

RESAMPLING BERDASARKAN ESTIMASI DENSITAS KERNEL BIVARIAT

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

e-mail : adi_setia_03@yahoo.com

Abstrak

Resampling berdasarkan estimasi densitas kernel sudah diusulkan dalam makalah Setiawan (2002). Dalam makalah ini dijelaskan bagaimana melakukan resampling berdasarkan estimasi kernel bivariat yang sekaligus juga dapat diperluas dalam kasus multivariat. Metode yang digunakan dijelaskan dengan menggunakan data bivariat karakteristik pH dan berat jenis sabun siri "X". Kelebihan metode ini dibandingkan metode bootstrap adalah bahwa titik-titik sampel yang diperoleh tidak harus bernilai sama dengan titik-titik sampel awal. Metode yang diusulkan relatif lebih baik dibandingkan dengan metode bootstrap.

Kata kunci : *estimasi densitas kernel, estimasi densitas kernel bivariat, koefisien korelasi Pearson, metode bootstrap.*

Pendahuluan

Dalam makalah Setiawan (2002) telah dijelaskan bagaimana membuat sampel baru berdasarkan estimasi densitas kernel univariat. Demikian juga metode yang sama digunakan dalam makalah Setiawan (2012). Metode bootstrap banyak digunakan dalam penentuan distribusi statistik yang sulit dihitung dengan menggunakan cara analitis. Beberapa makalah relatif baru yang berkaitan dengan penggunaan metode bootstrap ada dalam makalah Bishara & Hittner (2010), Maturi, Tahani & Elsayigh (2010) dan Mudelsee (2003). Dalam makalah ini akan dijelaskan bagaimana membuat sampel baru berdasarkan estimasi densitas kernel bivariat. Sampel yang diperoleh digunakan untuk menghitung statistik koefisien korelasi Pearson. Hasil yang diperoleh dibandingkan dengan sampel baru yang dibangkitkan dengan metode bootstrap.

Dasar Teori

Dalam dasar teori ini dijelaskan tentang estimasi densitas kernel multivariat yang bisa digunakan untuk menentukan estimasi densitas kernel bivariat bivariat.

Estimasi Fungsi Densitas Kernel Multivariat

Misalkan dimiliki sampel bivariat berukuran n yaitu X_1, X_2, \dots, X_n yang diambil dari suatu populasi dengan fungsi densitas f , maka estimasi fungsi densitas kernelnya adalah

$$\hat{f}(x; H) = \frac{1}{n} \sum_{i=1}^n K_H(x - X_i) \quad (1)$$

dengan H adalah matriks *bandwidth*, $x = (x_1, x_2)^T$, $X_i = (X_{i1}, X_{i2})^T$ untuk $i = 1, 2, \dots, n$, kernel $K_H(x) = |H|^{-1/2} K(H^{-1/2} x)$, $H = \begin{bmatrix} a^2 & b \\ b & d^2 \end{bmatrix}$ adalah matriks *bandwidth* yang simetris positif definit (*definite positive*) artinya semua eigen valuenya positif. Dalam hal ini, $a = \text{var}(X_{i1})$, $d = \text{var}(X_{i2})$, $b = \text{covar}(X_{i1}, X_{i2})$ dan $K(x) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} x^T x\right)$ adalah kernel normal baku bivariat. Pemilihan matriks *bandwidth* yang optimal mengacu pada paper Tarn Duong (2003). Dalam pemilihan matriks optimal ini digunakan perintah pada paket program *ks* yang ada pada CRAN-R.

Metode Bootstrap

Metode bootstrap secara praktis dapat dijelaskan berikut ini. Misalkan dimiliki sampel awal ukuran n yaitu X_1, X_2, \dots, X_n . Sampel baru ukuran m dibangkitkan dengan dengan cara mengambil sampel awal dengan pengembalian sehingga diperoleh sampel baru $X_1^*, X_2^*, \dots, X_m^*$. Statistik $T(X_1^*, X_2^*, \dots, X_m^*)$ dihitung dengan menggunakan sampel baru. Untuk sampel bivariat maka sampel baru yang terambil merupakan sampel berpasangan sehingga hal itu dapat dilakukan dengan cara mengambil angka antara 1 sampai n misalkan diperoleh i dengan $i=1, 2, \dots, n$ maka sampel baru ke-1 yaitu $X_1^* = X_i$, prosedur tersebut diulang sampai m kali sehingga diperoleh sampel baru ukuran m yaitu $X_1^*, X_2^*, \dots, X_m^*$.

Metode bootstrap dapat digambarkan dalam sampel bivariat berikut ini. Misalkan dimiliki sampel awal

(3.87, 1.0009), (3.81, 1.0029), (3.77, 1.0024), (3.79, 1.0021), (3.78, 1.0024).

Tiga sampel baru hasil metode bootstrap diantaranya adalah

(3.81, 1.0029), (3.81, 1.0029), (3.87, 1.0009), (3.77, 1.0024), (3.87, 1.0009);

(3.81, 1.0029), (3.81, 1.0029), (3.77, 1.0024), (3.78, 1.0024), (3.78, 1.0024);

(3.87, 1.0009), (3.87, 1.0009), (3.87, 1.0009), (3.81, 1.0029), (3.78, 1.0024).

Sampel baru pertama terambil titik sampel (2,2,1,3,1) artinya titik sampel kedua yang terambil pertama, titik sampel kedua yang terambil kedua, titik sampel pertama terambil ketiga, titik sampel ketiga yang terambil keempat dan akhirnya titik sampel pertama yang terambil kelima. Demikian juga dapat dijelaskan untuk sampel baru kedua dan sampel baru ketiga. Ketiga sampel baru ini digunakan untuk menghitung statistik koefisien korelasi Pearson yaitu berturut-turut diperoleh -0.83278, 0.9759 dan -0.9068.

Metode Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder pada proses produksi sabun sirih pada bulan September 2010 sampai dengan Desember 2010 sebanyak 200 titik sampel yang termuat dalam Pattihahuan (2012). Adapun karakteristik kualitas produk Sabun Sirih "X" yang digunakan dalam penelitian ini adalah pH dengan batas spesifikasi perusahaan adalah 3.5 sampai dengan 3.9 dan berat jenis dengan batas spesifikasi perusahaan adalah 0.9834 sampai dengan 1.0227.

Untuk membangkitkan sampel baru dari distribusi yang mempunyai densitas kernel bivariat seperti dinyatakan pada persamaan (1) digunakan langkah-langkah sebagai berikut :

1. Misalkan dimiliki sampel bivariat ukuran n yaitu X_1, X_2, \dots, X_n .
2. Pilihlah matriks *bandwidth* optimal H dengan perintah pada paket program *ks* untuk mengestimasi densitas kernel bivariat berdasarkan sampel X_1, X_2, \dots, X_n .
3. Sampel $X_i^* = (X_{i1}^*, X_{i2}^*)^T$ dibangkitkan di sekitar X_i dengan cara membangkitkan sampel ukuran 1 dari distribusi normal bivariat dengan mean X_i dan variansi sama dengan matriks *bandwidth* optimal H .
4. Langkah 3 diulang sebanyak ukuran sampel baru m yang ingin dibuat sehingga diperoleh $X_1^*, X_2^*, \dots, X_m^*$.
5. Sampel baru tersebut selanjutnya digunakan dalam perhitungan statistik $T(X_1^*, X_2^*, \dots, X_m^*)$.

Metode resampling berdasarkan estimasi densitas kernel dapat digambarkan dalam sampel bivariat berikut ini. Misalkan dimiliki sampel awal (3.87, 1.0009), (3.81, 1.0029), (3.77, 1.0024), (3.79, 1.0021), (3.78, 1.0024). *Bandwidth* optimal H berdasarkan sampel awal adalah

$$H = \begin{pmatrix} 5.7239 \times 10^{-4} & -8.2596 \times 10^{-6} \\ -8.2596 \times 10^{-6} & 2.0290 \times 10^{-7} \end{pmatrix}$$

Dengan menggunakan *bandwidth* optimal maka dilakukan resampling untuk mendapatkan sampel baru. Tiga sampel baru hasil metode resampling berdasarkan estimasi densitas kernel diantaranya adalah

(3.80, 1.0017), (3.78, 1.0022), (3.85, 1.0024), (3.83, 1.0026), (3.81, 1.0031);
 (3.84, 1.0022), (3.82, 1.0036), (3.89, 1.0004), (3.78, 1.0022), (3.87, 1.0008);
 (3.75, 1.0029), (3.77, 1.0029), (3.79, 1.0024), (3.83, 1.0025), (3.78, 1.0026).

Sampel baru yang diperoleh berada di sekitar titik-titik sampel awal sehingga tidak sama persis dengan titik-titik sampel awal. Ketiga sampel baru ini digunakan untuk menghitung statistik koefisien korelasi Pearson yaitu berturut-turut diperoleh 0.3197, -0.7117 dan -0.6882.

Untuk mengukur kebaikan sampel baru yang dibangkitkan berdasarkan estimasi densitas kernel bivariat dibandingkan dengan metode bootstrap digunakan jarak antara statistik koefisien korelasi Pearson sampel awal dan statistik koefisien korelasi Pearson dengan sampel baru yang dibangkitkan berdasarkan estimasi densitas kernel bivariat dengan jarak antara statistik koefisien korelasi Pearson sampel awal dan statistik koefisien korelasi Pearson dengan sampel baru yang dibangkitkan dengan menggunakan metode bootstrap. Bila prosedur ini diulang sebanyak bilangan besar B kali maka akan diperoleh ukuran kebaikan kedua metode tersebut.

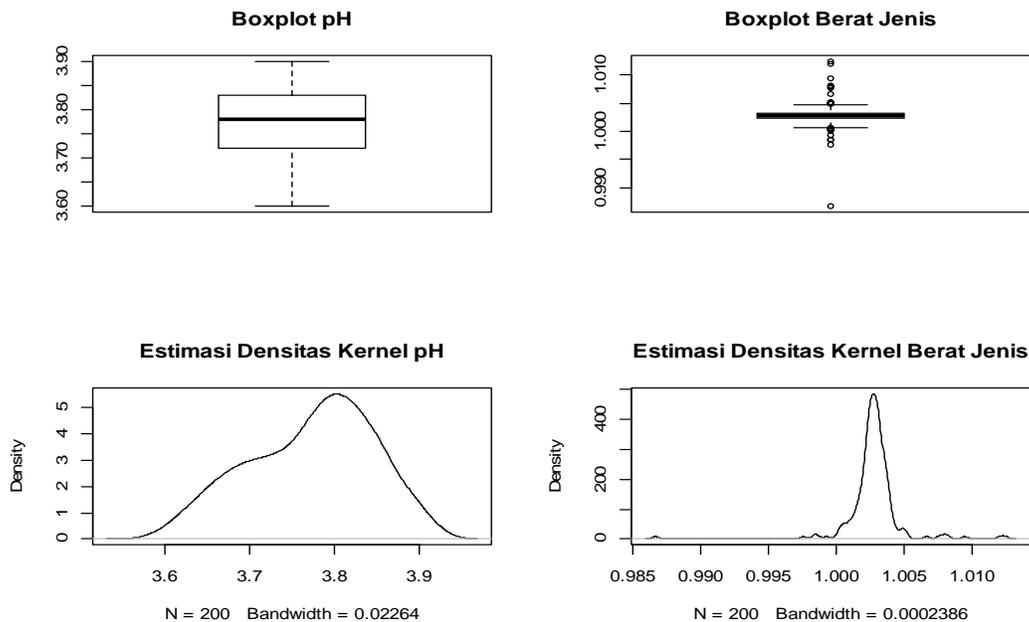
Hasil Analisis dan Pembahasan

Berdasarkan data diperoleh statistik deskriptif yang dinyatakan dalam Tabel 1. Di samping itu boxplot dan estimasi densitas kernel univariat untuk masing-masing karakteristik dapat dinyatakan pada Gambar 1. Terlihat bahwa karakteristik pH Sabun Sirih “X” lebih bervariasi dibandingkan karakteristik berat jenisnya. Hal itu juga

didukung oleh koefisien variasi karakteristik pH sebesar 0.0192 yang lebih besar dari koefisien variasi berat jenis yaitu sebesar 0.0020 serta gambar boxplot dan estimasi densitas kernel pada Gambar 1. Hal itu berarti bahwa data pH lebih menyebar dibandingkan data berat jenis.

Tabel 1. Statistik Deskriptif karakteristik pH dan berat jenis sabun sirih .

Statistik	pH	Berat Jenis
Minimum	3.6000	0.9867
Kuartil 1	3.7200	1.0020
Median	3.7800	1.0030
Mean	3.7730	1.0030
Kuartil 3	3.8300	1.0030
Maksimum	3.9000	1.0120



Gambar 1. Boxplot dan Estimasi Densitas Kernel (univariat) untuk pH dan Berat Jenis sabun sirih “X”.

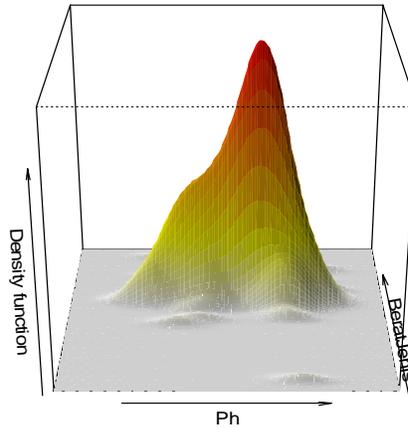
Korelasi Pearson antara kedua karakteristik tersebut adalah -0.1191 yang mempunyai nilai-p sebesar 0.093 sehingga jika digunakan tingkat signifikansi 10 % akan signifikan artinya pH berkorelasi negatif terhadap berat jenis. Jika berat jenis besar maka pH cenderung bernilai kecil dan sebaliknya jika berat jenis kecil maka pH cenderung bernilai besar.

Berdasarkan data bivariat pH dan berat jenis sabun sirih “X” diperoleh matriks *bandwidth* optimal *H* dengan menggunakan paket *ks* pada *software* R-2.15.1 yaitu

$$H = \begin{pmatrix} 7.1429 \times 10^{-4} & -2.0501 \times 10^{-6} \\ -2.0501 \times 10^{-6} & 2.5299 \times 10^{-6} \end{pmatrix}$$

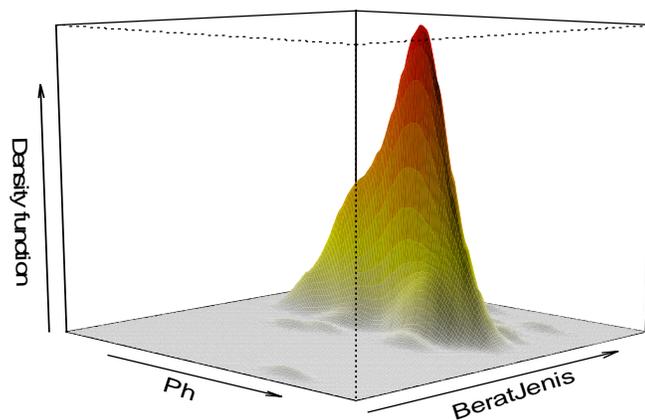
dengan eigen value $\lambda_1 = 7.1429 \times 10^{-4}$ dan $\lambda_2 = 2.4710 \times 10^{-7}$ sehingga *bandwidth H* positif definit. Estimasi densitas kernel bivariat untuk karakteristik pH dan berat jenis bila digunakan sudut rotasi horizontal (AZ) 0 derajat dan sudut elevasi vertikal (EL) 25 derajat dinyatakan pada Gambar 2. Terlihat bahwa ada beberapa titik yang terlihat jauh dibandingkan dengan kumpulan titik yang lain. Gambar 3 memperlihatkan estimasi densitas kernel bivariat untuk karakteristik pH dan berat jenis bila digunakan AZ 45 dan EL 10.

Estimasi Densitas Kernel Bivariat



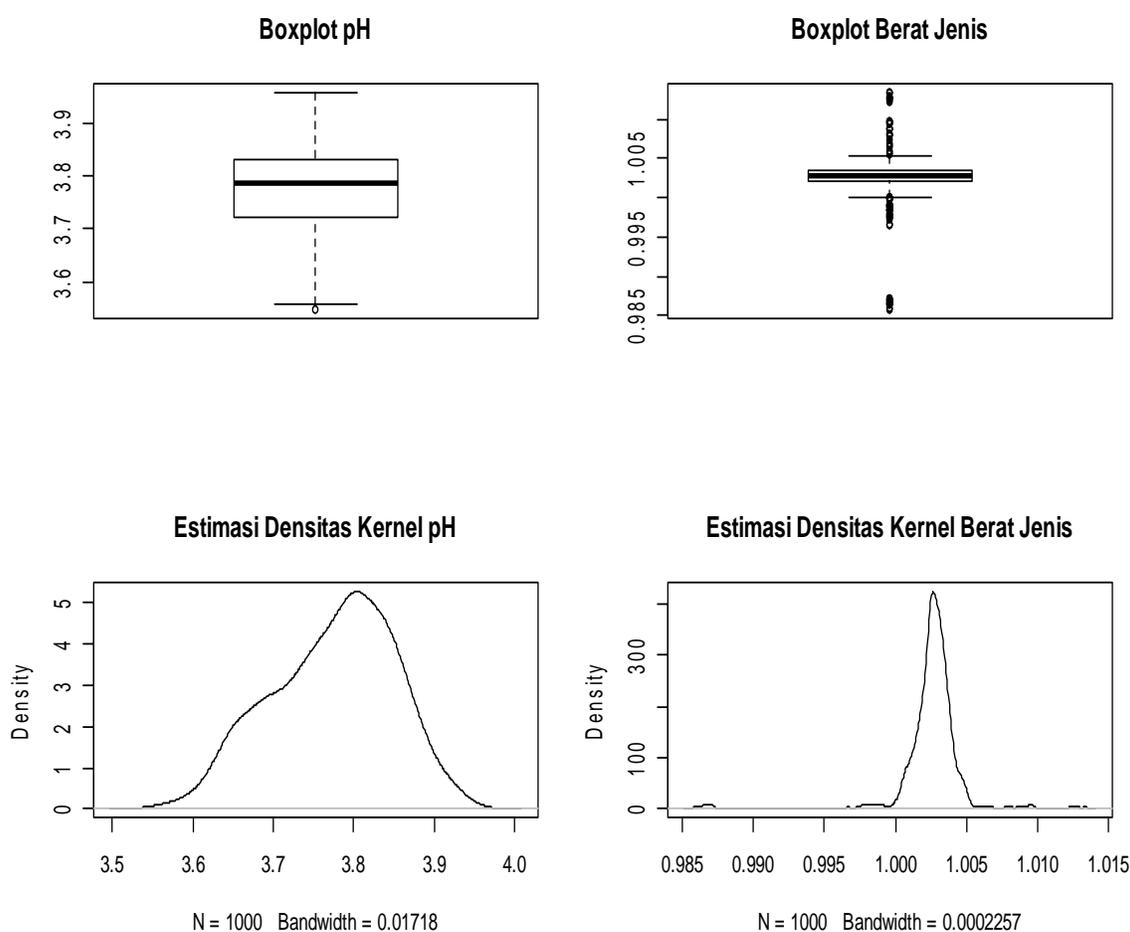
Gambar 2. Estimasi Densitas Kernel Bivariat untuk pH dan Berat Jenis bila dilihat dari AZ 0 dan EL 25.

Estimasi Densitas Kernel Bivariat



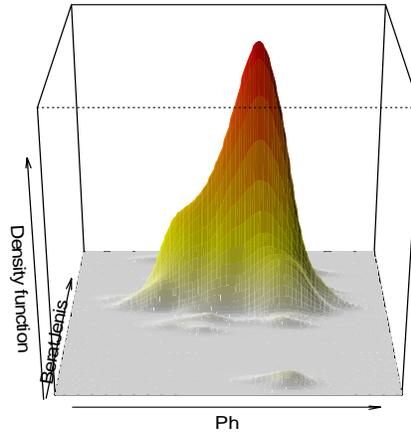
Gambar 3. Estimasi Densitas Kernel Bivariat untuk pH dan Berat Jenis bila dilihat dari AZ 45 dan EL 10.

Sampel baru ukuran m dibangkitkan dengan menggunakan metode resampling berdasarkan estimasi densitas kernel bivariat. Gambar 4 memperlihatkan boxplot dan estimasi densitas kernel univariat untuk pH dan berat jenis jika digunakan ukuran sampel $m=1000$. Terlihat bahwa kita bisa membangkitkan sampel dengan ukuran sampel lebih besar dari sampel asal dan mempunyai kemiripan sifat seperti sampel asal yaitu mempunyai mean (pH,berat-jenis) = (3.7747,1.0026) dan koefisien korelasi Pearson -0.1191 (nilai- p lebih kecil dari 0.0002). Dengan sampel yang sama koefisien variasi untuk pH adalah 0.0201 dan koefisien variasi untuk berat jenis adalah 0.0023. Hasil estimasi densitas kernel bivariat untuk sampel baru ukuran $m = 1000$ tersebut dinyatakan pada Gambar 5.



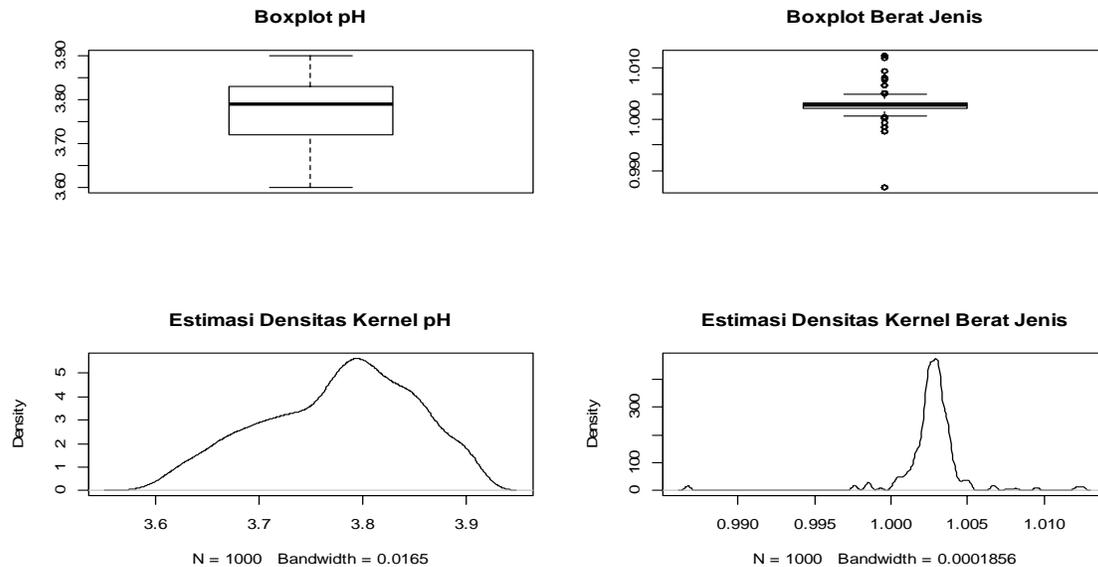
Gambar 4. Boxplot dan Estimasi Densitas Kernel (univariat) untuk pH dan Berat Jenis sabun sirih "X" hasil pembangkitan dengan metode resampling berdasar estimasi densitas kernel bivariat.

Estimasi Densitas Kernel Bivariat



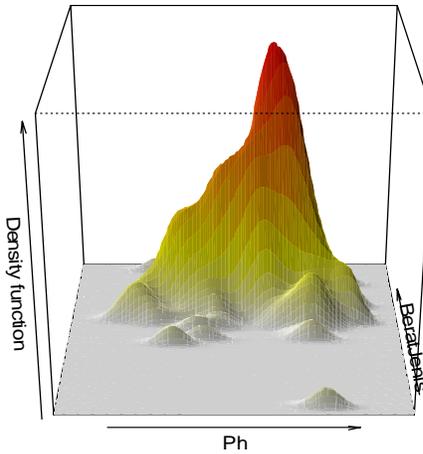
Gambar 5. Estimasi Densitas Kernel Bivariat untuk pH dan Berat Jenis bila dilihat dari AZ 0 dan EL 25.

Apabila digunakan metode bootstrap untuk mendapatkan sampel baru ukuran $m = 1000$. Gambar 6 memperlihatkan boxplot dan estimasi densitas kernel univariat untuk pH dan berat jenis sampel baru tersebut. Dalam hal ini sampel baru mempunyai mean $(\text{pH}, \text{berat-jenis}) = (3.7767, 1.0027)$ dan koefisien korelasi Pearson -0.1077 (nilai- p lebih kecil dari 0.0007). Dengan sampel yang sama koefisien variasi untuk pH adalah 0.0193 dan koefisien variasi untuk berat jenis adalah 0.0022 . Hasil estimasi densitas kernel bivariat untuk sampel baru ini dinyatakan pada Gambar 7.



Gambar 6. Boxplot dan Estimasi Densitas Kernel (univariat) untuk pH dan Berat Jenis sabun sirih "X" hasil pembangkitan dengan metode bootstrap.

Estimasi Densitas Kernel Bivariat



Gambar 7. Estimasi Densitas Kernel Bivariat untuk pH dan Berat Jenis bila dilihat dari AZ 0 dan EL 25.

Tabel 2. Proporsi jarak antara koefisien korelasi sampel awal dan sampel baru dengan metode estimasi densitas kernel lebih besar jarak antara koefisien korelasi sampel awal dan sampel baru dengan metode bootstrap.

B	<i>m</i> = 200	<i>m</i> = 500	<i>m</i> = 1000
50	0.440	0.480	0.440
100	0.500	0.540	0.520
200	0.550	0.465	0.495
500	0.482	0.504	0.480
1000	0.501	0.466	0.505

Untuk membandingkan kebaikan sampel baru yang dibangkitkan kedua metode tersebut di atas digunakan studi simulasi yang mengukur koefisien korelasi Pearson dari 2 sampel bivariat baru yang diperoleh berdasarkan kedua metode tersebut. Misalkan digunakan ukuran sampel baru $m = 200$, kemudian dibandingkan proporsi jarak antara koefisien korelasi yang dihasilkan oleh sampel baru dengan menggunakan metode resampling berdasarkan estimasi densitas kernel bivariat yang lebih besar dibandingkan jarak antara koefisien korelasi sampel asal dibandingkan dengan sampel baru yang diperoleh dengan menggunakan metode bootstrap. Hasil yang diperoleh untuk berbagai nilai B dinyatakan pada Tabel 2. Terlihat bahwa metode yang diusulkan memberikan rata-rata proporsi yang relatif lebih kecil yaitu 0,4912 sehingga relatif lebih baik dibandingkan dengan metode bootstrap.

Kesimpulan

Dalam makalah ini telah dijelaskan bagaimana membangkitkan sampel baru (*resampling*) berdasarkan estimasi densitas kernel bivariat. Kelebihan metode ini dibandingkan metode bootstrap adalah bahwa titik-titik sampel yang diperoleh tidak harus bernilai sama dengan titik-titik sampel awal. Metode yang diusulkan relatif lebih baik dibandingkan dengan metode bootstrap.

Daftar Pustaka

Bishara, Anthony J. & James B. Hittner (2012) Testing the Significance of a Correlation with Non-normal Data: Comparison of Pearson, Spearman, Transformation, and Resampling Approaches, *In Press*.

Maturi, Tahani A. & Anga Elsayigh (2010) A Comparison of Correlation Coefficients via a Three-Step Bootstrap Approach, *Journal of Mathematical Research* Vol. 2 No. 2.

Mudelsee, Manfred (2003) Estimating Pearson's Correlation Coefficient with Bootstrap Confidence Interval from Serially Dependent Time Series, *Mathematical Geology* Vol. 35 pp 651-665.

Pattihahuan, S. (2012) Penerapan Grafik Pengendali dan Studi Simulasi Estimasi Fungsi Densitas Kernel Bivariat, *Skripsi Mahasiswa Fakultas Sains dan Matematika, Universitas Kristen Satya Wacana Salatiga*.

Setiawan, Adi (2002) Simulasi Estimasi Fungsi Kepadatan Probabilitas, *Konferensi Nasional Matematika XI, FMIPA Universitas Negeri Malang 22-25 Juli 2002*.

Setiawan, Adi (2012) 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), *Prosiding Seminar Nasional Sains dan Pendidikan Sains*, Vol 3 No 1, ISSN 2087-0922