

二群以上の関係を把握する方法-2

2019年12月6日
データサイエンス研究所
伊藤嘉朗

1

Data Science Institute

本日の内容

- 主成分分析
主成分負荷量、寄与率
- 因子分析
因子抽出法、因子軸回転
因子負荷量、寄与率、共通性
- クラスター分析
階層的クラスター分析
非階層クラスター分析

2

Data Science Institute

主成分分析

◇英語, 一般常識, 数学, 技術知識の4つの試験科目の得点

	英 語	一般常識	数 学	技術知識
A	92	87	75	63
B	85	83	67	70
C	73	71	65	75
D	76	95	55	63
E	61	60	100	98
F	79	73	83	89
...

3

主成分分析 (SAS EG)

EG

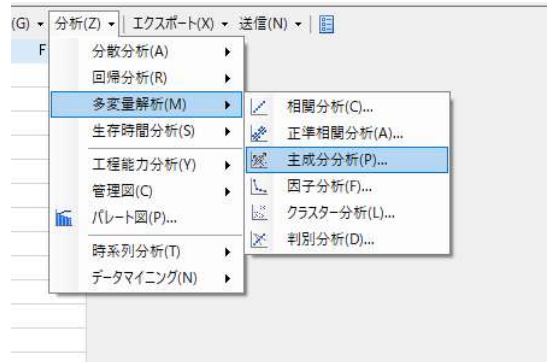
1) データを入力する。



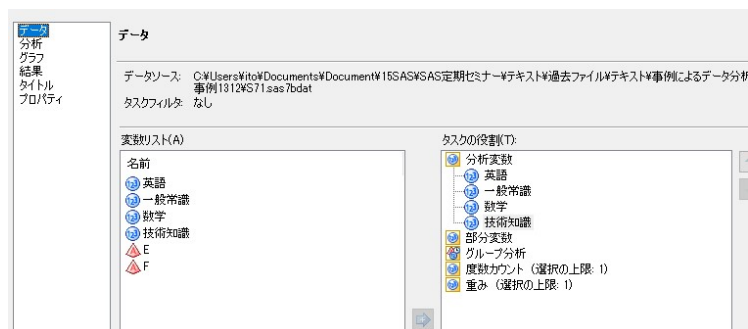
	英語	一般常識	数学	技術知識	E
1	92	87	75	63	
2	85	83	67	70	
3	73	71	65	75	
4	76	95	55	63	
5	61	60	100	98	
6	62	59	100	95	
7	79	73	83	89	
8	50	55	62	59	
9	89	90	92	91	
10	100	98	60	67	
11	89	90	80	80	
12	75	70	78	82	
13	50	52	63	69	
14	44	43	39	45	
15	93	90	89	90	
16	39	20	45	50	
17	63	70	25	30	
18	45	47	48	42	
19	51	49	75	77	
20	55	70	39	44	
21	92	87	75	63	
22	85	78	67	70	
23	73	71	65	75	

4

2) 「分析」-「多変量解析」-「主成分分析」を選択する。



3) 「データ」をクリックし、英語、一般常識、数学、技術知識を「分析変数」に設定する。



4) 「結果」をクリックし、「元のデータと主成分スコア…」を設定する。

7

主成分分析 結果

EG

	英語	一般常識	数学	技術知識
英語	1.0000	0.9323	0.3907	0.4144
一般常識	0.9323	1.0000	0.2980	0.3245
数学	0.3907	0.2980	1.0000	0.9653
技術知識	0.4144	0.3245	0.9653	1.0000

相関行列から

4つの科目は、英語と一般常識および数学と技術知識の2つのグループに分けられる。

8

	固有値	差	比率	累積
1	2.66457695	1.42601903	0.6661	0.6661
2	1.23855792	1.17598922	0.3096	0.9758
3	0.06256870	0.02827228	0.0156	0.9914
4	0.03429642		0.0086	1.0000

固有値 > 1以上 ⇒ 2個

	PRIN1	PRIN2	PRIN3	PRIN4
英語	0.511296	0.467672	-0.720935	-0.10604
一般常識	0.475712	0.544470	0.690118	0.031424
数学	0.501382	-0.502830	0.019046	0.703858
技術知識	0.510775	-0.481661	0.060230	-0.709566

- a) 第1主成分負荷量はすべてプラスの大きな値
⇒ 「基礎学力」
- b) 第2主成分負荷量は事務系の科目にプラスの値、
技術系の科目にマイナスの値
⇒ 「事務系対技術系学力」

「出カデータ」をクリックすると、主成分スコアが表示される。

	英語	一般常識	数学	技術知識	E	F	PRIN1	PRIN2	PRIN3	PRIN4
1	92	87	75	63			1.09386764	1.066187893	-0.255193729	0.5140362427
2	85	78	67	70			0.6806653205	0.6656801432	-0.276196948	-0.034815318
3	73	71	65	75			0.2703680109	0.096261691	-0.035098653	-0.295439809
4	76	95	55	63			0.3492659978	1.3721302699	0.6220163338	-0.155138257
5	61	60	100	98			1.1620083173	-1.952551597	0.1609936262	0.0407745537
6	62	59	100	95			1.0850152485	-1.877862342	0.0783998736	0.1511418821
7	79	73	83	89			1.3038135762	-0.496943239	-0.138568985	-0.198995949
8	50	55	62	59			-1.24703435	-0.434560928	0.2579475416	0.1891434034
9	89	90	92	91			2.2560349893	-0.06072256	0.0691353679	0.057665314
10	100	98	60	67			1.3128731724	1.8322952402	-0.191116057	-0.141526445
11	89	90	80	80			1.6638410074	0.5153988976	0.0229054252	0.0556512626
12	75	70	78	82			0.8107575933	-0.379268606	-0.112654216	-0.111589115
13	50	52	63	69			-1.028262629	-0.794782291	0.1880491402	-0.155981032
14	44	43	39	45			-2.639401126	0.0140376055	0.0145327803	-0.095603406
15	93	90	89	90			2.2655312872	0.139556859	-0.092285395	-0.010749532
16	39	20	45	50			-3.037238987	-1.008389705	-0.555614346	-0.108826914
17	63	70	25	30			-2.227912256	1.9495537415	0.1388889599	0.0145010025
18	45	47	48	42			-2.376997689	0.0015324261	0.1113973222	0.3384339295
19	51	49	75	77			-0.555240434	-1.350912432	0.0831829312	-0.04671616
20	55	70	39	44			-1.726139535	1.045245771	0.5076046801	-0.022322756

4つの変量の持つ情報を2つの主成分に縮約
情報を縮約する働きが主成分分析の重要な特徴

①主成分分析は、もとの幾つかの変量に共通して含まれる幾つかの要素を抽出するための統計的手法。

②この方法によって抽出される要素を主成分という。
主成分は互いに無相関である。

11

Data Science Institute

因子分析

因子分析とは

- 複数の項目間に共通して作用する潜在的な要因を探し出す。
- 相互の関連を数個の因子で説明する。

例) ブランドのイメージ調査

- アンケートからブランドイメージを構成する要因
- サービス、店のレイアウト、接客態度…に対する顧客の評価

12

Data Science Institute

(例) 採用試験の成績

英語、一般常識、数学、技術知識の成績

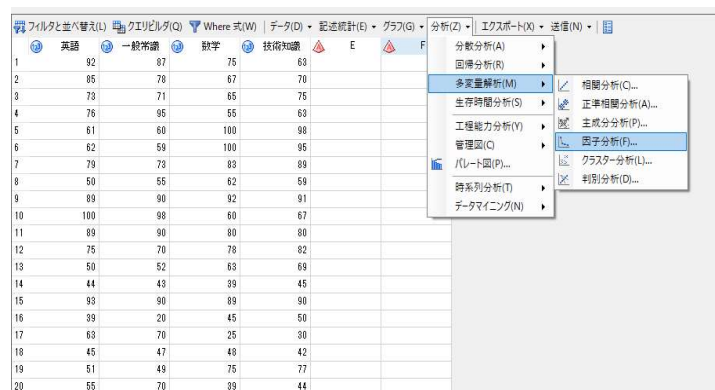
	英語	一般常識	数学	技術知識	E
1	92	87	75	63	
2	85	78	67	70	
3	73	71	65	75	
4	76	95	55	63	
5	61	60	100	98	
6	62	59	100	95	
7	79	73	83	89	
8	50	55	62	59	
9	89	90	92	91	
10	100	98	60	67	
11	89	90	80	80	
12	75	70	78	82	
13	50	52	63	69	
14	44	43	39	45	

13

Data Science Institute

因子分析 (SAS EG)

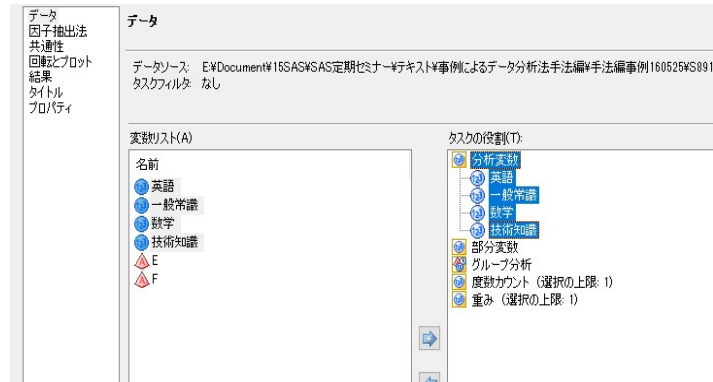
1) 「分析」-「多変量解析」-「因子分析」を選択する。



14

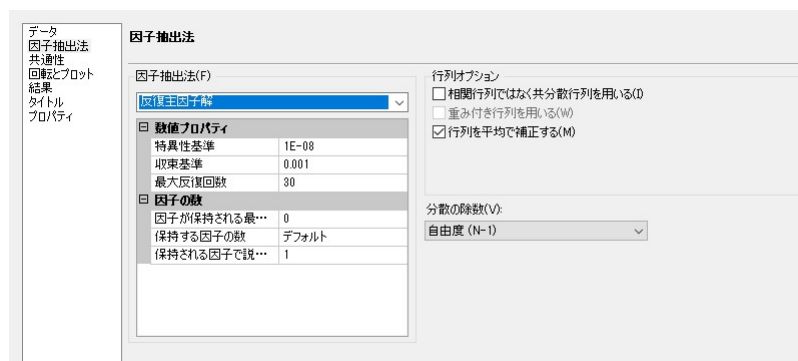
Data Science Institute

2) 「データ」をクリックし、「分析変数」を設定する。



15

3) 「因子抽出法」をクリックし、「反復主因子解」を選択する。



16

- 4) 「回転とプロット」をクリックし、「回転方法」に「直交バリマックス」を選択し、「表示するプロット」に「因子パターン」「回転前の因子パターン」を設定する。

データ
因子抽出法
共通性
回転とプロット
結果
タイトル
プロパティ

回転とプロット

回転方法(O)
直交/バリマックス

表示するプロット
 因子パターン (回転後)(A)
 回転前の因子パターン (回転前)(F)
 固有値のスクーププロット(E)

選択できるオプションはありません

すべての因子を使用する(L)
 プロットする因子の数(N):

17

- 5) 結果が表示される。

① 因子数の決定

	固有値	差	比率	累積
1	2.66457695	1.42601903	0.6661	0.6661
2	1.23855792	1.17598922	0.3096	0.9758
3	0.06256870	0.02827228	0.0156	0.9914
4	0.03429642		0.0086	1.0000

固有値が1以上の因子 = 2つ

18

② 回転前の因子負荷量（因子パターン）

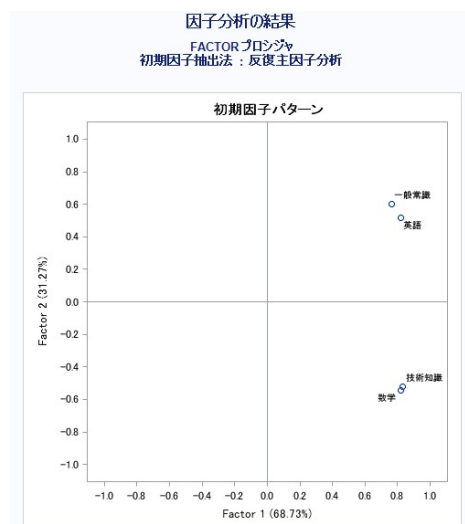
因子パターン		
	Factor1	Factor2
英語	0.81849	0.51504
一般常識	0.76237	0.59964
数学	0.81910	-0.54381
技術知識	0.83381	-0.51963

因子の分散	
Factor1	Factor2
2.6172962	1.1905792

最終的な共通性の推定値：合計 = 3.807875				
英語	一般常識	数学	技術知識	
0.93519744	0.94077733	0.96664541	0.96525517	

19

③ 回転前の因子負荷量の散布図



20

④ 回転後の因子負荷量

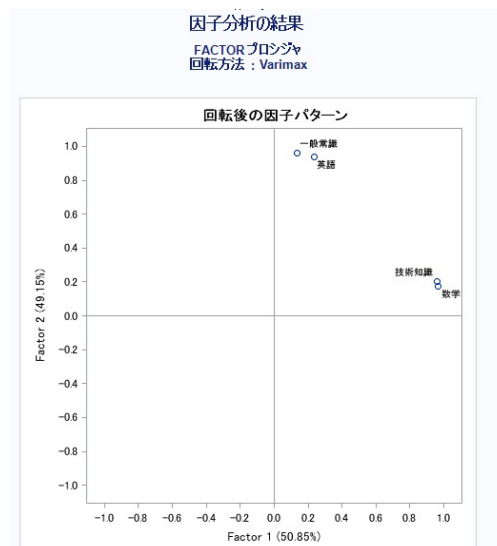
回転後の因子パターン		
	Factor1	Factor2
英語	0.23596	0.93783
一般常識	0.13693	0.96022
数学	0.96790	0.17270
技術知識	0.96183	0.20034

因子の分散	
Factor1	Factor2
1.9363696	1.8715058

最終的な共通性の推定値 : 合計 = 3.807875				
英語	一般常識	数学	技術知識	
0.93519744	0.94077733	0.96664541	0.96525517	

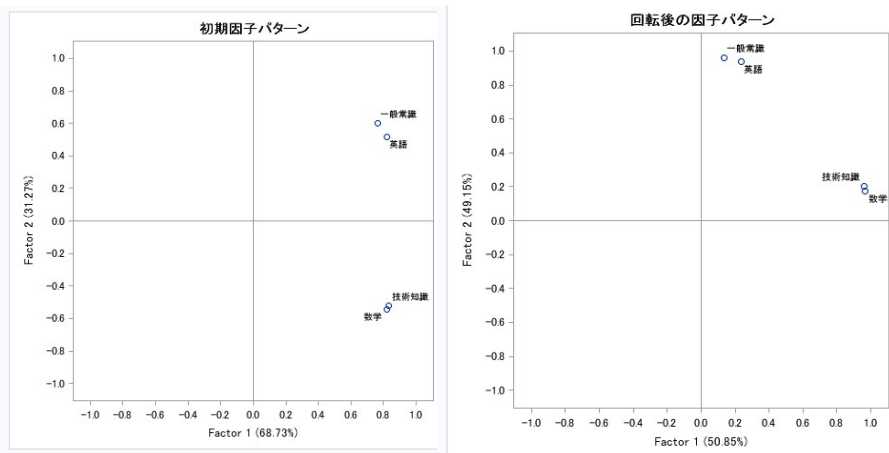
21

⑤ 回転後の因子負荷量の散布図



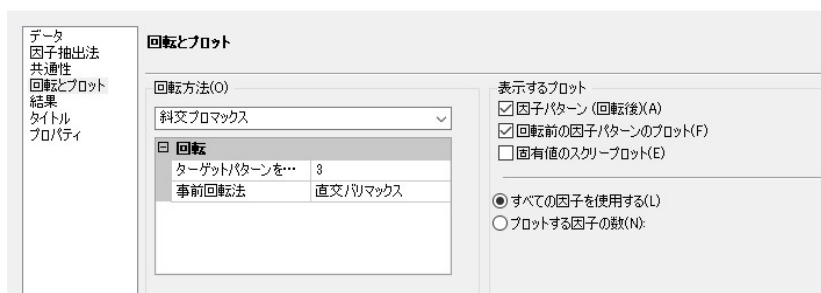
22

直交バリマックス回転前後の比較



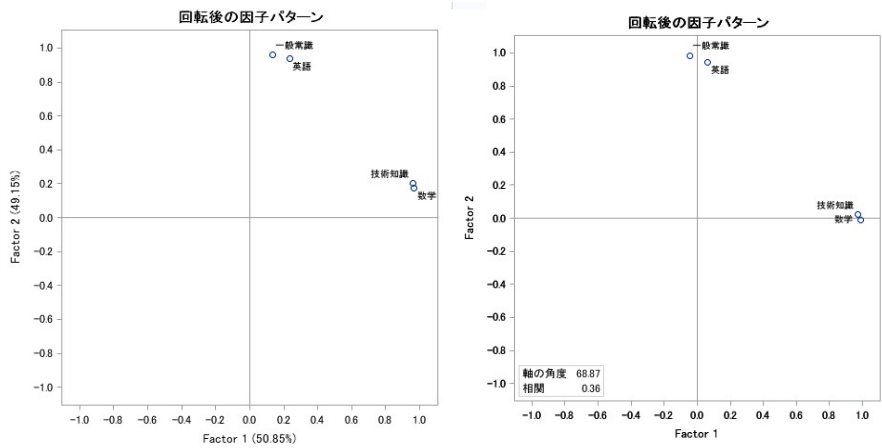
23

回転方法を「斜交プロマックス」に変更する。



24

回転方法の比較



直交バリマックス

斜交プロマックス

25

因子抽出法

1. 最尤法

モデルの尤度(尤もらしさ)を最大にする方法。
標本の大きさが大きいと最も良い推定法と言われる。
変数の正規性の前提があげられる。
推定精度が高い分、不適解も出やすいのが問題点。

2. 反復主因子法

古典的には良く用いられた方法。
安定した解が得られる。

26

回転方法

- 直交回転（バリマックス回転）
因子間に相関関係を仮定しない場合に選択する。
- 斜交回転（プロマックス回転）
因子間に相関関係を仮定する場合に選択する。

27

•固有値

因子数を決定する時に参考にする。
一般に固有値が1以上の因子を採用する。
(カイザー・ガットマン基準)

•因子負荷量（因子パターン）

各因子と各評定項目との関係の強さを示す。

28

共通性

因子によって説明される程度を示す指標。
(因子負荷量の2乗和)

- ・ 評価項目の取舍選択の際に、基準の1つとして使用。
- * 共通性の低い(例0.2未満)がある場合は、削除し、再び因子分析を行うのが一般的である。

因子寄与率 (因子の分散)

因子の説明力を示す指標。
(因子負荷量の2乗和)

* 斜交回転の場合は、使用できない。

29

Data Science Institute

(例) 日本、中国におけるブランドイメージ調査

- ・ 調査対象地域
日本：首都圏 中国：上海市
- ・ 調査対象企業ブランド
Microsoft Apple SONY Panasonic Lenovo
- ・ 調査方法
SD法によるアンケート調査
評価項目：30 尺度：7段階評価

30

Data Science Institute

◇SD法 (Semantic Differential Method)

	1	2	3	4	5	6	7	
嫌いな	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	好きな

非常に…1 かなり…2 やや…3 どちらでもない…4
やや…5 かなり…6 非常に…7

◇調査項目

- 1) 性別、 2) 年齢階層 (20歳代、30歳代…)
- 3) 形容詞イメージ評定15項目
- 4) 経営資源イメージ評定15項目

31

形容詞イメージ評定項目 (15項目)

「軽い～重い」	「暖かい～冷たい」
「女性的な～男性的な」	「モダンな～クラシックな」
「やわらかい～堅い」	「派手な～地味な」
「あっさりした～どい」	「弱い～強い」
「好きな～嫌いな」	「若々しい～ふけた」
「美しい～みにくい」	「明るい～暗い」
「しゃれた～野暮な」	「活発な～落ち着いた」
「楽しい～つまらない」	

32

経営資源イメージ評定項目15項目

「親近感」 「特徴的」 「高くても買いたい」
 「信頼性」 「先進性」 「国際性」
 「成長性」 「安定性」 「高級感」
 「個性」 「品質」 「技術力」
 「環境保護」 「経営理念」 「顧客大切度」

33

Data Science Institute

質問紙

『Microsoft』社についてお伺いします

Q1. 『Microsoft』社のイメージについてお伺いします。(ご存じない方は回答しないで下さい。)

下の言葉の組み合わせ(左右の形容詞)で、『Microsoft』社にふさわしいと思うのはどれでしょう。
 例えば、A(軽い……重い)で、「非常に軽い」と思われれば1に、「やや重い」と思われれば5に
 ○をつけてください。(それぞれ○はひとつ)

		非常に 軽い	かなり 軽い	どちら もやや	でも やや	かなり 重い	非常に 重い			
		1	2	3	4	5	6	7		
A	軽い								重い	A
B	洗練された								粗雑な	B
C	女性的な								男性的な	C
D	革新的な								保守的な	D
E	やわらかい								堅い	E
F	地味な								洗練された	F
G	あっさりした								くどい	G

34

Data Science Institute

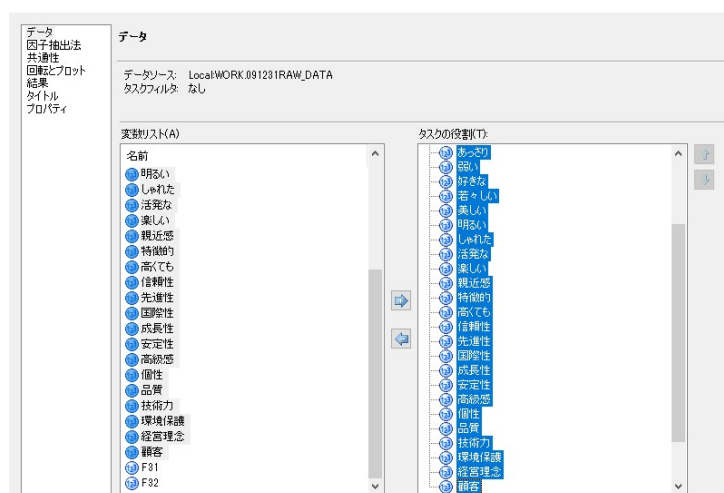
因子分析 (SAS EG)

- 1) データを入力する。
- 2) 「分析」-「多変量解析」-「因子分析」を選択する。



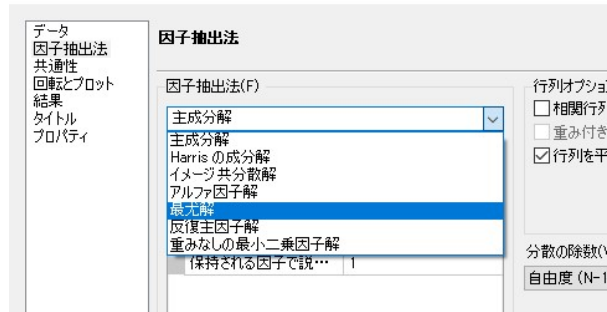
35

3) 分析変数を設定する。



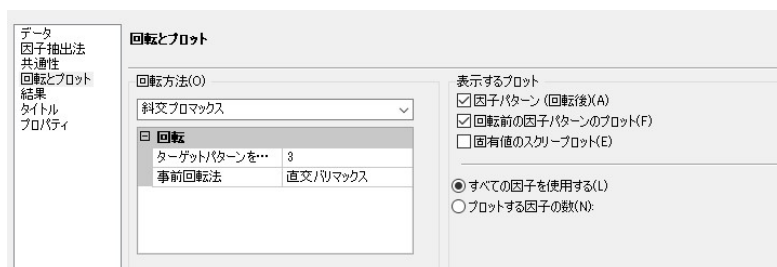
36

4) 「因子抽出法」をクリックし、「最尤解」を選択する。



37

5) 「回転とプロット」をクリックし、「回転方法」に「斜交プロマックス」を選択し、「表示するプロット」に「因子パターン」「回転前の因子パターン」を設定する。



38

6) 「結果」をクリックし、「因子パターン・大きい順に表示」を設定する。



39

① 固有値から因子数を設定する。

固有値の初期値 : 合計 = 40.210175 平均 = 1.34033917

	固有値	差	比率	累積
1	22.9121481	13.3104932	0.5698	0.5698
2	9.6016549	6.1046088	0.2388	0.8086
3	3.4970462	1.8205210	0.0870	0.8956
4	1.6765251	0.0809675	0.0417	0.9373
5	1.5955577	0.4494703	0.0397	0.9769
6	1.1460873	0.1339200	0.0285	1.0054
7	1.0121673	0.2668054	0.0252	1.0306
8	0.7453619	0.1330735	0.0185	1.0492
9	0.6122884	0.1288639	0.0152	1.0644

固有値1以上の因子は7つ。

40

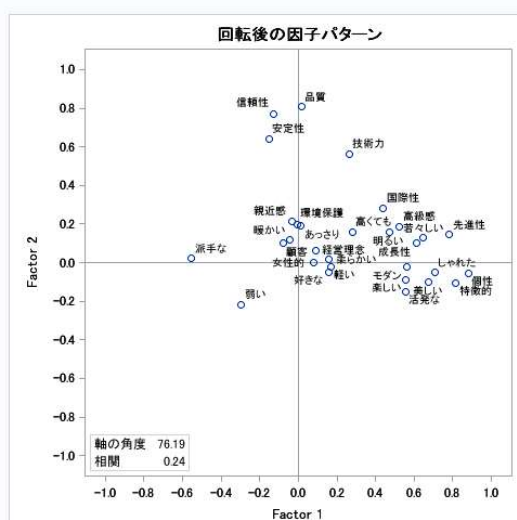
② 因子負荷量の検討

回転後の因子パターン (標準化回帰係数)						
	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
個性	0.88115	-0.05848	0.07409	0.18511	-0.22542	-0.06929
特徴的	0.81634	-0.10870	-0.00344	0.10863	0.00952	-0.01819
先進性	0.78324	0.14665	-0.05517	0.00092	-0.02839	0.20152
しゃれた	0.70612	-0.05236	0.02362	-0.06944	0.07699	-0.08954
美しい	0.67132	-0.10361	-0.02215	-0.10735	0.30859	-0.01338
若々しい	0.64739	0.12703	0.25318	-0.30242	-0.00024	0.12146
成長性	0.61291	0.10310	-0.10133	0.08226	0.12197	0.28454
モダン	0.56389	-0.02152	0.32828	0.05796	-0.11491	-0.01272
楽しい	0.55826	-0.09107	0.12897	0.04752	0.38105	0.04043
活発な	0.55360	-0.15004	0.23377	0.13707	-0.00867	-0.16438
高級感	0.52393	0.18464	-0.22386	0.00361	0.05103	-0.06361
明るい	0.47168	0.15494	0.27091	-0.11391	0.15549	-0.08848
国際性	0.44055	0.27918	-0.09020	0.01185	-0.10022	0.10391
高くても	0.28266	0.15794	-0.04396	0.07876	0.25978	0.02649
派手な	-0.55545	0.02057	-0.22247	0.14230	0.03307	0.29039
品質	0.01761	0.80673	0.11317	0.15007	-0.03949	0.02263
信頼性	-0.12933	0.77043	0.06886	0.07408	0.08717	0.00615
安定性	-0.15249	0.64123	-0.14647	0.01318	0.04028	0.04869
技術力	0.26307	0.56087	-0.01765	0.14698	-0.03734	0.14138
柔らかい	0.16017	0.01879	0.76523	0.03342	0.10640	0.10335
女性的	0.07672	-0.00214	0.68986	0.09287	-0.01674	0.11389
軽い	0.16611	-0.02111	0.60016	-0.10979	-0.03182	0.36707
弱い	-0.29864	-0.21928	0.54499	0.06513	-0.02022	0.49355
暖かい	-0.04727	0.11663	0.44307	0.11064	0.30895	-0.05015
経営理念	0.08902	0.06107	0.00078	0.71878	0.06662	-0.10795
顧客	-0.07873	0.10245	0.07306	0.64250	0.18713	-0.08253
環境保護	-0.00698	0.19914	0.07169	0.60475	-0.01322	-0.10623
好きな	0.15475	-0.05006	-0.03670	0.13177	0.80108	0.07164
親近感	-0.03200	0.21213	0.13116	0.11460	0.60060	0.01749
あっさり	0.01433	0.18952	0.14825	-0.22234	0.06182	0.36092

41

因子分析の結果

FACTORプロシヤ
回転方法 : Promax (power = 3)



42

因子の命名

第1因子

「個性」「特徴的」「先進性」
「しゃれた」「美しい」



『個性因子』

第2因子

「品質」「信頼性」「安定性」



『信頼因子』

43

中国における調査結果

固有値（中国）

	固有値	差	比率	累積
1	28.8750209	24.1373084	0.7333	0.7333
2	4.7377125	1.7771242	0.1203	0.8536
3	2.9605883	1.0660901	0.0752	0.9288
4	1.8944982	0.7448043	0.0481	0.9769
5	1.1496939	0.1094486	0.0292	1.0061
6	1.0402454	0.4097824	0.0264	1.0325
7	0.6304630	0.1024761	0.0160	1.0485
8	0.5279869	0.0834977	0.0134	1.0619
9	0.4444892	0.0494846	0.0113	1.0732

44

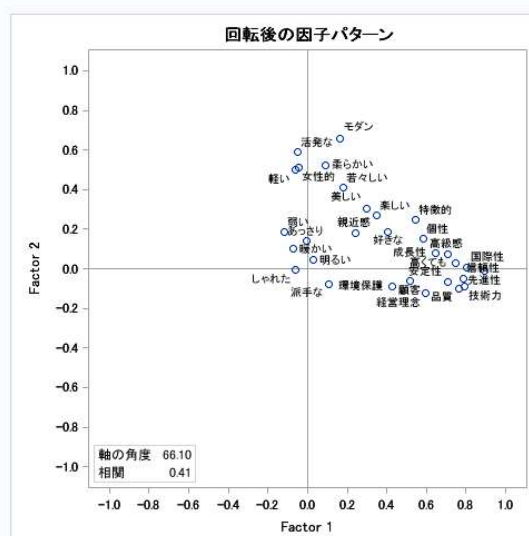
因子負荷量 (中国)

	回転後の因子パターン (標準化回帰係数)				
	Factor1	Factor2	Factor3	Factor4	Factor5
先進性	0.89484	-0.01107	-0.04194	-0.01436	-0.11852
国際性	0.80562	0.00582	0.07231	-0.13358	-0.13682
技術力	0.79287	-0.09208	-0.03982	0.06507	0.02908
信頼性	0.78701	-0.05195	-0.03501	0.05663	0.07210
品質	0.76320	-0.10344	0.01665	0.06275	0.14181
高くても	0.74599	0.03040	-0.09941	0.08070	-0.05947
安定性	0.70721	-0.06488	-0.01373	0.06929	-0.15459
高級感	0.70608	0.07190	0.03042	-0.01467	0.04298
成長性	0.64354	0.07702	0.03359	-0.03896	0.10383
経営理念	0.59742	-0.12087	0.02915	-0.01586	0.40075
個性	0.58253	0.15230	0.09068	-0.12081	0.23409
特徴的	0.54582	0.24846	-0.05370	-0.10889	0.10733
好きな	0.40638	0.18816	-0.16948	0.04319	0.23130
楽しい	0.34613	0.26939	-0.11587	-0.03720	0.29138
モダン	0.16564	0.65926	-0.05394	-0.07144	-0.16399
活発な	-0.05027	0.58813	0.16787	0.07435	0.12311
柔らかな	0.09008	0.52384	-0.11044	0.34346	-0.10069
女性的	-0.04450	0.51099	0.20188	0.24499	-0.22548
軽い	-0.06275	0.49986	-0.01779	0.00810	0.22941
若々しい	0.17983	0.40797	-0.12220	-0.05937	0.24689
美しい	0.29839	0.30600	-0.12207	-0.06368	0.24329
明るい	0.03062	0.04756	0.87148	-0.00063	-0.05013
しゃれた	-0.06371	-0.00314	0.79856	0.10346	0.13798
暖かい	-0.07362	0.10140	0.66493	-0.09715	-0.16738
派手な	0.10736	-0.07795	0.06433	0.74299	-0.02755
弱い	-0.11826	0.18528	-0.01677	0.65468	0.03367
あっさり	-0.00759	0.14120	-0.05497	0.60569	0.13912
環境保護	0.42499	-0.09205	-0.00330	0.11474	0.48789
顧客	0.51876	-0.06432	0.03188	0.01218	0.48423
親近感	0.23897	0.17757	-0.09682	0.03127	0.42744

45

因子分析の結果

FACTORプロシジャ
回転方法: Promax (power = 3)



46

因子の命名（中国）

第1因子

「技術力」「信頼性」「品質」
「高くても」「安定性」「高級感」



『経営基盤因子』

47

Data Science Institute

クラスター分析

クラスター分析とは、異なる性質のものが混ざりあっている集団（対象）の中から互いに似たものを集めて集落（クラスター）を作り、対象を分類しようという方法を総称したもの。

- ・変数クラスター分析（変数をグルーピング）
- ・サンプルクラスター分析

- ・階層的クラスター分析
- ・非階層的クラスター分析

48

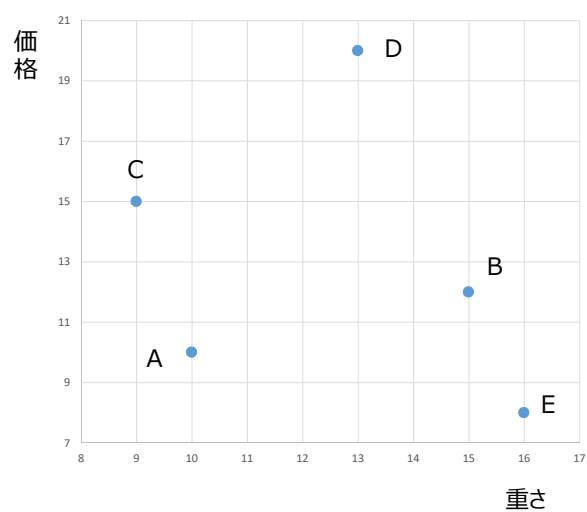
Data Science Institute

階層的クラスター分析

◇AからEの5つの企業が製造している製品を
クラスター分析（群平均法）により分類する。

企業	価格 (万円)	重さ (kg)
A	10	10
B	12	15
C	15	9
D	20	13
E	8	16

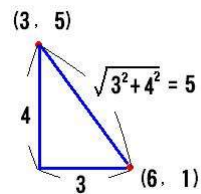
49



50

距離の計算法 (例)

- 2点の距離は三平方の定理で計算



- クラスター分析は複数の点の集合同士の距離を計算する必要。
⇒ 群平均法

51

- 1) 最初に5つの企業の距離をすべて計算する。

	A	B	C	D
B	5.39			
C	5.10	6.71		
D	10.44	8.25	6.40	
E	6.32	4.12	9.90	12.37

一番距離が短いのはBとEの4.12であるので、この2つの企業をまとめて1つのクラスターにする。

52

2) BEをクラスターとし、距離をすべて計算する。

	A	C	D
C	5.10		
D	10.44	6.40	
BE	5.90	8.31	10.31

AとBEの距離 : $(5.39 + 6.32) \div 2 = 5.90$
CとBEは8.31、DとBEは10.31

一番距離が短いのはAとCの5.10であるので、
この2企業をまとめて1つのクラスターにする。

53

Data Science Institute

3) BE、ACをクラスターにし、距離を計算する。

	D	BE
BE	10.31	
AC	8.42	7.08

BEとACの距離 : $(5.39+6.71+6.32+9.90) \div 4 = 7.08$

一番距離が短いのはBEとACの7.08であるので、
この4企業をまとめて1つのクラスターにする。

54

Data Science Institute

4) BEACを1つのクラスターにしてDとの距離を計算する。

	D
BEAC	9.37

DとBEACの距離: $(10.44+8.25+6.40+12.37)\div 4=9.37$

- ① BE
- ② AC
- ③ BEAC

55

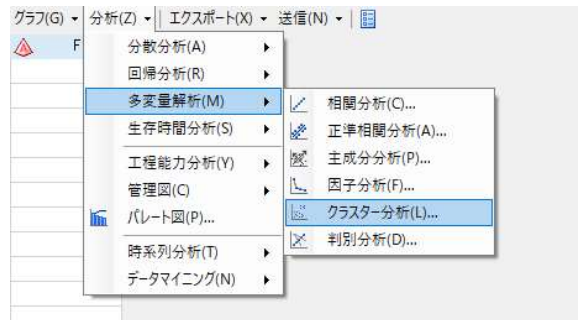
クラスター分析 (SAS EG)

1) データを入力する。

	企業	価格	重さ	D
1	A	10	10	
2	B	12	15	
3	C	15	9	
4	D	20	13	
5	E	8	16	
6		.	.	
7				

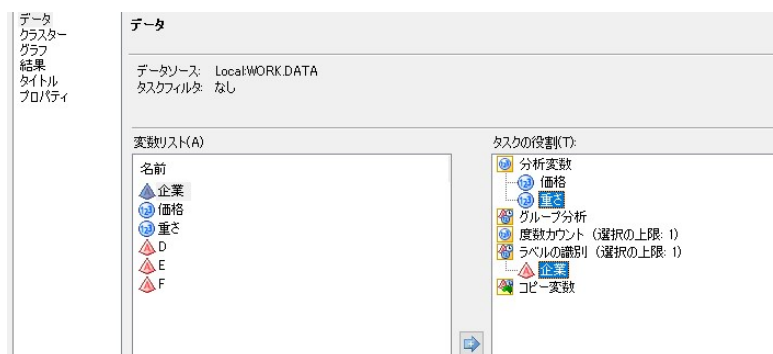
56

2) 「分析」-「多変量解析」-「クラスター分析」を選択する。



57

3) 「データ」をクリックし、価格、重さを「分析変数」に、「企業」を「ラベルの識別」に設定する。



58

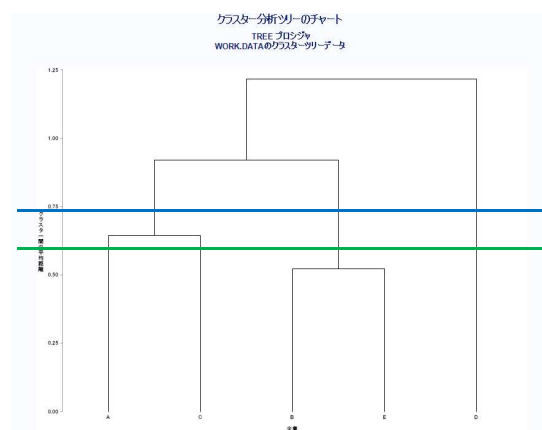
4) 「クラスター」をクリックし、「群平均法」を選択する。



59

5) デンドログラム（樹形図）が表示される。

EG



60

◇携帯電話15機種について、下記8項目の満足度（5点満点）を調査。（各機種の平均値）

商品名	価格	性能	デザイン	操作性	バッテリー	携帯性	拡張性	付属品
A	2.43	3.36	3.04	3.21	2.35	3.90	3.45	4.22
B	3.95	4.55	2.54	3.38	4.43	1.95	4.83	2.62
C	4.26	2.77	3.22	2.24	4.12	2.90	4.52	3.77
D	3.82	4.21	0.38	3.55	4.94	2.12	4.64	3.38
E	1.11	3.23	2.02	4.61	3.22	3.96	2.34	2.82
F	3.13	3.19	2.31	3.22	4.52	4.15	4.78	3.61
...
O	4.22	3.88	3.19	2.12	3.05	2.39	3.24	3.72

61

クラスター分析（SAS EG）

1) データを入力する。

	機種	価格	性能	デザイン	操作性	バッテリー	携帯性	拡張性	付属品
1	A	2.43	3.36	3.04	3.21	2.35	3.9	3.45	4.22
2	B	3.95	4.55	2.54	3.38	4.43	1.95	4.83	2.62
3	C	4.26	2.77	3.22	2.24	4.12	2.9	4.52	3.77
4	D	3.82	4.21	0.38	3.55	4.94	2.12	4.64	3.38
5	E	1.11	3.23	2.02	4.61	3.22	3.96	2.34	2.82
6	F	3.13	3.19	2.31	3.22	4.52	4.15	4.78	3.61
7	G	3.64	3.56	2.29	2.69	2.27	2.29	4.26	4.72
8	H	2.74	3.1	4.47	3.2	4.89	3.79	4.94	4.38
9	I	2.32	4.62	3.74	2.62	4.2	4.81	2.5	4.29
10	J	3.15	2.23	3.84	3.09	4.48	2.91	3.85	2.67
11	K	2.76	2.31	3.71	2.72	3.91	3.2	3.38	3.72
12	L	2.51	1.41	3.24	2.05	3.95	1.69	4.62	4.43
13	M	2.39	4.54	4.28	3.31	2.25	2.11	3.59	4.22
14	N	3.23	4.16	2.56	3.4	4.01	2.39	0.49	4.61
15	O	4.22	3.88	3.19	2.12	3.05	2.39	3.24	3.72

62

- 2) 「分析」-「多変量解析」-「クラスター分析」を選択する。
- 3) 「データ」をクリックし、価格～付属品を「分析変数」に、「機種」を「ラベルの識別」に設定する。



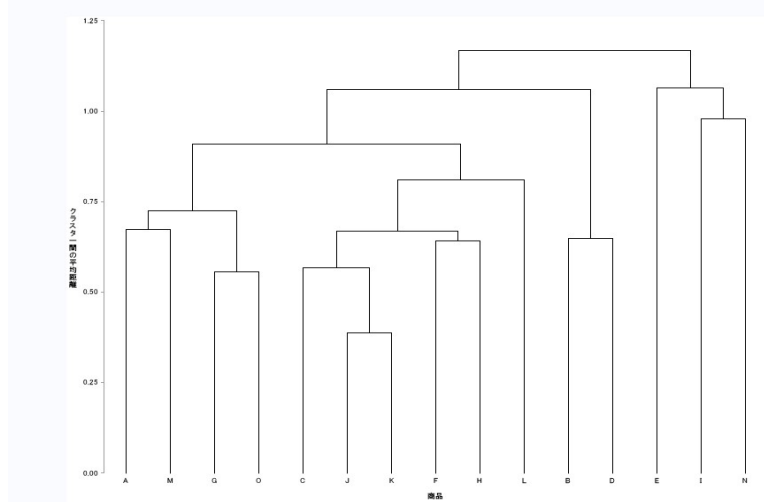
63

- 4) 「クラスター」をクリックし、「群平均法」を選択する。



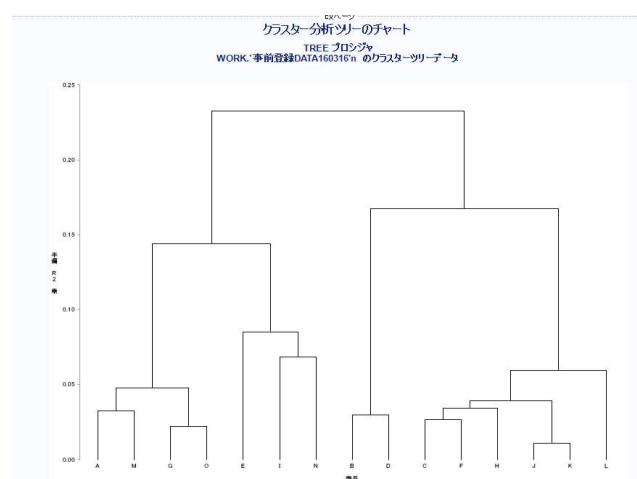
64

5) デンドログラム (樹形図) が表示される。



65

Ward法によるデンドログラム



66

- ・階層的クラスター分析では、ユークリッド距離と、クラスターが明確に算出されかつ分類感度も高いとされるウォード法(Ward法)が一般的である。

- ・クラスター距離には特定の単位がない。
相対的な関係性は示せる。

- ・使用した変数によりクラスター距離を算出する。
使用する変数が異なれば、クラスター距離は変わる。

67

Data Science Institute

非階層的クラスター分析

- ・代表的な方法 : [K-means法](#)

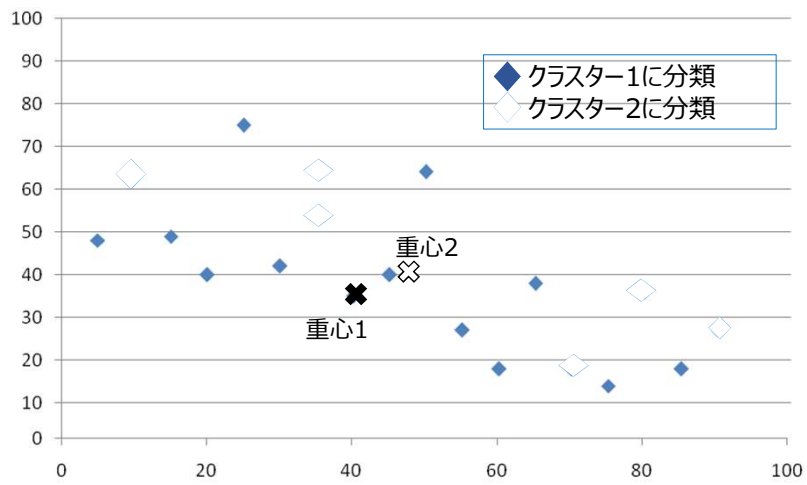
- ・階層的クラスター分析に比較して、計算量が圧倒的に少ない。

- 1) クラスターの数をあらかじめ設定する。
- 2) ランダムにすべての点をK個のクラスターに分ける。
- 3) K個の中心点の内、どの中心点との距離が一番短いかという基準で、すべての点を新たに再分類する。
- 4) 再び計算して中心点を求める。
- 5) 中心点からの距離に基づき分類⇒分類に基づき中心点を計算を繰り返す。

68

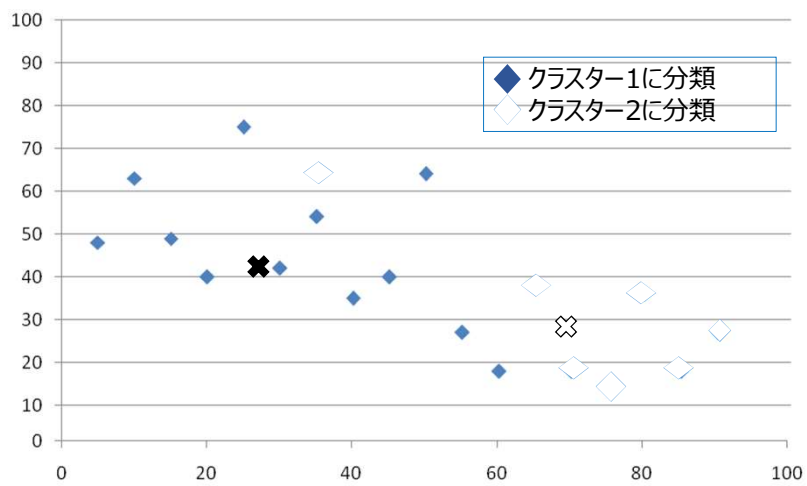
Data Science Institute

K-means法①



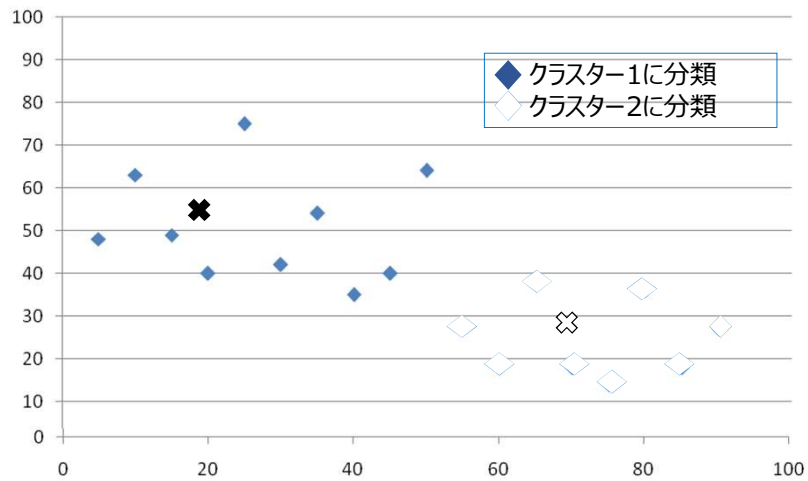
69

K-means法②



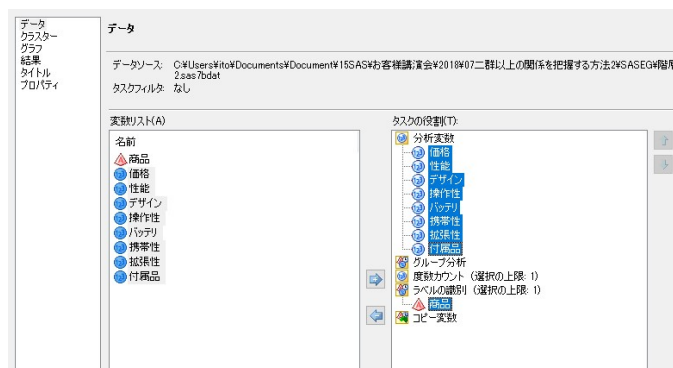
70

K-means法③



71

- 1) データを入力する。
- 2) 「分析」-「多変量解析」-「クラスター分析」を選択する。
- 3) 「データ」をクリックし、価格～付属品を「分析変数」に、「商品」を「ラベルの識別」に設定する。



72

4) 「クラスター」をクリックし、「K-means法」を選択し、最大クラスター数を設定する。



73

5) 「結果」をクリックし、「K-meansクラスター」をチェックする。



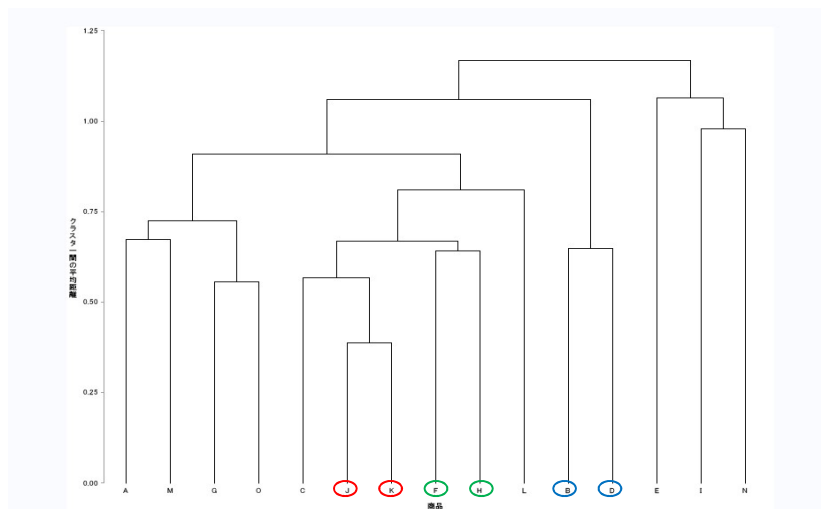
74

「K-means法」の結果（「出力データ」をクリック）

	商品	価格	性能	デザイン	操作性	バッテリー	携帯性	拡張性	付属品	CLUSTER	DISTANCE
1	A	2.43	3.36	3.04	3.21	2.35	3.9	3.45	4.22	1	0
2	B	3.95	4.55	2.54	3.38	4.43	1.95	4.83	2.62	4	1.198835862
3	C	4.26	2.77	3.22	2.24	4.12	2.9	4.52	3.77	7	1.4157251773
4	D	3.82	4.21	3.88	3.55	4.94	2.12	4.64	3.38	4	1.198835862
5	E	1.11	3.23	2.02	4.61	3.22	3.86	2.34	2.82	5	0
6	F	3.13	3.19	2.81	3.22	4.52	4.15	4.78	3.61	6	1.1670867702
7	G	3.64	3.56	2.29	2.89	2.27	2.29	4.26	4.72	7	1.4088599603
8	H	2.74	3.1	4.47	3.2	4.89	3.79	4.94	4.33	6	1.1670867702
9	I	2.92	4.62	3.74	2.62	4.2	4.81	2.5	4.29	9	0
10	J	3.15	2.23	3.84	3.09	4.48	2.91	3.85	2.67	10	0.7149650341
11	K	2.78	2.91	3.71	2.72	3.91	3.2	3.38	3.72	10	0.7149650341
12	L	2.51	1.41	3.24	2.05	3.35	1.89	4.62	4.43	3	0
13	M	2.99	4.54	4.28	3.81	2.25	2.11	3.59	4.22	8	0
14	N	3.23	4.16	2.56	3.4	4.01	2.39	0.49	4.61	2	0
15	O	4.22	3.88	3.19	2.12	3.05	2.39	3.24	3.72	7	1.065220692

75

階層式クラスター分析のデンドログラム



76

利用例

スーパーのポイント会員を、その購買の傾向が似ているクラスタに分割する。
会員を次のような属性で表現。

第1属性：先月の弁当・総菜の購入額
第2属性：先月の清涼飲料の購入額
第*n*属性：先月の野菜類の購入額

例：会員3番 =〈3000円,1200円,⋯,0円〉

77

Data Science Institute

見つかるクラスタの例

- クラスタ-1：主に日用品を購入する顧客
- クラスタ-2：主に昼食製品を購入する顧客

見つかったクラスタの利用例

- バーゲンなどでまとめ買いをするので、ダイレクトメールなどの販促策はクラスタ1の顧客を中心に
- その場で購入するものを決めるクラスタ2の顧客にはセットメニューによる客単価の向上などを中心に

クラスタ分析は、データを大まかに要約し、
その全体像を把握するための探索的手法

78

Data Science Institute

まとめ

- 主成分分析

 - 主成分負荷量、主成分寄与率

- 因子分析

 - 因子抽出法、反復主因子法、最尤法

 - 因子軸回転（直交バリマックス、斜交プロマックス）

 - 因子負荷量、寄与率、共通性

- クラスター分析

 - 階層的クラスター分析

 - 群平均法、Ward法

 - 非階層的クラスター分析

 - K-means法