

ESTIMASI VARIANS DENGAN PENDEKATAN METODE *RESCALED BOOTSTRAP*

Sapta Hastho Ponco^{1*}, Septiadi Padmadisastra², Gatot Riwi Setyanto²

Mahasiswa Program Magister Statistika Terapan, Universitas Padjadjaran, Bandung, Indonesia^{1*}
sapta13002@student.unpad.ac.id, sapta.hastho@gmail.com
Departemen Statistika, Universitas Padjadjaran, Bandung, Indonesia²

Abstrak

Untuk meningkatkan kualitas data hasil survei, Badan Pusat Statistik (BPS), khususnya Direktorat Statistik Kesejahteraan Rakyat, melakukan kegiatan pengembangan Susenas dengan kajian utama pada aspek kecukupan sampel rumah tangga yang digunakan. Kendala terbesar yang dihadapi dalam mengkaji kecukupan sampel rumah tangga survei berskala besar seperti Susenas adalah dibutuhkannya anggaran yang besar. Untuk mengatasi kendala tersebut, penggunaan metode resampling berupa metode bootstrap dapat disarankan. Penggunaan metode bootstrap standard untuk sampel yang independen dan berdistribusi identik (iid) ternyata menghasilkan ketidakkonsistenan dalam kasus ukuran sampel yang dibatasi pada strata. Oleh karena itu, bentuk modifikasi bootstrap berupa metode *rescaled bootstrap* lebih tepat diaplikasikan karena faktor juga diterapkan pada bobot survei sehingga penghitungan estimasi varians dapat berlaku untuk fungsi yang *smooth and non-smooth statistics*. Pada penelitian ini, kecukupan sampel rumah tangga Susenas dievaluasi dengan metode *rescaled bootstrap* melalui beberapa replikasi pada estimasi rata-rata lama sekolah dan kemudian hasilnya dibandingkan dengan hasil Susenas. Secara umum metode *rescaled bootstrap* dengan jumlah replikasi sebanyak 1000 kali menunjukkan hasil bahwa dengan sebanyak 8 sampel rumah tangga per blok sensus, presisi dan estimasi yang dihasilkan pada tiap strata ternyata hampir sama dengan hasil Susenas.

Kata kunci: Strata, Resampling, Bobot *Bootstrap*, *Bootstrap*, *Rescaled Bootstrap*, Susenas

1. Pendahuluan

Badan Pusat Statistik (BPS), berdasarkan Undang-Undang No. 16 tahun 1997 tentang Statistik dinyatakan sebagai lembaga pemerintah non departemen yang berhak menyelenggarakan kegiatan statistik dasar, memiliki kewenangan dalam melaksanakan kegiatan sensus dan survei dan bertanggung jawab atas tersedianya data yang diperlukan untuk menunjang perencanaan pembangunan sekaligus sebagai alat untuk memantau dan mengevaluasi keberhasilan program-program pembangunan yang telah dilaksanakan.

Salah satu sumber data yang diperlukan untuk perencanaan pembangunan khususnya di bidang sosial ekonomi penduduk adalah Survei Sosial Ekonomi Nasional (Susenas). Susenas merupakan survei berbasis rumah tangga yang dilaksanakan oleh BPS setiap tahun dan dirancang untuk mengumpulkan data sosial kependudukan yang relatif sangat luas. Selain itu, Susenas juga menjadi salah satu produk BPS yang selalu ditunggu oleh para pengguna data karena menghasilkan beberapa informasi penting terutama terkait program prioritas pemerintah seperti data kemiskinan, data konsumsi/ pengeluaran rumah tangga dan Indeks Pembangunan Manusia (IPM).

Metode penarikan sampel Susenas termasuk dalam survei kompleks, yakni penarikan sampel tiga tahap berstrata (*three-stage stratified sampling*) dimana unit sampling yang digunakan berupa kumpulan rumah tangga yang secara geografis berkumpul dalam satu klaster berupa blok sensus. Akibatnya, rumah tangga di blok sensus yang sama umumnya cenderung memiliki kesamaan sifat dan karakteristik (seperti pendapatan, tingkat pendidikan, pekerjaan, dan lain-lain). Pembentukan blok sensus bertujuan untuk mereduksi biaya pengumpulan data yang besar, tetapi korelasi antar unit sampling dalam blok sensus yang sama akan menaikkan varians sampel (menurunkan tingkat presisi) hasil survei (Yansaneh, 2005).

Untuk meningkatkan hasil kualitas data dan sejalan dengan peningkatan frekuensi permintaan data konsumsi/pengeluaran rumah tangga dan penghitungan data kemiskinan, Direktorat Statistik Kesejahteraan Rakyat selaku penanggung jawab terhadap pelaksanaan kegiatan Susenas melakukan kegiatan pengembangan Susenas. Kegiatan ini mencakup beberapa aspek, antara lain pengembangan instrumen pengumpulan data, organisasi dan prosedur operasional baku pendataan lapangan, dan jumlah sampel yang memadai untuk tingkatan estimasi wilayah tertentu.

Aspek utama yang menjadi perhatian adalah kecukupan sampel rumah tangga sebagai unit sampling. Sebelum tahun 2011, unit sampling yang digunakan oleh Susenas untuk setiap blok sensus adalah sebanyak 16 sampel rumah tangga kemudian sejak tahun 2011 terjadi perubahan jumlah unit

sampling yang cukup signifikan, yaitu menjadi 10 sampel rumah tangga untuk setiap blok sensus. Penurunan jumlah sampel rumah tangga tetap diiringi dengan penambahan jumlah sampel blok sensus sehingga jumlah sampel rumah tangga untuk tahun 2010 sampai tahun 2012 secara nasional tetap berjumlah sekitar 300.000 sampel rumah tangga.

Kajian mengenai kecukupan unit sampling pada survei berskala besar terkendala pada anggaran. Martinez, dkk (2011) mengatakan bahwa kendala tersebut dapat diatasi dengan memanfaatkan perkembangan teknik komputasi statistik yakni menggunakan metode resampling. Prosedur metode resampling dilakukan dengan cara mengambil sampel secara berulang dari data yang telah tersedia dengan menggunakan bantuan komputer. Cara pengambilan sampel dapat dilakukan dengan pengembalian (*with replacement*) atau tanpa pengembalian (*without replacement*) dan data sampel baru yang terbentuk bisa berukuran sama atau berbeda dari ukuran data sampel asli. Metode resampling memperlakukan sampel seolah-olah sebagai populasi terbatas (*finite population*) dan kemudian mengambil sampel baru secara berulang dari populasi itu (Lohr, 2010).

Salah satu metode resampling yang umum digunakan dalam penelitian adalah metode bootstrap. Metode bootstrap merupakan metode berbasis komputer-intensif dan diperkenalkan oleh Bradley Efron pada tahun 1979 guna mengestimasi kesalahan baku estimator dan interval kepercayaan suatu parameter dari distribusi yang tidak diketahui. Sebagai pengganti, metode bootstrap menggunakan distribusi empiris data sampel dan kemudian mengambil sampel baru berdasarkan distribusi empiris tersebut secara berulang kali dengan pengembalian yang selanjutnya digunakan untuk mencari estimator bootstrap. Metode bootstrap sangat potensial untuk diterapkan pada penanganan masalah keakuratan data dan untuk keperluan inferensia statistik (Efron dan Tibshirani, 1993).

Pada survei kompleks, penggunaan metode bootstrap umumnya dibatasi pada desain penarikan sampel satu tahap atau multistage dimana tahap pertama unit sampling dilakukan dengan pengembalian atau fraksi sampel pada tahap pertama bernilai kecil di sebagian strata (Preston, 2009). Sampai saat ini penggunaan metode bootstrap pada survei kompleks berkembang cukup pesat dan aplikasinya telah meluas dalam berbagai kondisi data sehingga muncul beberapa bentuk modifikasi.

Funaoka, dkk (2006) melakukan penelitian tentang formasi harga komoditas utama pada Survei Nasional Data Harga di Jepang tahun 1997 untuk mengestimasi rasio dan kuantil dengan menggunakan metode *Bernoulli Bootstrap* yang diaplikasikan pada desain penarikan sampel tiga tahap berstrata dimana unit sampling dipilih secara acak sederhana tanpa pengembalian pada setiap tahap dengan jumlah replikasi yang digunakan sebanyak 10.000 kali.

Preston (2009) melakukan studi simulasi *Rescaled Bootstrap* untuk melihat kinerja dari *Multistage Rescaled Bootstrap Variance Estimator* (MRBE) pada desain penarikan sampel dua tahap berstrata dimana unit sampling dipilih secara acak sederhana dari setiap 10 populasi amatan dengan jumlah replikasi sebanyak 100 kali. Selanjutnya hasil MRBE dibandingkan dengan *Single-stage Rescaled Bootstrap Variance Estimator* (SRBE) dan *Bernoulli Bootstrap Estimator* (BBE).

Saigo (2010) melakukan studi mengenai perbandingan 4 metode bootstrap yaitu *Bernoulli Bootstrap*, *Rescaling Bootstrap*, *Mirror-Match Bootstrap* dan *Without-Replacement Bootstrap* pada desain penarikan sampel tiga tahap berstrata dengan acak sederhana tanpa pengembalian dengan menggunakan jumlah replikasi sebanyak 1000 kali.

2. Metode Bootstrap

Metode bootstrap merupakan salah satu metode resampling dimana dari sekelompok data sampel terpilih dilakukan pengambilan kembali sejumlah sampel secara berulang kali (replikasi) dengan pengembalian untuk memperoleh suatu kesimpulan. Metode bootstrap diperkenalkan oleh Bradley Efron pada tahun 1979 salah satunya bertujuan untuk mengestimasi kesalahan baku estimator dalam situasi nonparametrik. Efron dan Tibshirani (1993) mengatakan bahwa metode bootstrap sangat potensial untuk diterapkan pada penanganan masalah keakuratan data dan untuk keperluan inferensia statistik. Davison dan Hinkley (1997) menambahkan bahwa metode bootstrap juga dapat digunakan untuk mengatasi situasi dimana asumsi standard tidak terpenuhi, seperti ukuran sampel kecil dan data tidak berdistribusi normal.

2.1 Prinsip Penggantian (Plug-in)

Metode plug-in merupakan metode sederhana yang digunakan untuk mengestimasi parameter dari data sampel. Prinsip dari metode plug-in adalah menggantikan distribusi populasi yang tidak diketahui dengan distribusi empiris berdasarkan data sampel. Misalnya dimiliki sampel acak berukuran n dari suatu distribusi peluang F , atau dinyatakan $F \rightarrow (x_1, x_2, \dots, x_n)$, maka fungsi distribusi empiris \hat{F} adalah daftar nilai-nilai yang diperoleh dari sampel $\mathbf{x} = (x_1, x_2, \dots, x_n)$ dengan peluang $1/n$ untuk setiap x_i , $i = 1, 2, \dots, n$ (Efron dan Tibshirani, 1993).

Parameter sebagai fungsi distribusi peluang F dituliskan $\theta = t(F)$. Notasi ini menekankan bahwa nilai parameter θ yang diamati didekati dengan prosedur numerik $t(\cdot)$ untuk distribusi F . Metode plug-in untuk mengestimasi parameter $\theta = t(F)$ adalah $\hat{\theta} = t(\hat{F})$. Dengan kata lain, fungsi $\theta = t(F)$ dari distribusi peluang F diestimasi oleh fungsi yang sama dari distribusi empiris \hat{F} sebagai $\hat{\theta} = t(\hat{F})$. Sebagai contoh,

jika F adalah distribusi peluang dari kondisi yang sebenarnya, maka estimasi plug-in untuk ekspektasi parameter $\theta = t(F) = E_F(x)$ adalah $\hat{\theta} = t(\hat{F}) = E_{\hat{F}}(x)$.

2.2 Pembentukan Sampel Bootstrap

Pada keadaan riil, F adalah distribusi peluang yang tidak diketahui dimana sekumpulan data berupa $\mathbf{x} = (x_1, x_2, \dots, x_n)$ diambil secara acak dan dihitung statistik $\hat{\theta} = s(\mathbf{x})$. Sementara pada bootstrap, distribusi F yang tidak diketahui digantikan dengan distribusi empiris \hat{F} dan parameternya didefinisikan melalui fungsi distribusi empiris tersebut. Selanjutnya dari sekumpulan data sampel $\mathbf{x} = (x_1, x_2, \dots, x_n)$ yang dianggap independen dan terdistribusi secara identik (iid) diambil kembali sampel acak secara berulang kali dengan pengembalian dan dituliskan $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$, atau dinyatakan sebagai $\hat{F} \rightarrow (x_1^*, x_2^*, \dots, x_n^*)$, dan diperoleh $\hat{\theta}^* = s(\mathbf{x}^*)$. Sekumpulan data sampel $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ ini disebut sebagai sampel bootstrap dan notasi bintang menunjukkan bahwa \mathbf{x}^* bukan sebagai data yang sebenarnya, namun berupa sekumpulan data yang telah mengalami penyampelan ulang (Efron dan Tibshirani, 1993).

Bila dimiliki sampel acak berukuran n , maka akan diperoleh kemungkinan sampel bootstrap sebanyak n^n dan dari tiap sampel bootstrap diperoleh $\hat{\theta}^*$. Seluruh kemungkinan sampel ini disebut dengan jumlah sampel ideal bootstrap.

2.3 Estimasi Kesalahan Baku Bootstrap

Efron dan Tibshirani (1993) mengemukakan bahwa metode bootstrap digunakan untuk mengestimasi kesalahan baku $\hat{\theta}$ sebagai ukuran sederhana dalam mengukur akurasi suatu statistik. Estimasi bootstrap untuk kesalahan baku $\hat{\theta}$, $se_{\hat{F}}(\hat{\theta})$, diperoleh melalui prinsip plug-in yang berdasarkan distribusi empiris \hat{F} untuk menggantikan distribusi peluang F yang tidak diketahui. Secara spesifik, estimasi bootstrap untuk $se_{\hat{F}}(\hat{\theta})$ didefinisikan menjadi:

$$se_{\hat{F}}(\hat{\theta}^*) = \sqrt{[E_{\hat{F}}\{(\hat{\theta}^* - E_{\hat{F}}(\hat{\theta}^*))^2\]}}$$

$se_{\hat{F}}(\hat{\theta}^*)$ disebut sebagai estimasi ideal bootstrap untuk kesalahan baku $\hat{\theta}$.

Perhitungan $\hat{\theta}^*$ atas semua kemungkinan sampel bootstrap memerlukan waktu yang cukup lama sehingga untuk memudahkan penyelesaian penghitungan estimasi kesalahan baku bootstrap digunakan pendekatan simulasi Monte Carlo dengan bantuan algoritma. Simulasi Monte Carlo menggunakan jumlah replikasi yang cukup besar tetapi jauh lebih kecil dibandingkan dengan jumlah sampel bootstrap ideal ($n \leq B \leq n^n$). Algoritma untuk mengestimasi kesalahan baku $\hat{\theta}$ adalah sebagai berikut (Efron dan Tibshirani, 1993):

1. Ambil sampel bootstrap secara independen dengan replikasi sebanyak B kali $\mathbf{x}^{*1}, \mathbf{x}^{*2}, \dots, \mathbf{x}^{*B}$, dimana tiap sampel yang diambil dilakukan pengembalian.
2. Evaluasi replikasi bootstrap yang bersesuaian pada setiap sampel bootstrap $\hat{\theta}^*(b) = s(\mathbf{x}^{*b})$ dimana $b = 1, 2, \dots, B$
3. Hitung estimasi kesalahan baku $se_{\hat{F}}(\hat{\theta})$ dengan mereplikasi standard deviasi empiris.

$$se_{\hat{F}} = \left\{ \sum_{b=1}^B \frac{[\hat{\theta}^*(b) - \hat{\theta}^*(.)]^2}{B-1} \right\}^{1/2}$$

dimana, $\hat{\theta}^*(.) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^*(b)$

Efron dan Tibshirani (1993) menyebutkan ada beberapa hal yang perlu diperhatikan berkaitan dengan jumlah replikasi pada bootstrap, yaitu:

1. Meskipun jumlah replikasi bootstrap kecil, misal $B = 25$ kali, biasanya sudah cukup untuk memberikan informasi tetapi dengan $B = 50$ kali sudah sangat cukup untuk memberikan estimasi $se_{\hat{F}}(\hat{\theta})$ yang akurat.
2. Jumlah replikasi bootstrap yang besar, misalnya $B = 200$ kali, biasanya tidak perlu dilakukan dalam mengestimasi kesalahan baku (jumlah replikasi bootstrap yang besar diperlukan dalam interval kepercayaan bootstrap).

Schmidheiny (2008) menyebut biasanya dengan jumlah replikasi sebanyak 100 kali dianggap sudah cukup. Sedangkan The Eurosystem Household Finance and Consumption Survey (2013) menggunakan jumlah replikasi paling sedikit 1000 kali untuk mendapatkan estimator bootstrap.

Limit $se_{\hat{F}}$ untuk B menuju tak hingga adalah estimasi ideal bootstrap untuk $se_{\hat{F}}(\hat{\theta})$, atau dinyatakan:

$$\lim_{B \rightarrow \infty} se_{\hat{F}} = se_{\hat{F}}(\hat{\theta}^*)$$

Fakta bahwa nilai $se_{\hat{F}}$ mendekati $se_{\hat{F}}(\hat{\theta}^*)$ dengan $B \rightarrow \infty$ mengindikasikan nilai standard deviasi empiris mendekati standard deviasi populasi dengan jumlah replikasi yang sangat besar. Populasi yang dimaksud disini adalah populasi dari nilai $\hat{\theta}^* = s(\mathbf{x}^*)$ dimana $\hat{F} \rightarrow (x_1^*, x_2^*, \dots, x_n^*) = \mathbf{x}^*$.

Estimasi ideal bootstrap untuk $se_{\hat{F}}(\hat{\theta}^*)$ dan pendekatannya, $se_{\hat{F}}$ sering juga disebut sebagai estimasi nonparametrik bootstrap karena estimasinya berdasarkan distribusi \hat{F} (Efron dan Tibshirani, 1993).

3. Metode *Rescaled Bootstrap* Pada Tahap Ketiga Susenas Tahun 2012

Tahap ketiga dari metode penarikan sampel Susenas adalah memilih sebanyak 10 sampel rumah tangga biasa secara sistematis berdasarkan hasil pemutakhiran rumah tangga pada blok sensus terpilih. Jumlah sampel rumah tangga yang digunakan oleh Susenas tersebut akan optimal jika estimasi dari parameter yang dihasilkan memiliki variansi sampling yang kecil.

Dengan mengikuti cara berpikir seperti Rao, Wu dan Yue (1992) dalam Preston (2009), maka dalam penelitian ini akan diterapkan metode resampling *Rescaled Bootstrap* pada tahap ketiga dari metode penarikan sampel Susenas tahun 2012 sebagai pendekatan untuk melihat kecukupan sampel rumah tangga yang digunakan. Efisiensi dari kecukupan sampel rumah tangga selanjutnya dilihat dengan cara membandingkan estimasi parameter yang dihasilkan dari data Susenas tahun 2012 dengan data hasil resampling.

Secara umum, alur/tahapan dari penelitian ini tetap berdasar pada metode penarikan sampel tiga tahap berstrata yang sama seperti metode penarikan sampel Susenas tahun 2012 lakukan. Hanya saja pada tahap ketiga dari metode penarikan sampel Susenas tersebut akan diaplikasikan metode resampling berupa metode *Rescaled Bootstrap* dimana rumah tangga yang dijadikan sebagai unit sampling dipilih secara acak dengan ukuran yang sama dari setiap sampel blok sensus ke-i strata h yang dianggap sebagai populasi (*finite population*). Strata yang digunakan dibedakan menurut blok sensus (BS) dan tipe daerah (perkotaan $h=1$, perdesaan $h=2$) di Kabupaten Bogor.

Misalkan B_{hi} adalah BS ke-i pada daerah h yang berisi muatan sekitar 10 rumah tangga dan akan diambil sampel rumah tangga berukuran n_{hi} . Prosedurnya dijelaskan sebagai berikut:

- 1) Secara independen pada setiap BS ke-i daerah h, ambil sampel bootstrap rumah tangga $n_{hi}^*(b) = n_{hi} - 1$ secara acak dengan pengembalian.
- 2) Hitung bobot bootstrap dengan mengikuti formula berikut ini:

$$w_{hij}^*(b) = w_{hij} \left[\left(1 - \sqrt{\frac{n_{hi}(b)}{n_{hi}-1}} \right) + \sqrt{\frac{n_{hi}(b)}{n_{hi}-1}} \left(\frac{n_{hi}}{n_{hi}^*(b)} \right) m_{hij}^*(b) \right]$$

dimana, $w_{hij}^*(b)$ adalah bobot bootstrap untuk rumah tangga ke-j di BS ke-i pada daerah h, w_{hij} bobot sampel original untuk rumah tangga ke-j di BS ke-i pada daerah h, $m_{hij}^*(b)$ banyaknya rumah tangga ke-j di BS ke-i daerah h yang terpilih sebagai sampel bootstrap dan $n_{hi}^*(b)$ adalah banyaknya sampel rumah tangga di BS ke-i daerah h.

Untuk menentukan besaran $n_{hi}^*(b)$, Girard (2009) mengemukakan bahwa:

- a. Jika $n_{hi}^*(b) > n_{hi} - 1$, maka kemungkinan bobot bootstrap bernilai negatif ($w_{hij}^*(b) < 0$).
- b. Jika $n_{hi}^*(b) \leq n_{hi} - 1$, maka bobot bootstrap akan selalu bernilai positif ($w_{hij}^*(b) > 0$) tetapi $n_{hi}^*(b) < n_{hi} - 1$ tidak terlalu memberikan hasil yang signifikan dalam simulasi (malah membutuhkan n_{hi} yang besar).

Oleh karena itu, untuk kemudahan di dalam penghitungan dan menjaga bobot bootstrap agar bernilai positif, maka digunakan $n_{hi}^*(b) = n_{hi} - 1$ sehingga persamaan di atas jika diuraikan akan menjadi seperti persamaan di bawah ini:

$$w_{hij}^*(b) = w_{hij} \left[\left(\frac{n_{hi}}{n_{hi}-1} \right) m_{hij}^*(b) \right]$$

- 3) Hitung estimasi total karakteristik rumah tangga Y dengan metode bootstrap, yaitu:

$$\hat{Y}^*(b) = \sum_{h=1}^H \sum_{i=1}^{B_h} \sum_{j=1}^{n_{hi}} w_{hij}^*(b) y_{hij}$$

Estimator bootstrap $\hat{\theta}$ merupakan fungsi dari hasil penghitungan estimasi total karakteristik rumah tangga Y tersebut, atau diekspresikan dengan: $\hat{\theta} = f(\hat{Y}^*)$

- 4) Secara independen ulangi langkah-langkah di atas dengan jumlah replikasi sebanyak 500 kali ($B = 500$) lalu hitung estimasi bootstrap:

$$\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, \dots, \hat{\theta}^{(500)}$$

- 5) Terakhir setelah diperoleh estimasi bootstrap, lakukan penghitungan estimasi variansi bootstrap dengan formula:

$$v(\hat{\theta}) = \frac{1}{500-1} \sum_{b=1}^{500} (\hat{\theta}^*(b) - \hat{\theta}^*(\cdot))^2$$

dimana $\hat{\theta}^*(\cdot) = \frac{1}{500} \sum_{b=1}^{500} \hat{\theta}^*(b)$

4. Pembahasan Hasil

4.1 Data Susenas tahun 2012

Sumber data yang digunakan berupa data Susenas pada tahun 2012 dengan cakupan di Kabupaten Bogor dengan maksud untuk memperoleh estimasi rata-rata lama sekolah.

Rata-rata lama sekolah merupakan indikator yang menggambarkan lamanya tingkat pendidikan yang ditempuh oleh penduduk usia 15 tahun ke atas dan diformulasikan oleh UNDP (*United Nations Development Programme*) pada tahun 1990 guna penyusunan Indeks Pembangunan Manusia sebagai

ukuran baku keterbandingan capaian hasil pembangunan manusia antar wilayah/antar negara. Rata-rata lama sekolah didefinisikan sebagai indikator yang menunjukkan rata-rata jumlah tahun efektif yang dicapai oleh penduduk untuk bersekolah. Jumlah tahun efektif adalah jumlah tahun baku yang harus dijalani untuk menamatkan suatu jenjang pendidikan.

Tabel 4.1 Rata-Rata Lama Sekolah di Kabupaten Bogor, 2012

	Estimasi	Kesalahan baku	Interval Kepercayaan 95%		Koefisien Variasi	Relative Standard Error
			Batas atas	Batas bawah		
Kota	8.15	0.236	7.696	8.622	0.029	2.9
Desa	5.59	0.345	4.909	6.263	0.062	6.2

Dari tabel di atas tampak bahwa estimasi rata-rata lama sekolah di Kabupaten Bogor untuk daerah perkotaan lebih tinggi dibandingkan dengan daerah perdesaan. Rata-rata lama sekolah penduduk di perkotaan sebesar 8.15 dan di perdesaan sebesar 5.59. Hal ini berarti secara rata-rata penduduk berusia 15 ke atas daerah perkotaan baru mencapai jenjang pendidikan kelas 2 SMP (kelas VIII) atau putus sekolah di kelas 3 SMP (kelas IX) sedangkan di perdesaan secara rata-rata baru menamatkan pendidikan dasar (kelas 6 SD).

Besaran *sampling error* dari suatu estimasi diukur melalui kesalahan baku. Semakin rendah nilai kesalahan baku, maka semakin tinggi tingkat presisi yang dihasilkan. Tampak dari tabel di atas, kesalahan baku untuk estimasi rata-rata lama sekolah daerah perkotaan dan perdesaan, masing-masing sebesar 0.236 dan 0.345.

Kesalahan baku dapat digunakan untuk menghitung interval kepercayaan guna melihat interval dari angka sebenarnya yang dapat menggambarkan populasi. Dengan tingkat kepercayaan 95 persen, dapat disajikan interval kepercayaan sebagai estimasi interval (*interval estimation*) dengan batas bawah sebesar nilai estimasi dikurangi dua kesalahan baku dan batas atas sebesar nilai estimasi ditambah dua kesalahan baku.

4.2 Resampling

Untuk melihat kecukupan unit sampling yang digunakan oleh Susenas, berikut ini dilakukan metode resampling dengan cara mengambil sampel dari data sampel rumah tangga Susenas tahun 2012 per BS (n_{RT}) pada kedua strata (perkotaan dan perdesaan) berturut-turut sebanyak 6, 7, 8, dan 9 unit dengan menggunakan *software* R secara acak pengembalian dengan jumlah replikasi sebanyak 500 kali dan 1000 kali.

Kecukupan unit sampling selanjutnya dilihat dengan membandingkan estimasi rata-rata lama sekolah hasil Susenas dengan hasil resampling.

4.2.1 Replikasi 500 Kali (B = 500 kali)

Tabel 4.2 Estimasi Rata-Rata Lama Sekolah di Kabupaten Bogor daerah Perkotaan, 2012

Banyak rumah tangga per BS	Estimasi	Kesalahan baku	Interval Kepercayaan 95%		Koefisien Variasi	Relative Standard Error
			Batas atas	Batas bawah		
$n_{RT} = 6$	8.51	0.267	7.979	9.038	0.031	3.1
$n_{RT} = 7$	8.08	0.255	7.577	8.587	0.032	3.2
$n_{RT} = 8$	8.34	0.263	7.818	8.859	0.032	3.2
$n_{RT} = 9$	8.03	0.245	7.543	8.517	0.031	3.1

Tabel 4.3 Estimasi Rata-Rata Lama Sekolah di Kabupaten Bogor daerah Perdesaan, 2012

Banyak rumah tangga per BS	Estimasi	Kesalahan baku	Interval Kepercayaan 95%		Koefisien Variasi	Relative Standard Error
			Batas atas	Batas bawah		
$n_{RT} = 6$	6.02	0.427	5.134	6.900	0.071	7.1
$n_{RT} = 7$	5.52	0.415	4.666	6.381	0.075	7.5
$n_{RT} = 8$	5.61	0.374	4.839	6.388	0.067	6.7
$n_{RT} = 9$	5.74	0.358	4.999	6.482	0.062	6.2

4.2.2 Replikasi 1000 Kali (B = 1000 Kali)

Berikut ini resampling dilakukan sebanyak 10000 kali. Semakin banyak jumlah replikasi yang digunakan, maka akan semakin tepat estimasi yang dihasilkan. Hasilnya dapat dilihat pada tabel di bawah ini.

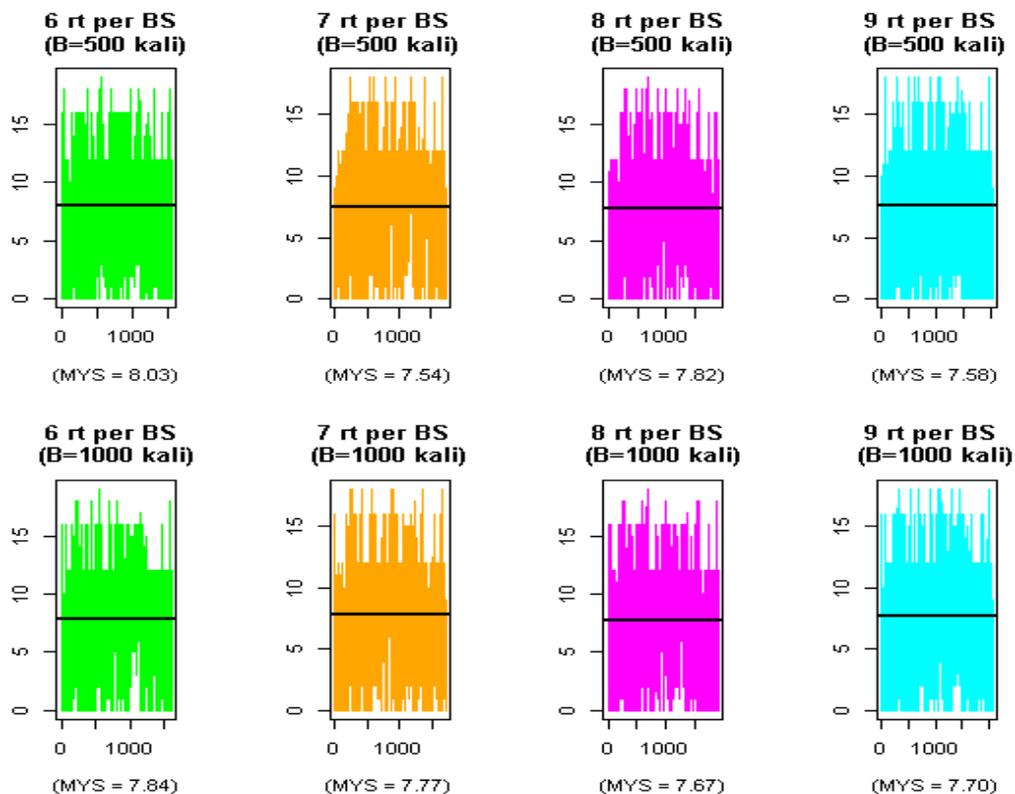
Tabel 4.4 Estimasi Rata-Rata Lama Sekolah di Kabupaten Bogor daerah Perkotaan, 2012

Banyak rumah tangga per BS	Estimasi	Kesalahan baku	Interval Kepercayaan 95%		Koefisien Variasi	Relative Standard Error
			Batas atas	Batas bawah		
$n_{rt} = 6$	8.36	0.262	7.848	8.888	0.031	3.1
$n_{rt} = 7$	8.26	0.240	7.781	8.734	0.029	2.9
$n_{rt} = 8$	8.18	0.247	7.689	8.671	0.030	3.0
$n_{rt} = 9$	8.21	0.243	7.726	8.690	0.030	3.0

Tabel 4.5 Estimasi Rata-Rata Lama Sekolah di Kabupaten Bogor daerah Perdesaan, 2012

Banyak rumah tangga per BS	Estimasi	Kesalahan baku	Interval Kepercayaan 95%		Koefisien Variasi	Relative Standard Error
			Batas atas	Batas bawah		
$n_{rt} = 6$	5.68	0.376	4.704	2.169	0.066	6.6
$n_{rt} = 7$	5.86	0.430	4.966	6.744	0.073	7.3
$n_{rt} = 8$	5.56	0.379	4.774	6.343	0.068	6.8
$n_{rt} = 9$	5.45	0.384	4.655	6.246	0.070	7.0

Plot estimasi rata-rata lama sekolah hasil resampling untuk total strata di Kabupaten Bogor ditampilkan pada gambar di bawah ini (replikasi 500 kali dan 1000 kali):



Gambar 4.1 Estimasi Rata-Rata Lama Sekolah di Kabupaten Bogor daerah Perkotaan dan Perdesaan, 2012

5. Kesimpulan

Berdasarkan hasil resampling di atas, secara umum resampling dengan menggunakan jumlah replikasi sebanyak 1000 kali memberikan hasil yang lebih baik dibandingkan dengan jumlah replikasi sebanyak 500 kali. Hal ini dapat dilihat dari ukuran efisiensi (kesalahan baku, koefisien variasi dan *Relative Standard Error*) yang dihasilkan.

Jika dilihat dari resampling data pada jumlah replikasi sebanyak 1000 kali, ternyata 8 sampel rumah tangga per blok sensus (BS) memiliki nilai estimasi yang hampir sama dengan estimasi rata-rata lama sekolah hasil Susenas 2012 baik daerah perkotaan maupun perdesaan.

6. Daftar Pustaka

- [1] BPS. 2012. *Pedoman Kepala BPS Prov dan Kab/kota Tahun 2012*. Jakarta: Badan Pusat Statistik.
- [2] Clark, R.G. 2007. Sampling Within Households in Household Surveys. *Journal of the Royal Statistical Society Series A*, 170, 63-82.
- [3] Davison, A.C. dan Hinkley, D.V. 1997. *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.
- [4] Efron, B. dan Tibshirani, R.J. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- [5] Funaoka, F., dkk. 2006. Bernoulli Bootstrap for Stratified Multistage Sampling. *Survey Methodology*, 32, 151–156.
- [6] Girard, Claude. 2009. The Rao-Wu Rescaling Bootstrap: From Theory to Practice. *Federal Committee on Statistical Methodology Research Conference*.
- [7] Groves, R.M., dkk. 2004. *Survey Methodology*. New York: John Wiley & Sons.
- [8] Jen, H.T., dkk. 2010. An Estimation of the Design Effect for the Two-stage Stratified Cluster Sampling Design. *International Research Conference (IRC-2010)*.
- [9] Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley & Sons.
- [10] Lohr, Sharon L. 2010. *Sampling: Design and Analysis*. Second edition. California: Duxbury Press.
- [11] Mach, L., dkk. 2007. Study of the Properties of the Rao-Wu Bootstrap Variance Estimator: What Happens When Assumptions Do Not Hold?. *Proceedings of the 2007 SSC Annual Meeting, Survey Methods Section, St-John's*.
- [12] Martinez, Gerardo Azor., dkk. 2011. Study of Variance Estimation Methods in the Spanish Labour Force Survey (EPA). *Instituto Nacional de Estadística*.
- [13] Preston, J. 2009. Rescaled Bootstrap for Stratified Multistage Sampling. *Survey Methodology*, 35, 227-234.
- [14] Saigo, H. 2010. Comparing Four Bootstrap Methods for Stratified Three-Stage Sampling. *Journal of Official Statistics*, 26, 193–207.
- [15] Schmidheiny, K. 2008. *Short Guides to Microeconometrics*. Melalui www.schmidheiny.name/teaching/upf/appliedl/topic10/bootstrap2up.pdf.
- [16] The Eurosystem Household Finance and Consumption Survey. 2013. *Methodological Report for the First Wave*. Germany: European Central Bank.
- [17] Yansaneh, Ibrahim S. 2005. *Overview of Sample Design Issues for Household Surveys in Developing and Transition Countries*. New York: Department of Economic and Social Affairs Statistics Division.
- [18] Undang-Undang No. 16 tahun 1997 tentang Statistik.