

ANALISIS KELOMPOK HIRARKI UNTUK PERBANDINGAN SAMPEL BANYAK

Elfitra¹

ABSTRAK

Biasanya dalam usaha perbandingan sampel banyak dari sebuah observasi, banyak metode yang digunakan. Tujuan dari beberapa metode ini adalah untuk menguji hipotesis kesamaan pasangan, akan tetapi sulit menyaring sifat-sifat khusus dari data yang disajikan. Sebuah pendekatan alternatif diperkenalkan dengan tidak melibatkan tes hipotesis untuk menguji kesamaan kelompok melainkan melihat perbedaan mean kemudian mengkategorikan mean dan sampel berbeda jika berada pada kelompok yang berbeda. Metode Analisis kelompok yang dikenalkan disini menggunakan algoritma secara hirarkidan mengenalkan model Informasi kriteria untuk melihat pasangan kelompok yang memiliki kesamaan. Secara umum dalam analisis kelompok diasumsi berdistribusi normal. Dengan mengabaikan asumsi normalitas analisis kelompok dikerjakan dengan distribusi power normal. Hasil analisis kelompok dengan power normal juga memiliki kesamaan gambaran seperti yang ditampilkan dalam grafik statistik.

Kata kunci: Analisis kelompok, Analisis kelompok Sampel Banyak, Kriteria Informasi, Perbandingan sampel banyak

¹ Elfitra, Alumni S2 Matematika, FMIPA USU, Dosen Universitas Negeri Medan

PENDAHULUAN

Keberagaman informasi sering dijumpai pada institusi, perusahaan, organisasi, medis dan bidang keilmuan lainnya. Karenanya, manajemen informasi merupakan hal yang penting dalam pengambilan keputusan. Informasi yang diperoleh disebut dengan data. Namun data yang diperoleh terkadang dapat menampilkan banyak karakter, sehingga dapat menimbulkan masalah dalam pengambilan keputusan seperti yang pernah dikemukakan oleh Trebuna (2013). Karenanya dibutuhkan suatu teknik analisis data sebagai upaya mengolah data sehingga karakteristik data dapat mudah dipahami dan bermanfaat dalam menjawab permasalahan yang terkait.

Ada beberapa teknik statistik yang dapat digunakan untuk menganalisis data. Tujuannya untuk mendapatkan informasi yang relevan. Teknik statistika yang paling sering digunakan untuk melihat hubungan data adalah Analisis of Variance (ANOVA) untuk data univariat dan atau Multivariat Analysis of variance (MANOVA) untuk data multivariat. Atau lebih dikenal uji statistik t atau uji statistik F.

Namun, menurut bozdogan (1986) usaha analisis data dengan ANOVA ataupun MANOVA tidak informatif. Penolakan hipotesis tidak mengindikasikan bahwa kelompok atau sampel adalah berbeda seperti yang seharusnya ditunjukkan. Hal yang sama juga dikemukakan oleh Almuitari [1]. Menurutnya, salah satu aspek

yang menjadi kontra dalam penelitian adalah tidak adanya tolak ukur pemilihan penggunaan tingkat signifikan α antara 1%, 5%, atau 10% selama uji tes. Permasalahan lain yakni ketika uji statistik pasangan, ada peluang penolakan minimal satu hipotesis dimana itu seharusnya itu tidak perlu terjadi.

Shimokawa (2011) menjelaskan bahwa grafik statistik juga dapat memberikan gambaran perbedaan sampel. Seperti yang pernah dilakukan oleh Tukey pada tahun 1977 yang dalam penelitiannya menggunakan grafik statistik untuk analisis data sampel banyak. Dalam hal ini Tukey merekomendasikan penjelasan data sampel banyak dengan menggunakan analisis kelompok.

Metode Analisis kelompok merupakan salah satu metode statistika multivariat. Metode ini menganalisis data tanpa menggunakan hipotesis melainkan dengan melakukan pengelompokan sampel dengan berdasarkan homogenitas dengan menggunakan algoritma pengelompokan yang dapat memaksimalkan perbedaan relatif kelompok terhadap variasi dalam kelompok. Dua metode umum dalam algoritma pengelompokan adalah metode hirarki dan non hirarki.

Dalam banyak kasus analisis data untuk perbandingan sampel banyak, data diasumsikan berdistribusi normal. Namun, menurut Shimokawa (2011) bukan hal yang mudah untuk memenuhi asumsi tersebut. Walaupun

dengan menggunakan versi nonparametrik, kita tidak perlu memenuhi distribusi normal dan memungkinkan untuk memperoleh informasi yang merupakan bagian dalam data (seperti, posisi, penyebaran dan bentuk). Kemudian Shimokawa mengembangkan analisis data multivariate dengan menggunakan distribusi power normal tanpa harus melakukan uji hipotesis.

PEMBAHASAN

Analisis data mempunyai peranan untuk memahami berbagai macam jenis data. Menurut Lexy J. Moleong (2000) Analisa Data adalah proses mengorganisasikan dan mengurutkan data kedalam pola, kategori, dan satuan uraian dasar sehingga dapat ditemukan tema dan dapat dirumuskan hipotesis kerja seperti yang disarankan oleh data. Analisis data diartikan sebagai upaya mengolah data menjadi informasi, sehingga karakteristik atau sifat-sifat data tersebut dapat dengan mudah dipahami dan bermanfaat untuk menjawab masalah-masalah yang berkaitan dengan kegiatan penelitian.

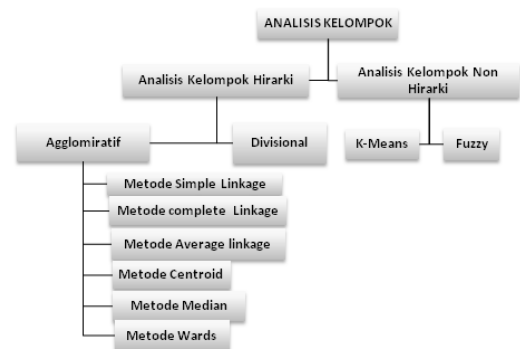
1. Analisis kelompok

Salah satu cara dalam analisis data adalah analisis kelompok. Analisis kelompok mengarah pada metode statistik multivariat. Analisis kelompok didefinisikan sebagai teknik logika umum, prosedur, yang kemudian diikuti pengelompokan variabel objek ke dalam grup kelompok yang berdasarkan pada persamaan atau perbedaan (Trebuna :2013) .

Analisis kelompok mengklasifikasikan objek sehingga setiap objek yang paling dekat kesamaannya akan berada dalam satu kelompok. kelompok yang memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi. Dilihat dari apa yang di kelompokkan , analisis kelompok dibagi atas pengelompokan observasi dan pengelompokan variabel.

Pengelompokan di sajikan dalam bentuk matrik x tipe $n \times p$ dimana n menyatakan nomor objek dan p menyatakan nomor variabel. Pada umumnya ketika objek di kelompokkan , perbedaan antara dua kelompok objek ditandai dengan jarak.

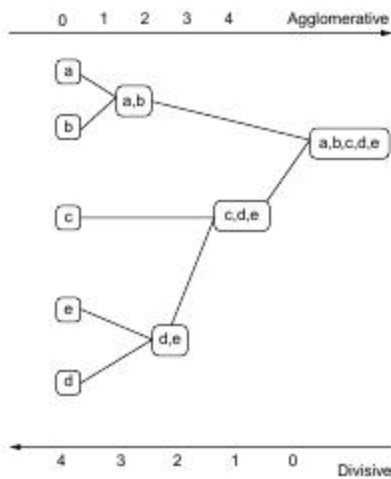
Metode dalam analisis kelompok diklasifikasikan dua bagian yakni metode hirarki dan non hirarki seperti yang tampak pada gambar berikut



Gambar. Klasifikasi Analisis Kelompok

Klasifikasi dari Analisis Kelompok Hirarki dalam metode ini adalah aglomerasi dan

divisional. Dalam metode aglomerasi tiap observasi pada mulanya dianggap sebagai kelompok tersendiri sehingga terdapat sebanyak jumlah observasi. kemudian dua kelompok yang terdekat kesamaannya digabung menjadi suatu kelompok baru, sehingga tiap tahap. Sebaliknya pada metode divisional dimulai dari satu kelompok besar yang mengandung seluruh observasi, selanjutnya yang mendekati kesamaan dipisah dan di bentuk kelompok yang lebih kecil.



Gambar. Alur Analisis Kelompok Hirarki

2. Analisis Kelompok Sampel Banyak.

Analisis kelompok sampel banyak merupakan metode alternatif untuk perbandingan sampel. Permasalahan dari berbagai proses perbandingan bisa dilihat dari pengelompokan rata-rata grup, sampel atau perlakuan. Analisis kelompok sampel banyak pertama kali dikenalkan oleh Bozdogan (1981,1986).

Berbeda dengan analisis kelompok pada umumnya yang menggunakan pendekatan jarak, Bozdogan menggunakan model seleksi kriteria untuk mengenalkan metode analisis kelompok sampel banyak sebagai alternatif perbandingan sampel. Dalam analisis kelompok sampel banyak, sekumpulan grup, sampel atau perlakuan dikelompokkan dalam himpunan yang memiliki kesamaan. Permasalahan ini lebih rumit dibandingkan pengelompokan individu atau objek kedalam satu kasus sampel.

Metode Analisis kelompok sampel banyak merupakan suatu pendekatan baru dan berbeda. Pada pendekatan ini model seleksi kriteria digunakan untuk memilih alternatif kelompok terbaik. Tahapan dalam analisis kelompok sampel banyak adalah mengelompokkan semua alternatif kelompok yang mungkin menggunakan algoritma kombinatorial. Kemudian informasi kriteria digunakan untuk menggolongkan perbedaan tanpa membuat pilihan sendiri selama alternative kelompok dibentuk. Alternatif kelompok dengan nilai informasi kriteria terkecil dipilih.

Penggunaan Kriteria informasi dalam teknik analisis data pernah diterapkan oleh Bozdogan (1986,2000) yang menjelaskan tentang kriteria informasi Akaike's untuk perbandingan sampel ganda. Tujuannya untuk membangun rencana baru untuk informasi yang luas. Konishi dan Kitagawa (1996) dalam penelitiannya menjelaskan bahwa informasi kriteria dibangun

untuk mengevaluasi model yang diperoleh dari berbagai macam cara ketika kelompok yang ditetapkan tidak memiliki distribusi yang membangkitkan data itu. Neath dan Cavanaugh (2006) melakukan pendekatan informasi kriteri bayes untuk menyelesaikan berbagai permasalahan perbandingan data.

3. Menghitung Alternatif Kelompok

Pada kasus sampel banyak setiap objek ditandai dengan ukuran p yang secara bersamaan berada dalam K kelompok. Dapat disajikan dalam bentuk matriks berikut

$$X(n \times p) = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_K \end{bmatrix} \begin{matrix} n_1 \times p \\ n_2 \times p \\ \vdots \\ n_K \times p \end{matrix}$$

Dimana $X_g(n_g \times p)$ mewakili pengamatan dari grup ke-g, $g=1,2,\dots,K$ dan $n = \sum_{g=1}^K n_g$. Tujuannya adalah untuk mengelompokkan K data kedalam k kelompok yang memiliki kesamaan, dengan k belum diketahui dan dapat bermacam-macam namun $k \leq K$.

Jika kita menggunakan teknik enumerasi komplit, kemudian pengelompokan dari K data kelompok kedalam k kelompok yang tak kosong, maka teorema adalah sebagai berikut (Bozdogan:1986):

Teorema 3.1. Banyaknya jalan untuk pengelompokkan K data sampel kedalam k

kelompok dimana $k \leq K$, dan tak satupun dari kelompok k yang kosong, diberikan sebagai berikut:

$$\sum_{g=0}^k (-1)^g \begin{bmatrix} k \\ g \end{bmatrix} (k-g)^K$$

Dimana setiap kelompok k tidak saling relevan.

Dari teorema, total jumlah pengelompokkan K data sampel ke dalama k kelompok himpunan bagian diberikan sebagai berikut:

$$w = S(K, k) = \frac{1}{k!} \sum_{g=0}^k (-1)^g \begin{bmatrix} k \\ g \end{bmatrix} (k-g)^K$$

Ini akan menghasilkam jumlah alternative kelompok. Jika jumlah alternative k tidak diketahui, maka jumlah alternative kelompok diberi sebagai berikut:

$$S(K, k) = kS(K-1, k) + S(K-1, k-1)$$

Dengan

$$S(1,1) = 1$$

$$S(1, k) = 0, k \neq 1$$

$$S(K, 2) = 2^{K-1} - 1$$

Contoh: Misalkan data suatu sampel memiliki $K=3$.Banyaknya alternative pengelompokan k adalah 1,2 dan 3 ($k \leq K$).

Dari teorema, total jumlah pengelompokkan K data sampel ke dalama k kelompok himpunan bagian diberikan sebagai berikut:

No alternatif pengelompokan	Alternatif	Pengelompokan	K
$\sum_{k=1}^{K=3} S(3, k) = 5$	1	(1,2,3)	1
	2	(1,2) (3)	2
	3	(1,3) (2)	2
	4	(2,3) (1)	2
	5	(1)(2)(3)	3

Tabel : Contoh alternative pengelompokan dari K-data sampel kedalam variasi kelompok k yang mungkin

4. Kriteria Informasi

Salah satu langkah penting dalam analisis kelompok sampel banyak setelah pemilihan algoritma adalah penggunaan criteria informasi yang digunakan untuk menggolongkan perbedaan tanpa membuat pilihan sendiri selama alternative kelompok dibentuk. Alternatif kelompok dengan nilai informasi kriteria terkecil yang dipilih.

Penggunaan Kriteria informasi dalam teknik analisis data juga pernah diterapkan oleh Bozdogan (1986,2000) yang menjelaskan tentang kriteria informasi Akaike's untuk perbandingan sampel ganda. Tujuannya untuk membangun rencana baru untuk informasi yang luas.

Konishi dan Kitagawa (1996) dalam penelitiannya menjelaskan bahwa informasi kriteria dibangun untuk mengevaluasi model yang diperoleh dari berbagai macam cara ketika

kelompok yang ditetapkan tidak memiliki distribusi yang membangkitkan data itu. Neath dan Cavanaugh (2006) melakukan pendekatan informasi kriteri bayes untuk menyelesaikan berbagai permasalahan perbandingan data.

Dalam Almutari (2011) disampaikan bahwa kriteria informasi pertama kali diperkenalkan oleh Akaike (1973). Ini memelopori penggabungan teori kemungkinan dan teori informasi untuk menghasilkan pendekatan secara signifikan dan langsung untuk model seleksi statistik. Banyak kriteria informasi telah diperkenalkan diantaranya Akaike Information Criterion (AIC), Consistent Akaike Information Criterion (CAIC), Schwarz Bayesian Criterion (SBC), Information Complexiry Information Criterion (ICOMP).

5. Analisis Kelompok Hirarki Untuk Sampel Banyak

- Analisis kelompok hirarki untuk sampel banyak berdasarkan distribusi normal

Tahapan pertama yang harus dilakukan dalam analisis kelompok sampel banyak adalah membuat algoritma sehingga K kelompok sampel dapat di optimalkan ke arah k himpunan bagian kelompok sampel ($k \leq K$). Kemudian pada masa algoritma berjalan model criteria informasi di masukkan sebagai acuan memilih pasangan kelompok yang memiliki sedikit perbedaan. Pasangan terpilih adalah yang

memiliki nilai kriteria minimum. Pada kasus ini diasumsikan bahwa data berdistribusi normal .

Adapun algoritmanya adalah sebagai berikut:

- Tahap 1 : Mulai dari k=1 alternatif himpunan bagian kelompok. Pada keadaan k =1 semua kelompok sampel berada dalam satu kelompok . Selanjutnya hitung nilai Aikake's Criterion Information (AIC).
- Tahap 2 : lanjut k=2 alternatif himpunan bagian kelompok. Banyak kelompok dapat diperoleh melalui rumus (w) . Kemudian hitung nilai AIC untuk semua alternatif himpunan bagian kelompok yang terbentuk. Kelompok yang terpilih adalah yang memiliki nilai AIC minimum.
- Tahap 3 : Ulangi tahap 2 untuk k=3,4,...K . Artinya hingga semua kelompok sampel terkelompok tunggal.

Pada kejadian ini diasumsikan kasus populasi data sampel banyak berdistribusi normal dengan nilai vektor mean berbeda (μ_g), $g= 1,2,\dots,K$ dan nilai varian (α) yang sama.

Untuk mendapatkan nilai Kriteria Informasi Akaike atau selanjutnya kita kenal dengan AIC diperoleh dengan rumus sebagai berikut:

$$AIC(k) = -2\log L[\hat{\theta}(k)] + 2m(k)$$

Dimana $L[\hat{\theta}(k)]$ adalah kemungkinan fungsi dari observasi, $\hat{\theta}(k)$ adalah nilai perkiraan maksimum dari parameter vektor $\hat{\theta}$ pada model M_k , m_k adalah parameter independen.

Nilai parameter $m = kp + p(p+1)/2$ sehingga rumus AIC sebagai berikut :

$$AIC(k) = -2\log L[\hat{\theta}(k)] + 2 \left[kp + \frac{p(p+1)}{2} \right]$$

Nilai fungsi log likelihood diperoleh sebagai berikut :

$$\begin{aligned} l(\{\mu_g, \sigma^2; X\}) &\equiv \text{Log}(\{\mu_g, \sigma^2; X\}) \\ &= -\left(\frac{np}{2}\right) \log(2\pi) - \left(\frac{n}{2}\right) \log \|\sigma^2\| - 1/2 \text{tr} \sigma^{-1} \\ &\quad - 1/2 \text{tr} \sigma^{-1} \sum_{g=1}^K n_g (\bar{x}_g - \mu_g)(\bar{x}_g - \mu_g) \end{aligned}$$

Dimana

$$\hat{\mu}_g = X_g, g = 1, 2, \dots, K$$

$$\sigma^2 = n^{(-1)}W$$

dengan memasukkan nilai fungsi maksimum likelihood, maka AIC dapat didefinisikan sebagai berikut (digunakan dalam algoritma analisis kelompok sampel banyak)

$$AIC = np \log(2\pi) + n \log \|\sigma^{(-1)}W\|$$

$$+np + 2 \left[kp + \frac{p(p+1)}{2} \right]$$

Dimana

n = jumlah sampel

$\|W\| = \text{DetetrminanMatriks Grup}$

k = Nomor alternatif kelompok

p = Jumlah variabel

Contoh yang akan dibahas disini studi kasus terhadap jenis padi yang datanya

sudah diberikan dalam Srivastava dan Carter (Bozdogan :1986). VArianas jenis

padi terdiri dari empat yang dinotasikan a,b,c dan d dengan setiap jenis teriri dari 5 macam.

Selama 6 minggu dilakukan pengukuran terhadap x1= pertumbuhan tinggi pohon, dan x2= jumlah tangkai per pohon. Ini berarti data diatas memiliki

$$n = \sum n_g = 20, n = 5, g = 1, 2, 3, 4$$

Untuk jenis kelompok varian padi diatas terdiri dari K=4 kelompok sampel yang akan dikelompokkan kedalam k=1,2,3 dan 4 kelompok himpunan bagian. Nilai p=1.Jumlah alternatif kelompok yang mungkin terjadi adalah 15.

No Alternatif pengelompokan	alternatif Kelompok	k	m	AIC
1	(A,B,C,D)	1	5	190.42*
2	(B,C,D)(A)	2	7	187.91
3	(A,C,D)(B)	2	7	190.42

4	(A,B,D)(C)	2	7	190.42
5	(A,B,C)(D)	2	7	190.57
6	(A,D)(B,C)	2	7	192.51
7	(A,C)(B,D)	2	7	182.75*
8	(A,B)(C,D)	2	7	195.70
9	(A,B)(C)(D)	3	9	196.01
10	(A,C)(B),(D)	3	9	185.80*
11	(A,D)(B),(C)	3	9	194.55
12	(B,C)(A),(D)	3	9	189.45
13	(B,D)(A),(C)	3	9	186.33
14	(C,D)(A),(B)	3	9	190.14
15	(A),(B),(C), (D)	4	11	189.24

Catatan :

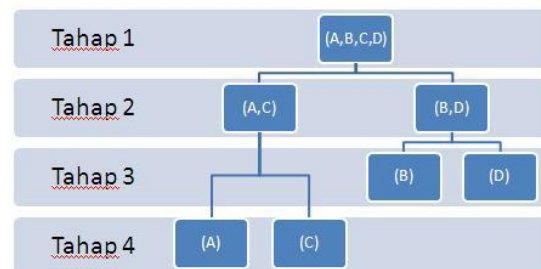
n=20 ; p=2 ; m==kp+p(p+1)/2 parameter

AIC = nploge(2_) + nlogekn(-1)Wk + np + 2m

* merupakan nilai minimum

dari hasil pengamatan tabel diatas dapat dilihat pada saat alternatif kelompok k=2, pasangan yang memiliki homogenitas terbaik adalah yang memiliki nilai AIC minimum yakni pada alternatif kelompok ke 7 untuk pasangan (A,C)(B,D). Berlanjut pada saat alternatif kelompok k=3, pasangan yang memiliki nilai AIC minimum pada alternatif kelompok ke 10 untuk pasangan (A,C),(B),(D).

Secara gambar dapat disajikan sebagai berikut :



Gambar : dendogram pembagian kelompok

- Analisis kelompok hirarki untuk sampel banyak berdasarkan distribusi power normal Pada banyak kasus analisis kelompok untuk perbandingan sampel banyak, observasi diasumsi harus memenuhi distribusi normal. Pada kenyataannya sulit untuk memenuhi asumsi tersebut. Karenanya dikenalkan suatu metode yang tidak memandang kebutuhan distribusi normal, namun menggunakan distribusi power normal yang merupakan distribusi yang ditetapkan sebelum perubahan bentuk normal oleh Box dan Cox tahun 1964 (Ishokawa ; 2012)

Secara umum variabel random x untuk transformasi power normal adalah

$$x^\lambda = \begin{cases} \frac{x^\lambda - 1}{\lambda} , & \lambda \neq 0 \\ \log x , & \lambda = 0 \end{cases}$$

Fungsi densitas probabilitas untuk Distribusi power normal sebagai berikut

$$f_{DPN}(x; \lambda, \mu, \sigma) = \frac{x^\lambda - 1}{A(\lambda)\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x^\lambda - \mu)^2}{2\sigma^2}\right\}$$

$$\hat{\sigma}_n = \sum_{k \in Q_n} \sum_{i=1}^{ki} (x_{ki}^{\hat{\lambda}_n} - \hat{\mu}_n)^2 / \sum_{k \in Q_n} ki$$

dimana μ adalah lokasi parameter , σ^2 adalah skala parameter dan $A(\lambda)$ peluang proporsional dari distribusi power normal yang didefinisikan sebagai berikut

$$A(\lambda) = \begin{cases} \phi[\text{sign}(\lambda)] \|\equiv\| , & \lambda \neq 0 \\ 1 , & \lambda = 0 \end{cases}$$

Dimana $\equiv = \left(\frac{\mu}{\sigma}\right)$ adalah titik pemotongan standar dari pemotongan distribusi normal dan ϕ adalah fungsi distribusi kumulatif dari distribusi normal. Untuk memenuhi kebutuhan bahwa itu power normal dengan menganggap $A(\lambda) \approx 1$ perubahan observasi adalah berdistribusi normal. Dalam mengubah model kriteria informasi ke bentuk power normal, maka dianggap setiap kelompok terdistribusi normal. Adapun bentuk AIC untuk distribusi power normal diberikan sebagai berikut

$$AIC_{DPN}^{(1)} = -2l_{DPN}(\hat{\lambda}_n, \hat{\mu}_n, \hat{\sigma}_n; x) + 2(3K)$$

Dengan

$$l_{DPN}(\hat{\lambda}_n, \hat{\mu}_n, \hat{\sigma}_n; x) = (\lambda_n - 1) \sum_{n=1}^N \sum_{k \in Q_n} \sum_{i=1}^{ki} \log x_{ki} - \log \sigma_n^2$$

n=1,2,...,K

$$\hat{\mu}_n = \sum_{k \in Q_n} \sum_{i=1}^{ki} x_{ki}^{\hat{\lambda}_n} / \sum_{k \in Q_n} ki$$

jika $\lambda_1 = \lambda_2 = \dots = \lambda_K = 1$, $\sigma_1 = \sigma_2 = \dots = \sigma_K = \sigma$ maka dapat diasumsikan bahwa hasil dari analisis kelompok dengan distribusi normal akan sama dengan hasil dari analisis kelompok distribusi power normal.

Algoritma yang akan digunakan juga sama dengan algoritma berdistribusi normal hanya letak perbedaannya pada parameter yang telah berubah bentuk power normal.

Jika pada distribusi normal kita menghitung nilai mean dari kelompok, pada distribusi power normal kita lokasi parameter dan skala parameter. Selanjutnya menghitung nilai AICDPN.

Contoh

Diberikan suatu data yang diambil dari penelitian Zelazo tahun 1972 (Shimokawa 2011) yang melakukan pengamatan terhadap penguatan berjalan dan refleks bayi yang baru lahir. Data ini mengacu pada umur bayi pertama kali jalan. Berikut data pengamatan.

Kelompok	Hasil Pengamatan (Bulan)	Mean
AK	9.00;9.50;9.75;10.00;13.00 ;9.50	10.12 5
PK	11.00;10.00;10.00;11.75;1 0.50; 15.00	11.37 5
NK	11.50;12.00;9.00;11.50;13.	11.70

	25; 13.00	8
EK	13.25;11.50;12.00;13.50;1 1.50	12.35 0

Tabel : Pengamatan awal bayi jalan

Dimana

AK = adalah kelompok bayi aktif latihan menerima stimulasi motorik empat kali sehari selama delapan minggu

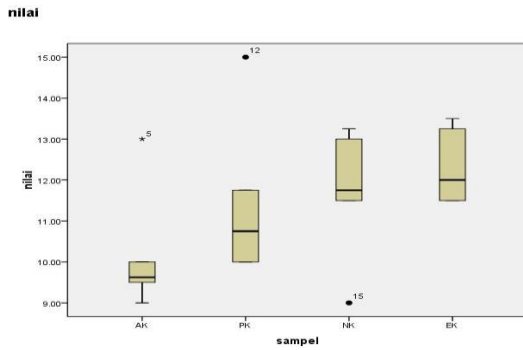
PK = adalah kelompok bayi pasif selama menerima stimulasi motorik yang sama empat kali sehari selama delapan minggu.

NK = adalah kelompok bayi yang tidak menerima latihan tetapi tetap di uji bersama bayi kelompok AK.

EK = adalah kelompok bayi dalam kelas kontrol delapan minggu untuk mengendalikan kemungkinan pemeriksaan ulang.

Dalam kasus ini akan dilakukan analisis kelompok hirarki berdasarkan distribusi normal dan power normal . Tapi sebelumnya akan ditampilkan perhitungan dengan menggunakan uji ANOVA, LSD dan akan kita bandingkan

Pada dasarnya ,perbedaan kelompok dapat dilihat dengan menggunakan tampilan grafik sebagai berikut:



Gambar : Diagram boxplot pengamatan bayi

Dari gambar tampilan grafik boxplot dapat diketahui bahwa bayi pada kelompok AK yang pertama berjalan. Usia bayi kelompok EK dan NK hampir sama.

Hasil yang berbeda akan terlihat apabila perbandingan kelompok sampel dilihat berdasarkan uji ANOVA.

ANOVA

nilai	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	14.778	3	4.926	2.142	.129
Within Groups	43.690	19	2.299		
Total	58.467	22			

Tabel: pengamatan bayi berjalan

Menurut hitungan ANOVA didapat nilai Signifikan > 0,05, maka H0 diterima dan H1 ditolak. Ini berarti seluruh sampel memiliki rata-rata yang sama secara statistik pada tingkat signifikansi 0,05. Tidak tampak ada perbedaan seperti yang seharusnya ditunjukkan ada tabel boxplot.

Multiple Comparisons

nilai LSD

(i) sampel	(j) sampel	Mean Difference (i-j)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
AK	PK	-1.25000	.87549	.170	-3.0824	.5824
	NK	-1.58333	.87549	.086	-3.4158	.2491
	EK	-2.22500*	.91822	.026	-4.1469	-.3031
PK	AK	1.25000	.87549	.170	-.5824	3.0824
	NK	-.33333	.87549	.708	-2.1658	1.4991
	EK	-.97500	.91822	.302	-2.8969	.9469
NK	AK	1.58333	.87549	.086	-.2491	3.4158
	PK	.33333	.87549	.708	-1.4991	2.1658
	EK	-.64167	.91822	.493	-2.5635	1.2802
EK	AK	2.22500*	.91822	.026	-.3031	4.1469
	PK	.97500	.91822	.302	-.9469	2.8969
	NK	.64167	.91822	.493	-1.2802	2.5635

*. The mean difference is significant at the 0.05 level.

Tabel : Pengamatan uji LSD

Namun apabila data diuji menggunakan uji LSD maka di peroleh bahwa kelompok sampel AK dan EK berbeda secara signifikan pada tingkat signifikansi 5%, karena nilai sig <0,05 (pada table pengamatan LSD). Namun tak tampak perbedaan signifikan untuk kelompok lain. Hasil yang diperoleh berbeda dengan perolehan dari perhitungan uji ANOVA.

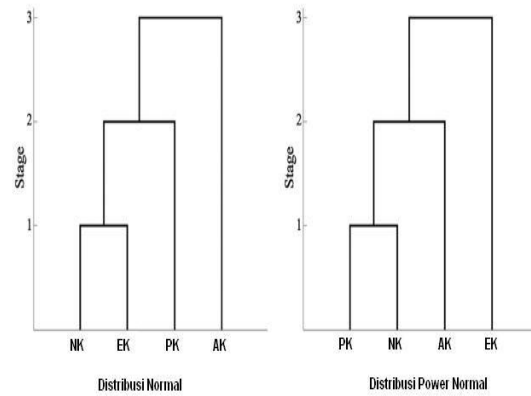
Kemudian diterapkan metode perbandingan untuk kelompok sampel diatas yakni metode analisis kelompok . Metode analisis kelompok dilakukan dengan menggunakan algoritma hirarki berdasarkan distribusi normal dengan menggunakan kriteria informasi AIC.

Hasil yang diperoleh dengan menggunakan analisis kelompok hirarki

berdistribusi normal didapat bahwa Kelompok (NK,PK) adalah yang pertama terkelompokkan (mempunyai homogenitas), di ikuti (AK) dan terakhir (EK). Itu dapat diartikan bahwa metode analisis kelompok hirarki menunjukkan bahwa dalam tahapan belajar berjalan tidak menyarankan adanya pelatihan .Tidak sejalan dengan grafik boxplot yang menunjukkan lebih awal berjalan pada kelompok AK (yang mendapat pelatihan dan stimulasi motorik).

Hasil yang berbeda ditunjukkan melalui metode analisis kelompok hirarki berdasarkan distribusi power normal. (NK,EK) adalah kelompok yang pertama terkelompokkan, kemudian tahap selanjutnya (PK) bergabung menjadi satu kelompok. Dari hasil pengelompokkan berdasarkan distribusi normal sangat mendukung dari grafik boxplot bila ditinjau dari rata-rata kelompok anak awal berjalan.

Berikut disajikan jelas perbedaan pengelompokkan dengan menggunakan dendogram melalui analisis kelompok yang terdistribusi normal dan power normal setelah melalui perhitungan menurut algoritma analisis kelompok sampel banyak



Gambar : Dendogram Analisis kelompok berdasarkan distribusi.

Dari beberapa metode perbandingan kelompok sampel dapat dilihat bahwa metode analisis kelompok hirarki berdistribusi power normal mendapatkan hasil yang lebih mendekati visual grafik boxplot dibandingkan metode analisis kelompok berdistribusi normal. Bahkan hasil perbandingan sampel dengan menggunakan metode perbandingan ANOVA, tidak dapat menunjukkan dengan jelas perbedaan kelompok sampel. Begitu pun hasil yang berbeda ditunjukkan dengan uji LSD .

KESIMPULAN DAN SARAN

Pada umumnya metode analisis data yang digunakan untuk prosedur perbandingan sampel adalah uji t, uji F, ANOVA atau MANOVA. Beberapa peneliti menganggap metode diatas tidak informatif dan memberikan gambaran yang jelas tentang karakteristik sampel.

Maka dikenal suatu pendekatan baru dalam prosedur perbandingan sampel tanpa

harus menguji hipotesis kesamaan pasangan kelompok yakni metode analisis kelompok.

Tahapan awal dalam metode kelompok adalah penentuan algoritma, dimana algoritma yang dipilih disini adalah secara hirarki agglomeratif. Kemudian untuk memilih pasangan kelompok terbaik selama algoritma pengelompokan model kriteria informasi AIC digunakan.

Sebenarnya secara grafik statistik dapat memberikan gambaran perbedaan kelompok sampel, namun dengan uji F atau ANOVA tidak memberikan penjelasan tentang perbedaan sampel. Bahkan tidak menunjukkan perbedaan seperti yang ditampilkan grafik statistik Begitu juga ketika metode analisis kelompok secara hirarki berdistribusi normal diterapkan. Hasil yang berbeda dan sesuai dengan grafik statistik tampak ketika analisis kelompok hirarki untuk sampel banyak berdistribusi power normal. Menurut Shimokawa(2011) karena distribusi power normal dibangun dari fungsi likelihood yang sama dengan distribusi normal, diasumsikan hasil analisis kelompok yang berdistribusi power normal dapat menutupi hasil analisis kelompok distribusi normal.

Analisis kelompok menjadi salah satu prosedur alternatif dalam perbandingan kelompok sampel

banyak. Harapan kedepannya akan ada yang mengevaluasi penggunaan model kriteria informasi yang digunakan dalam analisis kelompok, terutama yang berdistribusi power normal untuk melihat keterkaitan distribusi power normal dan distribusi normal lebih lanjut.

DAFTAR PUSTAKA

- Almuitari, Fahad. (2011). Algorithms for Multi-Sample Cluster Analysis. A Dissertation presented for the doctoral of philosophy degree the university of Tennessee, Knoxville
- Bozdogan, H., (1986). Multi-sample cluster analysis as an alternative to multiple comparison procedures. *Bulletin of Information and Cybernetics*, 22(1-2) :95-130.
- Bozdogan, H., (2000). Akaike's Information Criterion and Recent Developments in information Complexity. *Journal of Mathematical and Psychology*, 44 : 62-91.
- B.S. Everitt., S.Landau., M.Leese., Stahl, Daniel. (2011). *Cluster Analysis*. Willey, London.
- Ferreira, Laura. and D.B.Hitchcok. (2003). A Comparison of Hierarchical Methods for Clustering Functional Data. Department of Statistic University of South Carolina, : -.
- Ishogawa, Naoki. (2012). reditive Performance Bayesian Diagnoses. A Dissertation presented for the doctoral philosophy in engineering Osaka Uninersity
- Konishi, Sadanori. dan Kitagawa, Gensiro., (1996). Generalised Information Criteria In Model Selection. *Bulletin of Information and Cybernetics*, 22(1-2):95-130
- Neath, Andrew. dan Casanaugh, Joseph., (2006). A Bayesiayn Approach to Multiple Comparisons Problem. *Journal of data sains*, 4:131-146
- Shimokawa, Thosio. Goto, Massashi., (2011). Hierarchical Cluster Analysis for Multi-Sample Comparison Based on the Power Normal Distribution. *Behaviormatrika*, 36(2) : 125-138.
- Saracli, Sinan., Dogan, Nurhan., Dogan, Ismet. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation *Journal of Inequalities and applications* , 203.
- Trebuna, Peter. and Halcinova, Jana. (2013). *Mathematical Tools of Cluster Analysis*. Scientific Reasearsch, 4 : 814-816.
- Xu, Rui. dan Wunsch, Donald. (2005). Surveys of Clustering Algorithms. *IEEE Transaction of Neural Networks*, 16(3):645-678