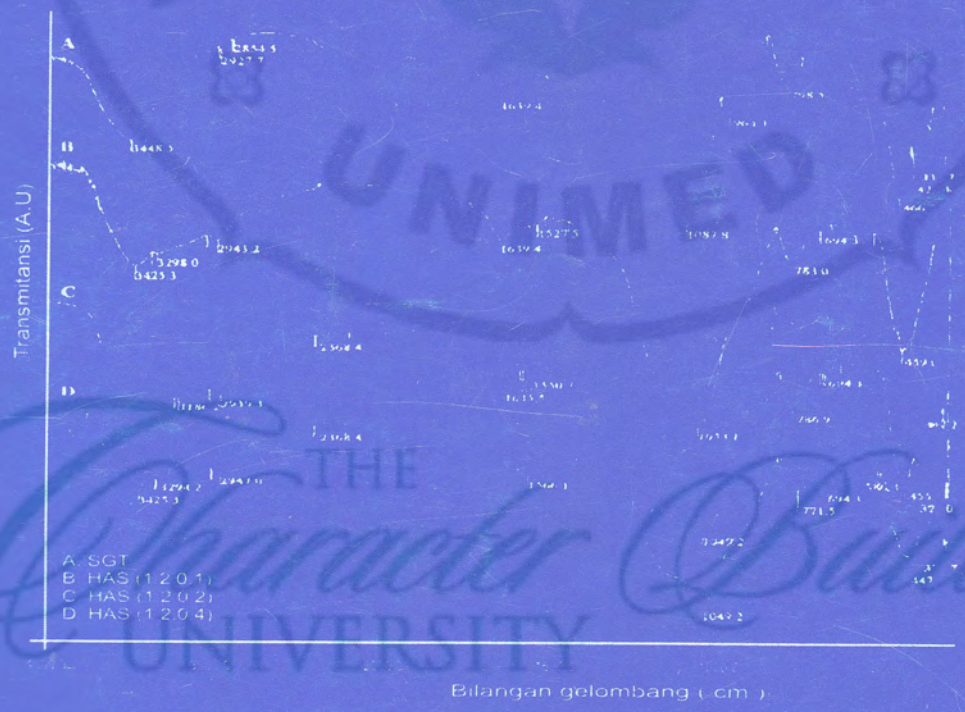
Volume 31 | Nomor 2 | Juli–Desember 2007 | ISSN 1978-3841

2A

Jurnal Sains Indonesia

Media Komunikasi Hasil Penelitian Sains dan Matematika



Diterbitkan Oleh
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Medan

ISSN 1978-3841

Jurnal Sains Indonesia

Media Komunikasi Hasil Penelitian Sains dan Matematika

Pembina
Prof. Dr. Djanius Djamin, S.H., M.S. (*Rektor Unimed*)
Drs. Syawal Gultom, M.Pd. (*Pembantu Rektor II*)
Prof. Dr. Albinus Silalahi, M.S. (*Pembantu Rektor III*)
Prof. Drs. Manihar Situmorang, M.Sc., Ph.D. (*Dekan FMIPA*)

Dewan Penyunting
Prof. Drs. Manihar Situmorang, M.Sc., Ph.D. (*Ketua*)
Drs. Pasar Maulim Silitonga, M.S. (*Wakil*)
Dra. Martina Restuati, M.Si. (*Wakil*)
Drs. Asrin Lubis, M.Pd. (*Anggota*)
Dr. Pargaulan Siagian, M.Pd. (*Anggota*)
Dr. Ridwan Abdul Sani, M.Si. (*Anggota*)
Prof. Dr. Suharta, M.Si. (*Anggota*)
Dr. rer. nat. Binari Manurung, M.Si. (*Anggota*)

Penyunting Ahli
Dr. Herbert Sipahutar, M.S., M.Sc.
Dr. Zainuddin M., M.Si.
Dr. A.K. Prodjosantoso
Dr. Ali Imron

Tata Usaha
Drs. Zulkifli
Dra. Sion Asmarida Purba
Tua P. Tambunan

Jurnal Sains Indonesia *(dahulu bernama Majalah Pendidikan Science) diterbitkan sejak tahun 1976, dengan SK Menteri Penerangan Republik Indonesia STT Penerbit Khusus tanggal 9 Desember 1976, No. 276/SK/Ditjen PPG/STT/1976. Jurnal ini diterbitkan untuk dapat digunakan sebagai media komunikasi bagi dosen, peneliti, mahasiswa semua strata bidang sains dan matematika. Pengelola menerima artikel hasil penelitian, catatan penelitian dan/atau telaah pustaka dalam bidang ilmu yang relevan. Petunjuk penulisan naskah dapat dilihat pada kulit belakang bagian dalam dari jurnal ini. Naskah dikirimkan ke alamat redaksi untuk dievaluasi dan disunting . Naskah yang tidak memenuhi persyaratan akan dikembalikan kepada penulis.*

Diterbitkan oleh:
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Medan

Alamat Redaksi:
Jurnal Sains Indonesia
Jl. Willem Iskandar Pasar V, Medan 20221
Telp. 061-6625970
E-mail: fmipa-unimed@medan .wasantara.net.id

Dari Pengelola

Volume 31 Nomor 2 Jurnal Sains Indonesia tahun 2007 ini hadir kembali di hadapan sejawat dengan 10 artikel hasil penelitian dari bidang biologi, kimia, fisika, dan matematika. Masing-masing artikel memuat informasi baru yang sangat menarik dan penting untuk dirujuk dalam upaya pengembangan sains dan matematika.

Pengelola menanti artikel-artikel orisinal yang bermutu dari kita semuanya. Namun demikian, sebelum menulis artikel, sangat diharapkan para penulis mempelajari terlebih dahulu, dan selanjutnya mengikuti, petunjuk penulisan artikel yang tertera pada sampul belakang bagian dalam dari jurnal ini. Pengelola juga mengharapkan agar gambar, foto, atau diagram yang akan dimuat dalam artikel dibuat dalam lembaran terpisah dan sudah dalam bentuk *camera ready*. Untuk memudahkan editing, sejawat penulis sangat diharapkan *hanya menggunakan* fasilitas konstruksi tabel standar yang tersedia pada *software* Microsoft Word atau Exel, dan *tidak menggunakan* fasilitas *drawing* yang ada pada software tersebut. Untuk memperkecil kerusakan file akibat gangguan virus, ada baiknya jika sejawat penulis mengirimkan *soft copy* dalam bentuk *rich text format* (rft) menyertai *hard copy* artikelnya masing-masing.

Untuk nomor berikutnya, artikel yang tidak memenuhi persyaratan berdasarkan petunjuk penulisan tersebut tidak akan dimuat dan akan dikembalikan kepada penulis untuk diperbaiki.

Selamat berkarya.

Medan, Desember 2007

Pengelola

THE Character Building UNIVERSITY

Volume 31 | Nomor 2 | Juli – Desember 2007
Jurnal
Sains
Indonesia
Media Komunikasi Hasil Penelitian Sains dan Matematika

Daftar Isi

Zul Amry	Analisis Bootstrap pada Model Regresi	51 – 54
Faiz Ahyaningsih	Menyelesaikan *Quadratic Assignment Problem* dengan Metode Heuristik Kelayakan	55 – 58
Adriana Y.D. Lbn Gaol	Gangguan Siklus Estrus Pasca Aktivitas Fisik Berat pada Mencit Merupakan Manifestasi Pemanjangan Diestrus	59 – 64
Erlintan Sinaga	Manfaat Kotoran Sapi terhadap Jumlah *Vermicomposting*	65 – 67
Mariaty Sipayung	Pengaruh Bahan Baku Terhadap Kecepatan dan Kualitas Kompos dengan Metode *Vermicomposting*	68 – 70
Jiman Girsang	Kandungan Logam Berat pada Tanaman Bayam (*Amaranthus sp*) yang Ditanam pada Media Tanam Kompos	71 – 73
Lisnawaty, Narsito dan Nuryono	Sintesis dan Karakterisasi Silika Gel-Aminopropil Trimetoksisilan dari Tetraetoksiortosilan	74 – 78
Asmin	Penerapan Metode Newton-Raphson dengan Cara Iterasi untuk Mengestimasi Parameter Kemampuan Responden untuk Model Logistik pada Teori Responsi Butir	79 – 83
Evita Mayasari	Kepentingan Asam Amino Glisin Bagi Kehidupan Sel	84 – 87
Simson Tarigan	Fenomena Fisik Teori Orbital Molekuler dalam Pembentukan Molekul-Molekul Diatomik Homonuklir	88 – 91

Jurnal Sains Indonesia
Volume 31(2): 51-54, 2007

Analisis Bootstrap pada Model Regresi

Zul Amry
Jurusan Matematika, FMIPA, Universitas Negeri Medan
Jl. Willem Iskandar Pasar V, Medan 20221

Diterima 31 Agustus 2007, disetujui untuk publikasi 08 September 2007

Abstract *The bootstrap analysis to represent of statistical analysis with way resampling towards the original data in endeavour obtain the best of estimator or approximation of distribution. In this paper will be refered the best approximation of normal distribution for $\sqrt{n}(\hat{\beta}-\beta)$ with bootsrap analysis. The methods to plan a program of computer simulation to compare between the approximation of distribution for $\sqrt{n}(\hat{\beta}^{*}-\hat{\beta})$ where β is parameter of population, $\hat{\beta}$ is least square estimator and $\hat{\beta}^{*}$ is bootstrap estimator toward a normal distribution according simulation.* [BOOTSTRAP ANALYSIS IN REGRESSION MODEL] *(J. Sains Indon., 31(2): 51-54, 2007)*

Kata kunci: Bootstrap approximation, normal distribution, regression model

Pendahuluan

Dalam inferensi statistika, pendekatan (hampiran) terhadap suatu distribusi sering dilakukan, karena distribusi yang eksak tidak selalu mudah didapat. Pada model regresi $Y = X\beta + \varepsilon$ permasalahan yang lazim dijumpai adalah bagaimana cara mengestimasi parameter β dengan estimator $\hat{\beta}$ berdasarkan sampel observasi $x_1, x_2, \ldots, x_n$. Salah satu cara untuk menentukan $\hat{\beta}$ adalah dengan metode kuadrat terkecil, yang menghasilkan $\hat{\beta} = (X^T X)^{-1}.X^T Y$.

Namun, yang menjadi permasalahan adalah seberapa dekatnya $\hat{\beta}$ dengan β, ini disebabkan oleh nilai deviasi random $(\hat{\beta}-\beta)$ yang sifatnya berubah-ubah. Karena itu diperlukan suatu estimasi langsung terhadap statistik $(\hat{\beta}-\beta)$. Dari pembahasan yang pernah dilakukan terhadap model regresi, diketahui bahwa statistik $\sqrt{n}(\hat{\beta}-\beta)$ menghampiri suatu distribusi normal dengan mean sama dengan nol. Sehubungan dengan itu, berarti permasalahannya sekarang adalah bagaimana mendapatkan hampiran distribusi normal yang terbaik untuk $\sqrt{n}(\hat{\beta}-\beta)$.

Metode *bootstrap* merupakan bentuk umum dari metode *jackknife* yang pertama kali diperkenaikan oleh Efron (1979). Cara kerja metode bootstrap adalah melakukan penyampelan kembali (*resample*) terhadap sampel aslinya sehingga diperoleh estimator atau hampiran distribusi yang lebih baik. Semenjak itu, metode bootstrap terus berkembang dan banyak digunakan dalam menganalisis model-model statistik, diantaranya Freedman (1984) yang menggunakan bootstrap pada masalah estimasi kuadrat terkecil model linear stasioner. Stine (1985) menggunakan bootstrap dalam memprediksi interval pada regresi. Sedangkan Thomb dan Scucany (1990) juga menggunakan bootsirap dalam memprediksi interval pada model autoregresif.

Estimator yang didapat melalui metode bootstrap disebut *estimator bootstrap*. Jika estimator bootstrap dinyatakan dengan $\hat{\beta}^{*}$, maka prinsip bootstrap pada model regresi adalah menentukan distribusi $\sqrt{n}(\hat{\beta}^{*}-\hat{\beta})$ yang dapat dihitung langsung dari data sampel sebagai hampiran untuk distribusi $\sqrt{n}(\hat{\beta}-\beta)$.

Secara teoritis, $\sqrt{n}(\hat{\beta}^{*}-\hat{\beta})$ dapat dihitung secara manual, tetapi dalam prakteknya sangat sukar untuk dikerjakan. Sebab untuk sampel random $X_1, X_2, \ldots, X_n$ dengan n ukuran data yang berbeda, akan terdapat sebanyak n^n sampel bootstrap yang mungkin. Bayangkan untuk n=10 akan terdapat 10^{10} sampel bootstrap yang dapat dienumerasi. Sebab itu masalah bootstrap biasanya dikerjakan dengan bantuan komputer.

Berdasarkan uraian di atas, dalam penelitian ini dicoba untuk mendapatkan hampiran distribusi normal terbaik untuk $\sqrt{n}(\hat{\beta}-\beta)$ melalui simulasi komputer dengan metode bootstrap.

Model regresi. Model regresi mempunyai bentuk $Y = X\beta + \varepsilon$, dimana:

$$Y = (y_1, y_2, \ldots, y_n)$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1,p-1} \\ 1 & x_{21} & x_{22} & \cdots & x_{2,p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{n,p-1} \end{bmatrix} \text{ matriks nxp}$$

dengan rank penuh $p \leq n$ $\beta = (\beta_0, \beta_1, \ldots, \beta_n)$ parameter yang dapat diestimasi dari data $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ residu random dengan $E(\varepsilon_i)=0$, $Var(\varepsilon_i) = \sigma^2 < \infty$ dan $cov(\varepsilon_i, \varepsilon_j)=0$ untuk semua i, j ; $i \neq j$.

Estimasi kuadrat terkecil. Berdasarkan data observasi model regresi dapat dituliskan dalam bentuk $Y=X\beta+e$. Estimasi kuadrat terkecil merupakan salah satu metode untuk mendapatkan suatu estimasi parameter β. Prinsipnya adalah dengan meminimumkan $\sum e_i^2$ terhadap β dari bentuk $e=Y - X\beta$, yang menghasilkan $\hat{\beta} = (X^T X)^{-1}.X^T.Y$.

Regresi bootstrap. Model regresi bootstrap dapat dinyatakan dengan himpunan data yang berbentuk :

$$\{(x_1, x_1\hat{\beta} + e_1^{*}), (x_2, x_2\hat{\beta} + e_2^{*}), \ldots, (x_n, x_n\hat{\beta} + e_n^{*})\}$$

Model ini merupakan resampel residu terpusat dengan $\hat{e} = Y - X\hat{\beta}$ dan mean:

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}\hat{e}_i \cdot$$

Misalkan $\hat{F}_n$ distribusi empiris $\hat{e}$ yang terpusat pada meannya, maka $\hat{F}_n$ berdistribusi dengan massa 1/n untuk setiap $\hat{e} - \hat{\mu}_n$ dan $\int x.d\hat{F}_n dx = 0$

Apabila Y diberikan dan $e_1^*, e_2^*, \ldots, e_n^*$ diasumsikan independen dengan distribusi bersama $\hat{F}_n$, maka: $Y^* = X\hat{\beta} + e^*$, dalam hal ini e^* diperoleh dengan resampling residu terpusat dan Y^* dibangun dari data. Selanjutnya, dengan diperolehnya data (X, Y^*) maka $\hat{\beta}^*$ dapat diperoleh dengan estimasi kuadrat terkecil, yaitu $\hat{\beta}^* = (X^T X)^{-1} X^T Y^*$ apabila dilakukan resample berukuran m, maka akan diperoleh $Y^*_{(m)} = X_{(m)}\hat{\beta} + e^*_{(k)}$ dan $\hat{\beta}^*_{(m)} = (X^T_{(m)} . X_{(m)})^{-1} X^T_{(m)} Y^*_{(m)}$ dengan $Y^*_{(m)} = (y^*_1, y^*_2, \ldots, y^*_m)$, $X_{(m)}$: matriks mxp dan $e^*_{(m)} = (e^*_1, e^*_2, \ldots, e^*_m)$.

Prinsip bootstrap pada model regresi adalah menghampiri distribusi $\sqrt{n}\,(\hat{\beta} - \beta)$ dengan distribusi $\sqrt{n}\,(\hat{\beta}^* - \hat{\beta})$.

Bahan dan Metode

Alat dan Bahan. Bahan yang digunakan dalam penelitian ini meliputi karaya-karya ilmiah berupa jurnal, buku teks, buletin, laporan penelitian maupun karya ilmiah lain yang relevan dengan penelitian yang akan duilakukan dan untuk keperluan pembahasan dilakukan analisis terhadap sekumpulan pasangan data (x_i, y_i). Sedangkan gsebagai alat bantu analisis digunakan software statistika S-PLUS.2000

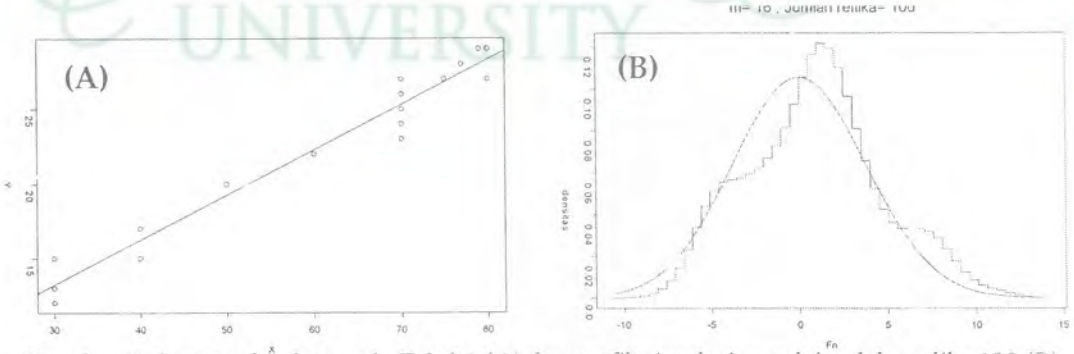
Prosedur Penelitian. Studi literatur dilakukan terhadap sejumlah karya ilmiah berupa jurnal, buku teks, hasil penelitian maupun karya ilmiah lain yang ada kaitannya dengan materi penelitian, dengan mengumpulkan dan mempelajari sejumlah konsep dalam model regresi yang menyangkut penentuan estimator dengan metode kuadrat terkecil; sifat-sifat estimasi kuadrat terkecil yang memuat masalah ketakbiasan, mean dan variansi dari estimator; teorema Gauss-Markov, teorema multivariate Lindberg-Levy, teorema limit sentral univariat dan multivariat, distribusi empiris sampai pada teorema Glivenko-Cantelli. Dalam metode bootsrap dipelajari konsep-konsep dari prinsip dasar pembentukan sampel bootstrap, metode ekspansi deret Taylor dan metode simulasi Monte Carlo; masalah asimtotis bootstrap yang menuntut pemahaman tentang metrik Mallows, ruang Banach dan ruang Hilbert. Sedangkan dalam software SPLUS-2000 dibahas konsep-konsep yang berkaitan dengan perintah-perintah cara memasukkan (input) data, perintah pemrosesan data baik pada model regresi maupun pada bootstrap. Selanjutnya, berbekal sebagian besar dari konsep-konsep tersebut, penelitian dilanjutkan dengan pambahasan lebih khusus pada model regresi linear sederhana dengan metode bootstrap yang dicobakan terhadap sekumpulan pasang data (x_i, y_i). Sedangkan program simulasi untuk melihat hampiran distribusi $\sqrt{n}\,(\hat{\beta} - \beta)$ dibuat dengan menggunakan software S-PLUS 2000.

Berdasarkan sampel random $X_1, X_2, \ldots, X_n$ berdistribusi F. Estimator bootstrap memerlukan sampel bootstrap $X_1^*, X_2^*, \ldots, X_n^*$ yang berdistribusi empiris $\hat{F}_n$. Untuk distribusi $R_n = R_n(X_1, X_2, \ldots, X_n; F)$, estimator bootstrap merupakan distribusi bersyarat $R_n^* = R_n(X_1^*, X_2^*, \ldots, X_n^*; \hat{F}_n)$ berdasarkan $(X_1, X_2, \ldots, X_n)$ dan pada dasarnya distribusi ini diketahui. Secara ringkas proses kerja simulasi yang akan dibuat, seperti langkah-langkah berikut: (a) konstruksi suatu sampel bootstrap iid $(X_1^*, X_2^*, \ldots, X_n^*)$ berdasarkan distribusi $\hat{F}_n$ dengan komputer, (b) berdasarkan $\hat{F}_n$ berarti $\theta(\hat{F}_n)$ diketahui dan $R_n^* = R_n(X_1^*, X_2^*, \ldots, X_n^*; \hat{F}_n)$ dapat ditentukan, (c) bagian (a) dan (b) sebanyak m kali, sehingga diperoleh $R^*_{n,1}, R^*_{n,2}, \ldots, R^*_{n,m}$, (d) dengan mengumpulkan nilai-nilai $R^*_{n,1}, R^*_{n,2}, \ldots R^*_{n,m}$, ditentukan distribusi empiris $\hat{F}_{n,m}(x)$, dan (e) membandingkan distribusi empiris terhadap suatu distribusi normal dengan menggunakan grafik.

Hasil dan Pembahasan

Berdasarkan pembahasan terhadap teori-teori yang terkait didalam penelitian ini dan software statistika S.PLUS 2000 serta berpedoman pada prosedur penelitian diatas maka dihasilkan sebuah program bootstrap yang dapat diaplikasikan untuk menganalisis suatu model regresi (terlampir).

Aplikasi program. Untuk mengaplikasikan



Gambar 1. Scatter plot data pada Tabel 1 (A) dan grafik simulasi untuk jumlah replika 100 (B).

program bootstrap yang telah dihasilkan, dilakukan analisis terhadap sekumpulan pasang data (xi, yi) seperti pada Tabel 1.

Tabel 1. Pasangan data (xi, yi).

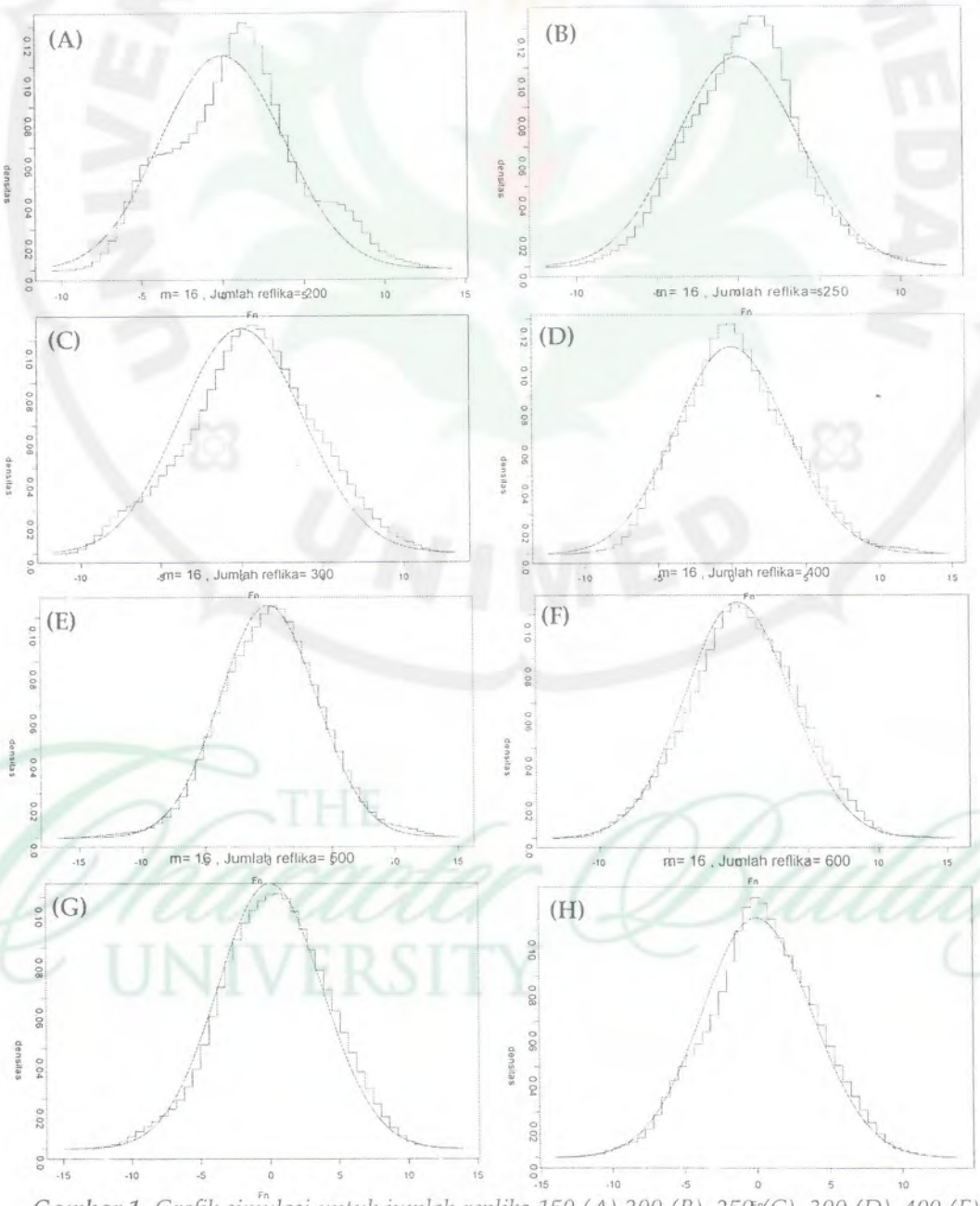
X	Y	X	Y
30	13	70	27
40	17	30	15
50	20	80	27
80	29	70	24
30	12	70	26
40	15	75	27
60	22	70	25
70	25	77	28
70	23	79	29

Scatter plot. Scatter plot dari 17 titik data dari Tabel 1 diperlihatkan pada Gambar 1A.

Berdasarkan scatter plot ini, dapat diambil suatu kesimpulan bahwa pasangan data di atas sudah memadai untuk analisis regresi dan model regresinya adalah $Y=\beta X+\varepsilon$ dengan p=2,

$$Y = \begin{bmatrix} 13 \\ 17 \\ 20 \\ 29 \\ \vdots \\ 29 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & 30 \\ 1 & 40 \\ 1 & 50 \\ 1 & 80 \\ \vdots & \vdots \\ 1 & 79 \end{bmatrix} \quad \text{dan} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

dan estimasi kuadrat terkecil menghasilkan:



Gambar 1. Grafik simulasi untuk jumlah replika 150 (A) 200 (B), 250 (C), 300 (D), 400 (F), 500 (G) dan 600 (H).

$$\hat{\beta}=\begin{bmatrix}4,2013\\0,3012\end{bmatrix}$$

Pembahasan lebih lanjut adalah membandingkan statistik $\sqrt{n}(\hat{\beta}^{*}-\hat{\beta})$ yang merupakan hampiran bagi statistik $\sqrt{n}(\hat{\beta}-\beta)$ dengan suatu distribusi normal untuk beberapa macam replika dengan m=16 melalui grafik simulasi sebagai berikut:

Jumlah replika = 100. Untuk replika = 100 statistik ditunjukkan oleh garis yang tidak mulus (Gambar 1B), terlihat grafiknya menceng ke kanan dan masih relatif jauh dari distribusi normal.

Jumlah replika = 150, 200, 250, 300, 400, 500 dan 600. Grafik simulasi untuk jumlah replika = 150, 200, 250, 300, 400, 500 dan 600 dapat dilihat pada Gambar 2.

Dari Gambar 1(B) dan Gambar 2 dapat dilihat bahwa pada jumlah replika 100 dan 150 distribusi dari statistik masih relatif melenceng jauh dari distribusi normal dan semakin membaik mendekati distribusi normal pada replika 200 dan 250, sedangkan pendekatan distribusi normal yang terbaik dicapai pada replika antara 300 dan 400, tetapi kembali melenceng menjauhi distribusi normal pada reflika 500 dan 600.

Penutup

Prinsip bootstrap pada analisis regresi $Y = X\beta + \varepsilon$ dalam tulisan ini adalah melakukan pendektan distribusi untuk statistik $\sqrt{n}(\hat{\beta}-\beta)$ dengan distribusi dari statistik $\sqrt{n}(\hat{\beta}^{*}-\hat{\beta})$ yang dapat dihitung dari data, dimana $\hat{\beta}$ estimator kuadrat terkecil dan $\hat{\beta}^{*}$ adalah estimator bootstrap, sehingga estimasi yang dilakukan adalah estimasi langsung terhadap statistik $(\hat{\beta}-\beta)$

Menurut perkiraan, semakin besar jumlah replika yang digunakan akan semakin baik hampiran distribusi normal untuk $\sqrt{n}(\hat{\beta}^{*}-\hat{\beta})$, tetapi dalam tulisan ini hampiran yang terbaik terjadi disekitar jumalah reflika 300 dan 400, sedangkan untuk jumlah reflika 500 dan 600 hampirannya mulai menjauh lagi dari distribusi nomal, untuk itu disarankan kepada peneliti lain untuk dapat menganalisis gejala ini lebih lanjut.

Daftar Pustaka

Bain, L.J, Engelhardt, M. (1992) *Introduction to probability and mathematical statistics (2nd ed.).* Belmont, California: Duxbury Press

Efron, B., Tibsirani, R.I. (1993) *An introduction to the bootstrap*. New York: Chapman & Hall

Everitt, B.S. (1994) *A handbook of statistical analyses using S-PLUS*. London: Chapman & Hall

Freedman, D.A. (1984) On bootstraping two-stage least-square estimate in stationary linear models. *Annals Stat.*, 12: 827-842

Searle, S.R. (1971). *Linear models*. New York: Jhon Wiley & Sons

Searle, S.R. (1982) *Matrix algebra useful for statistics*. New York: Jhon Wiley & Sons

Spector, P. (1994) *An introduction to S and S-PLUS*. California: Duxbury Press

StatSci Division MathSoft, Inc. (1995) *S-PLUS Guide to statistical and mathematical analysis Version 3.3.* Seattle, Washington: StatSci Division MathSoft, Inc.

Stine, R.A. (1985) Bootstrap prediction intervals for regression. *J. Am. Stat. Assoc.*, 392: 1026-1037

Thombs, L.R., Scucany, W.R. (1990) Bootstrap prediction interval for autoregresion. *J. Am. Stat. Assoc.*, 410: 486-492

THE Character Building UNIVERSITY

Volume 31 | Nomor 2 | Juli-Desember 2007

Jurnal Sains Indonesia

Media Komunikasi Hasil Penelitian Sains dan Matematika

Gambar sampul depan : *Spektra inframerah SGT dan HAS dari prekursor TEOS*
(lihat halaman 76)

ISSN 1978-3841