Laporan Penelitian Madya

Bidang Ilmu

Koefisien Determinasi sebagai Ukuran Kesesuaian Model pada Regresi Robust

Oleh:

Dra. Harmi Sugiarti,M.Si

Dra. Andi Megawarni,M.Ed

Pusat Penelitian dan Pengabdian kepada Masyarakat

UNIVERSITAS TERBUKA

2010

Koefisien Determinasi sebagai Ukuran Kesesuaian Model pada Regresi Robust

Harmi Sugiarti (harmi@mail.ut.ac.id), Andi Megawarni (mega@mail.ut.ac.id)

Abstrak

Dalam pemodelan, sebelum melakukan inferensi tentang parameter model regresi, dianggap perlu untuk mengetahui apakah model yang diperoleh sudah sesuai dengan data yang ada. Ada bebeberapa ukuran yang dapat dipergunakan, diantaranya adalah koefisien determinasi (R^2) yang diperoleh dengan metode OLS. Mengingat metode OLS sensitif terhadap pelanggaran asumsi-asumsi, maka adanya pengamatan pencilan *(outlier)* dalam data dapat mempengaruhi nilai R^2. Untuk mengatasi kelemahan-kelemahan dari metode yang ada, perlu dicoba untuk mendapatkan koefisien determinasi dengan metode lain yang bersifat tidak sensitif terhadap pelanggaran asumsi-asumsi, yaitu metode regresi robust *(robust regression)*. Penelitian ini bertujuan untuk menentukan koefisien determinasi garis regresi yang diperoleh dengan metode OLS, metode M *(maximum likelihood)*, dan metode LMS *(least median of square)* sebagai ukuran kesesuaian model.

Data yang digunakan dalam penelitian ini adalah data simulasi yang dibangkitkan dengan bantuan program MINITAB versi 13.1 dan SYSTAT. Hasil Penelitian menunjukkan bahwa untuk data yang tidak mengandung outlier, ke tiga metode yaitu metode OLS, metode M, dan metode LMS memberikan nilai koefisien determinasi (R^2) yang hampir sama. Sedangkan untuk data yang mengandung outlier, metode LMS tidak jauh berbeda dengan sebelumnya, namun metode OLS dan metode M memberikan nilai koefisien determinasi yang jauh lebih kecil dibanding sebelumnya

Kata kunci: koefisien determinasi, outlier, metode OLS, metode M, metode LMS

LEMBAR PENGESAHAN

Laporan Penelitian Madya
Bidang Ilmu
Lembaga Penelitian dan Pengabdian kepada Masyarakat UT

1.	a. Judul Penelitian	:	Koefisien Determinasi sebagai Ukuran Kesesuaian Model pada Regresi Robust
	b. Bidang Penelitian	:	Bidang Ilmu
	c. Klasifikasi Penelitian	:	Penelitian Mandiri
	d. Bidang Ilmu Statistika	:	Statistika
2.	Ketua Peneliti		
	a. Nama	:	Dra. Harmi Sugiarti, M.Si
	b. NIP	:	19670311 199202 2 001
	c. Golongan Kepangkatan	:	III/d
	d. Jabatan Akademik	:	Lektor
	e. Fakultas/Unit Kerja	:	FMIPA
3.	Anggota Tim Peneliti		
	a. Jumlah Anggota	:	1 (satu) orang
	b. Nama/Unit Kerja	:	Dra. Andi Megawarni, M.Ed / FMIPA
4.	Lama Penelitian	:	9 (sembilan) bulan
5.	Biaya Penelitian	:	Rp. 20.000.000,- (dua puluh juta rupiah)
6.	Sumber Biaya	:	LPPM-UT

Pondok Cabe, 29 Desember 2010

Mengetahui, Dekan FMIPA-UT Dr. Nuraini Soleiman,M.Ed NIP. 19540730 198601 2 001	Ketua Peneliti, Dra. Harmi Sugiarti, M.Si NIP. 19670311 199202 2 001
Mengetahui, Ketua LPPM- UT Drs. Agus Joko Purwanto, M.Si NIP. 19660508 199203 1 003	Menyetujui, Kepala Pusat Keilmuan LPPM-UT Dra. Endang Nugraheni, M.Ed,M.Si NIP. 19570422 198503 2 001

DAFTAR ISI

Halaman

BAB I
PENDAHULUAN

A. Latar Belakang

Pada pemodelan, sebelum melakukan inferensi tentang parameter model regresi, dianggap perlu untuk mengetahui apakah model yang diperoleh sudah sesuai dengan data yang ada. Ada bebeberapa ukuran yang dapat dipergunakan, diantaranya adalah koefisien determinasi, biasanya dinyatakan dengan R^2 dimana nilai ini menunjukkan proporsi variasi variabel dependen yang dijelaskan oleh variasi variabel independen.

Selain memberikan penaksir parameter $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ yang bersifat tak bias linear terbaik dari model regresi $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i$, $i = 1, 2, \ldots, n$, metode OLS (*ordinary least square*) juga memberikan ukuran koefisien determinasi R^2 yang sangat diperlukan dalam pemodelan.

Apabila terdapat asumsi yang tidak dipenuhi khususnya jika dalam data terdapat pencilan (*outlier*), maka patut dicoba metode yang bersifat tidak sensitif terhadap pelanggaran asumsi-asumsi, yakni regresi *robust*. Ada beberapa metode pendugaan/penaksiran koefisien garis regresi yang bersifat *robust* telah dikembangkan, diantaranya adalah metode pendugaan parameter regresi berdasarkan pada penduga M *(maximum likelihood estimators)* dan penduga LMS *(least median of square estimators)*.

Pada penelitian pendahuluan diperoleh hasil bahwa untuk data yang mengandung pencilan, secara umum metode regresi *robust* dengan penduga M dan penduga LMS lebih efisien dibanding metode OLS, sedangkan untuk data yang tidak mengandung pencilan, metode regresi *robust* dengan penduga M lebih efisien dibanding metode OLS dan metode LMS (Sugiarti, 2009). Dengan menghitung koefisien determinasi untuk masing-masing model yang dihasilkan oleh metode OLS dan metode regresi *robust* diharapkan dapat memberikan informasi yang lebih detil tentang hubungan yang ada di antara variabel independen dan variabel dependen dalam model regresi, dengan kata lain penentuan koefisien determinasi diharapkan dapat digunakan sebagai indikator untuk mengetahui apakah model yang diperoleh sudah sesuai dengan data.

B. Perumusan Masalah

Berdasarkan uraian di atas, permasalahan dalam penelitian ini dapat dinyatakan sebagai:

1. Bagaimana nilai koefisien determinasi untuk garis regresi dengan metode OLS?
2. Bagaimana nilai koefisien determinasi untuk garis regresi dengan metode M?
3. Bagaimana nilai koefisien determinasi untuk garis regresi dengan metode LMS?
4. Metode mana yang dipilih dalam menentukan nilai koefisien determinasi untuk garis regresi?

C. Tujuan Penelitian

Penelitian ini bertujuan untuk membandingkan koefisien determinasi garis regresi yang diperoleh dengan metode OLS, metode M, dan metode LMS sebagai ukuran kesesuaian model.

D. Manfaat Penelitian

Hasil penelitian ini dapat memberikan masukan bagi pengguna statistik dalam membangun suatu model jika data mengandung pengamatan pencilan.

BAB II

TINJAUAN PUSTAKA

A. Pengamatan Pencilan dan Berpengaruh

Model regresi $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} + \varepsilon_i$, atau ditulis dalam bentuk matriks $y_i = x_i' \beta + \varepsilon_i$ dengan x_i' menyatakan baris ke-i dari matriks rancangan X, β menyatakan parameter model dan ε_i menyatakan suku galat; seringkali dibangun dari data yang banyak mengandung kekurangan, diantaranya adalah adanya pengamatan pencilan yaitu pengamatan dengan sisaan yang cukup besar. Penolakan begitu saja suatu pencilan bukanlah prosedur yang bijaksana, karena adakalanya pengamatan pencilan memberikan informasi yang cukup berarti, misalnya karena pencilan timbul dari kombinasi keadaan yang tidak biasa yang mungkin saja sangat penting dan perlu diselidiki lebih lanjut. Pengamatan pencilan dapat merupakan pengamatan yang berpengaruh, artinya pengamatan yang dapat mempengaruhi hasil pendugaan koefisien regresi. Oleh karena itu tindakan membuang pengamatan yang berpengaruh akan mengubah secara berarti persamaan regresi serta kesimpulannya (Draper & Smith, 1981).

Untuk mendeteksi adanya pengamatan pencilan terhadap nilai-nilai X nya, dapat dilakukan dengan melihat matriks dugaan *(hat matrix)* yang didefinisikan sebagai $H = X(X'X)^{-1}X'$. Unsur ke-i pada diagonal utama matriks topi yakni h_{ii} biasanya dinamakan pengaruh *(leverage)* kasus ke-i yang dapat diperoleh dari $h_{ii} = x_i'(X'X)^{-1}x_i'$, dimana x_i' adalah vektor baris ke-i dari matriks X. Nilai h_{ii} terletak antara 0 dan 1 yang jumlahnya sama dengan p, yaitu banyaknya parameter regresi. Nilai *leverage* h_{ii} yang besar menunjukkan bahwa pengamatan ke-i berada jauh dari pusat semua pengamatan X. Suatu nilai leverage h_{ii} biasanya dianggap besar apabila nilainya lebih dari dua kali rataan semua leverage $(2p/n)$. Pada dasarnya nilai h_{ii} yang semakin besar menunjukkan semakin besar potensinya berpengaruh dalam pendugaan parameter regresi (Myers, 1990).

Sedangkan untuk mendeteksi adanya pengamatan yang berpengaruh, dapat digunakan nilai perbedaan dugaan peubah tak bebas terbakukan (*DFFITS*) yang dirumuskan sebagai:

$$(DFFITS)_i = \frac{\hat{y}_i - \hat{y}_{i,-i}}{s_{-i}\sqrt{h_{ii}}}$$

dimana: $\hat{y}_i$ = nilai pendugaan y_i, $\hat{y}_{i,-i}$ = nilai pendugaan y_i tanpa pengamatan ke-i, s_{-i} = dugaan simpangan baku tanpa pengamatan ke-i dan h_{ii} = unsur ke-i dari diagonal matriks topi. Jika p menyatakan banyaknya parameter dan n menyatakan banyaknya pengamatan, maka suatu pengamatan akan merupakan pengamatan berpengaruh dalam persamaan regresi apabila mempunyai nilai $|DFFITS|_i > 2\sqrt{(p/n)}$.

B. Metode Regresi *Robust* dengan Penduga M

Menurut Staudte & Sheather (1990), jika hubungan linear antara satu peubah respons dengan peubah-peubah bebasnya dimodelkan sebagai:

$$y_i = x_i'\beta + \varepsilon_i$$

dengan x_i' menyatakan baris ke-i dari matriks rancangan X, β menyatakan parameter model dan ε_i menyatakan suku galat. Nilai $\hat{y}_i$ dan sisaan e_i masing-masing didefinisikan sebagai:

$$\hat{y}_i = x_i'\beta \text{ dan } e_i = y_i - \hat{y}_i = y_i - x_i'\beta$$

Penduga M *(maximum likelihood estimator)* untuk model dengan p parameter ($\hat{\beta}_M$) diperoleh dengan cara meminimumkan fungsi konveks $\rho(x,e)$ yakni:

$$\min \sum_i \rho(x_i, e_i) = \min \sum_i \rho\left(x_i, y_i - x_i'\hat{\beta}_M\right)$$

atau mencari penyelesaian dari persamaan:

$$\sum_i x_i \Psi\left(x_i, y_i - x_i'\hat{\beta}_M\right) = 0$$

dengan $\Psi(x,e) = \rho'(x,e)$ untuk berbagai fungsi konveks $\rho(x,e)$ yang dapat diturunkan dan memenuhi $\Psi(x,0) = 0$. Penduga $\hat{\beta}_M$ yang diperoleh ini bukan merupakan skala *invariant*, yaitu jika sisaannya $\left(e_i = y_i - x_i'\hat{\beta}_M\right)$ digandakan dengan suatu konstanta akan diperoleh penyelesaian yang tidak sama seperti sebelumnya. Untuk mendapatkan skala *invariant*, digunakan nilai $\frac{e_i}{\sigma}$ sebagai pengganti e_i, dimana σ adalah faktor skala yang juga perlu diduga. Dengan demikian persamaan yang ada menjadi:

$$\sum_i x_i \, \Psi\left(x_i, \frac{e_i}{\sigma}\right) = \sum_i x_i \, \Psi\left(x_i, \frac{y_i - x_i' \, \hat{\beta}_M}{\sigma}\right)$$
$$= \sum_i x_i \left(y_i - x_i' \, \hat{\beta}_M\right) w_i = 0$$

dengan $w_i = w\left(x_i, \frac{y_i - x_i' \hat{\beta}_M}{\sigma}\right) = \frac{\Psi\left(x_i, \frac{e_i}{\sigma}\right)}{\frac{e_i}{\sigma}}$ adalah fungsi pembobot yang bernilai antara 0 dan 1. Secara umum fungsi pembobot dirumuskan sebagai berikut:

$$w_i = w\left(x_i, \frac{y_i - x_i' \hat{\beta}_M}{\sigma}\right) = \frac{\sigma v(x_i)}{e_i} \Psi\left(\frac{e_i}{\sigma v(x_i)}\right)$$

dimana Ψ adalah *influence function* dan $v(x_i)$ adalah suatu fungsi yang tidak diketahui dan tergantung pada x melalui nilai *leverage*. Dengan memilih fungsi Huber Ψ yang berbentuk:

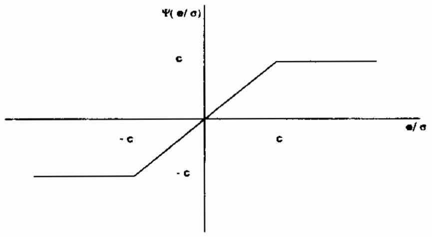
$$\Psi(\frac{e}{\sigma}) = \begin{cases} c \;, \text{ jika } \frac{e}{\sigma} > c \\ \frac{e}{\sigma} \;, \text{ jika } \left|\frac{e}{\sigma}\right| \le c \\ -c \;, \text{ jika } \frac{e}{\sigma} < -c \end{cases}$$

dan menentukan nilai $v(x_i) = \frac{(1 - h_{ii})}{\sqrt{h_{ii}}}$ serta $\hat{\sigma} = s_{(i)}$, nilai pembobot w_i menjadi tergantung pada kombinasi besarnya *leverage* dan *studentized residual* melalui $DFFITS$. Secara singkat nilai pembobot w_i dinyatakan dalam bentuk:

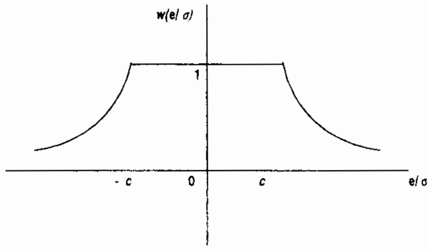
$$w\left(x_i, \frac{y_i - x_i' \hat{\beta}_M}{\sigma}\right) = w\left(x_i, \frac{e_i}{\sigma}\right) = \min\left(\frac{2\sqrt{p/n}}{|DFFITS|_i}, 1\right)$$

Dengan demikian persamaan $\sum_i \left(y_i - x_i' \hat{\beta}_M\right) w_i \, x_i = 0$ dapat dituliskan dalam bentuk matriks $X'WX\beta = X'WY$ yang kita kenal sebagai persamaan normal kuadrat terkecil tertimbang dengan W adalah matriks diagonal yang berisi pembobot. Solusi persamaan normal tersebut akan memberikan dugaan untuk

β yaitu $\hat{\beta}_M = (X'WX)^{-1}(X'WY)$ dan penduga-M untuk β diperoleh dengan cara melakukan iterasi sampai diperoleh suatu hasil yang konvergen, cara ini biasa dikenal sebagai metode kuadrat terkecil tertimbang secara iteratif *(iteratively reweighted least square).*



Gambar 1. Fungsi Huber



Gambar 2. Fungsi pembobot Huber

Berdasakan pembobot w_i dan penaksir-M parameter $\hat{\beta}_M$, matriks varians-kovarians untuk $\hat{\beta}_M$ yakni Σ_n dapat didekati dengan persamaan berikut:

$$\Sigma_n = \frac{1}{n-p}(X'D_1X)^{-1}(X'D_2X)(X'D_1X)^{-1}$$

dimana D_1 adalah matriks diagonal dengan elemen-elemen diagonalnya $\Psi'\left(\frac{e_i}{\sigma v(x_i)}\right)$ dan D_2 adalah matriks diagonal dengan elemen-elemen diagonalnya $w_i^2 e_i^2$ (Staudte,1990).

C. Metode Regresi *Robust* dengan Penduga LMS

Salah satu metode regresi *robust* yang juga sering digunakan adalah metode LMS *(least median of squares).* Metode ini mempunyai keuntungan untuk mengurangi pengaruh dari sisaan *(residual).* Menurut Rousseeuw dan Leroy (2003), penduga LMS diperoleh dengan mencari model regresi yang meminimumkan median dari h kuadrat sisaan (e_i^2) atau didefinisikan sebagai:

$$\hat{\beta}_{LMS} = \arg\min_{\beta}\ \underset{i}{median}\ e_i^2 \quad \text{dengan } e_i^2 = (y_i - x_i'\beta)^2\ ;\ i = 1,2,\ldots,n$$

Ukuran sebaran dari galat dapat ditaksir dengan cara menentukan dulu nilai awal:

$$s^0 = 1{,}4826\left(1 + \frac{5}{(n-p)}\right)\sqrt{\underset{i}{median}\ e_i^2(\hat{\beta}_{LMS})}$$

Faktor $1,4826 = \dfrac{1}{\Phi^{-1}(0,75)}$ digunakan karena $\dfrac{median_i\,|z_i|}{\Phi^{-1}(0,75)}$ merupakan penaksir konsisten untuk σ jika z_i berdistribusi $N\left(0,\sigma^2\right)$. Selanjutnya nilai awal s^0 digunakan untuk menentukan pembobot w_i untuk setiap pengamatan, yaitu:

$$w_i = \begin{cases} 1 & \text{jika } |e_i/\hat{\sigma}| \leq 2,5 \\ 0 & \text{jika } |e_i/\hat{\sigma}| > 2,5 \end{cases}$$

Berdasarkan pembobot w_i, maka nilai akhir taksiran σ untuk regresi LMS didefinisikan sebagai:

$$\hat{\sigma} = \sqrt{\frac{\left(\sum_{i=1}^{n} w_i\, e_i^2\right)}{\left(\sum_{i=1}^{n} w_i - p\right)}}$$

D. Koefisien Determinasi

Koefisien determinasi atau biasa disebut R^2 merupakan salah satu ukuran yang sederhana dan sering digunakan untuk menguji kualitas suatu persamaan garis regresi. Meskipun tidak direkomendasikan sebagai ukuran pemilihan model, namun R^2 cukup memberikan gambaran tentang kesesuaian variabel independen dalam memprediksi variabel dependen. Dalam kondisi asumsi klasik dipenuhi dan metode OLS digunakan untuk mendapatkan penaksir $\beta_0, \beta_1, \beta_2, ..., \beta_p$ dari model regresi $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_p x_{pi} + \varepsilon_i$, $i = 1,2,...,n$, maka menurut Myers (1990) koefisien determinasi dapat dihitung sebagai berikut:

$$R^2_{OLS} = \frac{JKR}{JKT} = 1 - \frac{JKS}{JKT} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

dimana JKR menyatakan jumlah kuadrat regresi, JKS menyatakan jumlah kuadrat sisaan, dan JKT menyatakan jumlah kuadrat total. Koefisien determinasi bernilai $0 < R^2_{OLS} < 1$.

Ukuran R^2_{OLS} sensitif terhadap pencilan, oleh karena itu dalam kondisi terdapat pencilan, Maronna (2006) mengusulkan koefisien determinasi dapat dihitung dengan rumus sebagai berikut:

$$R^2_M = 1 - \frac{\sum_{i=1}^{n} \rho\left(\frac{y_i - x_i'\hat{\beta}_M}{\hat{\sigma}}\right)}{\sum_{i=1}^{n} \rho\left(\frac{y_i - \hat{\mu}}{\hat{\sigma}}\right)}$$

dimana $\hat{\mu}$ merupakan penaksir M untuk $E(y)$ yakni solusi dari $\min \sum_{i=1}^{n} \rho\left(\frac{y_i - \mu}{\hat{\sigma}}\right)$, serta $\hat{\beta}_M$ dan $\hat{\sigma}$ masing-masing adalah penaksir M untuk β dan σ yang diperoleh berdasarkan fungsi ρ.

Selain itu, Rousseeuw (2003) mengusulkan koefisien determinasi untuk metode LMS adalah:

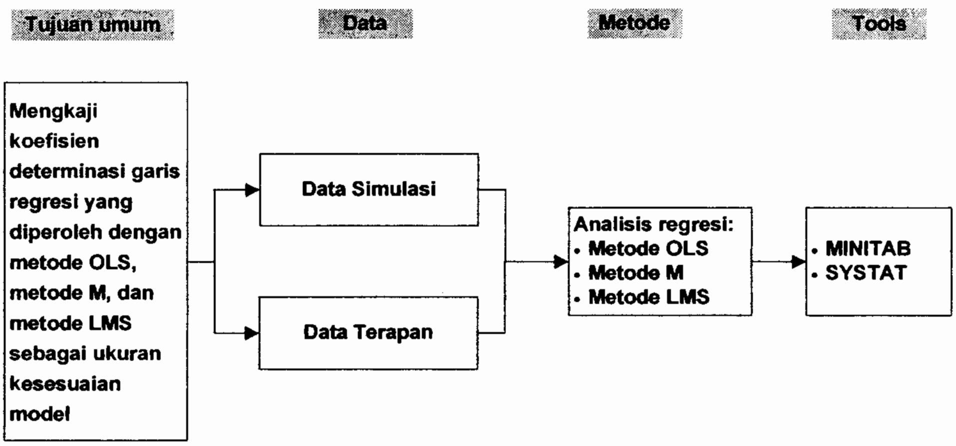
$$R^2_{LMS} = 1 - \left(\frac{med|e_i|}{mad(y_i)}\right)^2 = 1 - \left(\frac{med\left|y_i - x_i'\hat{\beta}_{LMS}\right|}{\underset{i}{med}\left\{\left|y_i - \underset{j}{med}\, y_j\right|\right\}}\right)^2$$

BAB III

METODE PENELITIAN

A. Desain Penelitian

Tujuan, data, metode, dan alat yang digunakan dalam penelitian ini disajikan dalam diagram berikut.



B. Data

Data yang digunakan dalam kajian metode ada dua macam, yaitu:

1. Data simulasi berupa data bangkitan yang diperoleh dengan bantuan program MINITAB versi 13.1.
2. Data terapan berupa nilai Tugas Tutorial Online (Tuton), Nilai Partisipasi Tuton, dan nilai UAS mata kuliah Metode Statistik I masa ujian 2008.1-2010.1

C. Tahapan Analisis

Adapun langkah-langkah yang akan dilakukan dalam penelitian ini adalah:

1. Membangkitkan sebanyak 40 pasang data sebagai peubah bebas (x_1, x_2) dan data galat (ε) dengan $\varepsilon \sim NIID(0,\sigma^2)$
2. Menentukan peubah tak bebas (y) melalui asumsi nilai $(\beta_0, \beta_1, \beta_2)$ tertentu untuk model $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$
3. Mendapatkan pengamatan pencilan (*outlier*) dengan mengganti sejumlah tertentu pengamatan y dengan nilai ekstrim sedemikian sehingga diperoleh pengamatan pencilan yang berpengaruh.

4. Mencari penaksir OLS, M, LMS untuk koefisien garis regresi untuk data simulasi dengan atau tanpa pencilan.
5. Menentukan koefisien determinasi R^2_{OLS}, R^2_{M}, dan R^2_{LMS} untuk data simulasi dengan atau tanpa pencilan.
6. Mencari penaksir OLS, M, LMS untuk koefisien garis regresi untuk data terapan.
7. Menentukan koefisien determinasi R^2_{OLS}, R^2_{M}, dan R^2_{LMS} untuk data terapan.

BAB IV
HASIL DAN PEMBAHASAN

Hasil Simulasi

Sebanyak empat puluh galat berdistribusi Normal dengan mean 0 dan variansi 1 dibangkitkan secara random dengan paket program MINITAB. Dengan mengasumsikan $\beta_0 = \beta_1 = \beta_2 = 1$, simulasi memberikan empat puluh pasang data (y, x_1, x_2). Penaksir koefisien garis regresi $(\hat{\beta})$ dan koefisien determinasi (R^2) untuk data simulasi tanpa *outlier* dapat dilihat pada Tabel 1.

Tabel 1. Penaksir Koefisien Garis Regresi dan Koefisien Determinasi untuk Data Tanpa *Outlier*

Koefisien	OLS	M	LMS
β_0	0,510	0,526	1,107*
β_1	0,973*	0,972*	0,969*
β_2	1,130*	1,129*	0,969*
R^2	0,961	0,994	0,982

* Signifikan pada $\alpha = 5\%$

Untuk data tanpa *outlier*, metode OLS memberikan koefisien determinasi $R^2_{OLS} = 0,961$ artinya 96,1% variabilitas dalam y dapat dijelaskan oleh x_1 dan x_2. Selain itu, metode OLS juga memberikan penaksir koefisien garis regresi masing-masing adalah $\hat{\beta}_0 = 0,510$, $\hat{\beta}_1 = 0,973$ dan $\hat{\beta}_2 = 1,130$. Secara simultan, model linear dengan $\hat{\beta}_0 = 0,510$, $\hat{\beta}_1 = 0,973$ dan $\hat{\beta}_2 = 1,130$ signifikan sebagaimana ditunjukkan oleh nilai $F = 457,21$ yang sangat tinggi dan nilai $p = 0,000$. Secara parsial, penaksir koefisien garis regresi $\hat{\beta}_1 = 0,973$ dan $\hat{\beta}_2 = 1,130$ signifikan pada $\alpha = 5\%$ sedangkan penaksir koefisien garis regresi $\hat{\beta}_0 = 0,510$ tidak signifikan. Hasil ini sesuai dengan koefisien garis regresi yang sudah ditentukan yakni $\beta_0 = \beta_1 = \beta_2 = 1$. Secara rinci, hasil olah Minitab dapat dilihat pada Tabel 2 berikut.

Tabel 2. Hasil Analisis Regresi untuk Data tanpa *Outlier*

```
The regression equation is
y = 0.510 + 0.973 x1 + 1.13 x2

Predictor        Coef     SE Coef         T        P        VIF
Constant        2.659       1.139      2.33    0.025
x1 sort        0.8048      0.1636      4.92    0.000        1.0
x2 sort        0.9693      0.1680      5.77    0.000        1.0

S = 1.006       R-Sq = 96.1%      R-Sq(adj) = 95.9%

Analysis of Variance

Source            DF          SS         MS        F        P
Regression         2      925.36     462.68   457.21    0.000
Residual Error    37       37.44       1.01
Total             39      962.80

Source       DF      Seq SS
x1            1      422.19
x2            1      503.16
```

Metode M memberikan koefisien determinasi $R_M^2 = 0,994$ artinya 99,4% variabilitas dalam y dapat dijelaskan oleh x_1 dan x_2. Penaksir koefisien garis regresi masing-masing adalah $\hat{\beta}_0 = 0,526$, $\hat{\beta}_1 = 0,972$ dan $\hat{\beta}_2 = 1,129$. Secara parsial, penaksir koefisien garis regresi $\hat{\beta}_1 = 0,972$ dan $\hat{\beta}_2 = 1,129$ signifikan pada $\alpha = 5\%$ sedangkan penaksir koefisien garis regresi $\hat{\beta}_0 = 0,526$ tidak signifikan. Hasil ini tidak jauh berbeda dengan hasil yang diperoleh metode OLS, yaitu sesuai dengan koefisien garis regresi yang sudah ditentukan bahwa $\beta_0 = \beta_1 = \beta_2 = 1$.

Sedikit berbeda dengan metode OLS dan metode M, metode LMS memberikan koefisien determinasi $R_M^2 = 0,982$ artinya 98,2% variabilitas dalam y dapat dijelaskan oleh x_1 dan x_2. Penaksir koefisien garis regresi masing-masing adalah $\hat{\beta}_0 = 1,107$, $\hat{\beta}_1 = 0,969$ dan $\hat{\beta}_2 = 0,969$. Secara parsial, penaksir koefisien garis regresi $\hat{\beta}_0 = 1,107$, $\hat{\beta}_1 = 0,969$ dan $\hat{\beta}_2 = 0,969$ signifikan pada $\alpha = 5\%$, hasil ini sesuai dengan koefisien garis regresi yang sudah ditentukan bahwa $\beta_0 = \beta_1 = \beta_2 = 1$.

Penaksir koefisien garis regresi $(\hat{\beta})$ dan koefisien determinasi (R^2) untuk data simulasi dengan 5% *outlier* dapat dilihat pada Tabel 3.

garis regresi $\hat{\beta}_0 = 0{,}994$, $\hat{\beta}_1 = 0{,}925$ dan $\hat{\beta}_2 = 1{,}017$ signifikan pada $\alpha = 5\%$. Hasil ini sesuai dengan koefisien garis regresi yang sudah ditentukan bahwa $\beta_0 = \beta_1 = \beta_2 = 1$.

Jika dilihat dari konsistensi nilai koefisien determinasi untuk masing-masing metode menunjukkan bahwa R^2_{LMS} lebih konsisten dibanding R^2_{OLS} dan R^2_M. Hal ini menunjukkan bahwa pada data tanpa *outlier* maupun data mengandung 5% *outlier,* ukuran koefisien determinasi yang diberikan oleh metode LMS lebih konsisten metode OLS dan metode M.

Tabel 4. Hasil Analisis Regresi untuk Data dengan 5% *Outlier*

```
The regression equation is
y = 2.66 + 0.805 x1 + 0.969 x2

Predictor        Coef     SE Coef          T        P        VIF
Constant        2.659       1.139       2.33    0.025
x1             0.8048      0.1636       4.92    0.000        1.0
x2             0.9693      0.1680       5.77    0.000        1.0

S = 3.334        R-Sq = 61.6%     R-Sq(adj) = 59.5%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         2      659.55      329.77     29.66    0.000
Residual Error    37      411.36       11.12
Total             39     1070.90

Source       DF      Seq SS
x1            1      289.37
x2            1      370.18

Unusual Observations
Obs       x1          y         Fit      SE Fit    Residual    St Resid
  1      2.0     20.000       4.268       0.965      15.732       4.93R
 40      2.0     20.000      10.084       0.742       9.916       3.05R

R denotes an observation with a large standardized residual
```

Penaksir koefisien garis regresi $(\hat{\beta})$ dan koefisien determinasi (R^2) untuk data simulasi dengan 10% *outlier* dapat dilihat pada Tabel 5. Pada data dengan 10% *outlier,* metode OLS memberikan koefisien determinasi $R^2_{OLS} = 0{,}445$ artinya 44,5% variabilitas dalam y dapat dijelaskan oleh x_1 dan x_2.

Tabel 5. Penaksir Koefisien Garis Regresi dan Koefisien Determinasi untuk Data dengan 10% *Outlier*

Koefisien	OLS	M	LMS
β_0	4,015*	2,700*	2,137*
β_1	0,774*	0,862*	0,866*
β_2	0,829*	0,956*	0,885*
R^2	0,445	0,663	0,979

* Signifikan pada $\alpha = 5\%$

Metode OLS memberikan penaksir koefisien garis regresi masing-masing adalah $\hat{\beta}_0 = 4{,}015$, $\hat{\beta}_1 = 0{,}774$ dan $\hat{\beta}_2 = 0{,}829$. Secara simultan, model linear dengan $\hat{\beta}_0 = 4{,}015$, $\hat{\beta}_1 = 0{,}774$ dan $\hat{\beta}_2 = 0{,}829$ signifikan sebagaimana ditunjukkan oleh nilai F yang masih cukup tinggi yaitu $F = 14{,}81$ dan nilai $p = 0{,}000$. Secara parsial, penaksir koefisien garis regresi $\hat{\beta}_0 = 4{,}015$, $\hat{\beta}_1 = 0{,}774$ dan $\hat{\beta}_2 = 0{,}829$ juga signifikan pada $\alpha = 5\%$. Hasil ini berbeda dengan koefisien garis regresi yang sudah ditentukan yakni $\beta_0 = \beta_1 = \beta_2 = 1$. Secara rinci, hasil olah Minitab dapat dilihat pada Tabel 6.

Tabel 6. Hasil Analisis Regresi untuk Data dengan 10% *Outlier*

```
The regression equation is
y = 4.02 + 0.774 x1 + 0.829 x2

Predictor          Coef      SE Coef          T        P        VIF
Constant          4.015        1.454       2.76    0.009
x1               0.7744       0.2089       3.71    0.001        1.0
x2               0.8295       0.2145       3.87    0.000        1.0

S = 4.258        R-Sq = 44.5%      R-Sq(adj) = 41.5%

Analysis of Variance

Source            DF          SS          MS          F        P
Regression         2      536.95      268.47      14.81    0.000
Residual Error    37      670.76       18.13
Total             39     1207.71

Source       DF       Seq SS
x1            1       265.84
x2            1       271.11

Unusual Observations
Obs         x1          y         Fit       SE Fit     Residual     St Resid
  1        2.0     20.000       5.564        1.232       14.436        3.54R
 18        2.0     20.000       7.223        0.972       12.777        3.08R
 20        7.0     20.000      11.095        0.976        8.905        2.15R
 40        2.0     20.000      10.541        0.948        9.459        2.28R

R denotes an observation with a large standardized residual
```

Metode M memberikan koefisien determinasi $R_M^2 = 0,588$ artinya 58,8% variabilitas dalam y dapat dijelaskan oleh x_1 dan x_2. Penaksir koefisien garis regresi masing-masing adalah $\hat{\beta}_0 = 2,700$, $\hat{\beta}_1 = 0,862$ dan $\hat{\beta}_2 = 0,956$. Secara parsial, penaksir koefisien garis regresi $\hat{\beta}_0 = 2,700$, $\hat{\beta}_1 = 0,862$ dan $\hat{\beta}_2 = 0,956$ signifikan pada $\alpha = 5\%$. Tidak jauh berbeda dengan metode OLS, metode M memberikan nilai yang berbeda dengan koefisien garis regresi yang sudah ditentukan adalah $\beta_0 = \beta_1 = \beta_2 = 1$.

Pada data yang mengandung 10% *outlier*, metode LMS memberikan koefisien determinasi $R_{LMS}^2 = 0,979$ artinya 97,9% variabilitas dalam y dapat dijelaskan oleh x_1 dan x_2. Penaksir koefisien garis regresi masing-masing adalah $\hat{\beta}_0 = 2,137$, $\hat{\beta}_1 = 0,866$ dan $\hat{\beta}_2 = 0,885$. Secara parsial, penaksir koefisien garis regresi $\hat{\beta}_0 = 2,137$, $\hat{\beta}_1 = 0,866$ dan $\hat{\beta}_2 = 0,885$ signifikan pada $\alpha = 5\%$. Hasil ini tidak sesuai dengan koefisien garis regresi yang sudah ditentukan bahwa $\beta_0 = \beta_1 = \beta_2 = 1$. Namun, jika dilihat dari konsistensi nilai koefisien determinasi untuk masing-masing metode menunjukkan bahwa R_{LMS}^2 lebih konsisten dibanding R_{OLS}^2 dan R_M^2. Hal ini menunjukkan bahwa pada data tanpa *outlier*, data mengandung 5% *outlier*, dan data mengandung 10% *outlier*, ukuran koefisien determinasi yang diberikan oleh metode LMS lebih konsisten metode OLS dan metode M.

Hasil ini sejalan dengan hasil kajian sebelumnya yang menunjukkan bahwa untuk data yang mengandung outlier, metode LMS lebih efisien dibanding metode M dalam menaksir koefisien garis regresi. Sedangkan untuk data yang tidak mengandung outlier, metode LMS kurang efisien dibanding metode M (Sugiarti, H dan Megawarni, A., 2009).

Hasil Terapan

Metode OLS memberikan koefisien determinasi $R_{OLS}^2 = 0,042$ artinya 4,2% variabilitas dalam UAS dapat dijelaskan oleh Tugas 1, Tugas 2, Tugas 3, dan Aktivasi mahasiswa.

BAB V

KESIMPULAN DAN SARAN

Secara umum dapat disimpulkan bahwa metode OLS, metode M, dan metode LMS memberikan nilai koefisien determinasi yang hampir sama pada data yang tidak mengandung *outlier.* Pada data yang mengandung *outlier,* metode LMS memberikan nilai koefisien determinasi yang tidak jauh berbeda dengan data yang tidak mengandung *outlier,* tetapi metode OLS dan metode M memberikan nilai koefisien determinasi yang jauh lebih kecil. Dengan kata lain, metode LMS lebih konsisten dibanding metode OLS dan metode M dalam memberikan nilai koefisien determinasi.

Secara khusus, metode LMS menunjukkan bahwa model linear sesuai untuk data terapan meskipun data terapan mengandung *outlier* sehingga analisis lanjut tentang koefisien garis regresi dapat dilakukan.

Sebagai saran, perlu dilakukan kajian lebih lanjut untuk mendapatkan penduga *robust* lainnya yang memungkinkan untuk mengatasi kendala yang ada.

DAFTAR PUSTAKA

Draper, N.R. & Smith, H. 1981. *Applied regression analysis.* 2nd ed. Wiley, New York.

Fleming,Ron. 2003. *Basic statistical concepts.* Diunduh dari http://www.uky.edu /Ag/Ag Econ .

Hoaglin,D.C., Mosteller,F., Tukey,J.W. 1983. *Understanding robust and exploratory data analysis.* Wiley, New York.

Maronna,R.A., Martin,R.D., Yohai,V.J. 1990. *Robust statistics: Theory and Methods.* Wiley, Chichester, West Sussex, UK

Myers, R.H. 1990. *Classical and modern regression with applications.* 2nd ed. PWS-Kent, Boston.

Rousseeuw,P.J. & Leroy,A.M. 2003. *Robust regression and outlier detection.* Wiley, New York.

Staudte, R.G. & Sheather, S.J. 1990. *Robust estimation and testing.* Wiley, New York.

Sugiarti, H. & Megawarni, A. 2009. Tingkat efisiensi penaksir M terhadap penaksir LMS dalam menaksir koefisien garis regresi. Laporan Penelitian LPPM Universitas Terbuka. Jakarta.

BIODATA

Nama	:	Andi Megawarni
Tempat, tanggal lahir	:	Makassar, 7 November 1953

Pengalaman Kerja

1988 – sekarang	:	Staf Jurusan Statistika FMIPA-UT
1989 - 1991	:	Penelaah materi dan format bahan ajar Jurusan Statistika FMIPA-UT
1995 - 2001	:	Koordinator Penulisan Bahan Ajar Jurusan Statistika FMIPA-UT
2001 – 2005	:	Sekretaris Jurusan Statistika FMIPA-UT
2005 - 2008	:	Asisten Pembantu Rektor II - UT

Riwayat Pendidikan

1987	:	S-1 Jurusan Matematika FMIPA - UI
2005	:	S-2 Jurusan Educatinal Psychology – UVvic,

Pelatihan dan Seminar

1998	:	Pelatihan Penulisan kisi-kisi soal, penulisan soal, item analysis dan revisi soal di UT
2001	:	*Workshop Loss Distribution* Jurusan Matematika FMIPA UI
2002	:	Peserta Seminar Nasional Matematika di FMIPA Universitas Indonesia
2003	:	Peserta Seminar Nasional Statistika di Universitas Brawijaya, Malang

Karya Tulis

2002	:	Sugiarti,H,& Megawarni,A, Selang Kepercayaan untuk Koefisien Garis Regresi Jika Ragam Galat Tidak Homogen dengan Metode OLS dan WLS
2003	:	Sugiarti,H, & Megawarni,A, Penggunaan Metode Regresi *Robust* untuk mencari Selang Kepercayaan Koefisien Garis Regresi Jika Ragam Galat Tidak Homogen
2004	:	Sugiarti,H,& Megawarni,A, Tingkat Efisiensi Metode Regresi *Robust* dalam Menaksir Koefisien Garis Regresi Jika Ragam Galat Tidak Homogen
2009	:	Sugiarti,H.& Megawarni,A. Tingkat Efisiensi Penaksir M terhadap Penaksir LMS dalam Menaksir Koefisien Garis Regresi