

## 21. クラスタ分析

クラスタ分析は、簡単に言うと、複数の変数について距離の近いグループをまとめ上げていく分類手法です。因子分析と同様に、さまざまな手法があり、データの種類と分析目的によって、手法を選択していく必要があります。

ここでは、手法としては単純な平方ユークリッド距離（デフォルト）の平均値によるクラスタ分析のシンタックスを紹介します。

基本的には、

```
cluster 変数リスト/ method baverage.
```

ですが、これだけだと、最も近い2ケースを結合させてその平均値をあらたなケースとみなし、つぎに再び最も近い2ケースを結合させてその平均値を新たなケースとみなし...これを、最後に一つにまとまるまで N-1 回繰り返すという結合過程が示されるだけで、その後の分析に利用できません。そこで、一定範囲のクラスタ数を指定して、その範囲のクラスタ解について、クラスタ番号を保存させる方法が、便利です。

その場合のシンタックスは、

```
cluster 変数リスト/save cluster (最小クラスタ数、最大クラスタ数)  
/method baverage (ルート名)
```

となります。

具体例として 1995 年国勢調査の東京都 54 市区町村データを主成分分析した結果（→ 20 因子分析の具体例参照）を利用して、3つの主成分得点（tk95a, tk95b, tk95c）をもとに、54 市区町村を分類する場合をとりあげます。保存するクラスタ数は 5 ~ 9、ルート名は clu95a とすると、シンタックスは、

```
cluster tk95a1 tk95a2 tk95a3 /save cluster (5,9) /method baverage (clu95a).
```

となります。

この場合の出力結果は、図 21.1 のように、結合過程が示されるだけですが、「記録」に見えるように、clu95a5 として、5 クラスタ解が、clu95a6 として 6 クラスタ解が、...clu95a9 として 9 クラスタ解が、SPSS データファイルに保存されています（図 21.2）。

どのクラスタ解を採用するかは、これらの変数を使ってクラスタの性格を分析したうえで、分析目的に応じて判断することになります。

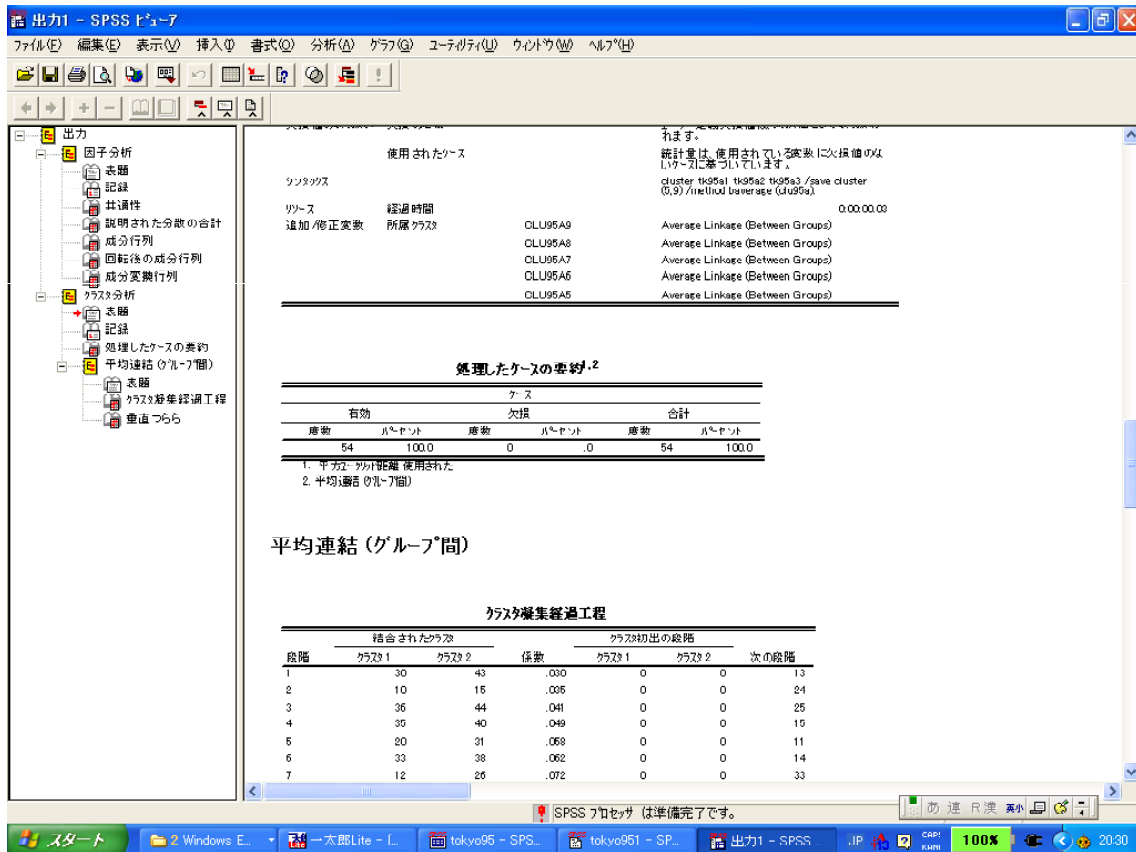


図 21.1 クラスター分析の出力例

	tk95a1	tk95a2	tk95a3	clu95a4	clu95a5	clu95a6	clu95a7	clu95a8	clu95a9	var	var	var	var	var	var	var
1	-2.33216	-.02425	-1.61703	1	1	1	1	1	1							
2	-1.05252	-.33957	-.57749	2	2	2	2	2	2							
3	-1.41563	-.12440	-1.57781	1	1	1	1	1	1							
4	-.30488	-1.32335	-.73965	2	2	2	2	2	2							
5	-.62063	-.73703	-1.40993	2	2	2	2	2	2							
6	-1.20469	-1.30505	.51187	3	3	3	3	3	3							
7	.12778	-1.31225	1.35025	4	4	4	4	4	4							
8	.77387	-.90411	1.18837	4	4	4	4	4	3							
9	.58319	-1.50375	.47402	4	4	4	4	4	3							
10	.26313	-1.24885	-.88184	2	2	2	2	2	2							
11	.49131	-.72888	.60079	4	4	4	4	4	3							
12	.12514	-.55215	-1.08378	2	2	2	2	2	2							
13	-.68726	-1.02540	-1.37380	2	2	2	2	2	2							
14	.37885	-1.78889	-.35514	5	5	2	2	2	2							
15	.33835	-1.07909	-.89724	2	2	2	2	2	2							
16	.15191	-2.02315	.06795	5	5	2	2	2	2							
17	.09179	-1.66497	.94942	4	4	4	4	4	3							
18	.04089	-1.61390	1.54678	4	4	4	4	4	3							
19	.94597	-1.22767	.80924	4	4	4	4	4	3							
20	.23140	.06008	-.34733	6	6	5	5	5	4							
21	.24380	-.49553	1.60828	4	4	4	4	4	3							
22	.50943	-.60223	1.36405	4	4	4	4	4	3							
23	1.10879	-.16098	1.24461	4	4	4	4	4	3							
24	.00403	1.61043	-.39553	6	6	5	5	5	4							
25	.46874	.52121	.17461	6	6	5	5	5	4							
26	-.03002	-.44015	-1.27145	2	2	2	2	2	2							
27	.08699	-.01057	-.77025	6	6	5	5	5	4							
28	.18330	1.89454	.47568	7	7	6	5	5	4							
29	.63943	.53159	-.40429	6	6	5	5	5	4							
30	.44815	.73101	.49562	8	7	6	5	5	4							
31	.27298	.04197	-.58420	6	6	5	5	5	4							
32	-.16094	1.11570	-.59349	6	6	5	5	5	4							
33	.28749	.43512	-1.24216	6	6	5	5	5	4							
34	.19828	.90269	-.84687	6	6	5	5	5	4							
35	.21547	.58197	-.70259	6	6	5	5	5	4							
36	-.14180	.49921	-.10960	6	6	5	5	5	4							
37	.03424	.73975	-1.51394	6	6	5	5	5	4							
38	.11175	.59679	-1.17449	6	6	5	5	5	4							
39	.36396	-.01660	-.14374	6	6	5	5	5	4							

図 21.2 SPSS データファイルに保存された各クラスター解のクラスター番号

クラスター番号順にケースを並べ替える方法は、2 2 をご覧ください。