

# Bab 5

## Analisis Diskriminan

Analisis Diskriminan adalah teknik Multivariat yang termasuk pada *Dependence Method*, dengan ciri adanya variabel dependen dan independen. Dengan demikian, ada variabel yang hasilnya tergantung pada data variabel independen. Ciri khusus analisis diskriminan adalah data variabel dependen harus berupa data kategori, sedangkan data untuk variabel independen justru berupa data rasio.

Secara teknis, analisis diskriminan mirip dengan analisis regresi, karena keduanya mempunyai variabel dependen dan variabel independen dalam modelnya. Hanya pada analisis regresi (sederhana maupun berganda), variabel dependen harus data rasio; sedangkan pada analisis diskriminan, jenis data untuk variabel dependen harus kategori.

Karena mempunyai model yang sama, secara dasar kegunaan, analisis diskriminan sama dengan analisis regresi. Dengan demikian, kegunaan utama dari analisis diskriminan ada dua. Pertama adalah kemampuan memprediksi terjadinya variabel dependen dengan masukan data variabel independen; kedua adalah kemampuan memilih mana variabel independen yang secara nyata memengaruhi variabel dependen dan mana yang tidak. Contoh sederhana berikut akan menjelaskan kegunaan analisis diskriminan dalam praktek.

Misalkan PT JAYA ABADI yang selama ini menjual telepon seluler merk JAX ingin mengetahui bagaimana sikap responden terhadap ponsel tersebut: apakah memang ada kelompok responden yang suka dan ada kelompok responden yang tidak suka? Jika ada, faktor apa yang membedakan kedua kelompok responden tersebut: apakah model ponselnya, apakah kualitas ponselnya, apakah garansi produknya, atautkah karena perbedaan tingkat penghasilan responden?

Pada kasus di atas, sikap 'suka' dan 'tidak suka' pada ponsel JAX adalah variabel dependen, karena data itulah yang menjadi 'kesimpulan' dari beberapa faktor (variabel independen). Jika digambarkan secara sederhana, akan menjadi:

SUKA/TIDAK SUKA PONSEL MERK JAX →.... MODEL PONSEL +....  
KUALITAS PONSEL + .....GARANSI PONSEL +..... TINGKAT  
PENGHASILAN RESPONDEN.

Kemudian sejumlah responden disurvei dengan sebuah kuesioner yang meminta opini responden pada sikap mereka terhadap ponsel JAX, serta pendapat mereka atas empat faktor yang telah dikemukakan di atas. Setelah sekian data didapat, dilakukan pengolahan data menggunakan analisis diskriminan. Hasilnya sebuah model diskriminan, yang jika diinterpretasikan dalam contoh-contoh pernyataan adalah:

- Apakah memang ada perbedaan yang jelas di antara sekian sikap responden terhadap ponsel JAX, sehingga mereka yang suka dapat dibedakan dengan jelas dengan mereka yang tidak suka ponsel tersebut?
- Jika memang ada perbedaan di antara kedua kelompok responden tersebut, faktor mana yang perbedaannya memang nyata dan faktor mana yang sesungguhnya tidak secara jelas berbeda?

Misal faktor MODEL yang jelas membedakan (mendiskriminasi) sikap kedua kelompok responden tersebut. Misalkan pula skor rata-rata sikap responden kelompok 1 (SUKA) untuk model ponsel JAX adalah 4,5, sedangkan skor rata-rata sikap responden kelompok 2 (TIDAK SUKA) untuk model ponsel JAX adalah 1,5. Perbedaan keduanya ( $4,5 - 1,5 = 3$ ) signifikan, dan dapat disimpulkan model ponsel JAX-lah yang membuat ada perbedaan sikap di antara responden. Dengan kata lain, variabel independen model ponsel secara nyata membedakan isi variabel dependen (konsumen yang suka atau tidak suka pada ponsel JAX).

Kemudian hasil dari faktor GARANSI ternyata tidak jelas membedakan sikap kedua kelompok responden tersebut. Misal skor rata-rata sikap responden kelompok 1 (SUKA) untuk garansi yang dijanjikan ponsel JAX adalah 3,5 sedangkan skor rata-rata sikap responden kelompok 2 (TIDAK SUKA) adalah 3,4. Perbedaan kedua angka sangat kecil ( $3,5 - 3,4 = 0,1$ ), sehingga dapat disimpulkan garansi ponsel JAX tidak membuat perbedaan sikap di antara responden.

Demikian seterusnya dapat dilihat dan disimpulkan faktor lain, yakni kualitas ponsel dan tingkat penghasilan konsumen.

- Setelah model diskriminan didapat, analisis dilakukan. Langkah selanjutnya yang tidak kalah penting adalah melihat sejauh mana ketepatan model diskriminan tersebut dapat memprediksi 'isi' variabel dependen dengan input data dari variabel independen.

Pada contoh di atas, misalkan ada responden bernama Ali yang mempunyai skor sikap tertentu pada variabel independen model, kualitas,

garansi dan ia menyebutkan tingkat penghasilan per bulannya. Pertanyaan yang muncul adalah secara skor total, Ali termasuk bersikap suka ataukah tidak suka terhadap ponsel merk JAX? Demikian seterusnya untuk responden Budi, Susi dan lainnya. Model diskriminan yang baik tentu dapat memprediksi secara tepat 'posisi' seorang responden. Jika dari skornya ia seharusnya masuk ke kelompok SUKA, maka dalam prakteknya ia juga memang bersikap suka pada ponsel merk JAX; sebaliknya, jika ia seharusnya masuk ke kelompok TIDAK SUKA, maka dalam prakteknya ia seharusnya ada di kelompok tersebut.

Tentu tidak ada model diskriminan yang menjamin ketepatan 100%, dalam artian bisa saja responden yang seharusnya masuk ke kelompok SUKA ternyata 'meleset' ke kelompok TIDAK SUKA, dan sebaliknya. Namun demikian, tingkat ketepatan prediksi model tersebut haruslah tinggi, misalkan di atas 70%, sehingga model tersebut dapat digunakan oleh manajer PT JAYA ABADI untuk merencanakan sejumlah strategi promosi tertentu bagi kelompok konsumen yang suka, agar mereka tetap loyal, dan strategi promosi yang lain kepada kelompok konsumen yang tidak suka, agar mereka dapat beralih menjadi menyukai ponsel JAX. Atau dapat dilakukan rencana-rencana lain berdasar hasil model diskriminan tersebut.

Selain dalam bidang pemasaran, analisis diskriminan juga populer dalam bidang manajemen keuangan, khususnya pengembangan model untuk memprediksi kesehatan keuangan perusahaan dengan variabel independen sejumlah rasio keuangan tertentu, dan variabel dependen adalah SEHAT atau TIDAK SEHAT. Pembuatan model diskriminan yang mempunyai kemampuan prediksi tinggi tentu sangat bermanfaat untuk membantu pengambilan keputusan yang tepat.

Berikut akan dijelaskan hal-hal umum yang perlu diketahui tentang analisis diskriminan dan beberapa contoh kasus penerapannya.

## 5.1 Hal-Hal Pokok Tentang Analisis Diskriminan

### Tujuan Analisis Diskriminan

Karena bentuk multivariat dari Analisis Diskriminan adalah Dependence, maka variabel Dependen adalah variabel yang menjadi dasar analisis diskriminan. Variabel Dependen bisa berupa kode grup 1 atau grup 2 atau lainnya, dengan tujuan diskriminan secara umum adalah:

- Ingin mengetahui apakah ada perbedaan yang jelas antar-grup pada variabel dependen? Atau bisa dikatakan apakah ada perbedaan antara anggota Grup 1 dengan anggota Grup 2?
- Jika ada perbedaan, variabel independen manakah pada fungsi diskriminan yang membuat perbedaan tersebut?
- Membuat Fungsi atau Model Diskriminan, yang pada dasarnya mirip dengan persamaan regresi.
- Melakukan klasifikasi terhadap objek (dalam terminologi SPSS disebut baris), apakah suatu objek (bisa nama orang, nama tumbuhan, benda atau lainnya) termasuk pada grup 1 atau grup 2, atau lainnya.

### **Proses Dasar dari Discriminant Analysis**

Proses dasar dari analisis diskriminan:

- Memisah variabel-variabel menjadi Variabel Dependen dan Variabel Independen.
- Menentukan metode untuk membuat Fungsi Diskriminan. Pada prinsipnya ada dua metode dasar untuk itu, yakni:
  1. SIMULTANEOUS ESTIMATION, di mana semua variabel dimasukkan secara bersama-sama kemudian dilakukan proses Diskriminan.
  2. STEP-WISE ESTIMATION, di mana variabel dimasukkan satu per satu ke dalam model diskriminan. Pada proses ini, tentu ada variabel yang tetap ada pada model, dan ada kemungkinan satu atau lebih variabel independen yang 'dibuang' dari model.
- Menguji signifikansi dari Fungsi Diskriminan yang telah terbentuk, menggunakan Wilk's Lambda, Pilai, F test dan lainnya.
- Menguji ketepatan klasifikasi dari fungsi diskriminan, termasuk mengetahui ketepatan klasifikasi secara individual dengan Casewise Diagnostics.
- Melakukan interpretasi terhadap Fungsi Diskriminan tersebut.
- Melakukan uji validasi Fungsi Diskriminan.

### **Jumlah Sampel pada Analisis Diskriminan**

Secara pasti tidak ada jumlah sampel yang ideal pada Analisis Diskriminan. Pedoman yang bersifat umum menyatakan untuk setiap variabel independen

sebaiknya ada 5-20 data (sampel). Dengan demikian, jika ada enam variabel independen, seharusnya minimal ada  $6 \times 5 = 30$  sampel. Secara terminologi SPSS, jika ada enam *kolom* variabel independen, sebaiknya ada 30 *baris* data.

Selain itu, pada analisis diskriminan sebaiknya digunakan dua jenis sampel, yakni **analysis sample** yang digunakan untuk membuat Fungsi Diskriminan, serta **holdout sample (spilt sample)** yang digunakan untuk menguji hasil diskriminan. Sebagai contoh, jika ada 70 sampel, maka sampel tersebut bisa dibagi dua, 35 untuk analysis sample dan 35 untuk holdout sample. Kemudian hasil fungsi diskriminan yang terjadi pada analysis sample dibandingkan dengan hasil fungsi diskriminan dari holdout sample, apakah terjadi perbedaan yang besar ataukah tidak. Jika ketepatan klasifikasi kedua sampel hampir sama besar, dikatakan fungsi diskriminan dari analysis sample sudah valid. Inilah yang disebut proses validasi silang (Cross Validation) dari fungsi diskriminan.

#### Asumsi pada Analisis Diskriminan

Asumsi penting yang harus dipenuhi agar model diskriminan bisa digunakan adalah:

- Multivariate Normality, atau variabel independen seharusnya berdistribusi normal. Jika data tidak berdistribusi normal, hal ini akan menyebabkan masalah pada ketepatan fungsi (model) diskriminan. Regresi Logistik (Logistic Regression) bisa dijadikan alternatif metode jika memang data tidak berdistribusi normal.
- Matriks kovarians dari semua variabel independen seharusnya sama (equal).
- Tidak ada korelasi antar-variabel independen. Jika dua variabel independen mempunyai korelasi yang kuat, dikatakan terjadi multikolinieritas.
- Tidak adanya data yang sangat ekstrem (outlier) pada variabel independen. Jika ada data outlier yang tetap diproses, hal ini bisa berakibat berkurangnya ketepatan klasifikasi dari fungsi diskriminan.

#### Model dari Analisis Diskriminan

Analisis diskriminan termasuk dalam *Multivariate Dependence Method*, dengan model:

$$\boxed{Y_1} = \boxed{X_1 + X_2 + \dots + X_n}$$

Non-Metrik

Metrik

**Keterangan:**

- Variabel Independen ( $X_1$  dan seterusnya) adalah data metrik, yakni data berjenis interval atau rasio, seperti Usia seseorang, tinggi sebuah pohon, kandungan zat besi dalam tubuh, dan sebagainya.
- Variabel Dependen ( $Y_1$ ) adalah Data Kategorikal atau Nominal, seperti Golongan Miskin (kode 1), Golongan Menengah (kode 2), Golongan Kaya (kode 3) dan sebagainya. Jika data kategorikal tersebut hanya terdiri atas dua kode saja (misal kode 1 untuk Daerah Banjir dan kode 2 Daerah Non-Banjir), maka model bisa disebut *Two-Group Discriminant Analysis*. Sedang jika kode lebih dari dua kategori, disebut dengan *Multiple Discriminant Analysis*.
- Dari keterangan di atas, perhatikan adanya perbedaan dalam penempatan data yang sekilas mirip. Seperti Usia seseorang (dalam tahun). Jika usia disebut secara langsung sekian tahun (17 tahun, 32 tahun dan sebagainya), maka data tersebut adalah rasio dan otomatis diperlakukan sebagai variabel independen. Namun, jika Usia seseorang dilakukan *penggolongan*, dan dimasukkan dalam kategori-kategori tertentu, seperti jika Usia seseorang antara 15-20 tahun, ia digolongkan *Remaja*, di atas 20 tahun digolongkan *Dewasa*, maka data orang yang berusia 17 tahun tidak akan ditulis langsung '17', namun akan ditulis *Remaja*. Data hasil kategorisasi ini adalah data nominal dan termasuk variabel Dependen.

Dengan demikian, usia 17 tahun bisa menjadi variabel dependen atau independen tergantung bagaimana data tersebut akan diperlakukan, langsung diinput apa adanya atau dilakukan penggolongan.

Berikut dibahas dua kasus, yakni analisis diskriminan untuk dua faktor dan tiga faktor.

## 5.2 Analisis Diskriminan Dua Faktor

### 5.2.1 Uji Variabel

Dengan Analisis Diskriminan, pada akhirnya akan dibuat sebuah model seperti regresi, yakni adanya *SEBUAH* variabel dependen, dan SATU ATAU LEBIH variabel independen. Perbedaan analisis diskriminan dengan model regresi adalah pada *jenis data* yang digunakan; pada analisis diskriminan, variabel dependen selalu data kategori.

Prinsip Diskriminan adalah ingin membuat model yang bisa secara jelas menunjukkan perbedaan (diskriminasi) antar isi variabel dependen, yang

dalam kasus berikut adalah perilaku Sedikit minum atau Banyak minum air mineral. Untuk itu, langkah pertama pada Analisis Diskriminan adalah menguji apakah semua variabel *independent* (bebas) berbeda secara nyata berdasar variabel dependen. Terkait dengan kasus di bawah, sebagai contoh akan diuji apakah variabel USIA konsumen berbeda pada mereka yang banyak minum dengan mereka yang sedikit minum.

## KASUS

Sebuah Perusahaan yang bergerak dalam penjualan Air Mineral mengumpulkan data sekelompok konsumen Air Mineral dengan variabel berikut.

- Tipe Konsumen dari banyaknya Air Mineral yang diminum, dengan kode:
  - Kode 0 = SEDIKIT (konsumen yang termasuk tipe sedikit minum air mineral)
  - Kode 1 = BANYAK (konsumen yang termasuk tipe banyak minum air mineral)
- Usia Konsumen (tahun)
- Berat Badan Konsumen (kilogram)
- Tinggi Badan Konsumen (centimeter)
- Pendapatan Konsumen (ribuan rupiah/bulan)
- Jam Kerja Konsumen dalam sehari (jam)
- Kegiatan Olahraga Konsumen dalam sehari (jam)

Variabel NAMA tidak disertakan dalam proses analisis diskriminan, karena berupa data STRING (berisi karakter dan bukannya angka).

Berikut sebagian tampilan data dengan SPSS.

	nama	minum	usia	berat	tinggi	income	jamkerja	olahraga
1	RUSDI	0.00	40.00	65.00	154.00	680.00	5.33	3.0
2	NINA	0.00	30.00	70.00	157.00	700.00	5.30	3.6
3	LANNY	0.00	25.00	60.00	158.00	580.00	5.27	3.5
4	CITRA	0.00	26.00	75.00	160.00	600.00	5.33	3.0
5	DINA	0.00	40.00	50.00	159.00	700.00	5.50	3.5
6	SISKA	1.00	28.00	62.00	158.00	440.00	5.00	2.2
7	LUSI	0.00	29.00	50.00	160.00	580.00	5.07	2.9
8	LENNY	0.00	40.00	52.00	165.00	800.00	5.13	4.0

**Gambar 5.1. Tampilan Data Kasus**

### Contoh Interpretasi Data

Pada baris pertama, konsumen dengan nama Rusdi ternyata termasuk orang yang sedikit minum air mineral. Ia berusia 40 tahun, berat badan 65 kilogram dengan tinggi 154 centimeter, penghasilan per bulan Rp.680.000,-, dalam sehari ia bekerja rata-rata 5,33 jam serta melakukan aktivitas berolahraga rata-rata 3 jam dalam sehari.

Demikian seterusnya untuk konsumen-konsumen yang lain.

Dari file DISKRIMINAN yang berisi Profil Pembeli Air Mineral Dalam Kemasan (AMDK) dari segala macam merk tersebut, akan dilakukan analisis Diskriminan untuk mengetahui:

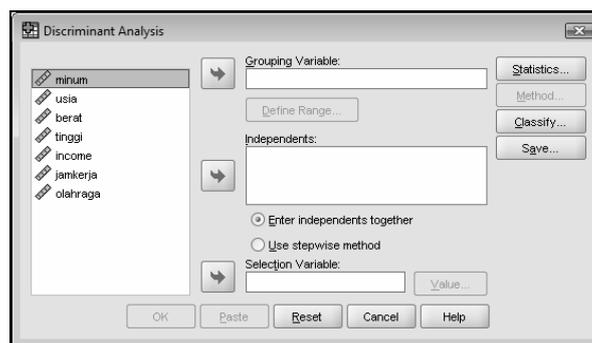
- Apakah ada perbedaan yang signifikan antara Mereka yang banyak minum AMDK dengan mereka yang sedikit meminumnya?
- Jika ada perbedaan yang signifikan, variabel apa saja yang membuat perilaku konsumsi air mineral mereka berbeda?
- Membuat model diskriminan dua faktor (karena hanya ada mereka yang SEDIKIT dengan yang BANYAK) untuk kasus tersebut.
- Menguji ketepatan model (fungsi) diskriminan.

### Langkah

Tahapan pengujian pada analisis Diskriminan.

#### 1. MENILAI VARIABEL YANG LAYAK UNTUK ANALISIS

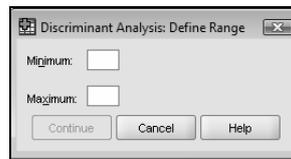
- Buka file **diskriminan**.
- Menu **Analyze** → **Classify** → **Discriminant....** Tampak di layar:



*Gambar 5.2. Kotak Dialog Discriminant Analysis*

Pengisian:

- Masukkan variabel **minum** ke bagian GROUPING VARIABLE. Hal ini berarti variabel MINUM berfungsi sebagai dependent variable (variabel tergantung), yang berciri data kategori. Karena itu, SPSS meminta masukan kode kategori yang dipakai.
- Untuk itu, buka ikon DEFINE RANGE..., hingga tampak di layar:



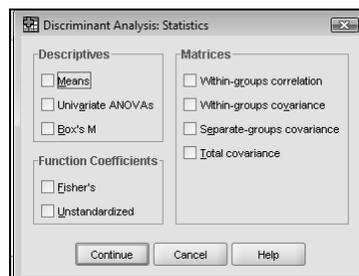
Sesuai kode untuk variabel MINUM, maka:

- o Masukkan angka **0** pada Minimum
- o Masukkan angka **1** pada Maximum

Tekan tombol CONTINUE untuk kembali ke kotak dialog utama.

Kode variabel MINUM dapat dilihat dengan menekan CTRL-T; lihat bagian (kolom) VALUES untuk variabel tersebut.

- Masukkan variabel **usia, berat, tinggi, income, jamkerja** dan **olahraga** ke dalam kotak INDEPENDENT. Hal ini berarti keenam variabel tersebut berfungsi sebagai variabel bebas (independen).
- Klik ikon STATISTICS. Tampak di layar:



*Gambar 5.3. Kotak Dialog Statistics*

Pengisian:

Pada bagian DESCRIPTIVES, aktifkan pilihan **Univariate ANOVAs** dan **Box's M**.

Abaikan bagian lain, dan tekan CONTINUE untuk kembali ke kotak dialog utama.

Abaikan bagian lain dan tekan OK untuk proses uji variabel, dengan Output:

(disimpan pada file **DISKRIMINAN UJI VARIABEL.spv**)

PERHATIKAN TIDAK SEMUA BAGIAN OUTPUT DITAMPILKAN. HANYA DUA BAGIAN OUTPUT YANG RELEVAN YANG AKAN DIBAHAS.

**Tests of Equality of Group Means**

	Wilks' Lambda	F	df1	df2	Sig.
usia	.945	4.247	1	73	.043
berat	.934	5.173	1	73	.026
tinggi	.946	4.186	1	73	.044
income	.894	8.656	1	73	.004
jamkerja	.999	.076	1	73	.783
olahraga	.946	4.183	1	73	.044

### Analisis

Tabel di atas adalah hasil pengujian untuk setiap variabel bebas yang ada.

Keputusan bisa diambil lewat dua cara.

- Dengan angka Wilk's Lambda

Angka Wilk's Lambda berkisar 0 sampai 1. Jika angka mendekati 0 maka data tiap grup cenderung berbeda, sedang jika angka mendekati 1, data tiap grup cenderung sama.

Dari tabel terlihat angka Wilk's Lambda berkisar antara 0,894 sampai 0,999 (mendekati 1). Dari kolom Sig bisa dilihat bahwa hanya variabel JAM KERJA yang cenderung tidak berbeda. Hal ini berarti Jam kerja untuk mereka yang sedikit atau banyak mengonsumsi air mineral ternyata tidak berbeda secara nyata. Hanya di sini kriteria 'mendekati angka 1' adalah sulit ditentukan secara pasti, karena hampir semua variabel di atas mempunyai angka Wilk's Lambda yang besar, namun hanya satu variabel yang tidak lolos. Untuk itu, pengujian dengan ANOVA, yang dibahas di bawah ini, lebih mudah dilakukan.

- Dengan F test

Lihat angka Sig.

Jika Sig. > 0,05 berarti tidak ada perbedaan antar-grup.

Jika Sig. < 0,05 berarti ada perbedaan antar-grup.

Contoh analisis menggunakan uji F:

- Variabel USIA, angka Sig. adalah di bawah 0,05 (0,043). Hal ini berarti ada perbedaan antar-grup, atau responden yang banyak membeli atau sedikit mengonsumsi air mineral terkait dengan usia responden tersebut. Mungkin mereka yang lebih tua akan lebih banyak mengonsumsi dibanding mereka yang masih muda usia, atau mungkin sebaliknya. Hal ini akan dibahas saat sudah terbentuk model diskriminan.
- Variabel BERAT, angka Sig. adalah di bawah 0,05 (0,026). Hal ini berarti berat badan responden memengaruhi banyak sedikitnya mereka mengonsumsi air mineral.
- Variabel JAM KERJA, angka Sig. adalah jauh di atas 0,05 (0,783). Hal ini berarti jumlah Jam Kerja seorang responden tidak memengaruhi banyak sedikitnya konsumsi air mineral mereka. Kesimpulan ini sama dengan jika berpatokan pada angka Wilk's Lambda yang hampir mendekati 1 untuk variabel JAM KERJA.

Dari enam variabel, ada lima variabel yang berbeda secara signifikan untuk dua grup diskriminan, yaitu USIA, BERAT, TINGGI, INCOME dan OLAHRAGA. Dengan demikian, sedikit atau banyaknya konsumsi seseorang akan air mineral dipengaruhi oleh usia responden tersebut, berat dan tinggi badan responden, tingkat penghasilan seorang responden, dan kegiatan olahraga responden yang bersangkutan.

#### Catatan

Pada beberapa analisis diskriminan, sebuah variabel yang tidak lolos uji tidak otomatis dikeluarkan. Seperti pada kasus di atas, variabel JAM KERJA walaupun tidak lolos uji, namun seharusnya tetap disertakan pada analisis diskriminan selanjutnya. Pandangan ini berdasar pada prinsip bahwa pada analisis multivariat, variabel-variabel dianggap suatu kesatuan, dan bukannya terpisah-pisah.

Test Results	
Box's M	28.145
F	Approx. 1.222
	df1 21
	df2 19569.371
	Sig. .220

Tests null hypothesis of equal population covariance matrices.

### Analisis

Jika analisis ANOVA dan angka Wilk's Lambda menguji means (rata-rata) dari setiap variabel, maka Box's M menguji varians dari setiap variabel. Analisis Diskriminan mempunyai asumsi bahwa:

- Varians variabel bebas untuk tiap grup seharusnya sama. Jika demikian, seharusnya varians dari Responden yang Sedikit mengonsumsi air mineral sama dengan varians dari Responden yang Banyak mengonsumsi air mineral.
- Varians *di antara* variabel-variabel bebas seharusnya juga sama. Jika demikian, seharusnya varians dari USIA sama dengan varians dari BERAT, sama dengan variabel OLAHRAGA dan sebagainya.

Kedua pengertian di atas bisa disimpulkan, seharusnya *group covariance matrices* adalah relatif sama, yang diuji dengan alat Box's M dengan ketentuan:

- HIPOTESIS

Ho: *group covariance matrices* adalah relatif sama

Hi: *group covariance matrices* adalah berbeda secara nyata

- Keputusan dengan dasar signifikansi (lihat angka Sig.)

Jika Sig. > 0,05 berarti Ho diterima.

Jika Sig. < 0,05 berarti Ho ditolak.

Dari tabel terlihat bahwa angka Sig. jauh di atas 0,05 (0,220) yang berarti *group covariance matrices* adalah sama. Hal ini berarti data di atas sudah memenuhi asumsi analisis diskriminan, sehingga proses bisa dilanjutkan.

### Catatan

- Sama tidaknya *group covariance matrices* juga bisa dilihat dari tabel output LOG DETERMINANT berikut (ada di atas tabel Box's M).

Log Determinants		
minum	Rank	Log Determinant
sedikit	6	14.085
banyak	6	14.258
Pooled within-groups	6	14.556

The ranks and natural logarithms of determinants printed are those of the group covariance matrices.

Terlihat angka Log Determinant untuk kategori SEDIKIT (14,085) dan BANYAK (14,258) tidak berbeda banyak, sehingga *group covariance matrices* akan relatif sama untuk kedua group.

- Bagaimana jika ternyata *group covariance matrices* adalah berbeda secara nyata?

Jika hal ini terjadi pada variabel dependen yang hanya ada dua kemungkinan, seperti pada kasus di atas (hanya ada kategori SEDIKIT dan BANYAK), maka proses lanjutan *seharusnya* tidak bisa dilakukan.

Namun, jika hal itu terjadi pada kasus dengan variabel dependen dengan kategori banyak (misal ada 5 kategori: SANGAT SEDIKIT, SEDIKIT, CUKUP BANYAK, BANYAK, SANGAT BANYAK), maka dengan melihat angka LOG DETERMINANT (lihat penjelasan terdahulu), di mana angka yang paling berbeda (misal tanda - sedang yang lain +) bisa dibuang dan proses uji diulang lagi. Jadi, -sebagai contoh- variabel CUKUP BANYAK- mungkin dikeluarkan dan proses diulang lagi. Dengan demikian, bisa saja nanti tinggal tiga kategori, namun asumsi sudah terpenuhi.

## 5.2.2 Melakukan Analisis Diskriminan

Setelah dilakukan pengujian atas variabel yang akan diproses (lihat kasus sebelumnya), langkah selanjutnya adalah melakukan proses pembuatan model diskriminan. Pada kasus ini akan dibahas model diskriminan yang paling sederhana, yakni hanya melibatkan dua kategori pada variabel dependen (Two-Group).

### KASUS (lanjutan)

Dari uji variabel file DISKRIMINAN yang berisi Profil Pembeli Air Mineral Dalam Kemasan (AMDK) pada kasus sebelumnya, diketahui ada lima variabel yang lolos uji, yakni USIA, BERAT, TINGGI, INCOME dan OLAHRAGA. Langkah selanjutnya adalah membuat model diskriminan dua faktor untuk kasus tersebut.

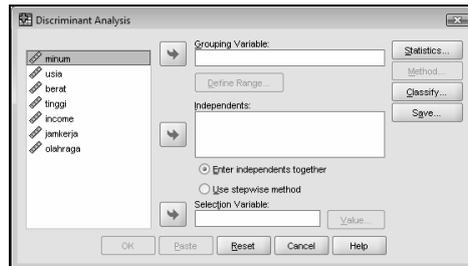
### Langkah

Tahapan lanjutan pada Analisis Diskriminan.

## 2. PROSES DISKRIMINAN

Karena hanya ada dua kategori pada variabel dependent (MINUM), maka yang akan dibuat adalah model satu faktor.

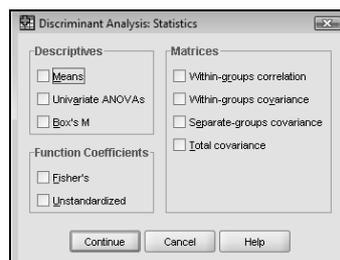
- Buka file **diskriminan**.
- Menu **Analyze** → **Classify** → **Discriminant....** Tampak di layar:



**Gambar 5.4. Kotak Dialog Discriminant Analysis**

Pengisian (mirip dengan pengisian pada kasus sebelumnya):

- Masukkan variabel **minum** ke bagian GROUPING VARIABLE.
- Kemudian buka ikon DEFINE RANGE. Pada kotak dialog, sesuai kode untuk variabel MINUM:
  - Masukkan angka **0** pada Minimum.
  - Masukkan angka **1** pada Maximum.
 Tekan tombol CONTINUE untuk kembali ke kotak dialog utama.
- Masukkan variabel **usia**, **berat**, **tinggi**, **income** dan **olahraga** ke dalam kotak INDEPENDENT.
- Klik ikon STATISTICS. Tampak di layar:



**Gambar 5.5. Kotak Dialog Statistics**

Pengisian:

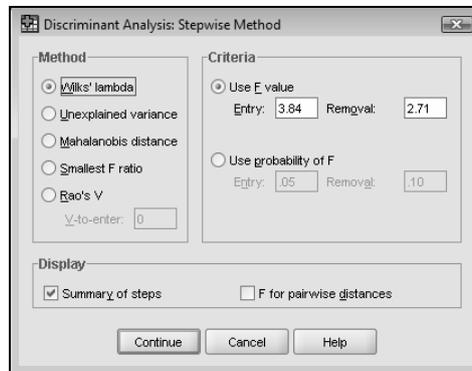
- Pada bagian DESCRIPTIVES, aktifkan pilihan **Means**.

- Pada bagian FUNCTION COEFFICIENTS, aktifkan pilihan **Fisher's** dan **Unstandardized**.

Abaikan bagian lain, dan tekan CONTINUE untuk kembali ke kotak dialog utama.

- Perhatikan bagian tengah kotak dialog utama DISCRIMINANT. Klik mouse pada pilihan **Use stepwise method** (yang terletak di tengah bawah), maka *secara otomatis* ikon METHOD yang ada di bagian kanan kotak dialog utama akan terbuka (aktif).

Klik ikon METHOD tersebut. Tampak di layar:



**Gambar 5.6. Kotak Dialog Stepwise Method**

Memilih STEP WISE Method berarti variabel akan dimasukkan satu per satu ke dalam model, dan bukannya sekaligus seperti jika pilihan adalah ENTER INDEPENDENT TOGETHER.

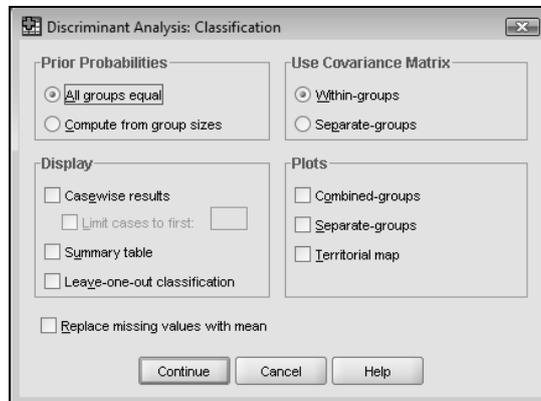
Pengisian:

- Pada bagian METHOD, pilih **Mahalanobis distance**.  
NB: metode untuk diskriminan menggunakan metode Mahalanobis.
- Pada bagian CRITERIA, pilih **Use Probability of F**, namun jangan mengubah isi yang sudah ada (default).

Di sini lolos tidaknya sebuah variabel akan diuji dengan uji F, dengan batasan signifikansi 5% (0,05).

Abaikan bagian yang lain, kemudian tekan CONTINUE untuk kembali ke kotak dialog utama.

- Klik ikon CLASSIFY. Tampak di layar:



**Gambar 5.7. Kotak Dialog Classification**

Kotak dialog CLASSIFY adalah pelengkap dari pembuatan model diskriminan, terutama cara penyajian model diskriminan, serta kelayakan model tersebut.

Pengisian:

- Pada bagian DISPLAY, aktifkan pilihan **Casewise results**.

Pilihan ini untuk membandingkan hasil kasus awal dengan model diskriminan.

- Masih pada bagian DISPLAY, aktifkan juga pilihan **Leave-one-out-classification**.

Pilihan ini untuk menampilkan data yang cocok dengan hasil proses diskriminan, dan mana yang tidak cocok (misclassified).

Abaikan bagian lain, dan tekan CONTINUE untuk kembali ke kotak dialog utama.

Abaikan ikon yang lain, dan tekan OK untuk proses data.

### 3. ANALISIS OUTPUT DISKRIMINAN

Output (simpan dengan nama **diskriminan 2 faktor**).

OUTPUT AKAN DITAMPILKAN PER BAGIAN UNTUK ANALISIS, DENGAN CATATAN TIDAK SEMUA OUTPUT DITAMPILKAN.

**Group Statistics**

minum		Mean	Std. Deviation	Valid N (listwise)	
				Unweighted	Weighted
sedikit	usia	31.8684	6.12558	38	38.000
	berat	57.2105	11.76609	38	38.000
	tinggi	158.4474	5.66496	38	38.000
	income	659.4737	106.79476	38	38.000
	olahraga	3.1974	.55385	38	38.000
banyak	usia	28.9459	6.15515	37	37.000
	berat	51.8108	8.48475	37	37.000
	tinggi	161.5676	7.44439	37	37.000
	income	578.9189	129.52442	37	37.000
	olahraga	2.9054	.67780	37	37.000
Total	usia	30.4267	6.27344	75	75.000
	berat	54.5467	10.56549	75	75.000
	tinggi	159.9867	6.74335	75	75.000
	income	619.7333	124.53127	75	75.000
	olahraga	3.0533	.63124	75	75.000

Tabel GROUP STATISTICS pada dasarnya berisi data statistik (deskriptif) yang utama, yakni Rata-rata dan standar deviasi, dari kedua grup Konsumen.

Sebagai contoh, Konsumen yang termasuk tipe **SEDIKIT** meminum air mineral, mempunyai Berat badan rata-rata 57,21 kilogram. Sedangkan mereka yang termasuk tipe **BANYAK** meminum air mineral, mempunyai Berat badan rata-rata lebih rendah, yakni 51,81 kilogram. Kedua angka ini tentu berbeda, namun apakah perbedaan tersebut *nyata* (signifikan)?

Demikian pula untuk variabel lainnya, semua mempunyai angka rata-rata dan standar deviasi yang berbeda untuk kedua Grup Konsumen. Dan semuanya akan diuji untuk mengetahui variabel mana yang mempunyai perbedaan yang signifikan.

Dari tabel di atas juga terlihat ada 38 responden yang tergolong sedikit mengonsumsi minum Air Mineral, sedangkan 37 responden lainnya tergolong banyak mengonsumsi Air Mineral. Jika melihat semua variabel (**USIA**, **BERAT** dan seterusnya) terisi angka 38 atau 37 semuanya, maka pada kasus ini tidak ada data yang hilang (missing), sehingga total data untuk semua variabel adalah 75 buah.

Tentu ini adalah keadaan yang ideal. Namun, jika pada sebuah kasus ada satu atau beberapa data yang hilang atau tidak tercatat, bisa dilakukan proses rata-rata untuk pengisian data yang kosong tersebut, atau menghilangkan data missing dari proses pembuatan model.

Variables Entered/Removed <sup>a,b,c,d</sup>							
Step	Entered	Min. D Squared					
		Statistic	Between Groups	Exact F			
				Statistic	df1	df2	Sig.
1	income	.462	sedikit and banyak	8.656	1	73.000	.004
2	berat	.776	sedikit and banyak	7.170	2	72.000	.001
3	tinggi	1.208	sedikit and banyak	7.343	3	71.000	.000

At each step, the variable that maximizes the Mahalanobis distance between the two closest groups is entered.

a. Maximum number of steps is 10.  
b. Maximum significance of F to enter is .05.  
c. Minimum significance of F to remove is .10.  
d. F level, tolerance, or VIF insufficient for further computation.

Tabel ini menyajikan variabel mana saja -dari lima variabel input- yang bisa dimasukkan (entered) dalam persamaan diskriminan.

Karena proses adalah stepwise (bertahap), maka akan dimulai dengan variabel yang mempunyai angka F hitung (statistic) terbesar.

#### Tahap pemasukan variabel bebas:

- Pada tahap pertama, angka F hitung variabel INCOME adalah yang terbesar, mencapai 8,656 maka pada tahap pertama ini, variabel INCOME terpilih.
- Pada tahap kedua, dengan variabel yang sudah berkurang satu, angka F hitung variabel BERAT adalah kedua terbesar, mencapai 7,170 maka pada tahap kedua ini, variabel BERAT terpilih.
- Pada tahap ketiga, dengan variabel yang sudah berkurang dua, angka F hitung variabel TINGGI adalah ketiga terbesar, mencapai 7,343 maka pada tahap ketiga ini, variabel TINGGI terpilih.

Perhatikan ketiga variabel tersebut tentunya mempunyai angka Sig. di bawah 0,05 seperti angka Sig. variabel INCOME yang adalah 4,367E-03 atau 0,004367, jauh di bawah 0,05.

Dengan demikian, dari lima variabel yang dimasukkan, hanya ada tiga variabel yang signifikan. Atau bisa dikatakan Income (pendapatan) Konsumen, Berat Badan serta Tinggi Badan konsumen memengaruhi perilaku konsumen tersebut untuk Sedikit atau Banyak minum air mineral. Bisa jadi mereka yang berusia lebih muda, atau mereka yang mempunyai Berat dan Tinggi badan di atas rata-rata akan mengonsumsi banyak air mineral. Atau berbagai kemungkinan lain yang akan diuraikan pada analisis selanjutnya.

**Variables in the Analysis**

Step		Tolerance	Sig. of F to Remove	Min. D Squared	Between Groups
1	income	1.000	.004		
2	income	.997	.004	.276	sedikit and banyak
	berat	.997	.026	.462	sedikit and banyak
3	income	.993	.004	.647	sedikit and banyak
	berat	.944	.008	.721	sedikit and banyak
	tinggi	.943	.012	.776	sedikit and banyak

Tabel di atas -dan tabel selanjutnya- sebenarnya hanyalah perincian (detail) dari proses stepwise pada tabel sebelumnya. Pada Step 1, variabel INCOME adalah variabel pertama yang masuk ke dalam Model Diskriminan. Hal ini disebabkan variabel tersebut mempunyai angka SIG. OF F TO REMOVE yang paling sedikit, yakni 0,004 (jauh di bawah 0,05).

Kemudian pada step 2, dimasukkan variabel kedua, yakni BERAT. Variabel tersebut memenuhi syarat, karena mempunyai angka SIG. OF F TO REMOVE di bawah 0,05 yakni 0,026.

Pada step 3 atau terakhir, dimasukkan variabel ketiga, yakni TINGGI. Variabel tersebut juga memenuhi syarat, dengan angka SIG. OF F TO REMOVE di bawah 0,05 yakni 0,012. Perhatikan perubahan angka pada variabel BERAT seiring dengan masuknya variabel ketiga, yakni variabel TINGGI.

**Variables Not in the Analysis**

Step		Tolerance	Min. Tolerance	Sig. of F to Enter	Min. D Squared	Between Groups
0	usia	1.000	1.000	.043	.227	sedikit and banyak
	berat	1.000	1.000	.026	.276	sedikit and banyak
	tinggi	1.000	1.000	.044	.223	sedikit and banyak
	income	1.000	1.000	.004	.462	sedikit and banyak
	olahraga	1.000	1.000	.044	.223	sedikit and banyak
1	usia	.318	.318	.542	.484	sedikit and banyak
	berat	.997	.997	.026	.776	sedikit and banyak
	tinggi	.997	.997	.042	.721	sedikit and banyak
	olahraga	.538	.538	.953	.462	sedikit and banyak
2	usia	.317	.316	.458	.812	sedikit and banyak
	tinggi	.943	.943	.012	1.208	sedikit and banyak
	olahraga	.537	.537	.876	.777	sedikit and banyak
3	usia	.316	.316	.524	1.238	sedikit and banyak
	olahraga	.523	.523	.798	1.213	sedikit and banyak

Tabel di atas adalah kebalikan dari tabel sebelumnya, di mana pada tabel ini justru yang ditayangkan adalah proses pengeluaran variabel secara bertahap.

- Pada step 0 (keadaan awal), kelima variabel secara lengkap ditayangkan dengan angka SIG. OF F TO REMOVE sebagai faktor penguji. Terlihat angka SIG. OF F TO REMOVE yang terkecil adalah pada variabel INCOME (angka 0,004). Maka variabel INCOME dikeluarkan dari step 0 tersebut, yang berarti variabel tersebut bukan termasuk variabel yang tidak dianalisis.
- Pada step 1, sekarang terlihat ada empat variabel, dan proses pengujian terus berjalan, dengan pedoman angka SIG. OF F TO REMOVE harus di bawah 0,05 dan jika mungkin diambil angka terkecil. Terlihat variabel BERAT sekarang mempunyai angka SIG. OF F TO REMOVE terkecil (0,026) sehingga variabel tersebut dikeluarkan.
- Pada step 2, sekarang terlihat ada tiga variabel, dan terlihat variabel TINGGI pada step ini mempunyai angka SIG. OF F TO REMOVE terkecil (0,012) sehingga variabel tersebut dikeluarkan.
- Pada step 3, sekarang terlihat hanya ada dua variabel, dan terlihat kedua variabel tersebut mempunyai angka SIG. OF F TO REMOVE di atas 0,05 (yakni 0,524 untuk USIA dan 0,798 untuk OLAHRAGA). Karena sudah tidak ada variabel yang memenuhi syarat maka proses pengeluaran variabel berhenti, dan kedua variabel sisa tersebut tidak dikeluarkan, yang berarti keduanya termasuk pada VARIABLE NOT IN THE ANALYSIS, atau variabel yang tidak dianalisis lebih lanjut.

Wilks' Lambda

Step	Number of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	1	.894	1	1	73	8.656	1	73.000	.004
2	2	.834	2	1	73	7.170	2	72.000	.001
3	3	.763	3	1	73	7.343	3	71.000	.000

Wilk's Lambda pada prinsipnya adalah varians total dalam discriminant scores yang tidak bisa dijelaskan oleh perbedaan di antara grup-grup yang ada. Perhatikan tabel di atas yang terdiri atas tiga tahap (step), yang terkait dengan tiga variabel yang secara berurutan dimasukkan pada tahapan analisis sebelumnya.

Pada step 1, jumlah variabel yang dimasukkan ada satu (INCOME), dengan angka Wilk's Lambda adalah 0,894. Hal ini berarti 89,4% varians tidak dapat dijelaskan oleh perbedaan antara grup-grup. Kemudian pada step 2, dengan

tambahan variable BERAT (lihat kolom NUMBER OF VARIABLES yang sekarang adalah 2), angka Wilk's Lambda turun menjadi 0,834. Dan pada step 3, angka itu turun lagi menjadi 0,763. Penurunan angka Wilk's Lambda tentu baik bagi model diskriminan, karena varians yang tidak bisa dijelaskan juga semakin kecil (dari 89,4% menjadi 76,3%).

Dari kolom F dan signifikansinya, terlihat baik pada pemasukan variable 1, 2, kemudian 3, semuanya adalah signifikan secara statistik. Hal ini berarti ketiga variable tersebut (INCOME, BERAT dan TINGGI) memang berbeda untuk kedua tipe konsumen.

### Summary of Canonical Discriminant Functions

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.310 <sup>a</sup>	100.0	100.0	.487

a. First 1 canonical discriminant functions were used in the analysis.

*Canonical Correlation* mengukur keeratan hubungan antara discriminant score dengan grup (dalam hal ini, karena ada dua tipe konsumen, maka ada dua grup). Angka 0,487 menunjukkan keeratan yang cukup tinggi, dengan ukuran skala asosiasi antara 0 sampai 1.

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	.763	19.321	3	.000

Tabel di atas menyatakan angka akhir dari Wilk's Lambda, yang sebenarnya sama saja dengan angka terakhir dari step 3 pembuatan model diskriminan (lihat table terdahulu). Angka Chi-Square sebesar 19,321 dengan tingkat signifikansi yang tinggi menunjukkan perbedaan yang jelas antara dua grup konsumen (mereka yang BANYAK minum dengan yang SEDIKIT minum).

Structure Matrix	
	Function
	1
income	.618
usia <sup>a</sup>	.521
berat	.478
olahraga <sup>a</sup>	.475
tinggi	-.430

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions  
Variables ordered by absolute size of correlation within function.

a. This variable not used in the analysis.

Tabel STRUCTURE MATRIX menjelaskan korelasi antara variable independen dengan fungsi diskriminan yang terbentuk. Terlihat variable INCOME paling erat hubungannya dengan fungsi diskriminan, diikuti oleh variable USIA, BERAT dan seterusnya. Hanya di sini variable USIA dan OLAHRAGA tidak dimasukkan dalam model diskriminan (perhatikan tanda huruf a di dekat variable tersebut). Perhatikan tanda korelasi yang sama dengan tanda koefisien pada model Discriminant Score.

Canonical Discriminant Function Coefficients	
	Function
	1
berat	.064
tinggi	-.093
income	.006
(Constant)	7.884

Unstandardized coefficients

Tabel di atas mempunyai fungsi yang hampir mirip dengan persamaan regresi berganda, yang dalam analisis diskriminan disebut sebagai FUNGSI DISKRIMINAN:

$$z \text{ Score} = 7,884 + 0,064 \text{ BERAT} - 0,093 \text{ TINGGI} + 0,006 \text{ INCOME}$$

Kegunaan fungsi ini untuk mengetahui sebuah case (dalam kasus ini adalah seorang konsumen) masuk pada Grup yang satu, ataukah tergolong pada grup yang lainnya.

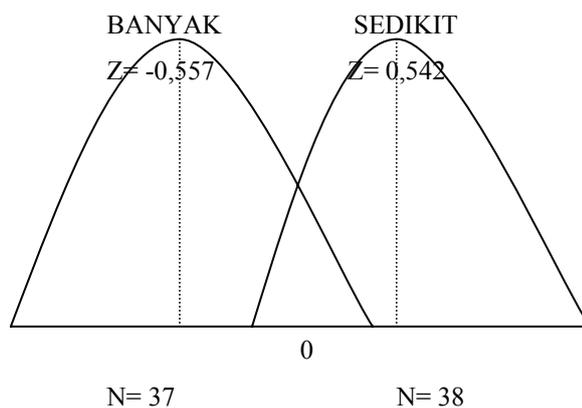
Selain fungsi di atas, dengan dipilihnya FISHER FUNCTION COEFFICIENT pada proses analisis, maka akan terbentuk pula Fungsi Diskriminan Fisher (lihat pembahasan selanjutnya).

Functions at Group Centroids	
	Function
minum	1
sedikit	.542
banyak	-.557

Unstandardized canonical discriminant functions evaluated at group means

Karena ada dua tipe Konsumen, maka disebut Two-Group Discriminant, di mana grup yang satu mempunyai Centroid (Group Means) negatif, dan grup yang satu mempunyai Centroid (Group Means) positif. Angka pada table menunjukkan besaran Z yang memisahkan kedua grup tersebut.

**Gambar:**



Terlihat distribusi anggota grup dengan kode 0 (SEDIKIT) dan kode 1 (BANYAK), di mana dari 75 responden, 37 orang ada pada Grup BANYAK,

dan 38 orang ada pada grup SEDIKIT (komposisi anggota lihat penjelasan selanjutnya).

Tampilan gambar di atas akan digunakan untuk menentukan apakah seorang Responden akan tergolong pada Grup SEDIKIT atau BANYAK (lihat penjelasan bagian CASEWISE RESULT).

**Prior Probabilities for Groups**

minum	Prior	Cases Used in Analysis	
		Unweighted	Weighted
sedikit	.500	38	38.000
banyak	.500	37	37.000
Total	1.000	75	75.000

Tabel di atas memperlihatkan komposisi ke 75 responden, yang dengan model diskriminan menghasilkan 37 responden ada di Grup BANYAK, sedang sisanya ada di grup SEDIKIT.

**Classification Function Coefficients**

	minum	
	sedikit	banyak
berat	.035	-.035
tinggi	3.587	3.690
income	.036	.030
(Constant)	-297.882	-306.557

Fisher's linear discriminant functions

Sama seperti tampilan Unstandardized (Canonical) sebelumnya, Fungsi Diskriminan dari Fisher pada prinsipnya membuat semacam persamaan regresi, dengan pembagian berdasar kode grup.

**Mereka yang minum air mineral dalam kategori SEDIKIT:**

$$\text{SCORE} = -297,882 + 0,03506 \text{ BERAT} + 3,587 \text{ TINGGI} + 0,03641 \text{ INCOME}$$

**Mereka yang minum air mineral dalam kategori BANYAK:**

$$\text{SCORE} = -306,557 - 0,03481 \text{ BERAT} + 3,690 \text{ TINGGI} + 0,03007 \text{ INCOME}$$

Selisih di antara grup SEDIKIT dengan BANYAK adalah:

$$\begin{aligned} &(-297,882+0,03506 \text{ BERAT}+3,587 \text{ TINGGI}+0,03641 \text{ INCOME})- \\ &(-306,557-0,03481 \text{ BERAT}+3,690 \text{ TINGGI}+0,03007 \text{ INCOME}) \end{aligned}$$

atau:

$$z \text{ Score} = 8,675+0,06987 \text{ BERAT}-0,103 \text{ TINGGI} +0,0064 \text{ INCOME}$$

Perhatikan a score dari Fungsi Fisher hampir sama dengan fungsi Unstandardized sebelumnya:

$$z \text{ Score} = 7,884 +0,064 \text{ BERAT}-0,093 \text{ TINGGI} +0,006 \text{ INCOME}$$

Untuk aplikasi pada penghitungan score pada Casewise Statistics, akan digunakan fungsi Unstandardized. Sedangkan fungsi Fisher sebenarnya bersifat proporsional dengan fungsi Unstandardized, yang untuk kasus ini, dengan mengalikan setiap koefisien dari Fisher dengan angka 0,9 maka akan diperoleh Fungsi z Score Unstandardized.

## CASEWISE STATISTICS

(hanya ditampilkan sebagian)

Casewise Statistics												
Case Number	Actual Group	Predicted Group	Highest Group				Second Highest Group			Discriminant Scores		
			P(O=d   G=g)				Squared Mahalanobis Distance to Centroid	Group	P(G=g   D=d)	Squared Mahalanobis Distance to Centroid	Function 1	Function 2
			g	df	P(G=g   D=d)							
Original	1	0	0	.304	1	.850	1.058	1	.150	4.527	1.571	
	2	0	0	.237	1	.870	1.397	1	.130	5.203	1.724	
	3	0	0	.811	1	.584	.057	1	.416	.739	.303	
	4	0	0	.520	1	.788	.413	1	.212	3.034	1.185	
	5	0	0	.783	1	.575	.076	1	.425	.677	.286	
	6	1	1	.858	1	.600	.032	0	.400	.847	-.378	
	7	0	1**	.970	1	.637	.001	0	.363	1.127	-.520	
	8	0	0	.895	1	.613	.017	1	.387	.936	.410	
	9	1	0**	.088	1	.923	2.917	1	.077	7.879	2.250	
	10	0	0	.078	1	.927	3.111	1	.073	8.196	2.306	
	11	0	0	.514	1	.789	.425	1	.211	3.067	1.194	
	12	0	0	.777	1	.714	.080	1	.286	1.910	.825	
	13	0	0	.667	1	.746	.186	1	.254	2.341	.973	
	14	1	0**	.760	1	.567	.094	1	.433	.629	.236	
	15	1	1	.739	1	.559	.111	0	.441	.586	-.223	
	16	1	1	.976	1	.639	.001	0	.361	1.144	-.527	
	17	0	0	.069	1	.931	3.303	1	.069	8.507	2.380	
	18	1	1	.933	1	.667	.007	0	.333	1.399	-.640	

Tabel CASEWISE pada prinsipnya ingin menguji apakah model diskriminan yang terbentuk akan mengelompokkan dengan tepat seorang responden pada kategori minum BANYAK atau SEDIKIT.

Sebagai contoh, lihat tampilan sebagian data responden mula-mula.

	nama	minum	usia	berat	tinggi	income	jamkerja	olahraga
1	RUSDI	0.00	40.00	65.00	154.00	680.00	5.33	3.0
2	NINA	0.00	30.00	70.00	157.00	700.00	5.30	3.6
3	LANNY	0.00	25.00	60.00	158.00	580.00	5.27	3.5
4	CITRA	0.00	26.00	75.00	160.00	600.00	5.33	3.0
5	DINA	0.00	40.00	50.00	159.00	700.00	5.50	3.5
6	SISKA	1.00	28.00	62.00	158.00	440.00	5.00	2.2
7	LUSI	0.00	29.00	50.00	160.00	580.00	5.07	2.9

**Gambar 5.8. Tampilan Data Kasus**

- Untuk Responden 1 dengan nama RUSDI yang mempunyai kategori minum SEDIKIT dan data variabel:

Berat=65 kilogram

Tinggi=154 centimeter

Income= Rp.680,-/bulan (dalam ribuan)

Maka perhitungan score untuk konsumen bernama Rusdi adalah:

$$z \text{ Score} = 7,884 + (0,064 * 65) - (0,093 * 154) + (0,006 * 680) = 1,802$$

Jika dilihat pada kolom terakhir dari tampilan CASEWISE STATISTICS, terlihat angka DISCRIMINANT SCORE untuk case number 1 (nama Rusdi) adalah 1,571. Perbedaan angka 1,802 dengan 1,571 disebabkan penyajian angka pada z score yang tidak lengkap (dilakukan pembulatan pada output SPSS).

Seharusnya fungsi diskriminan adalah:

$$z \text{ Score} = 7,884 + 0,0637 \text{ BERAT} - 0,093 \text{ TINGGI} + 0,0057 \text{ INCOME}$$

- Dengan demikian, konsumen Rusdi mempunyai score:

$$z \text{ Score} = 7,884 + (0,0637 * 65) - (0,093 * 154) + (0,0057 * 680) = 1,5785$$

NB: bandingkan dengan hasil di output sebesar 1,571.

- Untuk case number 2 (nama NINA) dengan data variabel bisa dilihat pada file DISKRIMINAN.sav:

$$z \text{ Score} = 7,884 + (0,0637 * 70) - (0,093 * 157) + (0,0057 * 700) = 1,732$$

NB: bandingkan dengan hasil pada kolom DISCRIMINANT SCORE di Output CASEWISE, yang adalah 1,724.

- Untuk case number 3 (nama LANNY) dengan data variabel bisa dilihat pada file DISKRIMINAN.sav:

$$z \text{ Score} = 7,884 + (0,0637 * 60) - (0,093 * 158) + (0,0057 * 580) = 0,3022$$

NB: bandingkan dengan hasil pada kolom DISCRIMINANT SCORE di Output CASEWISE, yang adalah 0,303.

Demikian seterusnya bisa dilakukan pembuatan score untuk masing-masing konsumen.

Score setiap konsumen (case) otomatis bisa ditampilkan pada DATA EDITOR atau file DISKRIMINAN.sav. Jika pada proses (kotak dialog) Diskriminan Analysis, buka option SAVE, dan aktifkan kotak DISCRIMINANT SCORE.

Hasil score tersebut selanjutnya akan dibandingkan dengan cut off score, untuk mengetahui apakah case (konsumen) masuk ke Grup SEDIKIT ataukah Grup BANYAK.

- Selanjutnya akan dijelaskan pembuatan cut off score (nilai batas).

Dari tabel PRIOR PROBABILITIES FOR GROUPS, didapat bahwa jumlah responden SEDIKIT adalah 38 orang, sedangkan responden BANYAK adalah 37 orang. Dengan demikian, dikaitkan dengan angka grup centroid:

$$(38 \times 0,542) + (37 \times -0,557) = 20,59 - 20,60 \text{ atau praktis sama dengan } 0.$$

Perhitungan  $Z_{CU}$  (angka kritis):

$$Z_{CU} = \frac{N_A Z_B + N_B Z_A}{N_A + N_B}$$

di mana:

$Z_{CU}$  = Angka Kritis, yang berfungsi sebagai cut off score

$N_A$  dan  $N_B$  = Jumlah sampel di grup A dan B, yang dalam kasus ini adalah grup SEDIKIT dan grup BANYAK.

$Z_A$  dan  $Z_B$  = angka centroid pada grup A dan B.

Perhitungan:

$$Z_{CU} = \frac{38 \cdot -0,557 + 37 \cdot 0,542}{38 + 37}$$

$$= -0,01483 \text{ atau praktis sama dengan } 0.$$

NB: perhatikan angka 38 dan 37 yang memang dibalik dibandingkan dengan angka grup centroid.

Penggunaan angka  $Z_{CU}$  ( Discriminating Z Score):

- Angka skor kasus di atas  $Z_{CU}$ , masuk ke grup SEDIKIT (kode 0).
- Angka skor kasus di bawah  $Z_{CU}$ , masuk ke grup BANYAK (kode 1).

Penggunaan angka 0 sebagai pembatas pada kasus ini karena *kebetulan didapat angka yang praktis sama dengan nol*. Pada banyak kasus lainnya, tentu angka pembatas bisa tidak sama dengan nol, seperti -2,4 atau +1,78 dan sebagainya.

Sebagai contoh:

- Case Number 1 (Rusdi) mempunyai score 1,571. Karena  $1,571 > 0$ , maka Rusdi masuk pada Grup 0 (SEDIKIT).
- Case Number 2 (Nina) mempunyai score 1,724. Karena  $1,724 > 0$ , maka Nina masuk pada Grup 0 (SEDIKIT). Atau bisa dikatakan Nina termasuk tipe konsumen yang hanya sedikit mengonsumsi air mineral.
- Case Number 6 (Siska) mempunyai score -0,378. Karena  $-0,378 < 0$  maka Siska masuk pada Grup 1 (BANYAK).

Demikian seterusnya semua responden bisa dikategorikan pada satu *dan satu-satunya* Grup tertentu, Grup SEDIKIT atau Grup BANYAK.

Selain dengan melihat angka Discriminant Score seperti di atas, pengelompokan kasus pada Grup tertentu bisa langsung dilihat pada actual dan predicted Grup.

#### **Penafsiran dengan melihat setiap Baris:**

- Pada baris 1 (Case 1):

Actual Grup=0. Hal ini berarti data awal DISKRIMINAN.sav menyatakan case 1 dikategorikan sebagai Grup 0 (SEDIKIT).

Predicted Grup=0. Hal ini berarti dari hasil perhitungan score, case 1 diprediksi masuk ke Grup=0. Karena sesuai dengan actual Grup yang juga 0, berarti fungsi diskriminan mampu mengategorikan case dengan tepat.

$P(G=g|D=d)=0.850$  pada Highest Group. Hal ini berarti kemungkinan case 1 tepat diklasifikasikan ke grup 0 adalah 85%.

$P(G=g|D=d)=0.150$  pada Second Highest Group. Hal ini berarti kemungkinan case 1 TIDAK tepat diklasifikasikan ke grup 0 adalah 15%.

Perhatikan jumlah kedua kemungkinan adalah (85%+15%) 100%.

- Pada baris 7 (Case 7):

Actual Grup=0. Hal ini berarti data awal DISKRIMINAN.sav menyatakan case 7 dikategorikan sebagai Grup 0 (SEDIKIT).

Predicted Grup=1. Hal ini berarti dari hasil perhitungan score, case 1 diprediksi masuk ke Grup=1. Karena TIDAK sesuai dengan actual Grup yang adalah 0, berarti fungsi diskriminan tidak mampu mengategorikan case dengan tepat. Hal ini ditandai dengan tanda \*\* pada angka 1 di case 7 tersebut.

Demikian seterusnya untuk kasus lainnya, dengan beberapa kasus tidak tepat diklasifikasikan, sehingga diberi tanda \*\*.

Setelah fungsi Diskriminan dibuat, kemudian klasifikasi dilakukan, maka selanjutnya akan dilihat seberapa jauh klasifikasi tersebut sudah tepat? Atau, berapa persen terjadi misklasifikasi pada proses klasifikasi tersebut, yang akan dijelaskan berikut ini.

Classification Results <sup>b,c</sup>						
		MINUM		Predicted Group Membership		Total
		.00 sedikit	1.00 banyak	.00 sedikit	1.00 banyak	
Original	Count	.00 sedikit	27	11	38	
		1.00 banyak	13	24	37	
	%	.00 sedikit	71.1	28.9	100.0	
		1.00 banyak	35.1	64.9	100.0	
Cross-validated <sup>a</sup>	Count	.00 sedikit	27	11	38	
		1.00 banyak	13	24	37	
	%	.00 sedikit	71.1	28.9	100.0	
		1.00 banyak	35.1	64.9	100.0	

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 68.0% of original grouped cases correctly classified.

c. 68.0% of cross-validated grouped cases correctly classified.

Pada bagian ORIGINAL, terlihat bahwa mereka yang pada data awal adalah tergolong SEDIKIT, dan dari klasifikasi Fungsi Diskriminan tetap pada kelompok SEDIKIT, adalah 27 orang. Sedangkan dengan model diskriminan, mereka yang awalnya masuk grup SEDIKIT, ternyata menjadi anggota grup BANYAK adalah 11 orang.

Demikian juga dengan grup BANYAK, yang tetap pada grup BANYAK sejumlah 24 orang, dan yang 'meleset' adalah 13 orang.

Dengan demikian, ketepatan prediksi dari model adalah:

$$(27+24)/75=0,68 \text{ atau } 68\%.$$

Karena angka ketepatan tinggi (68%) maka model diskriminan di atas sebenarnya bisa digunakan untuk analisis diskriminan. Atau penafsiran tentang berbagai tabel yang ada (lihat seluruh pembahasan di atas) valid untuk digunakan.

Pendapat lain mengatakan bahwa klasifikasi di atas terlalu optimis, dan tidak memperhitungkan berbagai bias yang mungkin terjadi. Untuk itu, disarankan juga penggunaan metode *Leave-one-out cross validation*, untuk mengurangi bias yang mungkin terjadi pada proses klasifikasi di atas. Dari keterangan tabel paling bawah didapat angka ketepatan klasifikasi data ke grup dengan metode *Leave-one-out cross validation*, yaitu tetap 68%, yang masih bisa dikategorikan ketepatan klasifikasi tetap tinggi.

Setelah terbukti bahwa Fungsi Diskriminan mempunyai ketepatan prediksi yang tinggi, maka fungsi diskriminan tersebut bisa digunakan untuk memprediksi sebuah kasus, apakah akan diklasifikasikan ke tipe SEDIKIT atukah tipe BANYAK.

### Contoh Kegunaan Fungsi Diskriminan

Jika ada seorang konsumen bernama LUCKY dengan berat badan 70 kilogram, tinggi badan 175 centimeter, dan mempunyai income rata-rata Rp.1.000.000,- per bulan, maka dengan memasukkan data tersebut ke dalam fungsi diskriminan:

$$z \text{ Score} = 7,884 + (0,0637 * 70) - (0,093 * 175) + (0,0057 * 1000) = +1,768$$

NB: perhatikan angka 1.000.000 dimasukkan sebagai 1000, karena data income dalam ribuan rupiah.

Karena angka skor (+1,768) di atas  $Z_{CU}$ , maka kasus masuk ke grup SEDIKIT. Atau, konsumen bernama LUCKY termasuk tipe konsumen yang sedikit meminum air mineral.

Demikian seterusnya, sejauh data mengenai berat badan, tinggi badan dan income seseorang diketahui, maka bisa dilakukan klasifikasi tipe minum orang tersebut.

#### 4. KESIMPULAN

Dari proses diskriminan, dimulai dari uji variabel sampai analisis output, didapat kesimpulan yang terkait dengan tujuan pada kasus awal (lihat kasus sebelumnya).

- ADA perbedaan yang signifikan antara Mereka yang banyak minum AMDK dengan mereka yang sedikit meminumnya. Hal ini dibuktikan pada analisis Wilk's Lambda.
- Variabel yang membuat perilaku konsumsi air mineral mereka berbeda adalah BERAT BADAN, TINGGI BADAN dan INCOME. Hal ini terlihat pada step analisis awal, baik pada bagian VARIABLE IN ANALYSIS maupun VARIABLE NOT IN ANALYSIS.
- Model atau fungsi diskriminan untuk kasus tersebut adalah:  
$$z \text{ Score} = 7,884 + 0,0637 \text{ BERAT} - 0,093 \text{ TINGGI} + 0,0057 \text{ INCOME}$$
- Model (fungsi) diskriminan di atas mempunyai ketepatan mengklasifikasi kasus sebesar 68%. Karena di atas 50%, ketepatan model dianggap tinggi, dan model di atas bisa digunakan untuk mengklasifikasi sebuah kasus pada tipe minum tertentu.

### 5.3 Analisis Diskriminan Tiga Faktor

Kasus ini merupakan pengembangan lebih lanjut dari model diskriminan terdahulu. Jika kasus sebelumnya hanya ada dua kategori untuk variabel dependen, sekarang akan dibahas pembuatan model diskriminan dan pengujiannya untuk tiga kategori (Multiple Discriminant Analysis).

#### KASUS

Data untuk kasus ini sama dengan kasus sebelumnya, yakni profil 75 pembeli Air Mineral Dalam Kemasan (AMDK), yang terdiri atas USIA, BERAT, TINGGI, INCOME, JAM KERJA dan OLAHRAGA. Namun kasus ini mengembangkan *Dependent Variable* atau variabel tergantung yang adalah variabel MINUM, dengan tiga kode, yakni:

- Kode 0 untuk kategori minum SEDIKIT
- Kode 1 untuk kategori minum SEDANG
- Kode 2 untuk kategori minum BANYAK

Analisis Diskriminan pada kasus ini bertujuan untuk mengetahui:

1. Apakah ada perbedaan yang signifikan antara ketiga tipe responden tersebut? Atau, apakah mereka yang tergolong minum sedikit air mineral mempunyai perbedaan (dalam usia, berat badan, tinggi badan, income per bulan, jam kerja dan waktu berolahraga) dengan mereka yang termasuk peminum sedang dan banyak?
2. Jika memang ketiga perilaku berbeda, variabel mana yang membedakan perilaku minum ketiga kelompok konsumen tersebut? Apakah Usia seseorang, Berat Badannya, atau yang lainnya.
3. Membuat fungsi diskriminan untuk menentukan apakah seseorang termasuk pada tipe peminum SEDIKIT, SEDANG atukah BANYAK.
4. Menguji apakah fungsi diskriminan yang terbentuk mempunyai tingkat ketepatan yang cukup tinggi untuk menggolongkan ketiga tipe konsumen tersebut. Sebagai contoh, jika pada data awal konsumen bernama Rusdi termasuk tipe peminum air mineral SEDIKIT, apakah fungsi diskriminan yang terbentuk mampu dengan tepat mengklasifikasikan Rusdi pada tipe SEDIKIT? Jika tidak semua konsumen bisa diklasifikasikan dengan tepat, berapa persen ketepatan prediksi fungsi diskriminan tersebut?

Keempat tujuan di atas sebenarnya merupakan tujuan dasar dari setiap analisis diskriminan, walaupun dalam pengolahan SPSS, hal itu tidak bisa dijelaskan secara *berurutan*. Juga berbeda dengan kasus sebelumnya untuk diskriminan dua faktor, pada kasus ini tidak disertakan proses pengujian variabel terlebih dahulu. Keenam variabel akan diproses secara bersama-sama, walaupun tetap menggunakan proses bertahap (stepwise).

## DATA

Data yang digunakan adalah **DISKRIMINAN 3 FAKTOR.sav**.

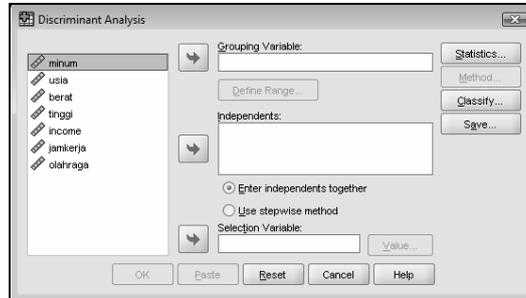
NB: Data ini hampir sama dengan data DISKRIMINAN.sav seperti kasus sebelumnya. Pengubahan input hanya ada pada variabel MINUM dan USIA.

## Langkah

### Proses Analisis Diskriminan

Karena ada tiga kategori pada variabel dependent (MINUM), maka yang akan dibuat adalah model dua faktor.

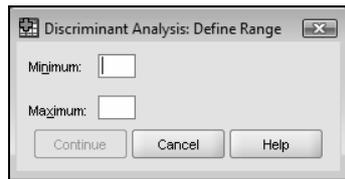
- Buka file **diskriminan 3 faktor**.
- Menu **Analyze → Classify → Discriminant....** Tampak di layar:



**Gambar 5.9. Kotak Dialog Discriminant Analysis**

Pengisian (mirip dengan pengisian pada kasus sebelumnya).

- Masukkan variabel **minum** ke bagian GROUPING VARIABLE.
- Kemudian buka ikon DEFINE RANGE..., hingga tampak di layar:



**Gambar 5.10. Kotak Dialog Define Range**

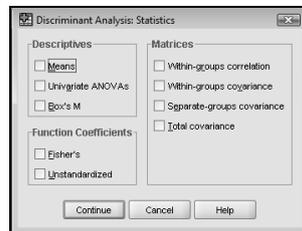
Sesuai kode untuk variabel MINUM:

- Masukkan angka **0** pada Minimum.
- Masukkan angka **2** pada Maximum.

Perhatikan pemasukan hanya angka minimum dan maksimum. Jika ada 5 kategori (angka 0, 1, 2, 3, dan 4), maka angka maksimum tentunya adalah 4.

Tekan tombol CONTINUE untuk kembali ke kotak dialog utama.

- Masukkan variabel **usia**, **berat**, **tinggi**, **income**, **jamkerja** dan **olahraga** ke dalam kotak INDEPENDENT.
- Klik ikon STATISTICS. Tampak di layar:



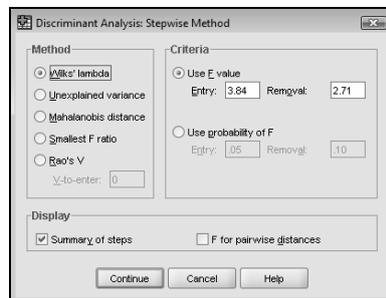
**Gambar 5.11. Kotak Dialog Statistics**

Pengisian:

- Pada bagian DESCRIPTIVES, aktifkan pilihan **Univariate ANOVAs**.
- Pada bagian FUNCTION COEFFICIENTS, aktifkan *hanya* pilihan **Unstandardized**.

Abaikan bagian lain, tekan CONTINUE untuk kembali ke kotak dialog utama.

- Perhatikan bagian tengah kotak dialog utama. Klik mouse pada pilihan **Use stepwise method**, maka secara otomatis ikon METHOD akan terbuka (aktif). Klik ikon METHOD tersebut. Tampak di layar:



**Gambar 5.12. Kotak Dialog Stepwise Method**

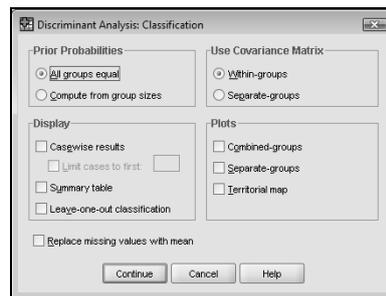
Pengisian:

- Pada bagian METHOD, pilih **Mahalanobis distance**.
- Pada bagian CRITERIA, pilih **Use Probability of F**, namun jangan mengubah isi yang sudah ada (default).
- Pada bagian DISPLAY, aktifkan pula kotak **F for pairwise distances**.

Pilihan ini untuk mengetahui kelompok mana yang paling berbeda (paling jauh jaraknya) dan mana yang paling dekat dengan kelompok lainnya.

Abaikan bagian yang lain, kemudian tekan CONTINUE untuk kembali ke kotak dialog utama.

- Klik ikon CLASSIFY. Tampak di layar:



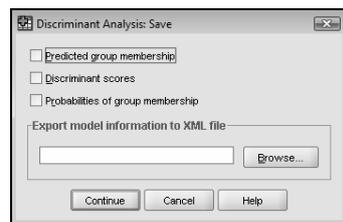
**Gambar 5.13. Kotak Dialog Classification**

Pengisian:

- Pada bagian DISPLAY, aktifkan pilihan **Casewise results**.
- Masih pada bagian DISPLAY, aktifkan juga pilihan **Leave-one-out-classification**.
- Pada bagian PLOTS, aktifkan pilihan **Territorial map**.

Abaikan bagian lain, dan tekan CONTINUE untuk kembali ke kotak dialog utama.

- Klik ikon SAVE. Tampak di layar:



**Gambar 5.14. Kotak Dialog Save**

Kotak dialog SAVE berfungsi untuk menyimpan hasil proses analisis diskriminan dalam bentuk variabel baru di DATA EDITOR.

Pengisian:

- Aktifkan pilihan **Prediction group membership**, dan **Discriminant scores**.

Abaikan bagian lain, dan tekan CONTINUE untuk kembali ke kotak dialog utama.

Abaikan ikon yang lain, dan tekan OK untuk proses data.

### Output dan Analisis

Ada dua jenis output yang dihasilkan.

1. Output berupa Data File baru (tipe sav) yang berisi data semula dengan tambahan tiga variabel (kolom) baru hasil proses diskriminan. Output tersebut bisa dilihat pada file **diskriminan 3 faktor hasil.sav**. Penjelasan ketiga kolom tersebut akan terkait dengan penjelasan output lainnya.
2. Output dengan ekstensi spo seperti standar output SPSS, yang bisa disimpan dengan nama **diskriminan 3 faktor.spv**.

OUTPUT AKAN DITAMPILKAN PER BAGIAN UNTUK ANALISIS, DENGAN CATATAN TIDAK SEMUA OUTPUT DITAMPILKAN.

### Uji Variable

Tests of Equality of Group Means					
	Wilks' Lambda	F	df1	df2	Sig.
usia	.904	3.821	2	72	.026
berat	.851	6.318	2	72	.003
tinggi	.984	.589	2	72	.557
income	.972	1.038	2	72	.360
jamkerja	.972	1.039	2	72	.359
olahraga	.898	4.089	2	72	.021

Tabel di atas berfungsi untuk menguji apakah ada perbedaan yang signifikan antar-grup untuk setiap variabel.

Pedoman (lihat angka Sig. untuk F test):

- Jika Sig. > 0,05 berarti tidak ada perbedaan antar-grup.
- Jika Sig. < 0,05 berarti ada perbedaan antar-grup.

Contoh analisis menggunakan uji F:

- Variabel USIA, angka Sig. adalah di bawah 0,05 (0,026). Hal ini berarti ada perbedaan antar-grup, atau konsumen yang sedikit, sedang atau banyak minum air mineral dipengaruhi oleh usia konsumen tersebut. Mungkin saja, makin tua seseorang, orang tersebut cenderung lebih banyak mengonsumsi air mineral. Atau kemungkinan lainnya yang akan dibahas pada penafsiran fungsi diskriminan.
- Variabel TINGGI, angka Sig. Di atas 0,05 (0,557). Hal ini berarti tidak ada perbedaan antar-grup, atau responden yang sedikit, sedang atau banyak minum air mineral tidak dipengaruhi oleh tinggi badan konsumen tersebut. Seseorang yang dianggap tinggi ternyata mengonsumsi air mineral yang relatif sama dengan mereka yang dianggap pendek atau sedang tinggi tubuhnya.

Dari enam variabel, hanya tiga variabel yang berbeda secara signifikan untuk tiga grup, yaitu USIA, BERAT, dan OLAHRAGA. Hal ini berarti perilaku minum air mineral dipengaruhi oleh usia seseorang, berat badannya dan waktu yang dihabiskan untuk melakukan aktivitas olahraga. Namun demikian, hal ini tidak menjamin apakah tiga variabel tersebut akan dimasukkan pada fungsi diskriminan. Untuk itu, dilakukan analisis diskriminan dengan *tetap menyertakan seluruh variabel yang ada*.

### Variabel yang Membentuk Fungsi Diskriminan

Variables Entered/Removed <sup>a,b,c,d</sup>							
Step	Entered	Min. D Squared			Exact F		
		Statistic	Between Groups	Statistic	df1	df2	Sig.
1	usia	.093	sedang and banyak	1.083	1	72.000	.302
2	berat	.162	sedang and banyak	.929	2	71.000	.400

At each step, the variable that maximizes the Mahalanobis distance between the two closest groups is entered.

a. Maximum number of steps is 12.  
 b. Maximum significance of F to enter is .05.  
 c. Minimum significance of F to remove is .10.  
 d. F level, tolerance, or VLN insufficient for further computation.

Dari tabel di atas, ternyata hanya dua variabel yang akan digunakan untuk membentuk fungsi diskriminan, yakni variabel USIA dan BERAT. Sedangkan variabel OLAHRAGA ternyata tidak masuk dalam fungsi diskriminan.

### Proses Pemasukan Variabel Dilihat dari Angka Wilk's Lambda

Wilks' Lambda									
Step	Number of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	1	.904	1	2	72	3.821	2	72.000	2.648E-02
2	2	.789	2	2	72	4.456	4	142.000	2.015E-03

Pada step 1, jumlah variabel yang dimasukkan ada satu (USIA), dengan angka Wilk's Lambda adalah 0,904. Hal ini berarti 90,4% varians tidak dapat dijelaskan oleh perbedaan antara grup-grup. Kemudian pada step 2, dengan tambahan variable BERAT (lihat kolom NUMBER OF VARIABLES yang sekarang adalah 2), angka Wilk's Lambda turun menjadi 0,789. Penurunan angka Wilk's Lambda tentu baik bagi model diskriminan, karena varians yang tidak bisa dijelaskan juga semakin kecil (dari 90,4% menjadi 78,9%).

Dari kolom F dan signifikansinya, terlihat baik pada pemasukan variable 1 ataupun 2, semuanya adalah signifikan secara statistik. Hal ini berarti kedua variabel tersebut (USIA dan BERAT) memang berbeda untuk ketiga tipe konsumen.

### Analisis Perbedaan Antar-Grup Konsumen

Pairwise Group Comparisons <sup>a,b</sup>					
Step	MINUM		.00 sedikit	1.00 sedang	2.00 banyak
1	.00 sedikit	F		3.041	7.306
		Sig.		.085	.009
	1.00 sedang	F	3.041		1.083
		Sig.	.085		.302
	2.00 banyak	F	7.306	1.083	
		Sig.	.009	.302	
2	.00 sedikit	F		4.441	8.697
		Sig.		.015	.000
	1.00 sedang	F	4.441		.929
		Sig.	.015		.400
	2.00 banyak	F	8.697	.929	
		Sig.	.000	.400	

a. 1, 72 degrees of freedom for step 1.  
b. 2, 71 degrees of freedom for step 2.

Tabel ini berkaitan dengan penjelasan angka Wilk's Lambda sebelumnya, di mana ada dua tahapan (proses) yang menghasilkan dua variabel pada pembentukan fungsi diskriminan.

Pada step kedua yang merupakan proses akhir, terlihat jarak (distance) antara Grup SEDIKIT (kode 0) dengan Grup BANYAK (kode 2) adalah yang terbesar, yakni 8,697. Sedangkan jarak terkecil adalah antara Grup SEDANG (kode 1) dengan Grup BANYAK (kode 2), yakni 0,929.

Dengan demikian, bisa dikatakan bahwa Konsumen di Grup SEDIKIT paling berbeda profilnya (Usia dan Berat Badannya) dibandingkan dengan mereka yang banyak minumannya. Sebaliknya, Usia dan Berat Badan konsumen yang konsumsi air mineralnya termasuk sedang, mempunyai perbedaan yang kecil dengan mereka yang konsumsinya termasuk banyak.

### Menguji Perbedaan Antar-Grup Konsumen

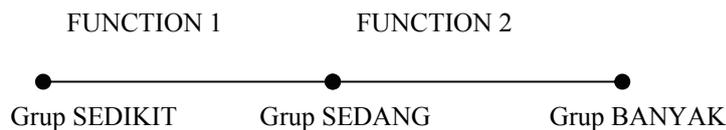
#### Summary of Canonical Discriminant Functions

Eigenvalues				
Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.266 <sup>a</sup>	99.7	99.7	.458
2	.001 <sup>a</sup>	.3	100.0	.030

a. First 2 canonical discriminant functions were used in the analysis.

Jika pada kasus dua grup seperti pada kasus sebelumnya, otomatis akan terbentuk satu fungsi diskriminan (function). Karena memang *sebuah* fungsi diskriminan berfungsi untuk menempatkan sebuah kasus pada pilihan *dua grup tertentu*, apakah akan masuk ke grup yang satu atau ke grup yang lain.

Dengan demikian, secara logika, jika ada tiga grup seperti pada kasus di atas, akan terbentuk dua fungsi diskriminan, dengan kriteria:



- Fungsi Diskriminan 1 untuk memilah mana yang masuk ke Grup SEDIKIT atau ke Grup SEDANG.
- Fungsi Diskriminan 2 untuk memilah mana yang masuk ke Grup SEDANG atau ke Grup BANYAK.

Dengan demikian, jika ada empat grup, maka akan terbentuk tiga fungsi diskriminan, dan seterusnya.

Sekarang perhatikan angka Canonical Correlation yang mengukur keeratan hubungan antara discriminant score dengan grup (dalam hal ini, karena ada

tiga tipe konsumen, maka ada tiga grup). Angka 0,458 menunjukkan keeratan yang cukup tinggi, dengan ukuran skala asosiasi antara 0 sampai 1. Dan walaupun angka canonical untuk fungsi kedua (function 2) dengan grup adalah lemah (0,03), namun kedua fungsi tetap digunakan untuk interpretasi selanjutnya. Hal ini bisa dilihat pada keterangan pada bagian bawah tabel output.

Wilks' Lambda				
Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.789	16.908	4	.002
2	.999	.065	1	.799

Dari tabel Wilk's Lambda, pada kolom TEST OF FUNCTION(S) 1 THROUGH 2 menguji hipotesa:

- Ho: Tidak ada perbedaan Rata-rata (Centroid) dari kedua fungsi diskriminan.
- Hi: Ada perbedaan Rata-rata (Centroid) yang jelas dari kedua fungsi diskriminan.

Untuk menguji Hipotesa, angka Wilk's Lambda ditransformasikan ke angka Chi-Square, dengan ketentuan:

- Angka Sig. > 0,05 maka Ho diterima.
- Angka Sig. < 0,05 maka Ho ditolak.

Terlihat angka Chi Square hitung adalah 16,908 dengan signifikansi 0,002, yang jauh di bawah 0,05. Karena itu, Ho ditolak, atau memang ada perbedaan yang nyata (signifikan) antara Rata-rata (centroid) dari kedua fungsi diskriminan yang telah terbentuk. Karena ada perbedaan yang nyata, maka perilaku minum untuk ketiga tipe konsumen memang berbeda (lihat tujuan pertama dari kasus di atas). Bisa dikatakan, usia dan berat dari mereka yang sedikit, sedang atau banyak minum air mineral memang berbeda.

Kemudian beralih ke baris 2. Interpretasi sama persis dengan baris 1, hanya di sini *fungsi diskriminan pertama dikeluarkan*, sehingga hanya fungsi kedua, yang mengategorikan antara mereka yang tipe minumannya SEDANG, dengan mereka yang tipe minumannya BANYAK.

### Hipotesa

Ho: Tidak ada perbedaan Rata-rata (Centroid) dari fungsi diskriminan kedua (variabel usia dan berat pada tipe SEDANG dengan tipe BANYAK).

Hi: Ada perbedaan Rata-rata (Centroid) yang jelas dari fungsi diskriminan kedua (variabel usia dan berat pada tipe SEDANG dengan tipe BANYAK).

Untuk menguji Hipotesa, angka Wilk's Lambda ditransformasikan ke angka Chi-Square. Karena hasil angka Sig. adalah 0,799 yang jauh di atas 0,05 maka Ho diterima, atau jika fungsi diskriminan kedua berdiri sendiri, justru tidak ada perbedaan antara usia dan berat dari tipe minum SEDANG dan BANYAK.

Dengan demikian, *seharusnya kedua fungsi diskriminan digunakan secara bersama-sama*, dan tidak bisa hanya mengambil fungsi diskriminan kedua saja.

### Penamaan Fungsi Diskriminan Satu dan Dua

Setelah pada analisis sebelumnya terbukti bahwa kedua fungsi diskriminan harus dimasukkan semuanya untuk interpretasi selanjutnya, analisis dilanjutkan dengan upaya menamakan kedua fungsi tersebut.

Structure Matrix		
	Function	
	1	2
berat	.812 <sup>*</sup>	-.583
tinggi <sup>a</sup>	.192 <sup>*</sup>	-.102
jamkerja <sup>a</sup>	.054 <sup>*</sup>	-.043
usia	.630	.776 <sup>*</sup>
income <sup>a</sup>	.446	.663 <sup>*</sup>
olahraga <sup>a</sup>	.199	.492 <sup>*</sup>

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions  
Variables ordered by absolute size of correlation within function.

\*. Largest absolute correlation between each variable and any discriminant function

a. This variable not used in the analysis.

Pada kasus ini, tabel STRUCTURE MATRIX menjelaskan korelasi antara variabel independen dengan dua fungsi diskriminan yang terbentuk. Perhatikan variabel-variabel dengan tanda huruf a di kiri atas nama variabel tersebut, yang menunjukkan variabel dimaksud tidak diikuti pada fungsi diskriminan.

Dari kriteria tersebut, hanya dua variabel independen yang lolos uji, yakni BERAT dan USIA. Dan jika dilihat dari besar korelasi (abaikan tanda - jika ada):

- Korelasi variabel BERAT dengan fungsi 1 (0,812) lebih besar daripada korelasi variabel tersebut dengan fungsi 2 (0,583). Dengan demikian, variabel BERAT masuk ke fungsi diskriminan 1.
- Korelasi variabel USIA dengan fungsi 2 (0,776) lebih besar daripada korelasi variabel tersebut dengan fungsi 1 (0,630). Dengan demikian, variabel USIA masuk ke fungsi diskriminan 2.

Perhatikan tanda \* pada angka yang lebih besar jika dibandingkan per baris.

Pengertian variabel BERAT dan USIA akan dijelaskan lebih lanjut di Territorial Map.

### Membuat Fungsi Diskriminan 1 dan 2

Fungsi diskriminan -untuk menjawab tujuan ketiga pada kasus ini- dibentuk dengan informasi berikut.

<b>Canonical Discriminant Function Coefficients</b>		
	Function	
	1	2
usia	.093	.129
berat	.079	-.064
(Constant)	-7.169	-.516

Unstandardized coefficients

Fungsi Diskriminan 1 (lihat angka pada kolom FUNCTION 1):

$$Z \text{ Score}_1 = -7,169 + (0,093 \text{ USIA}) + (0,079 \text{ BERAT})$$

Fungsi Diskriminan 2 (lihat angka pada kolom FUNCTION 2):

$$Z \text{ Score}_2 = -0,516 + (0,129 \text{ USIA}) - (0,064 \text{ BERAT})$$

Terkait dengan penjelasan mengenai fungsi dari Discriminant Function, maka hubungan kedua fungsi bisa diperlihatkan pada gambar berikut.

$$Z_1 = -7,1 + (0,09 \text{ U}) + (0,07 \text{ B}) \quad Z_2 = -0,5 + (0,12 \text{ U}) - (0,064 \text{ B})$$



## Penggunaan

Perhatikan sebagian data dari file **diskriminan 3 faktor hasil.sav** yang memuat tiga kolom tambahan sebagai hasil proses diskriminan.

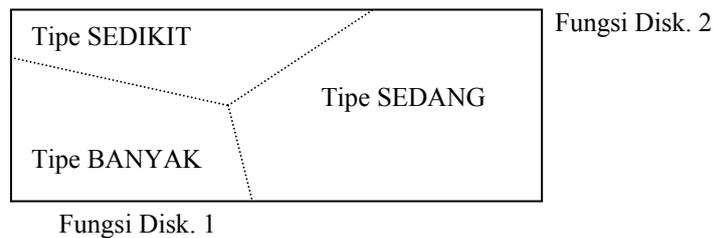
	nama	minum	usia	berat	tinggi	income	jamkerja	olahraga	Dis_1	Dis1_1	Dis2_1
1	RUSDI	0.00	40.00	65.00	154.00	680.00	5.33	3.0	0.00	1.64820	0.48060
2	NINA	0.00	40.00	70.00	157.00	700.00	5.30	3.6	0.00	2.04173	0.16096
3	LANNY	0.00	40.00	60.00	158.00	580.00	5.27	3.5	0.00	1.25486	0.80024
4	CITRA	0.00	39.00	75.00	160.00	600.00	5.33	3.0	0.00	2.34273	-0.28748
5	DINA	0.00	40.00	50.00	159.00	700.00	5.50	3.5	0.00	0.46759	1.43952
6	SISKA	1.00	28.00	62.00	158.00	440.00	5.00	2.2	0.00	0.30161	-0.87317

**Gambar 5.15. Hasil Proses Data (Diskriminan)**

- Untuk kasus konsumen dengan nama RUSDI dengan data:  
USIA= 40 tahun dan BERAT= 65 kilogram  
 $Z\ Score\_1 = -7,169 + (0,093 * 40) + (0,079 * 65) = 1,686$   
NB: perbedaan dengan hasil output pada kolom DIS1\_1 sebesar 1,648 karena adanya pembulatan.  
 $Z\ Score\_2 = -0,516 + (0,129 * 40) - (0,064 * 65) = 0.484$   
NB: perbedaan dengan hasil output pada kolom DIS2\_1 sebesar 0,480 karena adanya pembulatan.
- Penjelasan kolom DIS\_1 pada file **diskriminan 3 faktor hasil.sav**.  
Jika dilihat pada kasus konsumen dengan nama RUSDI, terlihat output (baris 1 kolom DIS\_1) menempatkan RUSDI pada konsumen dengan tipe SEDIKIT. Dari mana kesimpulan tersebut didapat?  
Seperti dilihat pada gambar di atas, fungsi Z score\_1 akan memilah konsumen dengan tipe SEDIKIT atau SEDANG, sedangkan fungsi Z score\_2 akan memilah konsumen dengan tipe SEDANG atau BANYAK. Pada kasus diskriminan 2 faktor, hal tersebut mudah dilakukan, karena hanya ada satu fungsi diskriminan, serta hanya ada dua kode (tipe). Dengan menetapkan  $Z_{CU}$  sebagai cutting score, maka pemasukan input segera menempatkan kasus pada tipe tertentu.  
Namun, jika kasus menggunakan tiga kode yang menghasilkan dua fungsi diskriminan, cara memilah dengan menetapkan  $Z_{CU}$  menjadi sangat kompleks. Untuk itu, pada kasus dengan lebih dari dua kode, penggunaan TERRITORIAL MAP akan lebih efektif untuk menentukan penempatan sebuah data pada kode (tipe) tertentu.

Territorial Map pada dasarnya memetakan (mapping) batas-batas setiap kode berdasar sumbu X (fungsi diskriminan 1) dan sumbu Y (fungsi diskriminan 2), sehingga dengan melihat koordinat sebuah kasus, dengan mudah akan dilihat kasus tersebut ada di teritori (daerah) mana.

Untuk dua fungsi dan tiga tipe konsumen seperti kasus, teritori map secara teori adalah:



#### Catatan

- Gambar di atas hanyalah contoh bagaimana melakukan mapping.
- Jika ada 4 tipe (kode) atau lebih, dimungkinkan terjadi fungsi diskriminan ketiga, yang membuat mapping menjadi kompleks karena adanya tiga sumbu (X, Y dan Z).

#### Mengetahui centroid (rata-rata tiap grup) dengan Tabel dan Grafik

Centroid adalah nilai rata-rata dari fungsi Z score dari setiap objek yang ada pada grup. Dalam hal ini, objek adalah para konsumen, dan karena ada tiga grup, maka akan ada tiga centroid. Jadi, ada satu centroid yang menyatakan nilai rata-rata dari semua data konsumen yang ada di grup SEDIKIT, satu centroid yang menyatakan nilai rata-rata dari semua data konsumen yang ada di grup SEDANG dan satu centroid yang menyatakan nilai rata-rata dari semua data konsumen yang ada di grup BANYAK.

Kegunaan centroid pada dasarnya untuk mengetahui bagaimana penyebaran data dari tiap grup, dan bagaimana kedekatan antar-centroid dari masing-masing grup. Pada satu grup, misal grup SEDIKIT, yang mempunyai titik centroid tertentu, seharusnya data-data konsumen di grup SEDIKIT akan 'bergerombol' di sekitar centroid-nya. Sedang di sisi lain, jika centroid grup SEDIKIT dengan grup BANYAK lebih jauh dibanding jarak centroid grup SEDIKIT dengan grup SEDANG, maka bisa dikatakan karakteristik data grup SEDIKIT dengan karakteristik grup BANYAK lebih berbeda dibandingkan antara grup SEDIKIT dengan grup SEDANG, atau mungkin antara grup BANYAK dengan grup SEDANG.

Pada output, centroid bisa ditampilkan dengan dua cara, yakni:

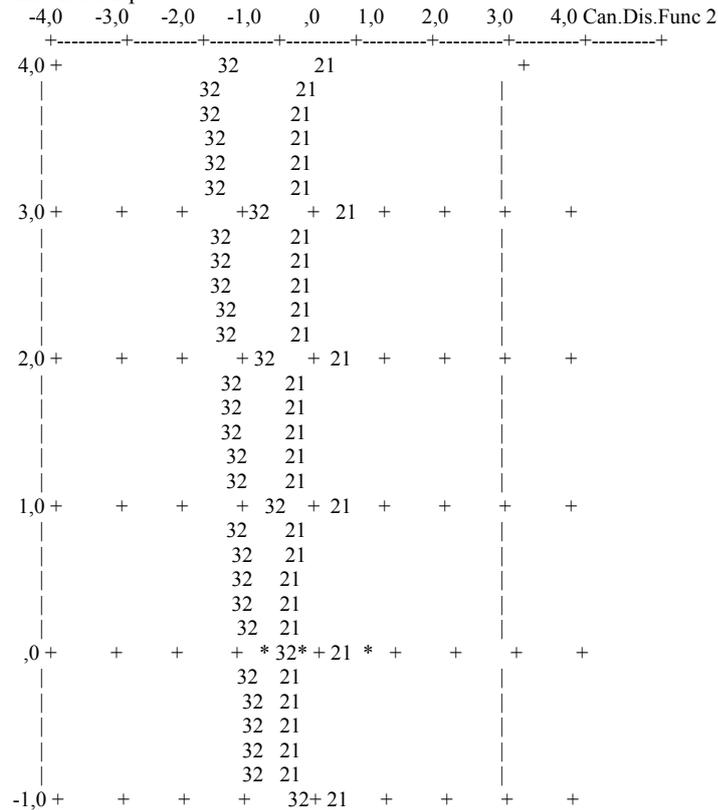
TABEL

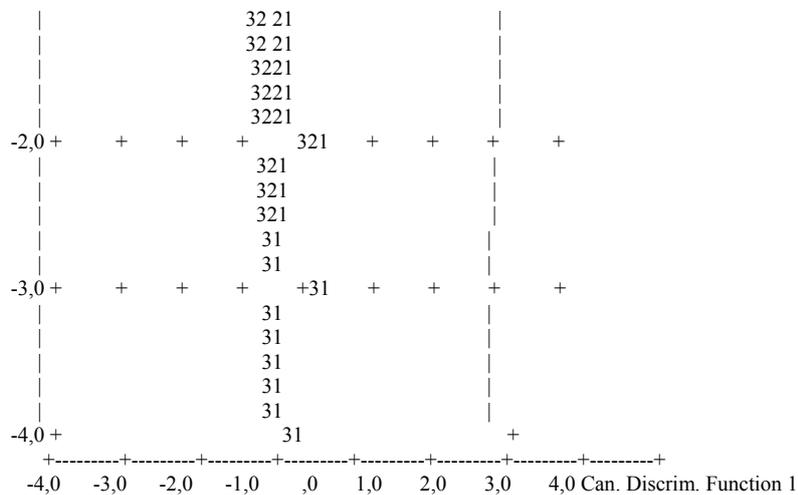
Functions at Group Centroids		
minum	Function	
	1	2
sedikit	.622	-.012
sedang	-.194	.039
banyak	-.590	-.032

Unstandardized canonical discriminant functions evaluated at group means

Dari tabel terlihat titik centroid untuk grup SEDIKIT adalah 0,622 pada function 1 dan -0,01182 pada function 2. Demikian seterusnya untuk centroid dua grup lainnya.

Territorial Map





Symbols used in territorial map

Symbol Group Label

Symbol	Group	Label
1	0	sedikit
2	1	sedang
3	2	banyak
*		Indicates a group centroid

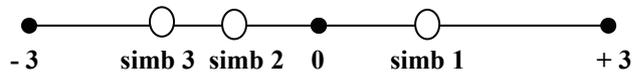
Untuk lebih jelas, angka-angka di atas dibuat grafik, yang disebut TERRITORIAL MAP, dengan function 1 adalah sumbu X, dan function 2 adalah sumbu Y.

Jika dilihat map di atas, terlihat tiga grup konsumen terbagi dari kiri ke kanan. Angka 32 yang tersebar dari kiri atas ke kiri bawah menunjukkan 'pembagian daerah' konsumen grup SEDANG (kode 1) dengan konsumen grup BANYAK (kode 2). Kode 3 dan 2 adalah simbol SPSS yang tentunya berbeda dengan kode input yang adalah 0, 1, dan 2. Lihat tabel di bawah map yang menampilkan simbol dan Grup serta label setiap tipe konsumen, di mana simbol 3 adalah grup BANYAK dan simbol 2 adalah grup SEDANG.

Dengan demikian, anggota grup BANYAK kebanyakan ada di bagian kiri sampai tengah map, sedang anggota grup SEDANG berada di tengah map. Demikian juga terlihat angka 21 yang ada di kanan atas sampai kanan bawah, yang berarti anggota grup SEDIKIT (simbol 1) kebanyakan ada di bagian kanan map.

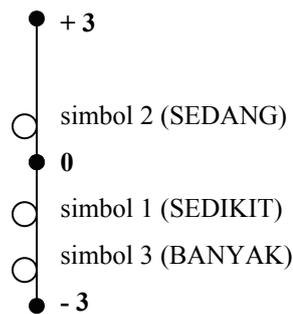
Sedang koordinat ketiga centroid bisa dilihat pada bagian tengah map (perhatikan tanda panah), yang *kebetulan* ketiga centroid terletak berdekatan. Map di atas memang tidak menampilkan sebaran tiap-tiap data berdasar function 1 dan function 2 kecuali centroid masing-masing grup. Untuk melihat sebaran tiap data (kasus), pada kotak dialog CLASSIFY dan pada bagian PLOT, pilih COMBINED-GROUPS dan SEPARATE-GROUPS, selain memilih TERRITORIAL MAP yang menghasilkan map di atas.

Interpretasi dilanjutkan dengan penjelasan STRUCTURE MATRIX sebelumnya, yang menyatakan bahwa variabel BERAT 'masuk' atau lebih dekat ke function 1, dengan koordinat (0,812, -0,583), sedangkan variabel USIA lebih dekat ke function 2, dengan koordinat (0,63, 0,776). Hal ini berarti, jika map di atas dibagi secara *horizontal* saja, dengan batas 0, maka:



Terlihat simbol 3 (grup BANYAK) jelas berbeda dengan simbol 1 (grup SEDIKIT) pada *Berat badannya*. Karena koordinat sumbu X untuk variabel BERAT adalah 0,812 yang *lebih dekat dengan simbol 1*, karena *keduanya positif*, maka makin besar nilai variabel BERAT, makin cenderung masuk ke simbol 1. Atau bisa dikatakan, makin berat seseorang, makin sedikit (simbol 1 atau kode 0) ia minum air mineral. Dan sebaliknya, makin rendah berat badannya, makin banyak (simbol 3 atau kode 2) konsumsinya air minumannya.

Sedang untuk variabel USIA, karena tipisnya perbedaan nilai ketiga centroid jika dilihat dari nilai sumbu yang berartinya, maka bisa dikatakan faktor USIA sebenarnya sulit untuk memilah seseorang masuk pada tipe mana. Walaupun demikian, jika tetap dipetakan koordinatnya, maka:



Karena nilai sumbu Y untuk USIA adalah positif (0,776), yang berarti dekat dengan simbol 2, maka makin tinggi usia seseorang, tipe konsumsi air mineralnya ada pada tingkat sedang. Sedangkan makin muda usia seseorang, maka makin banyak ia mengonsumsi air mineral.

Penilaian atau interpretasi variabel yang terkait dengan grup pada dasarnya dilakukan dengan membagi koordinat dengan sumbu X dan sumbu Y. Setelah itu, jika ada dua variabel terletak pada tempat yang berlawanan (positif dan negatif), maka penafsiran dilakukan dengan arah berlawanan pula.

### **KEGUNAAN TERRITORI MAP**

Kegunaan map di atas adalah untuk mengetahui sebuah data (kasus) apakah masuk pada grup **SEDIKIT**, **SEDANG** ataukah **BANYAK**. Sebagai contoh, untuk kasus pertama, dengan responden bernama **RUSDI**, terlihat nilai function untuknya adalah:

- Function 1(sumbu X pada map): 1,648 (angka pada kolom DIS1\_1 pada file **diskriminan tiga faktor hasil.sav**)
- Function 2(sumbu Y pada map): 0,48 (angka pada kolom DIS2\_1 pada file **diskriminan tiga faktor hasil.sav**)

Dengan demikian, koordinat **RUSDI** adalah (1,648, 0,48), yang jika diplot pada map di atas akan berada pada simbol 1 (grup 0). Dengan demikian, **Rusdi** termasuk konsumen dengan tipe **SEDIKIT**, atau dia hanya meminum sedikit air mineral.

**NB:** hal itu bisa langsung dilihat pada kolom DIS\_1 pada file **diskriminan tiga faktor hasil.sav**.

Demikian seterusnya untuk konsumen yang lain, penempatan seorang konsumen pada salah satu grup sesuai dengan angka function 1 dan function 2.

Jika sampel yang diambil (75 konsumen) bisa dianggap mewakili populasi, atau fungsi diskriminan tersebut benar-benar bisa membedakan perilaku ketiga kelompok, maka fungsi dan map di atas bisa digunakan untuk memprediksi seseorang akan masuk pada tipe yang mana.

Sebagai contoh, jika ada konsumen bernama **MULYADI**, dengan usia 20 tahun dan berat 40 kilogram, maka:

- $Z \text{ Score}_1 = -7,169 + (0,093 * 20) + (0,079 * 40) = -2,149$
- $Z \text{ Score}_2 = -0,516 + (0,129 * 20) - (0,064 * 40) = -0,496$

Berarti posisi **MULYADI** adalah (-2,149, -0,496), yang jika diplot pada map, akan berada pada daerah dengan simbol 3 (kode 2). Dengan demikian, **MULYADI** bisa dikategorikan orang yang **BANYAK** meminum air mineral.

## KOMPOSISI ANGGOTA GRUP

**Prior Probabilities for Groups**

minum	Prior	Cases Used in Analysis	
		Unweighted	Weighted
sedikit	.333	28	28.000
sedang	.333	26	26.000
banyak	.333	21	21.000
Total	1.000	75	75.000

Terlihat konsumen yang menjadi anggota grup **SEDIKIT** adalah terbanyak (26 orang dari total 75), dan paling sedikit adalah anggota grup **BANYAK** (21 dari total 75). Dengan demikian, kebanyakan konsumen termasuk sedikit dalam mengonsumsi air mineral.

## MENILAI KELAYAKAN FUNGSI DISKRIMINAN

Classification Results <sup>b,c</sup>						
MINUM			Predicted Group Membership			Total
			.00 sedikit	1.00 sedang	2.00 banyak	
Original	Count	.00 sedikit	19	4	5	28
		1.00 sedang	7	9	10	26
		2.00 banyak	4	4	13	21
	%	.00 sedikit	67.9	14.3	17.9	100.0
		1.00 sedang	26.9	34.6	38.5	100.0
		2.00 banyak	19.0	19.0	61.9	100.0
Cross-validated <sup>a</sup>	Count	.00 sedikit	18	4	6	28
		1.00 sedang	8	6	12	26
		2.00 banyak	4	4	13	21
	%	.00 sedikit	64.3	14.3	21.4	100.0
		1.00 sedang	30.8	23.1	46.2	100.0
		2.00 banyak	19.0	19.0	61.9	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 54.7% of original grouped cases correctly classified.

c. 49.3% of cross-validated grouped cases correctly classified.

Perhatikan keterangan dengan kode b pada tabel di atas, yang menyatakan bahwa 54,7% dari data telah terklasifikasi dengan benar. Hal ini berarti 54,7% dari 75 data yang diolah telah dimasukkan pada grup yang sesuai dengan data semula. Sebagai contoh, jika pada baris 1 konsumen bernama RUSDI dikategorikan masuk ke grup **SEDIKIT**, maka setelah proses pembuatan fungsi diskriminan dan map teritori, posisi RUSDI tetap dikategori minum **SEDIKIT**.

Jika dilihat dari validasi silang (cross validated), yang ada pada kode c, maka angka tersebut -yang tentunya lebih kecil dari 54,7%- adalah 49,3%. Semakin tinggi nilai validasi, termasuk cross-validated groups tentu semakin bagus, karena semakin tepat fungsi diskriminan membedakan ketiga grup konsumen. Namun demikian, karena kedua angka tersebut bisa dianggap ada di sekitar 50%, maka fungsi diskriminan yang telah dibentuk, juga map teritori yang telah dibuat, *sudah layak* untuk membedakan ketiga grup konsumen tersebut.

### KESIMPULAN

Dari berbagai tabel output dan analisis di atas, kesimpulan yang didapat dengan mengacu pada empat pertanyaan pada awal kasus:

1. ADA perbedaan yang jelas antara konsumen yang mengonsumsi SEDIKIT air mineral, dengan mereka yang termasuk SEDANG atau BANYAK mengonsumsi air mineral.
2. Variabel yang membedakan perilaku mengonsumsi air mineral adalah USIA seseorang dan BERAT badannya.
3. Ada dua fungsi diskriminan yang terbentuk, yakni:  
**Z Score<sub>1</sub> = -7,169 + (0,093 USIA) + (0,079 BERAT)**  
**Z Score<sub>2</sub> = -0,516 + (0,129 USIA) - (0,064 BERAT)**
4. Karena hasil validasi mendekati angka 50%, maka fungsi diskriminan yang terbentuk dianggap tepat untuk menggolongkan seorang konsumen berdasar usia dan berat badannya, ke dalam grup SEDIKIT, SEDANG ataukah BANYAK.

Pada kasus di mana ada dua (*Two-Group Discriminant*) atau tiga grup (termasuk *Multiple Discriminant*), pembuatan grafik atau map masih dimungkinkan untuk menggolongkan kasus tertentu hingga terletak pada grup tertentu. Namun, jika sudah ada empat grup atau lebih, pembuatan map menjadi kompleks, karena melibatkan banyak dimensi. Walaupun demikian, tahapan proses dan interpretasi output tetap berdasar pada berbagai penjelasan yang telah diuraikan di atas.

## 5.4 Logistic Regression

Pada dasarnya *logistic regression* (regresi logistik) sama dengan analisis diskriminan; perbedaan ada pada jenis data dari variabel dependen. Jika pada

analisis diskriminan variabel dependen adalah data rasio, maka pada regresi logistik variabel dependen adalah data nominal. Data nominal di sini lebih khusus adalah data *binary*.

Dengan demikian, tujuan regresi logistik adalah pembuatan sebuah model regresi untuk memprediksi besar variabel dependen yang berupa sebuah variabel *binary* menggunakan data variabel independen yang sudah diketahui besarnya.

### **Catatan**

Variabel Binary adalah data jenis nominal dengan dua kriteria saja, seperti:

1 = Membeli

0 = Tidak membeli

Atau contoh yang lainnya, seperti gagal-sukses, risiko-tidak risiko.

Data yang digunakan pada regresi logistik adalah data kuantitatif. Jika data adalah kualitatif (pada umumnya adalah data jenis kategori atau nominal), maka data tersebut akan diperlakukan sebagai binary variable.

Kasus berikut akan menjelaskan penggunaan regresi logistik.

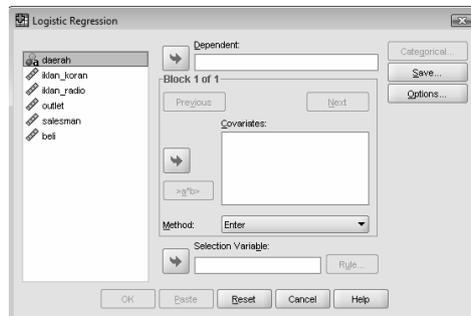
### **KASUS**

Manajer PT DUTA MAKMUR ingin mengetahui apakah perilaku membeli atau tidak membeli produk perusahaan ditentukan oleh kegiatan berikut ini (sebagai variabel bebas):

- Iklan di koran (variabel: iklan\_koran)
- Iklan di Radio (iklan\_radio)
- Jumlah Outlet penjualan di seluruh daerah (outlet)
- Jumlah salesman yang ada (salesman)

### **Pengolahan data dengan SPSS:**

- Buka file **regresi\_binary**.
- Menu **Analyze → Regression → Binary Logistic....** Tampak di layar:



**Gambar 5.16. Kotak Dialog Logistics Regression**

Pengisian:

- **Dependent** atau variabel tergantung. Pilih variabel **beli**.

Variabel 'beli' yang berupa sebuah kategori dengan tanda 1 untuk 'membeli' dan 0 untuk 'tidak membeli'.

- **Covariates**. Pilih variabel **iklan\_koran**, **iklan\_radio**, **outlet** dan **salesman**.
- Klik tombol **OPTIONS** kemudian aktifkan pilihan **Homer-Lemeshow goodness of fit** dan **Iteration history**.

Pilihan *iteration history* akan mencatat perubahan nilai  $-2\text{Log Likelihood}$  ( $-2LL$ ) yang menentukan kelayakan model regresi.

Abaikan bagian yang lain, dan ketik **CONTINUE** untuk kembali ke kotak dialog utama.

- **Method**. Pilih **Enter**.

ABAIKAN BAGIAN YANG LAIN, lalu tekan **OK** untuk proses data.

Output disimpan dengan nama **regresi\_binary1**.

## Output dan Analisis

### Menilai Kelayakan Model Regresi

#### Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	5.161	8	.740

Perhatikan output dari Hosmer and Lemeshow dengan urutan interpretasi.

### Hipotesis

Ho: Tidak ada perbedaan yang nyata antara klasifikasi yang diprediksi dengan klasifikasi yang diamati.

Hi: Ada perbedaan yang nyata antara klasifikasi yang diprediksi dengan klasifikasi yang diamati.

### Dasar Pengambilan Keputusan

Perhatikan nilai *goodness of fit test* yang diukur dengan nilai Chi-Square pada bagian bawah uji Hosmer and Lemeshow.

- Jika probabilitas  $> 0,05$  Ho diterima.
- Jika probabilitas  $< 0,05$  Ho ditolak.

### Keputusan

Karena angka probabilitas adalah 0,7402 yang adalah  $> 0,05$  maka Ho diterima. Hal ini berarti model regresi binary layak dipakai untuk analisis selanjutnya, karena tidak ada perbedaan yang nyata antara klasifikasi yang diprediksi dengan klasifikasi yang diamati.

### Menilai Keseluruhan Model (Overall Model Fit)

Output pada *Beginning Block* (*Block = 0*).

### Block 0: Beginning Block

Iteration History<sup>a,b,c</sup>

Iteration	-2 Log likelihood	Coefficients	
		Constant	
Step 0	1	40.381	.400
	2	40.381	.405
	3	40.381	.405

- a. Constant is included in the model.
- b. Initial -2 Log Likelihood: 40,381
- c. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

Output pada *Block 1*.

**Block 1: Method = Enter**

Iteration History<sup>a,b,c,d</sup>

Iteration	-2 Log likelihood	Coefficients				
		Constant	iklan koran	iklan radio	outlet	salesman
Step 1 1	32.180	3.422	-.145	.134	-.095	.009
2	31.808	4.002	-.179	.186	-.130	.016
3	31.802	4.060	-.184	.194	-.136	.019
4	31.802	4.061	-.184	.194	-.136	.019

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 40,381

d. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Omnibus Tests of Model Coefficients

	Chi-square	df	Sig.
Step 1 Step	8.578	4	.073
Block	8.578	4	.073
Model	8.578	4	.073

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	31.802 <sup>a</sup>	.249	.336

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Perhatikan angka - 2 Log Likelihood, di mana pada awal (Block Number = 0) angka - 2LL adalah 40,381, sedangkan pada Block Number = 1 angka - 2LL *turun* menjadi 31,802. Penurunan ini, di mana Likelihood pada regresi binari mirip dengan pengertian 'sum of squared error' pada model regresi, menunjukkan model regresi yang lebih baik.

**Menguji Koefisien Regresi**

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup> iklan_koran	-.184	.081	5.169	1	.023	.832
iklan_radio	.194	.197	.974	1	.324	1.214
outlet	-.136	.284	.229	1	.632	.873
salesman	.019	.310	.004	1	.950	1.020
Constant	4.061	4.092	.985	1	.321	58.018

a. Variable(s) entered on step 1: iklan\_koran, iklan\_radio, outlet, salesman.

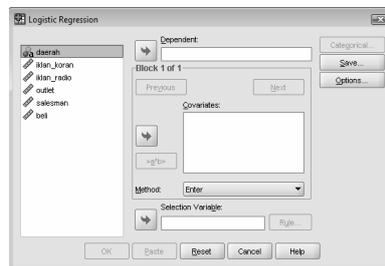
Dengan uji t yang terlihat pada bagian akhir output, terlihat bahwa hanya variabel **iklan\_koran** yang signifikan secara statistik, terlihat dari angka Sig. pada variabel iklan\_koran (0,0230) yang di bawah 0,05. Sedangkan variabel lain mempunyai angka probabilitas (signifikansi) di atas 0,05.

Untuk itu, model regresi *akan diulang lagi*, dengan sekarang *hanya* memasukkan variabel **iklan\_ko** sebagai variabel independen.

## MODEL REGRESI BINARY KE-2

Proses ulangan:

- Buka file **regresi\_binary**.
- Menu **Analyze** → **Regression** → **Binary logistic...**  
Klik tombol **reset** untuk menghapus semua input terdahulu.  
Kemudian lakukan pengisian pada kotak dialog *Binary Logistic*.



**Gambar 5.17. Kotak Dialog Logistics Regression**

Pengisian:

- **Dependent.** Pilih variabel **beli**.
- **Covariates.** Pilih variabel **iklan\_koran**.

Sekarang variabel iklan\_radio, salesman dan outlet tidak dimasukkan lagi.

- Klik tombol **OPTIONS** kemudian aktifkan pilihan **Homer-Lemeshow goodness of fit** dan **Iteration history**.

Abaikan bagian yang lain, dan ketik **CONTINUE** untuk kembali ke kotak dialog utama.

- **Method.** Pilih **Enter**.

Abaikan bagian yang lain dan tekan **OK** untuk proses data.

Output disimpan dengan nama **regresi\_binary2**.

## Output dan Analisis

Menilai Kelayakan Model Regresi

**Hosmer and Lemeshow Test**

Step	Chi-square	df	Sig.
1	8.157	8	.418

Perhatikan nilai goodness of fit test yang diukur dengan nilai Chi-Square pada bagian bawah uji Hosmer and Lemeshow. Terlihat angka probabilitas 0,418 yang adalah  $> 0,05$ . Hal ini berarti model regresi binary layak dipakai untuk analisis selanjutnya, karena tidak ada perbedaan yang nyata antara klasifikasi yang diprediksi dengan klasifikasi yang diamati.

Classification Table<sup>a</sup>

Observed		Predicted		
		perilaku beli		Percentage Correct
		Tidak membeli	Membeli	
Step 1	perilaku beli	8	4	66.7
	Tidak membeli			
	Membeli	4	14	77.8
Overall Percentage				73.3

a. The cut value is ,500

Selain itu, jika dilihat dari OVERALL CLASSIFICATION TABLE, terlihat peningkatan overall hit ratio, yaitu dari 70% pada model regresi 1 (lihat output regresi\_binary1) ke 73,3% pada model regresi 2.

### Menilai Keseluruhan Model (Overall Model Fit)

Blok awal  $\rightarrow$  Block number = 0.

#### Block 0: Beginning Block

Iteration History<sup>a,b,c</sup>

Iteration		-2 Log likelihood	Coefficients
			Constant
Step 0	1	40.381	.400
	2	40.381	.405
	3	40.381	.405

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 40,381

c. Estimation terminated at iteration number 3 because parameter estimates changed by less than ,001.

Block Number = 1.

#### Block 1: Method = Enter

Iteration History<sup>a,b,c,d</sup>

Iteration		-2 Log likelihood	Coefficients	
			Constant	iklan koran
Step 1	1	33.218	4.987	-.154
	2	33.014	5.953	-.184
	3	33.013	6.030	-.186
	4	33.013	6.030	-.186

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 40,381

d. Estimation terminated at iteration number 4 because parameter estimates changed by less than ,001.

Perhatikan angka - 2 Log Likelihood, di mana pada awal (Block Number = 0) angka - 2LL adalah 40,381, sedangkan pada Block Number = 1 angka - 2LL turun menjadi 33,013. Penurunan ini menunjukkan model regresi yang lebih baik.

### Menguji Koefisien Regresi

Variables in the Equation						
	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>						
iklan_koran	-.186	.077	5.772	1	.016	.830
Constant	6.030	2.400	6.312	1	.012	415.837

a. Variable(s) entered on step 1: iklan\_koran.

- Uji t untuk menguji signifikansi konstanta dan setiap variabel independen.

#### Hipotesis

$H_0$  = Koefisien regresi tidak signifikan.

$H_1$  = Koefisien regresi signifikan.

#### Pengambilan Keputusan

Berdasarkan Probabilitas.

- Jika probabilitas > 0,05 maka  $H_0$  diterima.
- Jika probabilitas < 0,05 maka  $H_0$  ditolak.

#### Keputusan

Terlihat bahwa pada kolom **Sig/significance**:

Dengan uji t yang terlihat pada bagian akhir output, konstanta dan variabel **iklan\_koran** adalah signifikan secara statistik (lihat angka Sig. pada konstanta yang 0,012 dan variabel iklan\_koran yang 0,016 di mana keduanya di bawah 0,05). Dengan demikian, sekarang model regresi layak digunakan untuk memprediksi variabel beli.

### Penafsiran dan Prediksi

- Persamaan regresi binary model ke-2 (lihat kolom B pada bagian VARIABLES IN THE EQUATION di bagian akhir output).

$$\text{beli} = 6,0298 - 0,1860 \text{ iklan\_ko}$$

Penafsiran:

Cara menafsir regresi dengan variabel dependen binary adalah dengan pendekatan *probabilitas*. Dan karena tidak ada probabilitas 'negatif' atau 'lebih dari satu', maka jika:

- Angka negatif, dianggap probabilitas 0.
  - Angka positif lebih dari satu, dianggap probabilitas 1
  - Angka positif di antara 0 sampai 1, probabilitas sesuai angka yang tertera.
- Konstanta sebesar 6,0298 menyatakan bahwa jika tidak ada iklan di koran, konsumen tetap membeli Produk Roti DUTA MAKMUR.
- NB: Di sini angka lebih dari satu, maka dianggap probabilitas = 1.
- Koefisien sebesar - 0,1860 menyatakan bahwa setiap perubahan Rp.1 juta Rupiah (satuan iklan\_ko dalam juta rupiah) dalam iklan produk di koran, maka kemungkinan konsumen untuk membeli roti malah berkurang (karena tanda -) sebesar 18,60%!

#### **Contoh**

Jika bulan Juni 2000 perusahaan menganggarkan iklan di radio sebesar Rp.30 juta, maka:

$$\text{beli} = 6,0298 - [ 0,1860 ( 30 ) ] = 0,4498$$

Arti: probabilitas konsumen membeli roti dengan iklan di radio Rp.30 juta sebulan adalah 44,98%. Karena tanda negatif pada koefisien, maka semakin tinggi anggaran iklan di radio, semakin turun probabilitas konsumen untuk membeli roti.

Tafsiran seperti ini sebenarnya secara logika agak 'aneh', karena seharusnya semakin banyak diiklankan, harusnya penjualan roti semakin bertambah, atau probabilitas konsumen membeli roti semakin bertambah.

Dengan output seperti itu, sebenarnya perusahaan malah tidak perlu melakukan kegiatan periklanan, karena tanpa iklan pun roti tetap terjual, dan membuat iklan malah ada kemungkinan penjualan menurun!