

数量化Ⅱ類

1. 目的

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。

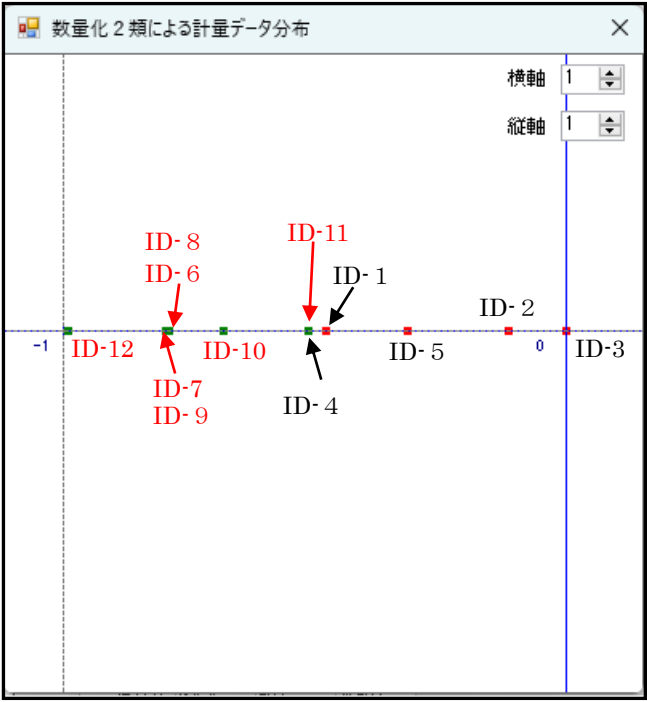
$\alpha_0 \sim \alpha_7$ を適切に定義できたとして、12 店舗の数量 ($S_1 \sim S_{12}$) が計算されたとします。

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。

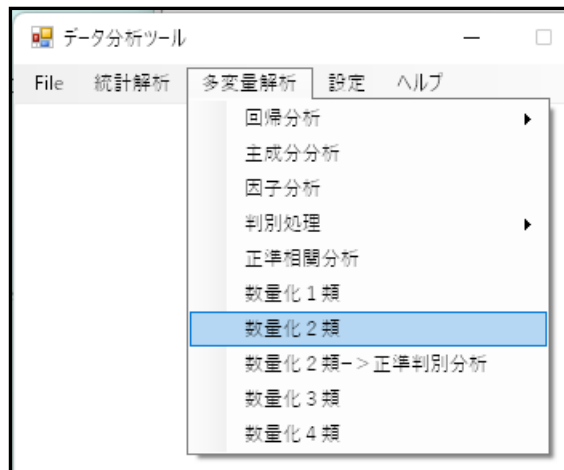
No.	Item	Catg		第 1 軸	第 2 軸	第 3 軸	第 4 軸
0	定数項	-	α_0	-0.5650727	0.2894817	-0.1627176	-0.1263791
1	立地条件	駅周辺	α_1	0.1674877	-0.2309712	0.4398736	0.04010426
2	立地条件	ビル街	α_2	0.05336358	0.4047922	-0.04873468	0.1036861
3	立地条件	商店街	α_3	-0.1431835	-0.185251	-0.2249364	-0.1070114
4	競合店	ない	α_4	0.2792342	-0.04109341	-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 予測を可能にします。

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」を選択します。



(2) パネルが表示されます。

The screenshot shows the '数量化2類' (Quantitative Class 2) panel. It has a title bar with '数量化2類' and standard window controls. The panel contains several sections:

- カテゴリースコアの推定** (Estimation of Category Scores): Includes buttons for '固有値のスクリーンプロット' (Eigenvalue Screen Plot), 'サンプル数量化表示' (Sample Quantification Display), and '推計式による予測' (Prediction by Estimation Formula).
- カテゴリースコア :** A table with columns 'No.', 'Item', 'Catg', and 'Score1'.
- 重相関係数、偏相関係数 :** A table with columns 'No.', 'Item', and 'Coef1'.
- 要因、カテゴリの定義** (Definition of Factors and Categories): Includes buttons for '表データを貼り付け' (Paste Table Data) and '先頭行をラベルとして使用' (Use First Row as Label). Below is a table with columns 'NO', 'Item1', 'Item2', 'Item3', 'Item4', and 'Item5'.
- 群、各要因のカテゴリ値の設定** (Setting Category Values for Groups and Factors): Includes buttons for '表データを貼り付け' (Paste Table Data), '先頭行をラベルとして使用' (Use First Row as Label), and 'クリア' (Clear). Below is a table with columns 'NO', 'ID', 'Group', 'Item1', 'Item2', 'Item3', 'Item4', and 'Item5'.

(3) データの入力

入力対象のグリッドは2種あります。

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、カテゴリの定義

表データを貼り付け

☐ 先頭行をラベルとして使用

	N0	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
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、カテゴリの定義

表データを貼り付け

☒ 先頭行をラベルとして使用

	N0	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。

群、各要因のカテゴリ値の設定

表データを貼り付け

☐ 先頭行をラベルとして使用

クリア

	N0	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。

群、各要因のカテゴリ値の設定

表データを貼り付け

☒ 先頭行をラベルとして使用

クリア

	N0	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	ID-4	黒字	商店街	ない	少ない
	5	ID-5	黒字	ビル街	ない	少ない
	6	ID-6	赤字	商店街	ある	多い
	7	ID-7	赤字	ビル街	ある	少ない

(4) 計算実行

カテゴリスコアの推定

ボタンを押すと計算されます。

カテゴリスコアの推定

固有値のスクリーンプロット

サンプル数量化表示

推計式による予測

カテゴリスコア :

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...
3	立地	商店街	-0.1431...	-0.1852...	-0.24403...	0.04996...

重相関係数、偏相関係数 :

No.	Item	P6C1	P6C2	P6C3
0	重相関...	0.820917	0.74380...	0.81030...
1	立地	0.53541...	-0.1507...	-0.5016...
2	競合店	0.76144...	0.696878	0.75900...
3	通行量	0.44049...	0.46139...	0.20059...

上記の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。

ここをクリックして Ctrl-C を押す

カテゴリスコア :

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...
3	立地	商店街	-0.1431...	-0.1852...	-0.24403...	0.04996...

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.	Item	Catg	Score1	Score2	Score3	Score4
0	定数項	-	-0.5650727	0.2894817	-0.18827	0.08368457
1	立地	駅周辺	0.1674877	-0.2309712	0.4366605	0.0665184
2	立地	ビル街	0.05336358	0.4047922	-0.0224563	-0.1123459
3	立地	商店街	-0.1431835	-0.185251	-0.2440313	0.04996564
4	競合店	ない	0.2792342	-0.04109341	-0.1131954	0.1620048
5	競合店	ある	-0.199453	0.02935243	0.08085383	-0.1157177
6	通行量	多い	0.1183509	-0.01741713	-0.1351952	-0.3122078
7	通行量	少ない	-0.08453636	0.01244081	0.09656799	0.2230055

ここでは4つの次元 (Score1~Score4) について、パラメータ
 $\alpha_0 \sim \alpha_7$ を縦方向に表示しています。

ここをクリックして Ctrl-C を押す

No.	Item	PGC1	PGC2	PGC3
0	重相関...	0.820917	0.74380...	0.81030...
1	立地	0.53541 ...	-0.1507 ...	-0.5016 ...
2	競合店	0.76144 ...	0.696878	0.75900...
3	通行量	0.44249...	0.46192...	0.38859...

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.	Item	PGC1	PGC2	PGC3	PGC4
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

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。

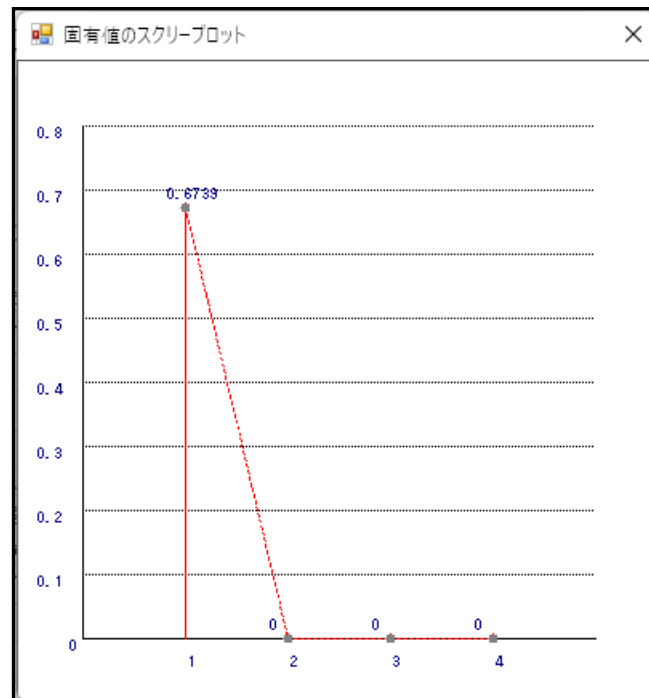
いずれも、推定の精度を示すもので、1.0に近いほど当てはまりの程度がよいことを示します。

ここでも4つの次元（PGC1～PGC4）について表示しています。

一番左の列（PGC1）に注目すると、重相関係数は0.82で、全体としての精度は良いとしても、偏相関係数は通行量については0.44ですので、精度として高いとはいえません。

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。



理論上は4つの自由度で固有値は計算されますが、この図を見ると有効な固有値は1つとなります。

言い換えると、現象の表現に有効な指標の数は1つということです。

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。

数量化結果の表示

表示次元数: 2 分布表示

	No.	I D	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...
	5	ID-5	黒字	0.0170...	0.55550...	0.0070...	0.25524...

グリッドの表示枠が小さい為、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。

ここをクリックして Ctrl-C を押す



数量化結果の表示

表示次元数 : 2 分布表示

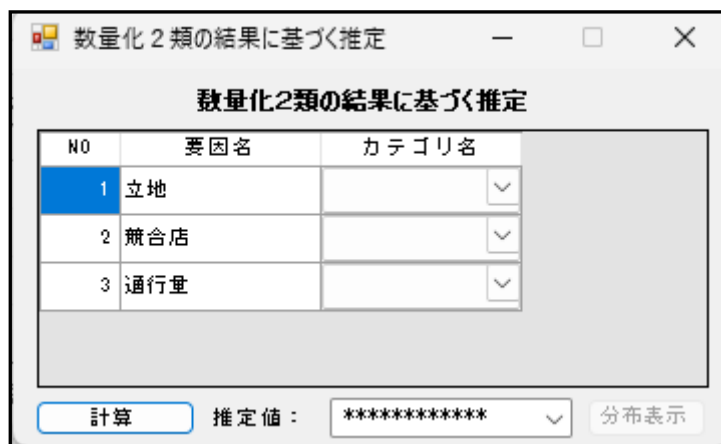
No.	ID	収益	1	2	3	4
1	ID-1	黒字	-0.4786 ...	0.07044 ...	0.19404 ...	-0.2777 ...
2	ID-2	黒字	-0.1141 ...	0.63576 ...	-0.4591 ...	-0.1788 ...
3	ID-3	黒字	1.38777 ...	-3.4694 ...	2.77555 ...	0
4	ID-4	黒字	-0.5135 ...	0.07557 ...	-0.4489 ...	0.51866 ...
5	ID-5	黒字	-0.3170 ...	0.66562 ...	-0.2273 ...	0.3563 ...

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が実行されます。



数量化2類の結果に基づく推定

NO	要因名	カテゴリ名
1	立地	
2	競合店	
3	通行量	

計算 推定値 : ***** 分布表示

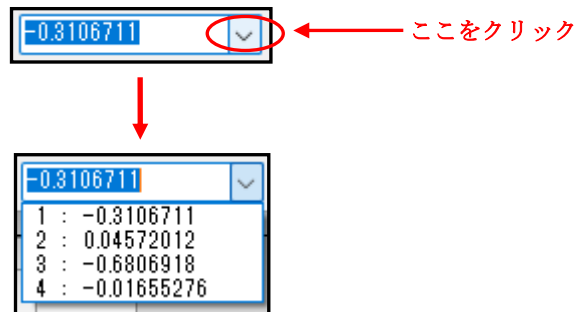
各要因について定義し、計算ボタンを押すと、

数量化2類の結果に基づく推定

NO	要因名	カテゴリ名
1	立地	商店街
2	競合店	ない
3	通行量	多い

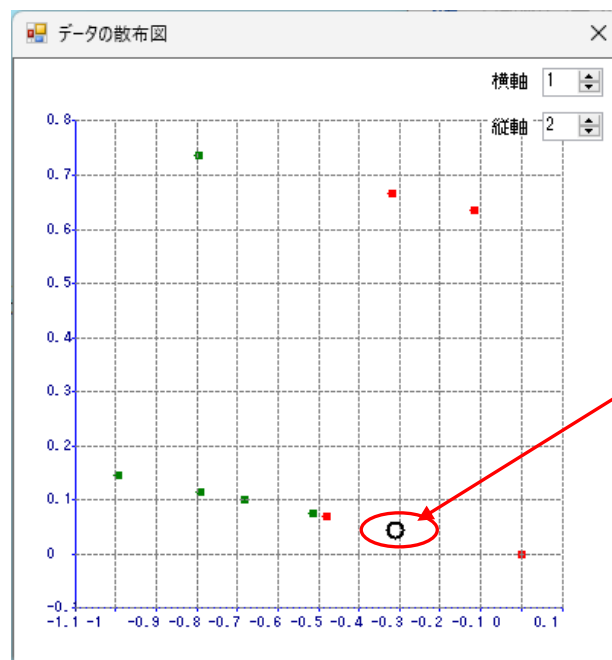
計算 推定値: -0.3106711 分布表示

推定値の欄に 数字が表示されます。



分布表示

をクリックすると、グラフ上に表現されます。



定義した対象の
位置が黒丸で
表示される