**PENAKSIR RASIO YANG EFISIEN UNTUK RATA-RATA POPULASI MENGGUNAKAN KOEFISIEN REGRESI ROBUST PADA SAMPING ACAK SEDERHANA**


**Mia Okto Morika[1], Arisman Adnan[2], Haposan Sirait[2]**

miaoktomoo@yahoo.co.id

[1]Mahasiswa Program S1 Matematika
[2]Dosen Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Riau
Kampus Binawidya Pekanbaru, 28293, Indonesia.



## ABSTRACT

The ratio estimators proposed by Kadilar, Candan and Cingi [4] for the population mean in simple random sampling using robust regression has been reviewed. This robust regression is defined by Huber M-estimators. These estimators are biased so that the *Mean Square Error* is calculated for each estimators to obtain which one is more efficient. This comparison shows that the ratio estimator involving of coefficient curtosis the more efficient then others.



**Key words:** *Simple Random Sampling, Robust Regression, Mean Square Error*


## 1. PENDAHULUAN

Metode kuadrat terkecil merupakan suatu metode penaksir parameter dengan prinsip meminimumkan jumlah kuadrat kesalahan pengamatan. Namun pada saat ditemukan adanya data pencilan dalam pengamatan, metode kuadrat terkecil akan memberikan fungsi yang kurang baik atau nilai penduga parameternya bersifat bias. Oleh karena itu, untuk memperbaiki garis regresi tersebut dapat dimanfaatkan metode regresi robust. Metode robust yang akan dibahas dalam penelitian ini merupakan suatu regresi robust dalam bentuk M-estimasi untuk meminimalkan error yang didefinisikan dalam fungsi Huber. M-estimasi Huber digunakan melalui fungsi $\psi(.)$dan fungsi yang akan diduga berbentuk $y = a + bx + e$.

Bentuk umum penaksir rasio sederhana untuk rata-rata populasi adalah $\bar{y}_r = \frac{\bar{y}}{\bar{x}}\bar{X}$, dimana $\bar{y}_r$ adalah penaksir rasio sederhana untuk rata-rata populasi, $\bar{y}$ dan $\bar{x}$ adalah rata-rata sampel, dan $\bar{X}$ adalah rata-rata populasi $X$. Penaksir rasio yang digunakan dalam penelitian ini merupakan penaksir yang bias. Sehingga untuk mendapatkan penaksir rasio yang efisien adalah dengan menghitung *MSE* untuk masing-masing penaksir. Semakin kecil *MSE* yang diperoleh maka penaksir tersebut semakin efisien.

Ketiga penaksir yang akan dibandingkan yaitu

$$\bar{y}_{a1} = \frac{\bar{y} + b_{rob}\left(\overline{X} - \bar{x}\right)}{\bar{x}}\overline{X} \tag{1}$$

$$\bar{y}_{a2} = \frac{\bar{y} + b_{rob}\left(\overline{X} - \bar{x}\right)}{\bar{x} + C_x}\left(\overline{X} + C_x\right) \tag{2}$$

$$\bar{y}_{a3} = \frac{\bar{y} + b_{rob}\left(\overline{X} - \bar{x}\right)}{\bar{x} + \beta_2(x)}\left[\overline{X} + \beta_2(x)\right] \tag{3}$$

dengan :

$b_{rob}$ adalah koefisien regresi robust,

$\beta_2(x)$ adalah koefisien kurtosis,

$C_x$ adalah koefisien variasi.

## 2. REGRESI LINEAR SEDERHANA

Bentuk umum model regresi linear sederhana dinyatakan dalam persamaan

$$Y_i = a + bX_i + \varepsilon_i \,. \tag{4}$$

Dari persamaan (4) maka kekeliruan ke-*i* ditulis

$$\varepsilon_i = Y_i - a - bX_i \,. \tag{5}$$

Dengan demikian jumlah kuadrat kekeliruan data terhadap garis regresi ditulis

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - a - bX_i\right)^2 \,. \tag{6}$$

Dengan meminimumkan persamaan pada (6), maka didapat nilai $\hat{a}$ dan $\hat{b}$ yaitu :

$$\hat{a} = \overline{Y} - \hat{b}\overline{X} \,,$$

dan

$$\hat{b} = \frac{s_{XY}}{s_{XX}} \,.$$

Sehingga nilai estimasi $Y_i$ pada persamaan (4) dapat ditulis

$$\hat{Y}_i = \hat{a} + \hat{b}X_i . \qquad (7)$$

Persamaan (7) membentuk suatu garis regresi. Pada saat adanya data pencilan dalam pengamatan, maka garis regresi yang didapat pada (7) akan memberikan fungsi yang kurang baik. Oleh karena itu, untuk memperbaiki garis regresi tersebut dapat dimanfaatkan metode regresi robust.

## 3. REGRESI ROBUST LINEAR SEDERHANA

Prosedur regresi robust ditujukan untuk mengakomodir data yang kekeliruannya berdistribusi tidak normal, inilah yang disebut dengan data pencilan dalam pengamatan. Regresi robust yang akan digunakan merupakan suatu regresi robust dalam bentuk M-estimasi yang didefinisikan dalam fungsi Huber [3].

Diberikan fungsi Huber yaitu

$$h(\varepsilon_i) = \begin{cases} \varepsilon_i^2 & -k \le \varepsilon_i \le k, \\ 2k|\varepsilon_i| - k^2 & \varepsilon_i < -k \ atau \ k < \varepsilon_i , \end{cases}$$

Model regresi linear sederhana dalam regresi robust dinyatakan dalam persamaan

$$Y_i = a_{rob} + b_{rob}X_i + \varepsilon_i . \qquad (8)$$

Dari persamaan (8) maka kekeliruan ke-*i* ditulis

$$\varepsilon_i = Y_i - a_{rob} - b_{rob}X_i . \qquad (9)$$

Parameter $a_{rob}$ dan $b_{rob}$ dalam regresi robust dapat diselesaikan dengan menggunakan M-estimasi melalui prosedur iterasi yang disebut *iteratively reweighted least squares* (IRLS).

M-estimasi berdasarkan fungsi $h(\varepsilon_i)$ merupakan solusi dari persamaan berikut:

$$\sum_{i=1}^{n} \psi(\varepsilon_i) \frac{\partial \varepsilon_i}{\partial j} = 0, \ j = \{a, b\}, \qquad (10)$$

dengan $\psi(\varepsilon_i) = \dfrac{d\rho(\varepsilon_i)}{d\varepsilon_i}$ adalah fungsi *influence*

Selanjutnya didefinisikan fungsi *weight* yaitu

$$w(\varepsilon_i) = \frac{\psi(\varepsilon_i)}{\varepsilon_i} .$$

Dengan demikian persamaan (10) menjadi

$$\sum_{i=1}^{n} w(\varepsilon_i)\varepsilon_i \frac{\partial \varepsilon_i}{\partial j} = 0,\ j = \{a, b\}. \tag{11}$$

Dengan menggunakan fungsi *weight*, maka nilai koefisien $a_{rob}$ dan $b_{rob}$ sebagai berikut

Untuk

$$h(\varepsilon_i) = \varepsilon_i^{\,2},\ \text{untuk } -k \le \varepsilon_i \le k$$

Diperoleh koefisien $a_{rob}$ dan $b_{rob}$ yaitu

$$\hat{a}_{rob} = \overline{Y} - \hat{b}_{rob}\overline{X},$$

$$\hat{b}_{rob} = \frac{\sum_{i=1}^{n} X_i Y_i - \left(\sum_{i=1}^{n} Y_i\right)\left(\sum_{i=1}^{n} X_i\right) \Big/ n}{\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2 \Big/ n},$$

kemudian

$$h(\varepsilon_i) = 2k|\varepsilon_i| - k^2,\ \text{untuk } \varepsilon_i < -k \text{ or } k < \varepsilon_i,$$

diperoleh koefisien $a_{rob}$ dan $b_{rob}$ yaitu

$$a_{rob} = \frac{k\left(\sum Y_i Z_i\right)\left(\sum X_i^2 Z_i\right) - k\sum (Y_i X_i) Z_i \left(\sum X_i Z_i\right)}{\sum \left(X_i^2 Z_i\right)\left(\sum Z_i\right) - \left(\sum X_i Z_i\right)^2} \tag{12}$$

dan

$$b_{rob} = \frac{k\left(\sum Y_i X_i\right) Z_i \left(\sum Z_i\right) - k\sum (Y_i Z_i)\left(\sum X_i Z_i\right)}{\sum \left(X_i^2 Z_i\right)\left(\sum Z_i\right) - \left(\sum X_i Z_i\right)^2} \tag{13}$$

Dengan demikian koefisien $a_{rob}$ dan $b_{rob}$ pada persamaan (12) dan persamaan (13) merupakan nilai $a_{rob}$ dan $b_{rob}$ untuk yang pertama. Setelah itu dapat kembali digunakan untuk mendapatkan nilai $a_{rob}$ dan $b_{rob}$ yang baru. Cara ini terus diulangi hingga $n$ kali pengulangan, sampai diperoleh perkiraan kesalahan antara $n-1$ dan $n$ mendekati sama[1].

## 4. SAMPLING ACAK SEDERHANA

Penarikan sampel acak sederhana adalah suatu metode untuk mengambil *n* unit dari populasi berukuran *N*, dimana setiap elemen mempunyai kesempatan yang sama untuk diambil menjadi anggota sampel. Pengambilan sampel dapat dilakukan dengan pengembalian atau tanpa pengembalian.

**Teorema 2.1 [2 :h.27]** Apabila sampel diambil secara acak sederhana tanpa pengembalian, maka variansi dari rata-rata $\bar{y}$ dari sampel acak sederhana adalah

$$\text{Var}\ (\bar{y}) = \frac{S_y^2}{n}\frac{N-n}{N} = \frac{S_y^2}{n}(1-f)$$

dengan $f = \dfrac{n}{N}$ adalah fraksi penarikan sampel.

**Bukti.** Bukti dari teorema ini dapat dilihat pada **[2]** .

**Deret Taylor untuk Dua Variabel [6 :h.47].** Misalkan $f(x,y)$ adalah suatu fungsi dua variabel dan $f^{(n+1)}$ ada pada $I$ dengan $(x_0, y_0) \in I$ . Jika $(x_0 + a, y_0 + b) \in I$ , maka

$$\begin{aligned} f(x_0 + a, y_0 + b) = f(x_0, y_0) &+ \frac{1}{1!}\left(a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y}\right) f(x_0, y_0) + \cdots \\ &+ \frac{1}{n!}\left(a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y}\right)^n f(x_0, y_0) \\ &+ \frac{1}{(n+1)!}\left(a\frac{\partial}{\partial x} + b\frac{\partial}{\partial y}\right)^{n+1} f(x_0 + \theta a, y_0 + \theta b) \end{aligned} \tag{14}$$

dengan $0 < \theta < 1$.

Deret taylor untuk dua variabel dapat ditulis

$$f(\bar{x}, \bar{y}) \approx f(\bar{X}, \bar{Y}) + \left.\frac{\partial f(\bar{x}, \bar{y})}{\partial \bar{x}}\right|_{\bar{X},\bar{Y}} (\bar{x} - \bar{X}) + \left.\frac{\partial f(\bar{x}, \bar{y})}{\partial \bar{y}}\right|_{\bar{X},\bar{Y}} (\bar{y} - \bar{Y}) \tag{15}$$

## 5. BIAS DAN *MSE* PENAKSIR RASIO UNTUK RATA-RATA POPULASI

Untuk meningkatkan presisi penaksiran digunakan suatu metode penaksir yaitu metode rasio. Suatu penaksir dikatakan efisien apabila penaksir tersebut mempunyai *MSE* yang minimum. Masing-masing penaksir yang dibahas merupakan penaksir yang bias. Kemudian akan ditentukan bias dan *MSE* dari masing-masing penaksir. Selanjutnya akan ditentukan penaksir yang efisien dengan membandingkan *MSE* dari setiap penaksir.

Bias dan *MSE* penaksir rasio pada sampling acak sederhana adalah

$$B(\bar{y}_R) \approx \frac{1-f}{n}\left(\bar{Y}C_x^2 - \bar{Y}C_{yx}\right)$$

$$MSE\left(\bar{y}_R\right) \approx \frac{1-f}{n}\left(R^2 S_x^2 - 2RS_{xy} + S_y^2\right)$$

Bias dan *MSE* penaksir rasio untuk rata-rata populasi menggunakan koefisien regresi robust pada sampling acak sederhana $\bar{y}_{a1}$ adalah

$$B(\bar{y}_{a1}) \approx \frac{1-f}{n}\frac{1}{\bar{X}}\left(B_{rob}S_x^2 - \rho S_y S_x + RS_x^2\right)$$

$$MSE\left(\bar{y}_{a1}\right) \approx \frac{1-f}{n}\left(B_{rob}{}^2 S_x^2 + 2B_{rob}R_{a1}S_x^2 + R_{a1}^2 S_x^2 - 2B_{rob}S_{xy} - 2R_{a1}S_{xy} + S_y^2\right)$$

Bias dan *MSE* penaksir rasio untuk rata-rata populasi menggunakan koefisien regresi robust pada sampling acak sederhana $\bar{y}_{a2}$ adalah

$$B(\bar{y}_{a3}) \approx \frac{1-f}{n}\frac{1}{\bar{X}_{a3}}\left(B_{rob}S_x^2 - \rho S_y S_x + RS_x^2\right)$$

$$MSE\left(\bar{y}_{a2}\right) \approx \frac{1-f}{n}\left(B_{rob}{}^2 S_x^2 + 2B_{rob}R_{a2}S_x^2 + R_{a2}^2 S_x^2 - 2B_{rob}S_{xy} - 2R_{a2}S_{xy} + S_y^2\right)$$

Bias dan *MSE* penaksir rasio untuk rata-rata populasi menggunakan koefisien regresi robust pada sampling acak sederhana $\bar{y}_{a3}$ adalah

$$B(\bar{y}_{a2}) \approx \frac{1-f}{n}\frac{1}{\bar{X}_{a2}}\left(B_{rob}S_x^2 - \rho S_y S_x + RS_x^2\right)$$

$$MSE\left(\bar{y}_{a3}\right) \approx \frac{1-f}{n}\left(B_{rob}{}^2 S_x^2 + 2B_{rob}R_{a3}S_x^2 + R_{a3}^2 S_x^2 - 2B_{rob}S_{xy} - 2R_{a3}S_{xy} + S_y^2\right)$$

Selanjutnya akan ditentukan penaksir yang efisien diantara ketiga penaksir yang diajukan, yaitu

1. Penaksir rasio $\bar{y}_{a1}$ dengan penaksir rasio $\bar{y}_{a2}$ .

Diperoleh bahwa penaksir rasio $\bar{y}_{a2}$ lebih efisien dari penaksir rasio $\bar{y}_{a1}$ jika

$$R_{a2_{(1)}} < R_{a2} < R_{a2_{(2)}}$$

Sebaliknya, penaksir rasio $\bar{y}_{a1}$ akan lebih efisien dari pada penaksir rasio $\bar{y}_{a2}$ jika

$$R_{a2} < R_{a2_{(1)}} \text{ atau } R_{a2} > R_{a2_{(2)}} .$$

2. Penaksir rasio $\bar{y}_{a1}$ dengan penaksir rasio $\bar{y}_{a3}$.

Diperoleh bahwa penaksir rasio $\bar{y}_{a3}$ lebih efisien dari penaksir rasio $\bar{y}_{a1}$ jika

$$R_{a3_{(1)}} < R_{a3} < R_{a3_{(2)}}$$

Sebaliknya, penaksir rasio $\bar{y}_{a1}$ akan lebih efisien dari pada penaksir rasio $\bar{y}_{a3}$ jika

$$R_{a3} < R_{a3_{(1)}} \text{ atau } R_{a3} > R_{a3_{(2)}} .$$

3. Penaksir rasio $\bar{y}_{a2}$ dengan penaksir rasio $\bar{y}_{a3}$.

Diperoleh bahwa penaksir rasio $\bar{y}_{a3}$ lebih efisien dari penaksir rasio $\bar{y}_{a2}$ jika

$$R_{a3_{(1)}} < R_{a3} < R_{a3_{(2)}}$$

Sebaliknya, penaksir rasio $\bar{y}_{a2}$ akan lebih efisien dari pada penaksir rasio $\bar{y}_{a3}$ jika

$$R_{a3} < R_{a3_{(1)}} \text{ atau } R_{a3} > R_{a3_{(2)}} .$$

## 6. CONTOH

Pertimbangkan data dari daerah Marmara Turkey yang digunakan oleh Kadilar dan Chingi [5] . Data berkaitan dengan taraf penghasilan apel dan jumlah dari pohon apel pada 106 desa di daerah Marmara tahun 1999.

Dengan menggunakan data tersebut akan ditentukan penaksir rasio yang paling efisien untuk menaksir rata-rata produksi apel yakni dengan menggunakan syarat penaksir lebih efisien yang diperoleh sebelumnya. Secara umum dapat ditunjukkan dengan menghitung MSE dari masing-masing penaksir. Untuk itu terlebih dahulu ditentukan nilai yang dibutuhkan. Dengan bantuan Microsoft Excel diperoleh nilai-nilai sebagaimana yang tertera pada Tabel 1.

**Tabel 1**

Nilai-nilai yang diperlukan untuk menghitung perbandingan *MSE*

| | | | |
|---|---|---|---|
| $N$ | 106 | $S_{xy}$ | 663757881.2 |
| $n$ | 30 | $\beta_2(x)$ | 34.57 |
| $\rho$ | 0.86 | $C_x$ | 2.09544335 |
| $B_{rob}$ | 5.02 | $C_y$ | 5.22080683 |
| $\bar{X}$ | 27421.6981 | $R_{A1}$ | 0.08068722 |
| $\bar{Y}$ | 2212.5943 | $R_{A2}$ | 0.08068156 |
| $S_x$ | 57460.6149 | $R_{A3}$ | 0.07977981 |
| $S_y$ | 11551.5276 | | |

Dengan menggunakan informasi pada Tabel 1, diperolehlah bahwa

1. Penaksir Rasio $(\bar{y}_{a2})$ lebih efisien dari penaksir $(\bar{y}_{a1})$, jika

$$\frac{1}{\bar{Y}R_{a2_{(2)}}} - \bar{X} < C_x < \frac{1}{\bar{Y}R_{a2_{(1)}}} - \bar{X}$$

2. Penaksir Rasio $(\bar{y}_{a3})$ lebih efisien dari penaksir $(\bar{y}_{a1})$, jika

$$\frac{1}{\bar{Y}R_{a3_{(2)}}} - \bar{X} < \beta_2(x) < \frac{1}{\bar{Y}R_{a3_{(1)}}} - \bar{X}$$

3. Penaksir Rasio $(\bar{y}_{a3})$ lebih efisien dari penaksir $(\bar{y}_{a2})$, jika

$$\frac{1}{\bar{Y}R_{a3_{(2)}}} - \bar{X} < \beta_2(x) < \frac{1}{\bar{Y}R_{a3_{(1)}}} - \bar{X}$$

## DAFTAR PUSTAKA

[1] Birkes, D. and Y, Dodge. 1993. Alternative Methods of Regression (John Wiley & Sons), New York.

[2] Cochran, W.G. 1991. *Teknik Penarikan Sampling*, Edisi Ketiga. Terjemahan *Sampling Techniques*, Oleh Rudiansyah dan E.R Osman. Penerbit Universitas Indonesia, Jakarta.

[3] Huber, P. 1981. *Robust Statistics*. Wiley, New York.

[4] Kadilar, C, Chandan, M. and Cingi, H. 2007. *Ratio Estimator using Robust Regression*. Hacettepe Journal of Mathematics and Statistics **36**, 181-188.

[5] Kadilar, C. & H. Cingi. 2004. Ratio Estimator in Simple Random Sampling, *Applied Mathematics and Computation* **151**: 893-902.

[6] Philips, G.M dan P.J, Taylor. 1972. *Theory and Applications of Numerical Analysis.* Second Edition. Academic Press, New York.