Kmeans: 5 variabili socioeconomiche

Applichiamo il metodo Kmeans al dataset relativo ai distretti della città di Los Angeles richiedendo 3 gruppi. Allo scopo utilizziamo la procedura FASTCLUS del SAS.

Cluster Summary

Cluster	Frequency	RMS Std Deviation	Maximum Distance from Seed to Observation	Radius Exceeded	Nearest Cluster
1	4	0.6835	2.0347		2
2	5	0.5337	1.5333		1
3	3	0.4817	1.1024		2

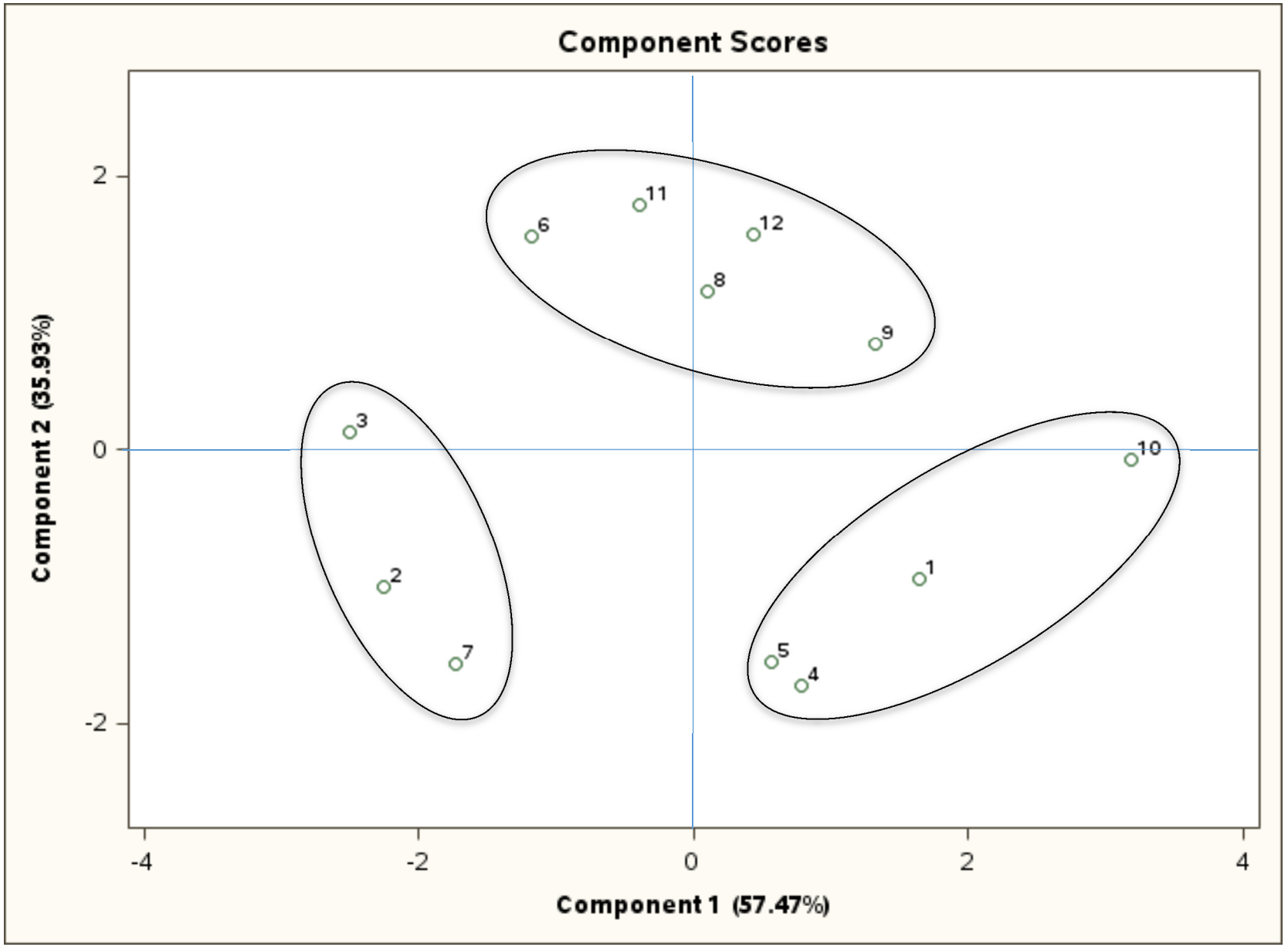
Pseudo F Statistic = 11.97

Cluster Means

Cluster	POP	SCH	OCC	SER	VCASE
1	-0.135659141	0.998202399	0.013427741	0.993379769	1.256373884
2	0.871609978	-0.437529836	0.794922254	-0.216078227	-0.502549554
3	-1.271804442	-0.601720138	-1.342774078	-0.964375980	-0.837582589

Cluster Standard Deviations

Cluster	POP	SCH	OCC	SER	VCASE
1	0.781603326	0.275639817	0.745693115	1.045344192	0.000000000
2	0.189065577	0.942126824	0.400407935	0.433313111	0.391043218
3	0.387112744	0.772224380	0.246134554	0.000000000	0.594569345


Component Scores
Component 2 (35.93%)
Component 1 (57.47%)
2
0
-2
-4
-2
0
2
4
11
6
12
8
9
3
10
2
1
7
5
4