PROSIDING

SEMINAR NASIONAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA

Kontribusi Matematika dan Pendidikan Matematika dalam Menumbuhkembangkan [illegible] dan Kepercayaan Diri Bangsa

JURUSAN PENDIDIKAN MATEMATIKA
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
2009

ISSN : 1693-0800

PROSIDING
SEMINAR NASIONAL
MATEMATIKA DAN PENDIDIKAN MATEMATIKA
(Kontribusi Matematika dan Pendidikan Matematika dalam Menumbuhkembangkan Kemandirian dan Kepercayaan Diri Bangsa)
Bandung, 19 Desember 2009

Penanggung jawab Umum : Dra. Siti Fatimah, M.Si., Ph.D

Penanggung jawab Khusus : Dra. Hj. Rini Marwati, M.Si

Editor : Dewi Rachmatin, S.Si., M.Si
Ahmad Solihin

Reviewer : Drs. Turmudi, M.Ed., M.Sc., Ph.D
Dra. Siti Fatimah, M.Si., Ph.D
Dr. Rizky Rosjanuardi, M.Si
Dr. H. Tatang Mulyana, M.Pd
Dr. Endang Mulyana M.Pd.
Drs. C. Jacob, M.Pd
Dr. Dadang Juandi, M.Si
Dr. Jarnawi Afgani Dahlan, M.Kes
Dr. Kusnandi, M.Si
Dr. Elah Nurlaelah, M.Si
Dra. Nurjanah, M.Pd
Dra. Entit Puspita, M.Si
Drs. Bambang Avip P. M., M.Si
Dra. Encum Sumiaty, M.Si
Dewi Rachmatin, S.Si., M.Si
Al Jupri, S.Pd., M.Sc

Sekretariat : Jurusan Pendidikan Matematika
FPMIPA UPI
Jl. Dr. Setiabudhi 229 Bandung
Telp/Fax : (022) 2004508
Website: http://matematika.upi.edu

Susunan Panitia

Panitia Pengarah

1) Drs. Turmudi, M.Ed.,M.Sc., Ph.D.
2) Dr. Dadang Juandi, M.Si. (Kaprodi Pendidikan Matematika)
3) Dr. Rizky Rosjanuardi, M.Si. (Kaprodi Matematika)
4) Dra. Siti Fatimah, M.Si., Ph.D.(Ketua IndoMS Wil. Jawa Barat, Banten dan DKI Jakarta)

Panitia Pelaksana

1) Ketua Pelaksana : Dr. Kusnandi, M.Si
2) Wakil Ketua : Dr. Endang Cahya MA., M.Si
3) Sekretaris : Al Jupri S.Pd., M.Sc
4) Bendahara : Fitriani Agustina, S.Si., M.Si & Dra. Entit Puspita, M.Si
5) Kesekretariatan : Dewi Rachmatin, S.Si, M.Si & Dra. Hj. Rini Marwati, M.Si.
6) Humas : Dra. Suhendra, M.Ed &Drs. & Drs. Bambang Avip P.,M.Si.
7) Seksi Acara : Drs. H. Cece Kustiawan, M.Si & Drs. Suhendra, M.Ed
8) Seksi Pameran Media Pembelajaran: Dra. Hj. Ade Rohaeti, M.Pd
9) Seksi Akomodasi : H. Asep Syarif M.Si
10) Seksi Sponsorship : Dr. Dadang Juandi, M.Si & Drs. M. Rahmat.M.Kes
11) Seksi Sarana dan Prasarana : Drs. H. Asep Syarif. M.Si
12) Seksi Konsumsi : Tia Purniati, S.Pd., M.Pd, Dra. Nurjanah, M.Pd
13) Seksi Dokumentasi : Ahsan, Januar
14) Seksi Pengadaan Seminar Kit: Drs. Endang Dedy, M.Si
15) Pembantu Umum : Para Dosen & Mahasiswa Jurusan Pendidikan Matematika

RENCANA KERJA PANITIA SEMINAR NASIONAL MATEMATIKA DAN PENDIDIKAN MATEMATIKA 2009

A. PERSIAPAN AWAL

No.	Kegiatan	Penanggung jawab	Ket
1.	Pembuatan dan Pendistribusian brosur	Al Jupri, S.Pd., M.Sc.	
2.	Penyusunan Proposal	Drs. Bambang Avip Priatna, M.Si Tia Purniati, S.Pd, M.Pd	
3.	Pencarian Sponsor	Dr. Dadang Juandi, M.Si, Drs. Rahmat, M.Kes	
4.	Menghubungi Pembicara Utama	Dra. Siti Fatimah, M.Si, Ph.D Dr. Nanang Priatna, M.Pd	
5.	Kesekretariatan	Al Jupri, S.Pd., M.Sc	

B. PERSIAPAN PELAKSANAAN SEMINAR

No.	Kegiatan	Penanggung jawab	Ket
1.	Pendaftaran	Dewi Rachmatin, S.Si., M.Si & Dra. Hj. Rini Marwati, M.Si	
2.	Review Makalah	Dra. Encum Sumiaty, M.Si & Dr. Elah Nurlaelah, M.Si	
3.	- Pembuatan Prosiding - Jurnal Terakreditasi	Dr. Endang Cahya MA., M.Si Dra. Siti Fatimah,	

		M.Si, Ph.D	
4.	Pengadaan Seminar Kit	Drs. Endang Dedi, M.Si & Drs. Suhendra, M.Ed	
5.	Pengadaan Sertifikat	Aljupri, S.Pd., M.Sc	
6.	Pengadaan Media Pembelajaran	Dra. Hj. Ade Rohaeti, M.Pd Dra. Nurjanah, M.Pd	
7.	Pengadaan Ruangan dan Perlengkapannya	Drs. H. Asep Syarif, M.Si	
8.	Pembuatan Spanduk	-	
9.	Kehumasan	Drs. H. Asep Syarif, M.Si & Drs. Suhendra, M.Ed	

C. PELAKSANAAN SEMINAR

No.	Kegiatan	Penanggung jawab	Ket
1.	Acara Pelaksanaan Seminar/Persidangan	Drs. H. Cece Kustiawan, M.Si, Drs. Suhendra, M.Ed	
2.	Penjemputan Pembicara Utama	Dr. Nanang Priatna, M.Pd	
3.	Regristasi dan Seminar Kit	Ririn Sispianti, S.Si, Kartika Yulianti, M.Si	
4.	Persiapan Ruangan dan Perlengkapannya	Drs. H. Asep Syarif, M.Si	
5.	Pameran Media Pemb.	Dra. Hj. Ade Rohaeti, M.Pd	
6.	Dokumentasi	Pak Sandi	
7.	Konsumsi	Tia Purniati, S.Pd., M.Pd	
8.	Pembuatan Sertifikat	Aljupri, S.Pd, M.Sc	

Daftar Isi

Bidang Matematika Analisis

Bidang Kajian Matematika Komputasi

Bidang Kajian Matematika Aljabar

Bidang Kajian Matematika Terapan

Bidang Kajian Statistika

Penerapan Prosedur Lachenbruch Pada Kasus *Quadratic Discriminant Analysis*

Dewi Rachmatin
Jurusan Pendidikan Matematika FPMIPA UPI
sujanadewi@yahoo.com

Kania Sawitri
Teknik Elektro ITENAS
kania@itenas.ac.id

ABSTRAK

Hasil-hasil penelitian tentang *Linear Discriminant Analysis* (*LDA*) maupun *Quadratic Discriminant Analysis* (*QDA*) kebanyakan menggunakan metode *Apparent Error Rate* (*APER*) dalam mengevaluasi aturan pengelompokkan dalam Analisis Diskriminan. Metode *APER* ini mempunyai kelebihan yaitu mudah dihitung, tetapi sayangnya cenderung menaksir terlalu rendah *Actual Error Rate* (*AER*), kecuali jika ukuran sampel populasi-populasi yang akan dikelompokkan sangat besar. Oleh karena itu pada penelitian ini diterapkan suatu metode yang disebut Prosedur Lachenbruch, untuk mengatasi hal tersebut. Pada prosedur ini sampel dibagi menjadi dua bagian yaitu sampel yang digunakan untuk membentuk aturan pengelompokkan (*training sample*) dan sampel yang digunakan untuk mengevaluasi hasil pengelompokkan (*validating sample*).

Prosedur Lachenbruch ini diterapkan pada data dua spesies lalat pengigit (*biting fly*) dengan genus *Leptoconos*, yang sama secara morpologi dan selama beberapa tahun kedua spesies ini dianggap sama. Hasil analisis *QDA* terhadap data ini menunjukkan bahwa kedua spesies ternyata berbeda.

Setelah diterapkan prosedur Lachenbruch's pada data *biting fly*, diperoleh hasil sebagai berikut $APER = 4/70$ dan $\hat{E}(AER) = 10/70$ (ekspektasi *Actual Error Rate*). Dapat dilihat bahwa nilai $APER < \hat{E}(AER)$ atau nilai *APER* menaksir terlalu rendah *AER*. Akan tetapi menurut Johnson (1982), kedua nilai *APER* dan $\hat{E}(\mathrm{AER})$ ini tidak akan jauh berbeda jika kedua ukuran sampel sangat besar. Walaupun demikian nilai ekspektasi *AER* sebesar 10/70 ini lebih realistis (Rencer, 2002).

Kata Kunci : *Quadratic Discriminant Analysis* (*QDA*), Prosedur Lachenbruch, *Apparent Error Rate* (*APER*) dan *Actual Error Rate* (*AER*).

1. Pendahuluan

Analisis diskriminan adalah teknik multivariat yang berkenaan dengan pemisahan kumpulan objek-objek yang berbeda dan mengalokasikan suatu objek yang baru ke dalam kelompok yang ada. Pengelompokkan objek-objek yang baru tersebut berdasarkan aturan tertentu yang dibuat sebelumnya berdasarkan kumpulan objek-objek yang sudah ada.

Analisis diskriminan secara luas digunakan juga dalam berbagai bidang, seperti bidang kesehatan, ekonomi, psikologi, sosial dan lain-lain. Salah satu contoh penggunaan analisis diskriminan antara lain di bidang medis, misalnya seorang dokter dihadapkan dengan pengambilan suatu keputusan yang sulit, antara mengambil keputusan melakukan pembedahan atau tidak melakukan pembedahan untuk penyakit tertentu seperti kanker. Pengelompokkan pasien (antara tindakan operasi atau tidak) hanya dibuat setelah penilaian-penilaian klinis terhadap fisik secara menyeluruh, dan dokter mempertimbangkan hasil penilaian pra-operasi tersebut untuk melakukan tindakan yang tepat bagi pasiennya (MacLachan, 1992).

Pada penelitian yang telah dilakukan oleh penulis, analisis diskriminan diterapkan terhadap masalah yang terjadi di bidang biologi, yaitu dua spesies lalat penggigit (*biting fly*) dengan genus *Leptoconos* sama secara morpologi. Selama beberapa tahun kedua spesies ini dianggap sama.

Hasil-hasil penelitian tentang *Linear Discriminant Analysis* (*LDA*) maupun *Quadratic Discriminant Analysis* (*QDA*) kebanyakan menggunakan metode *Apparent Error Rate* (*APER*) dalam mengevaluasi aturan pengelompokkan dalam Analisis Diskriminan. Metode *APER* ini mempunyai kelebihan yaitu mudah dihitung, tetapi sayangnya cenderung menaksir terlalu rendah *Actual Error Rate* (*AER*), kecuali jika ukuran sampel populasi-populasi yang akan dikelompokkan sangat besar. Oleh karena itu, pada penelitian ini yang menjadi rumusan masalahnya adalah Bagaimana hasil penerapan prosedur Lachenbruch dengan metode AER pada data studi kasus *biting fly* dibandingkan dengan hasil metode APER?

Tujuan Penelitian ini adalah membandingkan hasil penerapan prosedur Lachenbruch dengan metode AER dengan metode APER.

Pengolahan data untuk analisis diskriminan yang dilakukan pada penelitian ini menggunakan software MINITAB versi 13 dan SPLUS 2000. Untuk pengelompokan dengan metode *APER* dapat digunakan kedua software tersebut, tetapi untuk pengelompokan dengan menerapkan prosedur Lachenbruch (metode *AER*) digunakan software SPLUS, dan hasil pengelompokan untuk kedua metode *APER* dan *AER* dibandingkan.

2. Teori Dasar

Ada beberapa kasus analisis diskriminan yang diketahui, di antaranya :

1. Analisis Diskriminan Linier (*Linear Discriminant Analysis/LDA*)
   Analisis diskriminan linier digunakan jika data berdistribusi normal multivariat dan setiap kelompoknya memiliki matriks varians kovarians yang sama.
2. Analisis Diskriminan Kuadratik (*Quadratic Discriminant Analysis/QDA*)
   Analisis diskriminan kuadratik digunakan jika data berdistribusi normal multivariat tetapi matriks varians kovariansnya tidak sama dalam setiap kelompoknya.
3. Analisis Diskriminan Fisher (*Fisher Discriminant Analysis/FDA*)
   Analisis diskriminan Fisher digunakan jika data tidak berdistribusi normal multivariat tetapi matriks varians kovariansnya sama dalam setiap kelompoknya.
4. Analisis Diskriminan Nonparametrik (*Nonparametric Discriminant Analysis/NDA*)
   Analisis diskriminan nonparametrik digunakan jika data tidak berdistribusi normal multivariat tidak sama dalam setiap kelompoknya.

Aturan pengelompokan yang akan diterapkan pada penelitian ini, diuraikan secara lengkap, baik untuk *QDA* maupun untuk *LDA*. Sedangkan untuk *FDA* dan *NDA* tidak akan dibahas oleh penulis. Perhatikan, uraian berikut mengenai penentuan aturan pengelompokkan yang diuraikan oleh Johnson (1982).

Misalkan ada sebuah data sampel multivariat, dari data tersebut akan dilihat apakah ada pengelompokan atau tidak. Misalkan $f_i(\mathbf{x})$ adalah fungsi kepadatan peluang yang berhubungan dengan populasi π_i, $i = 1, 2, \cdots, g$. Jika distribusi data sampel tersebut normal multivariat, maka $f_i(\mathbf{x})$ adalah fungsi kepadatan peluang untuk distribusi normal multivariat dengan vektor mean $\boldsymbol{\mu}_i$ dan matriks kovarians $\boldsymbol{\Sigma}_i$, yakni

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \mid \boldsymbol{\Sigma}_i \mid^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\right\}, i = 1, 2, \cdots, g \qquad (2.1)$$

Misalkan p_i adalah peluang prior dari populasi π_i, $i = 1, 2, \cdots, g$ dan $c(k \mid i)$ adalah resiko salah mengelompokkan suatu data anggota populasi π_k, padahal kenyataannya data ini berasal dari π_i untuk $k, i = 1, 2, \cdots, g$. Misalkan R_k adalah himpunan semua $\mathbf{x}$ yang dikelompokkan sebagai π_k dan

$$P(\ k \mid i\ ) = P(\text{ mengelompokkan suatu data sebagai } \pi_k \mid \pi_i) = \int_{R_k} f_i(\mathbf{x})\ d\mathbf{x}$$

untuk $k, i = 1, 2, \cdots, g$ dengan $P(\ i \mid i\ ) = 1 - \sum_{\substack{\ell=1 \\ \ell \neq i}}^{g} P(\ell \mid i)$.

Nilai harapan bersyarat dari salah mengelompokkan suatu data $\mathbf{x}$ yang berasal dari π_1 ke dalam π_2, atau π_3, ... , atau π_g adalah

$$\begin{aligned} \text{ECM}(1) &= P(2 \mid 1)c(2 \mid 1) + P(3 \mid 1)c(3 \mid 1) + \cdots + P(g \mid 1)c(g \mid 1) \\ &= \sum_{\ell=2}^{g} P(\ell \mid 1)c(\ell \mid 1) \end{aligned}$$

(2.2)

Dengan cara yang sama dapat diperoleh nilai harapan bersyarat dari salah mengelompokkan $\text{ECM}(2), \cdots, \text{ECM}(g)$. Kalikan setiap ECM bersyarat dengan peluang priornya dan dijumlahkan menghasilkan ECM total :

$$\text{ECM} = p_1\ \text{ECM}(1) + p_2\ \text{ECM}(2) + \cdots + p_g\ \text{ECM}(g)$$

$$\begin{aligned} &= p_1\left(\sum_{\ell=2}^{g} P(\ell \mid 1)c(\ell \mid 1)\right) + p_2\left(\sum_{\substack{\ell=1 \\ \ell\neq 2}}^{g} P(\ell \mid 2)c(\ell \mid 2)\right) \\ &\qquad + \cdots + p_g\left(\sum_{\ell=1}^{g-1} P(\ell \mid g)c(\ell \mid g)\right) \qquad (2.3) \\ &= \sum_{i=1}^{g} p_i\left(\sum_{\substack{\ell=2 \\ \ell\neq i}}^{g} P(\ell \mid i)c(\ell \mid i)\right) \end{aligned}$$

Daerah pengelompokkan yang meminimumkan ECM total diperoleh dengan mengalokasikan $\mathbf{x}$ ke dalam π_k, dengan $\sum_{\substack{i=1 \\ i\neq k}}^{g} p_i f_i(\mathbf{x})\ c(k \mid i)$ yang terkecil (Anderson, 1958). Andaikan resiko salah mengelompokkan sama (tanpa kehilangan perumuman dapat dimisalkan nilainya sama dengan 1), sehingga menggunakan hasil tersebut, dapat dikelompokkan $\mathbf{x}$ ke dalam populasi π_k, $k = 1, 2, \cdots, g$, di mana

$$\sum_{\substack{i=1 \\ i \neq k}}^{g} p_i \; f_i(\mathbf{x}) \tag{2.4}$$

terkecil. Jadi persamaan (2.4) menjadi yang terkecil ketika suku $p_k \; f_k(\mathbf{x})$ diabaikan, $p_k \; f_k(\mathbf{x})$ ini yang terbesar. Akibatnya jika resiko salah pengelompokan nilainya sama, resiko harapan minimum dari aturan salah pengelompokan mempunyai bentuk yang lebih sederhana, yaitu:

Kelompokkan $\mathbf{x}$ ke dalam π_k,

$$\text{jika } p_k \; f_k(\mathbf{x}) > p_i \; f_i(\mathbf{x}) \text{ untuk setiap } i \neq k \tag{2.5}$$

atau ekuivalen dengan

Kelompokkan $\mathbf{x}$ ke dalam π_k

$$\text{jika } \ln p_k \; f_k(\mathbf{x}) > \ln p_i \; f_i(\mathbf{x}) \text{ untuk setiap } i \neq k \tag{2.6}$$

Aturan tersebut mempunyai tiga komponen, yakni: peluang prior, resiko salah pengelompokan, dan fungsi kepadatan peluang. Ketiga komponen ini harus ditaksir atau dikhususkan sebelum aturan diterapkan.

Kasus khusus penting terjadi jika

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \, |\boldsymbol{\Sigma}_i|^{1/2}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^t \, \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \right\}, \quad i = 1, 2, \cdots, g \tag{2.7}$$

adalah fungsi kepadatan normal multivariat dengan vektor-vektor mean $\boldsymbol{\mu}_i$ dan matriks kovarians $\boldsymbol{\Sigma}_i$. Lebih lanjut $c(i|i) = 0$, $c(k|i) = 1$, $k \neq i$ (atau secara ekuivalen, harga salah pengelompokan semua sama), persamaan (2.6) menjadi:

Jadi menurut aturan (2.6), jika

$$\begin{aligned} \ln p_k \; f_k(\mathbf{x}) &= \ln p_k - \left(\tfrac{p}{2}\right) \ln(2\pi) - \tfrac{1}{2} \ln |\boldsymbol{\Sigma}_k| - \tfrac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_k)^t \, \boldsymbol{\Sigma}_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k) \\ &= \underset{i}{\text{maks}} \; \ln p_i \; f_i(\mathbf{x}) \end{aligned} \tag{2.8}$$

Konstanta $\left(\frac{p}{2}\right) \ln(2\pi)$ dapat diabaikan dalam persamaan (2.8) karena ini sama untuk semua populasi. Kemudian didefinisikan skor diskriminan kuadratik untuk populasi ke-i menjadi:

$$d_i^Q(\mathbf{x}) = -\tfrac{1}{2} \ln |\boldsymbol{\Sigma}_i| - \tfrac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)^t \, \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) + \ln p_i \qquad i = 1, 2, \cdots, g \tag{2.9}$$

Skor kuadratik, $d_i^Q(\mathbf{x})$, tersusun dari kontribusi variansi $|\boldsymbol{\Sigma}_i|$, peluang prior p_i, dan jarak kuadrat dari $\mathbf{x}$ ke mean populasi $\boldsymbol{\mu}_i$. Menggunakan skor diskriminan aturan pengelompokan dari persamaan (2.8) menjadi sebagai berikut:

Kelompokkan $\mathbf{x}$ ke π_k jika

$$\text{Skor kuadratik } d_k^Q(\mathbf{x}) = \text{terbesar dari } d_1^Q(\mathbf{x}), d_2^Q(\mathbf{x}), \cdots, d_g^Q(\mathbf{x}) \tag{2.10}$$

di mana $d_i^Q(\mathbf{x})$ diberikan oleh persamaan (2.9), $i = 1, 2, \cdots, g$.

Dalam prakteknya, $\boldsymbol{\mu}_i$ dan $\boldsymbol{\Sigma}_i$ tidak diketahui, tetapi data sampel yang dicobakan dikelompokkan secara benar tersedia untuk pengkonstruksian taksiran $\boldsymbol{\mu}_i$ dan $\boldsymbol{\Sigma}_i$, kuantitas sampel yang relevan untuk populasi π_i adalah

$\overline{\mathbf{x}}_i$: vektor mean sampel

$\mathbf{S}_i$: matriks kovarians sampel dan

n_i : ukuran sampel

Taksiran dari skor diskriminan kuadratik $\hat{d}_i^Q(\mathbf{x})$ adalah

$$\hat{d}_i^Q(\mathbf{x}) = -\tfrac{1}{2}\ln|\mathbf{S}_i| - \tfrac{1}{2}(\mathbf{x}-\overline{\mathbf{x}}_i)^t\, \mathbf{S}_i^{-1}\,(\mathbf{x}-\overline{\mathbf{x}}_i) + \ln p_i \tag{2.11}$$

dan aturan pengelompokan didasarkan pada sampel adalah sebagai berikut:
Kelompokkan $\mathbf{x}$ ke π_k jika

$$\text{Skor kuadratik } \hat{d}_k^Q(\mathbf{x}) = \text{terbesar dari } \hat{d}_1^Q(\mathbf{x}), \hat{d}_2^Q(\mathbf{x}), \cdots, \hat{d}_g^Q(\mathbf{x}) \tag{2.12}$$

dimana $\hat{d}_i^Q(\mathbf{x})$ diberikan oleh persamaan (2.11), $i = 1, 2, \cdots, g$.

Penyederhanaan dimungkinkan jika matriks kovarians populasi, $\boldsymbol{\Sigma}_i$, adalah sama. Jika $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$, untuk $i = 1, 2, \cdots, g$, skor diskriminan pada persamaan (2.9) menjadi

$$d_i^Q(\mathbf{x}) = -\tfrac{1}{2}\ln|\boldsymbol{\Sigma}| - \tfrac{1}{2}\mathbf{x}^t\,\boldsymbol{\Sigma}_i^{-1}\,\mathbf{x} + \boldsymbol{\mu}_i{}^t\,\boldsymbol{\Sigma}_i^{-1}\,\mathbf{x} - \tfrac{1}{2}\boldsymbol{\mu}_i^t\,\boldsymbol{\Sigma}_i^{-1}\,\boldsymbol{\mu}_i + \ln p_i$$

Dua suku pertama sama pada persamaan di atas dapat diabaikan untuk pengelompokan. Sisanya terdiri dari konstanta $c_i = \ln p_i - \tfrac{1}{2}\boldsymbol{\mu}_i^t\,\boldsymbol{\Sigma}_i^{-1}\,\boldsymbol{\mu}_i$ dan kombinasi linear dari komponen-komponen x. Definisikan skor diskriminan linear

$$d_i(\mathbf{x}) = \boldsymbol{\mu}_i{}^t\,\boldsymbol{\Sigma}_i^{-1}\,\mathbf{x} - \tfrac{1}{2}\boldsymbol{\mu}_i^t\,\boldsymbol{\Sigma}_i^{-1}\,\boldsymbol{\mu}_i + \ln p_i \tag{2.13}$$

sehingga diperoleh bentuk berikut untuk aturan pengelompokan.

Taksiran $\hat{d}_i(\mathbf{x})$ dari skor diskriminan linear $d_i(\mathbf{x})$ didasarkan pada taksiran matriks kovarians gabungan (*pooled estimate*) dari Σ, yaitu

$$\mathbf{S}_{pooled} = \frac{(n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2 + \cdots + (n_g-1)\mathbf{S}_g}{n_1+n_2+\cdots+n_g-g} \tag{2.14}$$

dan

$$\hat{d}_i(\mathbf{x}) = \overline{\mathbf{x}}_i{}^t\,\mathbf{S}_{pooled}^{-1}\,\mathbf{x} - \tfrac{1}{2}\overline{\mathbf{x}}_i^t\,\mathbf{S}_{pooled}^{-1}\,\mathbf{x}_i + \ln p_i \tag{2.15}$$

Akibatnya, diperoleh aturan sebagai berikut :
Kelompokkan $\mathbf{x}$ ke π_k jika

$$\text{Skor diskriminan linear } \hat{d}_k(\mathbf{x}) = \text{terbesar dari } \hat{d}_1(\mathbf{x}), \hat{d}_2(\mathbf{x}), \cdots, \hat{d}_g(\mathbf{x}) \tag{2.16}$$

di mana $\hat{d}_i(\mathbf{x})$ diberikan oleh persamaan (2.15), $i = 1, 2, \cdots, g$.

Aturan pengelompokan yang serupa untuk kasus matriks kokovarians populasi sama dapat diperoleh dari persamaan (2.9) dengan mengabaikan suku konstanta, $-\tfrac{1}{2}\ln|\boldsymbol{\Sigma}|$. Hasil yang diperoleh kemudian diinterpretasikan dalam jarak kuadrat $\mathbf{x}$ dengan vektor mean sampel $\overline{\mathbf{x}}_i$, sebagai berikut:

$$D_i^2(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}})^t\ \mathbf{S}_{pooled}^{-1}\ (\mathbf{x} - \bar{\mathbf{x}}) \tag{2.17}$$

Sehingga aturan pengelompokan menjadi:

Tetapkan $\mathbf{x}$ ke dalam populasi π_i untuk $-\frac{1}{2} D_i^2(\mathbf{x}) + \ln p_i$ terbesar.
(2.18)

Dapat dilihat bahwa aturan ini menetapkan $\mathbf{x}$ ke dalam populasi yang terdekat. Jika peluang prior tidak diketahui, prosedur yang biasa dimisalkan $p_1 = p_2 = \cdots = p_g = \frac{1}{g}$. Suatu pengamatan kemudian dikelompokkan ke dalam populasi yang terdekat.

Untuk mengelompokkan data sampel yang terdiri dari 2 populasi, fungsi pengelompokan sampel secara prinsip dievaluasi oleh nilai "*Actual Error Rate* (*AER*)",

$$\text{AER} = p_1 \int_{\hat{R}_2} f_1(\mathbf{x})\, d\mathbf{x} + p_2 \int_{\hat{R}_1} f_2(\mathbf{x})\, d\mathbf{x} \tag{2.19}$$

di mana $\hat{R}_1$ dan $\hat{R}_2$ menyatakan masing-masing daerah pengelompokan yang ditentukan oleh ukuran sampel n_1 dan n_2. Untuk kasus kedua matriks kovarians sama daerah $\hat{R}_1$ dan $\hat{R}_2$ adalah sebagai berikut:

$$\hat{R}_1 :\ (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t\ \mathbf{S}_{\text{pooled}}^{-1}\ \mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t\ \mathbf{S}_{\text{pooled}}^{-1}\ (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \geq \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

$$\hat{R}_2 :\ (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t\ \mathbf{S}_{\text{pooled}}^{-1}\ \mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^t\ \mathbf{S}_{\text{pooled}}^{-1}\ (\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) < \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

atau

$$R_1 :\ -\tfrac{1}{2}\mathbf{x}^t\left(\Sigma_1^{-1} - \Sigma_2^{-1}\right)\mathbf{x} - \left(\boldsymbol{\mu}_1^t\ \Sigma_1^{-1} - \boldsymbol{\mu}_2^t\ \Sigma_2^{-1}\right)\mathbf{x} - \text{k} \geq \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

$$R_2 :\ -\tfrac{1}{2}\mathbf{x}^t\left(\Sigma_1^{-1} - \Sigma_2^{-1}\right)\mathbf{x} - \left(\boldsymbol{\mu}_1^t\ \Sigma_1^{-1} - \boldsymbol{\mu}_2^t\ \Sigma_2^{-1}\right)\mathbf{x} - \text{k} < \ln\left[\left(\frac{c(1|2)}{c(2|1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

(2.20)

di mana

$$k = \tfrac{1}{2}\ln\left(\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_2|}\right) - \tfrac{1}{2}\left(\boldsymbol{\mu}_1^t\ \boldsymbol{\Sigma}_1^{-1}\ \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^t\ \boldsymbol{\Sigma}_2^{-1}\ \boldsymbol{\mu}_2\right) \tag{2.21}$$

Ada ukuran yang tidak bergantung pada distribusi populasi induk dan dapat dihitung untuk sebarang aturan pengelompokan. Ukuran ini disebut "*Apparent Error Rate (APER)*", yang didefinisikan sebagai proporsi anggota-anggota dalam sampel yang diujicobakan yang salah dikelompokkan. Rumus APER untuk kasus dua populasi sebagai berikut:

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \tag{2.22}$$

di mana

n_{1M} : jumlah anggota-anggota π_1 yang salah dikelompokkan sebagai anggota π_2

n_{2M} : jumlah anggota-anggota π_2 yang salah dikelompokkan sebagai anggota π_1.

Rumus ini dapat diperumum untuk g populasi.

Kelebihan metode APER adalah mudah dihitung. Sayangnya, cenderung menaksir terlalu rendah AER, kecuali jika ukuran sampel n_1 dan n_2 sangat besar. Pada dasarnya, taksiran ini terjadi karena data digunakan untuk membangun fungsi pengelompokan juga digunakan untuk mengevaluasi itu, Johnson (1982) dan Rencer (2002).

Taksiran *error-rate* dapat dikonstruksi yang lebih baik daripada *apparent-error*, tetap secara relatif mudah dihitung, dan tidak memerlukan asumsi distribusional. Satu prosedur untuk membagi sampel total ke dalam sampel yang diujicobakan dan sampel yang diujikan. Sampel yang diujicobakan digunakan untuk mengkonstruksi fungsi pengelompokan dan sampel yang diujikan digunakan untuk mengevaluasi. *Error-rate* ditentukan oleh proporsi salah pengelompokan dalam sampel yang diujikan. Meskipun metoda ini mengatasi masalah bias dengan penggunaan data yang berbeda untuk membangun dan menilai fungsi pengelompokan, jadi ada dua hal yang harus diperhatikan: Johnson (1982) dan Rencer (2002).

(i) Dibutuhkan sampel yang besar, yang mungkin tidak tersedia.

(ii) Tidak dapat mengevaluasi fungsi pengelompokan, taksiran galat hanya berdasarkan pada ½
sampel yang mungkin hasilnya bervariasi dibandingkan hasil yang diperoleh berdasarkan data keseluruhannya.

Jadi lebih baik menggunakan semua atau hampir semua data untuk mengkonstruksi fungsi pengelompokan untuk meminimumkan variansi dari taksiran error rate.

Holdout method adalah prosedur membagi sampel untuk mengatasi dua kelemahan tersebut. Dalam prosedur *Holdout method* ini semua kecuali satu observasi digunakan untuk mengkonstruksi aturan pengelompokan/fungsi pengelompokan dan aturan ini digunakan untuk mengelompokkan observasi yang dikecualikan. Prosedur inilah yang diterapkan oleh Lachenbruch's. Prosedur Lachenbruch's untuk LDA/QDA adalah sebagai berikut

1. Mulai dengan grup yang pertama (π_1). Keluarkan 1 pengamatan dari π_1 dan bangun fungsi pengelompokan tanpa pengamatan ini, jadi semuanya hanya $n_1 - 1$, n_2 pengamatan.
2. Kelompokkan pengamatan yang dikeluarkan tadi menggunakan LDA atau QDA.
3. Ulangi langkah 1 dan 2 hingga semua pengamatan dari π_1 sudah dikeluarkan semuanya.
4. Ulangi langkah 1 hingga 3 untuk π_2.
5. Misalkan $n_{1M}^{(H)}$ adalah jumlah pengamatan π_1 yang salah dikelompokkan dan $n_{2M}^{(H)}$ adalah jumlah pengamatan π_2 yang salah dikelompokkan.

Taksiran $\hat{p}(2|1)$ dan $\hat{p}(1|2)$ adalah peluang kondisional salah pengelompokan dalam persamaan () dan () kemudian diberikan oleh

$$\hat{p}(2|1) = \frac{n_{1M}^{(H)}}{n_1} \text{ dan } \hat{p}(1|2) = \frac{n_{2M}^{(H)}}{n_2}$$

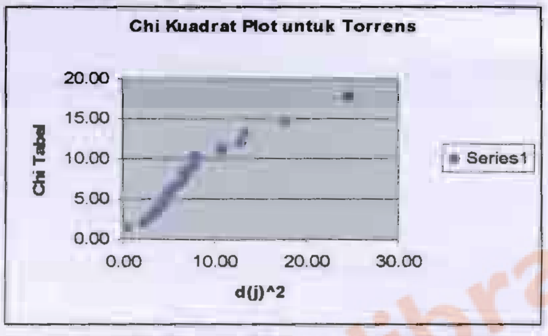
dan proporsi total salah pengelompokan $\left(n_{1M}^{(H)} + n_{2M}^{(H)}\right)/\left(n_1 + n_2\right)$ untuk sampel sedang, taksiran yang hampir takbias dari ekspektasi *error-rate* sebenarnya, dengan

$\hat{E}(\text{AER}) = \dfrac{n_{1M}^{(H)} + n_{2M}^{(H)}}{n_1 + n_2}$ (Johnson, 1982). $\hat{E}(AER)$ ini mereduksi bias dari metode APER (Rencer, 2002), yaitu dengan cara membagi sampel menjadi dua bagian yang disebut terdahulu, yaitu sampel yang dicobakan (*training sample*) dan sampel yang diujikan (*validating sample*).

3. Hasil Analisa Studi Kasus

Sebelum diterapkan analisis diskriminan yang sesuai, terlebih dahulu dilakukan estimasi pendahuluan berupa pengujian kenormalan data *biting fly* dan pengujian kesamaan variansi kedua spesies.

Untuk pengujian kenormalan dilakukan dengan uji khi-kuadrat, dan diperoleh bahwa lebih dari 50% data nilai jarak malahanobisnya kurang dari khi-kuadrat tabel. Sehingga dapat disimpulkan bahwa kedua populasi dari mana kedua spesies ini berasal dapat dianggap berdistribusi normal (perhatikan plot jarak malahanobis terhadap nilai khi-kuadrat tabel mendukung kesimpulan ini).



Selanjutnya diuji kesamaan variansi kedua populasi, dengan menguji hipotesis berikut :

$H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ melawan $H_1 : \mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$.

Statistik uji untuk pengujian kesamaan variansi ini adalah uji rasio log likelihood. Misalkan C adalah matriks kovarians gabungan dengan $C = \dfrac{\sum \mathbf{S}_i}{n-k}$, dan $\mathbf{S}_i$ adalah matriks kovarians sampel dari populasi ke-i. H_0 diuji menggunakan statistic rasio log likelihood yang dimodifikasi :

Misalkan $M = (n-k)\ln|C| - \sum (n_i - 1)\ln|C_i|$ dengan $C_i = \dfrac{\mathbf{S}_i}{n_i - 1}$, dan

$$h = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)}\left(\sum \frac{1}{n_i - 1} - \frac{1}{n-k}\right).$$

Perkalian Mh adalah chi-kuadrat dengan derajat kebebasan p(p+1)(k-1)/2.

Jika semua n_i sama, h direduksi menjadi $h = 1 - \frac{(2p^2 + 3p - 1)(k+1)}{6(p+1)(k-1)}$. Jika nilai khi-kuadrat pengamatan (Mh) lebih besar dari nilai kritis maka H_0 ditolak.

Dari hasil perhitungan diperoleh Mh=64,54 lebih besar dari $\chi^2_{28}(0.05) = 41.34$ sehingga hipotesis nol matriks varians kovarians kedua populasi sama ditolak. Ini berarti benar terdapat perbedaan antara kedua spesies tersebut karena kedua matriks kovarians sampel kedua spesies berbeda.

Untuk menentukan aturan pengelompokan dan mengelompokkan objek-objek yang baru dari kedua spesies ini digunakan *Quadratic Discriminant Analysis* (*QDA*), karena kedua sampel spesies tersebut berdistribusi normal dan matriks kovarians sampel keduanya berbeda.

Dengan aturan pengelompokkan (2.11) dan (2.12) diperoleh hasil untuk metode APER :

Tabel 3.1 Pengelompokan untuk Data Studi Kasus dengan Metoda APER

Grup Aktual	Jumlah Pengamatan	Grup Prediksi	
		1	2
Torrens (1)	35	34	1
Carteri (2)	35	3	32

Jadi kesalahan pengelompokan untuk metode APER adalah

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} = \frac{1+3}{35+35} = \frac{4}{70}.$$

Setelah diterapkan prosedur Lachenbruch pada data *biting fly*, diperoleh nilai harapan kesalahan pengelompokan dengan metode *AER* adalah sebagai berikut:

$$\hat{E}(\text{AER}) = \frac{n_{1M}^{(H)} + n_{2M}^{(H)}}{n_1 + n_2} = \frac{6+4}{70} = \frac{10}{70}.$$

Tabel 3.2 Pengelompokan untuk Data Studi Kasus dengan Metoda AER

	π_1	π_2
π_1	29	6
π_2	4	31

4. Kesimpulan dan Saran

Dari hasil analisa studi kasus, diperoleh hasil bahwa nilai APER < $\hat{E}(\text{AER})$ atau nilai APER menaksir terlalu rendah AER. Akan tetapi menurut Johnson, kedua nilai APER dan $\hat{E}(\text{AER})$ ini tidak akan jauh berbeda jika ukuran sampel kedua populasi sangat besar. Walaupun demikian *error rate* sebesar 10/70 ini lebih realistis (Rencer, 2002). Penelitian ini hanya diterapkan pada salah satu kasus di bidang biologi, untuk bidang-bidang lainnya menjadi kajian penelitian penulis berikutnya, terutama di bidang medis. Sehingga alangkah lebih baiknya jika penelitian ini dikembangkan oleh penulis pada bidang kajian lain, dan untuk kasus *NDA* akan menjadi kajian penulis yang berikutnya pula.

5. Daftar Pustaka

Anderson, T.W. (1958). "*An Introductions to Multivariate Statistical Methods*". New York: John Wiley.

ED231A. (1998). " *Hipothesis Testing: Equality of Population Covariance Matrices*". Online. Tersedia: http://www.gseis.ucla.edu/courses/ed231a1/notes3/covar.html [11 Desember 2008].

Johnson, R.A. & Wichern, D.W. (1982). "*Applied Multivariate Statistical Analysis*". New York: Prentice-Hall.Inc.

McLachlan, G.J. (1992). "*Discriminant Analysis and Statistical Pattern Recognition*". New Jersey: John Wiley & Sons. Inc.

Rencher, A.C. (2002). "*Methods of Multivariate Analysis*". Second Edition. New Jersey: John Wiley & Sons. Inc.

Wikipedia."*Leptoconops torrens*".
Online.Tersedia: http://en.wikipedia.org/wiki/Leptoconops. [31 Oktober 2008].