

PENENTUAN DISTRIBUSI SKEWNESS DAN KURTOSIS DENGAN METODE *RESAMPLING* BERDASAR DENSITAS KERNEL (STUDI KASUS PADA ANALISIS INFLASI BULANAN KOMODITAS BAWANG MERAH, DAGING AYAM RAS DAN MINYAK GORENG DI KOTA SEMARANG)

Adi Setiawan

Program Studi Matematika, Fakultas Sains dan Matematika
Universitas Kristen Satya Wacana Jl. Diponegoro 52-60 Salatiga 50711

Email korespondensi : adi_setia_03@yahoo.com

ABSTRAK

Skewness merupakan statistik yang digunakan dalam memberikan gambaran distribusi data apakah miring ke kiri, ke kanan atau simetris sedangkan *kurtosis* merupakan statistik yang digunakan dalam memberikan gambaran apakah distribusi data cenderung rata atau runcing. Dalam makalah ini dibahas tentang bagaimana menentukan distribusi statistik *skewness* dan *kurtosis* dengan metode *resampling* berdasarkan densitas kernel maupun lebar interval kepercayaan *skewness* dan *kurtosis* populasi. Metode yang dijelaskan digunakan dalam kasus inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng di kota Semarang. Hasil yang diperoleh dibandingkan dengan lebar interval kepercayaan dengan metode *bootstrap*. Untuk ukuran sampel yang kecil, lebar interval kepercayaan *skewness* dan *kurtosis* dengan metode *bootstrap* relatif lebih pendek dibandingkan dengan lebar interval kepercayaan *skewness* dan *kurtosis* dengan metode *resampling*, namun untuk ukuran sampel besar cenderung berlaku sebaliknya.

Kata kunci : IHK, inflasi komoditas, metode *resampling* berdasar densitas kernel, metode *bootstrap*.

PENDAHULUAN

Skewness merupakan statistik yang digunakan dalam memberikan gambaran distribusi data apakah miring ke kiri, ke kanan atau simetris sedangkan *kurtosis* merupakan statistik yang digunakan dalam memberikan gambaran apakah distribusi data cenderung rata atau runcing. Interval kepercayaan *bootstrap* untuk *skewness* dan *kurtosis* dibahas dalam paper Ankarali *et al.* (2009) sedangkan metode *resampling* berdasarkan densitas kernel dijelaskan dalam makalah Setiawan (2002). Dalam makalah ini dibahas tentang bagaimana menentukan distribusi statistik *skewness* dan *kurtosis* dengan metode *resampling* berdasarkan densitas kernel maupun lebar interval kepercayaan *skewness* dan *kurtosis* populasi. Hasil yang diperoleh dibandingkan dengan lebar interval kepercayaan dengan metode *bootstrap*. Metode yang dijelaskan digunakan dalam kasus inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng di kota Semarang.

DASAR TEORI

Skewness dari suatu variable random X yang dinotasikan dengan $Skew[X]$ didefinisikan sebagai

$$\tau = Skew[X] = \frac{E[(X - \mu)^3]}{(E[(X - \mu)^2])^{3/2}}$$

dengan $\mu = E[X]$. *Skewness* ini juga dinamakan *skewness* populasi. *Skewness* merupakan ukuran dari kesimetrisan atau lebih tepatnya kurang-simetrisan. Suatu distribusi dikatakan simetris jika distribusi tersebut nampak sama antara sebelah kanan dan sebelah kiri titik pusatnya. Distribusi yang simetris misalnya distribusi normal, distribusi t dan distribusi seragam. Distribusi yang mempunyai *skewness* positif misalnya distribusi eksponensial, distribusi chi-kuadrat, distribusi Poisson dan distribusi Binomial dengan $p > 0.5$ sedangkan

distribusi yang mempunyai *skewness* negatif misalnya distribusi Binomial dengan $p < 0.5$ (lihat Tabel 1). Jika dimiliki sampel X_1, X_2, \dots, X_n yang diambil dari suatu populasi maka *skewness* distribusi populasinya dapat diestimasi dengan *skewness* sampel yaitu

$$\hat{\tau} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{3/2}}$$

Kurtosis dari suatu variable random X didefinisikan sebagai

$$\kappa = \frac{E[(X - \mu)^4]}{(E[(X - \mu)^2])^2}$$

Kurtosis merupakan ukuran apakah distribusi X lebih rata secara relatif dari distribusi normal atau sebaliknya. Distribusi yang mempunyai *kurtosis* lebih kecil dari 3 maka kurang rata (*flat*) dibandingkan dengan distribusi normal. Dengan kata lain, distribusi yang mempunyai distribusi yang mempunyai *kurtosis* lebih dari 3 misalnya distribusi eksponensial, chi-kuadrat, distribusi t , distribusi Binomial dan distribusi Poisson, sedangkan yang mempunyai kurang dari 3 misalnya distribusi seragam (lihat Tabel 1). *Kurtosis* dari sampel X_1, X_2, \dots, X_n yang didefinisikan sebagai

$$\hat{\kappa} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2}$$

dapat digunakan untuk mengestimasi *kurtosis* populasi. Pada Tabel 1 berikut ini diberikan *skewness* dan *kurtosis* populasi untuk berbagai macam distribusi yang biasa digunakan (de Gunst dan van der Vaart, 1993).

Tabel 1. *Skewness* dan *Kurtosis* Populasi untuk Beberapa Distribusi.

Distribusi	<i>Skewness</i> τ	<i>Kurtosis</i> κ
Binomial Binom(n, p)	$\frac{1-2p}{\sqrt{np(1-p)}}$	$3 + \frac{1-6p(1-p)}{np(1-p)}$
Poisson Pois(μ)	$\mu^{-1/2}$	$3 + \mu^{-1}$
Normal N(μ, σ^2)	0	3
Seragam U(a, b)	0	9/5
Distribusi t t_ν	0 ($\nu > 3$)	$3 + \frac{6}{\nu-4}$ ($\nu > 4$)
Chi-kuadrat χ_ν^2	$2(2/\nu)^{1/2}$	$3 + \frac{12}{\nu}$
Ekspensial Exp(λ)	2	9

Metode *bootstrap* dapat dijelaskan berikut ini. Misalkan dimiliki sampel X_1, X_2, \dots, X_n dan sampel baru ukuran n dibangkitkan dengan cara mengambil satu demi satu dengan pengembalian dari sampel X_1, X_2, \dots, X_n sampai diperoleh sampel ukuran n . Sampel baru yang diperoleh yaitu $X_1^*, X_2^*, \dots, X_n^*$ digunakan dalam pernghitungan statistik *skewness* dan *kurtosis*. Sedangkan metode *resampling* berdasarkan densitas kernel dapat dijelaskan berikut ini. Misalkan dimiliki sampel X_1, X_2, \dots, X_n dan berdasarkan sampel tersebut dapat ditentukan fungsi densitas kernel berikut :

$$\hat{f}(t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K\left(\frac{t-X_i}{h}\right)$$

dengan $K(t)$ merupakan inti (*kernel*) merupakan fungsi kepadatan probabilitas (*probability density function*) dengan mean 0 dan variansi yang berhingga dan tidak nol. Lebar pita (*bandwidth*) h optimal dipilih dengan metode yang dijelaskan dalam Sheather dan Jones (1991). Sampel X_1^* dibangkitkan dengan cara membangkitkan sampel ukuran 1 dari distribusi yang mempunyai fungsi kepadatan probabilitas

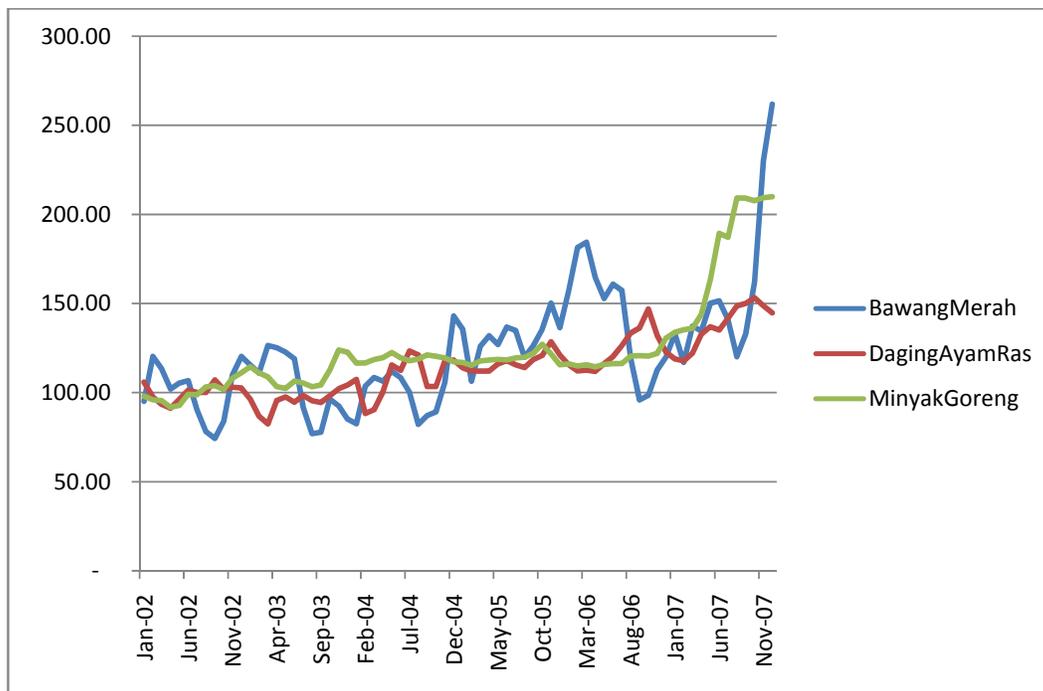
$$f_i(t) = \frac{1}{h} K\left(\frac{t - X_i}{h}\right)$$

dengan X_i dipilih secara acak dari X_1, X_2, \dots, X_n . Apabila prosedur tersebut diulang sebanyak n kali maka akan diperoleh sampel baru hasil *resampling* tersebut yaitu $X_1^*, X_2^*, \dots, X_n^*$. Sampel baru tersebut digunakan dalam perhitungan statistik *skewness* dan *kurtosis*.

METODE PENELITIAN

Data yang digunakan adalah data IHK komoditas bawang merah, daging ayam ras dan minyak goreng di kota Semarang pada periode bulan Januari 2002 sampai dengan Desember 2007 berdasarkan data BPS. Berdasarkan data IHK komoditas bawang merah, daging ayam ras dan minyak goreng, ditentukan inflasi bulanan komoditas masing-masing komoditas, *skewness* dan *kurtosis* untuk periode bulan Februari 2002 sampai dengan Desember 2007. Selanjutnya ditentukan distribusi *skewness* dan *kurtosis* untuk inflasi bulanan masing-masing komoditas yang menjadi perhatian tersebut dengan menggunakan metode *resampling* dan dibandingkan dengan metode *bootstrap*. Kemudian berdasarkan distribusinya ditentukan interval kepercayaan persentil dengan koefisien kepercayaan 95 %.

HASIL DAN PEMBAHASAN



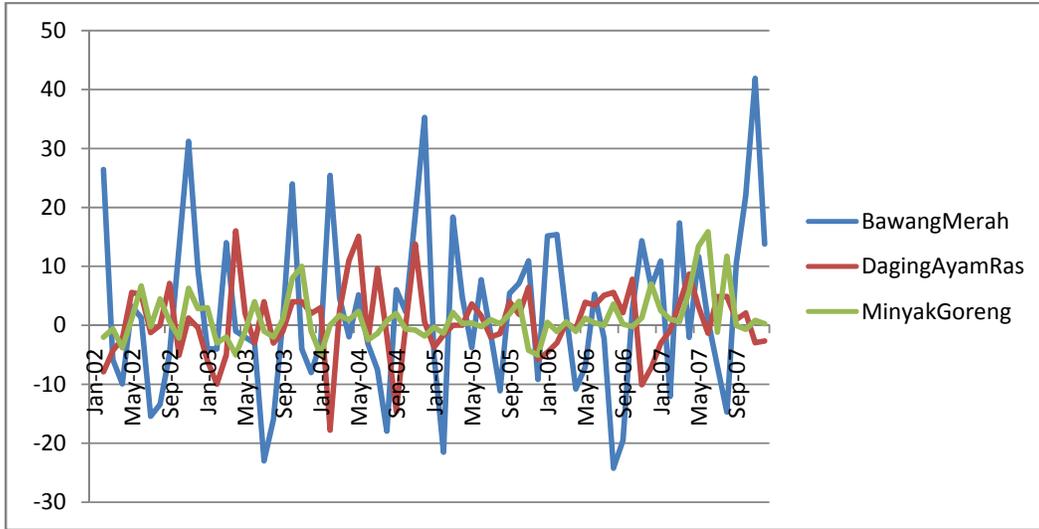
Gambar 1. IHK komoditas bawang merah, daging ayam ras dan minyak goreng periode Februari 2002 sampai dengan Desember 2007.

Gambar 1 menyatakan IHK komoditas yang menjadi perhatian yaitu bawang merah, daging ayam ras dan minyak goreng di kota Semarang. Terlihat bahwa IHK daging

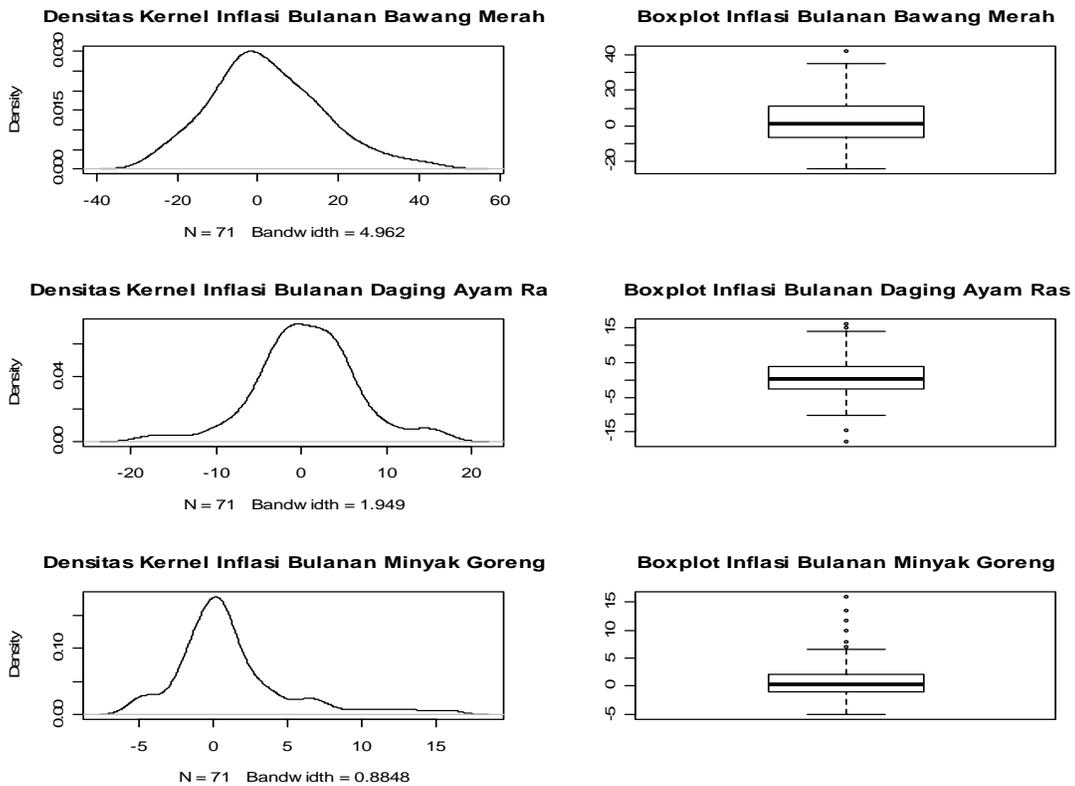
ayam ras cenderung stabil dibandingkan IHK mintak goreng dan bawang merah. Tabel 2 menyatakan statistik deskriptif dari ketiga komoditas yang menjadi perhatian yaitu bawang merah, daging ayam ras dan minyak goreng. Inflasi bulanan komoditas bawang merah mempunyai jangkauan yang lebih besar dibandingkan dengan yang lain. Karena untuk ketiga komoditas tidak terlalu banyak mempunyai *outlier* (Gambar 3) maka akan lebih baik menggunakan mean untuk menyatakan kecenderungan memusat dari data inflasi bulanan ketiga komoditas tersebut. Hal itu berarti, inflasi bulanan bawang merah, daging ayam ras dan minyak goreng berturut-turut berkisar pada berkisar pada 2,331, 0,620 dan 1,152 persen. *Skewness* dari inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng berturut-turut adalah 0,468, -0,125 dan 1,542 sehingga berarti kemencengan untuk inflasi bulanan komoditas bawang merah dan minyak goreng bernilai positif sedangkan *skewness* inflasi bulanan komoditas daging ayam ras bernilai negatif. Hal itu berarti bahwa inflasi bulanan komoditas bawang merah dan minyak goreng lebih banyak yang bernilai relatif kecil dibandingkan yang bernilai besar sedangkan untuk komoditas daging ayam ras berlaku sebaliknya. Hal tersebut juga didukung oleh densitas kernel dari inflasi bulanan untuk ketiga komoditas tersebut yang diberikan pada Gambar 3. *Kurtosis* dari inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng berturut-turut adalah 3,114, 4,122 dan 6,043 sehingga berarti ketebalan ekor (*kurtosis*) distribusi inflasi bulanan ketiga komoditas tersebut lebih besar dari ketebalan ekor (*kurtosis*) distribusi normal yaitu 3. Hal itu berarti inflasi bulanan ketiga komoditas ada relatif banyak yang bernilai besar.

Tabel 2. Statistik deskriptif inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng periode Februari 2002 sampai dengan Desember 2007.

Statistik	Bawang Merah	Daging Ayam Ras	Minyak Goreng
Minimum	-24,230	-17,780	-5,040
Kuartil 1	-6,410	-2,805	-1,005
Median	1,000	-1,005	0,390
Mean/Rata-rata	2,331	0,620	1,152
Kuartil 3	10,920	4,000	2,085
Maksimum	41,900	16,010	15,880
Simpangan baku	13,774	5,999	3,936
Koefisien variasi	5,909	9,682	3,418
Skewness	0,468	-0,125	1,542
Kurtosis	3,114	4,122	6,043



Gambar 2. Inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng periode Februari 2002 sampai dengan Desember 2007.

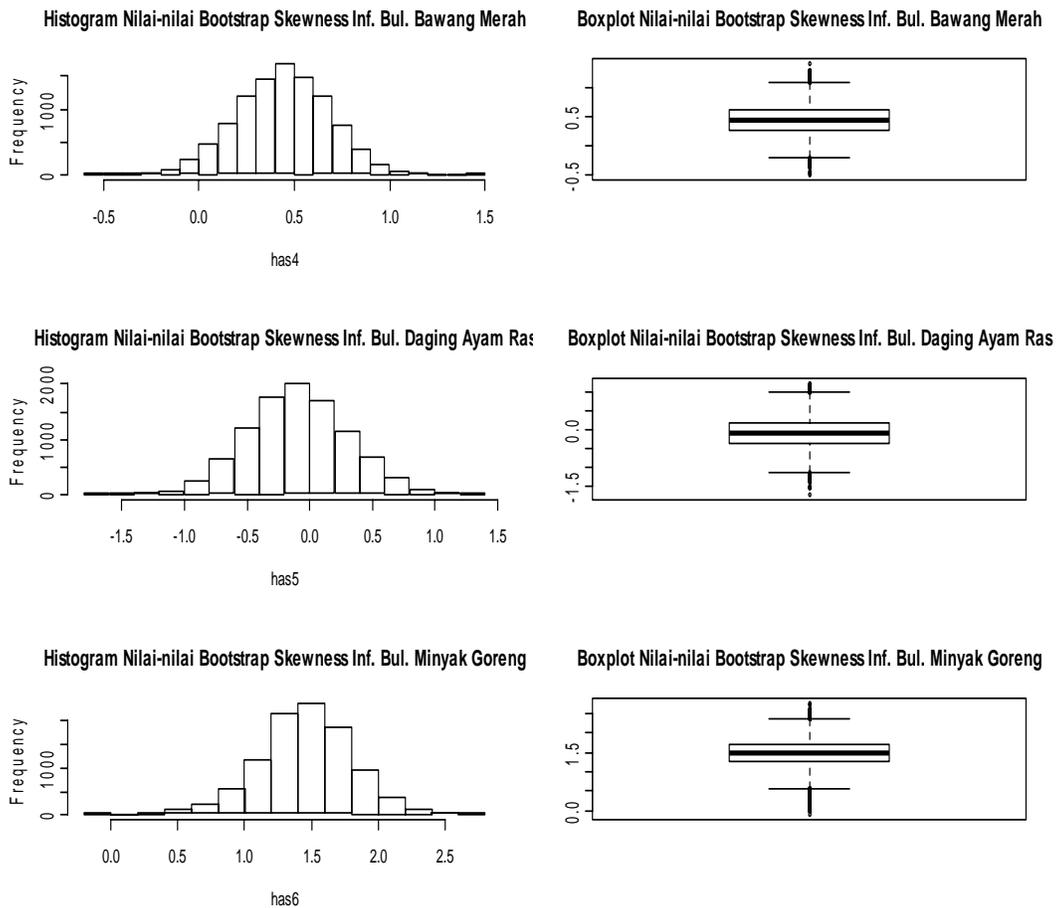


Gambar 3. Densitas kernel dan boxplot inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng periode Februari 2002 sampai dengan Desember 2007.

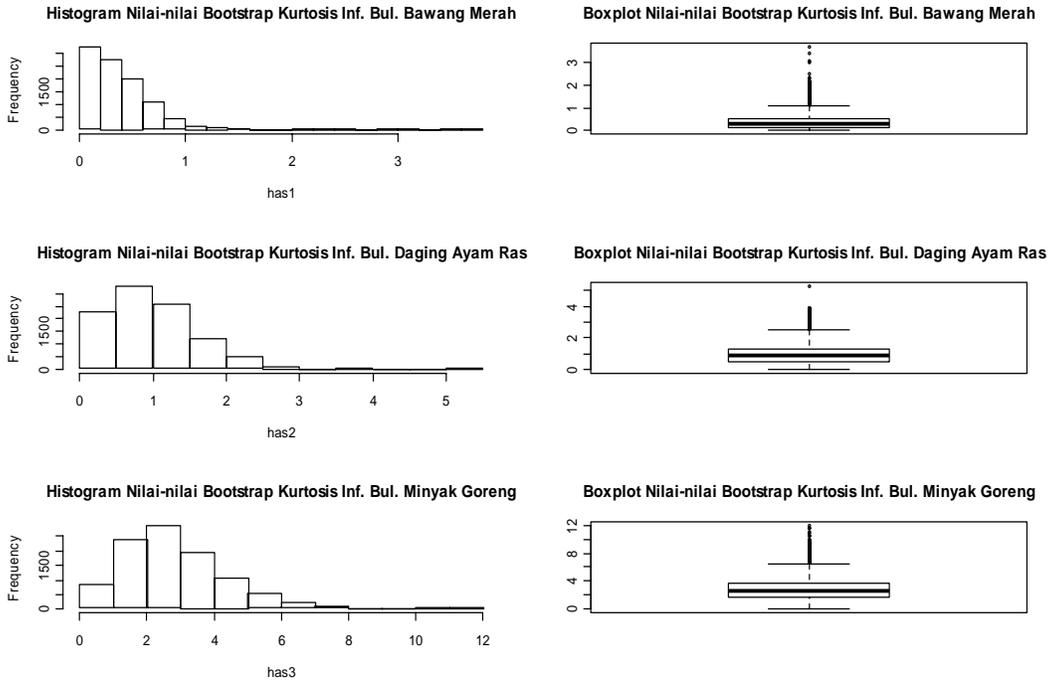
Gambar 4 dan Gambar 5 masing-masing menyatakan histogram dan boxplot nilai-nilai bootstrap dari *skewness* dan *kurtosis* dari inflasi bulanan sedangkan Gambar 6 dan Gambar 7 masing-masing menyatakan histogram dan boxplot nilai-nilai bootstrap dari *skewness* dan *kurtosis* dari inflasi bulanan untuk ketiga komoditas. Nilai-nilai bootstrap maupun nilai-nilai resampling untuk *skewness* dan *kurtosis* menghasilkan cukup banyak

outlier. Tabel 3 memperlihatkan interval kepercayaan *bootstrap* dan interval kepercayaan *resampling* untuk *skewness* dan *kurtosis* dari inflasi bulanan ketiga komoditas yang menjadi perhatian. Terlihat bahwa lebar interval kepercayaan *bootstrap* yang diperoleh, relatif lebih kecil dibandingkan dengan lebar interval kepercayaan *resampling*.

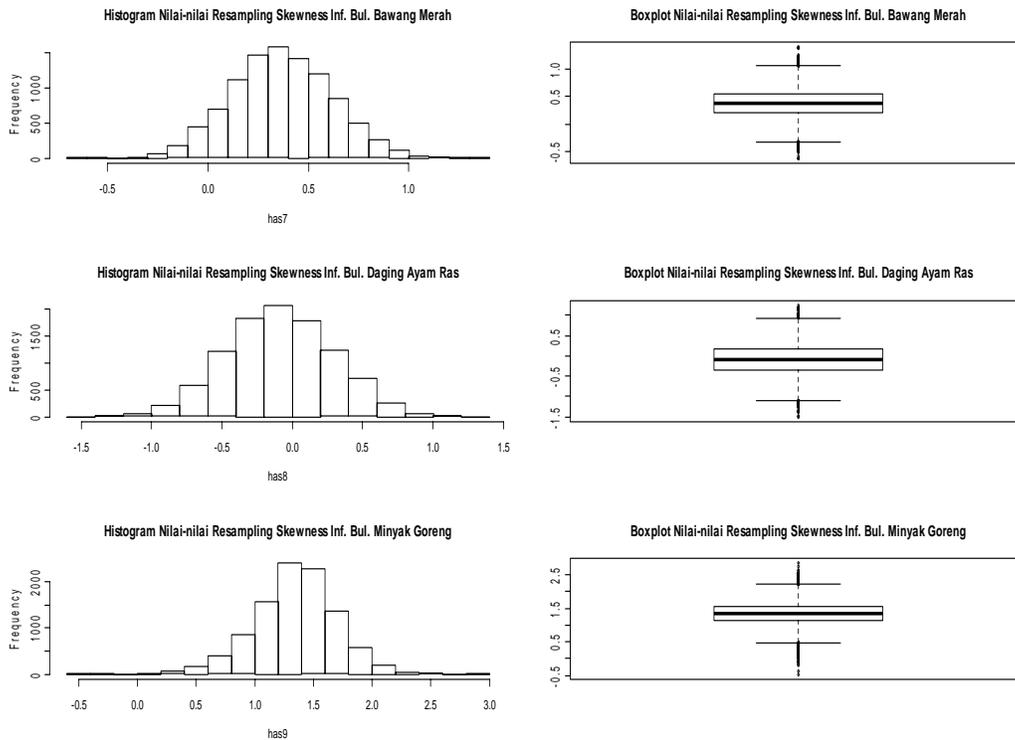
Distribusi normal baku mempunyai *skewness* 0 dan *kurtosis* 3, sedangkan distribusi eksponensial dengan mean 1 mempunyai *skewness* 2 dan *kurtosis* 9. Untuk distribusi chi-kuadrat dengan derajat bebas 1, akan mempunyai *skewness* $2\sqrt{2}$ dan *kurtosis* 15. Pada Tabel 4 dipresentasikan perbandingan antara lebar interval *bootstrap* dan lebar interval *resampling* untuk *skewness* dan *kurtosis* untuk berbagai nukuran sampel $n = 50, 100, 500, 1000$ dan 5000 dari distribusi normal baku, distribusi eksponensial dengan mean 1 dan distribusi chi-kuadrat dengan derajat bebas 1. Terlihat bahwa untuk ukuran sampel yang relatif kecil, lebar interval kepercayaan *bootstrap* lebih kecil, sedangkan untuk ukuran sampel yang lebih besar lebar interval kepercayaan *resampling* relatif lebih kecil.



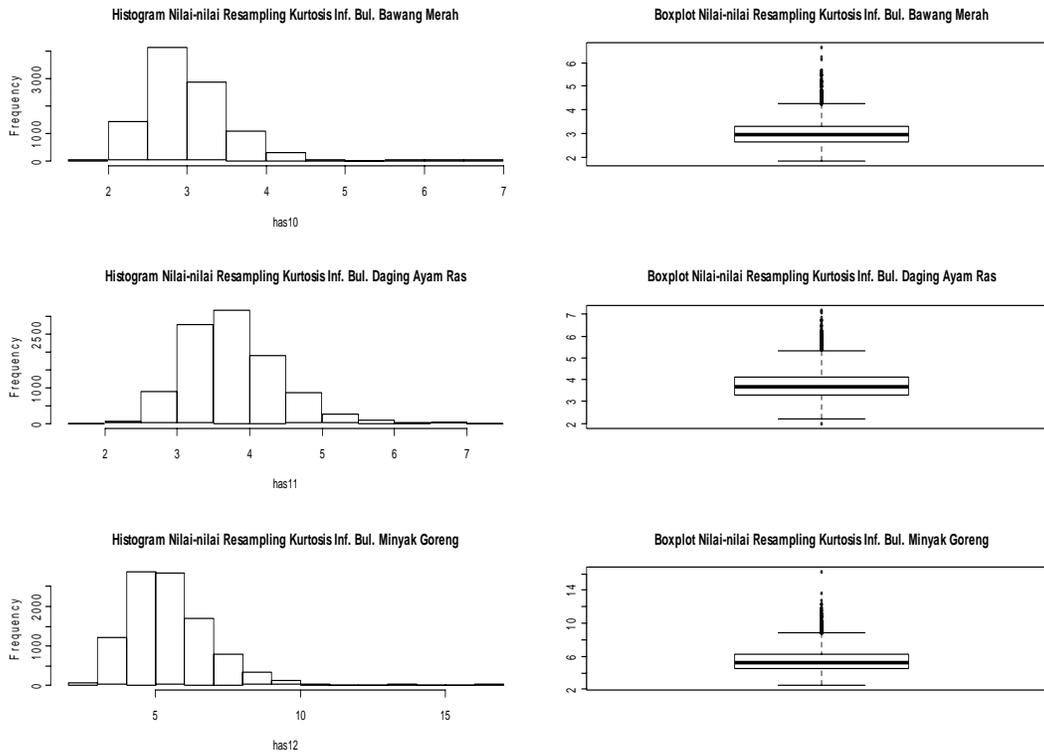
Gambar 4. Histogram dan boxplot nilai-nilai bootstrap *skewness* dari inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng periode Februari 2002 sampai dengan Desember 2007.



Gambar 5. Histogram dan boxplot dari nilai-nilai bootstrap *kurtosis* inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng periode Februari 2002 sampai dengan Desember 2007.



Gambar 6. Histogram dan boxplot dari nilai-nilai resampling skewness inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng periode Februari 2002 sampai dengan Desember 2007.



Gambar 7. Histogram dan boxplot dari nilai-nilai resampling kurtosis inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng periode Februari 2002 sampai dengan Desember 2007.

Tabel 3. Perbandingan antara estimasi interval bootstrap dan estimasi interval resampling untuk *skewness* dan *kurtosis*.

	skewness				kurtosis			
<i>Bootstrap</i>	BB	BA	Jarak antara BB dan BA	BB	BA	Jarak antara BB dan BA		
Bawang Merah	0.4677	-0.0333	0.9039	0.1145	0.0159	1.154		
Daras	-0.1253	-0.8587	0.7026	1.1215	0.0747	2.3749		
Migor	1.5416	0.6892	2.1199	1.4307	3.0423	0.4643		
	skewness				kurtosis			
<i>Resampling</i>	BB	BA	Jarak antara BB dan BA	BB	BA	Jarak antara BB dan BA		
Bawang Merah	0.4677	-0.1536	0.8668	1.0204	0.1145	2.2256		
Daras	-0.1253	-0.8264	0.6562	1.4826	1.1215	2.6486		
Migor	1.5416	0.5793	2.0046	1.4253	3.0423	3.3322		

KESIMPULAN

Dalam makalah ini, telah dijelaskan bagaimana mendapatkan distribusi statistik *skewness* dan *kurtosis* dengan metode *resampling* berdasar densitas kernel. Hasil yang diperoleh dibandingkan dengan lebar interval kepercayaan dengan metode *bootstrap*.

Metode yang dijelaskan, digunakan dalam kasus inflasi bulanan komoditas bawang merah, daging ayam ras dan minyak goreng di kota Semarang.

Tabel 4. Perbandingan antara lebar interval bootstrap dan lebar interval resampling untuk *skewness* dan *kurtosis* untuk berbagai ukuran sampel $n = 50, 100, 500, 1000$ dan 5000 dari distribusi normal baku, distribusi eksponensial dengan mean 1 dan distribusi chi-kuadrat dengan derajat bebas 1.

Skewness	n	bootstrap	resampling	Kurtosis		bootstrap	resampling
Normal	50	1.4165	1.8191	Normal	50	1.9337	2.2273
	100	0.5245	0.8217		100	0.6215	0.9273
	500	0.2469	0.3765		500	0.3907	0.4679
	1000	0.2177	0.3110		1000	0.3029	0.3692
	5000	0.0913	0.1334		5000	0.137	0.1405
Eksponensial	50	0.9007	1.1398	Eksponensial	50	4.5708	4.6079
	100	1.5915	2.2208		100	6.5032	6.5949
	500	0.6713	0.9076		500	2.8012	2.8065
	1000	0.405	0.6009		1000	1.9008	1.8633
	5000	0.4487	0.5966		5000	2.8375	2.9382
Chisq	50	0.8305	1.0134	Chisq	50	6.8797	7.0088
	100	1.1261	1.7104		100	11.6541	11.8004
	500	0.9647	1.2159		500	7.5008	7.3441
	1000	1.3948	1.6533		1000	10.1157	10.4477
	5000	0.3704	0.5364		5000	2.9392	2.893

DAFTAR PUSTAKA

- [1] Ankarali, Handan, Ayse Cananyazici, Seyit Ankarali, 2009, A bootstrap Confidence Interval for Skewness and Kurtosis and Properties of t-test in Small Sample from Normal Distribution, Presented at the XI. *National Congress of Biostatistics, May 27-30, 2008, Malatya, Turkey.*
- [2] de Gunst, M. C. M., 1994, *Statistische Data Analyse*, Faculteit Wiskunde en Informatica, Vrije Universiteit Amsterdam.
- [3] Sheather, S. J. and Jones M. C. , 1991, A reliable data-based bandwidth selection method for kernel density estimation. *J. Roy. Statist. Soc. B*, 683–690
- [4] Setiawan, Adi , 2002, Simulasi Estimasi Fungsi Kepadatan Probabilitas, *Konferensi Nasional Matematika XI, FMIPA Universitas Negeri Malang 22-25 Juli 2002.*