

Metode *Robust Regression on Ordered Statistics* (ROS) pada Data Tersensor Kiri dengan *Outlier*

Mashitah, Arief Wibowo, dan Diah Indriani

Departemen Biostatistika dan Kependudukan FKM UNAIR

Fakultas Kesehatan Masyarakat Universitas Airlangga

Jl. Mulyorejo Kampus C Unair Surabaya 60115

Alamat Korespondensi:

Mashitah

itah_emoet@yahoo.com

ABSTRACT

Robust Regression is the regression method that used if distribution from error abnormal and or some outlier that affected to model. This method is important tool to analyze data that affected by outlier with the result model that robust or resistant to outlier. Resistant estimate relatively not influence by large change on little part of data or little change on large part of data. Some estimation method in robust regression are M-estimation, Least Trimmed Square (LTS), MM estimation, S-estimation and Least Mean Square (LMS). This research use M-Estimation and Least Trimmed Square (LTS) estimation method. Best method determine with compare value of determination coefficient and value of Sum of Square Error (SSE) at approach that used. With founding best method to this robust regression could be predict patient age of onset to repetitive strain injury. Commonly, RSI patient don't know when the onset of RSI. Variables that used to find model are patient age of onset when diagnosed RSI and work duration. Base on analysis results, was found that the best model to predict patient age of onset to Repetitive Strain Injury is: $\hat{Y} = -8.0283 + 1.2751 X_1$. This best model found with Least Trimmed Square (LTS) approach.

Keywords: age of onset, work duration, RSI, robust regression

ABSTRAK

Regresi Robust adalah metode regresi yang digunakan jika distribusi error abnormal atau beberapa outlier berpengaruh terhadap model. Metode ini adalah alat penting untuk menganalisa data yang dipengaruhi oleh outlier dengan hasil model yang resisten terhadap outlier. Resistensi artinya tidak dipengaruhi oleh perubahan besar pada sebagian kecil data atau perubahan sedikit pada sebagian besar data. Beberapa metode estimasi dalam regresi adalah M-estimation, Least Trimmed Square (LTS), estimasi MM, S-estimasi dan Least Mean Square (LMS). Penelitian ini menggunakan M-Estimasi dan Least Trimmed Square (LTS) Metode terbaik ditentukan dengan membandingkan nilai koefisien determinasi dan nilai Sum of Square Error (SSE) pada pendekatan yang digunakan. Metode terbaik untuk regresi ini dapat memprediksi usia pasien RSI. Umumnya, pasien RSI tidak tahu kapan terjadinya RSI. Variabel yang digunakan untuk menemukan model adalah usia pasien ketika didiagnosis RSI dan lama kerja. Berdasarkan hasil analisis, ditemukan bahwa model terbaik untuk memprediksi umur pasien untuk Repetitive Strain Injury adalah: $\hat{Y} = -8,0283 + 1,2751 X_1$. Model terbaik ditemukan dengan Least Trimmed Square (LTS).

Kata kunci: usia onset, lama kerja, RSI, regresi robust

PENDAHULUAN

Data tersensor adalah data yang diperoleh sebelum hasil yang diinginkan dari pengamatan terjadi, sedangkan waktu pengamatan telah berakhir atau oleh sebab lain. Data yang mengalami penyensoran hanya memuat sebagian informasi mengenai variabel random yang diperhatikan, namun berpengaruh terhadap pengertian-pengertian dan perhitungan statistik.

Menurut Miiler (1998) data dikatakan tersensor jika pengamatan waktu survival hanya sebagian, tidak sampai *failure event*. Penyebab terjadinya data tersensor antara lain: *Loss to follow up* terjadi bila objek pindah, meninggal atau menolak untuk berpartisipasi. *Drop Out* terjadi bila perlakuan dihentikan karena alasan tertentu. *Termination of study* terjadi bila masa penelitian berakhir sementara objek yang diobservasi belum mencapai *failure event*.

Terdapat beberapa jenis penyensoran. Jenis penyensoran berdasarkan saat mulai pengamatan dibagi menjadi dua, yaitu penyensoran tunggal dan penyensoran progresif. Jenis penyensoran berdasarkan saat mulai dan mengakhiri pengamatan terbagi atas tiga jenis penyensoran, yaitu penyensoran tipe I, penyensoran tipe II dan penyensoran tipe III. Kemudian jenis penyensoran yang lain ada penyensoran ke kanan, penyensoran ke kiri dan penyensoran interval. (Kleinbaum. 2005).

Penyensoran ke kiri dilakukan bila nilai pengamatan yang pasti tidak diketahui, hanya diketahui lebih kecil daripada nilai tertentu. Analisis data tersensor kiri biasanya di gunakan pada penelitian di ilmu lingkungan dan industri, data dari penelitian mungkin tidak lengkap karena keterbatasan alat, metodologi untuk pengukuran dan ketidakmampuan untuk mengamati data. Beberapa metode statistik yang dapat digunakan untuk menganalisis data tersensor kiri yaitu *Kaplan-Meier estimator*, *Maximum Likelihood estimator* dan *metode Robust Regression on Ordered Statistics (ROS)*. Tiga metode ini masing-masing merupakan metode non parametrik, parametrik dan semi parametrik.

Contoh data tersensor kiri adalah data pada penderita *Repetitive Strain Injury*. Biasanya pasien yang didiagnosa mengalami *Repetitive Strain Injury* tidak mengetahui secara pasti kapan mulai mengalaminya. Ketidaktahuan pasien secara pasti mengenai kapan pertama kali mereka merasakan keluhan nyeri itulah yang termasuk dalam data tersensor ke kiri.

Overuse Injury/Repetitive Strain Injury (RSI) adalah kerusakan otot, tendon, tulang atau sistem saraf yang dicetuskan atau di esterbasi oleh pengerahan tenaga yang kuat secara berulang, postur canggung berkelanjutan untuk waktu yang lama, tekanan kontak permukaan, getaran atau dingin. Faktor risiko yang memiliki peranan terhadap terjadinya gangguan muskuloskeletal akibat pekerjaan yaitu beratnya, durasi, dan frekuensi pekerjaan. (Maffulli, 2005)

Pasien dengan RSI pada Anggota Gerak Atas (AGA) biasanya muncul dengan keluhan nyeri, biasanya pada leher, bahu, lengan atau tangan; kelelahan, baik keseluruhan ataupun terlokalisir; kelemahan; gangguan rasa (parestesi); kehilangan ketangkasan; depresi dan gangguan tidur. Pada

sebagian besar pasien, hal tersebut berhubungan dengan durasi dan intensitas pekerjaan mereka. Gejala dapat berkembang dalam hitungan minggu, bulan atau bertahun-tahun, dan pasien umumnya tidak dapat menentukan kapan pertama kali gejala ini mulai muncul. Pada beberapa pasien awalnya muncul seperti kelelahan sederhana. Perbedaan antara kelelahan sederhana dan yang berulang adalah terkait dengan durasi waktu dan intensitas dari gejala.

Pada penelitian ini, akan digunakan metode *Robust Regression on Ordered Statistics (ROS)* dalam menganalisis data tersensor kiri, data yang digunakan adalah catatan rekam medis pasien penderita *Repetitive Strain Injury* yang menjalani terapi di Instalasi Rehabilitasi Medik RSUD Dr. Soetomo Surabaya.

Dari beberapa penelitian yang dilakukan untuk menganalisis data tersensor kiri, ada beberapa metode yang digunakan. Metode yang umumnya digunakan adalah Kaplan Meier, Maksimum Likelihood dan *Robust Regression*. Pada penelitian yang dilakukan oleh Samuel (2009), masing-masing metode memiliki kekuatan dan kelemahan pada tiap ukuran sampel dan tingkat penyensoran. Dari penelitian yang dilakukan dengan menggunakan data simulasi, didapatkan kesimpulan bahwa metode ROS memiliki kekuatan yang lebih baik dibanding kedua metode lainnya dalam menganalisis data tersensor kiri pada kombinasi yang berbeda pada tingkat sensor dan ukuran sampel.

Oleh karena itu akan dilakukan penerapan metode *Robust Regression on Ordered Statistics* untuk data tersensor kiri pada data rekam medis pasien *Repetitive Strain Injury*.

Bagaimanakah kekuatan uji metode *Robust Regression on Ordered Statistics (ROS)* dalam menganalisis data tersensor kiri pada pasien *Repetitive Strain Injury*?

METODE PENELITIAN

Jenis penelitian adalah analitik dengan menggunakan data sekunder. Variabel penelitian ini adalah survival time (interval waktu antara prediksi waktu pertama kali pasien merasakan nyeri sakit sampai pasien didiagnosis mengalami RSI), lama kerja dan waktu pasien merasakan nyeri pertama kali. Waktu pasien merasakan nyeri

pertama kali merupakan *event* untuk menentukan apakah pasien termasuk dalam data tersensor kiri atau tidak. *Event* pada data tersensor kiri adalah kejadian yang terjadi sebelum dimulainya penelitian dan tidak diketahui kapan event itu terjadi.

Data yang digunakan adalah data sekunder rekam medis pasien *Repetitive Strain Injury* pada periode Maret 2010–Februari 2011 di Instalasi Rehabilitasi Medik RSUD Dr. Soetomo Surabaya. Data pasien yang diambil adalah pasien dengan *diagnosis Repetitive Strain Injury* Anggota Gerak Atas (AGA) dan yang bekerja untuk mendapatkan penghasilan.

Metode *Robust Regression on Ordered Statistics* (ROS) digunakan dalam analisis data tersensor kiri pada pasien *Repetitive Strain Injury*. Metode ini digunakan karena pada data penelitian terjadi error tidak berdistribusi normal, atau terdapat pencilan (*outlier*), atau terdapat titik data yang memengaruhi hasil regresi. Langkah-langkah analisisnya adalah.

1. Menentukan model regresi awal dengan menggunakan metode OLS. Kemudian melakukan uji asumsi klasik untuk model regresi yang ada.
2. Setelah ditemukan data yang *outlier* maka dilakukan beberapa metode estimasi pada regresi robust guna mendapatkan model terbaik. Metode *Robust Regression On Ordered Statistics* (ROS) memiliki beberapa metode estimasi diantaranya *M-Estimation*, *Least Trimmed Square* (LTS), *MM estimation*, *S estimation*, *Least Mean Square* (LMS). Pada penelitian ini yang akan digunakan adalah metode *Least Trimmed Square* (LTS) dan *M-Estimation*, di mana LTS dan *M-Estimation* merupakan metode regresi robust yang sering digunakan.
3. Menaksir parameter regresi dan didapatkan nilai koefisien determinasi dan *Sum of Square Error* (SSE).
4. Mendapatkan model terbaik dengan membandingkan nilai koefisien determinasi dan SSE dari beberapa metode yang digunakan. Model terbaik didapatkan dari model yang memiliki nilai koefisien determinasi tertinggi dan nilai SSE terkecil, di mana model tersebut dianggap sebagai model yang paling baik dalam melakukan prediksi

usia pertama pasien mengalami *Repetitive Strain Injury*.

HASIL PENELITIAN

Jumlah pasien yang didiagnosis mengalami *Repetitive Strain Injury* di Instalasi Rehabilitasi Medik RSUD Dr. Soetomo Surabaya periode Maret 2010–Februari 2011 sebanyak 58 pasien. Dari 58 pasien ini, sebanyak 55 pasien termasuk dalam data tersensor kiri, sedangkan 3 pasien tidak termasuk dalam data tersensor kiri. Pasien tidak termasuk dalam data tersensor kiri karena mengetahui kapan pertama kali merasakan nyeri (*event*).

Nilai minimum usia pasien pertama kali didiagnosis mengalami *Repetitive Strain Injury* adalah pada usia 23 tahun dan usia maksimum adalah 61 tahun. Nilai rata-rata usia pasien 43 tahun dengan standar deviasi 10,021. Sedangkan pada variabel lama kerja nilai minimum sebesar 3 tahun dan nilai maksimum sebesar 37 tahun. Rata-rata lam kerja 21,53 tahun dengan standar deviasi 8,070.

Model dengan Estimasi OLS

Persamaan regresi linear untuk mencari hubungan dan pengaruh variabel bebas terhadap variabel terikat. Model regresi dengan estimasi OLS adalah: $\hat{Y} = -5,48802 + 1,18359 X_1$ di mana: Y = Survival time, X_1 = lama kerja. Model regresi tersebut mempunyai nilai koefisien determinasi $R^2 = 0,9086$, $F = 556.54$ dan peluang 0,000. Sehingga disimpulkan bahwa ada pengaruh yang signifikan antara lama kerja pasien bekerja di tempat kerjanya sampai didiagnosa menderita *Repetitive Strain Injury* terhadap survival time pasien. Uji secara parsial juga menunjukkan hasil peluang variabel lama kerja sebesar $< 0,0001$ yang menunjukkan bahwa lama kerja secara parsial mempengaruhi survival time.

Asumsi klasik model regresi terdiri dari linieritas, heteroskedastisitas, autokorelasi, normalitas eror dan independensi eror. Pemeriksaan yang digunakan untuk mendeteksi adanya heteroskedastisitas yaitu dengan melihat plot antara residual dengan prediksinya. Jika plot antara residual dengan prediksi tidak ada yang keluar dari batas -2 sampai 2, maka dapat dikatakan asumsi homogenitas terpenuhi

atau dengan kata lain tidak terjadi gejala heteroskedastisitas. Dari Gambar 1a dapat dilihat bahwa plot antara residual dengan prediksi hanya terdapat satu titik yang keluar dari batas -2 hingga 2. Oleh karena itu, maka asumsi homogenitas masih terpenuhi sehingga dapat disimpulkan bahwa tidak terjadi gejala heteroskedastisitas dalam penelitian ini.

Pemeriksaan yang digunakan untuk mendeteksi adanya autokorelasi yaitu dengan melihat plot antara residual dengan *time order*. Jika plot antara residual dengan *time order* tidak ada yang keluar dari batas -2 sampai 2, maka dapat dikatakan asumsi autokorelasi terpenuhi atau dengan kata lain tidak terjadi gejala autokorelasi. Dari Gambar 1b dapat dilihat bahwa plot antara residual dengan *time order* hanya terdapat satu titik yang keluar dari batas -2 hingga 2. Oleh karena itu, maka asumsi autokorelasi masih terpenuhi sehingga dapat disimpulkan bahwa tidak terjadi gejala autokorelasi dalam penelitian ini. Uji yang akan digunakan untuk mendekteksi autokorelasi yaitu dengan menggunakan uji *Durbin-Watson*. Nilai uji statistik *Durbin-Watson* berkisar antara 0 dan 4. Apabila nilai uji statistik *Durbin-Watson* lebih kecil dari 1 atau lebih besar dari 3, maka *residuals* atau eror dari model regresi berganda

tidak bersifat independen atau dengan kata lain terjadi autokorelasi. Dari output SAS terlihat bahwa nilai *Durbin-Watson* sebesar 1.731 yang berarti lebih besar dari 1 dan lebih kecil dari 3. Hal ini menandakan bahwa tidak terjadi autokorelasi dalam penelitian ini.

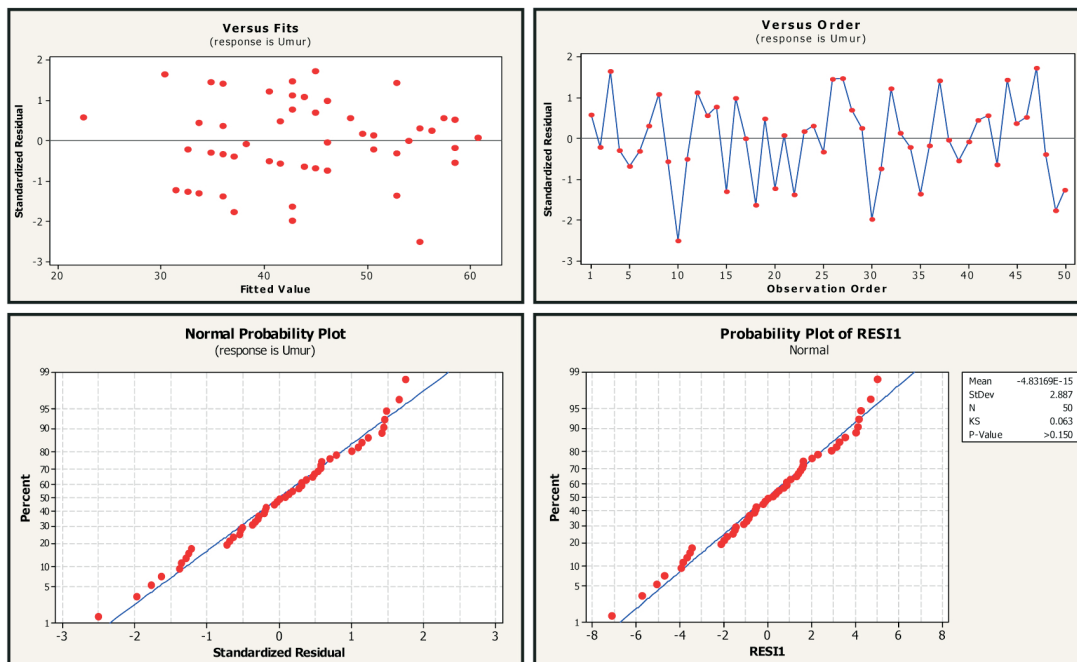
Uji normalitas dilakukan untuk melihat apakah sebuah data berdistribusi normal atau tidak.Uji yang digunakan untuk melihat normalitas yaitu dengan melihat P-P Plot di mana jika sudutnya mendekati 45°, maka dikatakan normal. Dari Gambar 1c dan 1d dapat dilihat bahwa P-P plot berada di dekat garis 45° atau dengan kata lain mendekati sudut 45°. Oleh karena plot mendekati sudut 45°, maka dapat dikatakan bahwa asumsi normalitas telah dipenuhi.

Deteksi Outlier

Salah satu alat untuk mendeteksi adanya outlier, yaitu melalui perbandingan nilai DFFITS, dan dapat dikatakan pengamatan tersebut sebagai *outlier* dikarenakan memiliki nilai

$$|DFFITS| > 2\sqrt{(k+1)/n}$$

Nilai mutlak DFFOTS pada penelitian ini melebihi batas



Gambar 1. Grafik Uji Asumsi Klasik Regresi Metode OLS

$$|DFFITS| > 2\sqrt{(k+1)/n} = 2\sqrt{\frac{1+1}{58}} = 0,37$$

sebesar 0,4609 dan 0,5755 pada pengamatan ke-3 dan ke-10, maka dapat dikatakan ada 2 data yang merupakan *outlier*. Sehingga ada indikasi adanya outlier dan perlu dilakukan dengan pendekatan regresi robust.

Model dengan Metode ROS Estimasi LTS

LTS merupakan suatu metode pendugaan parameter regresi *robust* untuk meminimumkan jumlah kuadrat h residual (fungsi objektif). Model regresi dengan metode LTS adalah $\hat{Y} = -8,0283 + 1,2751 X_1$ di mana: Y = Survival time, X_1 = lama kerja. Model regresi tersebut mempunyai nilai koefisien determinasi $R^2=0,9321$ dan peluang 0,0001 maka dapat disimpulkan bahwa ada pengaruh yang signifikan antara lama kerja pasien bekerja di tempat kerjanya sampai didiagnosis menderita *Repetitive Strain Injury* terhadap survival time pasien.

Uji asumsi klasik pada model estimasi regresi *robust* LTS (*Least Trimmed Square*) diperoleh suatu kesimpulan bahwa asumsi kenormalan, non heteroskedasitas dan non autokorelasi dipenuhi.

Model dengan Metode ROS Estimasi M-Estimation-Tukey Bisquare

Pada kasus adanya pencilan, dapat digunakan estimasi regresi linear dengan menggunakan IRLS dengan fungsi pembobot Tukey Bisquare. Model regresi dengan metode estimasi M-Estimation-Tukey Bisquare adalah $\hat{Y} = -5,7307 + 1,1948 X_1$ di mana: Y = Survival time, X_1 = lama kerja.

Persamaan regresi linear dengan estimasi melalui metode M-Estimator dengan fungsi pembobot Tukey Bisquare mempunyai nilai koefisien determinasi $R^2 = 0,8071$ dan nilai peluang 0,0001. Sehingga dapat disimpulkan bahwa ada pengaruh yang signifikan antara lama kerja pasien bekerja di tempat kerjanya sampai didiagnosis menderita *Repetitive Strain Injury* terhadap survival time pasien.

Uji asumsi klasik pada model estimasi M-Estimation dengan fungsi pembobot Tukey Bisquare diperoleh suatu kesimpulan bahwa

asumsi kenormalan, non heteroskedasitas dan non autokorelasi dipenuhi.

Model dengan ROS Estimasi M-Estimation-M Huber

Pada kasus yang sama, dapat digunakan IRLS dengan fungsi pembobot yang berbeda dengan sebelumnya yaitu dengan fungsi pembobot Huber. Model regresi yang diperoleh adalah $\hat{Y} = -5,7453 + 1,1959 X_1$ di mana: Y = Survival time, X_1 = lama kerja. Model tersebut mempunyai nilai koefisien determinasi $R^2 = 0,8429$ dengan peluang 0,0001. Sehingga dapat disimpulkan bahwa ada pengaruh yang signifikan antara lama kerja pasien bekerja di tempat kerjanya sampai didiagnosis menderita *Repetitive Strain Injury* terhadap survival time pasien.

Uji asumsi klasik pada model estimasi M-Estimation dengan fungsi pembobot Huber diperoleh suatu kesimpulan bahwa asumsi kenormalan, non heteroskedasitas dan non autokorelasi dipenuhi.

Kekuatan Metode Robust Regression on Ordered Statistics (ROS) pada Data Tersensor Kiri

Kekuatan metode ROS ini dapat kita lihat dengan membandingkan nilai *Sum Of Square Error* (SSE) dari ketiga estimator ROS yang dipergunakan pada penelitian ini. Secara lengkap disajikan dalam Tabel 1.

Berdasarkan Tabel 1, maka estimator dengan kekuatan terbaik dalam penerapan metode ROS pada data tersensor kiri adalah dengan pendekatan LTS. Di mana nilai SSE dari pendekatan LTS terkecil dengan nilai 1,2028 dan koefisien determinasi terbesar dengan nilai 0,9321 yang artinya sebesar 93,21% variasi lama

Tabel 1. Perbandingan Nilai SSE

Metode	SSE	Koef. Determinasi
LTS	1.2028	0,9321
M-Estimation-Tukey Bisquare	1.2644	0,8071
M-Estimation-M Huber	1.3225	0,8429

kerja dapat menjelaskan variasi yang terjadi pada survival time.

Prediksi Usia Pertama Pasien Mengalami Repetitive Strain Injury

Prediksi terhadap usia pertama pasien mengalami Repetitive Strain Injury, dapat dilakukan dengan memasukkan data usia pertama pasien didiagnosis mengalami *Repetitive Strain Injury*, survival time duga dan data lama bekerja ke dalam model terbaik yang telah didapatkan sebelumnya. Model terbaik yang dipakai yaitu model dengan pendekatan LTS. Model tersebut yaitu: $\hat{Y} = -8,0283 + 1,2751 X_1$ di mana: Y = Survival time, X_1 = lama kerja.

Misalkan pasien dengan usia pertama didiagnosis yaitu 32 tahun, survival time duga 9 tahun dan lama bekerja 12 tahun. Berdasarkan data yang ada maka kita masukkan ke dalam model terbaik di atas, yaitu:

$$\begin{aligned}\hat{Y} &= -8,0283 + 1,2751 X_1 \\ &= (-8,0283) + 1,2751 (12) \\ &= (-8,0283) + 15,3012 \\ &= 7,2729\end{aligned}$$

Jadi pada pasien dengan usia pertama didiagnosis yaitu 32 tahun dan lama bekerja 12 tahun, survival time dari pasien tersebut adalah 7,2729. Untuk mengetahui prediksi umur pertama pasien ini mengalami *Repetitive Strain Injury* dapat dihitung sebagai berikut:

$$\begin{aligned}\text{Umur pertama sakit} &= \text{umur pertama diagnosis} \\ &\quad \text{survival time} \\ &= 32 - 7,2729 \\ &= 24,7271\end{aligned}$$

Berdasarkan hasil perhitungan, dapat diprediksi usia pertama pasien mengalami *Repetitive Strain Injury* adalah pada usia 25 tahun.

PEMBAHASAN

Data dikatakan tersensor jika pengamatan waktu survival hanya sebagian, tidak sampai *failure event*. Terdapat beberapa jenis penyensoran, salah satunya adalah penyensoran ke kiri. Penyensoran ke kiri dilakukan bila nilai pengamatan yang pasti tidak diketahui, hanya diketahui lebih kecil daripada nilai tertentu.

Analisis data tersensor kiri biasanya di gunakan pada penelitian di ilmu lingkungan dan industri, di mana data dari penelitian mungkin tidak lengkap karena keterbatasan alat, metodologi untuk pengukuran dan ketidakmampuan untuk mengamati data. Pada data tersensor kiri biasanya ditemukan pencilan (*Outlier*).

Metode pendugaan parameter yang paling sering dipergunakan di dalam analisis regresi adalah metode kuadrat terkecil (*least squares*), metode ini mempunyai kelemahan jika diterapkan pada data yang mengandung pengamatan berpengaruh (*inflential observation*), persamaan regresi yang dihasilkan oleh metode kuadrat terkecil cenderung mudah berubah-ubah dengan adanya pengamatan berpengaruh.

Untuk mengatasi kelemahan metode kuadrat terkecil ini dapat dilakukan dengan dua cara yaitu:

1. Mengeluarkan titik yang berpengaruh yang dapat dideteksi dengan *dffit*, *cook distance*, *dfbetas*, setelah itu tetap menggunakan metode kuadrat terkecil
2. Tetap menggunakan seluruh data, tetapi dengan memberikan bobot yang kecil untuk pengamatan yang berpengaruh, metode ini dikenal dengan nama metode regresi *robust*.

Regresi *robust* merupakan alat yang penting untuk menganalisis data yang terkontaminasi oleh pencilan. Regresi *robust* digunakan untuk mendeteksi pencilan dan memberikan hasil yang resisten terhadap adanya pencilan (Chen, 2002).

Kriteria yang dapat digunakan untuk menentukan model regresi terbaik, yaitu R^2 dan standar error s . Model terbaik akan mempunyai R^2 *adjusted* terbesar dan s terkecil. Seperti penelitian yang dilakukan oleh Cahyawati (2009) di mana hasil penelitian menunjukkan bahwa untuk berbagai ukuran sampel yang diamati, pendugaan parameter MRR-W menghasilkan model yang lebih baik dari MKT. Hal ini dilihat berdasarkan nilai R^2 *adjusted* hasil MRR-W yang lebih besar dan nilai *root mean square error* (RMSE) selalu lebih kecil dari model hasil pendugaan MKT.

Pada penelitian yang telah dilakukan oleh Hanum (2011) pada jurnalnya yang berjudul Perbandingan Metode *Stepwise*, *Best Subset Regression*, dan Fraksi dalam Pemilihan Model Regresi Berganda Terbaik, pemilihan model

terbaik kriteria didasarkan pada penambahan R^2 dan R^2 -adj, pengurangan S^2 , atau kedekatan nilai C-p Mallow dengan jumlah peubah dalam model. Berdasarkan kriteria tersebut didapatkan hasil bahwa Pemilihan model regresi terbaik dengan metode Stepwise dan *Best Subset Regression* tidak mempertimbangkan masalah multikolinear. Akibatnya model terbaik pilihan kedua metode tersebut masih mengandung masalah multikolinear. Sementara Metode Fraksi lebih fokus untuk menghindari masalah multikolinear, sehingga menghasilkan model terbaik tanpa ada masalah multikolinear. Tetapi keharusan untuk memuat fraksi dari jumlah peubah bebas dalam data menjadi penghambat untuk lebih bebas menentukan model terbaik.

Fathurahman (2009) melakukan penelitian dengan judul Pemilihan Model Regresi Terbaik Menggunakan Metode *Akaike's Information Criterion* dan *Schwarz Information Criterion*. Penelitian ini bertujuan mengkaji pemilihan model regresi terbaik menggunakan metode AIC dan SIC pada kasus faktor-faktor yang mempengaruhi nilai ujian nasional (UNAS) siswa Sekolah Menengah Kejuruan Negeri (SMKN) 1 Samarinda Menurut metode AIC dan SIC, model regresi terbaik adalah model regresi yang mempunyai nilai AIC dan SIC terkecil.

Berdasarkan hasil analisis didapatkan model terbaik untuk memprediksi usia pertama pasien mengalami Repetitive Strain Injury yaitu model dengan pendekatan LTS. Di mana LTS merupakan suatu metode pendugaan parameter regresi *robust* untuk meminimumkan jumlah kuadrat h residual (fungsi objektif).

Metode *Least Trimmed Squares* sebagai salah satu metode penaksiran parameter model regresi yang *robust* terhadap kehadiran nilai pencilan. Adapun tujuan yang ingin dicapai adalah mendapatkan nilai parameter model regresi yang *robust* terhadap kehadiran nilai pencilan.

Analisis regresi *robust* telah digunakan selama ratusan tahun, tapi tidak dengan serius sampai akhir-akhir ini. Metode ini merupakan metode alternatif yang sesuai untuk data yang terkontaminasi nilai pencilan, bahkan bisa menyaingi prosedur biasa yang asumsi standarnya terpenuhi.

Metode ini dikembangkan oleh Rousseeuw dan Leroy (1987).

'Ketika menggunakan alat-alat analisis, biasanya langkah pertama adalah mencoba menghapus pencilan kemudian mencocokkan data yang sudah bagus dengan menggunakan metode kuadrat terkecil, tetapi analisis *robust* mencocokkan model regresi dengan sebagian besar data dan kemudian mengatasi titik-titik pencilan yang memiliki nilai residu yang besar sebagai solusi *robust* tersebut.' (Rousseeuw dan Leroy, 1987).

Jadi metode ini tidak membuang bagian dari data melainkan menemukan model fit dari mayoritas data.

Pada penelitian yang dilakukan oleh Soemartini (2007), didapatkan kesimpulan untuk menanggulangi pencilan pada data, yaitu dengan mengeluarkan atau membuang observasi ke- i pada data yang diduga merupakan pencilan. Kemudian dilakukan pengujian kembali untuk mendeteksi terdapat atau tidaknya pencilan pada data sampai tidak terdapat lagi pencilan pada data tersebut.

Meskipun pencilan identik dengan data yang tidak bagus, akan tetapi ia merupakan bagian terpenting dari data, karena menyimpan informasi tertentu. Untuk itu, alternatif yang dapat diambil terhadap data yang terkontaminasi pencilan adalah dengan menggunakan metode *Least Trimmed Square* dalam penaksiran model regresi.

Penelitian yang dilakukan oleh Ardiyanti (2011), mengambil simulasi pada suatu kasus dengan menggunakan data dari BPS (Badan Pusat Statistik) Provinsi Jawa Tengah yaitu data tentang produksi padi tiap kabupaten atau kota di Provinsi Jawa Tengah tahun 2007, di mana variabel-variabel tersebut meliputi jumlah produksi padi, luas panen, dan jumlah penduduk. Proses analisis dimulai dengan menggunakan metode kudrat terkecil, identifikasi *outlier*, dan analisis dengan dua metode *robust*.

Dalam menilai hasil kedua metode dengan membandingkan standar error kedua metode dengan OLS yang terdapat *outlier*. Apabila standar error yang dihasilkan metode regresi *robust* lebih kecil dari OLS, maka regresi *robust* dapat menganalisis data tanpa membuang *outlier* dan menghasilkan estimasi yang resisten terhadap

outlier. Sehingga dapat dikatakan regresi robust dapat mengatasi kelemahan OLS terhadap pengaruh *outlier*.

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa baik Estimasi-M maupun Estimasi-MM mempunyai keefektifan yang sama dalam mengatasi *outlier* pada OLS, karena keduanya dapat mengecilkan standar error yang dihasilkan OLS. Dilihat dari efek *breakdown point*, Estimasi-M kurang efektif daripada Estimasi-MM dalam mengatasi pengaruh *outlier* pada variabel prediktor.

Persamaan regresi linear dengan estimasi melalui metode LTS pada model terbaik yang telah dipilih, menunjukkan bahwa nilai konstanta adalah negatif sebesar 8,0283. Nilai tersebut mempunyai arti bahwa survival time pasien memiliki rata-rata sebesar 8,0283 dengan menganggap variabel lama kerja konstan.

Sedangkan nilai regresi pada variabel lama kerja memiliki koefisien regresi yang positif sebesar 1,2751 yang berarti apabila lama kerja pasien bekerja di tempat kerjanya sampai didiagnosis menderita *Repetitive Strain Injury* meningkat maka akan menimbulkan peningkatan survival time pasien sebesar 1,2751 tahun.

Setelah menetapkan model kemudian melakukan uji signifikansi model yang tujuannya untuk mengetahui apakah model yang didapatkan secara umum telah sesuai atau tidak. Diketahui $P\text{ Value} = 0,0001$ lebih kecil daripada $\alpha = 0,05$ maka dapat disimpulkan bahwa H_0 ditolak, yang berarti ada pengaruh yang signifikan antara lama kerja pasien bekerja di tempat kerjanya sampai didiagnosis menderita *Repetitive Strain Injury* terhadap survival time pasien.

Model di atas dipilih berdasarkan nilai koefisien determinasi tertinggi dan nilai SSE terkecil setelah membandingkan nilai koefisien determinasi dan SSE antara metode OLS, LTS, M-estimator – Tukey Bisquare dan M-estimator – M-Huber. Di mana nilai koefisien determinasi tertinggi yaitu pada metode LTS, dengan nilai koefisien determinasi sebesar 0,9321. Hasil ini menunjukkan bahwa sumbangan variabel lama kerja dalam menjelaskan variabel survival time sebesar 93,21 persen dan sebesar 6,79 persen dijelaskan oleh variabel lain yang tidak masuk dalam model.

Kekuatan uji juga dilakukan pada metode dari regresi robust yang digunakan. Di mana nilai *Sum of Square Error (SSE)* terkecil juga terdapat pada metode LTS, dengan nilai SSE sebesar 1.2028. Sedangkan untuk metode M-Estimation-Tukey Bisquare dan M-Estimation-M-Huber memiliki nilai SSE masing-masing sebesar 1.2644 dan 1.3225. Hal ini menunjukkan bahwa dari model maupun kekuatan uji yang terbaik yaitu pada metode LTS jika dibandingkan dengan metode M-estimator-Tukey Bisquare dan M-Estimator-M-Huber pada data pasien *Repetitive Strain Injury*.

Overuse Injury/Repetitive Strain Injury (RSI) adalah kerusakan otot, tendon, tulang atau sistem saraf yang dicetuskan atau di esterbasi oleh pengerahan tenaga yang kuat secara berulang, postur canggung berkelanjutan untuk waktu yang lama, tekanan kontak permukaan, getaran atau dingin. Faktor risiko yang memiliki peranan terhadap terjadinya gangguan muskuloskeletal akibat pekerjaan yaitu beratnya, durasi, dan frekuensi pekerjaan. (Maffulli, 2005)

Patofisiologi dasar dari *Repetitive Strain Injury* adalah *stress* jaringan yang melebihi waktu adaptasi jaringan. *Stress* yang terjadi dapat berupa *shear*, *tension*, kompresi, *impingement*, vibrasi dan kontraksi. Kelelahan jaringan akan mengakibatkan perubahan pada jaringan tendon, ligamen, saraf, dan jaringan lunak lainnya. Jaringan yang kelelahan/*fatigue* akan mengalami proses penyembuhan dan adaptasi jika mendapatkan waktu yang cukup, namun pada *repetitive strain injury*, cedera yang terjadi melebihi kemampuan jaringan untuk melakukan proses penyembuhan dan adaptasi. (Sharma, 2005)

Gejala dapat berkembang dalam hitungan minggu, bulan atau bertahun-tahun, dan pasien umumnya tidak dapat menentukan kapan pertama kali gejala ini mulai muncul. Seperti pada contoh pasien nomor 2 dan 3 yang mengalami *Repetitive Strain Injury*. Pasien nomor 2 dengan usia pertama didiagnosis yaitu 32 tahun, survival time duga 9 tahun dan lama bekerja 3 tahun, umur pertama pasien ini mengalami *Repetitive Strain Injury* diprediksi sekitar 25 tahun. Sedangkan pada pasien nomor 3 dengan usia pertama didiagnosis yaitu 35 tahun, survival time duga 12

tahun dan lama bekerja 10 tahun, umur pertama pasien ini mengalami *Repetitive Strain Injury* diprediksi sekitar 30 tahun.

Dengan didapatkannya model prediksi diharapkan dapat digunakan oleh tenaga medis khususnya dokter untuk melakukan prediksi usia pertama pasien mengalami *Repetitive Strain Injury*, dan memberikan penanganan yang lebih tepat kepada pasien. Kemudian dapat digunakan pula sebagai salah satu tindakan preventif bagi perusahaan maupun pekerja guna mengurangi risiko para pekerja mengalami *Repetitive Strain Injury*.

SIMPULAN DAN SARAN

Kesimpulan

Jumlah pasien yang didiagnosis mengalami *Repetitive Strain Injury* selama periode Maret 2010–Februari 2011 ditemukan sebanyak 58 pasien, dari 58 pasien ini, sebanyak 55 pasien termasuk dalam data tersensor kiri, sedangkan 3 pasien tidak termasuk dalam data tersensor kiri. Pasien tidak termasuk dalam data tersensor kiri karena mengetahui kapan pertama kali merasakan nyeri (event). Berdasarkan data yang ada nilai minimum usia pasien pertama kali didiagnosis mengalami *Repetitive Strain Injury* adalah pada usia 23 tahun dan usia maksimum adalah 61 tahun. Dengan nilai rata-rata 43, SE Mean sebesar 1,316 dan SD sebesar 10,021. Sedangkan pada variable lama kerja nilai minimum sebesar 3 tahun dan nilai maksimum sebesar 37 tahun. Mean 21,53, SE Mean sebesar 1,060 dan SD sebesar 8,070. Berdasarkan hasil analisis yang telah dilakukan, didapatkan bahwa nilai *Sum of Square Error* (SSE) terkecil terdapat pada pendekatan *Least Trimmed Square* (LTS) dengan nilai SSE sebesar 1.2028. Hasil tersebut menunjukkan bahwa kekuatan uji *Robust Regression On Ordered Statistics* (ROS) dengan pendekatan LTS merupakan yang memiliki kekuatan uji terbaik jika dibandingkan dengan metode M-estimator-Tukey Bisquare dan M-estimator-M-Huber pada data pasien *Repetitive Strain Injury*. Berdasarkan hasil analisis yang telah dilakukan telah didapatkan model terbaik untuk memprediksi usia pertama pasien

mengalami *Repetitive Strain Injury*, yaitu: $\hat{Y} = -8,0283 + 1,2751 X_1$. Model terbaik ini didapatkan dengan melakukan pendekatan *Least Trimmed Square* (LTS).

Saran

Sebaiknya dalam penelitian selanjutnya, variabel independen lain yang mempengaruhi dapat ditambahkan. Dalam pengembangan penelitian selanjutnya sebaiknya digunakan metode estimasi lain untuk data tersensor kiri dengan *outlier*, sehingga dapat diketahui perbedaan hasil pendugaan parameter yang diberikan dan efektivitasnya dalam hal mengatasi data tersensor kiri dengan *outlier*.

DAFTAR PUSTAKA

- Adnan, R., M.N. Mohamad, dan H. Setan. 2003. Multiple Outliers Detection Procedures in Linear Regression. URL: <http://www.fs.utm.my/voldfs/images/stories/matematika/200319105.pdf>. Diakses pada tanggal 22 Januari 2011.
- Aris, M. 2006. Estimasi Parameter untuk Data Waktu Hidup yang Berdistribusi Rayleigh pada Data Tersensor Tipe II Beserta Simulasinya. *Skripsi* (Tidak dipublikasikan). Semarang: Universitas Negeri Semarang.
- Ardiyati, Hanna. 2011. Perbandingan Keefektifan Metode Regresi Robust Estimasi-M dan Estimasi-MM Karena Pengaruh Outlier dalam Analisis Regresi Linear (Contoh Kasus Data Produksi Padi di Jawa Tengah Tahun 2007). *Skripsi* (Tidak Dipublikasikan). Semarang: Universitas Negeri Semarang.
- Chen, Colin. 2002. *Robust Regression and Outlier Detection with the ROBUSTREG procedure*. SUGI Paper 265-27. SAS Institute: Cary, NC. SAS *OnLineDoc*. SAS Institute, Cary, NC: IML Robust Regression, Available from URL: <http://v8doc.sas.com/sashtml>
- Curwin S.L. 2005. Rehabilitation after tendon injuries. In: Maffuli N. *et al* (eds). *Tendon Injuries, Basic science and clinical medicine*. New York: Springer-Verlag: 242–61.
- Drapper N.R., Smith, H. 1996. *Applied Regression Analysis, 2nd edition*. New York: John Wiley & Sons. Chapman and Hall.

- Fathurahman. 2009. Pemilihan Model Regresi Terbaik Menggunakan Metode Akaike's Information Criterion dan Schwarz Information Criterion. *Skripsi* (Tidak Dipublikasikan). Samarinda: Universitas Mulawarman.
- Fox J. 2002. *Robust Regression*. Diakses tanggal 1 Oktober 2008.
- Hampel F.R., E.M., Ronchetti, P.J., Rousseeuw and Stahel, W.A. 2005. *Robust Statistics: The Approach Based on Influence Functions*.
- Hanum, Herlina. 2011. Perbandingan Metode Stepwise, Best Subset Regression, dan Fraksi dalam Pemilihan Model Regresi Berganda Terbaik. *Tesis* (Tidak Dipublikasikan). Palembang: Universitas Sriwijaya.
- Hendon H.H. 2003. Indonesian Rainfall Variability: Impacts of ENSO and Local Air–Sea Interaction. *J.Climate*, Volume 16: 1775–1790; 28–45.
- Kleinbaum, D.G., Klein M. 2005. *Survival Analysis—A Self Learning Text*, Second Edition. New York: Springer-Verlag: 6–7.
- Kuzmic, Petr. 2004. Practical Robust Fit of Enzyme Inhibition Data. *Methods in Enzymology*. Volume 383: 366–381.
- Laker, S.,R., Sullivan W.,J. 2008. Overuse Injury. eMedicine. Available from URL: <http://emedicine.medscape.com/article/313121-overview>. Diakses tanggal 28 April 2009.
- Maffulli, N. 2005. *Tendon Injuries: Basic Science and Clinical Medicine*. New York: Springer-Verlag.
- Maganaris, C.N, Narici, M.V. 2005. Mechanical Properties of Tendons. In: Maffuli N. *et al* (eds). *Tendon Injuries, Basic science and clinical medicine*. New York: Springer-Verlag: 14–19.
- Montgomery, D.C, Peck, E.A. 1992. *Introduction to Linear Regression Analysis*. New York: A Wiley-Interscience Publication.
- Myers, R.H. 1990. *Classical and Modern Regression with Applications*. Boston: PWS.
- Paavola, M. 2005. Epidemiology of tendon problems in sport. In: Maffuli N. *et al* (eds). *Tendon Injuries, Basic science and clinical medicine*. New York: Springer-Verlag: 33.
- Rousseeuw, R.J., Leroy, A.M. 1987. *Robust Regression and Outlier Detection*. John Wiley & Sons. New York.
- Ryan, T.P. 1997. *Modern Regression Methods*. New York: A Wiley-Interscience Publication.
- Sharma, P., Maffuli, N. 2005. Tendon Injury and Tendinopathy: Healing and Repair. *J Bone Joint Surg Am*, Volume 87: 187–202. <http://www.ejbs.org>. Cited on Jan 2005. Diakses tanggal 16 Mei 2009.
- Soemartini. 2007. *Pencilan (Oulier)*. Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam. Jatinangor: Universitas Padjajaran.
- Stromberg, A.J. 2004. Why write statistical software? The case of robust statistical methods. *Journal of Statistical Software*.
- Willems, G., Aelst, S.V. 2005. Fast and robust bootstrap for LTS. *Journal of Computational Statistics & Data Analyst*, Volume 48: 703–715.
- Wilson, J., Best, T.M. 2005. Common overuse tendon problem: A review and recommendations for treatment. The American Family Physician. URL: <http://www.aafp.org/afp/20050901/811.html>. Diakses tanggal 16 Mei 2009.