

PENDUGA RATAAN GEOMETRIK PADA SAMPEL HIMPUNAN TERURUT UNTUK DISTRIBUSI NORMAL

Sukma Adi Perdana

sukmaperdana@umrah.ac.id

Universitas Maritim Raja Ali Haji

2018

Abstrak

Pada kajian ini, dikembangkan sebuah estimator atau penduga untuk rata-rata geometrik pada desain sampel himpunan terurut. Sampel himpunan terurut adalah sekelompok unit sampel yang diambil dari populasi dimana anggota populasinya diurutkan sebelumnya berdasarkan variabel tertentu yang memiliki korelasi yang kuat dengan populasi yang dibicarakan sebelum pengukuran yang nyata terhadap anggota populasi yang dibicarakan dimana hal ini akan mengurangi biaya dan waktu perhitungan. Kajian dilakukan untuk menginvestigasi bias dan efisiensi relatif dari penduga rata-rata geometrik yang dikembangkan dan perbandingan efisiensi relatif dilakukan untuk data dari populasi yang berdistribusi normal. Hasil kajian menunjukkan penduga pada sampel himpunan terurut mengungguli penduga pada sampel acak sederhana.

Kata kunci: Sampel Himpunan Terurut, Sampel Acak Sederhana, Rataan Geometrik, Efisiensi Relatif

Abstract

In this study, an estimator for geometric mean of population in ranked set sampling design is developed. Ranked set sampling is a set of sampling units which is drawn from population that can be ranked by using another certain mean without the actual measurement of the variable of interest which is to measure this variable is more costly and time consuming. The study investigates the bias and relative efficiency of the proposed estimator and the efficiency comparison is made for normal distribution. It is shown that the new estimators outperform its competitor in the literature.

Keywords: Ranked Set Sampling, Simple Random Sampling, Geometric Mean, Relative Efficiency

I. Pendahuluan

Salah satu teori sampling yang merupakan bagian dari statistika adalah memastikan bahwa data sampel memang mewakili populasi yang diamati. Usaha dalam memastikan hal ini mengarahkan ke beberapa

metode yang digunakan dalam pengambilan data sampel. Data sampel adalah himpunan bagian dari populasi yang diambil dengan tujuan untuk mengestimasi parameter dari populasi. Sampel acak sederhana adalah salah satu metode yang digunakan dalam pengambilan data sampel,

namun metode ini relatif tidak mencakup populasi secara penuh. Hal ini penting karena kekurangan dari sampel acak sederhana ini dapat mengurangi keakuratan dari penduga dalam menduga parameter populasi. Untuk menyikapi masalah tersebut ada metode alternatif yang dapat digunakan dalam pengambilan data sampel yaitu sampel himpunan terurut.

Rataan geometrik adalah salah satu perhitungan yang dapat digunakan sebagai nilai tengah dari variabel acak di dalam kehidupan nyata. Sebagai contoh, estimasi tingkat pertumbuhan yang akan datang di dalam bidang ekonomi (Spizman, 2008), menjelaskan tentang konsentrasi bakteri *fecal coliform* untuk kualitas air, dan untuk mengevaluasi investment returns dan tingkat fluktuasi bunga.

Mempertimbangkan kegunaan dari sampel himpunan terurut yang mampu meningkatkan akurasi dari penduga dalam menduga parameter serta aplikasi dari rata-rata geometrik maka perlu dilakukan sebuah penelitian untuk penduga rata-rata geometrik pada sampel himpunan terurut, sehingga tujuan dari penelitian ini adalah untuk mengenalkan penduga untuk rata-rata geometrik pada sampel himpunan terurut serta melakukan evaluasi terhadap penduga yang diusulkan dengan menelaah bias dan efisiensi relatif dari penduga yang diusulkan pada sekelompok data yang memiliki populasi yang berdistribusi normal.

Seperti diketahui bersama bahwa sampel adalah sebagian anggota dari populasi yang diambil untuk mewakili populasi berkaitan dengan proses pendugaan parameter populasi. Salah satu desain sampel yang sering digunakan adalah sampel acak sederhana. Sampel acak sederhana adalah desain sampel yang digunakan untuk memperoleh n data unit sampel dari N data unit populasi sehingga setiap kombinasi $\binom{N}{n}$ dari data sampel tertentu memiliki kesempatan atau peluang yang sama untuk dipilih (Cochran, 1977). Wolfe (2012) memberikan definisi bahwa sampel acak sederhana dari n data yang berasal distribusi peluang tertentu dengan fungsi kepadatan peluang $f(x)$ dan fungsi kumulatif

peluang $F(x)$ adalah kumpulan dari variabel acak X_1, \dots, X_n yang memenuhi dua properti sebagai berikut:

- (i) Setiap $X_i, i = 1, \dots, n$ memiliki distribusi peluang yang sama.
- (ii) N variabel acak X_1, \dots, X_n adalah saling bebas.

Penduga dari rata-rata aritmatik pada sampel acak sederhana dinotasikan \bar{x} seperti pada persamaan 1 berikut:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \dots \dots \dots (1)$$

dan Varians dari penduga dapat dihitung dengan persamaan berikut:

$$V(\bar{x}) = \frac{s^2}{n} \left(\frac{N-n}{N} \right) \dots \dots \dots (2)$$

Dimana

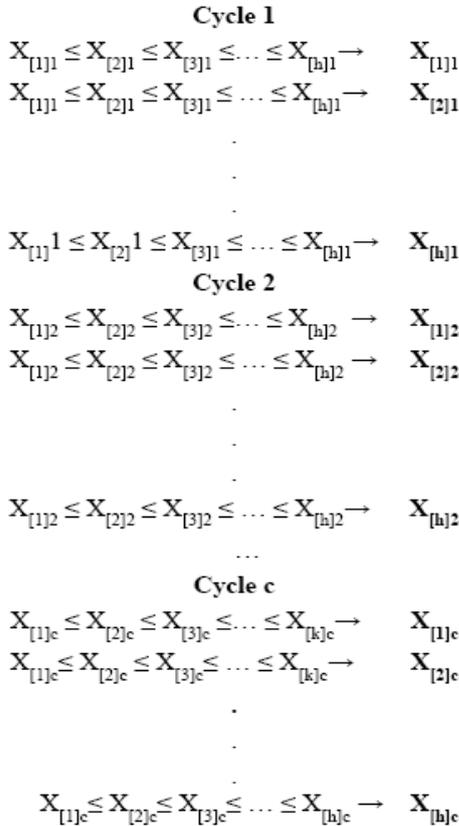
$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \dots \dots \dots (3)$$

Dan dapat dibuktikan bahwa penduga \bar{x} adalah penduga yang tidak bias dari μ seperti yang dijelaskan sebagai berikut:

$$\begin{aligned} E(\bar{x}) &= E\left(\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{1}{n} \sum_{i=1}^n E(x_i) = \frac{1}{n} \sum_{i=1}^n \mu \\ &= \frac{1}{n} n \mu = \mu \end{aligned}$$

Setelah kita memahami sampel acak sederhana selanjutnya didiskusikan mengenai sampel himpunan terurut. Dari (Chen, Bai, and Sinha, 2003), bentuk asli dari Sampel Himpunan Terurut yang diusulkan oleh McIntyre dapat dideskripsikan sebagai berikut. Pertama, pilih sampel berukuran h unit dengan sampel acak sederhana dari populasi dan h unit sampel tersebut diurut menurut concomitant variable dari X , tanpa pengukuran aktual dari X . Dan kemudian, urutan pertama dari unit ini diidentifikasi dan diambil untuk perhitungan X . Sisa unit dari sampel yang diambil tadi lalu dibuang. Selanjutnya, proses sama yang lain dilakukan dan lagi unit-unit dari sampel diurut dengan *Concomitant variable* dari X , unit urutan kedua diambil untuk perhitungan dari X dan sisa unit dibuang. Proses ini dilakukan sebanyak h kali dan akan didapat unit sampel sebanyak h unit. Langkah ini disebut sebagai satu siklus

sampel himpunan terurut. Siklus ini dapat diulang sebanyak c kali dan akan menghasilkan unit sampel sebanyak $n = c.h$ unit. Proses dari sampel himpunan terurut dapat diilustrasikan pada gambar 1.



Gambar 1. Proses Sampel Himpunan Terurut

Penduga dari rata-rata aritmatik pada sampel himpunan terurut yang dinotasikan μ_{RSS} dapat dilihat pada persamaan 4 dan varians dari penduga dapat dilihat pada persamaan 5.

$$\hat{\mu}_{RSS} = \frac{1}{h} \sum_{i=1}^h \frac{1}{c} \sum_{j=1}^c X_{[i]j} = \frac{1}{h} \sum_{i=1}^h \bar{X}_{[i]} \dots \dots \dots (4)$$

$$Var(\hat{\mu}_{RSS}) = Var(\bar{X}) - \frac{1}{h^2} \sum_{i=1}^h (\mu_{[i]} - \mu)^2 \quad (5)$$

Dimana $X_{[i]j}$ adalah variabel X setelah diurutkan, $\bar{X}_{[i]}$ adalah rata-rata X diantara siklus, dan μ adalah parameter rata-rata.

Selanjutnya akan dibuktikan jika $\hat{\mu}$ pada sampel himpunan terurut adalah penduga yang tidak bias untuk menduga μ dari populasi sebagaimana proses di bawah ini.

$$\begin{aligned} E(\hat{\mu}) &= E\left(\frac{1}{h} \sum_{i=1}^h \frac{1}{c} \sum_{j=1}^c X_{[i]j}\right) \\ &= \frac{1}{h} \sum_{i=1}^h \frac{1}{c} \sum_{j=1}^c E(X_{[i]j}) \\ &= \frac{1}{h} \sum_{i=1}^h E(X_{[i]}) \\ &= \frac{1}{h} \sum_{i=1}^h \left\{ \int_{-\infty}^{\infty} x \frac{h!}{(i-1)!(h-i)!} [F(x)]^{i-1} [1 - F(x)]^{h-i} f(x) dx \right\} \\ &= \frac{1}{h} \sum_{i=1}^h \left\{ \int_{-\infty}^{\infty} hx \frac{(h-1)!}{(i-1)!((h-1)-(i-1))!} [F(x)]^{i-1} [1 - F(x)]^{h-i} f(x) dx \right\} \\ &= \frac{1}{h} \sum_{i=1}^h \left\{ \int_{-\infty}^{\infty} hx \binom{h-1}{i-1} [F(x)]^{i-1} [1 - F(x)]^{h-i} f(x) dx \right\} \\ &= \int_{-\infty}^{\infty} xf(x) \left\{ \sum_{i=1}^h \binom{h-1}{i-1} [F(x)]^{i-1} [1 - F(x)]^{h-i} \right\} dx \end{aligned}$$

Karena diketahui dari distribusi binomial jika

$$\left\{ \sum_{i=1}^h \binom{h-1}{i-1} [F(x)]^{i-1} [1 - F(x)]^{h-i} \right\} = 1$$

Sehingga terbukti bahwa $\hat{\mu}$ merupakan penduga yang tidak bias dari μ atau

$$E(\hat{\mu}) = \int_{-\infty}^{\infty} xf(x) dx = \mu$$

Misra dan Gupta (2010) menyatakan bahwa terdapat kondisi dimana geometrik mean lebih sesuai digunakan dari pada *arithmetic mean*, contohnya adalah data dengan distribusi miring ke kanan atau distribusi dengan kemiringan positif. Geometrik mean secara spesifik dapat digunakan sebagai nilai tengah pada rata-rata rasio, persentase dan tingkat perubahan dalam satu periode dibandingkan dengan dengan periode lainnya. Menurut definisi secara matematik, geometrik mean adalah akar ke n dari produk perkalian n bilangan. Geometrik mean dari N unit dari populasi data dapat dihitung dengan persamaan berikut.

$$G = \sqrt[N]{X_1 \cdot X_2 \dots X_N} \dots\dots\dots(6)$$

Penduga dari rata-rata geometrik pada sampel himpunan terurut dapat dilihat pada persamaan 7.

$$\bar{X}_{GRSS} = \sqrt[c \cdot h]{\prod_{i=1}^c \prod_{j=1}^h X_{[j]i}} \dots\dots\dots(7)$$

Dimana \bar{X}_{GRSS} adalah penduga dari rata-rata geometrik pada sampel himpunan terurut, c adalah ukuran siklis dan h adalah ukuran himpunan yang diambil.

II. Metode Penelitian

Bukti empiris dari efisisensi penduga yang dibangun dilakukan dengan simulasi. Bias dan relatif efisiensi dari penduga diamati dan dilakukan pengamatan untuk data yang berdistribusi normal.

Proses simulasi dilakukan dengan menggunakan matlab. Pembangkitan data dan perhitungan dari rata-rata geometrik, rata-rata aritmatika, dan juga perhitungan standar deviasi untuk sampel acak sederhana maupun sampel

himpunan terurut dilakukan dengan Matlab.

Detil keterangan mengenai proses simulasi dijelaskan sebagai berikut:

1. Dibangkitkan 10.000 data bivariat dari data yang berdistribusi normal dengan koefisien korelasi $\rho = 0.5, 0.75, 0.9, 1$.
2. Setelah data diperoleh, data disimulasi 10.000 kali untuk metode sampel acak sederhana dan sampel himpunan terurut untuk setiap kombinasi, $n = c \cdot h$, dimana banyak siklis (c) = 2, 4, 10 dan banyak himpunan (h) = 3, 4, 5.
3. Setelah data sampel dari rata-rata geometrik didapat untuk setiap kombinasi siklis dan himpunan, standar deviasi untuk setiap rata-rata yang dihasilkan dihitung satu persatu.
4. Untuk menghitung Mean Square Error (MSE), maka digunakan persamaan 8.

$$MSE = bias^2 + variance \dots\dots\dots(8)$$

5. Setelah MSE dari setiap kombinasi siklis dan himpunan didapatkan, MSE digunakan untuk menghitung Efisiensi Relatif (RE) dengan persamaan 9.

$$RE_{GM} = \frac{MSE_{SRS_{GM}}}{MSE_{RSS_{GM}}} \dots\dots\dots(9)$$

III. Hasil dan Pembahasan

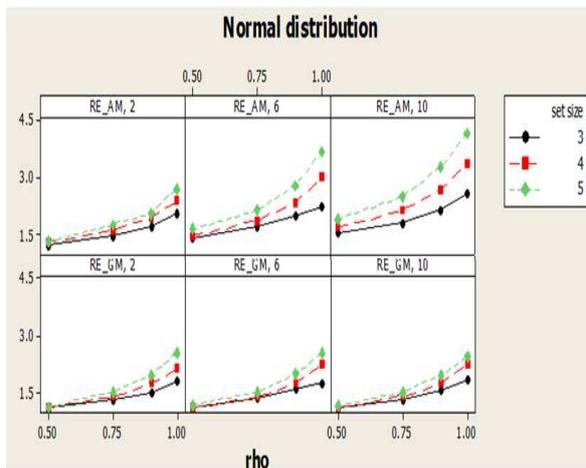
Hasil dari proses simulasi dihimpun pada tabel efisiensi relatif pada tabel 1 dan tabel bias pada tabel 2. Lalu nilai dari efisiensi relatif juga digambarkan pada gambar 2.

Pada tabel 1 mengenai hasil perhitungan efisien relatif, diberikan informasi hasil efisiensi relatif dari empat kelompok korelasi yaitu 0.5, 0.75, 0.9, dan 1. Dimana pada masing-masing kelompok korelasi tersebut dipisahkan juga dengan berbagai kategori menurut ukuran himpunan dan siklus himpunan nya. Siklus yang digunakan adalah 2, 6, dan 10 serta masing-masing siklus mempunyai ukuran himpunan sebesar 2, 3, dan 5.

Tabel 1. Nilai dari efisiensi relatif atau RE

cycle size	set size	ρ	AM	GM	ρ	AM	GM
2	3	0.5	1.218	1.141	0.9	1.690	1.549
2	4	0.5	1.278	1.149	0.9	1.917	1.774
2	5	0.5	1.290	1.167	0.9	2.049	1.981
6	3	0.5	1.402	1.171	0.9	1.994	1.636
6	4	0.5	1.455	1.133	0.9	2.341	1.769
6	5	0.5	1.624	1.206	0.9	2.807	2.028
10	3	0.5	1.529	1.139	0.9	2.123	1.585
10	4	0.5	1.673	1.132	0.9	2.690	1.783
10	5	0.5	1.888	1.189	0.9	3.277	1.997
2	3	0.75	1.430	1.329	1	2.039	1.851
2	4	0.75	1.606	1.427	1	2.383	2.150
2	5	0.75	1.753	1.552	1	2.708	2.566
6	3	0.75	1.696	1.408	1	2.254	1.799
6	4	0.75	1.857	1.422	1	3.012	2.263
6	5	0.75	2.127	1.552	1	3.657	2.571
10	3	0.75	1.791	1.327	1	2.569	1.868
10	4	0.75	2.125	1.431	1	3.396	2.293
10	5	0.75	2.495	1.538	1	4.179	2.476

Sedangkan distribusi normal dapat terlihat pada gambar di bawah ini.



Gambar 2. Grafik perbandingan efisiensi relatif

Tabel 2. Nilai bias dari hasil simulasi

ρ	c	h	Normal (20,16)	
			AM	GM
0.5	2	3	0.000686	0.029653
0.5	2	4	1.96E-06	1.23E-02
0.5	2	5	7.29E-06	0.008538
0.5	6	3	6.56E-05	2.30E-03
0.5	6	4	1.09E-05	1.84E-03
0.5	6	5	1E-06	0.000847
0.5	10	3	3.36E-05	0.001498
0.5	10	4	1.21E-06	4.58E-04
0.5	10	5	1.61E-04	6.08E-05
0.75	2	3	3.88E-04	2.27E-02
0.75	2	4	1.44E-06	0.009683
0.75	2	5	2.16E-04	7.96E-03
0.75	6	3	1.00E-06	1.73E-03
0.75	6	4	1.85E-05	5.71E-04
0.75	6	5	1.96E-06	4.20E-04
0.75	10	3	1.44E-06	0.000807
0.75	10	4	1.94E-05	5.24E-04
0.75	10	5	1.76E-05	4.41E-04
0.9	2	3	3.28E-04	1.69E-02
0.9	2	4	4.41E-06	0.007225
0.9	2	5	3.25E-05	4.50E-03
0.9	6	3	1.00E-06	0.00131
0.9	6	4	1.60E-07	6.97E-04
0.9	6	5	1.00E-06	3.72E-04
0.9	10	3	2.89E-06	7.56E-04
0.9	10	4	3.60E-07	0.000286
0.9	10	5	1.00E-08	2.13E-04
1	2	3	5.33E-05	1.26E-02
1	2	4	8.1E-07	0.004277
1	2	5	1.21E-06	2.13E-03
1	6	3	4.90E-05	1.99E-03
1	6	4	7.29E-06	3.42E-04
1	6	5	5.76E-06	4.49E-04
1	10	3	3.61E-06	0.000543
1	10	4	1.30E-04	5.06E-04
1	10	5	2.56E-06	9.22E-05

Pada tabel 1 dan gambar 2 terdapat informasi mengenai efisiensi relatif dari hasil simulasi data yang memiliki distribusi normal. Berdasarkan tabel dan gambar, terlihat bahwa setiap nilai dari efisiensi relatif adalah besar dari 1. Dimana untuk nilai korelasi antara variabel Populasi dan variabel pendamping sebesar 0.5 maka hasil maksimal efisien relatif adalah 1.888 untuk penduga rataan aritmatik dan 1.206 untuk penduga rataan geometrik nya. Lalu untuk nilai korelasi antara variabel.

Populasi dan variabel pendamping sebesar 0.75 maka hasil maksimal efisien relatif adalah 2.495 untuk penduga rataan aritmatik dan 1.552 untuk penduga rataan geometrik nya. Serta Untuk nilai korelasi antara variabel. Populasi dan variabel pendamping sebesar 0.9 maka hasil maksimal efisien relatif adalah 3.277 untuk penduga rataan aritmatik dan 2.028 untuk penduga rataan geometriknya.

Selanjutnya untuk nilai korelasi antara variabel Populasi dan variabel pendamping sebesar 1 maka hasil maksimal efisien relatif adalah 4.179 untuk penduga rataan aritmatik dan 2.571 untuk penduga rataan geometrik nya. Ini mengindikasikan bahwa penduga dengan sampel himpunan terurut lebih efisien dari penduga dengan menggunakan sampel acak sederhana.

Pada gambar 2 juga terlihat bahwa ketika nilai dari koefisien korelasi ditingkatkan maka nilai dari efisiensi relatif juga meningkat sehingga dapat dikatakan koefisien korelasi mempengaruhi nilai dari efisiensi relatif. Dari gambar juga dapat diindikasikan bahwa ukuran himpunan mempengaruhi efisiensi relatif, artinya semakin besar ukuran himpunan maka semakin besar juga efisiensi relatifnya.

Pada tabel 2 terdapat nilai-nilai bias dari hasil proses simulasi. Dari tabel 2 tersebut dapat dilihat jika nilai bias untuk setiap kombinasi sangat kecil sekali baik untuk rataan aritmatika maupun untuk rataan geometriknya. Ini menunjukkan bahwa penduga yang digunakan efektif dalam pendugaan parameter populasi.

IV. Penutup

Pada penelitian ini penduga yang dibangun untuk rataan geometrik pada sampel himpunan terurut ditunjukkan dengan persamaan 4 dan telah dilakukan evaluasi terhadap penduga tersebut dengan menggunakan simulasi komputer untuk data yang berdistribusi normal. Berdasarkan dari hasil simulasi, pendugaan rataan geometrik dengan menggunakan sampel himpunan terurut memberikan efisiensi yang lebih baik dibandingkan dengan sampel acak sederhana untuk data dari populasi yang berdistribusi normal. Lalu dapat juga disimpulkan bahwa efisiensi relatif akan meningkat jika nilai korelasi antara variabel populasi dan variabel pendamping ditingkatkan. Ini berarti korelasi secara signifikan mempengaruhi hasil efisiensi relatif. Penduga rataan geometrik pada sampel himpunan terurut dapat disimpulkan sebagai penduga yang tidak bias untuk data populasi yang menyebar pada distribusi normal.

Referensi

- Chen, Z., Bai, Z., & Sinha, B. K. (2003). *Ranked set sampling, theory and applications*. New York: Springer.
- Cochran, W. G. (1977). *Sampling techniques*. Massachusetts: John Wiley & Son.
- Dell, T. R., Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 28, 545-555.
- Misra, S., & Gupta, R. K. (2010). Estimation of geometric mean of skewed populations with and without using auxiliary information. *Journal of Probability and Statistical Studies*, 3(2), 35-46.
- Scheaffer, R. L., Mendenhall. W., & Ott. L. (1990). *Elementary survey sampling*. California: Duxbury Press.
- Spizman, L. (2008). A note on utilizing the geometric mean: when, why and how the forensic economist should employ the geometric mean. *Journal of Legal Economics*, 15(1), 43-55.

Wolfe, D. N. (2012). Ranked set sampling; its relevance and impact on statistical inference. *International Scholarly Research Network, 2012*(2012). Retrieved April 14, 2014, from <http://www.hindawi.com/journals/isrn/2012/568385/>.

