

因子分析

1. 目的

例題を提示して説明します。

20人の新入社員の人物評定が示されています。

各項目について 10点満点で 評価されたものです。

No.	氏名	所属	外向的	社交的	積極的	論理的	正確さ	冷静さ
1	浅野	研究	6	6	7	8	7	7
2	江川	研究	6	5	7	5	6	6
3	及川	開発	6	7	5	4	6	5
4	香取	開発	5	7	6	5	5	5
5	木下	研究	6	6	6	7	6	6
6	木場	開発	6	5	6	6	5	5
7	佐藤	開発	5	5	6	5	4	5
8	鈴木	開発	6	6	5	5	6	5
9	瀬川	開発	6	6	6	4	4	5
10	立川	開発	6	6	7	4	5	5
11	千城	営業	5	5	4	4	5	3
12	土屋	営業	6	6	7	6	4	4
13	名取	営業	4	6	6	3	5	4
14	西川	研究	4	5	5	5	6	6
15	野村	開発	5	4	4	5	7	5
16	羽山	研究	5	5	4	5	5	6
17	船橋	開発	3	4	4	5	4	4
18	穂田	研究	5	6	4	5	6	6
19	前川	研究	4	4	3	6	5	6
20	目黒	開発	5	3	4	3	5	4

評価内容は

- ・外向的
- ・社交的
- ・積極的
- ・論理的
- ・正確さ
- ・冷静さ

の6つで、各点数を $x_{i,j}$ で表します。

添え字の i は社員、 j は評価項目を表すとします。

上表の場合 $i = 1 \sim 20$, $j = 1 \sim 6$ で、例えば $x_{3,5}$ は
「及川」の「正確さ」の評定 6 を表します。

評価項目は6つですが、実は潜在的に共通な要因があつて、これら6つの評価項目を共通な要因でシンプルに表現出来ないかということを考えます。

例えば、2つの要因（以降 因子といいます）

- ・「人の良さ」因子
- ・「頭の良さ」因子

で説明できないかというようなことです。

今 「人の良さ」因子を f_1 、「頭の良さ」因子を f_2 として、おおむね以下の式で表現します。

$$\begin{array}{ll} \text{外向的} & X_1 \cong a_{1,1} f_1 + a_{2,1} f_2 \\ \text{社交的} & X_2 \cong a_{1,2} f_1 + a_{2,2} f_2 \\ \text{積極的} & X_3 \cong a_{1,3} f_1 + a_{2,3} f_2 \\ \text{論理的} & X_4 \cong a_{1,4} f_1 + a_{2,4} f_2 \\ \text{正確さ} & X_5 \cong a_{1,5} f_1 + a_{2,5} f_2 \\ \text{冷静さ} & X_6 \cong a_{1,6} f_1 + a_{2,6} f_2 \end{array}$$

$a_{k,j}$ ($k = 1 \sim 2, j = 1 \sim 6$) は 各評価項目($j = 1 \sim 6$) に対する各因子($k = 1 \sim 2$)の重みを表します。

実際のデータは 20人分のデータとなりますので、上記の概略式を以下のように数理モデルで詳細化します。

$$\begin{array}{l} i = 1 \sim 20 \text{ として、} \\ x_{i,1} = a_{1,1} f_{i,1} + a_{2,1} f_{i,2} + e_{i,1} \\ x_{i,2} = a_{1,2} f_{i,1} + a_{2,2} f_{i,2} + e_{i,2} \\ x_{i,3} = a_{1,3} f_{i,1} + a_{2,3} f_{i,2} + e_{i,3} \\ x_{i,4} = a_{1,4} f_{i,1} + a_{2,4} f_{i,2} + e_{i,4} \\ x_{i,5} = a_{1,5} f_{i,1} + a_{2,5} f_{i,2} + e_{i,5} \\ x_{i,6} = a_{1,6} f_{i,1} + a_{2,6} f_{i,2} + e_{i,6} \\ \quad (e_{i,1} \sim e_{i,6} \text{ は 誤差を表します}) \end{array}$$

因子分析は、上式の

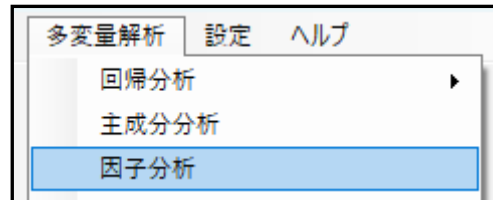
$$\begin{array}{ll} \cdot a_{k,j} & (k = 1 \sim 2, j = 1 \sim 6) \\ \cdot f_{i,k} & (i = 1 \sim 20, k = 1 \sim 2) \\ \cdot e_{i,j} & (i = 1 \sim 20, j = 1 \sim 6) \end{array}$$

を求めることです。

2. 使用法

(1) メニューの選択

メニューの「多変量解析→因子分析」を選択します。



(2) パネルが表示されます。

因子分析

1. 因子数の決定 固有値のスクリーンプット 因子数 : 0 使用法 考え方

2. 因子負荷量の計算 計算実行 回転前分布 回転後分布

	変量	共通性	因子-1	因子-2	因子-3	因子-4	因子-5

3. 因子スコア 因子スコアの表示 誤差表示

	ID	Group	因子-1	因子-2	因子-3	因子-4	因子-5

4. 集計

	項目	数	合計	平均	(標本)分散

表データを貼り付け ☐ 先頭行をラベルとして使用 クリア データ分布 相関行列表示

	NO	ID	Group	X1	X2	X3	X4	X5
▶*								

(3) データの入力

パネルのグリッド（下の部分）にデータを入力します。

☐ 先頭行をラベルとして使用

	N0	ID	Group	X1	X2	X3	X4	X5
▶								

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。

氏名	所属	外向的	社交的	積極的	論理的	正確さ	冷静さ
浅野	研究	6	6	7	8	7	7
江川	研究	6	5	7	5	6	6
及川	開発	6	7	5	4	6	5
香取	開発	5	7	6	5	5	5
木下	研究	6	6	6	7	6	6
木場	開発	6	5	6	6	5	5

データの項目は、

(1) ID

(2) Group （所属を示す文言）

これは何でもいいのですが、種類であったり、所属するクラス
だったり、データの種別を示すものです。

全部 同じ扱いでよければ 同じで構いません。

(3) X1 : データ 1

(4) X2 :

というものです。

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。

4. 集計

	項目	数	合計	平均	(標本)分散
▶	外向的	20	104	5.2	0.76
	社交的	20	107	5.35	1.0275

☒ 先頭行をラベルとして使用

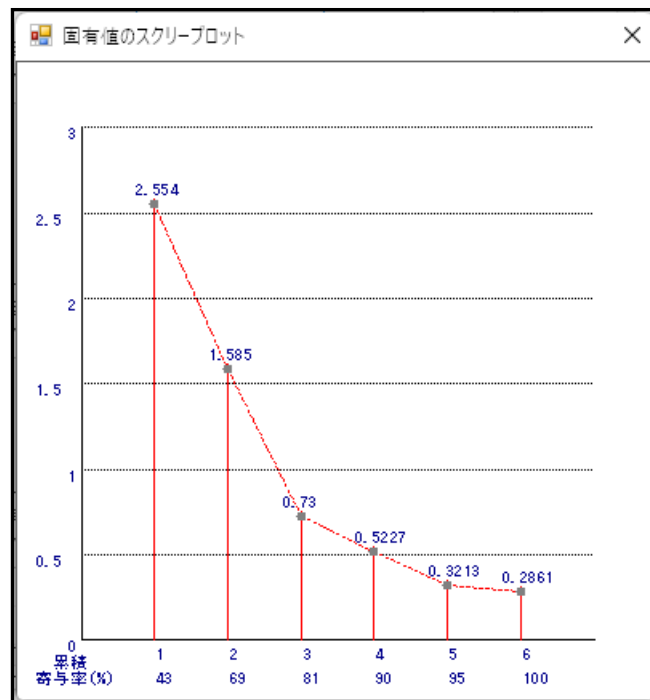
	N0	氏名	所属	外向的	社交的	積極的	論理的	正確さ	冷静さ
▶	1	浅野	研究	6	6	7	8	7	7
	2	江川	研究	6	5	7	5	6	6
	3	及川	開発	6	7	5	4	6	5
	4	香取	開発	5	7	6	5	5	5

下のグリッドには入力データがそのまま反映されます。
上のグリッドには、変量毎（評価項目毎）の集計が表示されます。

（４）固有値のスクリープロット

例題のデータは、６つの変量で表現されています。
したがって、最大６つの因子が定義できます。
しかし６つの変量を６つの因子で説明されても、うれしくありません。
せいぜい２～３程度の少ない因子で、データの特徴が捉えられる方が
良いからです。
ここでは ２つの因子（人の良さ、頭の良さ）で表現しようと考えてます。

固有値のスクリープロットをクリックすると、その目安を判断できます。



６つの変量ですので、６つの固有値が計算されます。
大きい順に、 2.554、1.585、0.73、0.5227、0.3213、0.2861 です。
この場合の固有値は、データの分散を表現していると言えます。
下に累積寄与率（％）が表示されていて、左から 43、69、81、90、95、100
となっています。
これは、最初の 2.554 で、全体の傾向のおおむね 43％が表現できている。
続く 1.585 で、最初と合わせて、全体の傾向の おおむね 69％が表現できて
いることを意味します。

(5) 因子数の決定

デフォルトでは、0 が設定されています。

因子数 : 0

ここで、因子数に 2 を設定します。

因子数の設定をいくらにすべきかは、いろいろ議論がありますが、ここではとりあえず2つの因子で見てみよう ということです、2を設定します。

因子数 : 2

(6) 計算実行

因子数を決定した段階で、**計算実行** を押して、計算実行します。

2. 因子負荷量の計算 **計算実行** 回転前分布 回転後分布

	変数	共通性	因子-1	因子-2	回転後 因子-1	回転後 因子-2
▶	外向的	0.5239613	0.63920	-0.33967	0.69411	0.20537
	社交的	0.4299812	0.52838	-0.38833	0.64910	0.09301
	積極的	0.7431124	0.64056	-0.57689	0.86124	0.03703
	論理的	0.406642	0.56286	0.29972	0.19172	0.60818

3. 因子スコア 因子スコアの表示 誤差表示

	氏名	所属	因子-1	因子-2
▶	浅野	研究	1.09841	1.87246
	江川	研究	0.93899	0.82113
	及川	開発	0.41794	-0.05875
	斎藤	開発	0.69282	-0.97440

上記2つの表で 計算結果が示されました。

いづれも グリッドが小さいため、データの全部を見ることはできません。スクロールして、見ることはできますが、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

	変数	共通性	因子-1	因子-2	回転後 因子-1	回転後 因子-2
▶	外向的	0.5239613	0.63920	-0.33967	0.69411	0.20537
	社交的	0.4299812	0.52838	-0.38833	0.64910	0.09301
	積極的	0.7431124	0.64056	-0.57689	0.86124	0.03703
	論理的	0.406642	0.56286	0.29972	0.19172	0.60818

表計算ソフト上で、Ctrl-V を押すと、データがそのままコピーされます。

変量	共通性	因子-1	因子-2	回転後因子-1	回転後因子-2
外向的	0.5240	0.6392	-0.3397	0.6941	0.2054
社交的	0.4300	0.5284	-0.3883	0.6491	0.0930
積極的	0.7431	0.6406	-0.5769	0.8612	0.0370
論理的	0.4066	0.5629	0.2997	0.1917	0.6082
正確さ	0.3314	0.4301	0.3826	0.0389	0.5744
冷静さ	0.9346	0.7464	0.6144	0.1023	0.9613
寄与率	-	0.3595	0.2021	0.2823	0.2793
固有値	-	2.1570	1.2126	1.6936	1.6761

因子-1、因子-2 は 数理モデルでの（回転前の） $a_{k,j}$ を表しています。
このままでも 利用は可能ですが、2つの因子を回転させて、見た目上
分かりやすくする操作を加えたものが、回転後因子-1、回転後因子-2 に
表示されている（回転後の） $a_{k,j}$ です。

因子負荷量と呼びます。

ここでは、直交回転のバリマックス回転を行っています。

後述の因子スコアの計算で後者の回転後の因子負荷量を利用します。

共通性とは観測されているデータが、共通の2つの因子によって説明できる
度合と考えられます。

通常 1 以下の値を取ります。

上表では各評価項目の共通性について

- ・外向的 0.5240
- ・社交的 0.4300
- ・積極的 0.7431
- ・論理的 0.4066
- ・正確さ 0.3314
- ・冷静さ 0.9346

となっており、「積極的」と「冷静さ」は 共通因子で かなり表現出来そうですが、
「正確さ」は むずかしい ようです。

共通性は1.0以下の値を取りますが、データの条件が不適切な場合に
1.0を超えることがあります。不適切な場合とは、

- ・データのサンプル数が少ない
- ・項目間で相関が高いもの同士がある

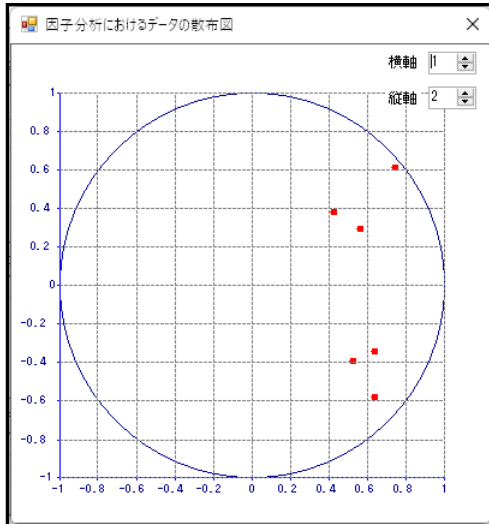
といったことです。

回転前分布

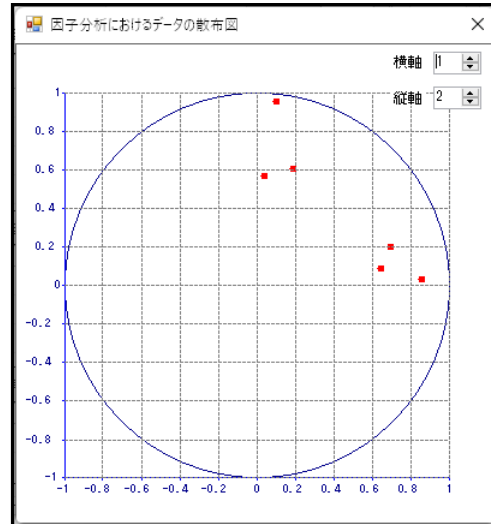
回転後分布

を押すと、因子負荷量を表示できます。

回転前の因子負荷量



回転後の因子負荷量

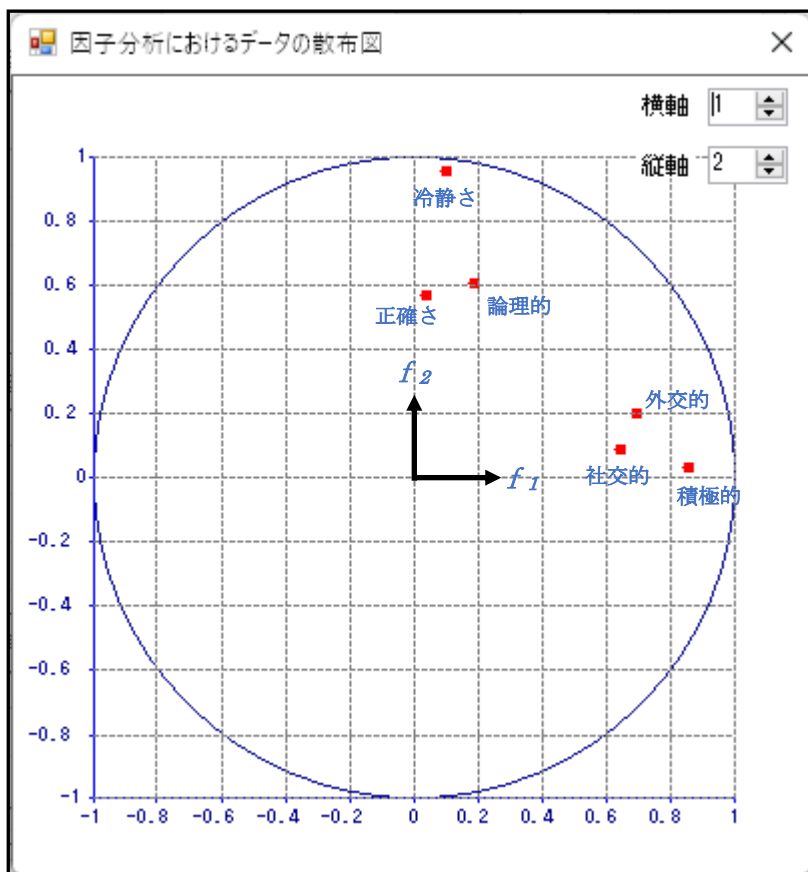


左の因子負荷量 $a_{k,j}$ を解釈しやすいように回転させたものが
右の因子負荷量 になります。

「人の良さ」因子は横軸で、「頭の良さ」因子は縦軸で表されています。

「外交的」、「社交的」、「積極的」が「人の良さ」軸の周りに、

「論理的」、「正確さ」、「冷静さ」が「頭の良さ」軸の周りに散らばっています。



次に説明するのは因子スコアです。

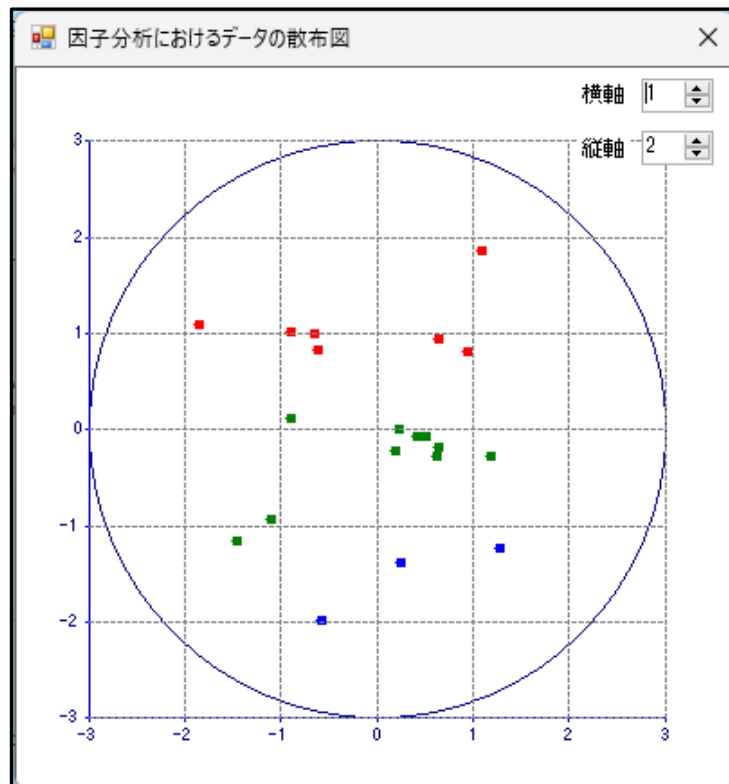
下の表は 因子スコアを表示しています。

因子スコアとは 数理モデルでの $f_{i,k}(i = 1\sim 20, k = 1\sim 2)$ を表します。

$k = 1$ は「人の良さ」因子、 $k = 2$ は「頭の良さ」因子を表し、 $i = 1 \sim 20$ は個々のデータを表しています。

因子スコアの表示

を押すと、因子スコア を表示できます。



赤い点は 研究部に所属する人たちを表し、
緑の点は 開発部に所属する人たちを表し、
青い点は 営業部に所属する人たちを表しています。

誤差表示

を押すと、誤差（独自因子とも呼ばれる）を表示します。

数理モデルでの $e_{i,j}$ ($i = 1 \sim 20, j = 1 \sim 6$) に相当します。

ID	Group	Org-1	...	Org-6	Cmp-1	...	Cmp-6	Err-1	...	Err-6
浅野	研究	6.000	...	7.000	6.200	...	6.904	-0.200	...	0.096
江川	研究	6.000	...	6.000	5.915	...	5.935	0.085	...	0.065
及川	開発	6.000	...	5.000	5.442	...	5.087	0.558	...	-0.087
香取	開発	5.000	...	5.000	5.528	...	4.911	-0.528	...	0.089
木下	研究	6.000	...	6.000	5.761	...	6.014	0.239	...	-0.014
木場	開発	6.000	...	5.000	5.498	...	5.091	0.502	...	-0.091
佐藤	開発	5.000	...	5.000	5.276	...	4.922	-0.276	...	0.079
鈴木	開発	6.000	...	5.000	5.341	...	5.129	0.659	...	-0.129
瀬川	開発	6.000	...	5.000	5.559	...	4.999	0.441	...	0.001
立川	開発	6.000	...	5.000	5.875	...	4.975	0.125	...	0.025
千城	営業	5.000	...	3.000	4.497	...	3.247	0.503	...	-0.247
土屋	営業	6.000	...	4.000	5.755	...	4.102	0.245	...	-0.102
名取	営業	4.000	...	4.000	5.101	...	3.874	-1.101	...	0.126
西川	研究	4.000	...	6.000	4.978	...	5.800	-0.978	...	0.200
野村	開発	5.000	...	5.000	4.679	...	5.127	0.321	...	-0.127
羽山	研究	5.000	...	6.000	4.844	...	5.944	0.156	...	0.056
船橋	開発	3.000	...	4.000	4.115	...	3.914	-1.115	...	0.086
穂田	研究	5.000	...	6.000	4.985	...	5.953	0.015	...	0.047
前川	研究	4.000	...	6.000	4.274	...	5.918	-0.274	...	0.082
目黒	開発	5.000	...	4.000	4.376	...	4.153	0.624	...	-0.153

上表の Org-1～Org-6 は、入力データそのものです。

続く Cmp-1～Cmp-6 は、数理モデルで計算されたもの、

続く Err-1～Err-6 は、入力データ - 数理モデルの誤差です。