Statistika, Vol. 10 No. 1, 43 – 49
Mei 2010

Perluasan Uji Kruskal Wallis untuk Data Multivariat

TETI SOFIA YANTI

Program Studi Statistika, Universitas Islam Bandung,
Jl. Purnawarman No. 60 Bandung.
E-mail: buitet@yahoo.com

ABSTRAK

Andaikan terdapat *k* buah populasi saling bebas dimana dari masing-masing terdapat p buah variat yang diperhatikan, selanjutnya akan diuji apakah rata-rata p buah variat dari k buah populasi tersebut mempunyai rata-rata yang sama. Dalam statistika parametrik pengujian seperti itu dijelaskan dalam Multivariat Analisis Varians (MANOVA) melalui uji F. Uji F dalam MANOVA mengasumsikan data harus berdistibusi normal dan variabel yang diperhatikan dari k buah populasi harus mempunyai varians yang homogen. Apabila kedua asumsi tersebut tidak terpenuhi maka prosedur statistika nonparametrik harus dilalui, salah satu pengujian yang bisa dilakukan adalah melalui perluasan uji Kruskal Wallis untuk data multivariat.

Kata Kunci: sampel saling bebas; rangking; data multivariat; Uji Kruskal Wallis.

1. PENDAHULUAN

Dalam analisis Statistika terdapat dua macam konsep dan prosedur analisis yaitu parametrik dan nonparametrik. Analisis parametrik sangat bergantung pada pola distribusi populasi yang diamati, sedangkan analisis nonparametrik tidak perlu memperhatikan hal tersebut sehingga untuk skala pengukuran yang paling rendahpun dapat dianalisis.

Dalam analisis parametrik, untuk pengujian hipotesis apakah terdapat perbedaan antara k buah rata-rata sampel yang independen digunakan Analisis Varians (ANAVA) dengan statistik uji F. Prosedur dari uji tersebut mengasumsikan bahwa sampel berasal dari distribusi Normal dengan varians yang homogen. Apabila kedua asumsi tersebut tidak terpenuhi atau skala pengukurannya ordinal, maka analisis nonparametrik dapat mengatasinya melalui ANAVA dengan uji Kruskal Wallis.

Misalnya akan dilakukan pengujian apakah prestasi mahasiswa k buah program studi di sebuah perguruan tinggi sama atau tidak, dimana variabel yang diamati adalah IPK. Apabila untuk pengujian di atas prosedur parametrik tidak dapat dipenuhi maka pengujiannya menggunakan uji Kruskal Wallis. Selanjutnya jika yang diperhatikan selain IPK adalah IQ dan Skor Toefl, maka uji Kruskal Wallis tidak dapat dilakukan, maka melalui tulisan ini akan dibahas Perluasan Uji Kruskal Wallis untuk Data Multivariat.

2. UJI KRUSKAL WALLIS UNTUK DATA UNIVARIAT

Andaikan terdapat *k* buah populasi saling bebas dimana dari masing-masing hanya memperhatikan satu buah variabel, selanjutnya akan diuji apakah rata-rata variabel atau perilaku dari *k* buah populasi tersebut sama? atau dengan kata lain apakah ke-*k* buah populasi mempunyai distribusi yang identik?

Asumsi yang harus dipenuhi dalam analisis ini adalah:

1. Data terdri atas k sampel acak berukuran $n_1, n_2,\ldots,n_k$
2. Observasi saling bebas di dalam sampel maupun antar sampel
3. Variabel pengamatan adalah kontinu
4. Skala pengukuran minimal Ordinal

5. Populasi-populasi identik kecuali dalam hal lokasi yang mungkin berbeda minimal untuk satu populasi

Hipotesisnya adalah:

H_0: $F_1(X) = F_2(X) = F_3(X) = \cdots = F_k(X)$: Semua k populasi mempunyai distribusi identik.

H_1: $F_k(X) \neq F_l(X)$ untuk beberapa $k \neq l$: Paling sedikit satu tanda sama dengan tidak berlaku.

Jika H_0 benar maka sampel berukuran n berasal dari populasi identik, atau dengan kata lain perilaku setiap populasi yang diwakili oleh sampel tidak berbeda. Struktur datanya sebagai berikut:

Tabel 1. Stuktur Data *k* Buah Sampel Bebas
dengan Variabel Univariat

Pengamatan	Sampel				
	1	2	3	...	K
1	x_{11}	x_{12}	x_{131}	...	x_{1k1}
2	x_{21}	x_{22}	x_{231}	...	x_{2k1}
...	...	...	...	...	...
n_j	X_{n11}	X_{n22}	x_{n33}	...	x_{nkk}

Statistik uji yang digunakan adalah:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1) \tag{1}$$

Dengan:

R_i = Jumlah rank sampel ke-i

$N = n_1 + n_2 + \ldots + n_k$

Statistik uji H mengikuti distribusi Chi Kuadrat (x^2), sehingga kriteria pengujiannya adalah tolak H_0 jika $H \geq x^2$ dengan derajat bebas k-1 pada taraf nyata atau taraf signifikansi a.

3. PERLUASAN UJI KRUSKAL WALLIS UNTUK DATA MULTIVARIAT

Andaikan terdapat k buah populasi saling bebas dimana dari masing-masing terdapat p buah variat yang diperhatikan, selanjutnya akan diuji apakah rata-rata p buah variat dari k buah populasi tersebut mempunyai rata-rata yang sama atau tidak. Dalam statistika parametrik pengujian seperti itu dijelaskan dalam Multivariat Analisis Varians (MANOVA) melalui uji F. Uji F dalam MANOVA mengasumsikan data harus berdistibusi normal dan variabel yang diperhatikan dari k buah populasi harus mempunyai varians yang homogen. Apabila kedua asumsi tersebut tidak terpenuhi maka prosedur statistika nonparametrik harus dilalui, salah satu pengujian yang bisa dilakukan adalah melalui Perluasan Uji Kruskal Wallis untuk Data Multivariat. Asumsi yang harus dipenuhi sama sepeti dalam uji Kruskal Wallis untuk data univariat.

Tabel 2. Stuktur Data k Buah Sampel Bebas dengan Variabel Multivariat

	Sampel																
	1				2				3				...	K			
	Variabel				Variabel				Variabel					Variabel			
No.	1	2	...	p	1	2	...	p	1	2	...	P	...	1	2	...	P
1	x_{111}	x_{112}	...	x_{11p}	x_{121}	x_{122}	...	x_{12p}	x_{131}	x_{132}	...	x_{13p}	...	x_{1k1}	x_{1k2}	...	x_{1kp}
2	x_{211}	x_{212}	...	x_{21p}	x_{221}	x_{222}	...	x_{22p}	x_{231}	x_{232}	...	x_{23p}	...	x_{2k1}	x_{2k2}	...	x_{2kp}
3	x_{311}	x_{312}	...	x_{31p}	x_{321}	x_{322}	...	x_{32p}	x_{331}	x_{332}	...	x_{33p}	...	x_{3k1}	x_{3k2}	...	x_{3kp}
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
n_j	x_{n111}	x_{n112}	...	x_{n11p}	x_{n221}	x_{n222}	...	x_{n22p}	x_{n331}	x_{n332}	...	x_{n33p}	...	x_{nkk1}	x_{nkk2}	...	x_{nkkp}

Seperti pengujian pada prosedur statistika nonparametrik yang lain, yang diperhatikan dalam perluasan Uji Kruskal Wallis untuk data multivariat adalah penggunaan Rank. Untuk lebih jelasnya, struktur datanya Terlihat pada Tabel 2.

Hipotesisnya adalah:

H_0: $F_1(\underline{X}) = F_2(\underline{X}) = F_3(\underline{X}) = \cdots = F_k(\underline{X})$: Semua k populasi mempunyai distribusi identik

H_1: $F_k(\underline{X}) \neq F_l(\underline{X})$ untuk beberapa k≠l : Paling sedikit satu tanda sama dengan tidak berlaku.

Jika H0 benar maka sampel berukuran n berasal dari populasi identik, atau dengan kata lain perilaku setiap populasi yang diwakili oleh sampel tidak berbeda.

Untuk mendapatkan statistik uji perhatikan uraian berikut. Data pengamatan dari k sampel yang saling bebas tersebut terlebih dahulu diberi tanda (sign). Dengan menggunakan rumus sign(x(i) –x(j)). Setelah pemberian tanda (sign) dilakukan, dapat dihitung pusat dan skala ranking yang didefinisikan sebagai berikut:

$$R(x^{(i)}) = \frac{1}{n}\sum_{j=1}^{n} sign(x^{(i)} - x^{(j)}), i \neq j \qquad (2)$$

dimana $R(x^{(i)})$ adalah jumlah hasil pengurangan tanda untuk masing-masing pengamatan, melalui

$$sign(x^{(i)} - x^{(j)}) = \begin{cases} -1 & \text{jika } (x^{(i)} - x^{(j)}) < 0 \\ 0 & \text{jika } (x^{(i)} - x^{(j)}) = 0 \\ 1 & \text{jika } (x^{(i)} - x^{(j)}) > 0 \end{cases}$$

Selanjutnya akan ditentukan Taksiran Kovarian Vektor Rangking berdasarkan taksiran gabungan

$$\hat{\sum} = \frac{1}{N-k}\sum_{k=1, i\in n_k}^{k}\sum R_{n_k}(x^{(i)})R_{n_k}(x^{(i)})' \qquad (3)$$

dengan:

n_k = banyaknya pengamatan pada sampel ke-k (tiap sampel)

k = banyaknya sampel yang diamati

$R_{n_k}(x^{(i)})$ adalah vektor ukuran skala rangking sampel ke n_k, yaitu $R_{n_k}(x^{(i)}) = \begin{bmatrix} R(^{x(1)}) \\ R(^{x(2)}) \\ \vdots \\ R(^{x(n_k)}) \end{bmatrix}$

Statistik uji yang digunakan berdasarkan statistik uji Kruskal Wallis yaitu:

$$KW = \sum_{k=1}^{k} n_k \left(\overline{R}^{(k)}\right)' \hat{\sum}^{-1} \overline{R}^{(k)} \qquad (4)$$

dimana $\left(\overline{R}^{(k)}\right)$ adalah rata-rata sampel ke-k, yaitu:

$$\left(\overline{R}^{(k)}\right) = \frac{1}{n_k}\sum_{i\in n_k} R(x^{(i)})$$

$$\left(\bar{R}^{(k)}\right)=\frac{1}{n_k}\left[R(x^{(1)})+R(x^{(2)})+\cdots+R(x^{(nk)})\right]$$

Statistik uji (KW) mengikuti distribusi Chi Kuadrat (x^2), sehingga kriteria pengujiannya adalah tolak H_0 jika KW ≥ x^2 dengan derajat bebas k-1 pada taraf nyata atau taraf signifikansi α.

Jika H_0 ditolak maka terdapat perbedaan antara kelompok sampel dilihat dari masing-masing variat. Untuk selanjutnya perlu dilihat kelompok sampel mana saja yang berbeda satu dengan yang lainnya. Pengujiannya dilakukan terhadap ke-k sampel untuk masing-masing variat melalui uji perbandingan berganda.

Untuk perbandingan berganda menggunakan rumus:

1) Jika ukuran sampel berbeda

$$\left|\bar{R}_i-\bar{R}_j\right| \le z_{(1-\{\alpha/k(k-1)\})}\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \tag{5}$$

dimana $N=n_1+n_2+\ldots+n_k$

2) Jika ukuran sampel sama

$$\left|\bar{R}_i-\bar{R}_j\right| \le z_{(1-\{\alpha/k(k-1)\})}\sqrt{k(N+1)/6} \tag{6}$$

4. CONTOH PENERAPAN

Untuk contoh penerapan adalah ingin dilihat apakah perilaku seseorang dalam memilih suatu produk dalam hal ini motor Yamaha Mio antara tiga wilayah di Kota Bandung sama atau tidak. Variabel yang diperhatikan ada tiga yaitu diferensiasi produk (DP), bauran promosi (BP) dan ekuitas merek (EM). Sampel diambil dari pengguna motor Yamaha Mio yang membayar pajak pada tahun 2009 di Samsat Kota Bandung. Samsat kota Bandung dibagi menjadi tiga wilayah yaitu: Samsat Bandung Timur, Samsat Bandung Tengah dan Samsat Bandung Barat. Setiap responden diberi beberapa pertanyaan untuk masing-masing variabel, skala pengukuran yang digunakan adalah ordinal. Data ini diperoleh dari Regina P.

Tabel 3. Jawaban Responden Pengguna Motor Yamaha Mio

Jawaban Responden								
Samsat Bandung Timur			Samsat Bandung Tengah			Samsat Bandung Barat		
P	P	M	P	P	M	P	P	M
9	3	8	8	6	6	7	0	3
6	7	8	2	6	0	7	2	8
4	2	3	4	0	2	1	5	1
8	7	3	0	8	6	0	7	6
9	8	9	9	1	0	5	5	0
6	6	9	9	6	2	5	4	6
7	4	0	9	4	8	6	1	6
			9	1	2			

Sumber: Regina P.

1) Perhitungan nilai pusat dan Skala Ranking untuk masing-masing sampel
   a. Skala rangking Bandung Timur

$$R(x^{(1)}) = \frac{1}{7}\sum_{j=1}^{7} sign(x^{(i)} - x^{(j)})$$
$$= \frac{1}{7}\left[sign(x^{(1)} - x^{(1)}) + sign(x^{(1)} - x^{(2)}) + \cdots + sign(x^{(1)} - x^{(7)})\right]$$
$$= \frac{1}{7}\left[\begin{pmatrix}0\\0\\0\end{pmatrix} + \begin{pmatrix}1\\-1\\0\end{pmatrix} + \begin{pmatrix}1\\-1\\1\end{pmatrix} + \begin{pmatrix}1\\-1\\1\end{pmatrix} + \begin{pmatrix}0\\1\\1\end{pmatrix} + \begin{pmatrix}-1\\-1\\-1\end{pmatrix} + \begin{pmatrix}-1\\-1\\-1\end{pmatrix}\right]$$
$$= \begin{bmatrix}0{,}142857\\-0{,}57143\\0{,}142857\end{bmatrix}$$

$$R(x^{(2)}) = \frac{1}{7}\sum_{j=1}^{7} sign(x^{(i)} - x^{(j)})$$
$$= \begin{bmatrix}-0{,}57143\\0{,}285714\\0{,}142857\end{bmatrix}$$

$$\vdots$$

$$R(x^{(7)}) = \frac{1}{7}\sum_{j=1}^{7} sign(x^{(i)} - x^{(j)})$$
$$= \begin{bmatrix}0{,}857143\\0{,}571429\\0{,}857143\end{bmatrix}$$

   b. Skala rangking Bandung Tengah

$$R(x^{(1)}) = \frac{1}{8}\sum_{j=1}^{8} sign(x^{(i)} - x^{(j)})$$
$$= \frac{1}{8}\left[sign(x^{(1)} - x^{(1)}) + sign(x^{(1)} - x^{(2)}) + \cdots + sign(x^{(1)} - x^{(7)})\right]$$
$$= \frac{1}{8}\left[\begin{pmatrix}0\\0\\0\end{pmatrix} + \begin{pmatrix}-1\\0\\1\end{pmatrix} + \begin{pmatrix}-1\\-1\\1\end{pmatrix} + \begin{pmatrix}-1\\-1\\1\end{pmatrix} + \begin{pmatrix}-1\\1\\1\end{pmatrix} + \begin{pmatrix}-1\\0\\1\end{pmatrix} + \begin{pmatrix}-1\\1\\1\end{pmatrix} + \begin{pmatrix}-1\\-1\\1\end{pmatrix}\right]$$
$$= \begin{bmatrix}-0{,}875\\-0{,}125\\0{,}875\end{bmatrix}$$

$$R(x^{(2)}) = \frac{1}{8}\sum_{j=1}^{8} sign(x^{(i)} - x^{(j)})$$

$$= \begin{bmatrix} 0{,}625 \\ 0{,}000 \\ -0{,}500 \end{bmatrix}$$

$$\vdots$$

$$R(x^{(8)}) = \frac{1}{8}\sum_{j=1}^{8} sign(x^{(i)} - x^{(j)})$$

$$= \begin{bmatrix} \text{-}0{,}250 \\ 0{,}875 \\ 0{,}125 \end{bmatrix}$$

c. Skala rangking Bandung Barat

$$R(x^{(1)}) = \frac{1}{7}\sum_{j=1}^{7} sign(x^{(i)} - x^{(j)})$$

$$= \frac{1}{7}\left[sign(x^{(1)} - x^{(1)}) + sign(x^{(1)} - x^{(2)}) + \cdots + sign(x^{(1)} - x^{(7)}) \right]$$

$$= \frac{1}{7}\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} \right]$$

$$= \begin{bmatrix} \text{-}0{,}71429 \\ -0{,}85714 \\ 0{,}00000 \end{bmatrix}$$

$$R(x^{(2)}) = \frac{1}{7}\sum_{j=1}^{7} sign(x^{(i)} - x^{(j)})$$

$$= \begin{bmatrix} \text{-}0{,}71429 \\ -0{,}28571 \\ \text{-}0{,}85714 \end{bmatrix}$$

$$\vdots$$

$$R(x^{(7)}) = \frac{1}{7}\sum_{j=1}^{7} sign(x^{(i)} - x^{(j)})$$

$$= \begin{bmatrix} 0{,}57143 \\ -0{,}57143 \\ 0{,}57143 \end{bmatrix}$$

2) Taksiran kovarian vektor rangking

$$\hat{\Sigma} = \begin{bmatrix} 0,118153 & 0,058002 & 0,073040 \\ 0,058002 & 0,108485 & 0,006444 \\ 0,073040 & 0,006444 & 0,125671 \end{bmatrix}$$

$$\hat{\Sigma}^{-1} = \begin{bmatrix} 20,6254 & -10,3469 & -11,4569 \\ -10,3469 & 14,4366 & 5,2733 \\ -11,4569 & 5,2733 & 14,3456 \end{bmatrix}$$

3) Rata-rata sampel

$$\left(\bar{R}^1\right) = \begin{bmatrix} 0,00000086 \\ 0,00000086 \\ 0,00000086 \end{bmatrix} \left(\bar{R}^2\right) = \begin{bmatrix} 0,138393 \\ 0,053714 \\ 0,279018 \end{bmatrix} \left(\bar{R}^3\right) = \begin{bmatrix} -0,00000057 \\ 0,02040786 \\ 0,02040886 \end{bmatrix}$$

4) Statistik uji

KW = 5,3676

5) Keputusan pengujian

Kriteria pengujiannya, dengan tingkat signifikansi ditetapkan sebesar 0,05 maka tolak H_0 jika KW ≥ 5,99. Karena KW< 5,99 maka Ho tidak ditolak yang berarti perilaku seseorang dalam memilih suatu produk dalam hal ini motor Yamaha Mio antara tiga wilayah di Kota Bandung tidak berbeda. Karena keputusan pengujian tidak menolak H_0 maka tidak perlu dilakukan ujian perbandingan berganda.

DAFTAR PUSTAKA

[1]. Choi, K. And Marden. J. 1987. An Approach to Multivariate Rank Test Multivariate Analysis of Variance. Journal of Th American Statistical Association.
[2]. Conover. 1971. Practical Nonparametric Statistic. New York, Wiley.
[3]. Daniel, Wayne W. 1989. Statistik Nonparametrik Terapan. PT Gramedia. Jakarta.1989.
[4]. Gibbon. 1971. Nonparametric Statistical Inference. New York. Mc Graw-Hill., Inc.