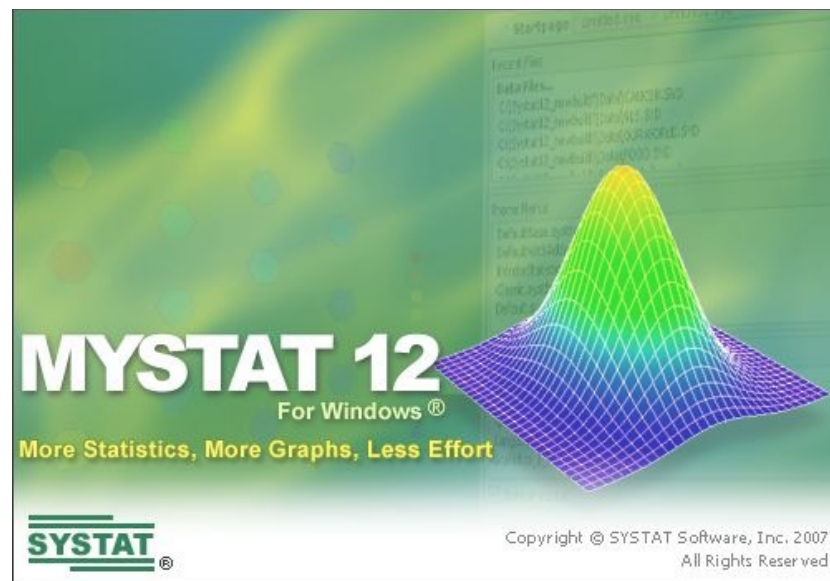


MODUL 2

ANALISIS DATA MULTIVARIAT MENGUNAKAN MYSTAT 12



Oleh:
Bertho Tantular

DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PADJADJARAN
2023

KATA PENGANTAR

Saat ini analisis data statistika merupakan hal yang penting dalam dunia pendidikan terutama yang berkaitan dengan penelitian atau riset. Untuk melakukan suatu penelitian seorang peneliti membutuhkan data yang akan mendukung pada analisis dan pengambilan kesimpulan penelitiannya. Analisis statistika merupakan salah satu cara dalam mengolah data sehingga dapat diperoleh kesimpulan atas penelitian yang dilakukan seorang peneliti, terutama penelitian yang melibatkan data kuantitatif.

Beragam aplikasi software komputer telah dibuat untuk memudahkan seseorang untuk mengolah data dan melakukan analisis statistika. Software yang pada saat ini banyak digunakan oleh banyak kalangan adalah SPSS karena sifatnya yang *user friendly* atau mudah digunakan. Selain itu software yang juga cukup banyak peminatnya adalah Minitab, Statistica, S-Plus dan SAS meskipun lebih banyak digunakan oleh kalangan statistikawan.

Kendala terbesar dari software-software tersebut adalah lisensi yang harus dibayar kepada perusahaan pemegang lisensi software tersebut masih terlalu mahal, terutama bagi kalangan mahasiswa. Belum lagi apabila software-software tersebut memberikan jangka waktu lisensinya yang akan semakin memberatkan.

Beberapa pemegang lisensi software telah membuat software yang bisa digunakan secara gratis oleh kalangan mahasiswa atau untuk kepentingan pendidikan diantaranya R, Openstat dan MYSTAT 12.

Software R merupakan software statistika yang sangat "*powerfull*" dan merupakan software statistika yang perkembangannya paling cepat, sehingga banyak peneliti yang menggunakan software ini bahkan ikut berkontribusi terhadap software ini karena sifatnya yang *open source*. Akan tetapi software R bukan software yang *user friendly* karena membutuhkan pengetahuan tentang *syntax* dan bahasa pemrogramannya.

MYSTAT 12 merupakan software yang *user friendly* dan dapat dipergunakan secara gratis oleh kalangan mahasiswa. Software ini merupakan *student version* dari software SYSTAT. Selain *user friendly* software ini juga mempunyai tahap-tahap analisis yang mirip dengan SPSS yang saat ini (mungkin) paling banyak digunakan dikalangan mahasiswa di Indonesia. Meskipun demikian software MYSTAT 12 memiliki kemampuan analisis yang terbatas karena memang di-*setting* khusus untuk kalangan mahasiswa.

DAFTAR ISI

KATA PENGANTAR

BAB 1 ANALISIS KLASTER

1. Pendahuluan
2. Analisis Klaster dalam MYSTAT
3. Analisis Klaster Hierarki
4. Analisis Klaster Non-Hierarki (K-means Cluster)
5. Contoh Penggunaan

BAB 2 MULTIDIMENSIONAL SCALLING

1. Pendahuluan
2. Multidimensional Scalling dalam MYSTAT 12/SYSTAT
3. Contoh Penggunaan

BAB 3 ANALISIS KONJOIN

1. Pendahuluan
2. Analisis Konjoin Dalam MYSTAT 12/SYSTAT
3. Contoh Penggunaan

PENUTUP

DAFTAR PUSTAKA



BAB I

ANALISIS KLASTER

I. Pendahuluan

Secara sederhana Analisis Klaster digunakan untuk menentukan pengelompokan objek atau pengamatan yang didasarkan atas kemiripan objek. Objek yang berada dalam satu kelompok merupakan objek yang mirip satu sama lain dan sebaliknya yang berbeda kelompok merupakan objek yang tidak mirip satu sama lain. Berbagai metode digunakan untuk menentukan keanggotaan kelompok dalam analisis klaster. Secara umum analisis klaster dibagi menjadi dua yaitu

1. Analisis Klaster Hierarki: prosedurnya adalah membagi objek-objek kedalam beberapa kelompok melalui pautan hierarki, biasanya digambarkan dalam bentuk dendogram.
2. Analisis Klaster Non-Hirarki (*K-means Cluster*): prinsipnya adalah memisahkan suatu set objek kedalam sejumlah kelompok dengan memaksimalkan variasi antar klaster dan meminimumkan variasi didalam klaster.

Data yang digunakan dalam Analisis Klaster dapat berupa data pengamatan yang berdistribusi Multivariat atau bisa berupa sebuah matriks simetris *proximity* yaitu kemiripan antar objek (*similarity*) atau ketakmiripan antar objek (*dissimilarity*). Matriks korelasi bisa digunakan sebagai matriks *proximity* sebagai ukuran kemiripan sedangkan matriks jarak (misalnya *Euclidian Distance*) sebagai ukuran ketakmiripan.

II. Analisis Klaster dalam MYSTAT

A. Analisis Klaster Hierarki

Seperti telah dijelaskan sebelumnya bahwa analisis klaster hierarki biasanya digambarkan dalam diagram pohon atau yang biasa disebut **dendogram**. Secara umum kita dapat mengelompokkan objek pengamatan (*cases*) atau variabel.

Prosedurnya bahwa semua objek pada awalnya didefinisikan berada pada klaster yang berbeda, kemudian digabungkan objek-objek yang memiliki jarak terdekat dalam satu klaster dan seterusnya secara bertahap hingga diperoleh satu klaster.



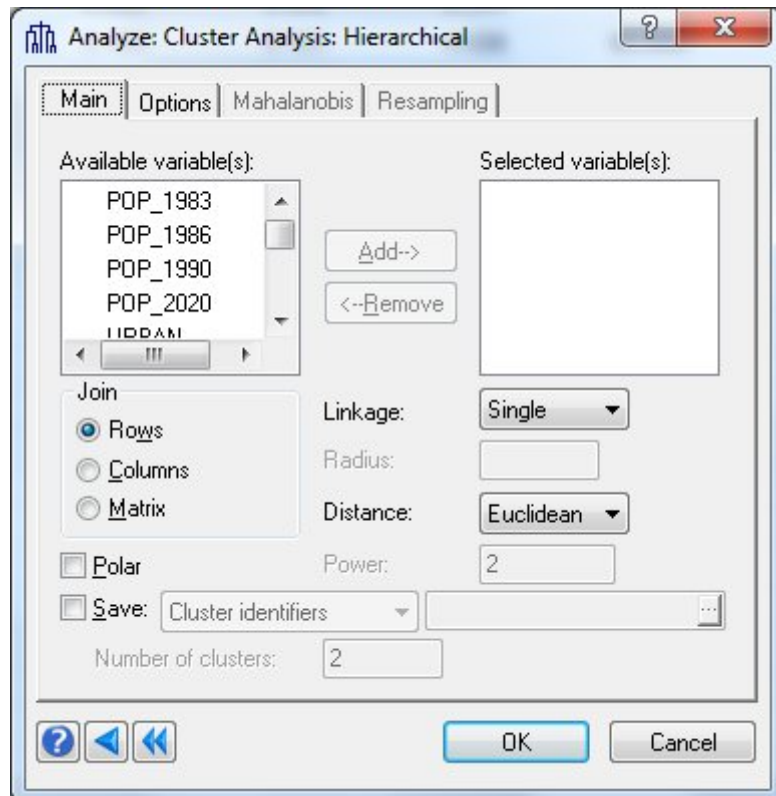
Dalam MYSTAT 12 analisis kluster hierarki dapat dibuka pada menu:

Analysis

Cluster Analysis

Hierarchical...

sehingga akan muncul jendela berikut ini



Pada gambar terlihat bahwa pada kotak sebelah kiri berisi variabel-variabel yang ada pada data (*Available variable(s)*). Pilih variabel-variabel terkait kemudian klik **Add-->** maka variabel-variabel terpilih akan pindah ke kotak sebelah kanan. Pilih apa yang akan diklusterkan pada kotak pilihan **Join**.

- **Rows:** apabila kita akan mengklusterkan objek pengamatan
- **Columns:** apabila kita akan mengklusterkan variabel
- **Matriks:** apabila kita akan mengkalsterkan objek pengamatan dan variabel

Selanjutnya pada menu **Lingka** memberikan pilihan algoritma yang digunakan dalam menggabungkan objek kedalam kluster, dengan kata lain mendefinisikan bagaimana jarak antar kluster diukur. Menu ini terdiri atas



- *Average*. Average linkage averages all distances between pairs of objects in different clusters to decide how far apart they are.
- *Centroid*. Centroid linkage uses the average value of all objects in a cluster (the cluster centroid) as the reference point for distances to other objects or clusters.
- *Complete*. Complete linkage uses the most distant pair of objects in two clusters to compute between-cluster distances. This method tends to produce compact, globular clusters. If you use a similarity or dissimilarity matrix from a SYSTAT file, you get Johnson's "max" method.
- *Median*. Median linkage uses the median distances between pairs of objects in different clusters to decide how far apart they are.
- *Single*. Single linkage defines the distance between two objects or clusters as the distance between the two closest members of those clusters. This method tends to produce long, stringy clusters. If you use a SYSTAT file that contains a similarity or dissimilarity matrix, you get clustering via Johnson's "min" method.
- *Ward*. Ward's method averages all distances between pairs of objects in different clusters, with adjustments for covariances, to decide how far apart the clusters are.

For some data, the last four methods cannot produce a hierarchical tree with strictly increasing amalgamation distances. In these cases, you may see stray branches that do not connect to others. If this happens, you should consider Single or Complete linkage.

Pilihan dalam menu **Distance** menspesifikasi ukuran jarak yang digunakan dalam penggabungan antar objek. Dalam menu ini terdapat pilihan sebagai berikut:

- *Chi-square*. Adalah jarak yang dihitung berdasarkan ukuran chi-kuadrat dari independensi baris dan kolom dalam tabel kontingensi $2 \times n$. Jarak ini digunakan untuk data hasil membilang (*counting*) dari objek atau kejadian.
- *Euclidean*. (default) adalah nilai jarak berdasarkan akar rata-rata kuadrat jarak antar dua objek. Ukuran jarak ini digunakan pada variabel kuantitatif. Bila ada data hilang tidak diikutsertakan dalam perhitungan.
- *Gamma*. yaitu jarak yang dihitung menggunakan 1 dikurangi koefisien korelasi Goodman-Kruskal Gamma. Ukuran jarak ini digunakan pada data



berskala ukur ordinal. Dalam ukuran jarak ini data hilang tidak diikutsertakan dalam perhitungan.

- *Minkowski*. adalah Jarak klaster yang dihitung menggunakan akar rata-rata koordinat kuasa jarak. Ukuran jarak ini digunakan pada variabel kuantitatif. Bila ada data hilang tidak diikutsertakan dalam perhitungan. Nilai kuasa (power) p dapat dispesifikasi pada Power text box.
- *Pearson*. adalah ukuran jarak yang dihitung menggunakan 1 dikurangi koefisien korelasi Pearson product-moment untuk setiap pasangan objek. Ukuran jarak ini digunakan pada variabel kuantitatif. Bila ada data hilang tidak diikutsertakan dalam perhitungan.
- *Percent* adalah jarak klaster yang diperoleh dari persentase perbandingan nilai hasil ketidaksesuaian dalam dua profil. Ukuran ini digunakan pada data kategori dengan skala ukur nominal.
- *Phi-square*. adalah jarak yang dihitung sebagai phi-square (χ^2/total) yang diukur dari tabel kontingensi $2 \times n$. Jarak ini digunakan untuk data hasil membilang (*counting*) dari objek atau kejadian.
- *RSquared*. adalah jarak yang dihitung menggunakan 1 dikurangi kuadrat dari koefisien korelasi Pearson product-moment dari setiap pasang objek. Ukuran jarak ini digunakan pada variabel kuantitatif. Bila ada data hilang tidak diikutsertakan dalam perhitungan.

Sebagai tambahan beberapa opsi yang dapat dipilih adalah sebagai berikut:

- ***Polar***: untuk menampilkan pohon klaster polar (*circular*).
- ***Save cluster identifiers***: digunakan untuk menyimpan hasil klastering setiap objek kedalam file. Dalam opsi ini dapat dipilih jumlah klaster yang diinginkan, secara *default* digunakan 2 klaster.

B. Analisis Klaster Non-Hierarki (*K-means Cluster*)

Penggunaan analisis klaster K-means berbeda dengan analisis klaster hierarki. Dalam K-means banyak klaster harus ditentukan diawal. Dalam MYSTAT 12 analisis *k-means klaster* dapat dibuka pada menu:

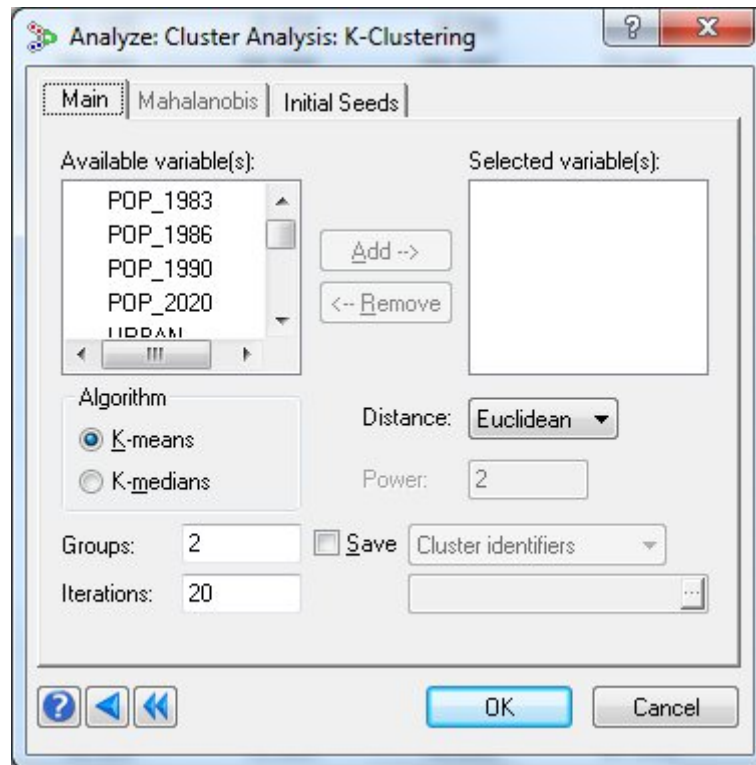


Analysis

Cluster Analysis

K-Clustering...

Sehingga muncul jendela seperti berikut ini



Secara umum penggunaan jendela ini sama dengan analisis kluster hierarki. kecuali untuk algoritma yang digunakan yaitu **K-means** atau **K-medians**.

Distance. Specifies the distance metric used to compare clusters.

Pilihan dalam menu **Distance** menspesifikasi ukuran jarak yang digunakan dalam penggabungan antar objek. Dalam menu ini terdapat pilihan sebagai berikut:

- *Chi-square.* Adalah jarak yang dihitung berdasarkan ukuran chi-kuadrat dari independensi baris dan kolom dalam tabel kontingensi $2 \times n$. Jarak ini digunakan untuk data hasil membilang (*counting*) dari objek atau kejadian.
- *Euclidean.* (default) adalah nilai jarak berdasarkan akar rata-rata kuadrat jarak antar dua objek. Ukuran jarak ini digunakan pada variabel kuantitatif. Bila ada data hilang tidak diikutsertakan dalam perhitungan.



- *Gamma*. yaitu jarak yang dihitung menggunakan 1 dikurangi koefisien korelasi Goodman-Kruskal Gamma. Ukuran jarak ini digunakan pada data berskala ukur ordinal. Dalam ukuran jarak ini data hilang tidak diikutsertakan dalam perhitungan.
- *Minkowski*. adalah Jarak klaster yang dihitung menggunakan akar rata-rata koordinat kuasa jarak. Ukuran jarak ini digunakan pada variabel kuantitatif. Bila ada data hilang tidak diikutsertakan dalam perhitungan. Nilai kuasa (power) p dapat dispesifikasi pada Power text box.
- *MW* adalah jarak yang dihitung sebagai penambahan simpangan jumlah kuadrat dalam klaster (*within sum of squares of deviations*) apabila suatu objek akan dimasukkan kedalam klaster. Sebuah objek dipindahkan kedalam klaster yang meminimumkan nilai simpangan jumlah kuadrat. Ukuran jarak ini digunakan pada variabel kuantitatif. Bila ada data hilang tidak diikutsertakan dalam perhitungan.
- *Pearson*. adalah ukuran jarak yang dihitung menggunakan 1 dikurangi koefisien korelasi Pearson product-moment untuk setiap pasangan objek. Ukuran jarak ini digunakan pada variabel kuantitatif. Bila ada data hilang tidak diikutsertakan dalam perhitungan.
- *Phi-square*. adalah jarak yang dihitung sebagai phi-square (chi-square/total) yang diukur dari tabel kontingensi $2 \times n$. Jarak ini digunakan untuk data hasil membilang (*counting*) dari objek atau kejadian.
- *RSquared*. adalah jarak yang dihitung menggunakan 1 dikurangi kuadrat dari koefisien korelasi Pearson product-moment dari setiap pasang objek. Ukuran jarak ini digunakan pada variabel kuantitatif. Bila ada data hilang tidak diikutsertakan dalam perhitungan.

Beberapa opsi tambahan yang dapat dipilih adalah sebagai berikut:

- **Groups**: untuk menentukan berapa banyak klaster yang digunakan. Secara default adalah 2 klaster
- **Iterations**: untuk menentukan banyak iterasi dalam perhitungan. Secara default adalah 20 iterasi.
- **Save cluster identifiers**: digunakan untuk menyimpan hasil klastering setiap objek kedalam file.



III. Contoh Penggunaan

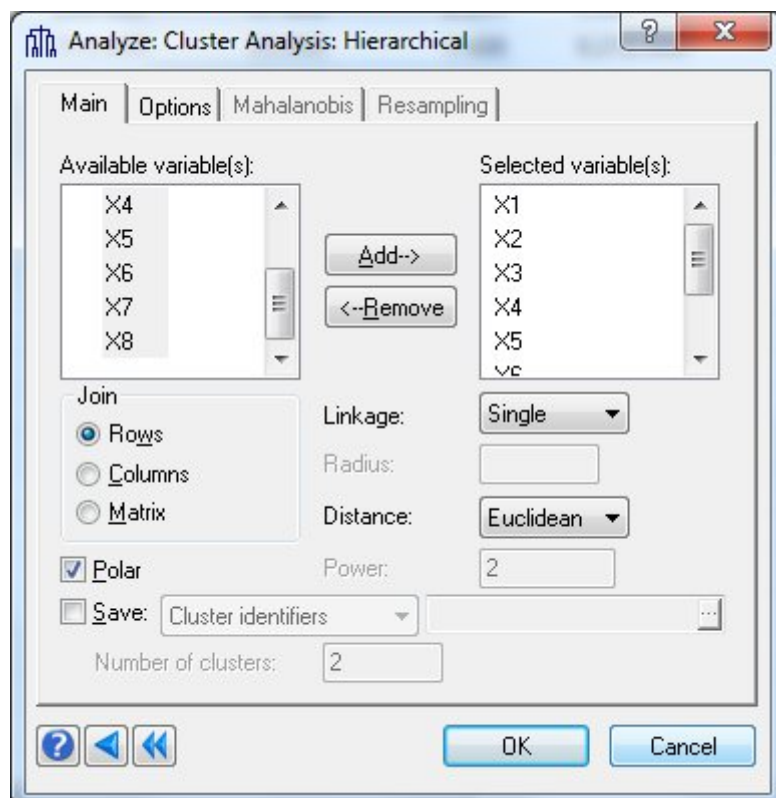
Data berikut dikumpulkan pada 22 perusahaan publik di Amerika Serikat pada tahun 1975, pada tiap perusahaan diukur variabel-variabel berikut ini:

- X_1 : Fixed-charge coverage ratio (income/debt)
- X_2 : Rate of Return on capital
- X_3 : Cost per KW capacity in place
- X_4 : Annual load factor
- X_5 : Peak kWh demand growth from 1974 to 1975
- X_6 : Sales (kWh use per year)
- X_7 : Percent nuclear
- X_8 : Total Fuel cost (cents per kWh)

Dengan menggunakan data tersebut perusahaan-perusahaan akan dikelompokkan berdasarkan kedelapan variabel yang telah diukur. Data dapat dilihat dalam Johnson & Wichern (2002) halaman 687.

Hal pertama yang harus dilakukan adalah mempersiapkan data agar dapat dianalisis. Dalam kasus ini semua variabel adalah kuantitatif sehingga data tinggal diinput kedalam sheet, ubah nama variabel dengan X1 sampai X8 kemudian simpan dengan nama **public.syz**.

Untuk metode hierarki: buka menu Hierarchical... seperti telah dijelaskan sebelumnya dan pilih semua variabelnya dan klik Add.





Untuk opsi **Linkage** pilih **Single** dan opsi **Distance** pilih **Euclidean**. Centang opsi **Polar** untuk menampilkan dendrogram, kemudian klik **OK**. Maka output yang dihasilkan adalah sebagai berikut

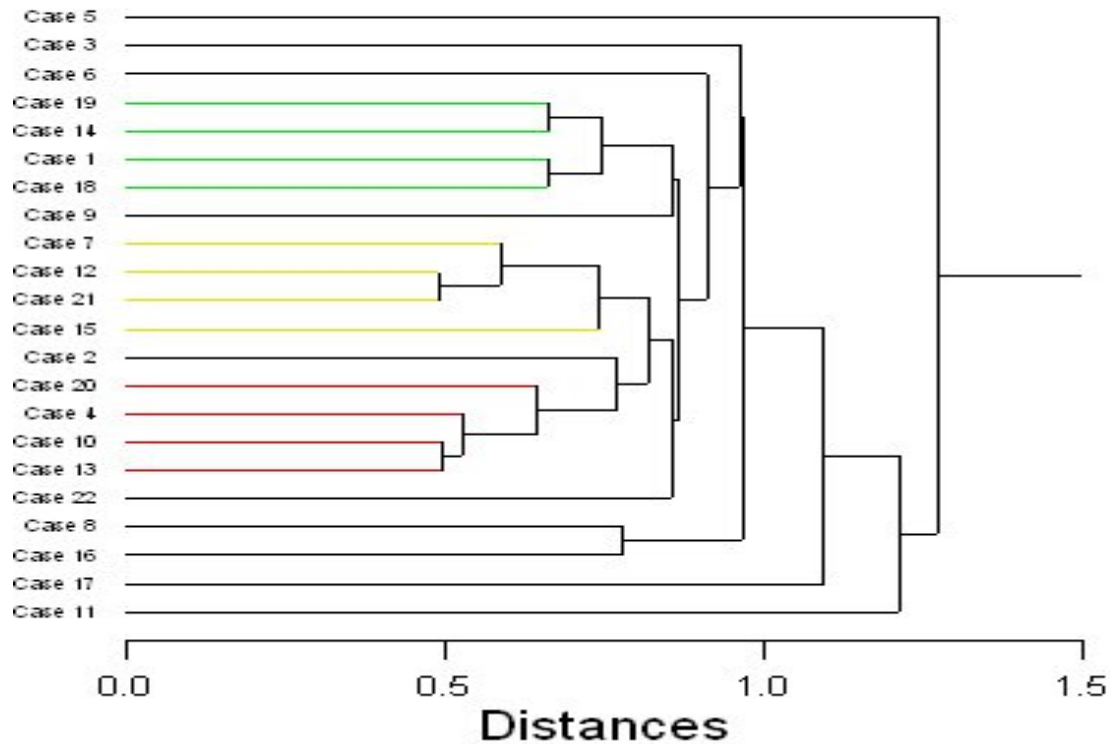
▼ Hierarchical Clustering

**Distance Metric is Euclidean Distance
Single Linkage Method (Nearest Neighbor)**

	Clusters Joining	at Distance	No. of Members
Case 21	Case 12	0.490	2
Case 13	Case 10	0.497	2
Case 13	Case 4	0.527	3
Case 21	Case 7	0.588	3
Case 20	Case 13	0.642	4
Case 19	Case 14	0.663	2
Case 18	Case 1	0.664	2
Case 15	Case 21	0.743	4
Case 19	Case 18	0.747	4
Case 2	Case 20	0.767	5
Case 16	Case 8	0.778	2
Case 15	Case 2	0.819	9
Case 22	Case 15	0.856	10
Case 19	Case 9	0.859	5
Case 22	Case 19	0.867	15
Case 22	Case 6	0.911	16
Case 22	Case 3	0.963	17
Case 22	Case 16	0.968	19
Case 17	Case 22	1.095	20
Case 17	Case 11	1.213	21
Case 5	Case 17	1.274	22

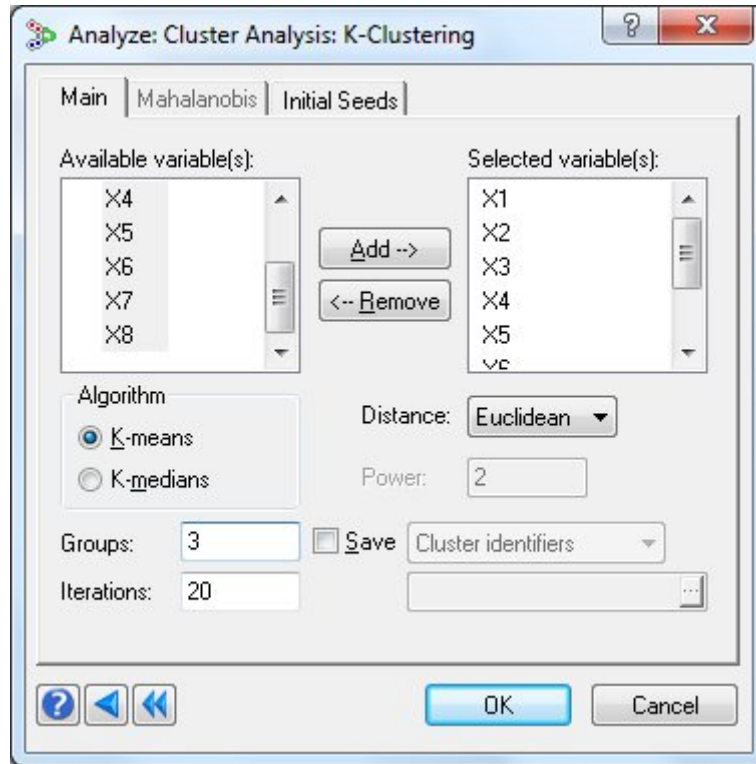


Cluster Tree



Dari dendrogram diatas terlihat ada tiga kelompok objek yaitu yang berwarna kuning, hijau dan merah, selainnya berwarna hitam berdiri sendiri. Memotong dendrogram pada jarak tertentu akan menghasilkan klaster yang diinginkan misalnya: jika dendrogram dipotong pada jarak 1.2 maka akan menghasilkan dua klaster, jika dendrogram dipotong pada jarak 1.1 maka akan menghasilkan 3 klaster dan seterusnya. Penggunaan opsi **Linkage** dan **Distance** yang berbeda akan menghasilkan nilai jarak yang berbeda dan dendrogram yang berbeda pula.

Untuk metode K-means: buka menu *K-Clustering* dan pilih semua variabel kemudian klik **Add**. Pada opsi **Algorithm** kita pilih **K-means** sedangkan pada opsi **Distance** kita pilih **Euclidean**.



Misalnya dalam kasus ini kita menginginkan banyak grup adalah 3 maka pada pilihan **Groups** isi dengan angka 3 dan iteration kita biarkan saja defaultnya yaitu 20. Klik OK maka akan muncul hasil output sebagai berikut

[▼K-Means Clustering](#)

**Distance Metric is Euclidean Distance
K-Means splitting cases into 3 groups**

Summary Statistics for All Cases					
Variable	Between SS	df	Within SS	df	F-ratio
X1	2.486	2	18.514	19	1.276
X2	9.341	2	11.659	19	7.612
X3	8.645	2	12.355	19	6.648
X4	6.602	2	14.398	19	4.356
X5	3.532	2	17.468	19	1.921
X6	15.540	2	5.460	19	27.038
X7	3.558	2	17.442	19	1.938
X8	16.581	2	4.419	19	35.641
** TOTAL	66.286	16	101.714	152	
**					



Cluster 1 of 3 Contains 12 Cases						
Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	Standard Deviation
Case 1	0.650	X1	-0.510	0.289	1.712	0.709
Case 3	0.910	X2	-0.685	0.593	2.078	0.756
Case 4	0.625	X3	-1.752	-0.484	0.748	0.858
Case 6	0.884	X4	-1.609	-0.346	0.767	0.735
Case 9	0.858	X5	-1.745	-0.216	1.270	0.971
Case 10	0.715	X6	-0.702	0.038	1.294	0.586
Case 13	0.972	X7	-0.715	0.347	2.275	1.181
Case 14	0.780	X8	-1.032	-0.488	0.371	0.495
Case 18	0.502					
Case 19	0.891					
Case 20	0.648					
Case 22	0.750					

Cluster 2 of 3 Contains 7 Cases						
Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	Standard Deviation
Case 2	0.644	X1	-1.919	-0.239	2.037	1.298
Case 5	1.179	X2	-1.932	-0.659	0.652	0.891
Case 7	0.827	X3	-0.781	0.256	0.870	0.579
Case 12	0.356	X4	-1.295	0.799	2.381	1.125
Case 15	0.600	X5	-1.071	-0.054	1.847	0.940
Case 17	1.077	X6	-1.581	-0.860	-0.358	0.389
Case 21	0.368	X7	-0.715	-0.288	0.792	0.593
		X8	0.533	1.250	1.828	0.478

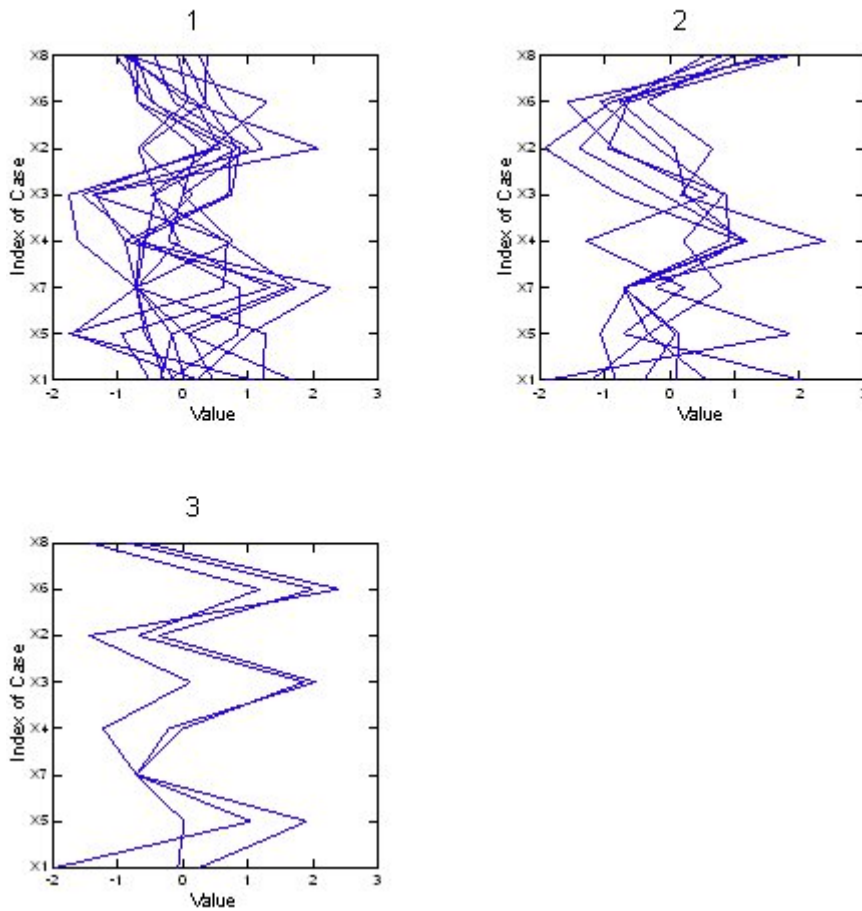
Cluster 3 of 3 Contains 3 Cases						
Members			Statistics			
Case	Distance	Variable	Minimum	Mean	Maximum	Standard Deviation
Case 8	0.552	X1	-1.973	-0.600	0.249	1.200
Case 11	0.770	X2	-1.442	-0.833	-0.373	0.550
Case 16	0.543	X3	0.117	1.339	2.035	1.062
		X4	-1.228	-0.481	0.005	0.657
		X5	0.019	0.992	1.911	0.947
		X6	1.174	1.857	2.402	0.625
		X7	-0.715	-0.715	-0.715	0.000
		X8	-1.427	-0.966	-0.600	0.422



Dari hasil diatas terlihat keanggotaan untuk masing-masing kluster. Kluster 1 mempunyai anggota paling banyak yaitu 11, kluster 2 mempunyai anggota 7 dan kluster 3 mempunyai anggota paling sedikit yaitu 3.

Selain itu dapat dilihat gambaran nilai antar variabel untuk seluruh anggota ketiga kluster pada diagram Parallel Kluster berikut

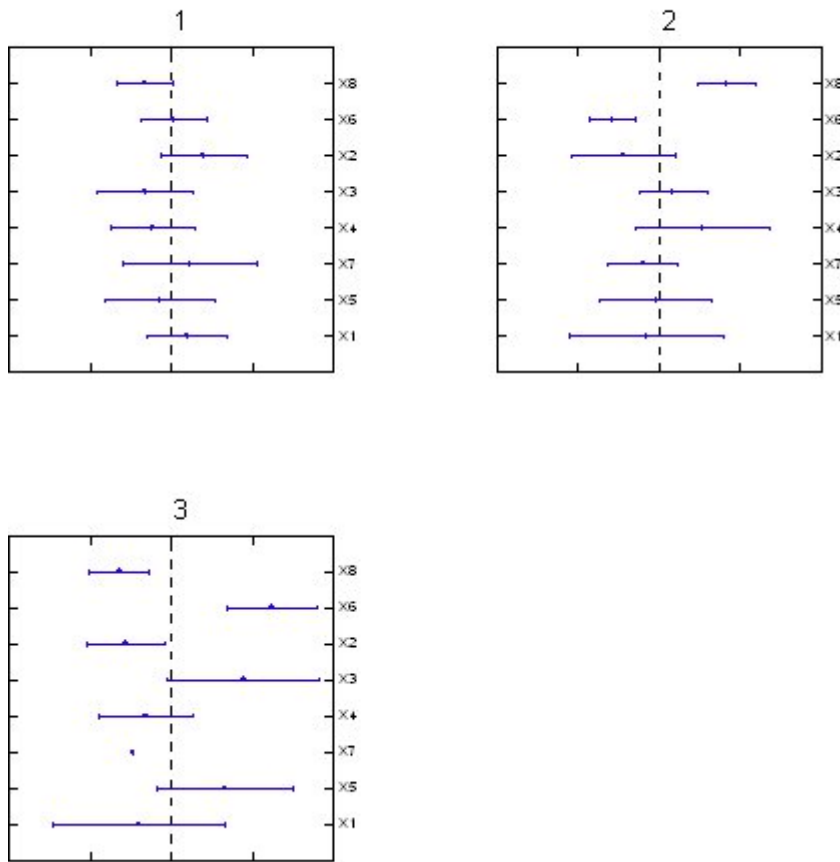
Cluster Parallel Coordinate Plots



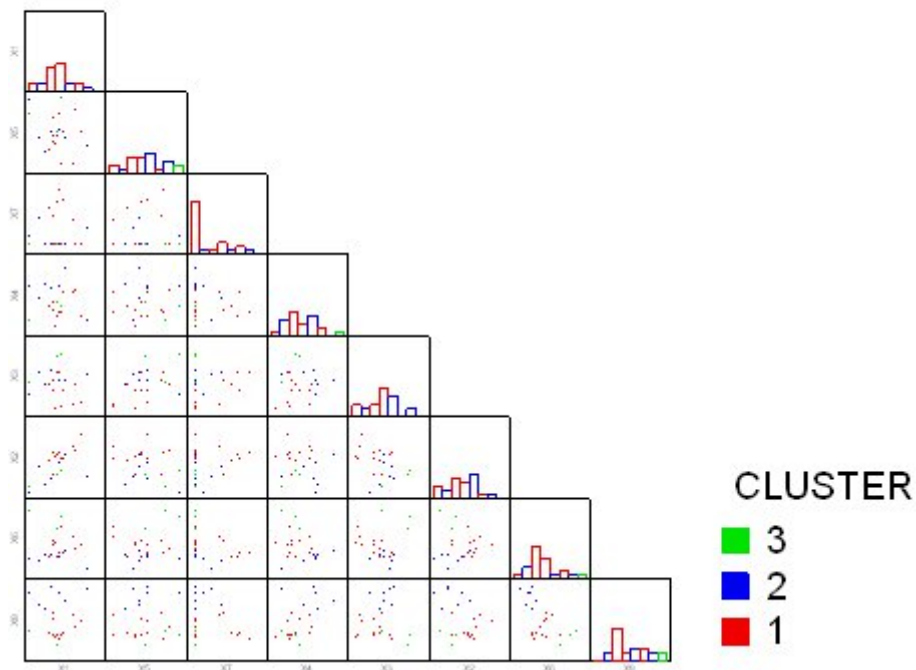
Lebih jelasnya dapat pula diperhatikan nilai rata-rata, dan batas-batas interval untuk setiap variabel dari ketiga kluster pada diagram profile cluster berikut ini:



Cluster Profile Plots



Cluster SPLOM





BAB 2

MULTIDIMENSIONAL SCALLING

I. Pendahuluan

Multidimensional Scalling (MDS) adalah suatu teknik statistika multivariat yang bertujuan untuk menganalisis kemiripan (*similarity*) dan ketakmiripan (*dissimilarity*) antar objek. Hasil dari analisis multidimensional scalling adalah berupa gambar titik-titik yang mana jarak antar titik menunjukkan kemiripan atau ketakmiripan. Selain itu penggunaan multidimensional scalling juga dapat memberikan petunjuk untuk mengidentifikasi peubah yang tidak diketahui atau faktor yang memengaruhi munculnya kemiripan atau ketakmiripan.

MDS bertujuan untuk membuat peta/konfigurasi posisi objek dalam ruang berdimensi rendah (umumnya 2 dimensi) berdasarkan data jarak antar objek atau data multivariate yang sebelumnya diubah dulu menjadi matriks jarak. Sehingga MDS akan sangat berguna untuk mendapatkan posisi relatif suatu objek dibandingkan objek lain. Dalam banyak kasus strategi bisnis, digunakan untuk menentukan pesaing dan benchmarking

Secara umum MDS dibagi menjadi dua yaitu

- MDS METRIK untuk data berskala interval atau rasio. Secara umum dalam metrik jarak dianggap bertipe rasio.
- MDS NON-METRIK untuk data berskala ordinal atau nominal. Dalam non-metrik jarak dianggap bertipe ordinal. Yang terpenting adalah urutannya benar, walaupun rasionya tidak sesuai maka masih diperbolehkan.

Hasil MDS dapat diukur tingkat kebaikannya menggunakan kriteria Stress sebagai berikut

Stress	Kriteria
> 20%	Kurang
10% - 20%	Cukup
5% - 10%	Baik
2.5% - 5%	Sangat Baik
< 2.5%	Sempurna



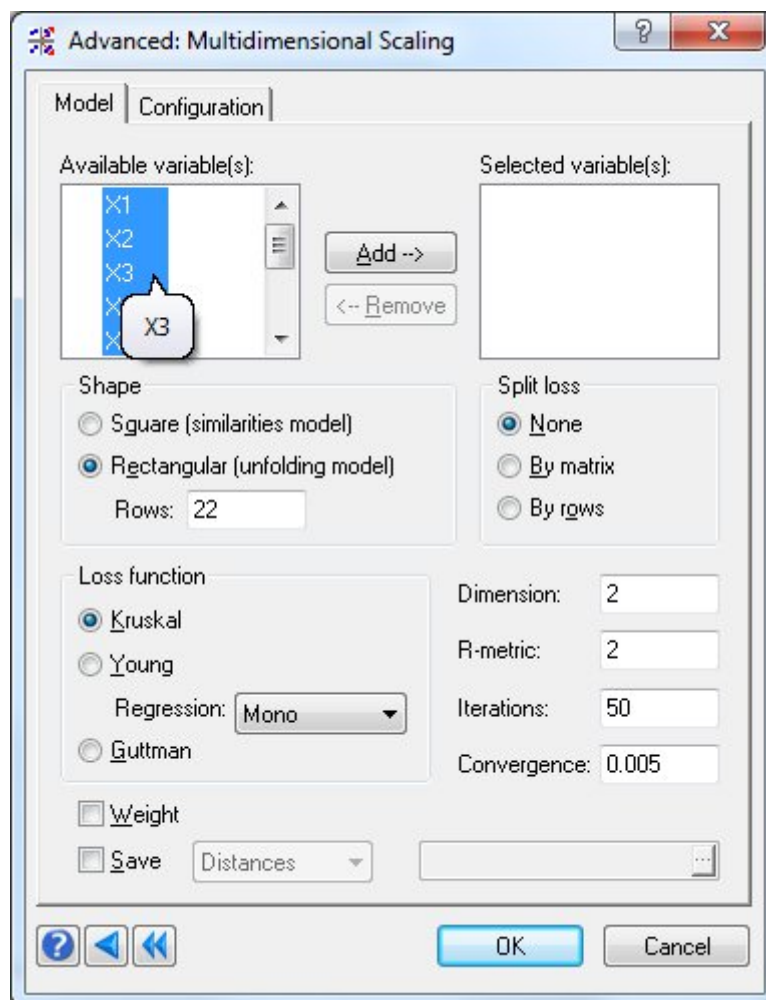
II. Multidimensional Scalling dalam MYSTAT/SYSTAT

Secara *default* MYSTAT 12 tidak menyediakan menu Multidimensional scalling. Untuk dapat menggunakan MDS kita perlu mendapatkan versi lengkap dari software ini yaitu SYSTAT 12. Prosedurnya adalah sebagai berikut

Advance

Multidimensional Scalling

maka akan muncul jendela dialog seperti dibawah ini



Dalam kotak dialog Multidimensional Scalling terdapat beberapa opsi yang dapat kita pilih yaitu:

- **Selected variable(s).** adalah kotak yang berisi variabel-variabel terpilih data matriks yang akan dianalisis.



- **Shape.** untuk spesifikasi input matriks. Pilih *square* untuk model similarities dan pilih *Rectangular* untuk model unfolding kemudian masukkan banyak baris dalam matriksnya.
- **Loss function.** adalah fungsi penskalaan matriks *similarity* dan *dissimilarity* menggunakan tiga fungsi yaitu:
 - *Kruskal*: Metode STRESS Kruskal (Formula 1)
 - *Guttman*. Menggunakan koefisien Guttman untuk metode penskalaan alienation.
 - *Young*: Metode penskalaan S-STRESS Young yang mengijinkan kita menggunakan fitur fungsi loss dalam ALSCAL.
- Apabila loss function Kruskal yang digunakan maka pilih fungsi yang menghubungkan jarak dengan similarities (atau dissimilarities) yaitu:
 - *Mono*. untuk penskalaan nonmetrick.
 - *Linear*. untuk penskalaan metrik.
 - *Log*. untuk spesifikasi fungsi log berupa penghalusan kurva linier untuk hubungan dissimilarities dengan jarak.
 - *Power*. untuk spesifikasi fungsi kuasa.
- **Split loss.** Untuk pembedaan individual dari model *unfolding*, membagi perhitungan fungsi loss berdasarkan baris-baris dari matriks (*by rows*) atau berdasarkan matriks (*by matrices*).
- **Dimension.** banyak dimensi yang akan diskalakan. Banyak dimensi harus bernilai bilangan bulat positif kurang dari atau sama dengan jumlah variabel yang akan diskalakan.
- **R-metric.** Nilai konstanta dari *Minkowski power metric* untuk menghitung jarak. Untuk jarak **Euclidean**, masukkan nilai 2. Untuk jarak **city-block**, masukkan nilai 1. Untuk nilai-nilai selain 1 atau 2, perhitungan akan lebih lambat karena logaritma dan eksponen digunakan. Rumus umum untuk menghitung jarak adalah

$$d_{jk} = \left(\sum_{i=1}^p |x_{ij} - x_{ik}|^r \right)^{\frac{1}{r}}$$

dengan r adalah nilai kuasa dan p adalah banyaknya dimensi.

- **Iterations.** Batas iterasi.
- **Convergence.** adalah kriteria berakhirnya iterasi. Iterasi berhenti pada saat terdapat perbedaan absolut antara suatu koordinat pada iterasi ke- i dengan



koordinat pada iterasi ke- $i-1$ kurang dari nilai yang ditetapkan. Nilai default untuk Convergence adalah 0.005.

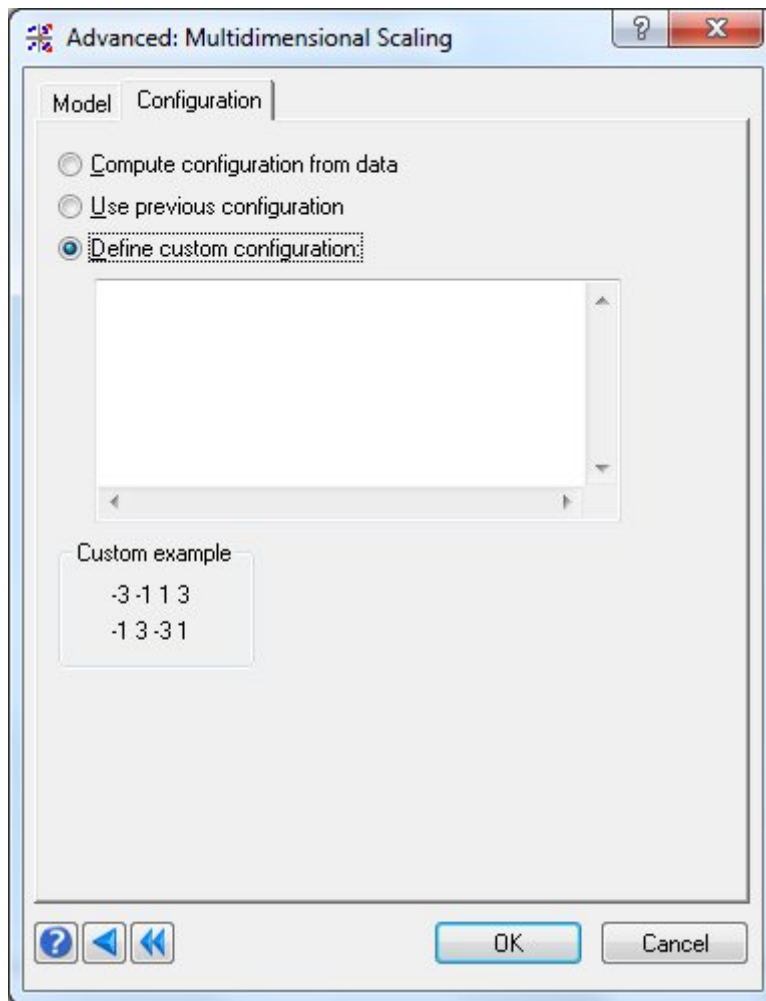
- **Weight.** adalah nilai bobot untuk setiap dimensi dan setiap subjek (matriks) untuk perhitungan nilai jarak secara terpisah yang digunakan dalam minimalisasi. Untuk model perbedaan individu pilih **Weight**.
- **Save.** Untuk menyimpan hasil perhitungan dalam file. Hasil yang dapat disimpan adalah:
 - *Configuration.* Untuk Konfigurasi akhir.
 - *Distances.* Untuk matriks jarak antar titik pada konfigurasi akhir.
 - *Residuals.* Untuk menyimpan data, jarak, taksiran jarak, residual, dan jumlah baris dan kolom dari jarak sebenarnya.

Residual dalam MDS menampilkan akar rata-rata kuadrat residual untuk setiap titik dalam output. Karena STRESS adalah fungsi jumlah kuadrat residual, akar rata-rata kuadrat residual adalah ukuran gangguan dari setiap titik dalam statistik STRESS.

Konfigurasi MDS

Dalam SYSTAT kita dapat menentukan beberapa alternatif untuk konfigurasi awal MDS. Untuk menentukan konfigurasi buka tab *configurations* kemudian pilih

- *compute configuration from data*
- *use previous configuration*
- *define custom configuration.*



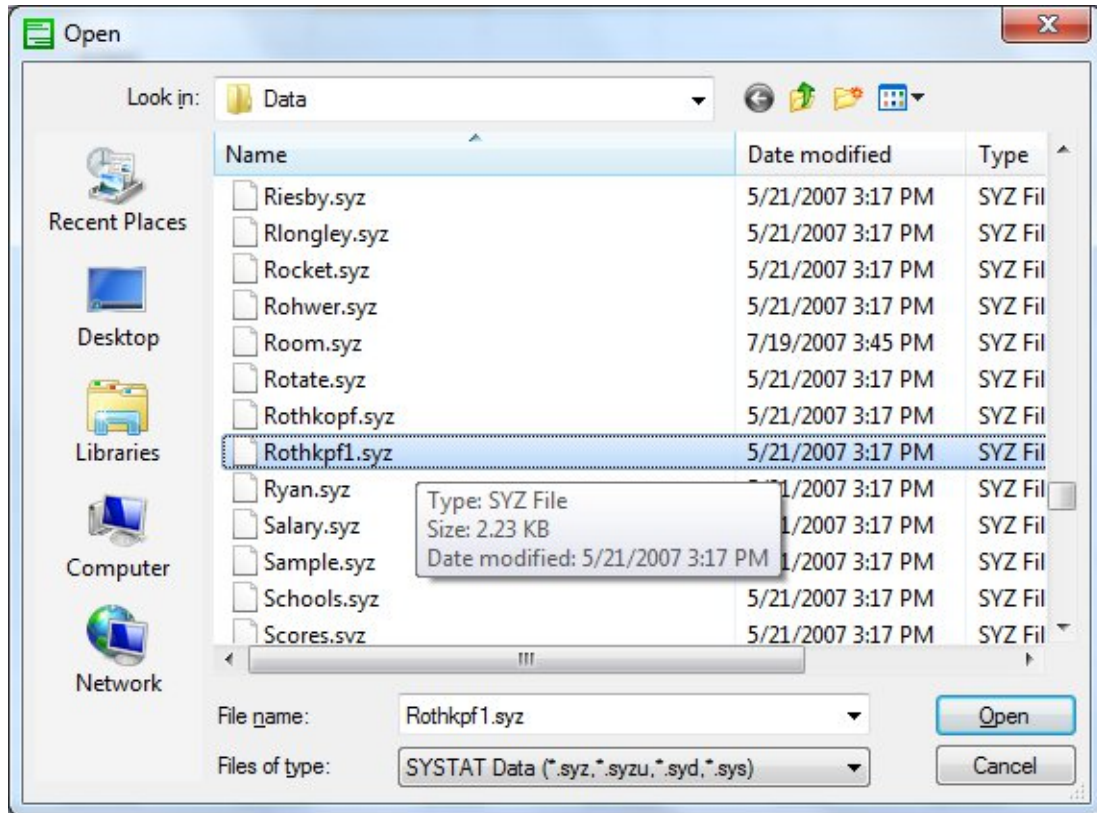
Secara *default* yang digunakan adalah *compute configuration from data* yaitu konfigurasi dihitung dari data dengan metode bergantung pada fungsi loss.

Apabila kita memilih *Use previous configuration* maka konfigurasi yang digunakan adalah konfigurasi pada penskalaan sebelumnya.

Apabila kita memilih *Define custom configuration* maka kita harus menginputkan konfigurasi sesuai dengan banyak dimensi, baris dan kolom dari matriks data yang akan diskalakan.

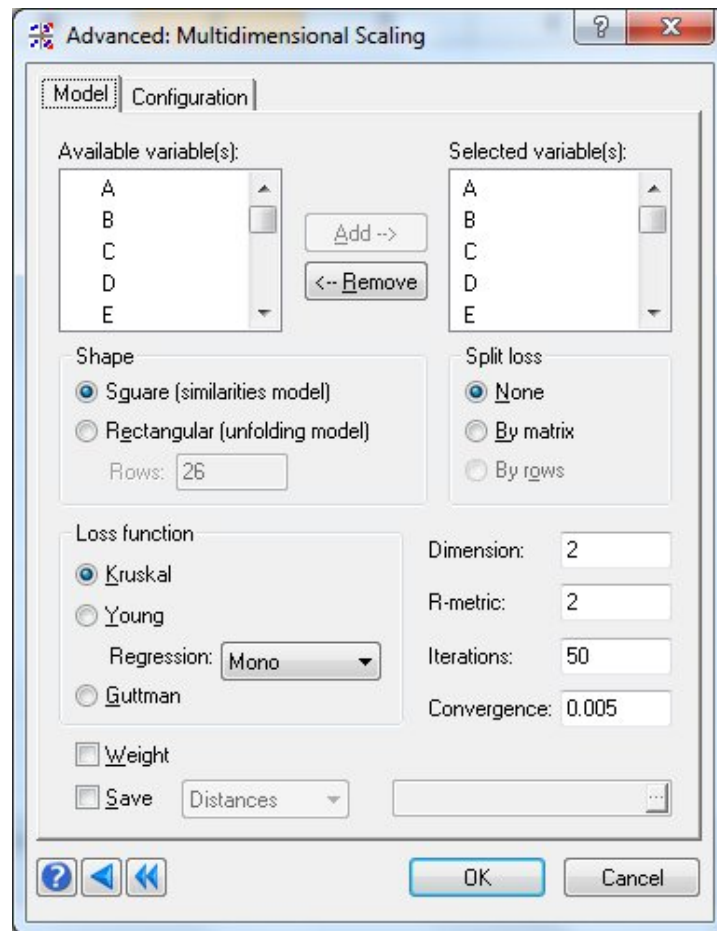
III. Contoh Penggunaan

Data yang digunakn dalam contoh ini adalah data ROTHKPF1 yang ada pada directory Data dalam software MYSTAT/SYSTAT. Klik ikon atau menu file > open dan pilih data "**Rothkpf1.syz**"



Data tersebut diadaptasi dari percobaan yang dilakukan oleh Rothkopf (1957). Data diperoleh dari 589 subjek yang menilai sinyal kode morse yang ditampilkan apakah sama atau tidak. Sinyal kode morse untuk huruf dan digit digunakan dalam percobaan tersebut, dan semua pasangan diuji dalam setiap dua urutan kemungkinan. Untuk keperluan *multidimensional scaling*, data untuk sinyal huruf dirata-ratakan diantara urutannya, dan diagonal matriks (yaitu pasangan dengan sinyal yang sama) diabaikan. Data tersebut pertama-tama di-skala-kan menggunakan Shepard.

Untuk menganalisis data ini buka menu *Advanced > Multidimensional Scalling* kemudian pilih semua variabel A sampai Z, dalam **Shape** gunakan *Square (similarities model)*, dalam **Split Loss** pilih *None*, yang lainnya biarkan sesuai dengan defaultnya



maka akan keluar output sebagai berikut

Number of Variables : 26
Number of Cases : 26

SYSTAT Similarity file C:\Program Files (x86)\SYSTAT 12\Data\Rothkopf.syz,
created Mon May 21 15:17:38 2007, contains variables:

A	B	C	D	E	F
G	H	I	J	K	L
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z				

Number of Variables : 27
Number of Cases : 26

SYSTAT Similarity file C:\Program Files (x86)\SYSTAT 12\Data\Rothkpf1.syz,
created Mon May 21 15:17:38 2007, contains variables:

CODE	A	B	C	D	E
\$					
F		G	H	I	J
L		M	N	O	P
R		S	T	U	V
X		Y	Z		



▼ Multidimensional Scaling

Monotonic Multidimensional Scaling

Kruskal Method

The data are analyzed as similarities

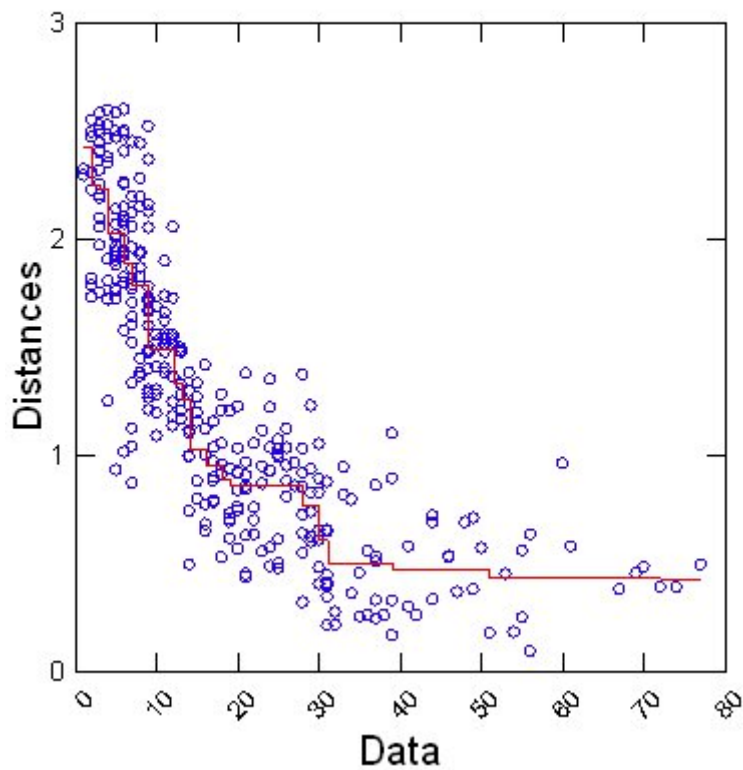
Minimizing Kruskal STRESS (form 1) in 2 dimensions

Iteration	History
Iteration	STRESS
n	
0	0.264
1	0.238
2	0.219
3	0.202
4	0.191
5	0.184
6	0.181
7	0.179
8	0.178

Stress of Final Configuration : 0.178

Proportion of Variance (RSQ) : 0.845

Shepard Diagram

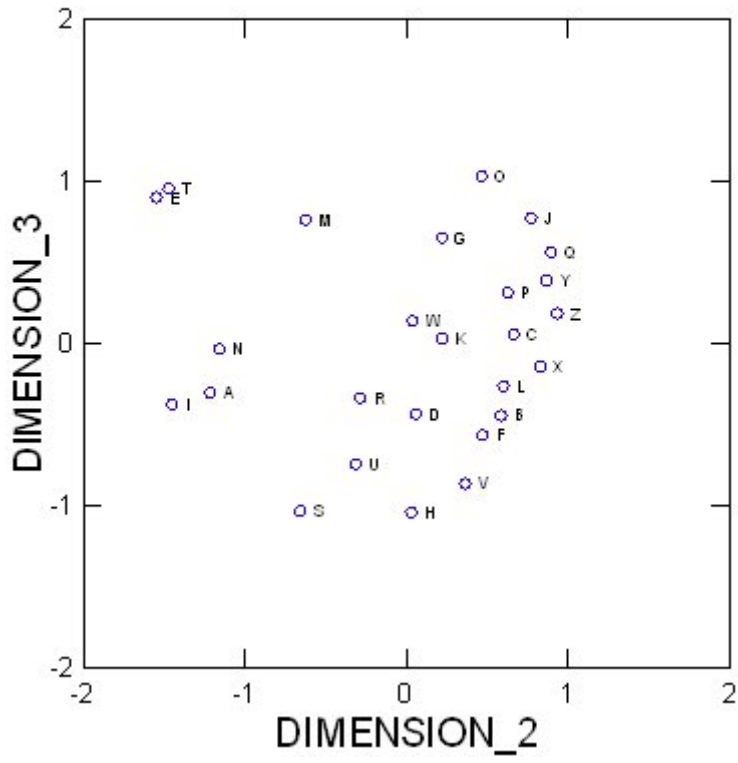




Coordinates in 2 Dimensions Variable	Dimension	
	1	2
	A	-1.211
B	0.588	-0.450
C	0.668	0.050
D	0.062	-0.440
E	-1.543	0.893
F	0.476	-0.572
G	0.224	0.646
H	0.032	-1.047
I	-1.447	-0.382
J	0.776	0.766
K	0.225	0.025
L	0.603	-0.270
M	-0.622	0.758
N	-1.154	-0.042
O	0.469	1.025
P	0.630	0.306
Q	0.897	0.556
R	-0.284	-0.344
S	-0.656	-1.039
T	-1.469	0.948
U	-0.311	-0.751
V	0.366	-0.870
W	0.042	0.131
X	0.833	-0.149
Y	0.871	0.382
Z	0.936	0.179



Configuration





MODUL 7

ANALISIS KONJOIN

1. Pendahuluan

Pengertian Analisis Konjoin telah didefinisikan oleh beberapa pakar diantaranya adalah sebagai berikut

- Hair et al (1998;392) : “Analisis Conjoint adalah salah satu teknik multivariate yang digunakan secara khusus untuk memahami bagaimana responden menentukan pilihan terhadap suatu produk atau jasa.
- Kotler (2000:339) “...suatu metode yang menurunkan nilai – nilai kegunaan dari berbagai level atribut produk yang diberikan pada konsumen”.
- Churchill dan Lacobucci (2002:748) "Analisis Conjoint menekankan pada kemampuan responden untuk membuat pertimbangan dari suatu stimuli. Stimuli ini menggambarkan beberapa kombinasi atribut yang telah ditetapkan sebelumnya".

Dari berbagai definisi tersebut Analisis Konjoin (*Conjoint Analysis*) dapat diartikan sebagai suatu metode analisis dalam analisis multivariat yang digunakan untuk membantu mendapatkan kombinasi atau komposisi atribut-atribut suatu produk atau jasa baik baru maupun lama yang paling disukai konsumen. Pada analisis ini konsumen akan diminta untuk membuat suatu pertimbangan pertukaran (*trade-off judgement*) atribut. Seberapa besar kesukaan konsumen terhadap suatu atribut dinilai cukup untuk mengorbankan atribut lain atau jika konsumen telah mempertimbangkan untuk mengorbankan suatu atribut untuk mendapatkan atribut lain, maka atribut mana yang akan dia ambil.

Dalam prosesnya analisis konjoin akan memberikan ukuran kuantitatif terhadap tingkat kegunaan (*utility*) dan kepentingan relatif (*relatif importance*) suatu atribut dibandingkan dengan atribut lain. Hal ini dilakukan melalui pertimbangan psikologis atau preferensi konsumen (Green & Tull, 1988). Lebih lanjut, nilai-nilai ini dapat digunakan untuk membantu menyeleksi atribut-atribut suatu produk yang akan ditawarkan.

Analisis konjoin umumnya ditujukan untuk mengidentifikasi atribut dan tarafnya. Atribut atau taraf atribut yang akan digunakan dalam merancang *stimuli* (kombinasi antar taraf atribut), dari sisi teori, sangat disarankan merupakan atribut yang memiliki peran dalam mempengaruhi preferensi konsumen dalam memilih produk. Selain itu analisis konjoin juga digunakan untuk mengidentifikasi kkala taraf atribut dan



model preferensi. Skala taraf atribut dan model preferensi merupakan salah satu masalah penting yang perlu diperhatikan dalam analisis konjoin. Hal ini akan berimplikasi pada proses optimalisasi kombinasi taraf antar atribut (*stimuli*). Dilihat dari sisi skala pengukurannya, atribut dapat berskala kualitatif atau kategori (*nominal* atau *ordinal*) atau kuantitatif (*interval* atau *ratio*). Merek (*brand*) umumnya berskala kualitatif, sedangkan harga berskala kuantitatif.

Dalam beberapa dekade terakhir analisis konjoin telah sedemikian populer, terutama dikalangan para peneliti pemasaran dan ahli ekonomi, untuk menganalisis preferensi konsumen berdasarkan beberapa atribut (Green and Srinivasan (1978, 1990), Crowe (1980), Louviere (1988)). Inti dari teknik ini adalah mengembangkan produk dengan berbagai atribut berdasarkan preferensi konsumen.

Dalam studi konjoin, data yang diperlukan dapat berupa *nonmetrik* (data dalam bentuk *nominal* atau *ordinal* atau kategori) maupun *metrik* (data berskala *interval* atau *rasio*). Untuk data berjenis *nonmetrik*, responden diminta untuk membuat *ranking* atau mengurutkan *stimuli* yang telah dibuat pada tahap sebelumnya. Secara teori perangkingan dapat dipandang sebagai evaluasi secara relatif terhadap taraf-taraf atribut. Nilai rangking ini dipercaya akan mencerminkan perilaku konsumen dalam situasi nyata. Pengurutan ini biasanya dimulai dari *stimuli* yang paling disukai sampai pada *stimuli* yang paling tidak disukai. Untuk *stimuli* yang paling disukai diberi nilai mulai dari 1 dan seterusnya hingga rangking terakhir bagi *stimuli* yang paling tidak disukai.

Untuk memperoleh data dalam bentuk *metrik*, responden diminta untuk memberikan *rating* atau nilai terhadap masing-masing *stimuli*. Melalui cara ini responden akan dapat memberikan penilaian terhadap masing-masing *stimuli* secara terpisah. Pemberian nilai atau *rating* dapat dilakukan melalui beberapa cara yaitu menggunakan skala *lickert* mulai dari 1 hingga 9 (1 = Paling tidak disukai, 9 = Paling disukai). Menggunakan nilai *ranking* terbalik, artinya untuk *stimuli* yang paling disukai diberi nilai tertinggi setara dengan jumlah *stimulinya*, sedangkan *stimuli* yang paling tidak disukai diberi nilai 1.

Saat ini terdapat beberapa metode atau prosedur yang dapat digunakan untuk menyelesaikan model dasar dari analisis konjoin. Umumnya metode-metode ini akan sangat bergantung pada tatacara pengumpulan data yang dilakukan. Metode yang umum digunakan dalam analisis konjoin adalah Regresi Dengan Variabel *Dummy*. Metode ini sangat populer digunakan untuk jenis data *nonmetrik* maupun *metrik* seperti yang telah diuraikan terdahulu, dimana data tersebut diperoleh melalui pengurutan maupun penilaian terhadap kombinasi atribut atau *stimuli* yang telah dirancang sebelumnya.



Beberapa variasi penggunaan metode regresi dengan variabel *dummy* (*dummy variable*) akan sangat tergantung pada tatacara penilaian atribut. Bila data yang digunakan berasal dari penilaian *stimuli* yang telah dirancang sebelumnya, dan penilaian dilakukan dengan menggunakan skala *metrik*, maka regresi dengan variabel *dummy* dapat dihitung langsung dengan menggunakan pendekatan *Ordinary Least Square* (OLS). Bila penilaian *stimuli* menggunakan urutan (*ranking*) *stimuli*, maka terlebih dahulu data tersebut harus dirubah menjadi skala *interval* dengan menggunakan *monotonic regression* atau menggunakan *Multidimensional Scaling* (MDS), cara ini dikenal juga sebagai *Linear Mapping* (LINMAP) yang dikombinasikan dengan *Monotonic Variance Analysis* (MONOVA) . Kemudian analisis dilanjutkan dengan regresi dengan variabel *dummy*. Bila data diperoleh melalui penilaian secara terpisah dari masing-masing atribut (taraf atribut), maka cara ini dikenal dengan istilah *discret choice*, dimana variabel *dependen*-nya umumnya berupa intensitas pilihan atau aktual pembelian. Analisis yang digunakan adalah LOGIT model. Variabel *dummy* adalah suatu bilangan yang dibangkitkan dari taraf-taraf atribut dengan ketentuan: suatu variabel diberi nilai 1 bila taraf yang bersangkutan ada 0 bila tidak ada. Jumlah variabel *dummy* dari suatu atribut ada sebanyak $p-1$, dengan p adalah banyaknya taraf dalam suatu atribut.

Setelah variabel dan jenis data ditentukan, tahap berikutnya adalah menentukan apakah data yang akan di analisis secara individual atau secara *agregat*. Untuk hal ini ada tiga metodologi yang dapat dipertimbangkan yaitu:

1. *Traditional conjoint*: Suatu metodologi yang memperkenalkan analisis dilakukan dalam *level* individu dengan jumlah atribut yang diteliti maksimal 9 atribut.
2. *Adaptive conjoint*: Seperti halnya *traditional conjoint* namun jumlah atributnya maksimal 30 atribut.
3. *Choice-based conjoint*: Suatu metodologi yang memperkenankan analisis dilakukan baik dalam *level* individu maupun keseluruhan (*aggregate*) dengan jumlah atribut maksimal 6 atribut.

II. Analisis Konjoin dalam MYSTAT/SYSTAT

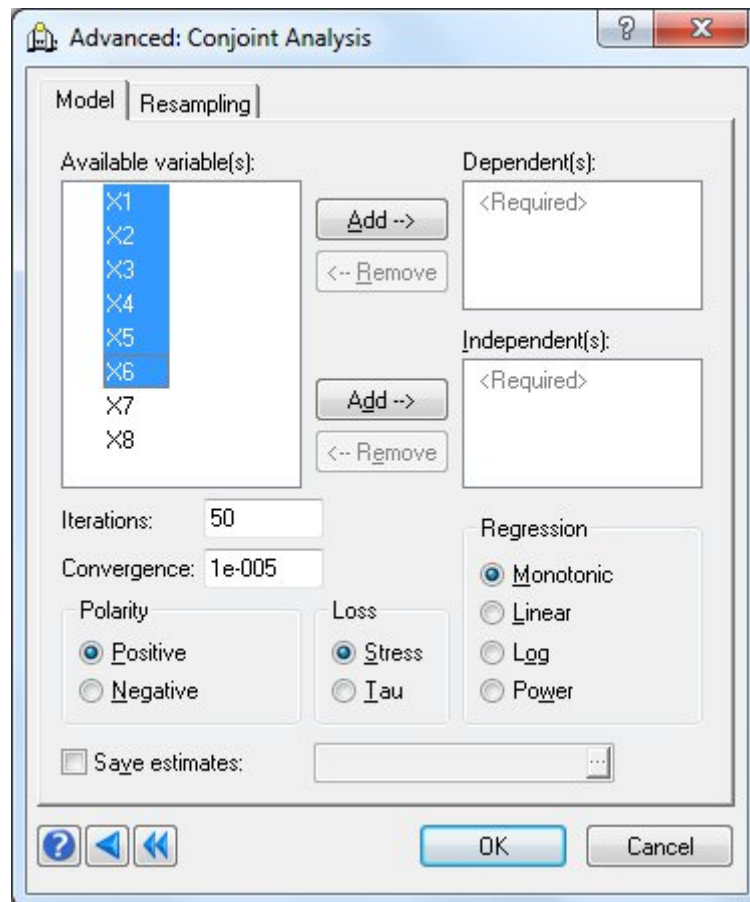
Secara *default* MYSTAT 12 tidak menyediakan menu *Conjoint Analysis*. Untuk dapat menggunakan Analisis Konjoin kita perlu mendapatkan versi lengkap dari software ini yaitu SYSTAT 12. Prosedurnya adalah sebagai berikut

Advance

Conjoint Analysis



maka akan muncul jendela dialog seperti dibawah ini



Kotak dialog *Conjoint Analysis* meliputi pemilihan model dan metode penaksiran ada dalam tab Model. Beberapa opsi yang dapat digunakan adalah

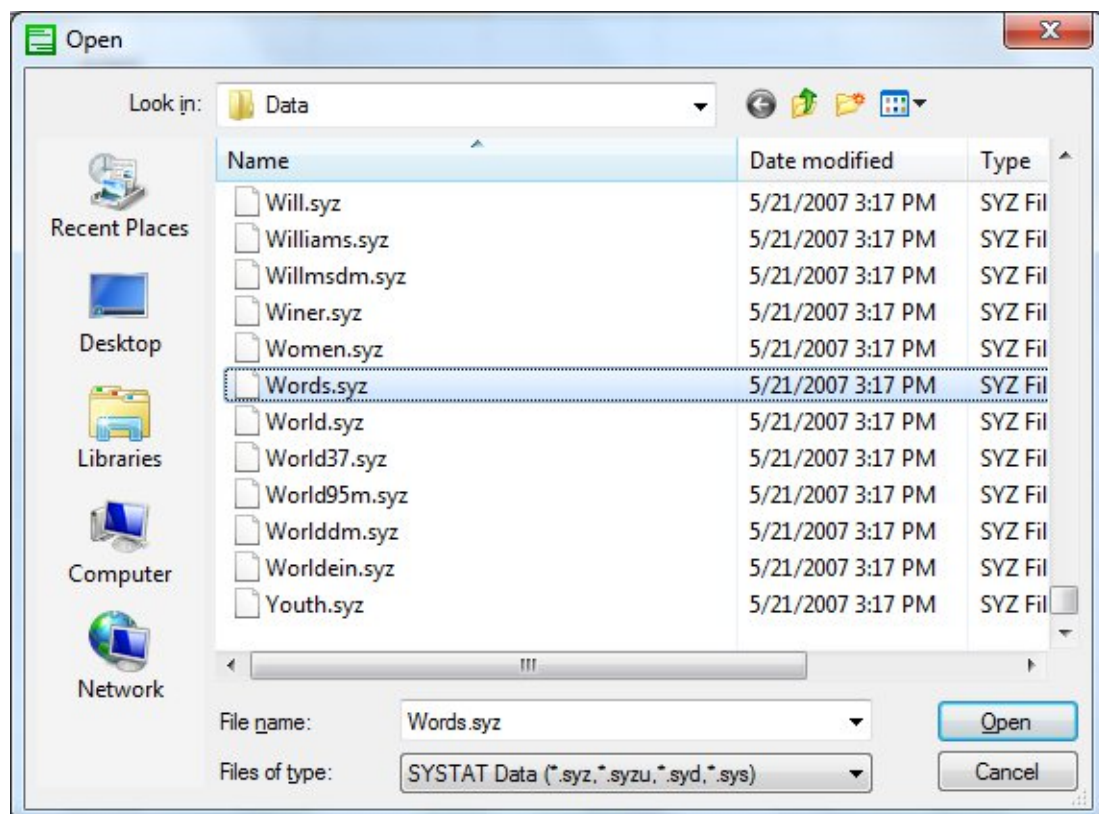
- **Dependent(s).** Untuk memilih variabel tak bebas (*dependent*) yang akan digunakan. Variabel tak bebas harus bernilai numerik (kontinu).
- **Independent(s).** Untuk memilih variabel bebas (*independent*). Variabel bebas bisa bernilai kontinu atau kategori.
- **Iterations.** Untuk menentukan maksimum banyak iterasi. Secara *default* bernilai 50.
- **Convergence.** Untuk menentukan kriteria dalam iterasi penaksiran parameter.
- **Polarity.** Untuk menentukan preferensi yang digunakan dalam rating.
 - **Positive:** apabila *stimuli* menggunakan skala 1 hingga 9 dengan ketentuan 1 adalah "Paling tidak disukai" dan 9 adalah "Paling disukai".



- **Negative** bila menggunakan nilai *ranking* terbalik, artinya untuk *stimuli* yang paling disukai diberi nilai tertinggi setara dengan jumlah *stimulinya*, sedangkan *stimuli* yang paling tidak disukai diberi nilai 1.
- **Loss.** Untuk menentukan fungsi loss dalam penaksiran model
 - *Stress.* Apabila menggunakan Kruskal's STRESS.
 - *Tau.* Apabila menggunakan Kendall's tau- τ .
- **Regression.** Untuk menentukan model yang digunakan:
 - *Monotonic.* Apabila menggunakan fungsi regresi monotonic. Apabila nilai **Loss** adalah **Stress**, maka model ini adalah model Kruskal's MONANOVA.
 - *Linear.* Apabila menggunakan fungsi regresi linier biasa.
 - *Log.* Apabila menggunakan fungsi regresi logaritma.
 - *Power.* Apabila menggunakan fungsi regresi $y = ax^c$. Hal ini akan berguna untuk model Box-Cox.
- **Save estimates.** Untuk menyimpan hasil penaksiran parameter dalam file.

III. Contoh Penggunaan

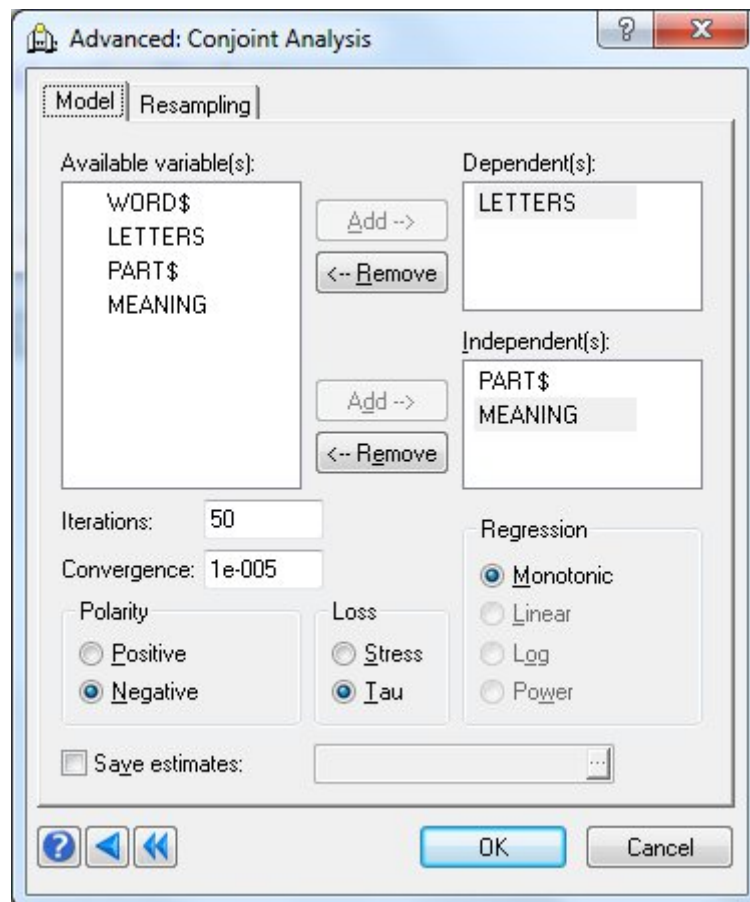
Data yang digunakan dalam contoh ini adalah data WORDS yang terdapat pada directory Data dalam software MYSTAT/SYSTAT. Data tersebut dapat dibuka melalui menu File > Open dan pilih "Words.syz"





Data berisi kata-kata yang frekuensinya paling banyak digunakan dalam bahasa Inggris Amerika (Carroll et al., 1971). Tiga pengukuran ditambahkan dalam data tersebut. Pertama adalah bagian lafal yang paling mirip (*most likely*) pada kolom PART\$, kedua banyak huruf pada kolom LETTERS dan ketiga arti dari kata pada kolom MEANING. Pengukuran ini mewakili jumlah kerugian dengan kriteria berikut (1 = sedikit, 4 = banyak) dengan mengabaikan kata dalam kalimat. Para ahli bahasa berargumen mengenai klasifikasi ini mereka mengungkapkan perbedaan mendasar. Meskipun ukuran yang digunakan dalam bentuk frekuensi, kita akan menggunakan nilai ranking untuk melihat apakah ada informasi yang cukup untuk kecocokan model. Dalam hal ini akan digunakan fungsi Loss Kendall's tau-b.

Untuk menganalisis data ini buka *Advance > Conjoint Analysis* kemudian pilih variabel LETTERS untuk *Dependent(s)* dan pilih variabel PART\$ dan MEANING untuk *Independent(s)*. Pada opsi **Polarity** kita pilih *Negative*, dan opsi **Loss** kita pilih *Tau*. Klik **OK**.



Output yang dihasilkan adalah sebagai berikut



▼ Conjoint Analysis

Iterative Conjoint Analysis

Monotonic Regression Model

Data are ranks

Loss Function is $1-(1+\tau)/2$

Factors and Levels	
P	M
a	1
a	2
c	3
p	
p	
v	

Iteration History

Convergence Criterion : 0.000

Maximum Iterations : 50

Iteration	Loss	Max Parameter Change
1	0.407	0.317
2	0.393	0.174
3	0.389	0.223
4	0.385	0.176
5	0.385	0.000
6	0.385	0.000

Parameter Estimates (Part Worth's)

	adjective	adverb	conjunction	preposition	pronoun	verb	MEANING(1)	MEANING(2)	MEANING(3)
	-0.168	-0.078	0.146	0.706	-0.024	-0.582	-0.036	0.246	-0.212

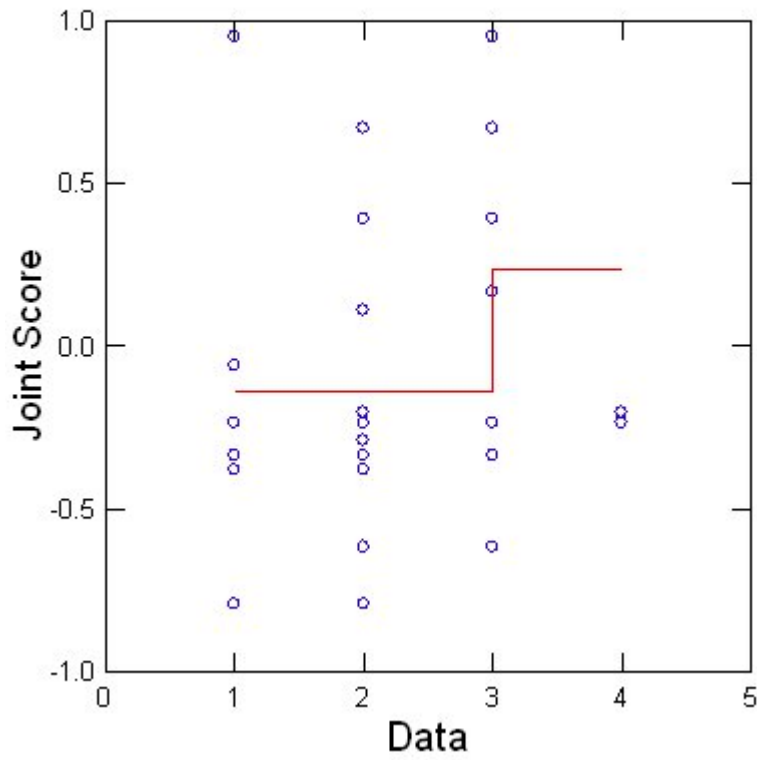
Goodness of Fit (Kendall tau)

LETTERS
0.231

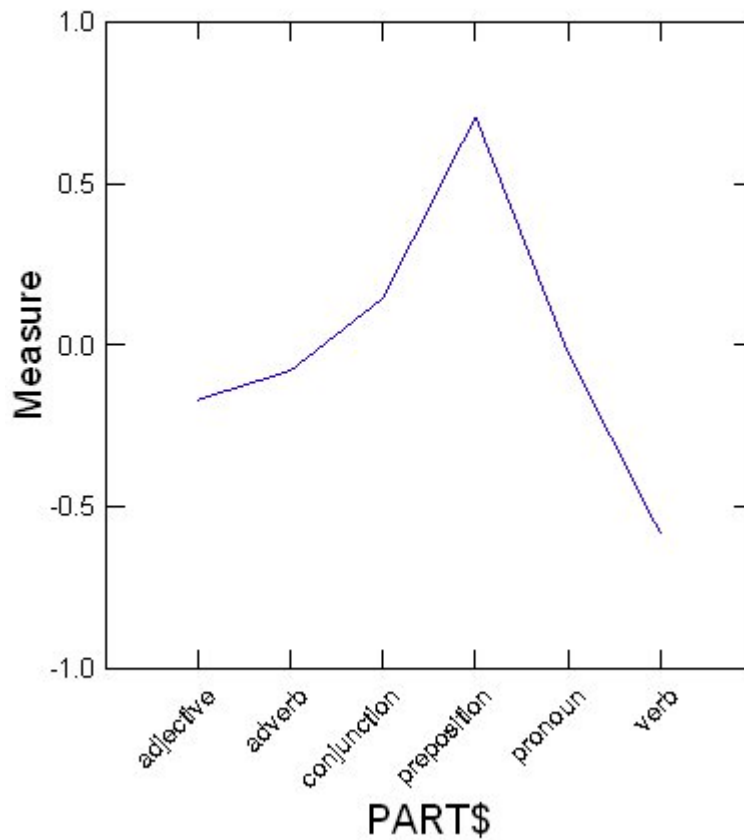
RMS Deleted Goodness of Fit Value, i.e. Fit when Parameter(i)=0									
	adjective	adverb	conjunction	preposition	pronoun	verb	MEANING(1)	MEANING(2)	MEANING(3)
	0.152	0.222	0.231	0.235	0.231	0.187	0.231	0.196	0.174



Shepard Diagram

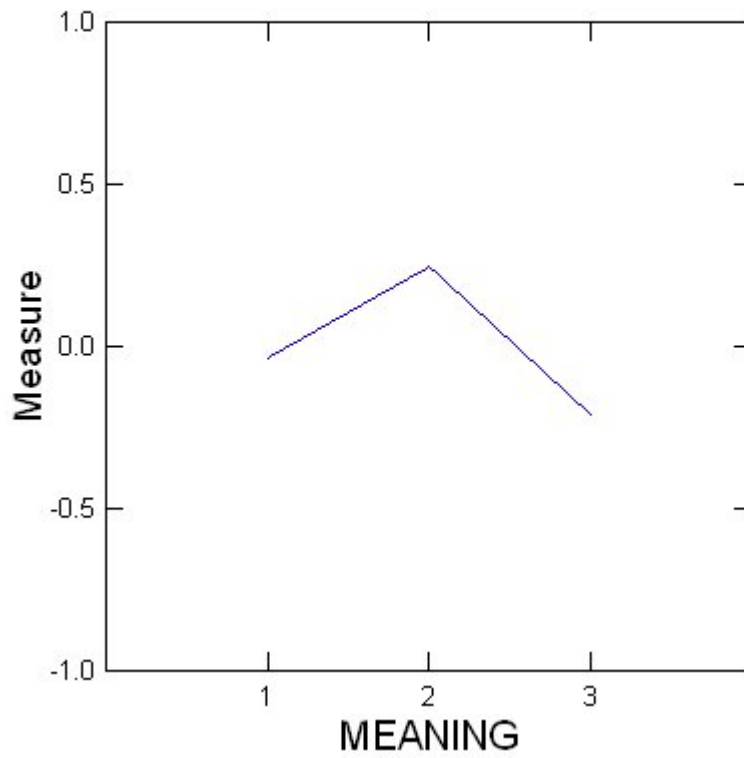


Profile Plot





Profile Plot





PENUTUP

Modul 1 dan Modul 2 Analisis Data Multivariat menggunakan MYSTAT 12 hanya menjelaskan penggunaan beberapa alat dalam analisis data multivariat secara praktis bukan penjelasan teori, metode dan analisis yang mendalam. Untuk dapat memahami teori, metode dan analisis statistika secara mendalam tentu harus dipelajari dari sumber yang tepat.

Modul ini merupakan penjelasan singkat mengenai penggunaan MYSTAT12 untuk beberapa analisis data multivariat, masih banyak analisis statistika yang belum dijelaskan dalam modul ini. Mudah-mudahan penjelasan mengenai penggunaan analisis statistika yang lain dapat dijelaskan dalam modul lanjutan yang direncanakan akan segera dibuat. Beberapa kekurangan yang terdapat dalam modul ini juga akan diperbaiki.



DAFTAR PUSTAKA

1. Johnson, Richard A, Wichern, Dean W. (2002) *Applied Multivariate Statistical Analysis*. Pearson Education International, New Jersey.
2. Manual Book: SYSTAT 11 Statistics. 2004. SYSTAT Software, Inc.