INFERENSI VEKTOR RATA – RATA

Disusun untuk memenuhi salah satu tugas mata kuliah multivariat

Disusun oleh:

Asti Aulia Rahman	(0607196)
Khaerunnisa Mahmudah	(060910)
Lucky Heriyanti Jufri	(0607103)
Risa Nur Vauzyah	(060933)
Syifa Insani	(060116)

PROGRAM STUDI MATEMATIKA

JURUSAN PENDIDIKAN MATEMATIKA

FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS PENDIDIKAN INDONESIA

2009

KATA PENGANTAR

Puji dan syukur kehadirat Allah SWT yang telah melimpahkan rahmat dan karunianya sehingga penyusun dapat menyelesaikan makalah ini dengan baik. Salam dan salawat selalu tercurahkan kepada junjungan kita nabi besar Muhammad SAW.

Pada makalah ini akan dibahas mengenai inferensi vektor rata – rata pada normal multivariat. Penyusun menyadari bahwa dalam makalah ini masih terdapat banyak kekurangan. Penyusun mengharapkan kritik dan saran demi kesempurnaan dalam penyusunan makalah selanjutnya.

Akhir kata semoga makalah ini dapat bermanfaat bagi penyusun dan para pembaca pada umumnya.

Bandung,   Juni 2009

Penyusun

DAFTAR ISI

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Ketika kita menggunakan statistika untuk menguji hipotesis maka muncullah dua macam hipotesis berupa hipotesis penelitian dan hipotesis statistika. Tepatnya hipotesis penelitian kita rumuskan kembali menjadi hipotesis statistika yang sepadan. Hipotesis statistika harus mencerminkan dengan baik maksud dari hipotesis penelitian yang akan diuji.

Pada hakikatnya ada dua jenis hipotesis statistika. Jenis pertama adalah apabila data kita berupa populasi yang kita peroleh melalui sensus. Dengan data populasi, hipotesis statistika cukup berbentuk H. Tidak diperlukan hipotesis H_0. Misalnya dalam hal rerata, hipotesis statistika itu berbentuk H: $\mu_X > 6$. Jika data populasi memiliki rerata di atas 6 maka hipotesis diterima dan jika tidak maka hipotesis ditolak. Karena seluruh populasi sudah dilihat maka keputusan ini menjadi kepastian.

Jenis kedua adalah apabila data kita berupa sampel yang kita peroleh melalui penarikan sampel. Biasanya sampel itu berupa sampel acak, baik dengan cara pengembalian maupun dengan cara tanpa pengembalian. Dengan data sampel, hipotesis statistika menjadi H_0 dan H_1. Misalnya dalam rerata, hipotesis statistika itu berbentuk H_0: $\mu_X = 6$ dan H_1: $\mu_X > 6$. Syaratnya adalah tiadanya pilihan ketiga.

Dalam hal data sampel, sering terjadi bahwa hipotesis penelitian dirumuskan kembali menjadi H_1. Pengujian hipotesis dilakukan melalui penolakan H_0. Selanjutnya dengan syarat tidak ada pilihan ketiga pada hipotesis, maka penolakan H_0 dapat diartikan sebagai penerimaan H_1. Jadi pengujian hipotesis penelitian dilakukan melalui cara tak langsung yakni melalui penolakan H_0 dan melalui tiadanya pilihan ketiga pada hipotesis.

Dalam makalah ini akan dibahas pengujian hipotesis tentang perbedaan antara vektor rata-rata dan vektor konstan. Mirip halnya dengan pengujian hipotesis pada situasi univariat. tentang perbedaan antara rata-rata dan konstan. Pada situasi multivariat juga diperlukan syarat-syarat agar rumus-rumus untuk pengujian hipotesis itu berlaku. Pada pengujian hipotesis untuk univariat disyaratkan bahwa populasi yang bersangkutan berdistribusi normal. Sesuai dengan itu, pada pengujian hipotesis untuk multivariat disyaratkan bahwa populasi yang bersangkutan berdistribusi normal multivariat.

Untuk memperoleh metode utama dalam menentukan inferensi dari sample, kita akan memperluas konsep interval kepercayaan *univariat* menjadi daerah kepercayaan *multivariate*. Berdasarkan penjelasan pada bab sebelumnya, telah dijelaskan inferensi sampel dengan menggunakan $\text{int}\,erval - T^2$ simultan. Namun seringkali kita jumpai interval yang lebih pendek untuk bilangan m yang kecil, yaitu ketika $m = p$. Dalam hal ini, akan lebih mudah untuk menggunakan dan menetapkan interval kepercayaan yang relatif pendek, yang dibutuhkan untuk membuat kesimpulan (*inference*).

Ketika ukuran sampel besar, pengujian hipotesis dan daerah kepercayaan untuk μ dapat dikonstruksi tanpa anggapan normalitas. Untuk jumlah n besar, kita dapat membuat taksiran tentang rata-rata populasi meskipun distribusi awalnya adalah diskrit.

Masalah lain yang timbul adalah ketika beberapa nilai observasi hilang. Pengestimasian terhadap nilai yang hilang perlu dilakukan untuk mempermudah pengolahan dan menemukan statiska cukupnya.

1.2 Permasalahan

1.2.1 Rumusan Masalah

1. Pada dasarnya pengujian hipotesis vektor rata-rata polpulasi multivariat membahas mengenai hubungan antara vektor rata-rata populasi multivariat dengan konsistensitas data. Oleh karena itu rumusan makalah yang dapat diambil adalah apakah suatu vektor rata-rata populasi multivariat akan selalu konsisten dengan data yang dimiliki?
2. Perbedaan pengujian hipotesis dengan menggunakan maksimum likelihood dan hotteling T^2 pada normal multivariate.
3. Menetapkan interval kepercayaan yang lebih pendek dari hotelling T^2, yaitu dengan metode banferroni.
4. Menentukan interval untuk sampel besar
5. Mengetahui cara estimasi dan prediksi dari beberapa observasi yang hilang.

1.2.2 Pembatasan masalah

Dalam makalah ini, masalah yang dibahas akan membahas pengujian hipotesis vektor rata-rata populasi multivariat serta landasan teori yang mendukungnya.

1.3 Tujuan Penulisan

Tujuan dari penulisan makalah ini adalah untuk mengetahui dengan melakukan pengujian hipotesis apakah vektor rata-rata populasi merupakan sebuah nilai *plausible* untuk rata-rata populasi normal. Perbedaan pengujian hipotesis dengan menggunakan maksimum likelihood dan hotteling T^2 pada normal multivariate.Menetapkan interval kepercayaan yang lebih pendek dari hotelling T^2, yaitu dengan metode banferroni. Menentukan interval untuk sampel besar. Mengetahui cara estimasi dan prediksi dari beberapa observasi yang hilang.

1.4 Metode Penulisan

Metode yang digunakan dalam penulisan makalah ini yaitu studi pustaka yang yang dilakukan di perpustakaan dan internet.

1.5 Sistematika Penulisan

Sistematika penulisan makalah ini yaitu :

a. BAB I Pendahuluan terdiri dari latar belakang masalah, rumusan masalah dan pembatasan masalah, tujuan penulisan, metode penulisan dan sistematika penulisan;
b. BAB II Landasan teori yang berisi matriks dispersi, distribusi normal multivariat, dan beberapa distribusi statisitik.
c. BAB III Isi yang membahas mengenai pengujian hipotesis apakah vektor rata-rata populasi merupakan sebuah nilai *plausible* untuk rata-rata populasi normal. Perbedaan pengujian hipotesis dengan menggunakan maksimum likelihood dan hotteling T^2 pada normal multivariate.Menetapkan interval kepercayaan yang lebih pendek dari hotelling T^2, yaitu dengan metode banferroni. Menentukan interval untuk sampel besar. Mengetahui cara estimasi dan prediksi dari beberapa observasi yang hilang.
d. BAB IV Penutup yang berisi kesimpulan dan saran.

BAB II
LANDASAN TEORI

2.1 Matriks Dispersi

Pada situasi univariat, jika variabel acak X mempunyai daerah harga (atau nilai-nilainya adalah) $X_1, X_2, \ldots, X_n$, maka rata-ratanya adalah $\mu_x = \frac{X_1, X_2, \ldots, X_n}{N}$ dan variansnya adalah $\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{N}\left(x_i - \mu_x\right)^2$.

Jika dari nilai-nilai X yang mungkin itu hanya tersedia satu sampel acaknya saja, misalnya $X_1, X_2, \ldots, X_n$, maka rata-rata dan varians yang dapat dihitung adalah rata-rata dan varians sampel saja, yang merupakan taksiran bagi rata-rata dan varians tersebut.

Rata-rata sampel adalah $\overline{X} = \frac{X_1, X_2, \ldots, X_n}{n}$ dan varians sampelnya adalah $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$.

Pada situasi multivariat yang melibatkan p variabel acak $X_1, X_2, \ldots, X_p$; misalkan X_{ij} menyatakan nilai ke-j dari variabel X_i, dimana $1 \le j \le N$.

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N1} & X_{N2} & \cdots & X_{Np} \end{pmatrix}$$

Jika μ_i menyatakan rata-rata dari variabel X_i, maka dapat disusun matriks rata-rata berorde $N \ x \ p$ sesuai dengan X di atas, yaitu

$$\mu \begin{pmatrix} \mu_1 & \mu_2 & \cdots & \mu_p \\ \mu_1 & \mu_2 & \cdots & \mu_p \\ \vdots & \vdots & \ddots & \vdots \\ \mu_1 & \mu_2 & \cdots & \mu_p \end{pmatrix}$$

dimana $\mu_i = \mu_{Xi} = \dfrac{X_{1i}, X_{2i}, \ldots, X_{Ni}}{N}$.

Ukuran yang mirip dengan σ_X^2 adalah Σ yang disebut matriks dispersi atau matriks varians-kovarians, dengan rumus

$$\Sigma = \frac{1}{n}(X-\mu)'(X-\mu)$$

Dapat dihitung:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_p^2 \end{pmatrix}$$

dimana $\sigma_i^2 = \dfrac{1}{N}\sum_{r=1}^{N}(x_{ri}-\mu_i)^2$

$$\sigma_{jk} = \frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N}(x_{sj}-\mu_j)(x_{tk}-\mu_k).$$

Telah kita kenal bahwa σ_i^2 disebut varians dari X_i sedang σ_{jk} disebut kovarians antara X_j dan X_k. Itulah sebabnya maka Σ disebut matriks varians-kovarians dari X.

Seperti yang telah ditunjukkan dalam bab 2, $\Sigma = \dfrac{1}{N}A$, dimana A adalah matriks Jumlah Kuadrat dan Hasil Silang (JKHS) dari X, dan dapat ditunjukkan bahwa

JKHS(X) = A

$$= (X-\mu)'(X-\mu)$$

$$\begin{pmatrix} \Sigma x_1^2 & \Sigma\Sigma x_1 x_2 & \cdots & \Sigma\Sigma x_1 x_p \\ \Sigma\Sigma x_2 x_1 & \Sigma x_2^2 & \cdots & \Sigma\Sigma x_2 x_p \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma\Sigma x_p x_1 & \Sigma\Sigma x_p x_2 & \cdots & \Sigma x_p^2 \end{pmatrix}$$

dimana $\sum x_i^2 = \sum_{r=1}^{N} \left(X_{ri} - \mu_i \right)^2$

dan $\sum \sum x_j x_k = \sum_{t=1}^{N} \sum_{s=1}^{N} \left(X_{sj} - \mu_j \right) \left(X_{tk} - \mu_k \right)$

perlu diingat bahwa $\sigma_{jk} = \rho \sigma_j \rho \sigma_k$,

dimana
ρ = koefisien korelasi antara X_j dan X_k;
σ_j = simpangan baku dari X_j;
σ_k = simpangan baku dari X_k;
σ_{jk} = kovarians antara X_j dan X_k.

Jika nilai-nilai dua variabel tersebut hanya tersedia sampel acak n nilai dari tiap-tiap variabel, maka terdapat matriks data

$$X = \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{pmatrix}$$

Taksiran untuk matriks rata-rata u adalah rata-rata sampel X, yaitu matriks berorde n x p.

$$\bar{X} = \begin{pmatrix} \bar{X}_1 & \bar{X}_2 & \cdots & \bar{X}_p \\ \bar{X}_1 & \bar{X}_2 & \cdots & \bar{X}_p \\ \vdots & \vdots & \ddots & \vdots \\ \bar{X}_1 & \bar{X}_2 & \cdots & \bar{X}_p \end{pmatrix}$$

dimana $\bar{X}_i = \dfrac{X_{1i}, X_{2i}, \ldots, X_{ni}}{n}$

Adapun taksiran untuk matriks dispersi, Σ , adalah matriks dispersi sampel, S ,yaitu matriks berorde p x p berikut ini

$$S = \frac{1}{n-1}\left(X - \bar{X}\right)'\left(X - \bar{X}\right)$$

$$= \begin{pmatrix} \frac{1}{n-1}\Sigma x_1^2 & \frac{1}{n-1}\Sigma\Sigma x_1 x_2 & \cdots & \frac{1}{n-1}\Sigma\Sigma x_1 x_p \\ \frac{1}{n-1}\Sigma\Sigma x_2 x_1 & \frac{1}{n-1}\Sigma x_2^2 & \cdots & \frac{1}{n-1}\Sigma\Sigma x_2 x_p \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n-1}\Sigma\Sigma x_p x_1 & \frac{1}{n-1}\Sigma\Sigma x_p x_2 & \cdots & \frac{1}{n-1}\Sigma x_p^2 \end{pmatrix}$$

$$= \begin{pmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{pmatrix}$$

dimana $\sum x_i^2 = \sum_{r=1}^{n}\left(X_{ri} - \bar{X}_i\right)^2$

$$\sum\sum x_j x_k = \sum_{t=1}^{n}\sum_{s=1}^{n}\left(X_{sj} - \bar{X}_j\right)\left(X_{tk} - \bar{X}_k\right)$$

$s_{ii} = s_i^2$ = varians sampel untuk X_i

$= \frac{1}{n-1}\ \Sigma\ x_i^2$

s_{jk} = kovarians sampel antara X_j dan X_k

$= \frac{1}{n-1}\ \Sigma\Sigma\ x_j\ .\ x_k$

2.2 Distribusi Normal Multivariat

Variabel acak X dikatakan berdistribusi Normal dengan rata-rata = μ, dan varians = τ^2, diamana $\tau > 0$, jika fungsi kepadatan probabilitas dari X tertentu oleh rumus

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}, \text{ untuk } -\infty < X < \infty$$

Grafik dari y = f(X) merupakan kurva atau garis lengkung, yang lazim dikatakan berbentuk lonceng (irisan bentuk lonceng).

Pada situasi mutivariat, terlibat lebih dari satu variabel. Sekelompok variabel $\left(X_1, X_2, \ldots, X_p\right)$ dikatakan berdistribusi normal p-variat dengan vektor rata-rata $\mu = \left(\mu_1, \mu_2, \ldots, \mu_p\right)'$ dan matriks varians-kovarians atau matriks dispersi Σ, jika fungsi kerapatan probabilitas bersama dari p-variabel itu tertentu oleh rumus.

$$f\left(X_1, X_2, \ldots, X_p\right) = \frac{1}{|\Sigma|^{1/2}\sqrt{(2\pi)^p}} e^{-1/2 K}$$

dimana

$$K = \left(X - \mu\right)' \Sigma^{-1} \left(X - \mu\right)$$

$$= \left(X_1 - \mu_1, X_2 - \mu_2, \ldots, X_p - \mu_p\right) \Sigma^{-1} \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{pmatrix}$$

Tampak adanya kemiripan antara rumus fungsi kerapatan probabilitas univariat dan multivariat.

Pada univariat : $|\Sigma|^{1/2} = \left(\sigma^2\right)^{1/2} = \sigma$, diketahui $p = 1$,

sehingga $\sqrt{(2\pi)^p} = \sqrt{2\pi}$, dan

$$K = \left(X - \mu\right)\left(\sigma^2\right)^{-1}\left(X - \mu\right)$$

$$= \left(\frac{X - \mu}{\sigma}\right)^2$$

Khususnya jika p = 2, terdapat

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_1 & \sigma_2^2 \end{pmatrix}$$

$$|\Sigma| = \left(1-\rho^2\right)\sigma_1^2\sigma_2^2;$$

$$\Sigma^{-1} = \frac{1}{\left(1-\rho^2\right)\sigma_1^2\sigma_2^2}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_2\sigma_1 & \sigma_1^2 \end{pmatrix}$$

$$K = \left(X_1-\mu_1, X_2-\mu_2\right)\Sigma^{-1}\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{pmatrix}$$

$$= \left(X_1-\mu_1, X_2-\mu_2\right)\left(\frac{1}{1-\rho^2}\right)\begin{pmatrix} \dfrac{1}{\sigma_1^2} & \dfrac{\rho}{\sigma_1\sigma_2} \\ \dfrac{\rho}{\sigma_2\sigma_1} & \dfrac{1}{\sigma_2^2} \end{pmatrix}\begin{pmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{pmatrix}$$

$$= \left(\frac{1}{1-\rho^2}\right)\left[\left(\frac{X_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{X_2-\mu_2}{\sigma_2}\right)^2 - 2\frac{\left(X_1-\mu_1\right)\left(X_2-\mu_2\right)}{\sigma_1\sigma_2}\right]$$

Fungsi kerapatan probabilitas Normal Bivariat, dan rumusnya adalah

$$f\left(X_1, X_2\right) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}Q}$$

dimana $$Q = \frac{1}{1-\rho^2}\left[\left(\frac{X_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{X_2-\mu_2}{\sigma_2}\right)^2 - 2\frac{\left(X_1-\mu_1\right)\left(X_2-\mu_2\right)}{\sigma_1\sigma_2}\right]$$

ρ = korelasi antara x_1 dan x_2;
μ_i = rata-rata dari X_i;
σ_i = simpangan baku dari X_i;

Grafik dari $z = f\left(X_1, X_2\right)$ merupakan luasan lengkung, mirip permukaan suatu lonceng. Kalau luasan lengkung ini dipotong dengan bidang datar yang sejajar dengan bidang $\left(X_1, X_2\right)$ maka irisannya adalah suatu elips.

Elips itu tertentu oleh suatu persamaan berbentuk Q = k, atau

$$\left(\frac{X_1-\mu_1}{\sigma_1}\right)^2+\left(\frac{X_2-\mu_2}{\sigma_2}\right)^2-2\frac{(X_1-\mu_1)(X_2-\mu_2)}{\sigma_1\sigma_2}=k$$

Elips demikian, untuk harga-harga k yang sesuai, merupakan batas daerah penolakan H_0 pada pengujian hipotesis dalam Analisis Bivariat dan disebut elips kerapatan sama.

2.3 Beberapa Distribusi Statistik

Pada Statistika Univariat sudah dikenal sifat bahwa apabila X berdiatribusi $N\left(\mu,\sigma^2\right)$, yaitu berdistribusi Normal dengan rata-rata = μ dan varians = σ^2, maka rata-rata sampel, yaitu X, berdistribusi $N\left(\mu,\frac{\tau^2}{n}\right)$ jika sampel itu adalah sampel acak sebesar n.

Dengan kata lain $\frac{X-\mu}{\left(\sigma/\sqrt{n}\right)}$ berdistribusi Normal Baku jika syarat-syarat tersebut dipenuhi.

Salah satu sifat yang telah terbukti secara matematis ialah bahwa apabila variabel v berdistribusi Normal Baku, sedang $w=v^2$, maka w berdistribusi χ^2 dengan derajat kebebasan 1. Berhubung dengan itu maka

$$\frac{\left(\bar{X}-\mu\right)^2 n}{\sigma^2} \quad \text{atau} \quad n\left(\bar{X}-\mu\right)\left(\sigma^2\right)^{-1}\left(\bar{X}-\mu\right)$$

berdistribusi χ^2 dengan derajat kebebasan 1 apabila syarat-syarat tersebut di atas terpenuhi.

Pada situasi multivariat terdapat sifat yang mirip dengan sifat tersebut.

Apabila $X_1,X_2,\dots,X_p$ berdistribusi Normal Multivariat $N\left(\mu,\Sigma\right)$, dimana $\mu=\left(\mu_1,\mu_2,\dots,\mu_p\right)$, sedang Σ adalah matriks dispersi, sedang $\bar{X}=\left(\bar{X}_1,\bar{X}_2,\dots,\bar{X}_p\right)'$, menyatakan vektor rata-rata dari sampel acak, dan apabila

$$W = n\left(\bar{X}_1 - \mu_1, \bar{X}_2 - \mu_2, \ldots, \bar{X}_p - \mu_p\right)\Sigma^{-1}\begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{pmatrix}$$

maka W berdistribusi χ^2 dengan derajat kebebasan p: dimana n menyatakan besarnya sampel.

Pada situasi univariat, apabila σ^2 tak diketahui maka distribusi $\bar{X}$ dapat ditinjau dalam hubungannya dengan varians sampel, yaitu bahwa $\dfrac{\bar{X} - \mu}{\left(s/\sqrt{n}\right)}$ berdistribusi t dengan derajat kebebasan $n-1$.

Juga telah dibuktikan bahwa apabila variabel v berdistribusi t dengan derajat kebebasan $n-1$, sedangkan $w = v^2$, maka W berdistribusi F dengan derajat kebebasan $(1, n-1)$. Berhubung dengan itu maka $\dfrac{\left(\bar{X} - \mu\right)n}{S^2}$ atau $n\left(\bar{X} - \mu\right)\left(s^2\right)^{-1}\left(\bar{X} - \mu\right)$ berdistribusi F dengan derajat kebebasan $(1, n-1)$.

Pada situasi multivariat terdapat pula sifat yang mirip dengan itu. Misalkan $\left(X_1, X_2, \ldots, X_p\right)$ berdistribusi denganvektor rata-rata $\mu = \left(\mu_1, \mu_2, \ldots, \mu_p\right)$, sedang $\bar{X} = \left(\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p\right)'$ menyatakan vektor rata-rata dari sampel acak sebesar n, dan apabila $W = n\left(\bar{X}_1 - \mu_1, \bar{X}_2 - \mu_2, \ldots, \bar{X}_p - \mu_p\right)S^{-1}\begin{pmatrix} \bar{X}_1 - \mu_1 \\ \bar{X}_2 - \mu_2 \\ \vdots \\ \bar{X}_p - \mu_p \end{pmatrix}$ maka W berdistribusi *Hotelling* T^2 dengan derajat kebebasan $\left(p, n-p\right)$. Dalam rumus tersebut S adalah matriks dispersi sampel. *Hotelling* telah membuktikan bahwa apabila variabel W berdistribusi T^2, dengan derajat kebebasan $\left(p, n-p\right)$ maka $\dfrac{n-p}{p(n-1)}$ W berdistribusi F dengan derajat kebebasan $\left(p, n-p\right)$.

Sifat-sifat dari distribusi statistik multivariat W tersebut dapat dimanfaatkan untuk menguji signifikansi perbedaan antara vektor rata-rata suatu populasi dan vektor konstan, atau perbedaan antara vektor-vektor rata-rata dua populasi.

Pada situasi univariat tentang selisih rata-rata dari dua sampel acak yang bebas, yaitu $\bar{X}_1 - \bar{X}_2$, diketahui bahwa statistik

$$\frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

berdistribusi t dengan derajat kebebasan $n_1 + n_2 - 2$, apabila

a) Sampel pertama berasal dari populasi yang berdistribusi Normal, dengan rata-rata = μ_1;

b) Sampel kedua berasal dari populasi yang berdistribusi Normal, dengan rata-rata = μ_2;

c) Kedua distribusi normal itu memeiliki varians yang sama;

d) n_1 = besarnya sampel pertama;

n_2 = besarnya sampel kedua;

e) s_1^2 = varians sampel pertama;

s_2^2 = varians sampel kedua.

Maka dapat dituliskan:

$$t = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\text{, atau}$$

$$t^2 = \frac{\left(n_1 + n_2 - 2\right)n_1 n_2}{n_1 + n_2}\left[\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)\right]^2\left[\left(n_1 - 1\right)s_1^2 + \left(n_2 - 1\right)s_2^2\right]^{-1}$$

Jika $W = t^2$, maka W berdistribusi F dengan derajat kebebasan $\left(1; n_1 + n_2 - 2\right)$.

Apabila σ_1^2 dan σ_2^2 berturut-turut menyatakan varians dari populasi pertama dan populasi kedua, maka

$$\frac{\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$ berdistribusi Normal Baku;

yang berarti bahwa $\left[\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)\right]^2 \left(\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)$ berdistribusi χ^2 dengan derajat kebebasan 1.

Hal ini berlaku untuk keadaan $\sigma_1^2 = \sigma_2^2$ maupun $\sigma_1^2 \neq \sigma_2^2$

Pada situasi multivariat, distribusi statistik mirip dengan distribusi di atas juga ada, asal dipenuhi syarat-syarat yang mirip dengan situasi univariat tersebut, yaitu

a) Populasi pertama berdistribusi Normal p-variat dengan vektor rata-rata

$$\mu_1 = \left(\mu_{11}, \mu_{12}, \ldots, \mu_{1p}\right)';$$

b) Populasi kedua berdistribusi Normal p-variat dengan vektor rata-rata

$$\mu_2 = \left(\mu_{21}, \mu_{22}, \ldots, \mu_{2p}\right)';$$

c) Kedua populasi memiliki matriks varians-kovarians yang sama.

Jika syarat-syarat itu dipenuhi, dan sampel pertama mempunyai vektor rata-rata $\bar{\bar{X}}_1 = \left(\bar{X}_{11}, \bar{X}_{12}, \ldots, \bar{X}_{1p}\right)$ dan matriks varians-kovarians S_1, sedang sampel kedua mempunyai vektor rata-rata $\bar{\bar{X}}_2 = \left(\bar{X}_{21}, \bar{X}_{22}, \ldots, \bar{X}_{2p}\right)$ dan matriks varians-kovarians S_2,

dan jika $W = \dfrac{n_1 \ n_2}{n_1 \ n_2}\left[\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)\right]' S_p^{-1} \left[\left(\bar{X}_1 - \bar{X}_2\right) - \left(\mu_1 - \mu_2\right)\right]$

maka W berdistribusi T^2 dengan derajat kebebasan $\left(p; n_1 + n_2 - p - 1\right)$ dimana

$$S_p = \left[(n_1 - 1) S_1 + (n_2 - 1) S_2\right]\left(\frac{1}{n_1 + n_2 - 2}\right).$$

Hal ini berarti pula bahwa $\left(\dfrac{n_1+n_2-p-1}{p\left(n_1+n_2-2\right)}\right)W$ berdistribusi T^2 dengan derajat kebebasan $\left(p;n_1+n_2-p-1\right)$.

Jika Σ_1 dan Σ_2, berturut-turut adalah matriks varians-kovarians dari populasi pertama dan populasi kedua, baik untuk keadaan $\Sigma_1=\Sigma_2$ maupun untuk keaadaan $\Sigma_1\neq\Sigma_2$, maka $\overline{X}_1-\overline{X}_2$ berdistribusi Normal p-variat dengan vektor rata-rata $\mu=\left(\mu_1-\mu_2\right)'$ dan matriks varians-kovarians $\Sigma=\dfrac{1}{n_1}\Sigma_1+\dfrac{1}{n_2}\Sigma_2$.

BAB III

ISI

Oleh : Khaerunnisa Mahmudah (060910)

3.1 *Plausibility* dari μ_0 sebagai sebuah nilai untuk sebuah rata-rata populasi normal.

Kita memulai dengan mengingat kembali teori univariat untuk menentukan jika sebuah nilai tertentu μ_0 adalah nilai *plausible* untuk rata-rata populasi μ. Dari segi pandang pengujian hipotesis, masalah ini dapat dirumuskan sebagai suatu uji bersaing hipotesis.

$H_0 : \mu = \mu_0$ melawan $H_1 : \mu \neq \mu_0$

Jika $X_1, X_2, \ldots, X_n$ adalah sample acak dari sebuah populasi normal pengujian statistik yang sesuai adalah

$$t = \frac{\left(\bar{X} - \mu_0\right)}{s/\sqrt{n}}, \quad \text{dimana} \quad \bar{X} = \frac{1}{n}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2 \quad \textit{dan} \quad s^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)^2$$

Uji statistik adalah mempunyai sebuah distribusi-t student's dengan derajat kebebasan n – 1. Kita tolak H_0, bahwa μ_0 adalah sebuah nilai *plausible* dari μ , jika diamati $|t|$ melebihi sebuah titik persentase tertentu dari sebuah distribusi dengan derajat n – 1.

Tolak H_0 ketika $|t|$ bernilai besar yang ekuivalen dengan menolak H_0 jika kuadratnya,

$$t^2 = \frac{\left(\bar{X} - \mu_0\right)^2}{s^2/n} = n\left(\bar{X} - \mu_0\right)\left(s^2\right)^{-1}\left(\bar{X} - \mu_0\right) \qquad (3 - 1)$$

bernilai besar. Variabel t^2 adalah kuadrat jarak dari rata-rata sampel $\bar{X}$ dengan nilai uji μ_0. Unit jarak yang dinyatakan dalam pernyataan dari $s/\sqrt{n}$ atau simpangan baku yang diperkirakan dari $\bar{X}$. Ketika $\bar{X}$ dan s^2 telah diamati, uji menjadi: Tolak H_0 menuju ke H_1, pada taraf signifikansi α, jika

$$n\left(\bar{x} - \mu_0\right)\left(s^2\right)^{-1}\left(\bar{x} - \mu_0\right) > t_{n-1}^2\left(\alpha/2\right) \qquad (3 - 2)$$

dimana $t_{n-1}(\alpha/2)$ menandakan batas atas $100(\alpha/2)th$ persentil dari distribusi-t dengan derajat kebebasan n – 1.

Jika H_0 tidak ditolak, kita menyimpulkan μ_0 adalah sebuah nilai *plausible* untuk rata-rata populasi normal. Apakah nilai lain dari μ akan selalu konsisten dengan data? Jawabannya ya! Pada kenyataannya selalu sebuah himpunan dari nilai *plausible* untuk sebuah rata-rata populasi normal. Dari yang diketahui hubungan antara daerah penerimaan untuk uji $H_0 : \mu = \mu_0$ melawan $H_1 : \mu \neq \mu_0$ dan interval kepercayaan untuk μ adalah

{Jangan menolak $H_0 : \mu = \mu_0$ pada level α } atau $\left|\frac{\bar{x} - \mu_0}{s/\sqrt{n}}\right| \leq t_{n-1}(\alpha/2)$

equivalen dengan

$\left\{ \mu_0 \text{ terletak pada interval kepercayaan } 100(1-\alpha) \ \bar{\mathrm{x}} \pm \mathrm{t}_{\text{n-1}}(\alpha/2)\frac{s}{\sqrt{n}} \right\}$

atau

$$\bar{x} - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}} \qquad (3 - 3)$$

Interval konfidensi memenuhi semua nilai μ_0 bahwa tidak akan ditolak oleh uji dari $H_0 : \mu = \mu_0$.

Sebelum sampel dipilih, interval konfidensi $100(1-\alpha)\%$ pada (3 - 3) adalah sebuah interval acak karena titik akhir tergantung pada variabel acak, $\bar{X}$ dan s. Kemungkinan bahwa interval memenuhi μ adalah $1-\alpha$; antar bilangan besar seperti interval independen, $100(1-\alpha)\%$ akan memenuhi μ.

Sekarang pertimbangkan masalah yang menentukan jika sebuah $p \times 1$ vektor μ_0 adalah sebuah nilai *plausible* untuk rata-rata dari sebuah distribusi normal multivariat. Kita akan berproses oleh analogi dari pengembangan univariat

Suatu generalisasi kuadrat jarak pada (3 - 1) adalah analog multivariat

$$T^2 = \left(\bar{X} - \mu_0\right)' \left(\frac{S}{n}\right)^{-1} \left(\bar{X} - \mu_0\right) = n\left(\bar{X} - \mu_0\right)' S^{-1} \left(\bar{X} - \mu_0\right) \qquad (3 - 4)$$

dengan

$$\underset{(p \times 1)}{\bar{X}} = \frac{1}{n}\sum_{j=1}^{n} X_j, \qquad \underset{(p \times p)}{S} = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \bar{X}\right)\left(X_j - \bar{X}\right)', \qquad dan \quad \underset{(p \times 1)}{\mu_0} = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \vdots \\ \mu_{p0} \end{bmatrix}$$

Statistik T^2 dinamakan *Hotelling's* T^2 sebagai penghormatan pada Harold Hotelling, seorang pelopor dalam analisis multivariat, yang pertama mengamati distribusi sampling. Disini $\left(1/n\right)S$ adalah penaksir matrik kovarians dari $\bar{X}$. Hal ini sesuai dengan teorema akibat yang menyatakan

> ” Diberikan $X_1, X_2, \ldots, X_n$ adalah sebuah sampel acak dari distribusi gabungan yang mempunyai rata-rata vektor μ dan kovarians matriks Σ. Maka $\bar{X}$ adalah estimator takbias dari μ dan kovarians matriksnya adalah $\frac{1}{n}\Sigma$”

Jika diamati umumnya jarak T^2 terlalu besar sehingga $\bar{x}$ terlalu jauh dari μ_0 maka hipotesis $H_0 : \mu = \mu_0$ akan ditolak. Pada langkah berikutnya tabel khusus dari persentase titik T^2 tidak diperlukan untuk uji formal hipotesis. Ini benar karena

$$T^2 \text{ akan berdistribusi } \frac{(n-1)\,p}{(n-p)} F_{p,n-p} \qquad (3 - 5)$$

dimana $F_{p,n-p}$ merupakan sebuah variabel acak dengan derajat kebebasan p dan n-p.

Untuk meringkas, disajikan sebagai berikut:

Diberikan $X_1, X_2, \ldots, X_n$ sebuah sampel dari sebuah populasi $N_p(\mu, \Sigma)$.

Maka dengan $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ dan $S = \frac{1}{(n-1)}\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})'$,

$$\alpha = P\left[T^2 > \frac{(n-1)\,p}{(n-p)} F_{p,\,n-p}(\alpha)\right]$$

$$= P\left[n\,(\bar{X} - \mu)' S^{-1}(\bar{X} - \mu) > \frac{(n-1)\,p}{(n-p)} F_{p,\,n-p}(\alpha)\right] \qquad (3 - 6)$$

apapun yang benar μ dan Σ. Disini $F_{p,\,n-p}(\alpha)$ adalah batas atas (100α)th persentil dari distribusi $F_{p,\,n-p}$.

Pernyataan (3 - 6) menunjukan sebuah uji untuk hipotesis $H_0 : \mu = \mu_0$ melawan $H_1 : \mu \neq \mu_0$. Pada taraf signifikansi α, tolak H_0 menuju H_1 jika

$$T^2 = n\left(\bar{X} - \mu_0\right)' S^{-1}\left(\bar{X} - \mu_0\right) > \frac{(n-1)\,p}{(n-p)} F_{p,n-p}(\alpha) \qquad (3 - 7)$$

Pada bagian sebelumnya kita gambarkan cara dimana distribusi Wishart generalisasi distribusi Chi-kuadrat. Dapat ditulis

$$T^2 = \sqrt{n}\left(\bar{X} - \mu_0\right)' \left(\frac{\sum_{j=1}^{n}\left(X_j - \bar{X}\right)\left(X_j - \bar{X}\right)'}{n-1}\right) \sqrt{n}\left(\bar{X} - \mu_0\right)$$

yang mana berbentuk

$$\begin{pmatrix}\text{vektor acak}\\ \text{normal multivariat}\end{pmatrix}' \left(\frac{\text{matrik acak Wishart}}{\text{derajat kebebasan}}\right)^{-1} \begin{pmatrix}\text{vektor acak}\\ \text{normal multivariat}\end{pmatrix}$$

Ini beranalogi pada

$$t^2 = \sqrt{n}\left(\bar{X} - \mu_0\right)\left(s^2\right)^{-1}\sqrt{n}\left(\bar{X} - \mu_0\right)$$

atau

$$\begin{pmatrix}\text{variabel}\\ \text{acak normal}\end{pmatrix}'\begin{pmatrix}\dfrac{\text{variabel acak Chi-kuadrat}}{\text{derajat kebebasan}}\end{pmatrix}^{-1}\begin{pmatrix}\text{variabel}\\ \text{acak normal}\end{pmatrix}$$

untuk kasus univariat. Karena normal multivariat dan variabel acak Wishart berdistribusi independen, dengan fungsi densitas gabungannya dari produk normal marginal dan distribusi Wishart. Dengan menggunakan kalkulus, distribusi T^2 seperti tersebut diatas dapat diperoleh dalam bentuk distribusi gabungan.

Adalah jarang, dalam keadaan multivariat, isi dengan sebuah uji $H_0 : \mu = \mu_0$, dimana semua komponen vektor rata-rata adalah tertentu dibawah hipotesis nol. Biasanya lebih baik mencari daerah dari nilai μ sehingga *plausible* untuk memecah data yang diamati.

Contoh 3.1

Diberikan data matrik untuk sebuah sampel acak berukuran n = 3 dari sebuah populasi normal bivariat

$$X = \begin{bmatrix} 6 & 10 & 8 \\ 9 & 6 & 3 \end{bmatrix}$$

Evaluasi yang diamati T^2 untuk $\mu_0' = [9,5]$ dan $\alpha = 0.05$. Apakah distribusi sampling dari kasus ini?

<u>Solusi</u>

$$\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} \dfrac{6+10+8}{3} \\ \dfrac{9+6+3}{3} \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$$

dan

$$s_{11}=\frac{(6-8)^2+(10-8)^2+(8-8)^2}{2}=4$$

$$s_{12}=\frac{(6-8)(9-6)+(10-8)(6-6)+(8-8)(3-6)}{2}=-3$$

$$s_{22}=\frac{(9-6)^2+(6-6)^2+(3-6)^2}{2}=9$$

jadi

$$S=\begin{bmatrix}4 & -3\\ -3 & 9\end{bmatrix}$$

sehingga

$$\mathrm{S}^{-1}=\frac{1}{(4)(9)-(-3)(-3)}\begin{bmatrix}9 & 3\\ 3 & 4\end{bmatrix}=\begin{bmatrix}\frac{1}{3} & \frac{1}{9}\\ \frac{1}{9} & \frac{4}{27}\end{bmatrix}$$

dan

$$\mathrm{T}^2=3[8-9,6-5]\begin{bmatrix}\frac{1}{3} & \frac{1}{9}\\ \frac{1}{9} & \frac{4}{27}\end{bmatrix}\begin{bmatrix}8-9\\ 6-5\end{bmatrix}=3[-1,1]\begin{bmatrix}-\frac{2}{9}\\ \frac{1}{27}\end{bmatrix}=\frac{7}{9}$$

Sebelum sampel dipilih, T^2 memiliki distribusi dari sebuah variabel acak

$$\frac{(3-1)2}{(3-2)}\mathrm{F}_{2,\,3-2}(0.05)=\ 4\ \mathrm{F}_{2,\,1}(0.05)=4(199.5)=798$$

Tolak H_0 jika $\mathrm{T}^2>\frac{(\mathrm{n}-1)p}{(n-p)}F_{\ p,n-p}(\alpha)$. Karena $T^2=0.778<798=\frac{(3-1)2}{(3-2)}\mathrm{F}_{2,\,3-2}(0.05)$ maka H_0 diterima sehingga $\mu_0'=[9,5]$ adalah sebuah nilai *plausible* untuk rata-rata populasi normal.

Contoh 3.2

Perspirasi dari 20 wanita sehat dianalisis. Tiga komponen, X_1 = *sweat rate*, X_2 = *sodium content*, dan X_3 = *potassium content*, telah diukur dan dinilai. Uji hipotesis $H_0:\mu'=[4,50,10]$ melawan $H_1:\mu'\neq[4,50,10]$ pada taraf signifikansi $\alpha=0.10$. Untuk datanya diberikan pada tabel berikut:

TABEL 3.1 *SWEAT DATA*

Individual	X_1 (*Sweat rate*)	X_1 (*Sodium*)	X_3 (*Potassium*)
1	3.7	48.5	9.3
2	5.7	65.1	8
3	3.8	47.2	10.9
4	3.2	53.2	12
5	3.1	55.5	9.7
6	4.6	36.1	7.9
7	2.4	24.8	14
8	7.2	33.1	7.6
9	6.7	47.4	8.5
10	5.4	54.1	11.3
11	3.9	36.9	12.7
12	4.5	58.8	12.3
13	3.5	27.8	9.8
14	4.5	40.2	8.4
15	1.5	13.5	10.1
16	8.5	56.4	7.1
17	4.5	71.6	8.2
18	6.5	52.8	10.9
19	4.1	44.1	11.2
20	5.5	40.9	9.4

Sumber : Courtesy of Dr. Gerald Bargman

Dari hasil perhitungan komputer diperoleh:

$$\bar{x} = \begin{bmatrix} 4.640 \\ 45.400 \\ 9.965 \end{bmatrix}, \qquad \mathrm{S} = \begin{bmatrix} 2.879 & 10.002 & -1.810 \\ 10.002 & 199.798 & -5.627 \\ -1.810 & -5.627 & 3.628 \end{bmatrix}$$

dan

$$\mathrm{S}^{-1} = \begin{bmatrix} 0.586 & -0.022 & 0.258 \\ -0.022 & 0.006 & -0.002 \\ 0.258 & -0,002 & 0.402 \end{bmatrix}$$

Sehingga akan diperoleh

$$T^2 = 20\left[4.640-4, 45.400-50, 9.965-10\right]\begin{bmatrix} 0.586 & -0.022 & 0,258 \\ -0.022 & 0.006 & -0.002 \\ 0.258 & -0.002 & 0.402 \end{bmatrix}\begin{bmatrix} 4.640-40 \\ 45.400-50 \\ 9,965-10 \end{bmatrix}$$

$$= 20\left[0.640, -4.600, -0.035\right]\begin{bmatrix} 0.467 \\ -0.042 \\ 0.160 \end{bmatrix}$$

$$= 9.74$$

Membandingkan yang diamati $T^2 = 9.74$ dengan nilai kritisnya

$$\frac{(\mathrm{n}-1)\,p}{(n-p)} F_{p,n-p}(\alpha) = \frac{(19)3}{17} F_{3,17}(0.10) = 3.353(2.44) = 8.18$$

karena $\mathrm{T}^2 = 9.74 > 8.18,$ maka H_0 ditolak pada taraf signifikansi 10%.

Kesimpulannya $\mu' = [4, 50, 10]$ merupakan suatu nilai *plausible* untuk μ .

Satu bentuk dari statistik-T^2 adalah invarians (tanpa perubahan) di bawah perubahan didalam unit pengukuran dari X dengan bentuk

$$\underset{(p\times 1)}{Y} = \underset{(p\times p)}{C}\ \underset{(p\times 1)}{X} + \underset{(p\times 1)}{d}, \qquad \text{C nonsingular} \qquad (3 - 8)$$

Sebuah transformasi dari pengamatan sesama muncul ketika sebuah konstanta b_i adalah yang dikurangidari variabel ke-i untuk membentuk $X_i - b_i$ dan hasil dari perkalian dengan konstanta $a_i > 0$ untuk mendapatkan $a_i\left(X_i - b_i\right)$. Sebelum perkalian yang berpusat dan berskala jumlahnya $a_i\left(X_i - b_i\right)$oleh setiap matrik nonsingular akan menghasilkan persamaan (3 - 8). Karena sebuah contoh, operasi yang melibatkan penggantian X_i dengan $a_i\left(X_i - b_i\right)$yang bersesuaian pada proses mengubah suhu dari Fahrenheit ke Celcius.

Diberikan pengamatan $x_1, x_2, \ldots, x_n$ dan transformasi pada (3 - 8), akan mengikuti suatu teorema akibat yaitu

”Kombinasi linier dalam AX pada

$$\begin{bmatrix} a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \\ a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p \\ \vdots \quad \vdots \quad \ddots \quad \vdots \\ a_{q1}X_1 + a_{q2}X_2 + \ldots + a_{qp}X_p \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1p} \\ a_{21} & a_{22} & \ldots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{q1} & a_{q2} & \ldots & a_{qp} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} = AX$$

memiliki vektor rata-rata sampel $A\overline{x}$ dan kovarians matriks ASA' ”

Sehingga

$$\overline{y} = C\,\overline{x} + d \quad dan \quad S_y = \frac{1}{n-1}\sum_{j=1}^{n}\left(y_i - \overline{y}\right)\left(y_i - \overline{y}\right)' = CSC'$$

Selanjutnya, oleh persamaan

$$\boxed{\begin{aligned} &E\left(X+Y\right) = E\left(X\right) + E\left(Y\right) \\ &E\left(AXB\right) = AE\left(X\right)B \end{aligned}}$$

dan persamaan kombinasi linier dari $Z = CX$ mempunyai

$$\boxed{\begin{aligned} &\mu_Z = E\left(Z\right) = E\left(CX\right) = C\mu_x \\ &\Sigma_Z = Cov\left(Z\right) = Cov\left(CX\right) = C\Sigma_X C' \end{aligned}}$$

maka akan dihasilkan

$$\mu_y = E\left(Y\right) = E\left(CX + d\right) = E\left(CX\right) + E\left(d\right) = C\mu + d$$

Oleh karena itu, T^2 dihitung dengan y’s dan sebuah nilai hipotesis $\mu_{Y,0} = C\mu_0 + d$ adalah

$$\begin{aligned} T^2 &= n\left(\overline{y} - \mu_{Y,0}\right)' S_Y^{-1}\left(\overline{y} - \mu_{Y,0}\right) \\ &= n\left(C\left(\overline{x} - \mu_0\right)\right)'\left(CSC'\right)^{-1}\left(C\left(\overline{x} - \mu_0\right)\right) \\ &= n\left(\overline{x} - \mu_0\right)' C'\left(CSC'\right)^{-1} C\left(\overline{x} - \mu_0\right) \\ &= n\left(\overline{x} - \mu_0\right)' C'\left(C'\right)^{-1} S^{-1}C^{-1}C\left(\overline{x} - \mu_0\right) \\ &= n\left(\overline{x} - \mu_0\right)' S^{-1}\left(\overline{x} - \mu_0\right) \end{aligned}$$

Persamaan yang terakhir dikenali sebagai nilai dari T^2 dihitung dengan x’s.

Oleh : Risa Nur vauzyah (060933)

3.2 Hotteling T^2 dan Uji Perbandingan Likelihood

Kita perkenalkan statistik-T^2 analogi dengan jarak kuadrat univariat, t^2. Ada sebuah prinsip umum untuk mengkontruksi langkah-langkah pengujian yang disebut *metode perbandingan likelihood* dan statistik-T^2dapat diperoleh sebagai uji rasio likelihood dengan $H_0 : \mu = \mu_0$. Uji rasio likelihood memiliki beberapa sifat optimal yang layak untuk sampel besar, dan terutama sekali untuk perumusan hipotesis dalam pernyataan parameter normal multivariat.

Kita ketahui bahwa maksimum likelihood normal multivariat sebagai μ dan Σ adalah bervariasi nilai kemungkinannya diberikan oleh

$$\max_{\mu,\Sigma} L(\mu,\Sigma) = \frac{1}{(2\pi)^{np/2}\left|\hat{\Sigma}\right|^{n/2}} e^{-np/2} \qquad \text{(3-9)}$$

dimana $\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})(x_j - \bar{x})'$ dan $\hat{\mu} = \bar{\mathrm{x}} = \frac{1}{\mathrm{n}}\sum_{j=1}^{n} x_j$

adalah penaksir maksimum likelihood. Sebagai pengingat bahwa penaksir maksimum likelihood $\hat{\mu}$ dan $\hat{\Sigma}$ dipilih dari μ dan Σ yang merupakan alasan terbaik untuk nilai yang diamati dari sampal acak.

Untuk hipotesis $H_0 : \mu = \mu_0$, normal likelihood mengkhususkan pada

$$\mathrm{L}(\mu_0,\Sigma) = \frac{1}{(2\pi)^{np/2}\left|\Sigma\right|^{n/2}} \exp\left(-\frac{1}{2}\sum_{j=1}^{n}(x_j - \mu_0)'\,\Sigma^{-1}(x_j - \mu_0)\right) \qquad \text{(3-10)}$$

Untuk menentukan apakah μ_0 adalah nilai yang tak mungkin untuk μ, maksimum $\mathrm{L}(\mu_0,\Sigma)$ dibandingkan dengan maksimum $\mathrm{L}(\mu,\Sigma)$ yang diperbolehkan. Hasil perbandingannya dinamakan *statistik perbandingan likelihood*.

Dengan menggunakan persamaan (5-9) dan (5-10) diperoleh,

$$\text{Rasio Likelihood} = \wedge = \frac{\max_{\mu,\Sigma} L(\mu,\Sigma)}{\max_{\mu,\Sigma} L(\mu_0,\Sigma)} = \left(\frac{\left|\hat{\Sigma}\right|}{\left|\hat{\Sigma}_0\right|}\right)^{\mathrm{n}/2} \qquad \text{(3-11)}$$

Padanan statistik untuk $\wedge^{2/n} = |\hat{\Sigma}| / |\hat{\Sigma}_0|$ disebut Wilks' lamda. Jika nilai pengamatan perbandingan likelihood ini terlalu kecil, hipotesis $H_0 : \mu = \mu_0$ tidak mungkin menjadi benar, oleh karena itu ditolak. Secara rinci, uji rasio likelihood untuk $H_0 : \mu = \mu_0$ melawan $H_1 : \mu \neq \mu_0$, tolak H_0 jika

$$\Lambda = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2} = \left(\frac{\left| \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})' \right|}{\left| \sum_{j=1}^{n} (x_j - \mu_0)(x_j - \mu_0)' \right|} \right)^{n/2} < c_\alpha \qquad (3\text{-}13)$$

dimana c_α adalah batas bawah $(100\alpha)th$ persentil dari distribusi Λ. (Catatan bahwa statistik uji rasio likelihood adalah sebuah kuasa perbandingan variansi yang diperumum).

Akibat 3.1.

Diberikan $X_1, X_2, \ldots, X_n$ adalah sampel acak dari populasi derdistribusi $N_p(\mu, \Sigma)$. Maka uji pada (5-7) merupakan dasar dati T^2 yang ekivalen dengan uji rasio likelihood dari $H_0 : \mu = \mu_0$ melawan $H_1 : \mu \neq \mu_0$, karena

$$\Lambda^{2/n} = \left(1 + \frac{T^2}{(n-1)} \right).$$

<u>Metode Perbandingan Likelihood Umum</u>

Kita sekarang akan mempertimbangkan metode perbandingan likelihood umum. Diberikan θ adalah sebuah vektor yang memenuhi semua parameter populasi yang diketahui, dan diberikan $L(\theta)$ adalah fungsi likelihood yang diperoleh dengan mengevaluasi kepadatan densitas dari $X_1, X_2, \ldots, X_n$ pada nilai yang diamati $x_1, x_2, \ldots, x_n$. Vektor parameter θ mengambil nilai dalam himpunan parameter Θ.

Uji rasio likelihood untuk $H_0 : \theta \in \Theta_0$ menuju ke $H_1 : \theta \notin \Theta_0$ jika

$$\Lambda = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} < c \qquad (\text{bab } 2\text{-}16)$$

dimana c adalah konstanta tertentu yang dipilih. Secara intuitif, kita tolak H_0 jika maksimum dari likelihood yang diperoleh dengan mempertukarkan θ pada himpunan Θ_0 yang lebih kecil dari maksimum likelihood yang dipenuhi oleh variasi θ untuk semua nilai pada Θ. Ketika maksimum pada pembilang dari persamaan (bab 2-16) lebih kecil dari maksimum penyebut, Θ_0 tidak memenuhi nilai plausibel untuk θ.

Pada setiap aplikasi dari metode perbandingan likelihood, kita akan memerlukan distribusi sampling dari statistik uji rasio likelihood Λ. Sehingga c dapat dipilih untuk menghasilkan sebuah uji dengan sebuah taraf signifikansi α tertentu. Bagaimanapun, ketika ukuran sampelnya besar dan kondisi keteraturan tertentu dipenuhi, distribusi sampling dari $-2\ln\Lambda$ yang didekati oleh sebuah distribusi chi-kuadrat.

Akibat 3.2

Ketika ukuran sampel n besar

$$-2\ln\Lambda = 2\ln\Lambda\frac{\max\limits_{\theta\in\Theta_0} L(\theta)}{\max\limits_{\theta\in\Theta} L(\theta)} \text{ adalah aproksimasi dari variabel acak } \chi^2_{\text{v-v}_0}$$

dengan derajat kebebasannya $v - v_0$ = (dimensi dari Θ) – (dimensi dari Θ_0).

3.3 Daerah Kepercayaan dan Perbandingan Simultan dari Komponen Rata-rata

Daerah yang ditentukan oleh sebuah data, untuk sementara, kita notasikan dengan R(X), dengan $X = [X_1, X_2, \ldots, X_n]$ adalah matriks data. Daerah R(X) dikatakana akan menjadi daerah kepercayaan $100(1-\alpha)\%$ jika sebelum sample dipilih,

$$P[R(X)\ akan\ mencakup\ nilai\ \theta\ yangsebenarnya] = 1-\alpha$$

Daerah kepercayaan untuk rata-rata μ dari dimensi-p yang berdistribusi normal diperoleh dari (2-6). Sebelum sampel dipilih,

$$P\left[n(\bar{X}-\mu)' S^{-1}(\bar{X}-\mu) \le \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)\right] = 1-\alpha$$

Untuk sebarang nilai μ dan $\sum$ tidak diketahui.

Untuk sample khusus, $\bar{x}$ dan S dapat dihitung dan ketaksamaan $n(\bar{x}-\mu)' S^{-1}(\bar{x}-\mu) \le (n-1)\,pF_{p,n-p}(\alpha)/(n-p)$ akan mendefinisikan daerah, R(X), dalam ruang dari semua nilai parameter yang mungkin. Dalam kasus ini, daerah akan menjadi ellipsoid dengan pusat $\bar{x}$. Ellipsoid ini adalah daerah kepercayaan $100(1-\alpha)\%$ untuk μ.

Daerah kepercayaan $100(1-\alpha)\%$ untuk rata-rata dari dimensi-p yang berdistribusi normal adalah himpunan yang ditentukan oleh semua μ sedemikian sehingga

$$n(\bar{x}-\mu)' S^{-1}(\bar{x}-\mu) \le \frac{(n-1)\,p}{(n-p)} F_{p,n-p}(\alpha)$$

dimana $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$, $S = \frac{1}{(n-1)}\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'$, dan $x_1, x_2, \ldots, x_n$ adalah sample pengamatan.

Untuk $p \ge 4$, kita tidak dapat menggambarkan daerah kepercayaan untuk μ. Akan tetapi, kita dapat menghitung sumbu-x dari ellipsoid kepercayaan dan panjang relatifnya. Hal ini ditentukan dari nilai eigen λ_i dan vector eigen e_i dari S. Seperti dalam persamaan $(x-\mu)' \Sigma^{-1}(x-\mu) = c^2$, arah dan panjang sumbu-x

dari $$n(\bar{x}-\mu)' S^{-1}(\bar{x}-\mu) \le c^2 = \frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha)$$

akan ditentukan oleh

$$\sqrt{\lambda_i}\, c/\sqrt{n} = \sqrt{\lambda_i}\sqrt{p(n-1)F_{p,n-p}(\alpha)/n(n-p)}$$

Unit sepanjang vector eigen e_i. Berawal di pusat $\bar{x}$, sumbu-x dari ellipsoid kepercayaan adalah

$$\pm\sqrt{\lambda_i}\sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)}\, e_i \quad \text{dimana} \quad Se_i = \lambda_i e_i \text{ , } i = 1, 2, \ldots, p$$

Perbandingan dari $\lambda_i' s$ akan membantu dalam mengidentifikasi jumlah relatif dari pemanjangan sepanjang pasangan sumbu-x.

Oleh : Lucky Heriyanti Jufri (0607103)

Pernyataan Kepercayaan Simultan

Ketika daerah kepercayaan $n\left(\overline{x}-\mu\right)' S^{-1}\left(\overline{x}-\mu\right) \le c^2$, dengan c adalah konstanta, dapat dilihat dengan tepat hubungan mengenai nilai plausible untuk μ, apa saja inti dari kesimpulan yang biasa dimasukkan dalam pernyataan kepercayaan tentang rata-rata komponen tunggal. Selanjutnya, kita gunakan aturan bahwa pernyataan kepercayaan yang terpisah, sebaiknya mempertahankan kesimultanaan-nya dengan tingginya probabilitas yang ditentukan. Hal ini merupakan jaminan dalam menentukan probabilitas terhadap banyaknya pernyataan salah yang menyebabkan interval kepercayaan simultan. Kita awali dengan mengingat pernyataan kepercayaan simultan yang berhubungan dengan daerah kepercayaan bersama berdasarkan statistik T^2.

Misalkan X berdistribusi $N_p(\mu,\Sigma)$ dan bentuk kombinasi liniernya yaitu

$$Z = \ell_1 X_1 + \ell_2 X_2 + \ldots + \ell_p X_p = \ell' X$$

Sebagaimana yang kita ketahui bahwa $\mu_z = E(Z) = \ell'\mu$ dan $\sigma_z^2 = Var(Z) = \ell'\Sigma\ell$. Selain itu, berdasarkan **akibat 4.2**, Z berdistribusi $N(\ell\mu, \ell'\Sigma\ell)$. Jika sample acak $X_1, X_2, \ldots, X_n$ dari populasi berdistribusi $N(\mu,\Sigma)$ adalah memungkinkan, maka sample $Z's$ dapat ditulis dengan menggunakan kombinasi linier yaitu. Jadi,

$$Z_j = \ell_1 X_{1j} + \ell_2 X_{2j} + \ldots + \ell_p X_{pj} = \ell' X_j \quad , j = 1, 2, \ldots, n$$

Rata-rata dan variansi dari $z_1, z_2, \ldots, z_n$ adalah $\overline{z} = \ell'\overline{x}$ dan $s_z^2 = \ell' S \ell$, dimana $\overline{x}$ dan S adalah vektor rata-rata dan matriks kovarians sample dari $x_j's$, berturut-turut.

Interval kepercayaan simultan dapat dikembangkan dengan pertimbangan dari interval kepercayaan $\ell'\mu$ untuk sebarang ℓ.

Untuk ℓ tertentu dan σ_z^2 tidak diketahui, interval kepercayaan $100\left(1-\alpha\right)\%$ untuk $\mu_z = \ell'\mu$ adalah berdasarkan rasio-*t student's*

$$t = \frac{\overline{z} - \mu_z}{s_x / \sqrt{n}} = \frac{\sqrt{n}\left(\ell'\overline{x} - \ell'\mu\right)}{\sqrt{\ell' S \ell}} \qquad (3\text{-}14)$$

Sehingga diperoleh pernyataan

$$\bar{z} - t_{n-1}\left(\alpha/2\right)\frac{s_z}{\sqrt{n}} \le \mu_z \le \bar{z} + t_{n-1}\left(\alpha/2\right)\frac{s_z}{\sqrt{n}}$$

atau

$$\ell'\bar{x} - t_{n-1}\left(\alpha/2\right)\frac{\sqrt{\ell' S \ell}}{\sqrt{n}} \le \ell'\mu \le \ell'\bar{x} + t_{n-1}\left(\alpha/2\right)\frac{\sqrt{\ell' S \ell}}{\sqrt{n}} \tag{3-15}$$

dimana $t_{n-1}\left(\alpha/2\right)$ adalah batas atas $100\left(1-\alpha\right)\%$ dari distribusi-t dengan derajat kebebasan (n-1).

Ketidaksamaan (3-5) dapat dinyatakan sebagai pernyataan mengenai komponen dari vektor rata-rata μ. Sebagai contoh, dengan $\ell' = [1, 0, \ldots, 0]$, $\ell'\mu = \mu_1$ dan ketidaksamaan (3-5) menghasilkan interval kepercayaan biasa untuk rata-rata dari populasi normal. Dalam kasus ini $\ell' S \ell = s_{11}$, jelasnya, kita akan menentukan beberapa pernyataan kepercayaan mengenai komponen μ, dengan menghubungkan koefisien kepercayaan $1-\alpha$, dengan memilih koefisien vector ℓ yang berbeda. Bagaimanapun, hubungan kepercayaan dengan semua pernyataan yang diambil bersama adalah bukan $1-\alpha$.

Berdasarkan intuisi, akan dihubungkan koefisien kepercayaan "kolektif" $1-\alpha$ dengan interval kepercayaan sehingga dihasilkan oleh semua pilihan ℓ. Nilai tersebut harus mengganti koefisien kepercayaan yang besar dengan sebaik-baiknya. Nilai tersebut ada dalam bentuk interval yang lebih luas dibandingkan dengan interval pada ketidaksamaan (3-15) untuk pilihan ℓ yang spesifik.

Diberikan data himpunan $x_1, x_2, \ldots, x_n$ dan ℓ tertentu, interval kepercayaan dalam ketidaksamaan (3-5) adalah himpunan dari nilai $\ell'\mu$ untuk

$$|t| = \left|\frac{\sqrt{n}\left(\ell'\bar{x} - \ell'\mu\right)}{\sqrt{\ell' S \ell}}\right| \le t_{n-1}(\alpha/2)$$

atau, ekivalen dengan

$$t^2 = \frac{n\left(\ell'\bar{x} - \ell'\mu\right)^2}{\ell' S \ell} = \frac{n\left(\ell'\left(\bar{x} - \mu\right)\right)^2}{\ell' S \ell} \le t_{n-1}(\alpha/2) \tag{3-16}$$

Daerah kepercayaan simultan diberikan oleh himpunan nilai $\ell'\mu$ yaitu t^2 relatif kecil untuk semua ℓ. Nampaknya pantas untuk menduga bahwa konstanta $t_{n-1}^2\left(\alpha/2\right)$ dalam persamaan (3-6) akan digantikan oleh nilai yang lebih besar yaitu c^2, ketika pernyataan dikembangkan untuk sembarang ℓ.

Mengingat nilai ℓ untuk $t^2 \leq c^2$, secara otomatis kita peroleh ketetapan :

$$\max_{\ell}\ t^2 = \max_{\ell} \frac{n\left(\ell'\left(\overline{x}-\mu\right)\right)^2}{\ell' S\ell}$$

Dengan menggunakan *Maximization lemma* : $\max\limits_{x\neq 0} \frac{\left(x'd\right)^2}{x'Bx} = d'B^{-1}d$, dimana $x=\ell$, $d=\left(\overline{x}-\mu\right)$, dan $B=S$, diperoleh :

$$\max_{\ell} \frac{n\left(\ell'\left(\overline{x}-\mu\right)\right)^2}{\ell' S\ell} = n\left[\max_{\ell} \frac{n\left(\ell'\left(\overline{x}-\mu\right)\right)^2}{\ell' S\ell}\right] = n\left(\overline{x}-\mu\right)' S^{-1}\left(\overline{x}-\mu\right) = T^2 \quad \text{(3-17)}$$

Untuk ℓ sepadan dengan $S^{-1}\left(\overline{x}-\mu\right)$.

Akibat 3.3

Misalkan $X_1, X_2, \ldots, X_n$ sample random dari populasi berdistribusi $N_p(\mu,\Sigma)$ Dengan Σ definit positif. Maka, kesimultanan untuk semua ℓ, interval

$$\left(\ell'\overline{X} - \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)\ell'S\ell}.\ell'\overline{X} + \sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)\ell'S\ell}\right)$$

akan memuat $\ell'\mu$ dengan probabilitas $1-\alpha$.

Bukti : Dari persamaan (bab 5-23),

$T^2 = n\left(\overline{x}-\mu\right)' S^{-1}\left(\overline{x}-\mu\right) \leq c^2$ termasuk $\frac{n\left(\ell'\overline{x}-\ell'\mu\right)^2}{\ell'S\ell} \leq c^2$ untuk setiap ℓ, atau $\ell'\overline{x} - c\sqrt{\frac{\ell'S\ell}{n}} \leq \ell'\mu \leq \ell'\overline{x} + c\sqrt{\frac{\ell'S\ell}{n}}$ untuk setiap ℓ. Dengan memilih

$c^2 = p(n-1)F_{p,n-p}(\alpha)/(n-p)$ memberikan interval yang akan memuat $\ell'\mu$ untuk semua ℓ, dengan probabilitas $1-\alpha = P\left[T^2 \le c^2\right]$.

Ini adalah tepat mengarahkan ke interval yang simultan dari **akibat 3.3** sebagai interval-T^2, karena pencakupan probalbilitas ditentukan oleh distribusi T^2. Berturut-turut kita pilih $\ell' = [1,0,\ldots,0]$, $\ell' = [0,1,\ldots,0]$, dengan demikian $\ell' = [0,0,\ldots,1]$ untuk interval-T^2 membolehkan kita untuk menyimpulkan

$$\begin{aligned}
\bar{x}_1 - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}} \le \mu_1 \le \bar{x}_1 + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{11}}{n}} \\
\bar{x}_2 - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}} \le \mu_2 \le \bar{x}_2 + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{22}}{n}} \\
\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
\bar{x}_p - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}} \le \mu_p \le \bar{x}_p + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{pp}}{n}}
\end{aligned} \qquad (3\text{-}18)$$

semua memperoleh kesimultanan dengan koefisien kepercayaan $1-\alpha$.

Catatan bahwa, tanpa modifikasi koefisien $1-\alpha$, kita dapat membuat pernyataan turunan dari $\mu_i - \mu_k$ sesuai dengan $\ell' = [0,\ldots,0,\ell_i,0,\ldots,0,\ell_k,0,\ldots,0]$, dimana $\ell_i = 1$ dan $\ell_k = -1$. Dalam kasus ini $\ell' S\ell = s_{ii} - 2s_{ik} + s_{kk}$, dan kita mempunyai pernyataan

$$\bar{x}_i - \bar{x}_k - \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii}-2s_{ik}+s_{kk}}{n}} \le \mu_i - \mu_k \le \bar{x}_i - \bar{x}_k + \sqrt{\frac{p(n-1)}{(n-p)}F_{p,n-p}(\alpha)}\sqrt{\frac{s_{ii}-2s_{ik}+s_{kk}}{n}} \qquad (3\text{-}19)$$

Kesimultanan Interval kepercayaan T^2 merupakan ide untuk “data snooping”. Koefisien kepercayaan $1-\alpha$ tetap tidak terganti untuk sebarang pemilihan ℓ, sehingga kombinasi linier dari komponen μ_i yang manfaat pemeriksaannya berdasarkan pemeriksaan dari data dapat dihitung.

Perbandingan Interval Kepercayaan Simultan dengan Interval Pada Satu Waktu

Sebagai alternatif, untuk meminimalisir terjadinya kesalahan dalam melakukan pendekatan untuk menentukan interval kepercayaan adalah dengan mempertimbangkan komponen μ_i pada satu waktu, seperti yang telah dijelaskan pada persamaan (3-5) dengan $\ell' = [0,\dots,0,\ell_i,0,\dots,0]$, dimana $\ell' = 1$. Pendekatan ini mengabaikan struktur kovarian dari variable-p dan membawa kita ke interval

$$\begin{aligned}
\bar{x}_1 - t_{n-1}\left(\alpha/2\right)\sqrt{\frac{s_{11}}{n}} &\le \mu_1 \le \bar{x}_1 + t_{n-1}\left(\alpha/2\right)\sqrt{\frac{s_{11}}{n}} \\
\bar{x}_2 - t_{n-1}\left(\alpha/2\right)\sqrt{\frac{s_{22}}{n}} &\le \mu_2 \le \bar{x}_2 + t_{n-1}\left(\alpha/2\right)\sqrt{\frac{s_{22}}{n}} \\
\vdots \qquad &\;\;\vdots \qquad\qquad \vdots \\
\bar{x}_p - t_{n-1}\left(\alpha/2\right)\sqrt{\frac{s_{pp}}{n}} &\le \mu_p \le \bar{x}_p + t_{n-1}\left(\alpha/2\right)\sqrt{\frac{s_{pp}}{n}}
\end{aligned} \qquad (3\text{-}20)$$

Walaupun sebelum pengambilan sampling, interval ke-i di atas memiliki probabilitas $1-\alpha$ meliputi μ_i, kita tidak tahu apa yang menyatakan secara umum, mengenai kemungkinan semua interval memuat masing-masing $\mu_i's$.

Untuk memberi pencerahan terhadap masalah ini, dengan mempertimbangkan kasus khusus dimana pengamatannya berdistribusi normal gabungan dan

$$\Sigma = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{pp} \end{bmatrix}$$

Karena pengamatan pada variable pertama adalah independent, begitupula untuk variable kedua, dan seterusnya. Aturan yang diperoleh yaitu untuk peristiwa independent dapat digunakan sebelum sampel dipilih,

$$\begin{aligned}
P\left[\, semua\ \text{int}erval - t\ \ pada\ (3-10)\ memuat\ \mu_i's \right] &= (1-\alpha)(1-\alpha)\cdots(1-\alpha) \\
&= (1-\alpha)^p
\end{aligned}$$

Untuk memastikan probabilitas $1-\alpha$ bahwa semua pernyataan mengenai komponen rata-rata secara umum, interval tunggal harus lebih luas dari interval terpisah. Luas interval ini bergantung pada p dan n, sebagaimana dalam $1-\alpha$.

Oleh : Asti Aulia Rahman (0607196)

3.4 Perbandingan Interval T^2 Simultan Dan Interval Bonferroni Dari Komponen Rata - Rata

Untuk memperoleh metode utama dalam menentukan inferensi dari sample, kita akan memperluas konsep interval kepercayaan *univariat* menjadi daerah kepercayaan *multivariate*. Berdasarkan penjelasan pada bab sebelumnya, telah dijelaskan inferensi sampel dengan menggunakan $\text{int}\,erval - T^2$ simultan. Namun seringkali kita jumpai interval yang lebih pendek untuk bilangan m yang kecil, yaitu ketika $m = p$. Dalam hal ini, akan lebih mudah untuk menggunakan dan menetapkan interval kepercayaan yang relatif pendek, yang dibutuhkan untuk membuat kesimpulan (*inference*). Sehingga kita dapat menetapkan nilai interval yang lebih pendek dari $\text{int}\,erval - T^2$**.** Metode seperti ini akan dibahas pada pembahasan berikut ini disertai dengan studi kasusnya.

Metode *Bonferroni* untuk Perbandingan Berganda

Seringkali perhatian kita terbatas pada bilangan yang kecil dari pernyataan kepercayaan tunggal. Dalam situasi seperti ini memungkinkan untuk melakukan sesuatu yang lebih baik dari kesimultanan interval dari **akibat 3.3**. Jika bilangan m dari komponen rata-rata khusus μ_i, atau kombinasi linier $\ell'\mu = \ell_1\mu_1 + \ell_2\mu_2 + \ldots + \ell_p\mu_p$, adalah kecil, interval kepercayaan simultan dapat dikembangkan menjadi lebih pendek (lebih tepat) dari pada interval-T^2 simultan. Metode alternatif untuk perbandingan berganda dinamakan "Metode *Bonferroni*" , karena ini dikembangkan dari kemungkinan yang membawa nama ketidaksamaan tersebut.

Andaikata, sebelum ke kumpulan data, pernyataan kepercayaan mengenai kombinasi linier m yaitu $\ell_1'\mu, \ell_2'\mu, \ldots, \ell_m'\mu$ adalah yang diharuskan. Misalkan C_i notasi dari pernyataan kepercayaan mengenai nilai dari $\ell_i'\mu$ dengan $P\left[C_i \;\; benar\right] = 1 - \alpha_i \quad , i = 1, 2, \ldots, m$.

$$
\begin{aligned}
P\left[semua\ C_i\ \ benar\right] &= 1 - P\left[\ paling\ sedikit\ satu\ C_i\ salah\right] \\
&\geq 1 - \sum_{i=1}^{m} P\left(C_i\ salah\right) = 1 - \sum_{i=1}^{m}\left(1 - P\left(C_i\ benar\right)\right) \qquad (3\text{-}21) \\
&= 1 - \left(\alpha_1 + \alpha_2 + \ldots + \alpha_m\right)
\end{aligned}
$$

Ketidaksamaan (3-21), kasus khusus dari ketidaksamaan *Bonferroni*, memenuhi pemeriksaan untuk mengontrol keseluruhan nilai kesalahan $\alpha_1 + \alpha_2 + \ldots + \alpha_m$, tanpa memperhatikan struktur korelasi di belakang pernyataan kepercayaan. Hal ini juga fleksibel dalam mengontrol nilai kesalahan untuk kelompok dari pernyataan penting dan seimbang dengan pilihan lain untuk pernyataan penting yang kurang.

Misalkan kita kembangkan estimasi interval kepercayaan untuk himpunan terbatas yang terdiri dari komponen μ_i dari μ. Tak cukup informasi dalam kepentingan yang relative dari komponen ini, kita mempertimbangkan interval *t*-tunggal $\bar{x}_i \pm t_{n-1}\left(\frac{\alpha_i}{2}\right)\sqrt{\frac{s_{ii}}{n}}\quad, i = 1, 2, \ldots, m$ dengan $\alpha_i = \alpha/m$. Karena $P\left[\ \bar{X}_i \pm t_{n-1}\left(\alpha/2m\right)\sqrt{s_{ii}/n}\ \ memuat\ \mu_i\ \right] = 1 - \alpha/m\quad, i = 1, 2, \ldots, m$, kita peroleh dari persamaan (3-11)

$$
\begin{aligned}
P\left[\ \bar{X}_i \pm t_{n-1}\left(\frac{\alpha}{2m}\right)\sqrt{s_{ii}/n}\ \ memuat\ \mu_i\ , semua\ i\ \right] &\geq 1 - \underbrace{\left(\frac{\alpha}{m} + \frac{\alpha}{m} + \cdots + \frac{\alpha}{m}\right)}_{bentuk\ m} \qquad (3\text{-}22) \\
&= 1 - \alpha
\end{aligned}
$$

Untuk itu, dengan keseluruhan tingkat kepercayaan lebih besar dari atau sama dengan $1-\alpha$, kita dapat membuat pernyataan $m = p$:

$$
\begin{aligned}
\bar{x}_1 - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}} &\leq \mu_1 \leq \bar{x}_1 + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{11}}{n}} \\
\bar{x}_2 - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{22}}{n}} &\leq \mu_2 \leq \bar{x}_2 + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{22}}{n}} \qquad (3\text{-}23) \\
\vdots \qquad & \quad \vdots \qquad \quad \vdots \\
\bar{x}_p - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}} &\leq \mu_p \leq \bar{x}_p + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{pp}}{n}}
\end{aligned}
$$

Pernyataan dalam ketidaksamaan (3-13) dapat dibandingkan dengan ketidaksamaan dalam (3-8). Nilai persentase $t_{n-1}\left(\alpha/2p\right)$ menggantikan $\sqrt{(n-1)\,pF_{p,n-p}(\alpha)/(n-p)}$, tapi sebaliknya intervalnya masih dalam struktur yang sama.

3.5. Inferensi Vektor Mean Populasi Untuk Sampel Besar

Ketika ukuran sampel besar, pengujian hipotesis dan daerah kepercayaan untuk μ dapat dikonstruksi tanpa anggapan normalitas. Untuk jumlah n besar, kita dapat membuat taksiran tentang rata-rata populasi meskipun distribusi awalnya adalah diskrit.

Keuntungan berasosiasi dengan sample besar yaitu kemungkinan kehilangan informasi dari statistic cukup $\bar{x}$ dan **S** adalah kecil. Selain itu, $\bar{x}$ dan **S** yang merupakan statistic cukup untuk populasi normal adalah hal yang mendasari populasi normal multivariate, dimana informasi tersebut akan digunakan untuk membuat taksiran.

Penaksiran μ untuk sample besar adalah mendekati distribusi χ^2. Sebagaimana kita tahu dari bab sebelumnya bahwa $(\bar{X}-\mu)'(S/n)^{-1}(\bar{X}-\mu)=n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu)$ mendekati distribusi χ^2 dengan derajat kebebasan adalah p, maka

$$P[n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu)\le \chi_p^2(\alpha)]=1-\alpha$$

Misalkan X_1, X_2, ..., X_n adalah sample acak dari populasi dengan mean μ dan kovarians Σ. Jika n-p besar, hipotesis H_0 : $\mu=\mu_0$ ditolak dengan alternative H_1 : $\mu\neq\mu_0$ pada taraf signifikansi α jika

$$n(\bar{X}-\mu)'S^{-1}(\bar{X}-\mu)>\chi_p^2(\alpha)$$

Misalkan X_1, X_2, ..., X_n adalah sample acak dari populasi dengan mean μ dan definit positif kovarians Σ. Jika n-p besar, maka

$$\ell'\bar{X}\pm\sqrt{\chi_p^2(\alpha)}\sqrt{(\ell'S\ell/n)}$$

Dimana setiap ℓ memuat $\ell'\mu$ dengan probabilitas 1 - α. Akibatnya kita dapat membuat interval konfidensi 100 (1-α)%

$$\bar{x}_1 \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{s_{11}}{n}} \qquad \text{memuat } \mu_1$$

$$\bar{x}_2 \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{s_{22}}{n}} \qquad \text{memuat } \mu_2$$

.

.

.

$$\bar{x}_p \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{s_{pp}}{n}} \qquad \text{memuat } \mu_p$$

Oleh : Syifa Insani (060116)

3.6 Penaksir Vektor Mean Ketika Beberapa Vektor Inferensi Hilang

Sering kali beberapa komponen dari vektor observasi tidak ada. Maka dalam menyelesaikan masalah tersebut dengan menggunakan teknik EM algorithm, disetiap iterasi memiliki dua langkah yakni :

- Prediksi
- Estimasi

  Menggunakan statistika cukup untuk estimasi parameter

Misal X_1, X_2 ,…, Xn adalah sampel acak berpopulasi normal p variate (μ , Σ). Algoritma prediksi dan estimasi berdasar pada statistika cukup sebagai berikut:

$$T_1 = \sum_{j=1}^{n} X_j$$

$$T_2 = \sum_{j=1}^{n} X_j X_j' = (n-1)S + n\bar{X}\,\bar{X}'$$

Langkah Prediksi :

Untuk setiap $X_j^{(1)}$ adalah komponen vektor yang hilang, dan $X_j^{(2)}$ adalah komponen vektor yang ada. Untuk penduga $\tilde{\mu}$ dan $\tilde{\Sigma}$ dari langkah estimasi digunakan mean distribusi bersyarat $x^{(1)}$ dan diberikan $x^{(2)}$ untuk menduga nilai yang hilang. Sehingga:

$$\tilde{x}_j^{(1)} = E(X_j^{(1)} | x_j^{(2)}; \tilde{\mu}, \tilde{\Sigma}) \quad = \tilde{\mu}^{(1)} + \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}(\tilde{x}_j^{(2)} - \tilde{\mu}^{(2)})$$

Menduga kontribusi $x_j^{(1)}$ untuk T_1 :

$$\widetilde{x_j^{(1)} x_j^{(1)}} = E(X_j^{(1)} X_j^{(1)}{}' | x_j^{(2)}; \tilde{\mu}, \tilde{\Sigma})$$

$$= \tilde{\Sigma}_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + \tilde{X}_j^{(1)} \tilde{X}_j^{(1)}$$

$$\widetilde{x_j^{(1)} x_j^{(2)}} = E(X_j^{(1)} X_j^{(2)}{}' | x_j^{(2)}; \tilde{\mu}, \tilde{\Sigma})$$

$$= \tilde{x}_j^{(1)} \tilde{x}_j^{(2)}$$

Menduga kontribusi $x_j^{(1)}$ untuk T_2 :

Kontribusi pertama dijumlahkan untuk setiap x_j dengan komponen hilang. Hasil ini digabung dengan data sampel menghasilkan T_1 dan T_2.

Langkah estimasi:

Dihitung penduga maksimum likelihood terevisi:

$$\tilde{\mu} = \frac{\tilde{T}_1}{n}$$

$$\tilde{\Sigma} = \frac{1}{n}\tilde{T}_2 - \tilde{\mu}\tilde{\mu}'$$

Contoh 5.7 halaman 204

Estimasilah populasi normal ini dengan mean μ dan Σ variansi, himpunan datanya sebagai berikut:

$$X_{(3,4)} = \begin{bmatrix} - & 7 & 5 & - \\ 0 & 2 & 1 & - \\ 3 & 6 & 2 & 5 \end{bmatrix}$$

Jawab:

Diperoleh rata-rata sampel adalah :

$$\tilde{\mu}_1 = 6 \qquad \tilde{\mu}_2 = 1 \qquad \tilde{\mu}_3 = 4$$

kemudian subsitusikan rata-rata tersebut ke nilai yang hilang, sehingga diperoleh estimasi terhadap variansi, yaitu :

$$\tilde{\sigma}_{11} = \frac{1}{2} \qquad \tilde{\sigma}_{22} = \frac{1}{2} \qquad \tilde{\sigma}_{33} = \frac{5}{2} \qquad \tilde{\sigma}_{12} = \frac{1}{4} \qquad \tilde{\sigma}_{23} = \frac{3}{4} \qquad \tilde{\sigma}_{13} = \frac{1}{1}$$

Langkah pertama adalah Prediksi, dalam memprediksi nilai yang hilang kita menggunakan estimasi terhadap $\tilde{\mu}$ dan $\tilde{\Sigma}$, disubsitusikan ke statistika cukup T_1 dan T_2. Komponen $x_1$yang hilang, dipartisi sehingga:

$$\tilde{\mu}_1 = \begin{bmatrix} \tilde{\mu}_1 \\ \hline \tilde{\mu}_2 \\ \tilde{\mu}_3 \end{bmatrix} = \begin{bmatrix} \tilde{\mu}^{(1)} \\ \hline \tilde{\mu}^{(2)} \end{bmatrix} \qquad \tilde{\Sigma} = \left[\begin{array}{c|cc} \tilde{\sigma}_{11} & \tilde{\sigma}_{12} & \tilde{\sigma}_{13} \\ \hline \tilde{\sigma}_{12} & \tilde{\sigma}_{22} & \tilde{\sigma}_{23} \\ \tilde{\sigma}_{13} & \tilde{\sigma}_{23} & \tilde{\sigma}_{33} \end{array}\right]$$

diduga

$$\tilde{x}_{11} = \tilde{\mu}^{(1)} + \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}(\tilde{X}_j^{(2)} - \tilde{\mu}^{(2)})$$

$$= 6 + \left[\frac{1}{4}, 1\right] \begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ \frac{3}{4} & \frac{5}{2} \end{bmatrix}^{-1} \begin{bmatrix} 0-1 \\ 3-4 \end{bmatrix} = 5.73$$

$$\tilde{x}_{11}^2 = \tilde{\sigma}_{11} - \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21} + \tilde{x}_{11}^2$$

$$= \frac{1}{2} - \begin{bmatrix} \frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{3}{4} \\ \frac{3}{4} & \frac{5}{2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{4} \\ 1 \end{bmatrix} + (5.73)^2 = 32.99$$

$$\widetilde{x_{11}[x_{21},x_{31}]} = \tilde{x}_{11}[x_{21},x_{31}] = 5.73[0,3] = [0,17,18]$$

Untuk data hilang pada komponen ke 4, dipartisi sebagai:

$$\tilde{\mu}_1 = \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \\ --- \\ \tilde{\mu}_3 \end{bmatrix} = \begin{bmatrix} \tilde{\mu}^{(1)} \\ ---- \\ \tilde{\mu}^{(2)} \end{bmatrix} \qquad \tilde{\Sigma} = \begin{bmatrix} \tilde{\sigma}_{11} & \tilde{\sigma}_{12} & | & \tilde{\sigma}_{13} \\ \tilde{\sigma}_{12} & \tilde{\sigma}_{22} & | & \tilde{\sigma}_{23} \\ --- & --- & & --- \\ \tilde{\sigma}_{13} & \tilde{\sigma}_{23} & | & \tilde{\sigma}_{33} \end{bmatrix}$$

Diduga :

$$\begin{bmatrix} \tilde{x}_{14} \\ x_{24} \end{bmatrix} = E\left(\begin{bmatrix} x_{14} \\ x_{24} \end{bmatrix} x_{34} = 5; \tilde{\mu}, \tilde{\Sigma} \right) = \begin{bmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{bmatrix} + \tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}(x_{34} - \tilde{\mu}_3)$$

$$= \begin{bmatrix} 6 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ \frac{3}{4} \end{bmatrix} \left(\frac{5}{2} \right)^{-1} (5-4) = \begin{bmatrix} 6.4 \\ 1.3 \end{bmatrix}$$

Kontribusi terhadap T_1:

$$\begin{bmatrix} \tilde{x}_{14}^2 & \widetilde{x_{14}x_{24}} \\ \widetilde{x_{14}x_{24}} & \tilde{x}_{24}^2 \end{bmatrix} = E\left(\begin{bmatrix} X_{14}^2 & X_{14}X_{24} \\ X_{14}X_{24} & X_{24}^2 \end{bmatrix} x_{34} = 5; \tilde{\mu}, \tilde{\Sigma} \right)$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} \end{bmatrix} - \begin{bmatrix} 1 \\ \frac{3}{4} \end{bmatrix} \left(\frac{5}{2} \right)^{-1} [1, \frac{3}{4}] + \begin{bmatrix} 6.4 \\ 1.3 \end{bmatrix} [6.4 \quad 1.3] = \begin{bmatrix} 41.06 & 8.27 \\ 8.27 & 1.97 \end{bmatrix}$$

$$\begin{bmatrix} x_{14} \\ x_{24} \end{bmatrix} (x_{34}) = E\left(\begin{bmatrix} X_{14}X_{34} \\ X_{24}X_{34} \end{bmatrix} \middle| x_{34} = 5; \tilde{\mu}, \tilde{\Sigma} \right) = \begin{bmatrix} \tilde{x}_{14} \\ x_{24} \end{bmatrix} (x_{34})$$

$$= \begin{bmatrix} 6.4 \\ .1.3 \end{bmatrix} (5) = \begin{bmatrix} 32.0 \\ 6.5 \end{bmatrix}$$

Penduga Statististika cukup:

$$\tilde{T}_1 = \begin{bmatrix} \tilde{x}_{11} + x_{12} + x_{13} + \tilde{x}_{14} \\ x_{21} + x_{22} + x_{23} + \tilde{x}_{24} \\ x_{31} + x_{32} + x_{33} + x_{34} \end{bmatrix} = \begin{bmatrix} 24.13 \\ 4.30 \\ 16.00 \end{bmatrix}$$

Selanjutnya adalah langkah estimasi , dengan menggunakan maksimum likelihood terevisi sebagai berikut:

$$\tilde{T}_2 = \begin{bmatrix} 148.05 & 27.27 & 101.18 \\ 27.27 & 6.97 & 20.50 \\ 101.18 & 20.50 & 74.00 \end{bmatrix}$$

$$\tilde{\mu} = \frac{\tilde{T}_1}{n} = \frac{1}{4}\begin{bmatrix} 24.13 \\ 4.30 \\ 16.00 \end{bmatrix} = \begin{bmatrix} 6.03 \\ 1.08 \\ 4.00 \end{bmatrix}$$

$$\tilde{\Sigma} = \frac{1}{n}\tilde{T}_2 - \tilde{\mu}\tilde{\mu} = \frac{1}{4}'\begin{bmatrix} 148.05 & 27.27 & 101.18 \\ 27.27 & 6.97 & 20.50 \\ 101.18 & 20.50 & 74.00 \end{bmatrix} - \begin{bmatrix} 6.03 \\ 1.08 \\ 4.00 \end{bmatrix}[6.03 \quad 1.08 \quad 4.00]$$

$$= \begin{bmatrix} 0.65 & 0.31 & 1.18 \\ 0.31 & 0.58 & 0.81 \\ 1.18 & 0.81 & 2.5 \end{bmatrix}$$

Terlihat $\tilde{\sigma}_{11} = 0.65$ dan $\tilde{\sigma}_{22} = 0.58$ lebih besar dari estimasi pertama observasi yang hilang. Sedangkan $\tilde{\sigma}_{33} = 0.25$ sama dengan estimasi awal. Dari hasil tersebut, kita harus melakukan iterasi yang sama sampai elemen elemen dan sama dan $\tilde{\mu}$ tidak $\tilde{\Sigma}$ diganti.

Estimasi $\tilde{\mu}$ dan $\tilde{\Sigma}$ berakhir ketika :

$$n(\hat{\mu} - \mu)'\hat{\Sigma}^{-1}(\hat{\mu} - \mu) \leq \chi_p^2(\alpha)$$

$\forall\mu$ memenuhi dengan kepercayaan elipsoide 100 (1-α)%.

BAB IV

PENUTUP

4.1 Kesimpulan

1) Dari analisis dan perhitungan yang telah dilakukan pada studi kasus dapat ditunjukkan $\mu_0' = [13, 7, 11]$ merupakan suatu nilai *plausible* untuk μ. Dengan kata lain vektor rata-rata populasi multivariat akan selalu konsisten dengan data yang dimiliki.

2) Pengujian hipotesis dengan menggunakan rumus perhitungan T^2 yang berbentuk $\mathrm{T}^2 = \left(\overline{X} - \mu_0\right)' \left(\frac{S}{n}\right)^{-1} \left(\overline{X} - \mu_0\right) = n\left(\overline{X} - \mu_0\right)' S^{-1} \left(\overline{X} - \mu_0\right)$ maupun

3) Dapat kita lihat dari pernyataan simultan di atas bahwa komponen μ_0 dari melodi, tempo dan meter tidak terbukti sebagai nilai yang mungkin untuk nilai akhir rata-rata.(dengan derajat kebebasan 90%, nilai yang kita tetapkan tepat dengan perhitungan atau tidak)

4.2 Saran

Agar kesalahan dapat terminimalkan maka penyusun memberi saran sebagai berikut:

a. Pergunakanlah software yang memadai dalam melakukan pengujian hipotesis terutama dalam perhitungan perkalian matriksnya. Software yang penyusun sarankan untuk menghitung perkalian matriks adalah Math Lab.
b. Diperlukan kehati-hatian dalam melakukan penginputan karena seringkali terjadi ketidakcocokan hasil perhitungan yang disebabkan kekeliruan memasukan data.

DAFTAR PUSTAKA

Johnson, Richard A. and Dean W. Wichern. Third Edition. *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall, Englewood Cliffs.

Suryanto, Dr. 1988. *Metode Statistika Multivariat*. Jakarta: Departemen Pendidikan dan Kebudayaan.