

**Statistik Terpangkas F_t menggunakan kriteria pemangkasan berdasarkan
penganggar skala teguh MAD_n**

Abstrak: Ketaknormalan data dan ketakhomogenan varians merupakan masalah yang sering dihadapi semasa pengujian hipotesis dijalankan. Kewujudan masalah ini akan memberi kesan terhadap ralat Jenis I. Salah satu cara bagi menangani kedua-dua masalah tersebut adalah dengan menggunakan kaedah teguh. Dalam kajian ini, kami telah mengubahsuai statistik F_t yang diperkenalkan oleh Lee dan Fung (1985) menggunakan min terpangkas. Min terpangkas ini bergantung kepada suatu kriteria pemangkasan (trimming criterion) yang dapat memangkas nilai-nilai ekstrim tanpa memerlukan penentuan awal (priori) peratus pemangkasan. Oleh yang demikian, penggunaan kriteria baru ini dapat mengelakkan pemangkasan yang tidak perlu. Kriteria pemangkasan yang diperkenalkan dalam kajian ini menggunakan penganggar skala teguh MAD_n . Kertas ini akan membincangkan secara terperinci tentang statistik F_t yang telah diubahsuai dan diilustrasikan dengan satu contoh.

1. Pengenalan

Statistik merupakan satu cabang ilmu yang mempunyai aplikasi meluas dalam kehidupan manusia. Walaupun demikian, terdapat beberapa isu yang sering dipertikaikan semasa penggunaan aplikasi ini. Antara perkara yang menjadi perhatian adalah masalah ketaknormalan data serta masalah ketakhomogenan varians yang merupakan syarat utama yang perlu dipenuhi apabila menjalankan analisis menggunakan kaedah statistik tradisional, terutamanya analisis varians (ANOVA). Kewujudan kedua-dua keadaan ini akan menyebabkan ANOVA yang biasa tidak lagi sesuai digunakan. Sekiranya kedua-dua masalah tersebut berlaku pada masa yang sama ia akan memberi kesan terhadap kebarangkalian ralat Jenis I serta kuasa bagi ujian yang dijalankan (Wilcox & Keselman, 2003). Apabila ralat jenis ini berlaku, maka ia akan menyebabkan hipotesis nol (H_0) yang benar terpaksa ditolak. Keadaan ini sekaligus akan menyebabkan pengurangan kuasa bagi ujian tersebut.

Pelbagai kajian telah dijalankan bagi mencari penyelesaian terhadap permasalahan ini. Salah satu cara bagi menangani masalah-masalah tersebut adalah dengan menggunakan kaedah teguh. Kaedah teguh merupakan kaedah yang paling sesuai digunakan kerana kaedah ini menggabungkan kedua-dua kaedah berparameter dan kaedah tidak berparameter (Hampel, Ronchetti, Rousseeuw & Stahel, 1986). Kaedah teguh ini secara umumnya tidak sensitif kepada andaian terhadap sifat keseluruhan data.

Perkataan teguh atau robos mempunyai pelbagai makna yang adakalanya tidak konsisten antara satu sama lain (Huber, 1981). Definisi bagi perkataan ini bergantung kepada pemahaman dan matlamat seseorang penyelidik semasa melakukan pengujian. Huber (1981) mendefinisikan keteguhan sebagai suatu keadaan yang tidak sensitif terhadap perubahan kecil yang berlaku kepada andaian yang dibuat, manakala Garthwaite, Jolliffe dan Jones (2002) pula menambahkan hanya sedikit sahaja perubahan yang akan berlaku akibat dari data pencilan pada takrifan Huber. Steven (1990) pula beranggapan, sesuatu ujian statistik akan dianggap teguh apabila aras keertian, α sebenar iaitu peratusan penolakan yang salah apabila satu atau lebih andaian tidak dipenuhi adalah sangat hampir dengan α nominal iaitu aras yang telah ditetapkan oleh penyelidik. α nominal juga adalah peratusan penolakan yang salah apabila diketahui bahawa hipotesis nol adalah benar dan kesemua andaian dipenuhi. Statistik teguh ini telah digunakan sejak bertahun-tahun yang lalu (Stigler, 1973) namun yang demikian, tiada kajian secara terperinci dilakukan oleh ahli statistik sehinggalah beberapa tahun kebelakangan ini (Staudte & Sheather, 1990).

Kaedah teguh secara umumnya tidak sensitif kepada andaian terhadap sifat keseluruhan data. Kaedah ini dapat menggantikan kaedah statistik yang menggunakan pengangkar kuasa dua terkecil yang menjadi tidak cekap apabila wujud masalah-masalah yang dinyatakan sebelum ini. Dalam kaedah ini, min dan varians akan digantikan dengan min terpankask dan varians Winsorized (Winsorized variance). Kedua-dua ukuran ini adalah tidak sensitif kepada kenormalan data dan keadaan varians yang tidak sama. Malah Wilcox, Keselman dan Kowalchuk (1998)

menyatakan bahawa kuasa ujian statistik tidak dipengaruhi oleh kenormalan yang menggunakan min terpankaskan dan varians Winsorized.

Keselman et al., (2004) di dalam artikelnya menyatakan bahawa kawalan yang baik terhadap ralat Jenis I dapat diperolehi dengan menggunakan peratus pemangkasan yang sederhana iaitu 15% atau 10% dari setiap ekor taburan. Walau bagaimanapun, Wilcox (1998) berdasarkan kajiannya menulis bahawa apabila lebih banyak pemangkasan dilakukan, maka lebih kurang kesan terhadap kepencongan. Berdasarkan kajian literatur, nilai yang paling optimum bagi penentuan awal nilai pemangkasan adalah antara 0 dan 0.25. Nilai yang terbaik adalah 0.20 bagi setiap ekor taburan (Wilcox dan Keselman, 2003). Terdapat pelbagai cadangan tentang nilai pemangkasan yang patut dilakukan sekiranya data adalah terpencong. Namun tiada suatu kaedah yang dapat memastikan nilai tersebut dapat memberikan keputusan yang baik. Selain itu, berdasarkan penentuan awal peratus pemangkasan, terdapat juga data yang dikehendaki tetapi terpankaskan. Oleh yang demikian, kajian ini mencadangkan satu kaedah pemangkasan yang tidak memerlukan penentuan awal peratus pemangkasan dan hanya memangkas data yang tidak diperlukan sahaja (outliers).

Kajian ini memfokuskan kepada pengubahsuaian statistik F_t yang diperkenalkan oleh Lee dan Fung (1985) menggunakan min terpankaskan biasa kepada suatu kriteria pemangkasan (trimming criterion) yang dapat memangkas nilai-nilai ekstrim tanpa memerlukan penentuan awal (priori) peratus pemangkasan. Kriteria pemangkasan yang diperkenalkan dalam kajian ini akan menggunakan penganggar skala teguh MAD_n . Penganggar skala teguh MAD_n dipilih berdasarkan nilai titik

musnah yang tinggi serta fungsi pengaruh terbatas. Kedua-dua ini merupakan asas kepada penilaian keteguhan (Wilcox, 2005).

2. Kaedah

Kajian ini menggunakan penganggar skala MAD_n sebagai kriteria pemangkasan bagi statistik F_t .

2.1 Penganggar Skala MAD_n

Ukuran skala adalah kuantiti yang dapat menerangkan tentang penyebaran bagi sesuatu taburan. Semasa proses pencarian ukuran bagi skala, nilai titik musnah (breakdown point) merupakan faktor penting yang perlu dipertimbangkan (Wilcox, 2005). Pada tahun 1993, Rousseeuw dan Croux telah memperkenalkan penganggar-penganggar yang mengambilkira nilai titik musnah ini. MAD_n merupakan antara penganggar skala terbaik dan sering digunakan kerana penganggar ini amat kurang dipengaruhi oleh nilai-nilai ekstrim.

MAD_n adalah median bagi sisihan mutlak (median absolute deviation) bagi median. Bagi sebarang data $y_1, y_2, \dots, y_n$,

$$MAD_n = b \operatorname{med}_i |y_i - \operatorname{med}_j y_j|$$

di mana $b = 1.482$ diperlukan untuk menjadikan penganggar konsisten bagi parameter yang diminati.

Bagi sebarang statistik teguh, secara teori, statistik ini dinilai oleh dua faktor iaitu nilai titik musnah dan fungsi pengaruh (influence function). Fungsi pengaruh perlu dibatasi dan nilai titik musnah perlu tinggi. MAD_n memenuhi kedua-dua syarat tersebut (Rousseeuw & Croux, 1993). Walau bagaimanapun, terdapat juga beberapa kelemahan bagi penganggar skala ini. Di antara kelemahannya adalah, kecukupan bagi MAD_n adalah sangat rendah iaitu hanya sebanyak 37% berdasarkan taburan Gauss jika dibandingkan dengan penganggar skala yang lain.

2.2 Kaedah F_t

Lee dan Fung (1985) memperkenalkan suatu statistik yang mampu untuk menangani masalah lokasi sampel apabila varians bagi populasi adalah berbeza. Statistik ini dinamakan statistik terpankaskan, F_t . Mereka juga mencadangkan penggunaan statistik baru ini bagi menangani masalah yang melibatkan ANOVA sehala dan menjadi kaedah alternatif kepada kaedah F tradisional. Selain itu, kaedah ini juga mudah untuk diprogramkan.

Anggapkan $X_{(1)j}, X_{(2)j}, \dots, X_{(n_j)j}$ adalah sampel yang disusun secara menaik bagi kumpulan ke- j dengan saiz sampel n_j . Maka min terpankaskan bagi kumpulan ke- j ditakrifkan seperti berikut,

$$\bar{X}_{tg} = \left[\sum_{j=k+1}^{n-k} X_j + (1-\varepsilon)(X_k + X_{n-k+1}) \right] / (n - 2gn), \quad \text{di mana } \varepsilon = gn - [gn].$$

Jumlah sisihan kuasa dua bagi g -Winsorized

$$SSD_{tg} = \sum_{j=k+1}^{n-k} (X_j - \bar{X}_{tg})^2 + k \left\{ [(1-\varepsilon)X_k + \varepsilon X_{k+1} - \bar{X}_{tg}]^2 + [(1-\varepsilon)X_{n-k+1} + \varepsilon X_{n-k} - \bar{X}_{tg}]^2 \right\}$$

Seterusnya statistik F_t diberikan oleh

$$F_t(j) = \frac{\sum_{j=1}^J (\bar{X}_{tj} - \bar{X}_t)^2 / (J-1)}{\sum_{j=1}^J SSD_{tj} / (H-J)}$$

di mana

$$J = \text{bilangan kumpulan, } h_j = n_j - g_{1j} - g_{2j}, \quad H = \sum_{j=1}^J h_j \quad \text{dan} \quad \bar{X}_t = \sum_{j=1}^J h_j \bar{X}_{tj} / H.$$

Berdasarkan prinsip penggantian (substitution principle), bentuk taburan bagi statistik

F_t adalah mengikut bentuk taburan statistik F asal iaitu $\sim \chi^2, F$

2.3 Pengkaedahan Baru

$$\bar{X}_{tj} = \frac{1}{(n_j - g_1 - g_2)_j} \left[\sum_{i=(g_1+1)_j}^{(n_j-g_2)_j} X_{(i)j} \right]$$

di mana

$$g_{1j} = \text{bilangan pemerhatian } X_{(i)j} \text{ bagi } \left(X_{(i)j} - \hat{M}_j \right) < -2.24 MAD_{n_j},$$

$$g_{2j} = \text{bilangan pemerhatian } X_{(i)j} \text{ bagi } \left(X_{(i)j} - \hat{M}_j \right) > 2.24 MAD_{n_j},$$

dan $\hat{M}_j$ = median bagi kumpulan ke- j .

Sekiranya $g_{1j} = g_{2j} = g$, maka jumlah sisihan kuasa dua Winsorized bagi pemangkasan simetri adalah,

$$SSD_{ij} = (g_j + 1)(X_{(g+1)j} - \bar{X}_{tj})^2 + (X_{(g+2)j} - \bar{X}_{tj})^2 + \dots + (X_{(n_j-g-1)j} - \bar{X}_{tj})^2 + (g_j + 1)(X_{(n_j-g)j} - \bar{X}_{tj})^2$$

Manakala jumlah sisihan kuasa dua Winsorized bagi pemangkasan yang tidak simetri, yakni $g_{1j} \neq g_{2j}$ diberikan seperti berikut,

$$SSD_{ij} = (g_{1j} + 1)(X_{g_{1j}+1} - \bar{X}_{tj})^2 + (X_{g_{1j}+2} - \bar{X}_{tj})^2 + \dots + (X_{n_j-g_{2j}-1} - \bar{X}_{tj})^2 + (g_{2j} + 1)(X_{n_j-g_{2j}} - \bar{X}_{tj})^2 - \left\{ (g_{1j})[X_{g_{1j}+1} - \bar{X}_{tj}] + (g_{2j})[X_{n_j-g_{2j}} - \bar{X}_{tj}] \right\}^2 / n_j$$

Seterusnya statistik F_t diberikan oleh

$$F_t(j) = \frac{\sum_{j=1}^J (\bar{X}_{tj} - \bar{X}_t)^2 / (J-1)}{\sum_{j=1}^J SSD_{ij} / (H-J)}$$

di mana

$$J = \text{bilangan kumpulan, } h_j = n_j - g_{1j} - g_{2j}, \quad H = \sum_{j=1}^J h_j \quad \text{dan} \quad \bar{X}_t = \sum_{j=1}^J h_j \bar{X}_{tj} / H$$

3. Contoh

Data yang digunakan sebagai ilustrasi untuk kaedah ini diberikan di dalam Jadual 1. Terdapat tiga kumpulan yang mempunyai saiz sampel yang berbeza iaitu $n_1 = 15$, $n_2 = 10$ serta $n_3 = 10$.

Jadual 1 : Data Contoh

Kumpulan	Data											
1	26	34	46	48	42	49	100	61	51	53	26	34
2	51	50	33	28	47	50	48	60	91	42		
3	52	64	19	54	58	53	77	56	63	59		

Ujian kenormalan data juga dijalankan terhadap data contoh. Ujian kenormalan yang dijalankan adalah ujian Kolmogorov-Smirnov (KS) serta Shapiro Wilk (SW). Berdasarkan ujian-ujian tersebut didapati bahawa set data bagi ketiga-tiga kumpulan adalah tidak normal. Ini dipamerkan di dalam Jadual 2. Selain itu, Jadual 2 juga menyenaraikan bilangan data yang perlu dipangkas bagi setiap kumpulan untuk dua kaedah F_t yang digunakan.

Jadual 2 : Ujian kenormalan data serta bilangan data yang dipangkas

Kumpulan	Normal (KS, SW)	F_t & MAD_n		F_t (per=0.2)	
		g_1	g_2	g_1	g_2
1	(Tidak, Tidak)	2	1	3	3
2	(Tidak, Tidak)	2	1	2	2
3	(Tidak, Tidak)	1	1	2	2

Di dalam ilustrasi ini kami membuat perbandingan terhadap kaedah F_t yang telah diubahsuai dengan menggunakan menggunakan penganggar skala MAD_n bagi menentukan bilangan nilai-nilai ekstrim yang perlu dipangkas, dengan kaedah F_t asal yang menggunakan peratus pemangkasan yang tetap iaitu sebanyak 20%. Dalam

kedua-dua jadual, Jadual 2 dan Jadual 3, kaedah pertama diwakili dengan singkatan F_t & MAD_n . Manakala kaedah kedua diwakili oleh F_t (per = 0.2). Jadual 3 juga menyenaraikan nilai min terpangkas, nilai jumlah sisihan kuasa dua Winsorized, nilai F_t serta nilai ralat Jenis I yang diperolehi.

Bagi mengetahui kaedah pengujian yang tidak sensitif kepada andaian terhadap sifat keseluruhan data, kriteria liberal Bradley (1978) dijadikan pengukuran tahap keteguhan sesuatu pengujian. Berdasarkan kriteria ini, sesuatu pengujian dikatakan teguh apabila kadar empirikal bagi ralat Jenis I ($\hat{\alpha}$) berada di antara julat $0.5\alpha \leq \hat{\alpha} \leq 1.5\alpha$. Oleh yang demikian, bagi aras keertian sebanyak 5% yang digunakan dalam ilustrasi ini, pengujian statistik yang dijalankan dikatakan teguh apabila kadar empirikal ralat Jenis I berada dalam julat $0.025 \leq \hat{\alpha} \leq 0.075$. Di dalam contoh ini, nilai ralat Jenis I bagi kaedah F_t yang telah diubahsuai adalah 0.0284 manakala 0.0205 bagi kaedah F_t yang asal.

Jadual 3 : Keputusan

Kaedah	Keputusan							
	Min terpangkas			Jumlah sisihan kuasa dua Winsorized			F_t	Ralat Jenis I
	Kump 1	Kump 2	Kump 3	Kump 1	Kump 2	Kump 3		
F_t & MAD_n	46.167	49.714	57.375	1145.73	399.600	212.500	4.148	0.0284
F_t (per=0.2)	45.111	48.000	57.167	685.296	144.000	169.611	4.858	0.0205

Berdasarkan kriteria liberal Bradley didapati bahawa nilai yang diperolehi bagi kaedah yang menggunakan penganggar skala teguh MAD_n berada dalam julat yang ditetapkan. Ini menunjukkan bahawa kaedah tersebut adalah tidak sensitif atau teguh terhadap andaian data. Manakala nilai yang diperolehi berdasarkan kaedah F_t yang asal adalah lebih kecil serta tidak berada dalam julat kriteria Bradley. Maka kajian ini memberi petunjuk bahawa penggunaan kaedah baru ini dapat memberikan kawalan ralat Jenis I yang lebih baik berbanding dengan kaedah asal.

Penghargaan

Penulis-penulis ingin merakamkan penghargaan bahawa penyelidikan yang menghasilkan penerbitan ini dibiayai oleh dana Fundamental Research Grant Scheme, Kementerian Pendidikan Tinggi, Malaysia.

Rujukan

1. Bradley, J.V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, **31**, 321 - 339.
2. Garthwaite, P. H., Jolliffe, I. T., & Jones, B. (2002). *Statistical inference (2nd ed.)*. Oxford, United Kingdom: Oxford University Press.
3. Lee, H., & Fung, K. Y. (1985). Behaviour of trimmed F and sine-wave F statistics in one-way ANOVA. *Sankhya: The Indian Journal of Statistics*, **47**, 186 - 201.
4. Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics*. New York: Wiley.
5. Huber, P. J. (1981). *Robust statistics*. New York: Wiley.
6. Keselman, H. J., Othman, A. R., Wilcox, R. R., & Fradette, K. (2004). The New and Improved Two-Sample t Test. *American Psychological Society*, **15**, 47-51.
7. Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, **88**, 1273 – 1283.
8. Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation and testing*. New York: Wiley.

9. Stevens, J.P. (1990). *Intermediate statistics: A modern approach*. Hillsdale, NJ: Erlbaum.
10. Stigler, S. M. (1973). Simon Newcombe, Percy Daniell, and the history of estimation 1885-1920. *Journal of the American Statistical Association*, **68**, 872-879.
11. Wilcox, R. R. (1998). The Goal and Strategies of Robust Methods. *British Journal of mathematical and Statistical Psychology*, 51, 1 – 39.
12. Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.
13. Wilcox, R. R., Keselman, H. J., & Kowalchuk, R. K. (1998). Can tests for treatment group equality be improved?: The bootstrap and trimmed means conjecture. *British Journal of Mathematical and Statistical Psychology*, **51**, 123-134.
14. Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: Measures of central tendency. *Psychological Methods*, 8, 254-274.

Menemui Matematik
(Discovering Mathematics),
2007, Vol. 29(1): 21 - 28

Zahayu Md Yusof,

Fakulti Sains Kuantitatif,

Universiti Utara Malaysia, 06010 Sintok, Kedah.

zahayu@uum.edu.my

Abdul Rahman Othman,

Pusat Pengajian Pendidikan Jarak Jauh,

Universiti Sains Malaysia, 11800 USM, Penang.

oarahman@usm.my

Sharipah Soaad Syed Yahaya,

Fakulti Sains Kuantitatif,

Universiti Utara Malaysia, 06010 Sintok, Kedah.

sharipah@uum.edu.my