BAB VIII

UJI HIPOTESIS

Setelah kita membahas mengenai beberapa aturan klasifikasi dan cara-cara klasifikasi, baik klasifikasi dalam dua populasi Normal multivariat maupun klasifikasi dalam lebih dari dua populasi Normal multivariat, maka dalam Bab ini kita akan membahas mengenai uji hipotesisnya.

Analisa Diskriminan juga bertujuan untuk menghasilkan ukuran tingkat perbedaan antara kelompok-kelompok yang terlibat. Dengan kata lain seberapa besarkah perbedaan antara populasi yang pertama dengan populasi yang berikutnya.

Misalkan bahwa fungsi diskriminan untuk dua kelompok π_1 dan π_2 dinyatakan dengan $y = \underset{\sim}{b}'X$, dimana $\underset{\sim}{b} = \mathbf{S}^{-1}(\mu_1 - \mu_2)$. Mean dari masing masing kelompok atau populasi adalah μ_1 dan μ_2, sedangkan matriks kovariansi sampel gabungan adalah $\mathbf{S}$. Kita akan menunjukkan uji hipotesisnya.

8.1. MATRIKS KOVARIANSI SAMA

Misalkan hipotesis yang kita lakukan adalah :

$H_0 : \mu_1 = \mu_2$ lawan

$H_1 : \mu_1 \neq \mu_2$

Statistiki uji yang kita gunakan adalah :

$$F = \frac{n_1 n_2}{n_1 + n_2} \cdot \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} \cdot D^2 ,$$

dengan derajat kebebasan $(p,\ n_1 + n_2 - p - 1)$, dimana :

n_1 = ukuran sampel dari populasi π_1

n_2 = ukuran sampel dari populasi π_2

p = banyak variabel

D^2 = jarak Mahalanobis

$= (\mu_1 - \mu_2)' \; S^{-1} \; (\mu_1 - \mu_2)$

Maksud daripada statistik uji ini adalah variabel F yang ditentukan dengan rumus diatas berdistribusi F, dengan derajat kebebasan $(p, n_1 + n_2 - p - 1)$, apabila H_o benar dan kedua populasi yang bersangkutan mempunyai matriks kovariansi yang sama. Kriteria pengujian adalah H_o ditolak apabila F_{hitung} lebih besar daripada F_{tabel} atau $F_{hitung} > F\alpha ; (p, n_1 + n_2 - p - 1)$, dimana α adalah taraf signifikan yang dipilih, dan $F\alpha ; (p, n_1 + n_2 - p - 1)$ adalah distribusi F dengan derajat kebebasan p, n_1+n_2-p-1.

8.2. MATRIKS KOVARIANSI BEBAS

Misalkan bahwa ukuran sampel I adalah n_1 dari π_1 (populasi berdistribusi Normal p variat) dengan mean sampel $\bar{\underset{\sim}{X}}_1 = (\bar{X}_{11}, \bar{X}_{12}, \ldots\ldots, \bar{X}_{1p})$. Sedangkan ukuran sampel II dari π_2 (populasi berdistribusi Normal p variat) mengenai variabel yang sejenis dengan sampel I, mean sampel II adalah :

$$\bar{\underset{\sim}{X}}_2 = (\bar{X}_{21}, \bar{X}_{22}, \ldots\ldots, \bar{X}_{2p})$$

Vektor mean populasi $\pi_1 = \mu_1$, vektor mean populasi $\pi_2 = \mu_2$

Kita akan mengadakan uji hipotesis

$H_o : \mu_1 = \mu_2$ lawan

$H_1 : \mu_1 \neq \mu_2$

Statistik uji yang digunakan adalah :

$$\chi^2 = (\bar{\underset{\sim}{X}}_1 - \bar{\underset{\sim}{X}}_2)' \; \Sigma^{-1} \; (\bar{\underset{\sim}{X}}_1 - \bar{\underset{\sim}{X}}_2) ,$$

dengan derajat kebebasan p, dimana :

Σ_1 = matriks kovariansi populasi π_1

Σ_2 = matriks kovariansi populasi π_2

$\Sigma = \frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2$

Kriteria keputusan ini adalah H_o ditolak jika χ^2_{hitung} lebih besar daripada χ^2_{tabel} atau $\chi^2 > \chi^2\alpha$; mp, dimana α adalah taraf signifikan yang dipilih, dan $\chi^2\alpha$; mp adalah distribusi χ^2 (chi kuadrat) dengan derajat kebebasan mp.

Pengujian ini lebih baik daripada uji F diatas, karena disini yang diperlukan hanya bahwa matriks kovariansi diketahui, tanpa memperhatikan apakah kedua matriks kovariansi sama atau tidak. Sedang pada statistik uji F diatas, matriks kovariansinya harus sama, maka barulah uji F tersebut benar.

8.3. UJI V BARTLETT

Selanjutnya untuk uji hipotesis dalam kasus Analisa Diskriminan Majemuk, kita akan menggunakan statistik uji V dari Bartlett, serta statistik Lambda dari Wilks. Misalkan akar-akar karakteristik dari matriks $D^{-1}A$ adalah λ_1, λ_2, ... λ_r, dimana $\lambda_1 > \lambda_2 > ...... > \lambda_r > 0$. Maka statistik Lambda dari Wilks dinyatakan dengan LW, yang ditentukan dengan rumus :

$$LW = \frac{1}{(1 + \lambda_1)\ (1 + \lambda_2)\ .....\ (1 + \lambda_r)},$$

Sedangkan statistik V dari Bartlett, ditentukan dengan rumus :

$$VB = -[\ n - 1 - 1/2\ (p + m)\ ]\ \ln LW\ ,$$

dimana :

m = jumlah populasi atau kelompok

p = jumlah variabel peramal

n = ukuran sampel

LW = statistik Lambda dari Wilks

Statistik dari VB ini berdistribusi mendekati distribusi χ^2 dengan derajat kebebasan p(m-1) ; jika H_o benar, yaitu jika terhadap fungsi diskriminan itu, vektor mean untuk semua populasi sama. Statistik VB dapat dituliskan sebagai berikut :

$$VB = [\, n - 1 - 1/2\ (p + m)\,] \sum_{i=1}^{r} \ln\ (1 + \lambda_i)$$

Karena akar-akar karakteristiknya adalah $\lambda_1, \lambda_2, \ldots, \lambda_r$.

Cara yang lain untuk uji hipotesis dalam Analisa Diskriminan Majemuk adalah sebagai berikut. Misalkan bahwa kita hanya akan menghitung satu harga dari akar karakteristiknya, misalnya akar karakteristik ke-j, maka :

$$VB_j = [\, n - 1 - 1/2\ (p + m)\,] \ln\ (1 + \lambda_j)$$

Statistik ini berdistribusi mendekati distribusi χ^2 dengan derajat kebebasan p + m - 2j .

Selanjutnya $VB-VB_1$, $VB-VB_1-VB_2$, $VB-VB_1-VB_2-VB_3$ dan seterusnya disebut Diskriminan Residual, sedangkan distribusinya mendekati distribusi χ^2 dengan derajat kebebasan berturut-turut (p - 1) (m - 2), (p - 2)(m - 3), (p - 3) (m - 4) dan seterusnya. Bentuk umumnya (p - i) (m - i-1) kita nyatakan dengan tanda d(i). Selanjutnya, apabila kita gunakan taraf signifikan, maka banyaknya fungsi diskriminan yang signifikan adalah w apabila terjadi dua kejadian, yaitu :

1. $VB - VB_1 - \ldots\ldots\ldots - VB_{w-1} > \chi^2\alpha$; d(w-1)
2. $VB - VB_1 - \ldots\ldots\ldots - VB_w \leq \chi^2\alpha$; d(w),

dimana α adalah taraf signifikan.