ISSN 0216-2393

Jurnal Gradien Vol. 10 No. 1 Januari 2014 : 963-966

Pendugaan Galat Baku Nilai Tengah Menggunakan Metode *Resampling Jackknife* dan *Bootstrap Nonparametric* dengan *Software* R 2.15.0

*Septiana Wulandari, Dian Kurniasari, Widiarti

[1]*Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung, Indonesia*
**Tyanjawa@gmail.com*

Diterima 3 Mei 2013; Disetujui 10 Juni 2013

Abstrak - Metode *Jackknife* dan *Bootstrap* merupakan metode berbasis *resampling*. Kedua metode ini biasa digunakan untuk inferensia statistic, yaitu pada pendugaan parameter dengan sebaran awal yang tidak diketahui. Metode *Jackknife* dan *Boostrap* merupakan metode yang tidak didasarkan pada pemenuhan asumsi tertentu, sehingga metode ini dapat diterapkan pada seluruh data kuantitatif. Secara teori dapat dibuktikan bahwa metode *Jackknife* merupakan penduga yang baik untuk data asal. Namun berdasarkan kreteria Kuadrat Tengah Galat (KTG), metode *Bootstrap* lebih tepat digunakan karena menghasilkan nilai KTG yang lebih kecil dibanding metode *jackknife*.

Kata Kunci: ***Jackknife, Bootstrap, Resampling***

1. Pendahuluan

Perputaran waktu dan perkembangan zaman menjadikan statistika mengalami berbagai kemajuan baik dari aspek teoritis yang terkait dengan penggalian berbagai pengetahuan baru maupun yang bersifat praktis berkenaan dengan keluasan bidang pekerjaan yang bisa menjadi lebih baik dan mudah dengan statistika [**1**]. Pada umumnya statistika digunakan untuk menganalisis suatu data dengan melakukan penarikan sampel. Keterbatasan dalam memperoleh sampel mengakibatkan penduga parameter yang diperoleh bias. Sehingga dapat diatasi dengan melakukan *resampling*. Ada 2 metode *resampling* yang digunakan, yaitu *jackknife* dan *bootstrap*. Dalam penelitian ini akan dilakukan perbandingan mengenai galat baku nilai tengah untuk kedua metode tersebut, dimana galat baku nilai tengah merupakan simpangan baku dari galat sampel *mean* terhadap *mean* sesungguhnya [**2**]. Selain itu juga akan dilakukan kajian empiris dengan *Software* R 2.15.0.

2. Metode Penelitian

Pada penelitian ini dilakukan perbandingan galat baku nilai tengah kedua metode *jackknife* dan *bootstrap*. Selanjutnya dilakukan kajian empiris dengan *software* R 2.15.0. Langkah yang pertama dalam melakukan penelitian ini adalah menentukan banyaknya contoh dari masing-masing metode dan selanjutnya melakukan evaluasi terhadap setiap pengulangan. Dengan demikian dapat dihitung besar galat baku untuk metode *jackknife* dan *bootstrap*.

3. Hasil dan Pembahasan

Metode Jackknife

Metode *jackknife* merupakan teknik *resampling* yang digunakan untuk pendugaan suatu galat baku dan memperkirakan selang kepercayaan untuk suatu parameter. Prinsip dasar dari metode *jackknife* terletak pada perhitungan suatu statistik secara berulang dengan mengeluarkan satu atau lebih pengamatan dari suatu sampel yang ditetapkan, sehingga akan menghasilkan sampel terpisah yang masing-masing memiliki besar ukuran n-1 atau n-d. Dari proses pengulangan tersebut, dapat dihitung dugaan suatu kebiasan dan varians [**3**]. Dapat didefinisikan misal $\theta_{(i)} = t(x_{(i)})$ adalah penduga yang dibangun dari data X untuk menduga parameter θ, maka dari himpunan $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n$ merupakan penduga galat baku *jackknife*, yaitu

$$S_{Jack} = \left[\frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{(i)} - \hat{\theta}_{(.)})^2\right]^{1/2}$$

dimana $\hat{\theta}_{(.)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(i)}$ dan $\hat{\theta}_{(i)}$ disebut sebagai ulangan *jackknife* ke-i dari $\hat{\theta}$.

Apabila statistika yang diinginkan adalah nilai tengah sampel maka $\hat{\theta}_{(i)} = \bar{x}_{(i)}$, sehingga diperoleh S_{Jack} sebagai berikut:

$$S_{Jack}(\bar{X}) = \left[\frac{n-1}{n}\sum_{i=1}^{n}(\bar{x}_{(i)} - \bar{x}_{(.)})^2\right]^{1/2} \ldots\ldots[4]$$

Selanjutnya akan dilakukan kajian secara teori terkait metode *jackknife*. Untuk metode *jackknife* misalkan diketahui suatu sampel acak berukuran n, X = $(x_1, x_2, \ldots, x_n)$ yang menyebar secara bebas stokastik identik dari suatu sebaran peluang F yang tidak diketahui.

Misalkan $x_{(i)}$ adalah data *resample* ke-i

$$x_{(i)} = \begin{bmatrix} x_2 & x & x_1 \cdots & x_1 & x_1 \\ \vdots & \vdots & \vdots\ \vdots & \vdots & \vdots \\ x_n & x_n & x_n \cdots & x_n & x_{n-1} \end{bmatrix} ; i=1,2,\ldots,n$$

Maka $x_{(i)}$ disebut sebagai sampel *jackknife*. Dengan demikian penduga galat baku *jackknife* dari nilai tengah yaitu:

$$S_{Jack}(\bar{x}) = \left[\frac{n-1}{n}\sum_{i=1}^{n}(\bar{x}_{(i)} - \bar{x}_{(.)})^2\right]^{1/2}$$

dimana,

$$\bar{x}_{(i)} = \frac{1}{n-1}\left[\left(\sum_{i=1}^{n} x_i\right) - x_i\right] \text{ dan } \bar{x}_{(.)} = \frac{1}{n}\left(\sum_{i=1}^{n}\bar{x}_{(i)}\right) ..[4]$$

Untuk melihat bahwa metode *jackknife* merupakan penduga yang baik, maka harus memenuhi penduga yang tak bias dan ragam minimum. Jika diketahui:

$$\bar{x}_{(i)} = \frac{1}{n-1}\left[\left(\sum_{i=1}^{n} x_i\right) - x_i\right]$$

$$\bar{x}_{(.)} = \frac{1}{n}\left(\sum_{i=1}^{n}\bar{x}_{(i)}\right) \text{ dan}$$

$\hat{\theta}_{(.)} = \frac{1}{n}\left(\sum_{i=1}^{n}\hat{\theta}_{(i)}\right)$ merupakan penduga dari θ

Berdasarkan definisi dan contoh yang diberikan oleh Tatik Widiharih (2007) dalam buku ajarnya Statistika Matematika II, syarat penduga tak bias sebagai berikut:

E($\bar{x}_{(.)}$) = $\bar{x}$ atau E($\hat{\theta}_{(.)}$) = θ sehingga,

$\text{Bias}(\bar{x}_{(.)}) = E(\bar{x}_{(.)}) - \bar{x}$ $\qquad$ $\text{Bias}(\hat{\theta}_{(.)}) = E(\hat{\theta}_{(.)}) - \theta$

$= \bar{x} - \bar{x}$ $\qquad$ $= \theta - \theta$

$= 0$ $\qquad$ $= 0$

$$\frac{n-1}{\theta}.S^2 \sim \chi^2_{(n-1)}$$

$$\text{var}\left[\frac{n-1}{\theta}.S^2\right] = 2(n-1)$$

$$\left(\frac{n-1}{\theta}\right)^2 \text{var}(S^2) = 2(n-1)$$

$$\text{var}(S^2) = \frac{2\theta^2}{n-1}$$

$$\text{var}(\hat{\hat{\theta}}) = \text{var}(S^2)$$

$$\text{var}(\hat{\theta}) = \text{var}\left(\frac{1}{n}.(n-1)S^2\right)$$

$$= \left(\frac{n-1}{n}\right)^2 \text{var}(S^2)$$

$$= \left(\frac{n-1}{n}\right)^2 . \frac{2\theta^2}{n-1}$$

$$= \frac{2(n-1)}{n^2}.\theta^2$$

$$KTG(\hat{\theta}) = \text{var}(\hat{\theta}) + [\text{Bias}(\hat{\theta})]^2$$

$$= \frac{2(n-1)}{n^2}.\theta^2 + 0$$

$$= \frac{2(n-1)}{n^2}.\theta^2$$

$$KTG(\hat{\hat{\theta}}) = \text{var}(\hat{\hat{\theta}}) + [\text{Bias}(\hat{\hat{\theta}})]^2$$

$$= \frac{2\theta^2}{n-1} + 0$$

$$= \frac{2\theta^2}{n-1}$$

$$KTG(\hat{\theta}) < KTG(\hat{\hat{\theta}})$$

$$\frac{2(n-1)}{n^2}.\theta^2 < \frac{2\theta^2}{n-1} \text{; untuk } n>1$$

∴ var ($\hat{\theta}$) varian minimum untuk metode *jackknife*.

Metode Bootstrap

Metode *bootstrap* adalah metode *resampling* untuk inferensia statistik yang biasanya digunakan untuk menduga selang kepercayaan. Namun dapat juga digunakan untuk menduga bias dan menduga ragam atau menentukan hipotesis. Metode ini diperkenalkan oleh Efron (1979), dimana dapat didefinisikan jika:

$$X=(x_1, x_2, \ldots, x_n)$$

adalah sampel acak dari suatu populasi maka:

$$\text{variabel } X^* = (x^*_1, x^*_2, \ldots, x^*_n)$$

adalah sampel acak *bootstrap* yaitu sampel yang diperoleh dari X secara acak dengan pengembalian.

Variabel $x^*_1, x^*_2, \ldots, x^*_n$ bebas dan berdistribusi bersyarat terhadap X. Tanda * mengindikasikan bahwa X^* bukan himpunan data X tetapi hasil resampel, berarti titik data x^*_1, $x^*_2, \ldots, x^*_n$ adalah sampel acak berukuran n dengan pengembalian dari populasi n $(x_1, x_2, \ldots, x_n)$, $x^*_1 = x_7$, $x^*_2 = x_3, \ldots, x^*_n = x_n$ [2].

Sama halnya dengan metode *jackknife*, untuk menunjukkan secara teori bahwa metode *bootstrap* merupakan penduga yang baik maka perlu ditunjukkan syarat ketakbiasan dan ragam minimum.

Misal diketahui $X^* \sim N(\overline{x^*}, \theta)$ dan diketahui $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i^{*2} - (\overline{x^*})^2$ dengan penduga lain dari θ adalah $\hat{\hat{\theta}} = \frac{1}{n-1}\sum_{i=1}^{n} x_i^{*2} - (\overline{x^*})^2$

Syarat penduga tak bias: $E(\hat{\theta}) = E(S^2) = \theta$

Bukti:

$$E(\hat{\theta}(x_i^*)) = E(\hat{\theta}(x_i^*)|X)$$
$$= E\left(\frac{1}{n}\sum_{i=1}^{n} x_i^{*2} - (\overline{x^*})^2\right)$$
$$= (E(x^*))^2 - \left(\frac{1}{n}E(x^{*2})\right) + \frac{n-1}{n}((E(x^*))^2)$$
$$= \frac{n-1}{n}(E(x^{*2}) - (E(x^*))^2)$$
$$= \frac{n-1}{n} \cdot \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2$$
$$= \frac{n-1}{n} \cdot S_x^2$$
$$E(\hat{\theta}) = \frac{n-1}{n} \cdot \theta$$

$\therefore$ penduga berbias untuk metode *bootstrap*

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$$
$$= \frac{1}{n} \cdot (n-1)\,\theta - \theta$$
$$= -\frac{\theta}{n}$$
$$\frac{n-1}{\theta} \cdot S^2 \sim \chi^2_{(n-1)}$$
$$\text{var}\left[\frac{n-1}{\theta} \cdot S^2\right] = 2(n-1)$$
$$\left(\frac{n-1}{\theta}\right)^2 \text{var}(S^2) = 2(n-1)$$
$$\text{var}(S^2) = \frac{2\theta^2}{n-1}$$
$$\text{var}(\hat{\hat{\theta}}) = \text{var}(S^2)$$
$$\text{var}(\hat{\theta}) = \text{var}\left(\frac{1}{n} \cdot (n-1)\, S^2\right)$$
$$= \left(\frac{n-1}{n}\right)^2 \text{var}(S^2)$$
$$= \left(\frac{n-1}{n}\right)^2 \cdot \frac{2\theta^2}{n-1}$$
$$= \frac{2(n-1)}{n^2} \cdot \hat{\theta}^2$$
$$KTG(\hat{\theta}) = \text{var}(\hat{\theta}) + \left[\text{Bias}(\hat{\theta})\right]^2$$
$$= \frac{2(n-1)}{n^2} \cdot \hat{\theta}^2 + \left(-\frac{\theta}{n}\right)^2$$
$$KTG(\hat{\hat{\theta}}) = \text{var}(\hat{\hat{\theta}}) + \left[\text{Bias}(\hat{\hat{\theta}})\right]^2$$
$$= \frac{2\theta^2}{n-1} + 0$$
$$= \frac{2\theta^2}{n-1}$$
$$KTG(\hat{\theta}) < KTG(\hat{\hat{\theta}})$$
$$\frac{2(n-1)}{n^2} \cdot \hat{\theta}^2 + \left(-\frac{\theta}{n}\right)^2 < \frac{2\theta^2}{n-1}\text{; untuk } n>1$$

$\therefore$ var $(\hat{\theta})$ varian minimum untuk metode *bootstrap*

Karena tidak memenuhi syarat ketakbiasan maka metode *bootstrap* bukan merupakan penduga yang baik.

Kajian Empiris

Untuk memperkuat kajian secara teori ,maka dilakukann kajian secara empiris yaitu dengan melakukan simulasi dengan program R.2.15.0. Dibawah ini adalah data penelitian yang diambil dari dari buku "*An Introduction to the Bootstrap(1993)*" dengan fokus pada group *treatment.*

Tabel 1. Data peneltian

Data	94	197	16	38	99	141	23
Sample Size	7						
Mean	86,86						
Estimated Standard Error	25,24						

Selanjutnya akan dilakukan *resample* sebanyak 10, 30, dan 100 sampel dengan pengulangan sebanyak 10, 100, dan 1000 yang tersaji pada Tabel 2.

Berdasarkan Tabel 2 dapat dilihat bahwa meanduga *jackknife* tepat menunjukkkan mean sampel awal yang telah diketahui, sedangkan untuk meanduga *bootstrap* menghasilkan mean yang beragam. Dimana jika dicermati ukuran sampel dan banyaknya pengulangan mempengaruhi hasil mean dari metode *bootstrap*, meskipun tidak terlihat jauh perbedaannya. Dengan demikian, jika dilihat secara keseluruhan dengan metode *bootstrap* sampel dengan replikasi 1000 menghasilkan dugaan nilai tengah yang rata-rata lebih mendekati parameter sampel awal. Untuk melihat metode mana yang terbaik, maka kita dapat membandingknnya dengan melihat pendugaan galat baku nilai tengah dengan kreteria kuadrat tengah galat untuk masing-masing metode. Dengan melihat hasil yang tetera pada Tabel 2, maka metode yang baik digunakan yaitu metode *Bootstrap* yaitu dengan menghasilkan besar galat yang minimum dari parameter sampel awal. Dimana ada komponen dasar dalam Kuadrat Tengah Galat yaitu *varians* yang mengukur variabilitas (*precision*) dan bias yang mengukur keakuratan (*accuracy*) dari suatu penduga.

Tabel 2. Hasil *Resample* dengan Program R 2.15.0

Jumlah sampel *bootstrap* (n)	Replikasi *Bootstrap*	Metode *jackknife*			Metode *Bootstrap*		
		Meanduga *Jackknife*	Galat baku *Jackknife*	KTG *Jackknife*	Meanduga *Bootstrap*	Galat Baku *Bootstrap*	KTG *Bootstrap*
10	**10**	86,86	25,24	636,83	84,35	9,07	88,60
	100				87,35	11,47	131,85
	1000				87,99	10,95	121,16
30	**10**	86,86	25,24	636,83	86,09	4,74	23,02
	100				86,84	5,26	27,63
	1000				86,72	5,86	34,38
100	**10**	86,86	25,24	636,83	87,35	2,63	7,15
	100				87,30	2,58	6,86
	1000				86,90	2,97	8,83

4. Kesimpulan

Berdasarkan penelitian yang telah dilaksanakan diperoleh kesimpulan sebagai berikut:

1. Ukuran Sampel yang digunakan pada metode *bootstrap* cukup mempengaruhi besarnya nilai KTG yang dihasilkan. Semakin besar sampel, umumnya akan semakin kecil nilai KTG yang dihasilkan.
2. Secara teori dapat dibuktikan bahwa metode *Jackknife* merupakan penduga yang baik karena nilai parameter yang dihasilkan tepat menunjukkan parameter sampel awal, sedangkan Metode *bootstrap* merupakan penduga yang bias.
3. Berdasarkan perbandingan yang diperlihatkan secara empiris dengan kreteria Kuadrat Tengah Galat (KTG) pada metode *Jackknife* dan *Bootstrap* maka, metode *Bootstrap* telah menunjukkan besar galat yang relatif lebih kecil. Dengan demikian untuk melihat keragaman yang minimum dari sampel awal metode *Bootsrap* lebih tepat digunakan. Sedangkan *Jackknife* akan tepat digunakan jika ingin mengambil keakuratan terhadap data sampel awal.

Daftar Pustaka

[1] Santosa, Budi., Purbayu., Muliawan Hamdani. 2007. *Statistika Deskriptif*. Jakarta: PT. Gelora Aksara Pratama.

[2] Kenney, J., Keeping, E.S. 1963. *Mathematics of Statistics.*Van Nostrand. p. 187.

[3] Efron, B., Tibshirani, R.J. 1993. *An Introduction to The Bootstrap*. New York-London: Chapman&Hall.

[4] Miller, R.G. 1964. A Trustworthy Jackknife. *Journal of the Institute of Mathematical Statistics*.1:9-21

[5] Widiharih, Tatik. 2007. *Buku Ajar Statistika Matematika II*. Semarang: Universitas Diponegoro.