MATEMATIKA, 2007, Volume 23, Number 1, 41–50

Penganggar Teguh Parameter Skala bagi Taburan Normal-Separuh
(Robust Estimators for the Scale Parameter of the Half-Normal Distribution)

[1]**Abdul Aziz Jemain, Marina Zahari & Wan Zawiah Wan Zin**
Pusat Pengajian Sains Matematik, Fakulti Sains dan Teknologi
Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
e-mail: [1]azizj@pkrisc.cc.ukm.my

Abstrak Kertas ini mengemukakan beberapa penganggar teguh parameter skala taburan normal separuh. Penganggar teguh yang dicadangkan adalah berdasarkan gabungan linear statistik tertib. Kecekapan setiap penganggar teguh berkenaan dikaji dengan membandingkannya dengan penganggar saksama varians minimum yang diperoleh daripada kaedah kebolehjadian maksimum, dan penganggar saksama kaedah momen. Didapati penganggar kebolehjadian maksimum mudah terjejas dengan kehadiran pencilan berbanding dengan penganggar kaedah momen. Kajian ini juga mendapati yang min kuasa dua ralat penganggar statistik tertib jauh lebih kecil berbanding dengan min kuasa dua ralat bagi kedua-dua penganggar kebolehjadian maksimum dan momen apabila data dicemari oleh pencilan. Keputusan yang kami peroleh menunjukkan penganggar gabungan linear statistik tertib lebih teguh dan mempunyai kecekapan yang tinggi dengan titik musnah 15%. Kecekapan relatifnya melebihi dua kali ganda lebih tinggi berbanding penganggar kebolehjadian maksimum.

Katakunci Statistik teguh; kecekapan; titik musnah; normal separuh.

Abstract *This paper proposed new robust estimators for scale parameter of half-normal distribution. The suggested estimators are based on linear combinations of order statistics. The efficiency of these estimators is compared with the unbiased minimum variance estimator of maximum likelihood method and unbiased method of moment. The results showed that the maximum likelihood estimator performed worse than the moment estimator in the presence of outliers. The study also showed that the mean square errors of the proposed estimators are much smaller than those obtained by maximum likelihood and moment methods. The results clearly indicate that estimators derived using order statistics are superiorly more robust and have high efficiency with breakdown point 15%. The relative efficiency is more than two times higher than the maximum likelihood estimator.*

Keywords *Robust statistics; efficiency; breakdown point; half-normal.*

1 Pengenalan

Kualiti bagi anggaran yang diperoleh daripada suatu penganggar bergantung kepada sejauhmana data yang diguna mematuhi andaian yang diperlukan oleh penganggar berkenaan. Kegagalan sebahagian kecil data memenuhi andaian penganggar boleh menyebabkan anggaran yang diperoleh pincang. Penganggar yang mudah dipengaruhi oleh gelagat segelintir data yang mempunyai nilai yang jauh menyimpang daripada kebanyakan data dikatakan penganggar tidak teguh. Data yang mempunyai nilai yang jauh berbeza daripada kebanyakan data yang lain pula dikenali sebagai pencilan. Data yang mengandungi pencilan dikatakan data tercemar, Barnett dan Lewis [1] memberikan huraian terperinci berkenaan masalah data terpencil.

Kehadiran pencilan boleh disebabkan oleh banyak faktor, antaranya akibat kesilapan mencatat, kecacatan dalam pengendalian ujikaji, atau sifat alami dalaman data tersebut. Bagi pencilan yang diakibatkan oleh kesilapan mencatat, pembaikan boleh dilakukan dengan penyemakkan tetapi bagi punca yang lain pembetulan sukar untuk dilakukan. Dalam khasus sedemikian, usaha mengatasi pencilan dilakukan dengan menggunakan penganggar teguh. Perbincangan berhubung dengan penganggar teguh bukan suatu yang baru, buku oleh Hoaglin et al. [2] dan Huber [3] merupakan antara bahan yang memberikan liputan yang lengkap berhubung dengan penganggar teguh.

Meskipun tumpuan kebanyakan perbincangan tentang penganggar teguh adalah berkenaan penganggar lokasi dan skala, namun huraian lebih terperinci tertumpu kepada penganggar lokasi. Keadaan sedemikian adalah kerana parameter lokasi lebih banyak digunakan berbanding parameter skala. Kebelakangan ini perbahasan berkenaan parameter skala mula mengambil tempat, antaranya kertas oleh Park dan Lindsay [4] dan Wang dan Suter [5].

Kertas ini akan memberikan perhatian tentang penganggar parameter skala bagi taburan normal separuh. Taburan normal separuh sering digunakan dalam penganggaran bilangan haiwan pada suatu kawasan menggunakan kaedah garis transek, untuk huraian lanjut penerapan taburan ini rujuk Buckland et al. [6] dan Burnham et al. [7]. Penganggar parameter skala untuk taburan normal separuh biasanya diperoleh dengan menggunakan kaedah kebolehjadian maksimum. Meskipun penganggar yang terhasil saksama dan variansnya minimum, tetapi tidak teguh. Penganggar kebolehjadian maksimum yang diperoleh memberikan anggaran yang pincang dan variansnya meningkat apabila terdapat nilai cerapan yang terpencil.

Dalam kertas ini, pelbagai pilihan statistik teguh yang boleh digunakan telah diselidiki. Seterusnya perbandingan dilakukan untuk meneliti kecekapan setiap penganggar dengan penganggar kebolehjadian maksimum dan penganggar kaedah momen. Usaha mendapatkan penganggar teguh bagi taburan normal separuh, khususnya dalam konteks kaedah garis transek masih belum diselidiki. Kebanyakan kerja lepas cenderung untuk mendapatkan penganggar berparameter model Pareto dan model eksponen (lihat Brazauskas dan Serfling [8] yang menggunakan kaedah kuantil statistik), dan penganggar tidak berparameter (Huang [9] dan Huang dan Brill [10] membincangkan penganggar kuantil bukan kernel berdasarkan statistik tertib menggunakan fungsi empirik pintasan-tahap). Dalam kertas ini, tumpuan diberikan untuk mendapatkan penganggar teguh parameter skala bagi taburan normal separuh. Penganggar teguh yang dikemukakan dalam kajian ini adalah berdasarkan statistik tertib. Ianya saksama secara asimptot, teguh dan mempunyai kecekapan relatif lebih 70% berbanding penganggar kebolehjadian maksimum.

Bahan yang disajikan dalam kertas ini dimulai dengan memperkenalkan taburan normal separuh dan penganggar parameter skala yang diperoleh dengan kaedah kebolehjadian maksimum dan momen. Setelah itu huraian mengenai statistik tertib diberikan, diikuti dengan huraian berkenaan statistik tertib dalam taburan normal separuh. Seterusnya pelbagai penganggar teguh bagi skala taburan normal separuh dibincangkan. Setelah mengemukakan beberapa penganggar teguh, kualiti setiap penganggar tersebut diselidiki dengan lebih lanjut. Setiap penganggar yang dicadangkan diselidiki kecekapan relatifnya dengan membandingkan varians penganggarnya terhadap varians penganggar kaedah kebolehjadian maksimum.

2 Taburan Normal Separuh dan Penganggar Sisihan Piawai

Katalah terdapat suatu sampel rawak n cerapan, $x_1, x_2, \ldots, x_n$ dari populasi normal separuh dengan fungsi ketumpatan

$$f(x;\sigma) = \frac{2}{\sigma\sqrt{2\pi}} \text{eksp}\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right), \quad x > 0$$

Penganggar parameter skala, σ, boleh diperoleh dengan kaedah kebolehjadian maksimum atau kaedah momen. Berasaskan kaedah kebolehjadian maksimum, penganggar σ bersamaan dengan

$$\hat{\sigma}_K = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \tag{1}$$

Penganggar ini saksama secara asimptot, manakala varians penganggar dengan mudah dapat ditunjukkan bersamaan dengan

$$Var(\hat{\sigma}_K) = \frac{\sigma^2}{2n} \tag{2}$$

Berdasarkan momen pertama, penganggar σ bersamaan dengan

$$\hat{\sigma}_M = \bar{x}\sqrt{\frac{\pi}{2}} \tag{3}$$

dengan $\bar{x}$ merupakan min sampel. Penganggar ini juga saksama dan dapat ditunjukkan bahawa variansnya bersamaan dengan

$$Var(\hat{\sigma}_M) = \left(\frac{\pi - 2}{2}\right)\frac{\sigma^2}{n} \simeq 0.5707\frac{\sigma^2}{n} \tag{4}$$

Berdasarkan persamaan (2) dan (4), $Var(\hat{\sigma}_K) < Var(\hat{\sigma}_M)$. Oleh itu $\hat{\sigma}_K$ adalah lebih cekap daripada $\hat{\sigma}_M$ dengan $\hat{\sigma}_M$ adalah 87.6% cekap berbanding dengan $\hat{\sigma}_K$. Penganggar berdasarkan kebolehjadian maksimum $\hat{\sigma}_K$ juga adalah saksama secara asimptot dan mempunyai varians minimum. Namun demikian kedua-dua penganggar tidak teguh dan amat dipengaruhi oleh kehadiran data terpencil. Nilai pencilan akan menyebabkan kedua-dua penganggar memberikan anggaran yang berlebihan.

3 Statistik Tertib Data Normal Separuh

Bagi mengatasi masalah terlebih anggar akibat data pencilan, suatu penganggar yang lebih teguh diperlukan. Usaha mendapatkan penganggar teguh dapat dilakukan dengan memanfaatkan statistik tertib. Berikut akan diberikan beberapa sifat penting bagi suatu statistik tertib.

Katalah $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$ mewakili sampel rawak n cerapan yang telah ditertibkan secara tidak menyusut. Jika $X_{r:n}$ mewakili pembolehubah rawak cerapan yang ke-r terbesar daripada n cerapan, tertib ke-r, David dan Nagaraja [11] memberikan hubungan berikut

$$E(X_{r:n}) \simeq Q_r$$

$$Kov(X_{r:n}, X_{s:n}) \simeq \left(\frac{p_r q_s}{f(Q_r)f(Q_s)}\right)\frac{1}{n}$$

dengan Q_r nilai kuantil $p_r = r/(n+1)$, $q_r = 1 - p_r$ dan $r \leq s$. Dapat ditunjukkan dengan mudah, bagi cerapan daripada fungsi ketumpatan normal separuh dengan varians σ^2, hubungan berikut sah,

$$E(X_{r:n}) \simeq Q_r = \sigma z_{\bar{r}} \tag{5}$$

$$Var(X_{r:n}) \simeq \left(\frac{p_r q_r}{\phi^2(z_{\bar{r}})}\right)\frac{\sigma^2}{4n} \tag{6}$$

$$Kov(X_{r:n}, X_{s:n}) \simeq \left(\frac{p_r q_s}{\phi(z_{\bar{r}})\phi(z_{\bar{s}})}\right)\frac{\sigma^2}{4n} \tag{7}$$

dengan $z_{\bar{r}} = \Phi^{-1}(\bar{r})$ dan $\bar{r} = \frac{1}{2}(p_r + 1)$. Manakala $\Phi(z)$ dan $\phi(z)$ setiap satunya adalah fungsi taburan kumulatif dan fungsi ketumpatan normal piawai.

4 Penganggar Teguh Parameter Skala

Terdapat banyak penganggar teguh berasaskan statistik tertib yang boleh diterbitkan. Penganggar tersebut boleh berdasarkan sebarang satu cerapan pada tertib tertentu atau sebarang gabungan beberapa statistik tertib. Sekiranya penganggar yang hendak digunakan berasaskan satu cerapan pada tertib ke-r maka

$$\hat{\sigma}_{r:n} = \frac{x_{r:n}}{z_{\bar{r}}} \tag{8}$$

Penganggar ini saksama iaitu boleh ditunjukkan dengan memanfaatkan persamaan (5), dan mempunyai varians bersamaan dengan

$$Var(\hat{\sigma}_{r:n}) = \left(\frac{p_r q_r}{4z_{\bar{r}}^2\phi^2(z_{\bar{r}})}\right)\frac{\sigma^2}{n} \tag{9}$$

Untuk penganggar gabungan dua nilai statistik tertib, penganggar berikut boleh digunakan

$$\hat{\sigma}_{r,s:n} = \frac{x_{r:n}z_{\bar{r}}^{\alpha-1} + x_{s:n}z_{\bar{s}}^{\alpha-1}}{z_{\bar{r}}^{\alpha} + z_{\bar{s}}^{\alpha}}, \qquad \alpha = 0, 1, \ldots \tag{10}$$

Varians penganggar bersamaan dengan

$$Var(\hat{\sigma}_{r,s:n}) = \mathbf{c}'_{r,s}\boldsymbol{\Sigma}_{r,s}\mathbf{c}_{r,s}, \tag{11}$$

dengan

$$\mathbf{c'}_{r,s} = \left(\frac{z_{\bar{r}}^{\alpha-1}}{z_{\bar{r}}^{\alpha} + z_{\bar{s}}^{\alpha}}, \frac{z_{\bar{s}}^{\alpha-1}}{z_{\bar{r}}^{\alpha} + z_{\bar{s}}^{\alpha}} \right)$$

dan

$$\mathbf{\Sigma}_{r,s} = \left[\frac{p_r q_s}{4\phi(z_{\bar{r}})\phi(z_{\bar{s}})} \right]_{2\times 2} \frac{\sigma^2}{n}$$

yang merupakan matriks kovarians bagi vektor rawak $(X_{r:n}, X_{s:n})'$. Keputusan yang diperoleh untuk gabungan dua pembolehubah di atas dapat diitlakkan untuk sebarang gabungan linear statistik tertib yang memberikan hubungan berikut,

$$\hat{\sigma} = \mathbf{c}'\underline{X} \quad \text{dengan} \quad Var(\hat{\sigma}) = \mathbf{c}'\Sigma\mathbf{c} \tag{12}$$

dimana $\mathbf{c}' = \left(z_{\bar{r}}^{\alpha-1}, z_{\bar{s}}^{\alpha-1}, \ldots\right) \Big/ \sum_{\bar{i}} z_{\bar{i}}^{\alpha}$, $\mathbf{\Sigma} = \left[\frac{p_r q_s}{4\phi(z_{\bar{r}})\phi(z_{\bar{s}})} \right] \frac{\sigma^2}{n}$, dan $\underline{X}$ mewakili vektor statistik tertib dengan $\bar{i}$ merujuk semua unsur $\mathbf{c}$.

Bagi gabungan dua statistik tertib, statistik berikut juga boleh digunakan sebagai penganggar,

$$\hat{\sigma}_{r,s:n} = \frac{x_{s:n} z_{\bar{s}}^{\alpha-1} - x_{r:n} z_{\bar{r}}^{\alpha-1}}{z_{\bar{s}}^{\alpha} - z_{\bar{r}}^{\alpha}}, \qquad r < s, \quad \alpha = 1, 2, \ldots \tag{13}$$

Varians untuk penganggar ini dapat diperoleh berdasarkan persamaan (12). Seperti khasus (10) di atas, dengan mengambil gabungan lebih dua statistik tertib juga boleh dibangunkan berasaskan hubungan pada (13).

5 Kecekapan Penganggar

Terdapat banyak penganggar teguh bagi parameter skala bagi taburan normal separuh yang telah dicadangkan. Bahagian ini akan menyelidiki kecekapan penganggar-penganggar tersebut. Sebagai permulaan tumpuan diberi kepada penganggar yang hanya menggunakan satu statistik tertib.

Jika $x_{(p_r)} = x_{r:n}$ mewakili kuantil p_r data, iaitu cerapan yang ke-r terbesar, $r = \lceil np_r \rceil$, maka lajur bertajuk penganggar dalam Jadual 1 memberikan penganggar parameter skala yang berkenaan, seperti persamaan (8). Lajur bertajuk faktor pula memberikan faktor varians yang merupakan faktor kepada σ^2/n, seperti persamaan (9). Lajur ini menyenaraikan nilai yang sepadan dengan ungkapan $fk = p_r q_r \big/ (4\phi^2(z_{\bar{r}}) z_{\bar{r}}^2)$. Kecekapan adalah nisbah varians penganggar kebolehjadian maksimum terhadap varians statistik teguh, dinyatakan dalam peratus, $100(0.5/fk)$.

Untuk menunjukkan dengan lebih jelas bagaimana isi Jadual 1 diperolehi, satu contoh berdasarkan $p = 0.5$ dengan $\bar{r} = \frac{1}{2}(p+1) = 0.75$ diberikan seperti berikut.

$$\text{Penganggar } \frac{x_{(0.50)}}{z_{\bar{r}} = \Phi^{-1}\left[\frac{1}{2}(p+1)\right]} = \frac{x_{(0.50)}}{0.675} = 1.48 x_{(0.50)}$$

Jadual 1: Faktor varians dan kecekapan penganggar teguh berasaskan statistik tertib pada kuantil yang dinyatakan

Penganggar	Faktor	Kecekapan	Penganggar	Faktor	Kecekapan
$7.96x_{(0.10)}$	9.0953	5.50	$1.19x_{(0.60)}$	1.0807	46.26
$5.29x_{(0.15)}$	5.8036	8.62	$1.07x_{(0.65)}$	0.9799	51.02
$3.95x_{(0.20)}$	4.1753	11.98	$0.96x_{(0.70)}$	0.8990	55.62
$3.14x_{(0.25)}$	3.2108	15.57	$0.87x_{(0.75)}$	0.8359	59.81
$2.60x_{(0.30)}$	2.5774	19.40	$0.78x_{(0.80)}$	0.7908	63.23
$2.20x_{(0.35)}$	2.1324	23.45	$0.69x_{(0.85)}$	0.7676	65.14
$1.91x_{(0.40)}$	1.8048	27.70	$0.68x_{(0.86)}$	0.7666	65.22
$1.67x_{(0.45)}$	1.5553	32.15	$0.66x_{(0.87)}$	0.7672	65.17
$1.48x_{(0.50)}$	1.3605	36.75	$0.64x_{(0.88)}$	0.7696	64.97
$1.32x_{(0.55)}$	1.2055	41.48	$0.61x_{(0.89)}$	0.7743	64.57

Faktor: $fk \times \sigma^2/n$ Kecekapan: $100(2 \times fk)^{-1}$

Faktor $\dfrac{0.5^2}{4\phi^2(0.675)(0.675)^2} = 1.3605$ diperolehi dengan

$$\phi^2(0.675) = \left[\frac{1}{\sqrt{2\pi}}\text{eksp}\left(\frac{-0.675}{2}\right)\right]^2$$

Kecekapan $36.75 = 100(2 \times 1.3605)^{-1}$

Berdasarkan hasil seperti yang dipamirkan dalam Jadual 1, penganggar yang terbaik adalah penganggar yang menggunakan persentil 86%. Penganggar ini mampu memberikan kecekapan 65% berbanding dengan penganggar kebolehjadian maksimum. Titik musnah bagi penganggar ini adalah 15%, ini bermakna penganggar ini teguh dan nilai anggaran tidak berubah selagi peratus data tercemar kurang daripada 15%.

Tumpuan selanjutnya ialah kecekapan dua statistik tertib. Untuk tujuan ini penganggar seperti pada persamaan (10) dan variansnya, persamaan (11) diberikan perhatian. Untuk ini tiga penganggar seperti yang ditentukan oleh nilai $\alpha = 0, 1, 2$ diteliti. Hasil yang diperoleh dipersembahkan pada Jadual 2.

Nilai pada lajur pasangan memaklumkan pasangan statistik tertib yang dijadikan asas dalam pembinaan penganggar teguh. Pasangan ($p_{\bar{r}} = 0.1, p_{\bar{s}} = 0.45$) bermaksud penganggarnya

$$\hat{\sigma} = \frac{7.96^{\alpha-1}x_{(0.10)} + 1.67^{\alpha-1}x_{(0.45)}}{7.96^{\alpha} + 1.67^{\alpha}}$$

seperti pada persamaan (10). Nilai yang tersenarai setentang dengan lajur bertajuk "Nilai α" memaparkan kecekapan penganggar teguh berbanding dengan penganggar kebolehjadian maksimum. Nilai tersebut diperoleh berasaskan $Var(\hat{\sigma}_K) \times 100/Var(\hat{\sigma})$ seperti pada persamaan (2) dan (11). Hanya pasangan yang memberikan kecekapan yang terbaik sahaja yang disenaraikan.

Jadual 2: Kecekapan penganggar teguh berasaskan pelbagai pasangan statistik tertib.

Pasangan		Nilai α			Pasangan		Nilai α		
p_r	p_s	0	1	2	p_r	p_s	0	1	2
0.10	0.50	15.62	33.30	36.85	0.40	0.70	49.19	55.31	57.75
0.15	0.50	21.00	33.07	36.77	0.45	0.70	51.88	55.89	57.77
0.20	0.50	25.26	33.26	36.58	0.50	0.70	53.94	56.36	57.69
0.10	0.55	16.26	38.01	41.66	0.45	0.75	55.83	61.59	63.46
0.15	0.55	22.18	37.72	41.69	0.50	0.75	58.35	62.12	63.58
0.20	0.55	26.99	37.85	41.63	0.55	0.75	60.19	62.45	63.51
0.15	0.60	23.25	42.62	46.67	0.50	0.80	62.73	67.98	68.96
0.20	0.60	28.61	42.72	46.75	0.55	0.80	65.03	68.43	69.23
0.25	0.60	33.07	43.08	46.74	0.60	0.80	66.58	68.60	69.20
0.30	0.65	39.34	48.71	52.09	0.60	0.85	71.80	74.51	74.14
0.35	0.65	42.84	49.29	52.10	0.65	0.85	72.98	74.50	74.27
0.40	0.65	45.68	49.89	52.05	0.70	0.85	73.18	73.91	73.80

Kajian ini mendapati pasangan dua statistik tertib yang terbaik mampu memberikan kecekapan sekitar 74%. Pasangan pada $\alpha = 2$ memberikan kecekapan yang lebih baik. Namun demikian kecekapannya hampir sama dengan apabila $\alpha = 1$. Dua penganggar berikut adalah pasangan penganggar yang terbaik, iaitu pasangan yang menggunakan kuantil 0.65 dan 0.85,

$$\hat{\sigma} = 0.57(x_{(0.65)} + x_{(0.85)}) \text{ dan } \hat{\sigma} = 0.66x_{(0.65)} + 0.43x_{(0.85)}$$

Titik musnah bagi kedua-dua penganggar ini adalah 15%, ini bermakna penganggar ini teguh dan nilai anggaran tidak berubah selagi peratus data tercemar kurang daripada 15%.

Nilai pada Jadual 3 membentangkan kesan kehadiran pencilan terhadap anggaran parameter yang nilai sebenarnya adalah satu. Data dalam contoh ini dijana secara rawak menggunakan penjana rawak normal perisian S-Plus versi 6.2. Nilai rawak normal separuh dijana berdasarkan nilai mutlak nombor rawak normal. Dalam kajian khasus yang disajikan pada Jadual 3, sampel diambil daripada nilai mutlak populasi normal piawai, manakala pencilan merupakan nilai mutlak sampel rawak daripada populasi normal min sifar dengan sisihan piawai 20. Bilangan pencilan dalam sampel diputuskan mengikut bilangan sampel rawak dan ini dinyatakan pada lajur pertama. Peratus tertinggi pencilan dalam sampel ialah 10%. Maklumat yang dipersembahkan pada jadual adalah untuk membandingkan anggaran parameter skala yang diperoleh berdasarkan statistik tertib persentil 50, 60, 70 dan 80 dengan anggaran yang diperoleh menggunakan kaedah momen dan kebolehjadian maksimum.

Untuk kes saiz sampel 20, tanpa pencilan, nilai anggaran untuk semua penganggar adalah hampir dengan nilai sebenar iaitu 1. Bagaimanapun, kehadiran satu pencilan telah menyebabkan nilai anggaran meningkat kepada 2.100 bagi kaedah momen dan 3.458 untuk kaedah kebolehjadian maksimum. Jelas menunjukkan kesan pencilan lebih ketara pada anggaran kebolehjadian maksimum iaitu nilai anggaran melebihi 3 kali ganda daripada nilai sebenar. Apabila bilangan pencilan meningkat, nilai anggaran kebolehjadian maksimum menjadi semakin besar daripada nilai sebenar yang bersamaan 1. Kesan kehadiran pencilan masih juga kelihatan meskipun saiz sampel telah ditingkatkan. Apabila sampel ditambah kepada 100, kesan pencilan semakin mengurang tetapi masalah terlebih anggar masih kekal, lebih-lebih lagi apabila bilangan pencilan meningkat.

Jadual 3: Anggaran parameter skala mengikut pelbagai penganggar, $T_{.\alpha}$ penganggar kuantil α, dan bilangan kehadiran pencilan untuk sampel bersaiz 20, 50, dan 100. Pencilan daripada populasi normal separuh dengan sisihan piawai 20.

N=20			Penganggar			
Pencilan	$T_{.50}$	$T_{.60}$	$T_{.70}$	$T_{.80}$	Momen	PKM
0	1.328	1.175	1.082	0.951	1.238	1.220
1	1.369	1.268	1.143	1.042	2.100	3.458
2	1.369	1.268	1.143	1.186	3.653	6.775

N=50			Penganggar			
Pencilan	$T_{.50}$	$T_{.60}$	$T_{.70}$	$T_{.80}$	Momen	PKM
0	0.998	1.054	1.110	1.116	1.162	1.157
1	0.998	1.054	1.150	1.218	1.683	3.303
2	1.009	1.133	1.187	1.298	1.877	3.498
3	1.009	1.133	1.187	1.298	3.096	8.030
4	1.009	1.133	1.188	1.370	3.694	8.779
5	1.009	1.237	1.255	1.420	4.105	9.111

N=100			Penganggar			
Pencilan	$T_{.50}$	$T_{.60}$	$T_{.70}$	$T_{.80}$	Momen	PKM
0	0.896	0.944	0.955	0.968	0.957	0.998
1	0.916	0.962	0.957	0.973	1.060	1.318
2	0.942	1.011	0.969	0.976	1.156	1.543
3	0.942	1.011	0.969	1.008	1.413	2.653
4	1.008	1.044	0.988	1.104	1.433	2.659
5	1.008	1.044	0.996	1.134	1.646	3.200
6	1.008	1.052	1.035	1.144	1.844	3.607
7	1.050	1.067	1.102	1.149	1.991	3.812
8	1.057	1.078	1.157	1.170	2.053	3.848
9	1.064	1.093	1.186	1.200	2.101	3.868
10	1.076	1.139	1.198	1.212	2.130	3.876

Jika dilihat dari segi kepincangan anggaran, kajian simulasi ini juga menunjukkan bahawa penganggar momen sentiasa lebih baik jika dibandingkan dengan kaedah kebolehjadian maksimum. Namun demikian kaedah momen masih tidak dapat menandingi kelebihan anggaran yang diperoleh berdasarkan statistik tertib. Kajian juga menunjukkan yang statistik tertib adalah lebih teguh terhadap gangguan pencilan dan anggaran yang diperoleh hampir dengan nilai sebenar. Peningkatan pincang pada anggaran kaedah kebolehjadian telah menyebabkan ia menjadi tidak cekap, suatu keadaan yang juga berlaku pada anggaran kaedah momen.

6 Kesimpulan

Tanpa kehadiran pencilan, kesemua penganggar memberikan anggaran yang hampir sama dengan nilai sebenar. Namun demikian, varians kaedah kebolehjadian maksimum memberikan kejituan yang lebih baik. Kehadiran pencilan telah memusnahkan kelebihan anggaran parameter skala taburan normal separuh kaedah kebolehjadian maksimum. Meskipun kaedah tersebut menawarkan penganggar saksama varians minimum, namun dengan kehadiran pencilan dalam peratusan yang rendah sahaja sudah memadai untuk memusnahkan kelebihan tersebut. Kehadiran pencilan menyebabkan bukan sahaja anggaran yang diperoleh jauh lebih anggar tetapi juga meningkatkan purata kuasa dua ralat anggaran. Kaedah momen juga mengalami kesan yang sama tetapi dalam magnitud yang lebih rendah. Peningkatan saiz sampel hanya mengurangkan sedikit kesan pencilan. Penganggar berasaskan statistik tertib adalah jauh lebih stabil dan mampu memberikan anggaran yang jauh lebih tepat. Oleh itu sekiranya data didapati tercemar maka diusulkan agar keutamaan kepada statistik tertib. Ujian kehadiran pencilan boleh diselidiki menggunakan ujian yang dicadangkan oleh David dan Nagarajah [11] atau Barnett dan Lewis [1].

Penghargaan

Kami mengucapkan terima kasih kepada Universiti Kebangsaan Malaysia, yang telah membiayai kajian ini di bawah biayaan Projek Jangka Pendek ST-014-2004.

Rujukan

[1] V. Barnett dan T. Lewis, *Outliers in Statistical Data*, 3rd. ed., John Wiley, Chichester, 1994.

[2] C.D. Hoaglin, F. Mosteller dan J.W. Tukey, *Understanding Robust and Exploratory Data Analysis*, Wiley-Interscience, New York, 2000.

[3] P.J. Huber, *Robust Statistics*, Wiley-Interscience, New York, 2003.

[4] C. Park & B. G. Lindsay, *Robust scale estimation and hypothesis testing based on quadratic inference function*, Technical reports ♯99-03, Center for likelihood studies, Department of Statistics, The Pennsylvania State University, 1999.

[5] H. Wang & D. Suter, *Robust scale estimation from true parameters of model*, Technical Reports MECSE-2-2003, Department of Electrical and Computer Systems Engineering, Monash University, Clayton, Victoria Australia, 2003.

[6] S.T. Buckland, D.R. Anderson, K.P. Burnham, J.L Laake, D.L. Borchers & L. Thomas, *Introduction to distance sampling: Estimating abundance of biological populations*, Oxford University Press, New York, 2001.

[7] K.P. Burnham, D.R. Anderson & J.L. Laake, *Estimation of density from line transect sampling of biological populations*, Wildlife Monographs, 72(1980), 1–202.

[8] V. Brazauskas & R. Serfling, *Robust estimation of tail parameters for two-parameter Pareto and Exponential models via Generalized Quantil statistics*, Extremes Statistical Theory and Applications, 3(2001), 231–249.

[9] M.L. Huang, *On distribution-free quantile estimator*, Computational Statistics and Data Analysis, 37(4)(2001), 477–486.

[10] M.L. Huang & P. H. Brill, *A level crossing quantile estimation method*, Statistics and Probability Letters, 45(2)(1999), 111–119.

[11] H.A. David & H.N. Nagaraja, *Order statistics*, 3rd. ed., Wiley-Interscience, New Jersey, 2003.