

Cluster Analysis

1. Import Data

```
> data1 <- read.csv("C:/Users/Indah/Downloads/data susenas.csv", sep =
",",
+                      header = TRUE)
> head(data1)
  Domisili  X1    X2    X3    X4    X5    X6
1  Selayar 142 13.13 32.45 68.08 114.85 8.01
2  Bulukumba 353 8.37 13.47 72.62 106.27 3.73
3  Bantaeng 461 9.68 19.59 74.59 107.12 5.71
4  Jeneponto 391 15.31 28.28 65.40 115.16 4.08
5  Takalar 501 9.62 20.37 70.30 111.67 4.24
6  Gowa 377 8.00 7.09 72.04 111.96 2.82
> data <- data1[-1]
```

2. Standardisasi Data

Data yang diamati memiliki satuan yang bervariasi sehingga proses *cluster* tidak bisa dilakukan. Oleh karena itu, perlu dilakukan standardisasi data dengan melakukan transformasi pada data asli.

```
> standardisasi.data <- scale(data, center = TRUE, scale = TRUE)
```

3. Statistika Deskriptif

```
> statistika.deskriptif <- summary(data)
> statistika.deskriptif
```

X1		X2		X3		X4	
Min.	: 39.0	Min.	: 4.48	Min.	: 1.040	Min.	: 65.40
1st Qu.:	144.2	1st Qu.:	8.15	1st Qu.:	7.585	1st Qu.:	71.11
Median :	190.5	Median :	9.65	Median :	13.770	Median :	72.36
Mean :	615.1	Mean :	10.42	Mean :	14.660	Mean :	72.16
3rd Qu.:	380.5	3rd Qu.:	13.32	3rd Qu.:	20.762	3rd Qu.:	73.81
Max.	: 8131.0	Max.	: 16.38	Max.	: 32.450	Max.	: 75.66

X5		X6	
Min.	: 105.9	Min.	: 2.410
1st Qu.:	106.3	1st Qu.:	3.737
Median :	109.1	Median :	4.580
Mean :	109.5	Mean :	4.832
3rd Qu.:	111.7	3rd Qu.:	5.657
Max.	: 116.1	Max.	: 8.010

4. Pengujian Asumsi

a. Asumsi Sampel Representatif

```
> library(psych)
warning message:
package 'psych' was built under R version 4.2.3
> KMO <- KMO(data)
> KMO
Kaiser-Meyer-Olkin factor adequacy
Call: KMO(r = data)
Overall MSA = 0.61
MSA for each item =
  X1  X2  X3  X4  X5  X6
0.84 0.64 0.54 0.57 0.58 0.60
```

Setiap variabel bernilai lebih dari 0.5 sehingga sampel telah representatif atau mewakili populasi.

b. Asumsi Nonmultikolinearitas

```
> korelasi <- cor(data)
> korelasi
```

	x1	x2	x3	x4	x5	x
6						
x1	1.0000000	-0.4348758	-0.3628717	0.2145422	-0.2419613	-0.379279
1						
x2	-0.4348758	1.0000000	0.6040912	-0.3160846	0.3208308	0.508689
9						
x3	-0.3628717	0.6040912	1.0000000	-0.4772784	0.1833344	0.126106
0						
x4	0.2145422	-0.3160846	-0.4772784	1.0000000	-0.5819319	-0.117051
9						
x5	-0.2419613	0.3208308	0.1833344	-0.5819319	1.0000000	0.298866
1						
x6	-0.3792791	0.5086899	0.1261060	-0.1170519	0.2988661	1.000000
0						

Nilai korelasi antarvariabel kurang dari 0.8 sehingga tidak terjadi multikolinearitas pada setiap variabel. Seluruh asumsi sudah terpenuhi, maka dapat dilanjutkan ke langkah berikutnya.

5. Pemilihan Metode Terbaik

```
> matriks.jarak1 <- dist(data)
>
> # 1. Single Linkage
> cluster.hierarki.single <- hclust(matriks.jarak1, method = "single")
> jarak.single <- cophenetic(cluster.hierarki.single)
> korelasi.single <- cor(matriks.jarak1, jarak.single)
>
> # 2. Average Linkage
> cluster.hierarki.average <- hclust(matriks.jarak1, method = "ave")
> jarak.average <- cophenetic(cluster.hierarki.average)
> korelasi.average <- cor(matriks.jarak1, jarak.average)
>
> # 3. Complete Linkage
> cluster.hierarki.complete <- hclust(matriks.jarak1, method = "complete")
> jarak.complete <- cophenetic(cluster.hierarki.complete)
> korelasi.complete <- cor(matriks.jarak1, jarak.complete)
>
> # 4. Centroid Linkage
> cluster.hierarki.centroid <- hclust(matriks.jarak1, method = "centroid")
> jarak.centroid <- cophenetic(cluster.hierarki.centroid)
> korelasi.centroid <- cor(matriks.jarak1, jarak.centroid)
>
> # 5. Ward
> cluster.hierarki.ward <- hclust(matriks.jarak1, method = "ward.D")
> jarak.ward <- cophenetic(cluster.hierarki.ward)
> korelasi.ward <- cor(matriks.jarak1, jarak.ward)
>
> ### Metode Terbaik
> korelasi.cophenetic <- data.frame(korelasi.single, korelasi.average,
+                                   korelasi.complete, korelasi.centroid,
+                                   korelasi.ward)
> korelasi.cophenetic
```

	korelasi.single	korelasi.average	korelasi.complete	korelasi.centroid
1	0.9975755	0.9985852	0.9975982	0.9985087
korelasi.ward				
1	0.9749552			

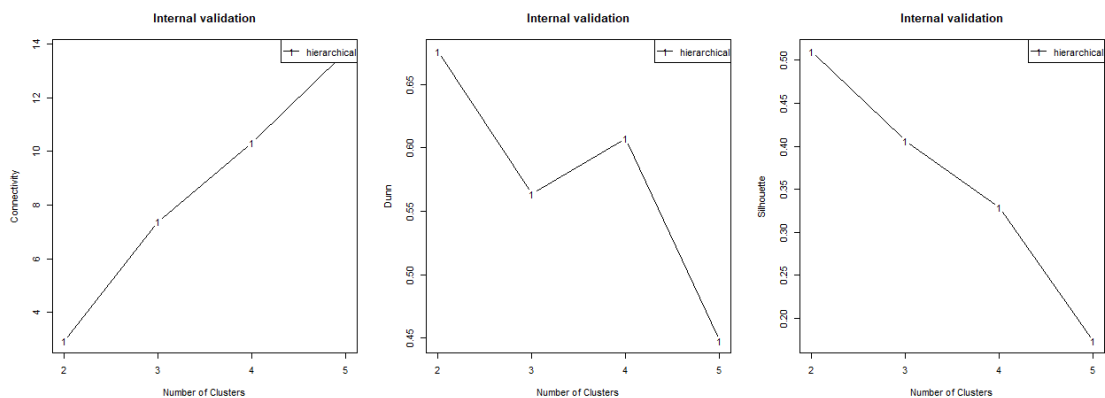
Nilai korelasi yang paling mendekati nilai 1 adalah metode *average linkage* sebesar 0.998 sehingga metode tersebut merupakan metode terbaik.

6. Pemilihan Jumlah *Cluster* yang Optimal

```
> library(clvalid)
Loading required package: cluster
Warning message:
package 'clvalid' was built under R version 4.2.3
> validasi <- clvalid(obj = data.frame(standardisasi.data), nClust = 2:5,
+                     clMethods = "hierarchical", validation = "internal",
+                     method = "average", maxitems = length(standardisasi.data),
+                     verbose = TRUE)
[1] "Finished internal validation, hierarchical 2 clusters"
[1] "Finished internal validation, hierarchical 3 clusters"
[1] "Finished internal validation, hierarchical 4 clusters"
[1] "Finished internal validation, hierarchical 5 clusters"
Warning message:
In clvalid(obj = data.frame(standardisasi.data), nClust = 2:5, clMethods = "hierarchical", :
  rownames for data not specified, using 1:nrow(data)
> optimalScores(validasi)
```

	Score	Method	Clusters
Connectivity	2.9289683	hierarchical	2
Dunn	0.6758966	hierarchical	2
Silhouette	0.5102026	hierarchical	2

```
> par(mfrow=c(1,3))
> plot(validasi)
```



Cluster yang optimal digunakan adalah sebanyak dua karena memiliki nilai terendah pada indeks connectivity sebesar 2.928 serta nilai tertinggi pada indeks dunn sebesar 0.675 dan silhouette sebesar 0.510.

7. Analisis *Cluster* Hierarki

a. Matriks Jarak

```
> matriks.jarak <- dist(standardisasi.data, method = "euclidean",
+                        diag = FALSE)
> matriks.jarak
```

	1	2	3	4	5	6
7						
2	4.9994516					
3	4.2728441	1.7629324				
4	2.9655435	4.8313853	5.0391566			
5	3.3317612	2.1068165	2.4512060	2.9985339		
6	5.0928729	1.9732608	3.0305527	4.4410233	1.9702149	
7	4.1909795	1.2983038	0.8381216	4.6961170	2.1492174	2.8273614
8	4.7636175	1.4343556	1.7896911	4.4566415	2.2146474	2.5944206
83128						1.62

9	2.1762331	4.4309268	4.0785246	2.0535847	2.7018001	4.0452935	3.99
75236							
10	3.9498435	1.9684967	2.4252090	4.0904623	2.0401003	2.5885611	1.77
95760							
11	4.1392135	1.4524551	2.0566304	3.8462411	1.7912130	2.7643254	1.38
13712							
12	3.9926847	2.1343330	2.0848259	4.5468903	2.0917753	2.2081543	1.90
23552							
13	4.2965447	1.7947756	2.2505014	4.3856216	1.6544991	1.2874012	2.08
04301							
14	4.9411771	2.2775097	2.7815186	4.9286066	2.2046074	1.2027236	2.79
87258							
15	4.8645148	1.5136097	2.3969354	4.5573162	1.8324396	0.6826432	2.24
33348							
16	4.5750706	2.4471442	1.6155428	5.0079292	2.8905022	3.0951962	1.99
67390							
17	4.1055191	2.8207437	1.8873511	4.5321809	2.8384062	3.8655837	2.15
10877							
18	4.5986051	2.9400700	2.0052344	5.6143214	3.5733600	3.9041707	2.09
70287							
19	4.3483288	2.0968023	2.2537033	3.9244635	2.3450762	3.1822333	1.92
85793							
20	4.0331025	1.2507838	1.9015577	4.0465184	1.1618898	1.7197629	1.47
86163							
21	4.0917333	4.3411969	3.5462793	5.1816008	3.9008966	4.1895688	3.74
13494							
22	8.4414720	5.2488563	5.8344656	8.0873505	6.0335585	5.3469281	5.81
13787							
23	6.3104187	1.7529291	2.6285449	6.2825878	3.4422665	2.4227606	2.56
56801							
24	4.3153629	1.6861738	1.9832898	4.5744666	1.9798763	1.8573230	1.72
54842							
	8	9	10	11	12	13	
14							
2							
3							
4							
5							
6							
7							
8							
9	3.9647039						
10	2.7142155	3.6138999					
11	1.4022531	3.7052966	1.8450788				
12	2.8585426	3.5730123	1.3868737	2.5791665			
13	2.5668298	3.6510371	1.8013754	2.4730410	1.0087683		
14	2.9906507	4.2493064	2.8312997	3.1924821	1.9595403	1.0665726	
15	2.1937731	3.9614638	2.2499022	2.4267197	1.7363541	0.8665004	1.03
81379							
16	1.9076642	3.7112537	2.9286007	2.6808081	2.4420739	2.6007861	3.10
72085							
17	1.6989034	3.6719490	3.5055554	2.3210174	3.4734455	3.4652446	3.91
66914							
18	3.0953269	4.3100626	2.5048042	3.1105564	2.1985448	2.9665539	3.76
91725							
19	1.0359564	3.5361913	2.6723144	1.2494543	3.1406796	3.0502806	3.68
59120							
20	2.0011202	3.6366161	1.5647939	1.5728272	1.5484982	1.0974482	1.75
55050							
21	4.3499239	3.3808295	3.4435155	4.4194336	2.6642885	3.3346381	3.98
42305							
22	5.8527978	7.8181030	6.0333941	6.0507115	5.8973082	5.6284391	5.54
56314							
23	2.7778571	5.6859649	3.1143134	3.1617041	2.7692185	2.4308754	2.43
44501							
24	2.5236545	3.7463939	1.3122026	2.3090292	0.6162017	0.8736577	1.82
59671							
	15	16	17	18	19	20	
21							

```

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16 2.5435251
17 3.3928855 1.9382609
18 3.3046158 1.8983821 3.1118500
19 2.8416595 2.1901827 1.6434286 3.1333253
20 1.3200011 2.6780871 2.9319314 3.0763856 2.4138345
21 3.8135678 2.8233828 4.2020164 2.4509685 4.3086478 3.9314811
22 5.3852394 6.2560418 6.8064770 6.5019292 6.4237981 5.7488192 7.32
01039
23 2.0686552 3.1309386 4.0360815 3.4530781 3.6128586 2.5566502 4.79
51064
24 1.3816785 2.3502782 3.3774305 2.3039358 2.8884174 1.3484539 3.00
75562
      22      23
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23 4.2651906
24 5.3947659 2.2863215

```

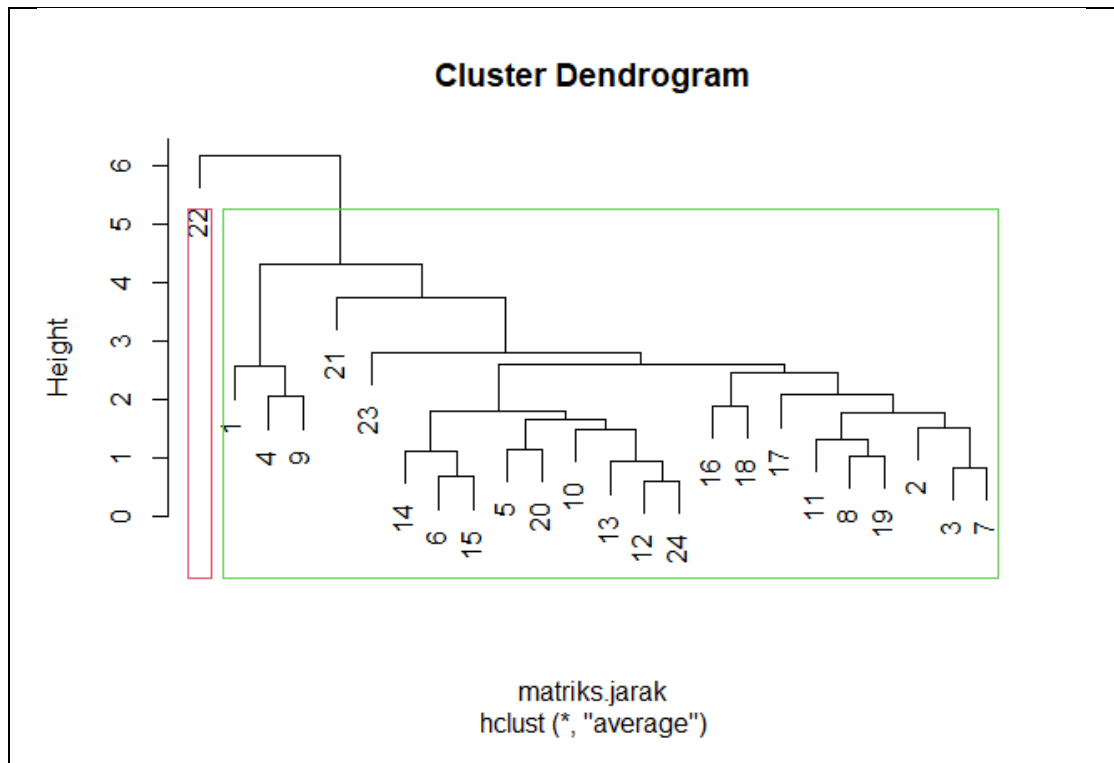
Kabupaten Soppeng dan Kota Palopo memiliki karakteristik yang lebih mirip daripada kabupaten/kota lain karena memiliki jarak paling kecil sebesar 0.616.

b. Cluster Dendogram

```

> cluster.hierarki <- hclust(matriks.jarak, method = "average")
> plot(cluster.hierarki)
> rect.hclust(cluster.hierarki, k = 2, border = 2:5)

```



c. Cluster Membership

```
> cluster <- cutree(cluster.hierarki, k = 2)
> cbind(data1[1], cluster)
```

	Domisili	cluster
1	Selayar	1
2	Bulukumba	1
3	Bantaeng	1
4	Jeneponto	1
5	Takalar	1
6	Gowa	1
7	Sinjai	1
8	Maros	1
9	Pangkep	1
10	Barru	1
11	Bone	1
12	Soppeng	1
13	Wajo	1
14	Sidrap	1
15	Pinrang	1
16	Enrekang	1
17	Luwu	1
18	Tana Toraja	1
19	Luwu Utara	1
20	Luwu Timur	1
21	Toraja Utara	1
22	Makassar	2
23	Parepare	1
24	Palopo	1

Pembentukan anggota dengan 2 cluster adalah sebagai berikut:

Cluster	Anggota
1	Selayar, Bulukumba, Bantaeng, Jeneponto, Takalar, Gowa, Sinjai, Maros, Pangkep, Barru, Bone, Soppeng, Wajo, Sidrap, Pinrang,

	Enrekang, Luwu, Tana Toraja, Luwu Utara, Luwu Timur, Toraja Utara, Pare-pare, dan Palopo
2	Makassar

8. Perbandingan *Cluster*

<pre>> aggregate(data, list(cutree(cluster.hierarki, k = 2)), mean)</pre>							
	Group.1	x1	x2	x3	x4	x5	x6
1	1	288.3043	10.67391	15.25174	72.06739	109.6809	4.936957
2	2	8131.0000	4.48000	1.04000	74.38000	106.0200	2.410000

1. *Cluster* 1 memiliki nilai rata-rata tertinggi pada presentase penduduk miskin, presentase rumah tangga yang tidak memiliki fasilitas BAB, angka partisipasi kasar sekolah dasar, dan presentase proporsi puskesmas.
2. *Cluster* 2 memiliki nilai rata-rata tertinggi pada kepadatan penduduk dan angka harapan hidup.

Berdasarkan nilai rata-rata, kelompok daerah dengan tingkat kesejahteraan yang kurang baik adalah *cluster* 2 dan yang baik adalah *cluster* 1.

Kesimpulan

Kesejahteraan berdasarkan sosial ekonomi rumah tangga di Provinsi Sulawesi Selatan terbagi ke dalam 2 *cluster* sebagai berikut:

1. *Cluster* 1 memiliki jumlah anggota tertinggi, yaitu Selayar, Bulukumba, Bantaeng, Jeneponto, Takalar, Gowa, Sinjai, Maros, Pangkep, Barru, Bone, Soppeng, Wajo, Sidrap, Pinrang, Enrekang, Luwu, Tana Toraja, Luwu Utara, Luwu Timur, Toraja Utara, Pare-pare, dan Palopo. *Cluster* ini dikategorikan sebagai wilayah dengan kesejahteraan yang kurang baik.
2. *Cluster* 2 memiliki jumlah anggota terendah, yaitu Makassar. *Cluster* ini dikategorikan sebagai wilayah dengan kesejahteraan yang baik.

Saran yang dapat saya berikan apabila terdapat perbedaan tingkat kesejahteraan sosial ekonomi yaitu pemerintah melakukan pemerataan pada fasilitas yang menunjang kehidupan sehari-hari masyarakat dan meningkatkan Sumber Daya Manusia (SDM) dengan meningkatkan mutu pendidikan.