

数量化Ⅱ類一＞正準判別分析

1. 目的

最初に 数量化Ⅱ類を実行して対象の数量化を行った後、正準判別分析を実行し、どの群に近いか（所属するか）の判定を行います。

数量化Ⅱ類と同様に、例題を提示して説明します。

某コンビニチェーンは、既存の店舗データを参考に新規出店を計画します。

店舗状況として以下を考えます。

- ・立地 : 駅周辺、ビル街、商店街 (3 択)
- ・競合店 : ある、ない (2 択)
- ・通行量 : 多い、少ない (2 択)

店舗状況を要因、各要因の分類をカテゴリと呼ぶことにします。
上記の例では、要因は3つ、「立地」、「競合店」、「通行量」です。
各要因のカテゴリは、

要因「立地」: 「駅周辺」、「ビル街」、「商店街」

要因「競合店」: 「ある」、「ない」

要因「通行量」: 「多い」、「少ない」

となります。

下表は、既存12店舗の収益を、要因とカテゴリを用いて、表現しています。

ID	収益	立地	競合店	通行量
ID-1	黒字	駅周辺	ある	多い
ID-2	黒字	ビル街	ない	多い
ID-3	黒字	駅周辺	ない	多い
ID-4	黒字	商店街	ない	少ない
ID-5	黒字	ビル街	ない	少ない
ID-6	赤字	商店街	ある	多い
ID-7	赤字	ビル街	ある	少ない
ID-8	赤字	商店街	ある	多い
ID-9	赤字	ビル街	ある	少ない
ID-10	赤字	駅周辺	ある	少ない
ID-11	赤字	商店街	ない	少ない
ID-12	赤字	商店街	ある	少ない

要因、カテゴリの違いが、収益面での黒字、赤字を左右することは自明です。
そこで、各要因、カテゴリの影響を数量化して、それが黒字、赤字の違いを表現できるならば、新規店舗の計画にあたっての収益性の予測に役立ちます。

数量化のモデル式は、この例では以下となります。

$$\begin{aligned} \text{数量化スコア } S = & \alpha_0 \text{ (定数項)} + \\ & \alpha_1 \text{ (立地:駅周辺)} * \delta_1 + \\ & \alpha_2 \text{ (立地:ビル街)} * \delta_2 + \\ & \alpha_3 \text{ (立地:商店街)} * \delta_3 + \\ & \alpha_4 \text{ (競合店:ある)} * \delta_4 + \\ & \alpha_5 \text{ (競合店:ない)} * \delta_5 + \\ & \alpha_6 \text{ (通行量:多い)} * \delta_6 + \\ & \alpha_7 \text{ (通行量:少ない)} * \delta_7 \end{aligned}$$

上式の $\alpha_0 \sim \alpha_7$ は未知の数値です。

$\delta_1 \sim \delta_7$ は 実際のデータに当てはめて 0 または 1 となります。

実際のデータに当てはめると、以下の表になります。

ID	立地 駅周辺 δ_1	立地 ビル街 δ_2	立地 商店街 δ_3	競合店 ある δ_4	競合店 ない δ_5	通行量 多い δ_6	通行量 少ない δ_7
ID-1	1	0	0	1	0	1	0
ID-2	0	1	0	0	1	1	0
ID-3	1	0	0	0	1	1	0
ID-4	0	0	1	0	1	0	1
ID-5	0	1	0	0	1	0	1
ID-6	0	0	1	1	0	1	0
ID-7	0	1	0	1	0	0	1
ID-8	0	0	1	1	0	1	0
ID-9	0	1	0	1	0	0	1
ID-10	1	0	0	1	0	0	1
ID-11	0	0	1	0	1	0	1
ID-12	0	0	1	1	0	0	1

数量化Ⅱ類は最適な $\alpha_0 \sim \alpha_7$ を求めるアルゴリズムです。

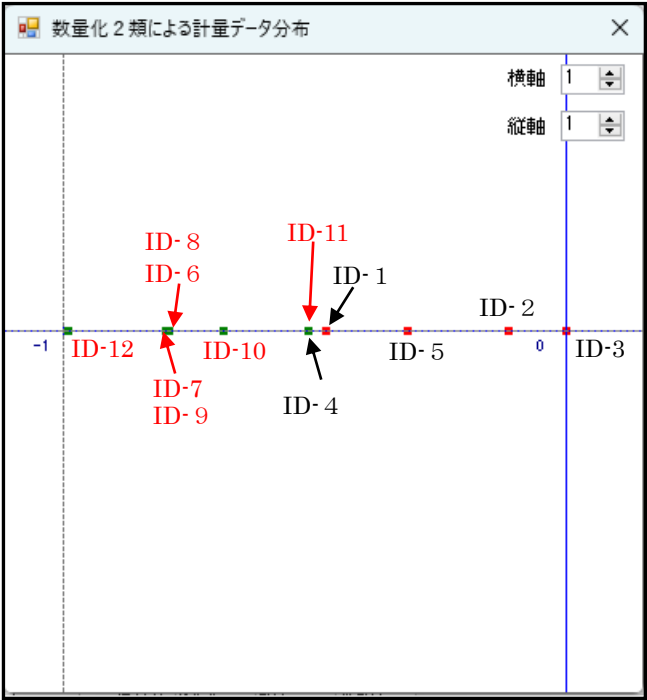
$\alpha_0 \sim \alpha_7$ を適切に定義できたとして、12 店舗の数量 ($S_1 \sim S_{12}$) が計算されたとします。

黒字店（ID-1～ID-5）の数量 $S_1 \sim S_5$ と、赤字店（ID-6～ID-12）の数量 $S_6 \sim S_{12}$ とが数値上 分離されて表現できれば、 $\alpha_0 \sim \alpha_7$ の定義は、成功していると考えられます。

実際、アルゴリズム適用した結果は以下となりました。

No.	ID	S	収益	第 1 軸
1	ID-1	S_1	黒字	-0.4786871
2	ID-2	S_2	黒字	-0.1141241
3	ID-3	S_3	黒字	1.39E-17
4	ID-4	S_4	黒字	-0.5135584
5	ID-5	S_5	黒字	-0.3170114
6	ID-6	S_6	赤字	-0.7893583
7	ID-7	S_7	赤字	-0.7956985
8	ID-8	S_8	赤字	-0.7893583
9	ID-9	S_9	赤字	-0.7956985
10	ID-10	S_{10}	赤字	-0.6815744
11	ID-11	S_{11}	赤字	-0.5135584
12	ID-12	S_{12}	赤字	-0.9922456

数量化結果を 第 1 軸 のみでグラフに表現すると以下です。



この図から黒字店（ID-1～ID-5）と赤字店（ID-6～ID-12）との分離がある程度できています。

しかし 黒字店 ID-4 と赤字店 ID-11 とが重なっています。
 ID-4 と ID-11 とは条件が同じ（商店街、競合店なし、交通量少ない）
 なのに 片方は黒字で、片方は赤字 となっているからです。

これは、カテゴリを細かく定義する、もしくは要因・カテゴリを増やして
 説明変数の自由度を上げる必要を示しています。
 ただし同時に、サンプルを増やす必要が生じるかもしれません。

数量化Ⅱ類は、対象の分類（黒字、赤字）が明確になるように、
 各データに1次元（必要なら2，3...次元）の座標を付与します。

また、数量化Ⅱ類は、各データの座標を計算する前に、数量化のモデル式の
 パラメータ $\alpha_0 \sim \alpha_7$ を計算します。

各対象のスコア（数値）は、以下の式で表現されました。

$$\begin{aligned} \text{スコア } S = & \alpha_0 \text{ (定数項)} + \\ & \alpha_1 \text{ (立地:駅周辺)} * \delta_1 + \\ & \alpha_2 \text{ (立地:ビル街)} * \delta_2 + \\ & \alpha_3 \text{ (立地:商店街)} * \delta_3 + \\ & \alpha_4 \text{ (競合店:ある)} * \delta_4 + \\ & \alpha_5 \text{ (競合店:ない)} * \delta_5 + \\ & \alpha_6 \text{ (通行量:多い)} * \delta_6 + \\ & \alpha_7 \text{ (通行量:少ない)} * \delta_7 \end{aligned}$$

例題では、以下のように パラメータ $\alpha_0 \sim \alpha_7$ が計算されます。

No.	Item	Catg		第1軸	第2軸	第3軸	第4軸
0	定数項	-	α_0	-0.5650727	0.2894817	-0.1627176	-0.1263791
1	立地条件	駅周辺	α_1	0.1674877	-0.2309712	0.4398736	0.04010426
2	立地条件	ビル街	α_2	0.05336358	0.4047922	-0.04873468	0.1036861
3	立地条件	商店街	α_3	-0.1431835	-0.185251	-0.2249364	-0.1070114
4	競合店	ない	α_4	0.2792342	-0.04109341	-0.07105615	-0.1844174
5	競合店	ある	α_5	-0.199453	0.02935243	0.05075439	0.1317267
6	通行量	多い	α_6	0.1183509	-0.01741713	-0.2060998	0.2706922
7	通行量	少ない	α_7	-0.08453636	0.01244081	0.1472142	-0.1933516

新規開店を予定している店舗の状況に応じて、 $\delta_1 \sim \delta_7$ を設定し、
 スコア S を計算することで 数量化できます。

以上は 数量化Ⅱ類による計算結果です。

次に 正準判別分析を実行して、新規開店を予定している店舗が、
どちらの群（黒字、赤字）に所属するかを決めます。

数量化Ⅱ類の実行により、既存店舗 ID-1～ID-12 の数量が計算されて
います。

No.	ID	収益	1	2	3	4
1	ID-1	黒字	-0.4786871	0.07044584	0.1940492	-0.2777225
2	ID-2	黒字	-0.1141241	0.6357633	-0.4591168	-0.1788643
3	ID-3	黒字	1.39E-17	-3.47E-18	2.78E-17	0
4	ID-4	黒字	-0.5135584	0.07557806	-0.4489287	0.5186606
5	ID-5	黒字	-0.3170114	0.6656213	-0.2273537	0.3563491
6	ID-6	赤字	-0.7893583	0.116166	-0.4866426	-0.2942753
7	ID-7	赤字	-0.7956985	0.7360671	-0.03330447	0.07862655
8	ID-8	赤字	-0.7893583	0.116166	-0.4866426	-0.2942753
9	ID-9	赤字	-0.7956985	0.7360671	-0.03330447	0.07862655
10	ID-10	赤字	-0.6815744	0.1003038	0.4258124	0.2574908
11	ID-11	赤字	-0.5135584	0.07557806	-0.4489287	0.5186606
12	ID-12	赤字	-0.9922456	0.1460239	-0.2548794	0.2409381

ここで新規開店を予定する店舗の属性（立地、競合店、交通量）を指定して、
それについての数量も計算されます。

例えば（立地、競合店、交通量）＝（商店、ない、多い）とした場合の
4つの数量は （-0.3106711、0.04572012、-0.6806918、-0.01655276）
となります。

この座標成分が、黒字店の構成する5つの点群と、赤字店の構成する7つの
点群との間で どちらに所属するかを計算するのが、正準判別分析です。

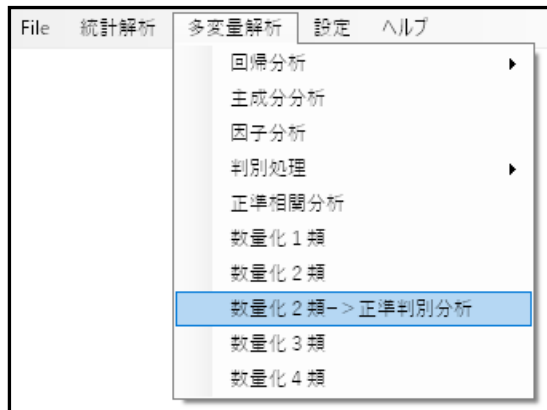
具体的には、各群の平均座標を計算し、各群の平均からの汎距離
（マハラノビス距離）を計算して、最も近い群に所属することにします。

なお、この例では2つの群のいずれに所属する？を問題にしていますが、
正準判別分析は3群以上のデータにも対応したものです。
正準判別分析 を参照してください。

2. 使用法

(1) メニューの選択

メニューの「多変量解析→数量化2類⇒正準判別分析」を選択します。



(2) パネルが表示されます。

数量化2類⇒正準判別分析

有意水準 α (%): 5 正準スコアの計算 固有値のスクリーンプロット サンプル数量化表示 群間差検定結果 推計式による予測

数量化のカテゴリースコア:

No.	Item	Catg	Score1
-----	------	------	--------

数量化の重相関係数、偏相関係数:

No.	Item	Coef1
-----	------	-------

サンプルの正準スコア: 表示次元数: 2 正準スコア表示

No.	ID	Z1
-----	----	----

サンプルと各群とのマハラノビス距離:

No.	ID	Mdst1
-----	----	-------

各群の平均値:

No.	Group	Count	Mean1
-----	-------	-------	-------

各群の正準スコアの平均座標:

No.	Group	Z1
-----	-------	----

要因、カテゴリーの定義

☐ 表データを貼り付け ☐ 先頭行をラベルとして使用

No.	Item1
-----	-------

サンプルデータの入力

☐ 表データを貼り付け ☐ 先頭行をラベルとして使用 クリア

No.	ID	Group	Item1	Item2	Item3	Item4	Item5
*							

(3) データの入力

入力対象のグリッドは2種あります。
これは 数量化Ⅱ類と同様です。

最初は、要因とそのカテゴリを定義するグリッドです。

要因、カテゴリーの定義

表データを貼り付け

☐ 先頭行をラベルとして使用

	No.	Item1
--	-----	-------

以下のように前もって表計算ソフトで定義しておきます。

要因1	要因2	要因3
立地	競合店	通行量
駅周辺	ない	多い
ビル街	ある	少ない
商店街		

データの赤線部分をコピーします。表題部分も取り込みますので、
「先頭行をラベルとして使用」にチェックを入れます。

表データを貼り付け

 をクリックすると、グリッド部分にコピーされます。

要因、カテゴリーの定義

表データを貼り付け

☒ 先頭行をラベルとして使用

	No	立地	競合店	通行量
▶	1	駅周辺	ない	多い
	2	ビル街	ある	少ない
	3	商店街		

このグリッドを、先に定義しないと、続く データの定義ができません。

続いて、パネルのグリッド（下の部分）にデータを入力します。

サンプルデータの入力

表データを貼り付け

☐ 先頭行をラベルとして使用

クリア

	No.	ID	Group	Item1	Item2	Item3	Item4	Item5
*								

以下のように前もって表計算ソフトで定義しておきます。

		要因1	要因2	要因3
ID	収益	立地	競合店	通行量
ID-1	黒字	駅周辺	ある	多い
ID-2	黒字	ビル街	ない	多い
ID-3	黒字	駅周辺	ない	多い
ID-4	黒字	商店街	ない	少ない
ID-5	黒字	ビル街	ない	少ない
ID-6	赤字	商店街	ある	多い
ID-7	赤字	ビル街	ある	少ない
ID-8	赤字	商店街	ある	多い
ID-9	赤字	ビル街	ある	少ない
ID-10	赤字	駅周辺	ある	少ない
ID-11	赤字	商店街	ない	少ない
ID-12	赤字	商店街	ある	少ない

データの赤線部分をコピーします。表題部分も取り込みますので、「先頭行をラベルとして使用」にチェックを入れます。

表データを貼り付け

をクリックすると、以下のようにグリッド部分にコピーされます。

サンプルデータの入力

表データを貼り付け

☒ 先頭行をラベルとして使用

クリア

	No.	ID	収益	立地	競合店	通行量
▶	1	ID-1	黒字	駅周辺	ある	多い
	2	ID-2	黒字	ビル街	ない	多い
	3	ID-3	黒字	駅周辺	ない	多い

(4) 計算実行

正準スコアの計算

ボタンを押すと計算されます。

数量化のカテゴリースコア:

No.	Item	Catg	Score1	Score2	Score3	Score4
0	定数項	-	-0.5650...	0.28948...	-0.18827	0.08368...
1	立地	駅周辺	0.16748...	-0.2309...	0.43666...	0.06651...
2	立地	ビル街	0.05336...	0.40479...	-0.0224...	-0.1123...

数量化の重相関係数、偏相関係数:

No.	Item	Coef1	Coef2	Coef3
0	重相関...	0.820917	0.74380...	0.81030
1	立地	0.53541...	-0.1507...	-0.5016

サンプルの正準スコア: 表示次元数: 2 正準スコア表示

No.	ID	Z1
1	ID-1	0.47837...
2	ID-2	2.497209
3	ID-3	3.129191

サンプルと各群とのマハラノビス距離:

No.	ID	黒字	赤字
1	ID-1	0.74830...	3.947947
2	ID-2	0.57828...	20.37425
3	ID-3	1.611132	28.14086

各群の平均値:

No.	収益	Count	Mean1
1	黒字	5	-0.2846...
2	赤字	7	-0.7653...

各群の正準スコアの平均座標:

No.	Group	Z1
1	黒字	1.552746
2	赤字	-1.1091...

上記の各グリッドに計算結果が表示されます。

表示枠が小さい為、スクロールして見ることになりますが、グリッドの左上隅のセルをクリックして、Ctrl-C を押すと、表形式のデータがコピーされます。その際、グリッドが 以下のように青くなるのを確認してください。

そのまま、表計算ソフトのシート上で Ctrl-V を押すと、表データがペーストされます。

以下 ①～⑥の表について、説明します。

① 数量化のカテゴリースコア

数量化モデル式の $\alpha_0 \sim \alpha_7$ を表示しています。

4次元までのカテゴリースコアを表示しています。

この $\alpha_0 \sim \alpha_7$ を用いて、既存店舗 ID-1～ID-12 の座標成分（4次元）を計算することになります。

No.	Item	Catg		第1軸	第2軸	第3軸	第4軸
0	定数項	-	α_0	-0.5650727	0.2894817	-0.1627176	-0.1263791
1	立地条件	駅周辺	α_1	0.1674877	-0.2309712	0.4398736	0.04010426
2	立地条件	ビル街	α_2	0.05336358	0.4047922	-0.04873468	0.1036861
3	立地条件	商店街	α_3	-0.1431835	-0.185251	-0.2249364	-0.1070114
4	競合店	ない	α_4	0.2792342	-0.04109341	-0.07105615	-0.1844174
5	競合店	ある	α_5	-0.199453	0.02935243	0.05075439	0.1317267
6	通行量	多い	α_6	0.1183509	-0.01741713	-0.2060998	0.2706922
7	通行量	少ない	α_7	-0.08453636	0.01244081	0.1472142	-0.1933516

② 数量化の重相関係数、偏相関係数

重相関係数、3つの偏相関係数（立地、競合店、通行量）の4つの指標が示されています。

いずれも、推定の精度を示すもので、1.0に近いほど当てはまりの程度がよいことを示します。

4つの次元（Coef1～Coef4）について表示しています。

No.	Item	Coef1	Coef2	Coef3	Coef4
0	重相関係数	0.820917	0.7438084	0.8103094	0.7429087
1	立地	0.5354126	-0.1507857	-0.5016249	-0.1407498
2	競合店	0.7614468	0.696878	0.7590048	0.6967023
3	通行量	0.4424946	0.4619254	0.3885985	-0.4596388

③ サンプルの正準スコア

既存店舗 ID-1～ID-12 の数量化結果を入力として、正準判別分析を行い、各サンプル（既存店舗 ID-1～ID-12）の 正準スコア を表示したものです。

この場合、正準スコアは 1変量で表現されます。

No.	ID	Z1
1	ID-1	0.4783757
2	ID-2	2.497209
3	ID-3	3.129191
4	ID-4	0.2852699
5	ID-5	1.373684
6	ID-6	-1.242021
7	ID-7	-1.277131
8	ID-8	-1.242021
9	ID-9	-1.277131
10	ID-10	-0.6451489
11	ID-11	0.2852699
12	ID-12	-2.365546

④ サンプルと各群とのマハラノビス距離

サンプル毎（既存店舗 ID-1～ID-12）に 各群（黒字、赤字）の
 正準スコアの平均との間の距離を表示しています。
 この場合の距離はマハラノビス距離です（ユークリッド距離を分散で
 割った距離）。
 この距離が一番小さい群に サンプルは属していると考えるのが
 妥当です。
 近い群の方が 黄色で表示されます。

No.	ID	黒字	赤字
1	ID-1	0.7483097	3.947947
2	ID-2	0.5782866	20.37425
3	ID-3	1.611132	28.14086
4	ID-4	1.041485	3.045885
5	ID-5	0.02078638	9.656823
6	ID-6	5.063662	0.02767679
7	ID-7	5.191689	0.04422965
8	ID-8	5.063662	0.02767679
9	ID-9	5.191689	0.04422965
10	ID-10	3.131751	0.3372154
11	ID-11	1.041485	3.045885
12	ID-12	9.953297	2.473087

⑤ 各群の平均値

群ごとの 数量化データ（数量化Ⅱ類により算出された座標）の平均が
 表示されています。
 また、ここでは、正準判別分析が1次元に縮約されるため、1次元のみ
 表示しています。

No.	収益	Count	Mean1
1	黒字	5	-0.2846762
2	赤字	7	-0.765356
3	全体	12	-0.5650727

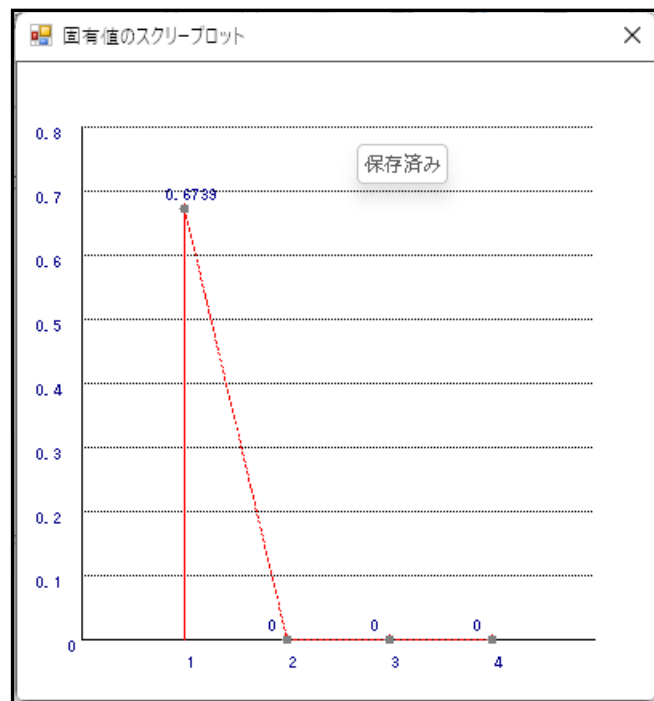
⑥ 各群の正準スコアの平均座標

各群の正準スコアの平均が表示されています。
ここでは、正準判別分析が1次元に縮約されるため、1次元のみ表示しています。

No.	Group	Z1
1	黒字	1.552746
2	赤字	-1.109104

固有値のスクリーンプロット

をクリックすると、計算に有効と思われる固有値の個数を見積もることができます。



サンプル数量化表示

をクリックすると、対象の数量化結果がグリッドに表示されます。

No.	ID	収益	1	2	3	4
1	ID-1	黒字	-0.4786...	0.07044...	0.19404...	-0.2777...
2	ID-2	黒字	-0.1141...	0.63576...	-0.4591...	-0.1788...
3	ID-3	黒字	1.38777...	-3.4694...	2.77555...	0
4	ID-4	黒字	-0.5135...	0.07557...	-0.4489...	0.51866...

そのまま、表計算ソフトのシート上で Ctrl-V を押すと、表データがペーストされます。

No.	ID	収益	1	2	3	4
1	ID-1	黒字	-0.4786871	0.07044584	0.1940492	-0.2777225
2	ID-2	黒字	-0.1141241	0.6357633	-0.4591168	-0.1788643
3	ID-3	黒字	1.39E-17	-3.47E-18	2.78E-17	0
4	ID-4	黒字	-0.5135584	0.07557806	-0.4489287	0.5186606
5	ID-5	黒字	-0.3170114	0.6656213	-0.2273537	0.3563491
6	ID-6	赤字	-0.7893583	0.116166	-0.4866426	-0.2942753
7	ID-7	赤字	-0.7956985	0.7360671	-0.03330447	0.07862655
8	ID-8	赤字	-0.7893583	0.116166	-0.4866426	-0.2942753
9	ID-9	赤字	-0.7956985	0.7360671	-0.03330447	0.07862655
10	ID-10	赤字	-0.6815744	0.1003038	0.4258124	0.2574908
11	ID-11	赤字	-0.5135584	0.07557806	-0.4489287	0.5186606
12	ID-12	赤字	-0.9922456	0.1460239	-0.2548794	0.2409381

推計式による予測 をクリックすると、新たな定義についての数量化を行い、更に正準スコアを計算します。

数量化、正準判別分析に基づく判定

判定対象のデータ入力:

N0	要因名	カテゴリ名
1	立地	駅周辺
2	競合店	ない
3	通行量	少ない

計算

入力データの数値化結果:

N0	変量名	変量値
1	1	-0.2028873

指定データの正準スコア:

N0	Z	スコア
1	1	2.005667

分布表示

判定対象と各群の平均との汎距離:

N0	群	汎距離
1	黒字	0.132989
2	赤字	15.19872

黒字側に近いことを示す

パネル表示

