】 解法 3. ラグランジュの未定乗数法を使う

(1) ラグランジュの未定乗数法

x の制約条件は、分散の式から

$$\bullet \frac{1}{3}(x_1^2 + x_2^2 + x_3^2) - 1 = 0$$

関数 F を下式で定義します。

$$F = Q - \frac{\lambda}{3}(x_1^2 + x_2^2 + x_3^2 - 3)$$

F, Q が最大になる条件は、

$$\bullet \frac{\partial F}{\partial x_1} = 2(x_1 - x_2) + 4(x_1 - x_3) - 6(x_3 - x_1) - \frac{2}{3}\lambda x_1 = 0$$

...①

$$\bullet \frac{\partial F}{\partial x_2} = -2(x_1 - x_2) + 2(x_2 - x_3) - \frac{2}{3}\lambda x_2 = 0 \dots ②$$

$$\bullet \frac{\partial F}{\partial x_3} = -4(x_1 - x_3) - 2(x_2 - x_3) + 6(x_3 - x_1) - \frac{2}{3}\lambda x_3 = 0$$

...③

【6】 解法 4. 結果的に固有方程式になる

①～③の式を整理していきます。

(1) ①②③式を整理

$$\bullet 6x_1 - 6x_2 - 5x_3 - \frac{1}{3}\lambda x_1 = 0$$

$$\bullet -x_1 + 2x_2 - x_3 - \frac{1}{3}\lambda x_2 = 0$$

$$\bullet -5x_1 - x_2 + 6x_3 - \frac{1}{3}\lambda x_3 = 0$$

(2) 固有方程式が結果的にできる

この式を行列表記すると、結果的、固有方程式ができます。

$\frac{1}{3}\lambda = a$ と置きます。

$$\begin{pmatrix} 6-a & -6 & -5 \\ -1 & 2-a & -1 \\ -5 & -1 & 6-a \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

(3) 固有値解を解く

上の固有方程式から、下の行列式=0 となる条件を解けば OK です。

$$\begin{vmatrix} 6-a & -6 & -5 \\ -1 & 2-a & -1 \\ -5 & -1 & 6-a \end{vmatrix}$$

行列式を解くと、

$$(6-a)(2-a)(6-a)+(-6)(-1)(-5)+(-5)(-1)(-1) \\ -(-5)(2-a)(-5)-(6)(-1)(6-a)-(6-a)(-1)(-1)=0$$

3次方程式をまとめると

$$a^3 - 14a^2 + 28a + 55 = 0$$

実は、因数分解できます！

$$(a-11)(a^2-3a-5)=0$$

よって a は

$$a = \frac{1}{3}\lambda = 11, \frac{3 \pm \sqrt{29}}{2}$$

$$\lambda = 33, 12.576, -3.579$$

が得られます。

ここまでで、固有値解が計算できました。次は固有ベクトルを計算してデータの関係性を確認します。

【7】 解法 5.最大の固有値解からデータの関係性を求める

固有値が3つ ($\lambda = 33, 12.576, -3.579$) 求まりましたので、それぞれの固有ベクトルを計算しましょう。

(1) 固有値 $\lambda = 33$ のとき

固有方程式は

$$\begin{pmatrix} -5 & -6 & -5 \\ -1 & -9 & -1 \\ -5 & -1 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

計算すると

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0.707 \\ 0 \\ -0.707 \end{pmatrix}$$

となります。

(2) 固有値 $\lambda = 12.576$ のとき

同様に解くと。

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0.380 \\ -0.521 \\ 0.764 \end{pmatrix}$$

(3) 固有値 $\lambda = -3.579$ のとき

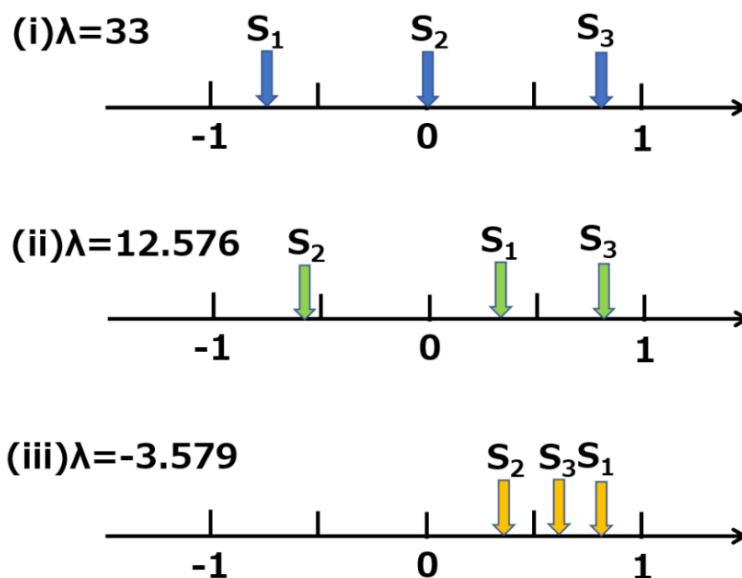
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0.725 \\ 0.402 \\ 0.559 \end{pmatrix}$$

となります。

(3) 固有ベクトルからわかること

解析結果をまとめると下表になり、その関係性を数直線に描いてみましょう。

–	$\lambda=33$	$\lambda=12.576$	$\lambda=-3.579$
x_1	0.707	0.38	0.725
x_2	0	-0.521	0.402
x_3	-0.707	0.764	0.559



どうでしょうか？ 数量化4類で分析すると、親近度の関係性が見えてきます。あとは、これをどう分析に使うかを考えていけばよいのです。

数量化4類は数量化3類と同じ解法で重みづけを加えて距離の平方和が最大になる条件をラグランジュの未定乗数法から求めます。その結果、固有方程式につながります。主成分分析、数量化3類と同じ解法の流れになりますね！

以上、「数量化4類の分析ができる」を解説しました。

- コレスポンデンス分析は各欄の値が 0 以上の整数において相関係数の最大化する条件を求めるもの
- コレスポンデンス分析は数量化 3 類と数量化 4 類の解法を合体させたものと理解しよう！

【1】コレスポンデンス分析とは

簡単にいうと

- 数量化 3 類の拡張版でカテゴリーを数量化と重みづけしたもの。重みづけは数量化 4 類の解法をベースに理解しよう！
- 数量化 3 類と数量化 4 類の解法を応用したもの

なので、数量化 3 類と数量化 4 類の解法がベースとなります。関連記事で事前に確認しましょう。まったく同じ解法で本記事を解説します。

【関連記事】数量化 3 類の分析ができる

<https://qcplanets.com/method/multivariate/theory-class3/>

★本冊子【数量化 4 類の分析ができる】

【2】コレスポンデンス分析の解き方

(1) 解法手順

では、コレスポンデンス分析の解法を解説します！ 次のステップで解いていきます。

1. データ表を用意
2. 相関係数が最大になる条件を求める
3. ラグランジュの未定乗数法を使う
4. 結果的に固有方程式になる
5. 最大の固有値解からデータの関係性を求める

よく見ると、

主成分分析と同じ解法！

数量化 3 類と同じ解法！

重みづけデータがあるのは数量化 4 類と同じ解法！

なので、注意したいのは、

固有値を計算することより、条件式がラグランジュの未定乗数法を用いた結果
固有方程式になるという意識が大事でしたね

主成分分析でも同じことを解説しています。関連記事で紹介します。

【関連記事】【まとめ】主成分分析が計算できる

<https://qcplanets.com/method/multivariate/parameters/>

では、実データを使いながら解説します。

【3】解法 1. データ表を用意

(1) データ表を用意

あるアンケートを取ったら、下表のようになったとしましょう。これをコレスポンデンス分析で分析しましょう。

データ	b_1	b_2	b_3	計
a_1	0	1	1	2
a_2	1	0	2	3
a_3	3	0	0	3
計	4	1	3	8

ここで、表の意味を説明します。

(a_i, b_j) においては、その点数が指定されています。

数量化3類は点数はなく、すべての (a_i, b_j) が1または0(空欄)でしたね。

1以上の整数が入るのが、コレスポネンシ分析と数量化3類との違いです。

その中で、表全体の相関係数が最大となる条件を考えます。

(2) データの平均と分散を0,1と標準化する

解析しやすくするために、

●平均0

●分散1

とします。

★平均

$$\bullet \bar{a} = \frac{2a_1 + 3a_2 + 3a_3}{8} = 0$$

$$\bullet \bar{b} = \frac{4b_1 + b_2 + 3b_3}{8} = 0$$

★分散V

$$\bullet V_a = \sum_{i=1}^8 \frac{(a_i - \bar{a})^2}{8} = \sum_{i=1}^8 \frac{a_i^2}{8} \\ = \frac{1}{8}(2a_1^2 + 3a_2^2 + 3a_3^2) = 1$$

$$\bullet V_b = \sum_{i=1}^8 \frac{(b_i - \bar{b})^2}{8} = \sum_{i=1}^8 \frac{b_i^2}{8} \\ = \frac{1}{8}(4b_1^2 + b_2^2 + 3b_3^2) = 1$$

まとめると、

$$\frac{1}{8}(2a_1^2 + 3a_2^2 + 3a_3^2) = 1$$

$$\frac{1}{8}(4b_1^2 + b_2^2 + 3b_3^2) = 1$$

後で使う式となります。

【4】 解法2.相関係数が最大になる条件を求める

解法は数量化3類と全く同じです。

(1) 相関係数を計算

相関係数 r は $r = \frac{S_{ab}}{S_a S_b}$ ですね。

分母はすでに分散のところで計算済なので、

$$\bullet S_a = (2a_1^2 + 3a_2^2 + 3a_3^2) = 8$$

$$\bullet S_b = (4b_1^2 + b_2^2 + 3b_3^2) = 8$$

分子を計算すると、

$$\bullet S_{ab} = \sum_{i=1}^8 (a_i - \bar{a})(b_i - \bar{b}) \\ = a_1 b_2 + a_1 b_3 + a_2 b_1 + 2a_2 b_3 + 3a_3 b_1$$

よって、相関係数 r は

$$r = \frac{1}{8}(a_1 b_2 + a_1 b_3 + a_2 b_1 + 2a_2 b_3 + 3a_3 b_1)$$

【5】 解法 3. ラグランジュの未定乗数法を使う

(1) ラグランジュの未定乗数法

a, b の制約条件は、分散の式から

$$\bullet \frac{1}{8}(2a_1^2 + 3a_2^2 + 3a_3^2) - 1 = 0$$

$$\bullet \frac{1}{8}(4b_1^2 + b_2^2 + 3b_3^2) - 1 = 0$$

関数 F を下式で定義します。今回変数が a, b の 2 種類があるので λ_1, λ_2 を使います。

$$F = \frac{1}{8}(a_1b_2 + a_1b_3 + a_2b_1 + 2a_2b_3 + 3a_3b_1)$$

$$- \frac{\lambda_1}{2} \left(\frac{1}{8}(2a_1^2 + 3a_2^2 + 3a_3^2) - 1 \right)$$

$$- \frac{\lambda_2}{2} \left(\frac{1}{8}(4b_1^2 + b_2^2 + 3b_3^2) - 1 \right)$$

相関係数 r が最大になる条件は、

$$\bullet \frac{\partial F}{\partial a_1} = \frac{1}{8}(b_2 + b_3) - \frac{\lambda_1}{2} \frac{4a_1}{8} = 0 \dots ①$$

$$\bullet \frac{\partial F}{\partial a_2} = \frac{1}{8}(b_1 + 2b_3) - \frac{\lambda_1}{2} \frac{6a_2}{8} = 0 \dots ②$$

$$\bullet \frac{\partial F}{\partial a_3} = \frac{1}{8}(3b_1) - \frac{\lambda_1}{2} \frac{6a_3}{8} = 0 \dots ③$$

$$\bullet \frac{\partial F}{\partial b_1} = \frac{1}{8}(a_2 + 3a_3) - \frac{\lambda_2}{2} \frac{8b_1}{8} = 0 \dots ④$$

$$\bullet \frac{\partial F}{\partial b_2} = \frac{1}{8}(a_1) - \frac{\lambda_2}{2} \frac{2b_2}{8} = 0 \dots ⑤$$

$$\bullet \frac{\partial F}{\partial b_3} = \frac{1}{8}(a_1 + 2a_2) - \frac{\lambda_2}{2} \frac{6b_3}{8} = 0 \dots ⑥$$

本記事のテーマはラグランジュの未定乗数法を使うことです。ここをよく意識しておいてください。

【6】 解法 4. 結果的に固有方程式になる

①～⑥の式を整理していきます。

(1) 相関係数 r と λ_1, λ_2 の関係式を作る

$$① \times a_1 + ② \times a_2 + ③ \times a_3$$

$$(a_1b_2 + a_1b_3 - 2\lambda_1a_1^2) + (a_2b_1 + 2a_2b_3 - 3\lambda_1a_2^2) +$$

$$(3a_3b_1 - 3\lambda_1a_3^2) = 0$$

$$(a_1b_2 + a_1b_3 + a_2b_1 + 2a_2b_3 + 3a_3b_1) -$$

$$\lambda_1(2a_1^2 + 3a_2^2 + 3a_3^2) = 0$$

$$8r - 8\lambda_1 = 0$$

より、

$$r = \lambda_1$$

とシンプルになります。

同様に

$$④ \times b_1 + ⑤ \times b_2 + ⑥ \times b_3$$

$$(a_2b_1 + 3a_3b_1 - 4\lambda_2b_1^2) + (a_1b_2 - \lambda_2b_2^2) +$$

$$a_1b_3 + 2a_2b_3 - 3\lambda_2b_3^2 = 0$$

$$(a_2b_1 + 3a_3b_1 + a_1b_2 + a_1b_3 + 2a_2b_3) -$$

$$\lambda_2(4b_1^2 + b_2^2 + 3b_3^2) = 0$$

$$8r - 8\lambda_2 = 0$$

より、

$$r = \lambda_2$$

とシンプルになります。

まとめると、 $r = \lambda_1 = \lambda_2$ の関係式を使っていきます。

(2) 固有方程式が結果的にできる

$\lambda_1 = \lambda_2 = \lambda$ として、①②③式から

$$\bullet a_1 = \frac{b_2 + b_3}{2\lambda} \quad \dots \textcircled{1}$$

$$\bullet a_2 = \frac{b_1 + 2b_3}{3\lambda} \quad \dots \textcircled{2}$$

$$\bullet a_3 = \frac{b_1}{\lambda} \quad \dots \textcircled{3}$$

④⑤⑥式に代入すると

$$\bullet \frac{b_1 + 2b_3}{3\lambda} + \frac{3b_1}{\lambda} - 4\lambda b_1 = 0 \quad \dots \textcircled{4}$$

$$\bullet \frac{b_2 + b_3}{2\lambda} - \lambda b_2 = 0 \quad \dots \textcircled{5}$$

$$\bullet \frac{b_2 + b_3}{2\lambda} + \frac{2b_1 + 4b_2}{3\lambda} - 3\lambda b_3 = 0 \quad \dots \textcircled{6}$$

この式を行列表記すると結果的、固有方程式ができます。

$$\begin{pmatrix} 10 - 12\lambda^2 & 0 & 2 \\ 0 & 1 - 2\lambda^2 & 1 \\ 4 & 11 & 3 - 18\lambda^2 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

(3) 固有値解を解く

上の固有方程式から、下の行列式=0 となる条件を解けば OK です。3 次方程式になりますが、頑張って解きます！

$$\begin{vmatrix} 10 - 12\lambda^2 & 0 & 2 \\ 0 & 1 - 2\lambda^2 & 1 \\ 4 & 11 & 3 - 18\lambda^2 \end{vmatrix} = 0$$

$\lambda^2 = t (t \geq 0)$ において、行列式を解くと、

$$\lambda^2 = t (t \geq 0)$$

$$(10 - 12t)(1 - 2t)(3 - 18t) + 0 \cdot 1 \cdot 4 + 2 \cdot 0 \cdot 11 - 2 \cdot (1 - 2t) \cdot 4 - 0 \cdot 0 \cdot (3 - 18t) - (10 - 12t) \cdot 1 \cdot 11 = 0$$

$$54t^3 - 81t^2 + 16t + 11 = 0$$

$$(t - 1)(54t^2 - 27t - 11) = 0$$

より、

$$t = 1, -0.2659, 0.766$$

$t \geq 0$ より、

$$t = 1, 0.766$$

数量化 3 類でありましたように、 $t=1$ は相関係数が 1 となるので、特例として扱わないとします。

よって、 $t=0.766$ 、 $\lambda=0.875$ の 1 つだけとなります。

ここまでで、固有値解が計算できました。次は固有ベクトルを計算してデータの関係性を確認します。

【7】 解法 5.最大の固有値解からデータの関係性を求める

固有値が 1 つ $\lambda=0.875$ 求まりましたので、それぞれの固有ベクトルを計算しましょう。

(1) 固有値 $\lambda=0.875$ のとき

固有方程式は

$$\begin{pmatrix} 0.808 & 0 & 2 \\ 0 & -0.532 & 1 \\ 4 & 11 & -10.788 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

計算すると

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} -0.758 \\ 0.575 \\ 0.306 \end{pmatrix}$$

a_1, a_2, a_3 は

$$a_1 = \frac{b_2 + b_3}{2\lambda} = 0.504$$

$$a_2 = \frac{b_1 + 2b_3}{3\lambda} = 0.150$$

$$a_3 = \frac{b_1}{\lambda} = -0.866$$

$$r = \lambda = 0.875$$

(2) 固有値 $\lambda=0.809$ からわかること

a_1, a_2, a_3 と

b_1, b_2, b_3 の

大きい順に並べると

$$\bullet b_2 = 0.575, b_3 = 0.306, b_1 = -0.758$$

$$\bullet a_1 = 0.504, a_2 = 0.150, a_3 = -0.866$$

の順になります。これを最初の表に適用すると、相関性の高い順に並び変わります。

元の表は、

データ	b_1	b_2	b_3	計
a_1	0	1	1	2
a_2	1	0	2	3
a_3	3	0	0	3
計	4	1	3	8

から下表に変化します。

データ	b_2	b_3	b_1	計
a_1	1	1	0	2
a_2	0	2	1	3
a_3	0	0	3	3
計	1	3	4	8

どうでしょうか？

左上から右下への対角線上にデータが乗るように、入れ替わりましたね！
これがコレスポンデンス分析で実施したいことです。

コレスポンデンス分析は
相関係数が最大になる条件を
ラグランジュの未定乗数法から求めます。
その結果、固有方程式につながります。
数量化 3 類と数量化 4 類の応用した解法ですね！

以上、「コレスポンデンス分析ができる」を解説しました。