

## تحلیل خوشه بندی سلسله مراتبی با حضور یک متغیر کمی

مثال: اطلاعات ۷ نفر از کارکنان یک شرکت در ارتباط با میزان حقوق دریافتی آنها در دست است. با توجه به داده‌های مربوطه و با استفاده از روش خوشه‌بندی سلسله‌مراتبی، کارکنان این شرکت را دسته‌بندی کنید. (داده‌ها در فایل تحت عنوان Employee data ذخیره شده است).

### Case Processing Summary<sup>a</sup>

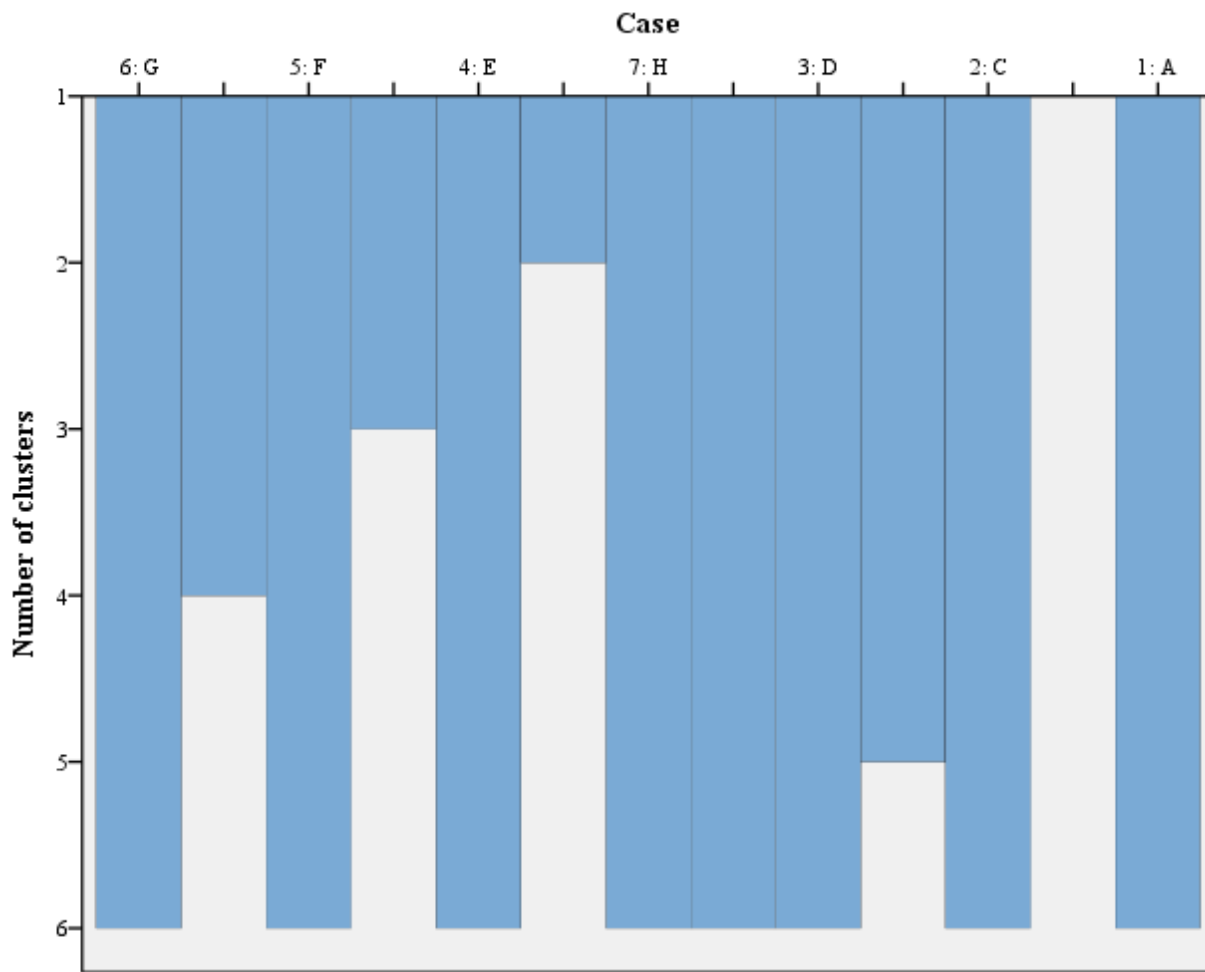
Valid		Cases Missing		Total	
N	Percent	N	Percent	N	Percent
7	100.0	0	.0	7	100.0

a. Single Linkage

### Proximity Matrix

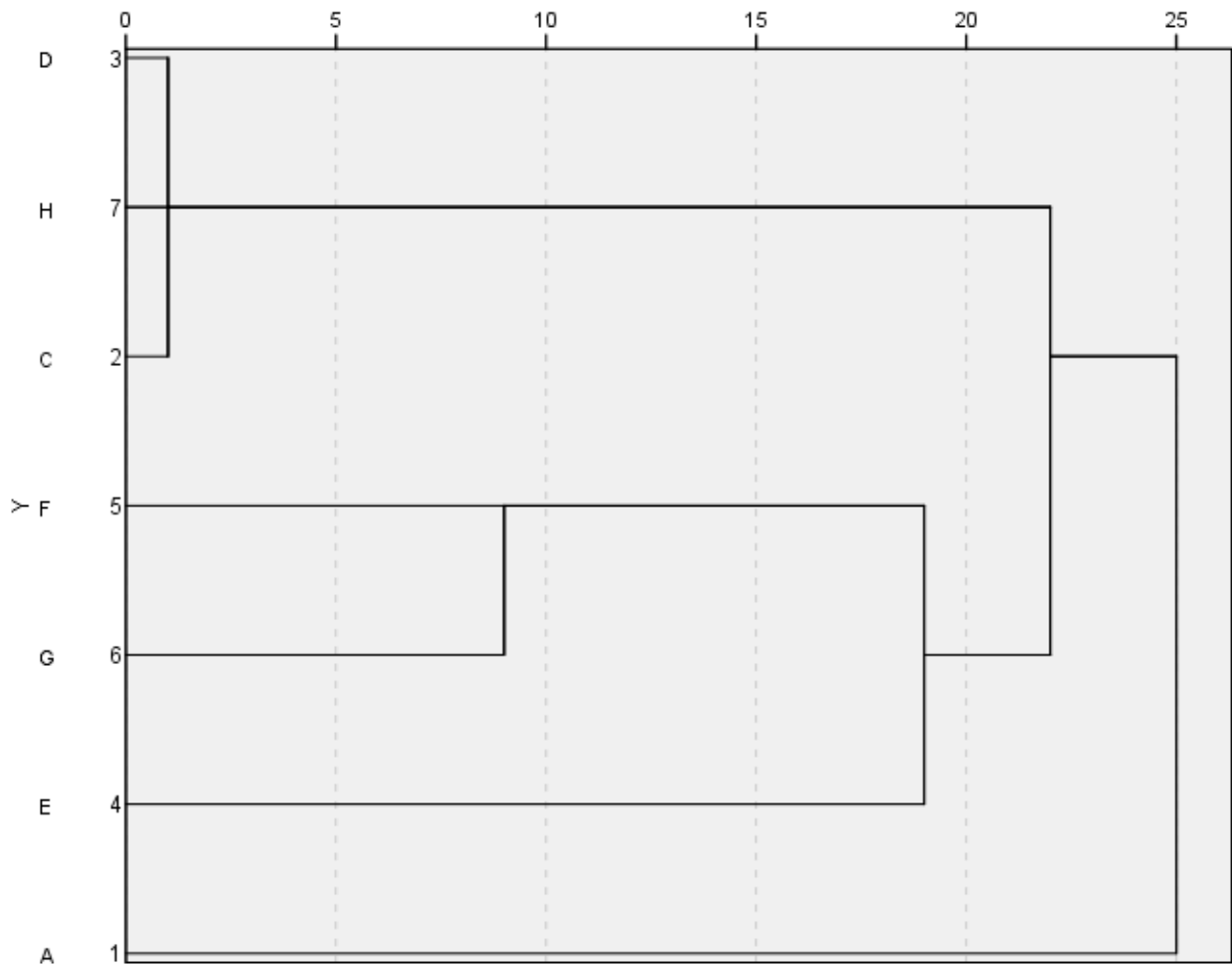
Case	Chebychev Distance						
	1:A	2:C	3:D	4:E	5:F	6:G	7:H
1:A	.000	35550.000	35100.000	12000.000	24900.000	21000.000	35100.000
2:C	35550.000	.000	450.000	23550.000	10650.000	14550.000	450.000
3:D	35100.000	450.000	.000	23100.000	10200.000	14100.000	.000
4:E	12000.000	23550.000	23100.000	.000	12900.000	9000.000	23100.000
5:F	24900.000	10650.000	10200.000	12900.000	.000	3900.000	10200.000
6:G	21000.000	14550.000	14100.000	9000.000	3900.000	.000	14100.000
7:H	35100.000	450.000	.000	23100.000	10200.000	14100.000	.000

This is a dissimilarity matrix



### Dendrogram using Single Linkage

Rescaled Distance Cluster Combine



**Cluster  
Membership**

Case	3 Clusters
1:A	1
2:C	2
3:D	2
4:E	3
5:F	3
6:G	3
7:H	2

**Cluster Membership**

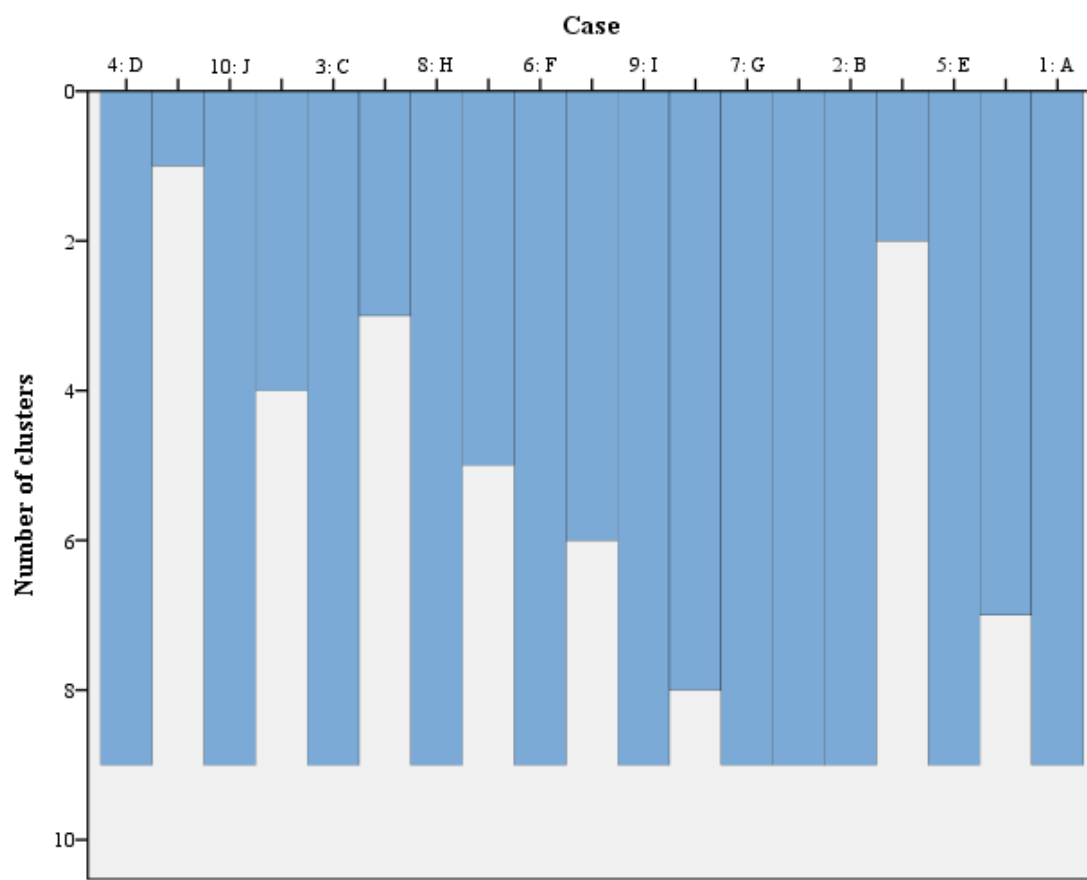
Case	3 Clusters	2 Clusters
1:A	1	1
2:C	2	2
3:D	2	2
4:E	3	2
5:F	3	2
6:G	3	2
7:H	2	2

مثال: اطلاعات ۱۰ نفر از کارکنان یک شرکت در ارتباط با متغیرهای میزان حقوق دریافتی، میزان حقوق آغازین، سطح تحصیلات، میزان تجربه‌ی قبلی و مدت زمان استخدام (برحسب ماه) در دست است. با توجه به داده‌های مربوطه و با استفاده از روش خوشه‌بندی سلسله‌مراتبی، کارکنان این شرکت را دسته‌بندی کنید. (داده‌ها در فایلی تحت عنوان Employee data\_1 ذخیره شده است).

### Proximity Matrix

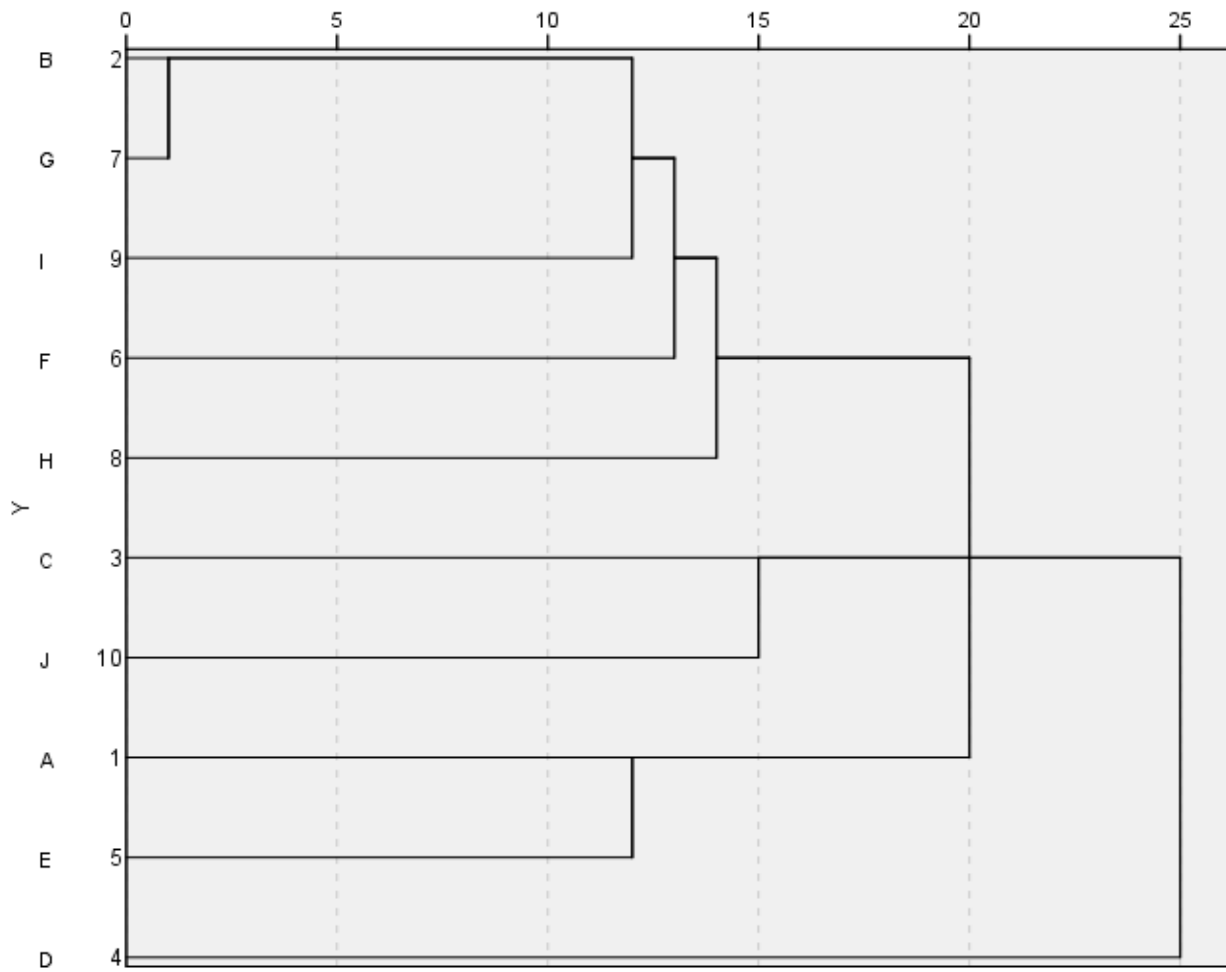
Case	Chebychev Distance									
	1:A	2:B	3:C	4:D	5:E	6:F	7:G	8:H	9:I	10:J
1:A	.000	2.016	2.995	2.957	1.145	2.577	2.609	3.292	2.720	2.780
2:B	2.016	.000	3.144	3.251	2.016	1.779	.711	2.016	1.145	2.016
3:C	2.995	3.144	.000	1.741	2.214	2.861	2.609	3.472	2.424	1.248
4:D	2.957	3.251	1.741	.000	2.845	2.845	2.845	1.731	2.845	1.625
5:E	1.145	2.016	2.214	2.845	.000	1.431	2.609	2.147	1.575	1.769
6:F	2.577	1.779	2.861	2.845	1.431	.000	2.372	1.219	1.186	1.613
7:G	2.609	.711	2.609	2.845	2.609	2.372	.000	2.609	1.186	2.609
8:H	3.292	2.016	3.472	1.731	2.147	1.219	2.609	.000	1.423	2.224
9:I	2.720	1.145	2.424	2.845	1.575	1.186	1.186	1.423	.000	1.423
10:J	2.780	2.016	1.248	1.625	1.769	1.613	2.609	2.224	1.423	.000

This is a dissimilarity matrix



### Dendrogram using Single Linkage

Rescaled Distance Cluster Combine



## خوشه بندی سلسله مراتبی برای داده‌های دودویی

مثال: اطلاعات ۱۲ زندانی در ارتباط با متغیرهای جنسیت، وضعیت تاهل، اشتغال، اقلیت مذهبی بودن، محکومیت دوم و دومین جرم خشونت آمیز ثبت شده است. با توجه به داده‌های مربوطه و با استفاده از روش خوشه‌بندی سلسله‌مراتبی، زندانیان این زندان را دسته‌بندی کنید. (داده‌ها در فایلی تحت عنوان Binary\_Data ذخیره شده است).

### Case Processing Summary<sup>b</sup>

Valid		Missing Value		Rejected		Total	
N	Percent	N	Percent	N	Percent	N	Percent
12	100.0	0	.0	0	.0	12	100.0

a. Value different from both 1 and 0.

b. Single Linkage

### Proximity Matrix

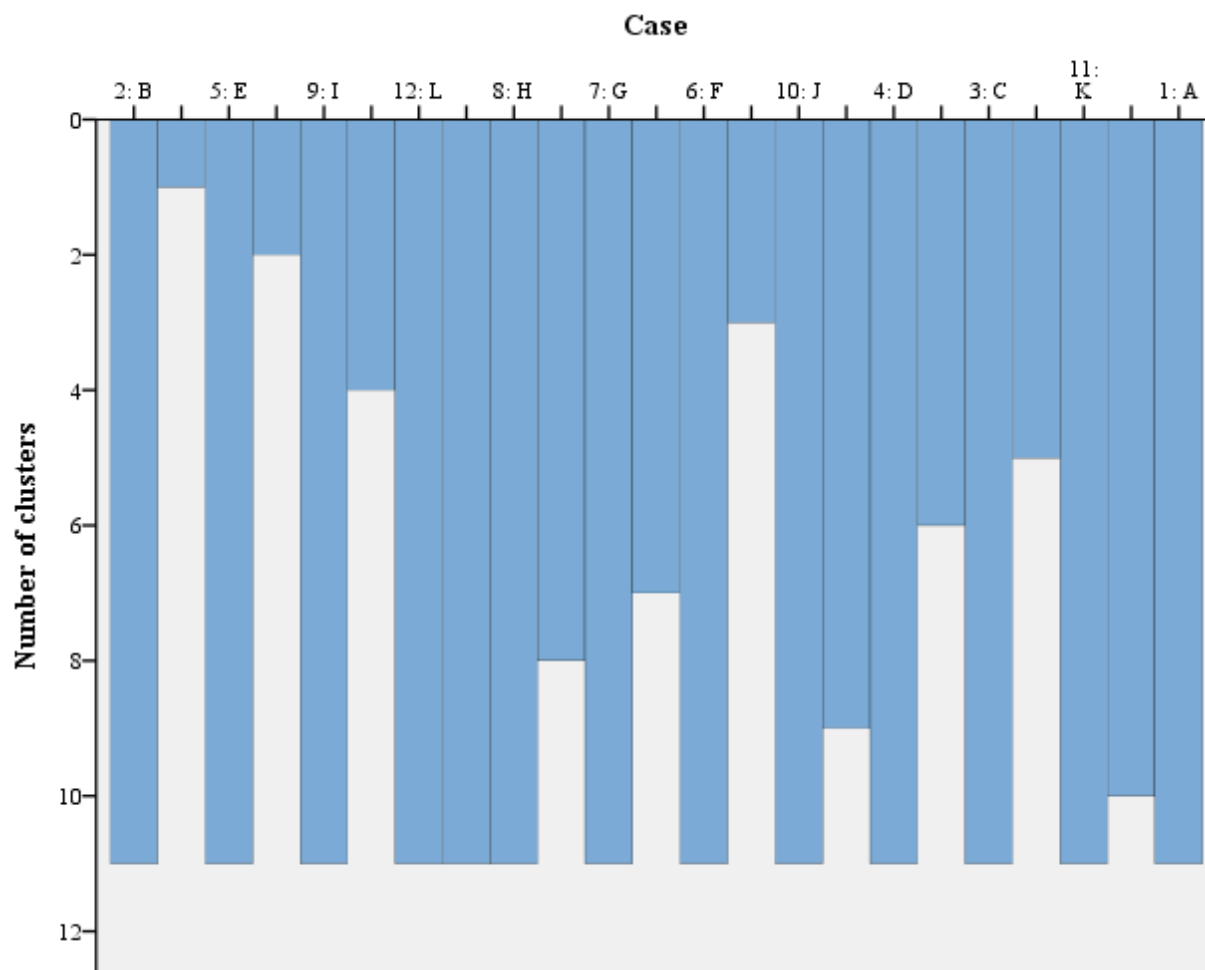
Case	Simple matching Measure											
	1:A	2:B	3:C	4:D	5:E	6:F	7:G	8:H	9:I	10:J	11:K	12:L
1:A	1.000	.500	.833	.500	.167	.333	.500	.667	.667	.667	1.000	.667
2:B	.500	1.000	.333	.333	.333	.500	.333	.500	.500	.500	.500	.500
3:C	.833	.333	1.000	.667	.333	.167	.333	.500	.500	.833	.833	.500
4:D	.500	.333	.667	1.000	.667	.500	.333	.167	.167	.833	.500	.167
5:E	.167	.333	.333	.667	1.000	.500	.333	.167	.500	.500	.167	.167
6:F	.333	.500	.167	.500	.500	1.000	.833	.667	.333	.333	.333	.667
7:G	.500	.333	.333	.333	.333	.833	1.000	.833	.500	.167	.500	.833
8:H	.667	.500	.500	.167	.167	.667	.833	1.000	.667	.333	.667	1.000
9:I	.667	.500	.500	.167	.500	.333	.500	.667	1.000	.333	.667	.667
10:J	.667	.500	.833	.833	.500	.333	.167	.333	.333	1.000	.667	.333
11:K	1.000	.500	.833	.500	.167	.333	.500	.667	.667	.667	1.000	.667
12:L	.667	.500	.500	.167	.167	.667	.833	1.000	.667	.333	.667	1.000

This is a similarity matrix



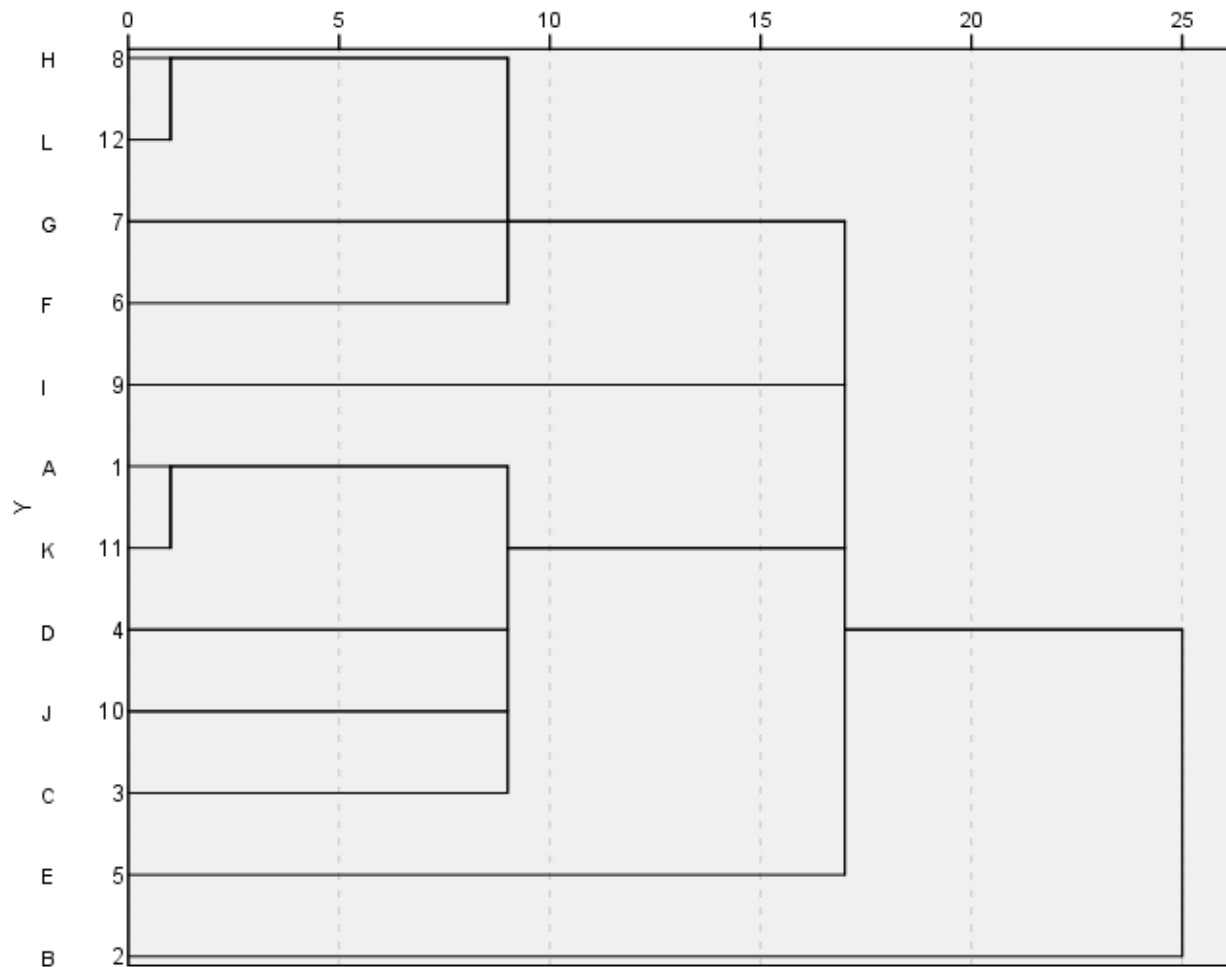
**Cluster Membership**

Case	3 Clusters
1:A	1
2:B	2
3:C	1
4:D	1
5:E	3
6:F	1
7:G	1
8:H	1
9:I	1
10:J	1
11:K	1
12:L	1



### Dendrogram using Single Linkage

Rescaled Distance Cluster Combine



تذکره مهم: چنانچه متغیرهای پاسخ برای آزمودنی‌ها از نوع شمارشی باشد، برای خوشه‌بندی باید از *Chi – Squared Measure* استفاده کرد.

مثال: اطلاعات مربوط به آمار ازدواج و طلاق در شهرستان‌های استان همدان مربوط به سال‌های ۹۱ تا ۹۵، در فایل داده *Chi\_Data* ذخیره شده است. با استفاده از روش خوشه‌بندی سلسله‌مراتبی، شهرستان‌های استان را به چندین خوشه به لحاظ آمار ازدواج و طلاق تقسیم کنید.

#### Case Processing Summary<sup>a,b</sup>

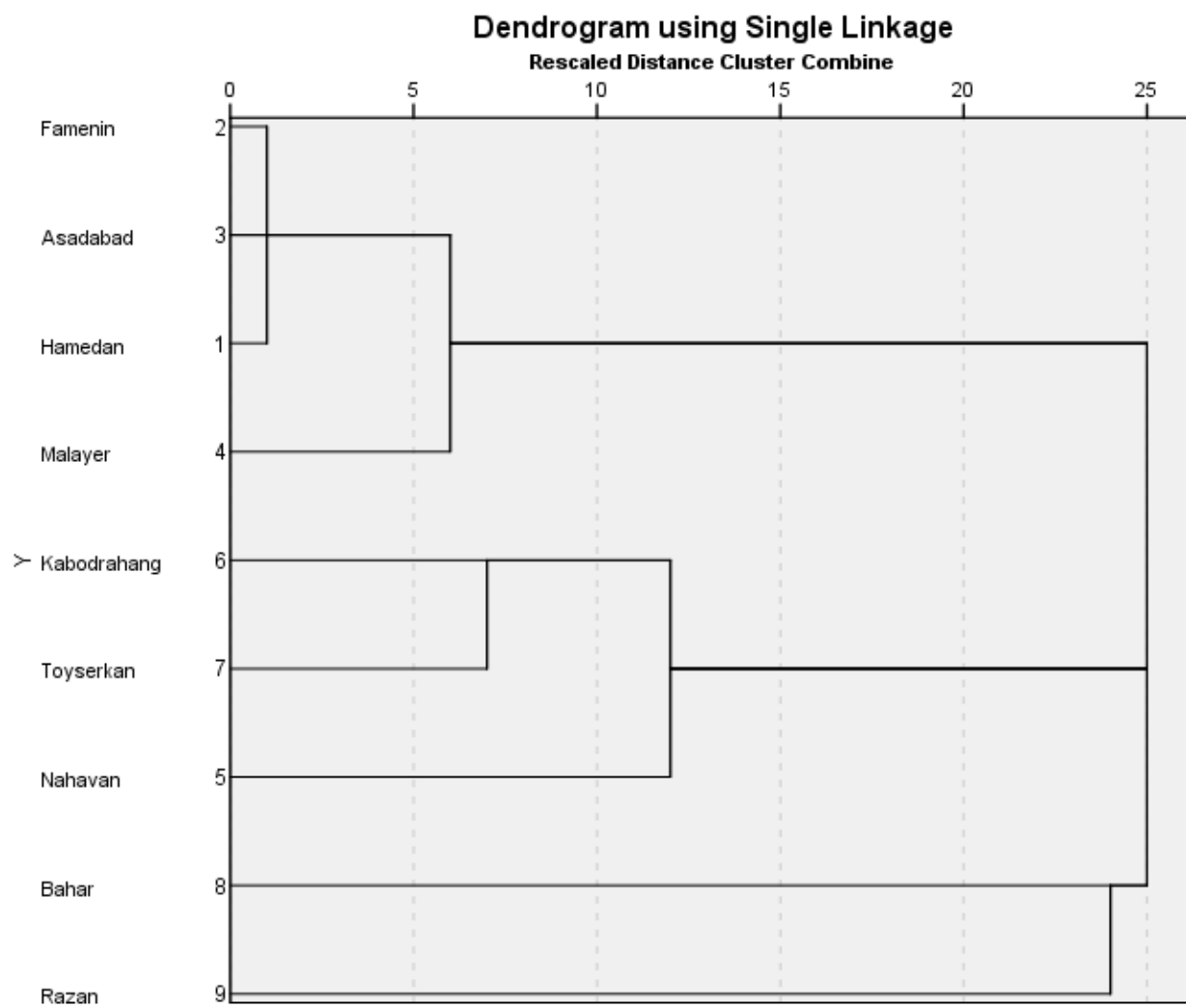
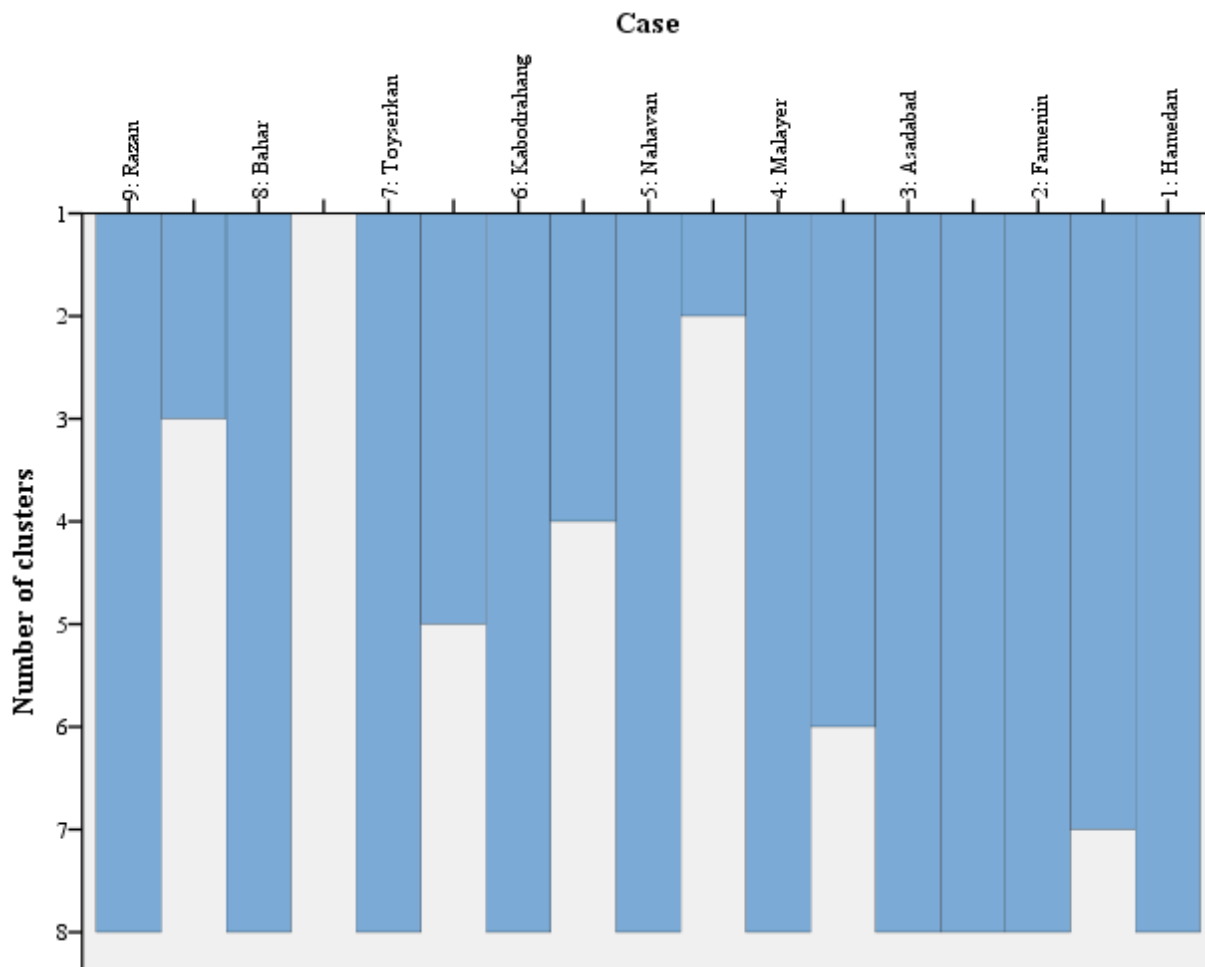
Valid		Missing Value		Negative Value		Total	
N	Percent	N	Percent	N	Percent	N	Percent
9	100.0	0	.0	0	.0	9	100.0

a. Chi-square between Sets of Frequencies used

b. Single Linkage

#### Cluster Membership

Case	3 Clusters
1:Hamedan	1
2:Famenin	1
3:Asadabad	1
4:Malayer	1
5:Nahavan	2
6:Kabodrahang	2
7:Toyserkan	2
8:Behabad	3
9:Razan	3



## خوشه بندی $k - means$

مثال: اطلاعات ۴۷۴ نفر از کارکنان یک شرکت در ارتباط با متغیرهای میزان حقوق دریافتی و میزان حقوق آغازین ثبت شده است. با توجه به داده‌های مربوطه و با استفاده از روش خوشه‌بندی  $k - means$ ، کارکنان این شرکت را به ۴ خوشه تقسیم کنید.

(داده‌ها در فایل‌های تحت عنوان Employee data\_2 ذخیره شده است).

تذکر: قبل از انجام هرگونه عملیاتی لازم است، متغیرها استاندارد شوند.

### Initial Cluster Centers

	Cluster			
	1	2	3	4
Zscore: Current Salary	2.66932	4.46281	-1.09334	5.89028
Zscore: Beginning Salary	-.16086	3.55548	-.86601	7.99985

### Iteration History<sup>a</sup>

Iteration	Change in Cluster Centers			
	1	2	3	4
1	1.799	1.599	.864	.000
2	.119	.268	.015	.000
3	.045	.101	.004	.000
4	.018	.071	.000	.000
5	.037	.132	.000	.000
6	.035	.106	.000	.000
7	.055	.132	.000	.000
8	.071	.080	.006	.000
9	.104	.106	.009	.000
10	.039	.176	.000	1.569

a. Iterations stopped because the maximum number of iterations was performed. Iterations failed to converge. The maximum absolute coordinate change for any center is 1.269. The current iteration is 10. The minimum distance between initial centers is 3.828.

### Final Cluster Centers

	Cluster			
	1	2	3	4
Zscore: Current Salary	1.11409	2.44157	-.41064	4.96791
Zscore: Beginning Salary	.94420	2.36790	-.38790	6.73057

**ANOVA**

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore: Current Salary	129.012	3	.183	470	705.365	.000
Zscore: Beginning Salary	129.793	3	.178	470	729.528	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

**Number of Cases in each Cluster**

Cluster	1	57.000
	2	34.000
	3	381.000
	4	2.000
Valid		474.000
Missing		.000

مثال: اطلاعات ۴۷۴ نفر از کارکنان یک شرکت در ارتباط با متغیرهای میزان حقوق دریافتی و میزان حقوق آغازین، جنسیت، رده بندی شغلی و اقلیت مذهبی بودن ثبت شده است. با توجه به داده‌های مربوطه کارکنان این شرکت را به چندین خوشه‌ی مناسب تقسیم کنید.

(داده‌ها در فایلی تحت عنوان Employee data\_3 ذخیره شده است).

تذکر: دقت کنید در این مساله با متغیرهای کمی و کیفی بطور همزمان مواجه هستید.

### Model Summary

Algorithm	TwoStep
Inputs	5
Clusters	4

### Cluster Quality

