

PENAKSIR RASIO REGRESI YANG EFISIEN UNTUK RATA-RATA POPULASI PADA SAMPLING ACAK SEDERHANA MENGGUNAKAN DUA KARAKTER TAMBAHAN

Rizki Wulandari^{1*}, Rustam Efendi², Haposan Sirait²

¹ Mahasiswa Program Studi S1 Matematika

² Dosen Jurusan Matematika

Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Riau
Kampus Binawidya Pekanbaru (28293), Indonesia

*rizkiwulandari07@gmail.com

ABSTRACT

Three regression ratio estimators are used to estimate the population mean $\bar{Y}$ under simple random sampling. Using two auxiliary variables X and Z , it should be conditioned that there is high correlation between Y and X , Y and Z , whereas there may or may not be any correlation between X and Z . These estimators are recommended by Singh, Upadhyaya and Premchandra [5] which is a review of the article “An Improved Version of Regression Ratio Estimator with Two Auxiliary Variables in Sample Surveys.” All estimators are biased. The efficient estimator is one with the least Mean Square Error (MSE), determined by comparing each type of estimator.

Keywords: *regression ratio estimator, simple random sampling, biased estimator and mean square error.*

ABSTRAK

Tiga penaksir rasio regresi linear digunakan untuk menaksir rata-rata populasi $\bar{Y}$ pada sampling acak sederhana. Penaksir rasio regresi tersebut menggunakan dua karakter tambahan yaitu X dan Z , dimana terdapat korelasi yang tinggi antara Y dan X , Y dan Z , sedangkan antara X dan Z tidak harus berkorelasi. Penaksir tersebut diajukan oleh Singh, Upadhyaya dan Premchandra [5] yang merupakan *review* dari artikel “An Improved Version of Regression Ratio Estimator with Two Auxiliary Variables in Sample Surveys.” Ketiga penaksir merupakan penaksir bias. Penaksir yang efisien merupakan penaksir yang memiliki *Mean Square Error (MSE)* terkecil yang diperoleh dengan membandingkan *MSE* dari masing-masing penaksir.

Kata kunci: penaksir rasio regresi, sampling acak sederhana, penaksir bias dan *mean square error*.

1. PENDAHULUAN

Pengambilan sampel dari populasi berkarakter Y merupakan suatu cara pengumpulan data yang hanya mengambil sebagian dari anggota populasi. Pengambilan sampel

bertujuan untuk menaksir nilai parameter dari populasi yang dikenal dengan penaksir parameter populasi. Ada beberapa metode yang sering digunakan pada sampling acak, salah satunya sampling acak sederhana. Sampling acak sederhana yaitu metode pengambilan sampel dari anggota populasi secara acak sehingga mempunyai kesempatan yang sama untuk dipilih [3: h. 21].

Untuk mendapatkan hasil yang lebih akurat diperlukan informasi tambahan yang berhubungan dengan Y . Informasi tambahan tersebut merupakan informasi pendukung yang dinotasikan dengan X . Misalkan terdapat informasi tambahan lainnya yaitu Z , dengan Z diketahui. Terdapat korelasi yang tinggi antara Y dan X , Y dan Z , sedangkan antara X dan Z tidak harus berkorelasi.

Metode yang digunakan untuk menaksir rata-rata populasi pada sampling acak sederhana yang lainnya adalah metode rasio regresi. Tujuan metode rasio regresi adalah untuk meningkatkan ketelitian penelitian dengan mengambil manfaat hubungan antara penaksir regresi linear Y dan X yang berkorelasi terhadap Z .

Penelitian ini akan membandingkan tiga penaksir rasio regresi yang merupakan *review* dari artikel yang diajukan oleh Singh, Upadhyaya dan Premchandra [5] yaitu $\hat{Y}_{r1}$, $\hat{Y}_{r2}$ dan $\hat{Y}_{r3}$ untuk mendapatkan penaksir rasio regresi yang efisien.

2. SAMPLING ACAK SEDERHANA

Pengambilan sampel acak sederhana merupakan suatu metode untuk mengambil n unit sampel dari N unit populasi, sehingga setiap elemen C_n^N sampel yang berbeda mempunyai kesempatan yang sama untuk dipilih sebagai unit sampel. Pengambilan sampel ini adalah pengambilan sampel acak tanpa pengembalian agar karakteristik unit-unit lebih akurat.

Materi pendukung yang digunakan untuk menentukan penaksir rasio regresi yang efisien untuk rata-rata populasi adalah ekspektasi matematika dan sifat-sifat ekspektasi dari variabel acak.

Teorema 1 [1: h. 27] Apabila sampel berukuran n diambil dari populasi berukuran N yang berkarakter Y , dengan sampling acak sederhana tanpa pengembalian maka variansi rata – rata sampel $\bar{y}$ dirumuskan sebagai berikut

$$V(\bar{y}) = \frac{S_y^2}{n} \left(\frac{N-n}{N} \right).$$

Bukti dari Teorema ini dapat dilihat pada [3: h. 27].

Teorema 2 [3: h. 29] Jika y_i, x_i adalah sebuah pasangan yang bervariasi ditetapkan pada unit dalam populasi berukuran N dan $\bar{y}, \bar{x}$ adalah rata-rata dari sampel acak sederhana berukuran n , maka kovariansinya adalah

$$\text{cov}(\bar{y}, \bar{x}) = \frac{(1-f)}{n} \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{Y})(x_i - \bar{X}).$$

Bukti dari Teorema ini dapat dilihat pada [3: h. 29].

Teorema 3 Deret Taylor [2: h.184] Misalkan $k \in \mathbb{N}$ dan $I = [a, b]$, dengan $f : I \rightarrow \mathbb{R}$ dan $f', f'', \dots, f^{(k)}$ kontinu pada I dan $f^{(k+1)}$ ada pada (a, b) . Jika $x_0 \in I$ maka untuk sembarang $x \in I$ terdapat suatu titik $c \in (x_0, x)$ sehingga

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \dots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \frac{f^{(k+1)}(c)}{(k+1)!}(x - x_0)^{k+1}$$

Bukti dari teorema ini dapat dilihat pada [2: h.184]

Dari Teorema 2.3 jika $x_0 = 0$, maka Deret Taylor disebut deret Maclaurin yaitu

$$f(x) = f(0) + x f'(0) + \frac{x^2}{2!} f''(0) + \dots + \frac{x^k}{k!} f^{(k)}(0) + \frac{x^{k+1}}{(k+1)!} f^{(k+1)}(c)$$

3. PENAKSIR REGRESI LINEAR

Bentuk umum model regresi linear sederhana Y atas X dinyatakan dalam persamaan

$$Y = A + BX + e, \quad (1)$$

dengan Y adalah variabel tak bebas, X adalah variabel bebas, A dan B adalah parameter (koefisien regresi), dan e adalah kesalahan penelitian sampel [3: h. 221].

Metode yang sering digunakan untuk menentukan taksiran dari parameter A dan B adalah metode kuadrat terkecil (*least square method*), yaitu suatu metode penaksir dengan prinsip meminimumkan jumlah kuadrat kesalahan pengamatan. Misalkan x_i, y_i adalah n pasangan data pengamatan $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, dengan demikian persamaan (1) dapat ditulis

$$y_i = A + BX_i + e_i, \text{ untuk } i = 1, 2, 3, \dots, n \quad (2)$$

Jumlah kuadrat kesalahan penelitian sampel dapat ditulis

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - A - BX_i)^2$$

Dengan meminimumkan jumlah kuadrat kesalahan diperoleh taksiran untuk B

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

dan taksiran untuk A didapat

$$a = \frac{\sum_{i=1}^n y_i}{n} - b \frac{\sum_{i=1}^n x_i}{n}.$$

$$a = \bar{y} - b\bar{x}. \quad (3)$$

Garis regresi linear yang melalui titik pangkal artinya jika $a=0$, maka persamaan (3) menjadi

$$\bar{y} = b\bar{x}. \quad (4)$$

Dari persamaan (2), koefisien b yang diperoleh dari sampel digunakan juga untuk menaksir parameter populasi, penaksir untuk rata-rata populasi dinotasikan dengan $\hat{\bar{Y}}$ dan dirumuskan dengan

$$\hat{\bar{Y}} = b\bar{X}. \quad (5)$$

Selanjutnya dilakukan pengurangan terhadap persamaan (5) dengan persamaan (4) secara aljabar maka diperoleh

$$\hat{\bar{Y}} = \bar{y} + b(\bar{X} - \bar{x}).$$

$\hat{\bar{Y}}$ disebut penaksir regresi linear untuk rata-rata populasi yang dinotasikan dengan $\hat{\bar{Y}}_{LR}$ dan dirumuskan sebagai berikut

$$\hat{\bar{Y}}_{LR} = \bar{y} + b(\bar{X} - \bar{x}).$$

4. BIAS DAN MSE PENAKSIR UNTUK RATA-RATA POPULASI

Selanjutnya akan dibahas bias dan MSE dari ketiga penaksir rasio regresi yang diajukan oleh Singh, Upadhyaya dan Premchandra [5]. Penaksir yang efisien dengan membandingkan MSE dari setiap penaksir dengan menggunakan efisiensi relatif. Suatu penaksir dikatakan efisien apabila mempunyai MSE yang terkecil.

Pertama, penaksir rasio regresi yang diajukan oleh Mohanty dan Sahoo [5]

$$\begin{aligned} \hat{\bar{Y}}_{lr1} &= \bar{y}_{lr} \frac{u}{\bar{K} - \alpha \bar{u} + \alpha} \\ &= \bar{y} + b_{yx}(\bar{X} - \bar{x}) \left(\frac{\frac{\bar{z}}{\bar{Z}}}{(1-\alpha)\frac{\bar{z}}{\bar{Z}} + \alpha} \right). \end{aligned} \quad (5)$$

Kedua, penaksir rasio regresi yang diajukan oleh Naik dan Gupta (1991 a) [5]

$$\begin{aligned} \hat{\bar{Y}}_{lr2} &= \bar{y}_{lr} \frac{u}{\bar{K} - \alpha \bar{u}^2 + \alpha u^{-1}} \\ \hat{\bar{Y}}_{lr2} &= \bar{y} + b_{yx}(\bar{X} - \bar{x}) \left(\frac{1 - \alpha \left(\frac{\bar{z}}{\bar{Z}} \right)^2 + \alpha \frac{\bar{z}}{\bar{Z}}}{\bar{K} - \alpha \bar{u}^2 + \alpha u^{-1}} \right)^{-1}. \end{aligned} \quad (6)$$

Ketiga, penaksir rasio regresi yang diajukan oleh Naik dan Gupta (1991 a) [5]

$$\hat{\bar{Y}}_{rlr3} = \bar{y}_{lr} \left[\bar{X} + \alpha \frac{\bar{z}}{\bar{Z}} \right]$$

$$\hat{\bar{Y}}_{rlr} = \bar{y} + b_{yx} (\bar{X} - \bar{x}) \left[\bar{X} + \alpha \frac{\bar{z}}{\bar{Z}} \right]. \quad (7)$$

$u = \frac{\bar{z}}{\bar{Z}}$ dan α adalah konstanta.

Berikut ini adalah bias dan *MSE* penaksir rasio regresi untuk rata-rata populasi pada sampling acak sederhana menggunakan dua variabel tambahan. Bias dan *MSE* dari persamaan (5) adalah

$$B(\hat{\bar{Y}}_{rlr1}) \approx \frac{\bar{Y}(\bar{X} - 3\alpha + \alpha^2)}{\bar{Z}^2} S_z^2 + \frac{\alpha}{\bar{Z}} S_{yx} - \frac{\alpha}{\bar{Z}} B_{yx} S_{xz}.$$

$$MSE(\hat{\bar{Y}}_{rlr1}) \approx \frac{1-f}{n} Y^2 \left[C_y^2 - \rho_{yx}^2 C_y^2 + \alpha^2 C_z^2 - 2\alpha \rho_{yx} \rho_{xz} C_y C_z + 2\alpha \rho_{yz} C_z C_y \right]. \quad (8)$$

Bias dan *MSE* dari persamaan (6) adalah

$$B(\hat{\bar{Y}}_{rlr2}) \approx \frac{\bar{Y}(\bar{X} - 3\alpha + \alpha^2)}{\bar{Z}^2} S_z^2 + \frac{(\alpha - 1)}{\bar{Z}} S_{yz} - \frac{2\alpha}{\bar{Z}} B_{yx} S_{xz}.$$

$$MSE(\hat{\bar{Y}}_{rlr2}) \approx \frac{1-f}{n} Y^2 \left[S_y^2 + b_{yx}^2 S_x^2 + \left(\frac{\bar{Y}(\alpha - 2)}{\bar{Z}} \right)^2 S_z^2 - 2b_{yx} \frac{\bar{Y}(\alpha - 2)}{\bar{Z}} S_{xz} \right. \\ \left. + 2\left(b_{yx} \right) S_{yx} + 2 \left(\frac{\bar{Y}(\alpha - 2)}{\bar{Z}} \right) S_{yz} \right]. \quad (9)$$

Bias dan *MSE* dari persamaan (7) adalah

$$B(\hat{\bar{Y}}_{rlr3}) \approx \bar{Y} + \alpha \bar{Y} - \frac{\alpha}{\bar{Z}} \text{Cov}(\bar{y}_{lr}, \bar{z}) - \bar{Y} \bar{Z}.$$

$$MSE(\hat{\bar{Y}}_{rlr3}) \approx \frac{1-f}{n} Y^2 \left[C_y^2 - \rho_{yx}^2 C_y^2 + \alpha^2 C_z^2 + 2\alpha \rho_{yx} \rho_{xz} C_y C_z - 2\alpha \rho_{yz} C_z C_y \right]. \quad (10)$$

5. PENAKSIR RASIO REGRESI YANG EFISIEN

Penaksir yang efisien dari penaksir yang bias, dapat ditentukan dengan cara membandingkan *MSE* dari masing-masing penaksir tersebut.

1. Perbandingan *MSE* dari penaksir $\hat{\bar{Y}}_{rlr1}$ dengan penaksir $\hat{\bar{Y}}_{rlr2}$ jika

i). kemungkinan pertama

$$\rho_{yz} > \frac{-\alpha C_z + C_z + \rho_{yx} \rho_{xz} C_y}{C_y} . \quad (11)$$

ii). Kemungkinan kedua

$$\rho_{yz} < \frac{-\alpha C_z + C_z + \rho_{yx} \rho_{xz} C_y}{C_y} .$$

2. Perbandingan *MSE* dari penaksir $\hat{Y}_{rlr1}$ dengan penaksir $\hat{Y}_{rlr3}$ jika

i). kemungkinan pertama

$$\rho_{yz} < \frac{\rho_{yx} \rho_{xz} C_y}{C_y} . \quad (12)$$

ii). Kemungkinan kedua

$$\rho_{yz} > \frac{\rho_{yx} \rho_{xz} C_y}{C_y} .$$

3. Perbandingan *MSE* dari penaksir $\hat{Y}_{rlr2}$ dengan penaksir $\hat{Y}_{rlr3}$ jika

i). kemungkinan pertama

$$\rho_{yz} > \frac{\alpha C_z - C_z + \alpha \rho_{yx} \rho_{xz} C_y}{\alpha C_y C_z - C_y} . \quad (13)$$

ii). Kemungkinan kedua

$$\rho_{yz} < \frac{\alpha C_z - C_z + \alpha \rho_{yx} \rho_{xz} C_y}{\alpha C_y C_z - C_y} .$$

6. CONTOH

Sebagai contoh dari pembahasan, diberikan data mengenai faktor-faktor yang mempengaruhi banyaknya tabungan (*Y*) di Kelurahan Damun Kecamatan Bengkalis tahun 2005 [4] dengan $n=20$ menggunakan variabel tambahan yaitu pendapatan/bulan (*X*) dan pengeluaran/bulan ($\bar{Z}$). Dengan menggunakan data tersebut akan ditentukan penaksir rasio regresi linear yang paling efisien untuk menaksir rata-rata jumlah tabungan yakni dengan menggunakan syarat penaksir lebih efisien yang diperoleh sebelumnya dan secara umum dapat ditunjukkan dengan menghitung *MSE* dari masing-masing penaksir yang diberikan. *MSE* dari masing-masing penaksir dihitung dengan menentukan terlebih dahulu nilai yang dibutuhkan, dengan menggunakan *Microsoft Excel* diperoleh nilai-nilai sebagai berikut dengan $\alpha = 0,5$.

$N = 35$	$n = 20$	$S_{yx} = 1,3582$	
$\bar{Y} = 933000$	$S_x = 2,32798$	$S_{yz} = 2,6745$	$\rho_{yz} = 0,325089$
$\bar{X} = 227965$	$S_y = 1,19467$	$S_{xz} = 8,9116$	$\rho_{xz} = 0,55568$
$\bar{Z} = 1069171$	$S_z = 6,88896$	$\rho_{yx} = 0,48864$	

Tabel 1. Nilai *Mean Square Error* dari Masing-Masing Penaksir

Penaksir Rasio	<i>MSE</i>
$\hat{Y}_{rlr1}$	$1,93607 \times 10^2$
$\hat{Y}_{rlr2}$	$4,797736 \times 10^3$
$\hat{Y}_{rlr3}$	$3,65494 \times 10^3$

Selanjutnya substitusikan nilai-nilai yang diperoleh dari tabungan, pendapatan dan pengeluaran masyarakat Kelurahan Damun tahun 2005 ke persamaan (10), (11), (12), diperoleh

1. $MSE(\hat{f}_{rlr1}) < MSE(\hat{f}_{rlr2})$ jika $\rho_{yz} > -0,25174$.
2. $MSE(\hat{f}_{rlr1}) < MSE(\hat{f}_{rlr3})$ jika $\rho_{yz} < 2,71536$.
3. $MSE(\hat{f}_{rlr2}) < MSE(\hat{f}_{rlr3})$ jika $\rho_{yz} > 0,53129$.

Berdasarkan Tabel 1, maka diketahui bahwa penaksir $\hat{f}_{rlr1}$ lebih efisien dari penaksir $\hat{f}_{rlr2}$ dan penaksir $\hat{f}_{rlr3}$.

7. KESIMPULAN

Setelah diperoleh nilai *MSE* dari penaksir rasio regresi menggunakan dua karakter tambahan untuk rata-rata populasi pada sampling acak sederhana, kemudian membandingkan *MSE* dari masing-masing penaksir, sehingga dapat disimpulkan berdasarkan data tabungan (*Y*), pendapatan (*X*) dan pengeluaran (*Z*) bahwa penaksir rasio regresi $\hat{f}_{rlr1}$ lebih efisien dibandingkan dengan penaksir $\hat{f}_{rlr2}$ dan penaksir $\hat{f}_{rlr3}$ jika syarat-syarat efisiensi terpenuhi.

DAFTAR PUSTAKA

- [1] Bain, L.J. & M. Engelhardt. 1991. *Introduction to Probability and Mathematical Statistics*, Second Edition. Duxbury Press, California.
- [2] Bartle, R.G. & D.R Sherbert. 2000. *Introduction to Real Analysis*, Third Edition. John Wiley & Sons Inc, New York.
- [3] Cochran, W.G. 1977. *Sampling Techniques*, Third Edition. John Wiley & Sons, New York.
- [4] Herlina. 2005. Faktor-faktor yang Mempengaruhi Jumlah Tabungan Masyarakat Kelurahan Damun Wilayah Kecamatan Bengkalis Kabupaten Bengkalis. Skripsi Fakultas Ekonomi Universitas Riau, Pekanbaru.

- [5] Singh, H.P., L.N. Upadhyaya, & Premchandra. 2009. An Improved Version of Regression Ratio Estimator With Two Auxiliary Variables in Sample Surveys, *Statistics In Transition*, 10: 85-100.