

LAPORAN PENELITIAN DOSEN STIS



***EFISIENSI PPS SAMPLING PADA AUXILARY VARIABEL YANG
BERKORELASI NEGATIF DENGAN PENDEKATAN METODE
MODIFIED HANSEN HURWITZ'S ESTIMATOR***

**Achmad Prasetyo
Ita Wulandari**



**PUSAT PENELITIAN DAN PENGABDIAN MASYARAKAT
POLITEKNIK STATISTIKA STIS
2018**

LEMBAR PENGESAHAN

Laporan Penelitian Dosen dengan judul:

EFISIENSI PPS SAMPLING PADA AUXILARY VARIABEL YANG BERKORELASI NEGATIF DENGAN PENDEKATAN METODE MODIFIED HANSEN HURWITZ'S ESTIMATOR

Nama Peneliti:
Achmad Prasetyo
Ita Wulandari

Dilaksanakan pada Maret 2018 sampai dengan November 2018

Telah disahkan oleh Kepala Pusat Penelitian dan Pengabdian Masyarakat (PPPM)
Politeknik Statistika STIS, pada tanggal **27 November 2018**

Menyetujui,
Kepala PPPM

Ketua Peneliti

Setia Pramana, S.Si., Ph.D
NIP. 197707222000031002

Achmad Prasetyo, S.Si., M.M
NIP197211171995121001

Mengetahui
Wakil Direktur I

Dr. Hardius Usman, S.Si., M.Si
NIP. 196704251989011002

Efisiensi PPS Sampling pada *Auxilliary Variable* yang Berkorelasi Negatif dengan Pendekatan Metode *Modified Hansen-Hurwitz's Estimator*

(Efficiency of PPS Sampling on Auxilliary Variable That are Negatively Correlated with Modified Hansen-Hurwitz's Estimator Method Approach)

Achmad Prasetyo¹ dan Ita Wulandari²

¹Politeknik Statistika STIS

²Politeknik Statistika STIS

¹praze@stis.ac.id

²ita.wulandari@stis.ac.id

ABSTRAK

Pada survei skala besar, dalam menentukan unit sampling yang yang akan dipilih selalu memanfaatkan *auxilliary variable* yang biasanya memiliki hubungan positif dengan variabel yang diteliti. Namun dalam kenyataannya sering juga ditemukan bahwa *auxilliary variabel* memiliki hubungan negatif dengan variabel yang diteliti sehingga apabila dipaksakan akan menyebabkan nilai estimasi menjadi tidak tepat. Untuk itu, dalam penelitian ini ingin dilihat keefisienan PPS Sampling ketika *auxilliary variable* memiliki korelasi negatif dengan pendekatan metode *Modified Hansen-Hurwitz's Estimator (MHH)*. Hasilnya kemudian dibandingkan dengan metode *simple random sampling*, *Hansen-Hurwitz's Estimator* dan *Product Estimator* dengan melihat nilai relatif standar error dan nilai *relative efficiency*. Sampel yang digunakan dalam penelitian ini sebanyak 6523 rumah tangga sampel dengan 58 variabel. Dari hasil penghitungan diperoleh 9 variabel yang bernilai negatif, kemudian dipilih 3 variabel, yaitu variabel umur KRT, lama bekerja KRT dan pengeluaran konsumsi untuk rokok yang masing-masing memiliki nilai korelasi sebesar -0,5; -0,15 dan -0,08. Hasil yang diperoleh menunjukkan bahwa keefisienan Estimator Metode MHH semakin baik ketika nilai korelasi antara *auxiliary variable* dengan variabel yang diteliti semakin mendekati -1. Hal ini menunjukkan bahwa estimator yang dihasilkan dengan Metode MHH dapat dijadikan sebagai alternatif untuk mengestimasi ketika dihadapkan pada nilai *auxilliary variable* yang berkorelasi negatif.

Kata kunci: *auxilliary variable*, *negatively correlated* dan Modified Hansen-Hurwitz's Estimator

ABSTRACT

Determining the sampling unit that will be selected in large scale surveys always use the auxiliary variable which usually has a positive correlation with the studied variable. But in reality it is often discovered auxiliary variables that have a negative correlation with the studied variables so that if forced it will cause the estimated value to be incorrect. For this reason, this study wants to see the efficiency of PPS sampling when the auxiliary variable has a negative correlation with the Modified Hansen-Hurwitz's Estimator (MHH) method approach. The results are then compared with the simple random sampling method, Hansen-Hurwitz's Estimator and Product Estimator by looking at the relative standard error values and relative efficiency values. The sample used in this study was 6523 household samples with 58 variables. The calculation results shows 9 variables that are negative. Then 3 variables are chosen, namely the KRT age variable, length of KRT work and consumption expenditure for cigarettes, each of which has a correlation value of -0.5; -0.15 and -0.08. The results obtained

show that the Estimator Efficiency of the MHH Method is better when the correlation value between the auxiliary variables and the variables studied is getting closer to -1. This shows that the estimator produced by the MHH Method can be used as an alternative to estimate when faced with an auxiliary variable value that is negatively correlated.

Key words: auxilliary variable, negatively correlated and Modified Hansen-Hurwitz's Estimator

A. PENDAHULUAN

Berbagai cara yang dilakukan dalam pengumpulan data, yaitu registrasi, sensus, survei dan eksperimen. Namun dari keempat cara pengumpulan data tersebut, survei merupakan cara pengumpulan data yang paling sering digunakan yaitu dengan hanya mengambil sebagian dari unit populasi yang kemudian digunakan untuk melihat gambaran populasinya. Dengan survei ini tentu memiliki keuntungan tersendiri dibanding dengan cara pengumpulan data lainnya. Keuntungan tersebut diantaranya adalah (1) dapat dilakukan penghematan dari sisi biaya, tenaga dan waktu; (2) cakupan variabel yang dikumpulkan lebih banyak; dan (3) dapat menekan *nonsampling error* sehingga akurasinya lebih baik.

Dalam pengambilan sampel yang bersifat konvensional, umumnya kita menggunakan estimator-estimator yang berbeda tergantung pada bagaimana sampel itu diperoleh. Jika pengambilan sampelnya dilakukan dengan sampel acak sederhana maka estimator yang digunakan adalah mean sampel, dengan asumsi parameternya adalah mean populasi. Di sisi lain, jika pengambilan sampelnya dilakukan dengan *probability proportional to size*, maka estimator yang digunakan adalah estimator Horvitz-Thompson (Horvits and Thompson, 1952). Dalam kegiatan survei, jenis sampling yang paling umum digunakan adalah *probability proportional to size* (PPS) dimana *size* dijadikan sebagai variabel pendukung (*auxilliary variable*). Prosedur ini menggunakan nilai-nilai *auxilliary variable* sedemikian rupa sehingga peluang unit sampling yang terpilih mengikuti besar kecilnya *size* tersebut. Semakin besar nilai *auxilliary variable* pada unit sampling maka peluang unit sampling itu akan terpilih sebagai sampel juga semakin besar. Penggunaan *auxilliary variable* terkait dengan variabel penelitian pada survei skala besar memainkan peran penting untuk mendapatkan estimator yang efisien. Namun demikian, sangat tidak ekonomis untuk dilakukan jika tujuan utamanya hanya memperkirakan satu parameter padahal kenyataannya banyak parameter lainnya (survei multi karakter) yang bisa dapat diperkirakan.

Dalam sampling PPS seperti yang dilakukan pada survei multi-karakter, beberapa karakteristik mungkin tidak terkait dengan *auxilliary variable* atau bahkan memiliki hubungan yang negatif dengan

variabel yang diteliti. Hal ini tentunya akan menimbulkan permasalahan pada estimator yang dihasilkan. Dan apabila dipaksakan akan menyebabkan nilai estimasi yang dihasilkan menjadi salah.

PKL STIS Tahun Akademik 2016/2017 mengangkat tema "*Multipurpose Survey dan Small Area Estimation: Studi Kemiskinan dan Pemerataan Pendapatan di Provinsi Kepulauan Bangka Belitung Tahun 2017*". Dengan teknik pengumpulan data yang digunakan berupa teknik wawancara langsung terhadap dua objek penelitian: (1) rumah tangga; (2) desa/kelurahan dan kecamatan. Untuk rumah tangga, metode penarikan sampel menggunakan *Stratified Three Stages Sampling* pada setiap kecamatan dengan rumah tangga sampel sebanyak 7080 rumah tangga yang diwawancarai dengan kuesioner pada *Computer Assisted Personal Interviewing (CAPI)* yang memuat 15 blok dan 263 butir pertanyaan. Sedangkan pada objek penelitian desa/kelurahan dan kecamatan, digunakan metode sensus dengan tujuan untuk mendapatkan karakteristik seluruh desa/kelurahan dan kecamatan sebagai variabel penyerta dalam *Small Area Estimation*.

Multipurpose survey adalah suatu survei yang dilakukan untuk meningkatkan efisiensi dengan cara menggabungkan beberapa tujuan survei yang berbeda. Pada PKL STIS Tahun Akademik 2016/2017, penerapan *multipurpose survey* menyebabkan tujuan dan populasi target dari masing-masing survei menjadi berbeda-beda, sehingga *desain sampling* yang akan dibuat harus mengakomodir berbagai tujuan dari survei-survei tersebut. Untuk mengatasi permasalahan ini, ditentukan suatu variabel pembantu (*auxilliary variable*) yang memiliki korelasi dengan variabel yang diteliti untuk semua tujuan survei. Hal ini mengakibatkan adanya kemungkinan korelasi yang negatif antara *auxilliary variabel* dengan variabel yang diteliti.

Survei skala besar dengan multi-karakter, yaitu suatu survei yang memperkirakan parameter populasi seperti total, mean dan varians dengan lebih dari satu indikator yang dikumpulkan dalam satu waktu. Misalkan suatu *auxilliary variable* tersedia dan hanya memiliki korelasi kuat dengan beberapa indikator namun memiliki korelasi yang tidak kuat dengan indikator lainnya, dalam kondisi seperti ini, sampel yang dipilih dengan peluang menggunakan *auxilliary variable* tersebut, dapat memberikan estimator yang efisien terhadap indikator-indikator yang memiliki korelasi kuat dengan variabel penelitian dan mungkin tidak memberikan estimator yang efisien untuk indikator yang kurang berkorelasi kuat dengan variabel penelitian.

Selama ini cara sederhana ketika menggunakan data survei yang memiliki korelasi negatif dari *auxilliary* variabel dan variabel yang diteliti adalah dengan *product estimator*. Namun estimator ini memiliki kelemahan yaitu estimator yang dihasilkan bias. Sehingga jarang sekali orang menggunakan estimator ini. Demikian juga dengan Sahoo, *dkk* (2012), dalam penelitiannya mendapatkan suatu estimator yang *unbiased* untuk hal ini yaitu menggunakan pendekatan *Modified Hansen-Hurwitz's Estimator*. Namun simulasi yang mereka lakukan hanya sebatas pada populasi kecil. Berkaitan dengan hal tersebut pertanyaan yang muncul pada peneliti ini adalah sejauh mana efisiensi yang diberikan metode *Modified Hansen-Hurwitz's Estimator* dibandingkan dengan metode *simple random sampling*, *Hansen-Hurwitz's Estimator* dan *product estimator*?

Hal menarik dalam memperkirakan parameter yang berkaitan dengan beberapa karakteristik adalah pentingnya penentuan *auxilliary variable* yang akan digunakan dalam memilih unit sampling utama (*primary sampling unit*). Beberapa karakteristik tersebut mungkin tidak terkait dengan *auxilliary variable* atau bahkan memiliki hubungan yang negatif dengan variabel yang diteliti. Apabila dipaksakan tentunya akan menimbulkan permasalahan pada estimator yang dihasilkan. Misalkan suatu *auxilliary variable* tersedia dan hanya memiliki korelasi kuat dengan beberapa indikator namun memiliki korelasi yang tidak kuat dengan indikator lainnya, dalam kondisi seperti ini, sampel yang dipilih dengan peluang menggunakan *auxilliary variable* tersebut, dapat memberikan estimator yang efisien terhadap indikator-indikator yang memiliki korelasi kuat dengan variabel penelitian dan mungkin tidak memberikan estimator yang efisien untuk indikator yang kurang berkorelasi kuat dengan variabel penelitian.

Selama ini cara sederhana ketika menggunakan data survei yang memiliki korelasi negatif dari *auxilliary variabel* dan variabel yang diteliti adalah dengan *product estimator*. Namun estimator ini memiliki kelemahan yaitu estimator yang dihasilkan *bias*. Sehingga jarang sekali orang menggunakan estimator ini. Demikian juga dengan Sahoo, *dkk* (2012), dalam penelitiannya mendapatkan suatu estimator yang *unbiased* untuk hal ini yaitu menggunakan pendekatan *Modified Hansen-Hurwitz's Estimator*. Namun simulasi yang mereka lakukan hanya sebatas pada populasi kecil. Berkaitan dengan hal tersebut maka dalam penelitian ini memiliki tujuan: (1) Mengetahui sejauh mana efisiensi yang diberikan metode *Modified Hansen-Hurwitz's Estimator*, dan (2) Membandingkan keefisienan metode *Modified Hansen-Hurwitz's Estimator* dengan metode *Simple Random Sampling*, *Hansen-Hurwitz's Estimator* dan *Product Estimator*.

Hensen dan Hurwitz (1943) memperkenalkan penggunaan unit sampling utama dengan peluang sebanding dengan beberapa ukuran *size* pada pengambilan sampel dari satu unit sampling utama per stratum. Kemudian Lahiri (1951) mengajukan skema sampling di mana sampel dipilih dengan peluang sebanding dengan total *size* dari variabel pendukung. Lahiri juga mempresentasikan metode pemilihan sampel sebenarnya, yang menghindari kebutuhan untuk daftar semua kemungkinan sampel dan mendapatkan total dan kumulatif *size*-nya. Grundy (1954) mengembangkan metode praktis untuk memilih unit sampling dengan peluang yang tepat sebanding dengan *size*, di mana baik penghitungan awal dan penambahan sejumlah besar *size* dihindari. Metode ini dianggap sebagai perpanjangan dari metode Lahiri.

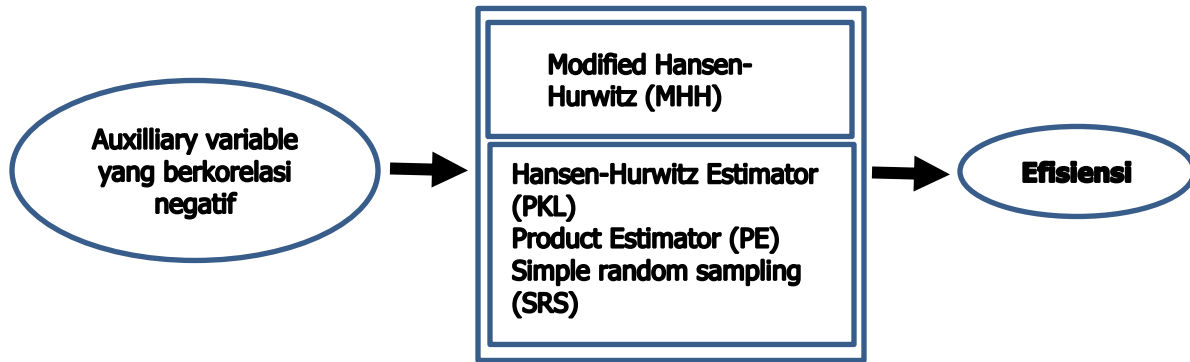
Bansal dan Singh (1985) mengemukakan estimator alternatif lain untuk variabel yang berkorelasi "buruk" dengan peluang pemilihan unit sampling. Estimator yang disarankan adalah *Probability Proportional to Size with replacement* (PPS-WR) dengan mempertimbangkan nilai kasar dari koefisien korelasi antara *auxiliary variable* (x) dengan variabel penelitian (y). Hal ini dikarenakan adanya fakta bahwa tidak biasa ditemui dalam praktik bila korelasi dalam populasi sama persis dengan nol.

Meskipun sejak tahun 1934 sejumlah besar makalah yang dipelajari telah ditulis memuji berbagai mode sampling yang *unequal probabilities*, namun tidak banyak yang telah dilakukan area sampling PPS dalam survei multi-karakter. Rao (1966) menyarankan suatu estimator alternatif dari total populasi untuk variabel yang berkorelasi buruk dengan peluang seleksi dalam skema PPS sampling untuk multi karakteristik. Dia lebih lanjut membandingkan penaksir alternatif ini dengan estimator konvensional di bawah model superpopulasi. Hal ini menunjukkan bahwa varians rata-rata dari alternatif estimator lebih kecil dari varians rata-rata dari estimator konvensional. Untuk membuat perbandingan efisiensi antara estimator biasa dengan penaksir alternatif yang dia usulkan, Rao menganggap populasi yang terbatas sebagai yang terpilih dari suatu superpopulasi tak terbatas dimana variabel penelitian, y , dan variabel tambahan, x , adalah independen. Hasil yang diperoleh tidak berlaku untuk populasi tunggal yang terbatas tetapi untuk rata-rata semua populasi terbatas yang dapat diambil dari super populasi.

Bedi dan Rao (1996) dalam jurnal penelitiannya mengatakan bahwa ada sebuah metode estimasi keefisienan *Probability Proportional to Size* (PPS) dengan transformasi variabel tambahan untuk situasi ketika antara variabel tambahan dan variabel studi memiliki korelasi negatif. Meskipun estimator dari *Hansen-Hurwitz* dan *Horvits-Thompson* adalah estimator yang *unbiased* namun akan memiliki varians yang besar ketika garis regresi x terhadap y jauh dari titik origin. Dalam situasi ini disarankan mentransformasi x ke x^* seperti $X_i^* = (X - X_i)$ dimana $i = 1, 2, \dots, N$. Nilai x^* umumnya lebih dari nol, sehingga kita dengan mudah melihat korelasi antara x^* dan y selalu positif sebanding dengan koefisien korelasi antara x dan y dimana $\sum X_i^* = (N-1)X$. Sehingga modifikasi peluang terpilihnya menjadi $p_i^* = \frac{(1-p_i)}{(N-1)}$ dengan $i = 1, 2, \dots, N$.

Singh dan Horn (1998) juga mengusulkan estimator alternatif untuk memperkirakan total populasi dalam survei multi-karakter ketika variabel-variabel tertentu memiliki korelasi positif yang buruk dan yang lain memiliki korelasi negatif yang buruk. Estimator ini memiliki definisi yang sama dengan penduga pada $\hat{\tau}_k = \frac{1}{n} \sum_1^N \frac{y_i}{g_k(x)}$ kecuali dalam peluang transformasi yang menggabungkan dua bentuk peluang seleksi yaitu: (1) untuk kasus korelasi positif adalah $p_i^+ = \frac{x_i}{X}; X = \sum_1^N x_i$ dan untuk kasus korelasi negatif adalah $p_i^- = \frac{z_i}{X}; X = \sum_1^N x_i$ dengan $z_i = \frac{X - nx_i}{N - n}$.

Sahoo dkk. (2012) menyarankan untuk mendapatkan estimator pada sampling PPS-WR dengan menggunakan mean harmonik dari *auxilliary variable* tanpa secara jelas mendefinisikan ukuran *size* normalnya dan karenanya tidak ada penerapan yang jelas untuk skema sampling PPS tersebut. Dari penjelasan diatas maka kerangka pikir yang terbentuk adalah sebagai berikut:



Gambar 1. Kerangka Pikir

Penelitian ini menggunakan data sekunder yaitu data PKL STIS Tahun Akademik 2016/2017 yang mengangkat tema “*Multipurpose Survey dan Small Area Estimation: Studi Kemiskinan dan Pemerataan Pendapatan di Provinsi Kepulauan Bangka Belitung Tahun 2017*”. Metode penarikan sampel yang digunakan adalah *Stratified Three Stages Sampling with PPS-WR* dengan *auxiliary variable* yang digunakan adalah jumlah balita disetiap rumah tangga sehingga diperoleh total sampel sebanyak 7080 rumah tangga.

B. METODOLOGI

Studi kemiskinan dan pemerataan pendapatan di Propinsi Bangka Belitung Tahun 2017 dilakukan di Provinsi Kepulauan Bangka Belitung yang terdiri dari 6 kabupaten dan 1 kota, 43 kecamatan, dan 361 Desa. Pelaksanaan pengumpulan datanya dilakukan pada tanggal 20 Pebruari sampai 3 Maret 2017 di 140 sampel desa yang tersebar di seluruh kecamatan di Provinsi Kepulauan Bangka Belitung. Populasi target pada survei tersebut adalah Seluruh rumah tangga biasa di Provinsi Kepulauan Bangka Belitung. Dengan desain sampling yang digunakan adalah *Stratified Three Stage Sampling* dimana terdapat dua strata di level kecamatan, yaitu strata perkotaan dan strata perdesaan dengan tahapan sebagai berikut:

- a) Tahap I. Pada masing-masing strata dilakukan pemilihan desa dari kerangka sampel daftar desa hasil Sensus Penduduk 2010 secara *Probability Proportional to Size With Replacement* (PPS-WR) dengan *size* jumlah rumah tangga yang memiliki balita dengan metode pemilihan kumulatif.

- b) Tahap II. Pada masing-masing desa yang terpilih pada tahap pertama dilakukan pemilihan blok sensus biasa dari kerangka sampel daftar blok sensus biasa hasil Sensus Penduduk 2010 secara PPS-WR dengan *size* jumlah rumah tangga yang memiliki balita dengan metode pemilihan kumulatif.
- c) Tahap III. Pada tahap ketiga dilakukan pemilihan rumah tangga dari blok sensus terpilih dari kerangka sampel daftar rumah tangga hasil listing pada blok sensus terpilih secara *Circular Systematic Sampling* dengan dua *implicit stratification* yaitu: kepemilikan balita dan tingkat pendidikan terakhir yang ditamatkan kepala rumah tangga.

Adapun variabel-variabel yang digunakan pada studi kemiskinan dan pemerataan pendapatan di Propinsi Bangka Belitung Tahun 2017 ada sebanyak 58 variabel diantaranya adalah : Keterangan Demografi, Keterangan Pendidikan, Keterangan bersekolah, Jenjang pendidikan tertinggi, dan sebagainya.

Korelasi

Dalam suatu penelitian, sering kali peneliti ingin mengetahui apakah dua himpunan memiliki hubungan atau tidak, atau ingin mengetahui besarnya kedua hubungan. Dengan mengetahui koefisien korelasi, kita dapat melihat besarnya hubungan antarvariabel secara signifikan kuat atau lemah. Selain itu, koefisien korelasi juga bisa digunakan untuk menguji reliabilitas observasi-observasi kita. Pengukuran ini akan menghasilkan ukuran-ukuran korelasi nonparametrik serta menampilkan statistik uji untuk menentukan kemungkinan yang berkaitan dengan adanya besar korelasi observasi dalam sampel.

Dalam penentuan korelasi, penghitungan korelasi untuk data skala pengukuran nominal dan sekurang-kurangnya ordinal berbeda. Dalam pengukuran koefisien korelasi skala ordinal, bisa digunakan beberapa uji, diantaranya: koefisien korelasi Rank Spearman, Ranking Kendall, Ranking Partial Kendall, dan koefisien Kendall W. Untuk skala pengukuran nominal, bisa digunakan korelasi koefisien kontingensi. Dalam makalah ini, hanya akan membahas tentang korelasi koefisien kontingensi saja.

Hubungan dua variabel atau dalam hal ini dua himpunan atribut disebut dengan korelasi. Tabel kontingensi merupakan tabel yang terdiri atas baris (himpunan pertama) dan kolom

(himpunan kedua) yang berisikan data kategorik. Sehingga Koefisien Korelasi Kontingensi merupakan suatu ukuran kadar hubungan (asosiasi) antara dua himpunan atribut yang bersifat kategorik dimana tabel tersebut merangkum frekuensi bersama dari observasi pada setiap himpunan atribut. Ukuran korelasi ini merupakan ukuran yang paling sederhana karena dapat digunakan ketika peneliti hanya memiliki informasi kategori (skala nominal) mengenai satu diantara himpunan-himpunan atribut atau kedua himpunan atribut itu. koefisien ini dapat dilakukan apabila atribut-atribut tersebut terdiri dari suatu rangkaian frekuensi yang tidak terurut. Koefisien kontingensi tak perlu membuat anggapan kontinuitas, bahkan dua himpunan atribut tersebut dengan susunan baris dan kolom yang berbeda akan memiliki harga perhitungan yang sama.

Simple Random Sampling (SRS)

Simple random sampling adalah teknik yang digunakan untuk memilih sampel dari populasi dengan cara sedemikian rupa sehingga setiap anggota populasi mempunyai peluang yang sama untuk diambil sebagai sampel. Sampel yang akan diambil bisa dengan cara pemulihan (*with replacement*) atau dengan tanpa pemulihan (*without replacement*). Misalkan elemen-elemen populasi yang terpilih secara *simple random sampling-with replacement* (SRS—WR) yang berukuran n ialah u_1, u_2, \dots, u_n dengan variabel, Y , sehingga sampel yang diperoleh yang mengandung nilai $y_1, y_2, y_3, \dots, y_n$. Total nilai suatu karakteristik tertentu dalam populasi berdasarkan sampel dapat dituliskan dengan rumus sebagai berikut:

$$\hat{Y} = \frac{N}{n} \sum_i^n y_i \tag{1}$$

dan perkiraan varians adalah:

$$v(\hat{Y}) = N^2 \frac{s^2}{n} \tag{2}$$

dan perkiraan *standard error*-nya adalah:

$$se(\hat{Y}) = \sqrt{v(\hat{Y})} \tag{3}$$

dengan perkiraan *relative standard error*-nya adalah:

$$RSe(\hat{Y}) = \frac{se(\hat{Y})}{\hat{Y}} \times 100\% \quad (4)$$

Yang membedakan rumus varians antara pengambilan sampel yang dilakukan secara SRS-WR dan SRS-WOR adalah terletak pada nilai $(1-f)$. Pada SRS-WOR, perkiraan variansnya yaitu dengan menambah nilai $(1-f)$ sehingga menjadi $v(\bar{y}) = \frac{N-n}{N} \frac{s^2}{n} = (1-f) \frac{s^2}{n}$. Nilai $(1-f)$ ini biasa disebut dengan istilah *finite population correction (fpc)* dan bisa diabaikan ketika nilai $f = \frac{n}{N}$ (*fraction*) kurang dari lima persen (Cochran, 1977).

Simple Random Sampling (SRS), baik yang dengan pengembalian atau tanpa pengembalian, merupakan metode pengambilan sampel yang paling mudah dipahami dan dilakukan, terutama kalau populasi tidaklah amat besar sehingga SRS yang dilakukan adalah berskala kecil. Dengan populasi yang tidak terlalu besar maka kerangka sampel yang berupa daftar semua unit pengamatan, yang sekaligus merupakan unit sampling, bisa diperoleh ataupun dimungkinkan untuk dibentuk terlebih dahulu.

Bahkan dengan tersedianya *random number generator*, maka proses mencari sejumlah angka random tidak lagi memerlukan TAR, sehingga menjadi amat mudah melakukan SRS. Dalam banyak penelitian berskala kecil, misal dengan populasi maksimum 1.000 unit pengamatan, yang juga sebagai unit sampling, maka SRS bisa dilakukan.

Kelebihan utama lainnya dari SRS adalah banyak metode analisis statistik yang mengasumsikan bahwa pengamatan yang dipunyai adalah pengamatan dengan variabel yang bersifat *independent* dan *identically distributed* dan asumsi ini terpenuhi secara penuh bila pengambilan sampel adalah SRS-WR. Bahkan bila N relatif amat besar dibandingkan dengan n, maka data yang berdasarkan SRS-WOR pun bisa dianggap menghasilkan data dengan variabel yang *independent* dan *identically distributed* (Asra dan Prasetyo, 2015).

Walaupun PSAS (*SRS*) merupakan teknik pengambilan yang paling sederhana dan mudah dimengerti, tetapi, bila ukuran populasi amat besar, *SRS* bisa menjadi mahal dan dalam prakteknya tidak mungkin untuk dilakukan (Asra dan Prasetyo, 2015). Alasan lain *SRS* jarang digunakan adalah bila yang menjadi fokus penelitian adalah sub-kelompok yang merupakan bagian kecil dari populasi.

Dengan mengambil sampel secara *SRS*, maka sampel dari kelompok ini bisa saja sangat kecil sehingga kurang memberikan keterwakilan yang cukup untuk memberikan gambaran tentang sub-kelompok tersebut. Untuk itu diperlukan beberapa metode sampling yang bisa lebih baik digunakan untuk keperluan survei tertentu (Levy dan Lemshow, 1999).

Probability Proportional to Size (PPS)

Probability proportional to size sampling (PPS Sampling) adalah suatu prosedur penarikan sampel dimana peluang terpilihnya suatu unit sampel sebanding dengan *auxiliary variable* dari unit sampling tersebut. *Auxiliary variable* yang dipertimbangkan sebagai dasar penarikan sampel harus memiliki korelasi yang kuat dengan variabel-variabel yang akan diteliti. Namun karena nilai *auxiliary variable* setiap unit sampling berbeda, maka peluang terpilihnya unit sampling pun menjadi berbeda pula, sehingga PPS Sampling termasuk dalam *unequal probability sampling*. Alasan penggunaan metode *PPS sampling* ini adalah karena *PPS Sampling*, sebagai suatu metode pemilihan sampel, akan memberikan penduga-penduga parameter populasi yang lebih efisien (Asra dan Prasetyo, 2015).

Pada suatu penarikan sampel sebanyak n unit yang diambil dari sebuah populasi dengan ukuran N unit secara PPS-WR dengan *auxiliary variable* sebesar x_i , maka besarnya peluang terpilihnya unit ke- i sebagai sampel adalah:

$$p_i = \frac{x_i}{\sum_{i=1}^N x_i} = \frac{x_i}{X} \text{ dimana } \sum_{i=1}^N p_i = 1 \tag{5}$$

Apabila y_i dianggap sebagai nilai variabel yang berpadanan dengan terpilihnya unit ke- i maka penduga yang *unbiased* bagi total adalah:

$$\hat{Y}_i = \frac{y_i}{p_i} = X \frac{y_i}{x_i} \tag{6}$$

\hat{Y}_i ($i = 1, 2, \dots, n$) merupakan penduga yang *unbiased* bagi total Y dan saling *independent*.

Maka kombinasi dari penduga tersebut merupakan penduga yang *unbiased* bagi total Y , yaitu:

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n \hat{Y}_i = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{p_i} = \frac{X}{n} \sum_{i=1}^n \frac{y_i}{x_i} \tag{7}$$

dengan perkiraan varians:

$$v(\hat{Y}) = \frac{1}{n(n-1)} \sum_i^n (\hat{Y}_i - \hat{Y})^2 = \frac{1}{n(n-1)} \sum_i^n \left(\frac{y_i}{p_i} - \hat{Y} \right)^2 = \frac{1}{n(n-1)} \left[\sum_i^n \left(\frac{y_i}{p_i} \right)^2 - n\hat{Y}^2 \right] \quad (8)$$

dan perkiraan *standard error*-nya adalah:

$$se(\hat{Y}) = \sqrt{v(\hat{Y})} \quad (9)$$

dengan perkiraan *relative standard error*-nya adalah:

$$RSe(\hat{Y}) = \frac{se(\hat{Y})}{\hat{Y}} \times 100\% \quad (10)$$

Hansen-Hurwitz's Estimator (HHE)

Dalam buku-buku teks tentang survey sampling yang ada sering merujuk pada estimator varians dari Sen–Yates–Grundy yang digunakan untuk desain *unequal probability sampling without replacement*. Namun estimator varians ini jarang sekali diimplementasikan karena kendala kompleksitas dalam menentukan *joint inclusion probabilities*-nya. Sehingga dalam prakteknya, estimator varians sederhana yang biasa digunakan adalah estimator Hansen-Hurwitz with replacement, walaupun sering kali menyebabkan varians yang dihasilkan cenderung *overestimate*.

Estimator Hansen-Hurwitz merupakan suatu *framework* untuk semua teknik sampling yang terkait dengan *unequal probability sampling with replacement* (Hansen and Hurwitz, 1943). Misalkan sampel berukuran n yang diambil secara with replacement dan pada setiap pengambilan memiliki peluang terpilihnya unit sampling ke- i dari populasi adalah sebesar p_i , maka estimator untuk total populasinya adalah:

$$\hat{\tau} = \frac{1}{n} \sum_i^n \frac{y_i}{p_i} \quad (11)$$

dimana y_i adalah setiap unit pengamatan ke- i dan besarnya nilai $p_i = \frac{x_i}{\sum_1^N x_i}$. Sehingga estimator

untuk varians dari total populasi yang *unbiased* adalah

$$v(\hat{t}) = \frac{1}{n(n-1)} \sum_1^n \left(\frac{y_i}{p_i} - \hat{t} \right)^2 \tag{12}$$

dan perkiraan *standard error*-nya adalah:

$$se(\hat{Y}) = \sqrt{v(\hat{Y})} \tag{13}$$

dengan perkiraan *relative standard error*-nya adalah:

$$RSe(\hat{Y}) = \frac{se(\hat{Y})}{\hat{Y}} \times 100\% \tag{14}$$

Modified Hansen-Hurwitz's Estimator (MHH)

Sebuah populasi terbatas Ω dari N yang teridentifikasi secara unik (1, 2, ..., N) yang didefinisikan sebagai dua nilai riil dari variabel x dan y dimana $x_i > 0$ dan $y_i > 0$. Sebuah sampel berukuran n diambil secara *with replacement* dari Ω dan variabel x maupun y memiliki korelasi negatif, maka transformasi untuk setiap pengambilan sampling memiliki peluang terpilihnya unit sampling ke- i dari populasi adalah sebesar $p_i' = \frac{z_i}{Z} = \frac{z_i}{\sum_1^N z_i}$ dimana $z_i = \frac{1}{x_i}$. Pembuktiannya adalah $y \propto \frac{1}{x} \Rightarrow y \propto z$,

dimana $z = \frac{1}{x}$, sehingga $p_i' = \frac{\frac{1}{x_i}}{\sum_1^N \frac{1}{x_i}} \Rightarrow p_i' = \frac{z_i}{Z}$. Di bawah transformasi tersebut, penamaan

$p_i' = \frac{z_i}{Z}$ dengan $z_i = \frac{1}{x_i}$ adalah realisasi peluang pemilihan untuk variabel yang berbanding terbalik

satu sama lain. Transformasi ini memiliki sifat-sifat seperti mean harmonik. Sebagai konsekuensi dari

transformasi ini, maka peluang pemilihan yang direalisasikan tersebut akan berbanding lurus dengan nilai Z, sehingga estimator ini dinamakan *Modified Hansen and Hurwitz (MHH) Estimator*.

Estimator untuk total populasinya adalah:

$$\hat{\tau}_{MHH} = \frac{1}{n} \sum_1^n \frac{y_i}{p_i} \quad (15)$$

estimator untuk varians dari total populasi yang *unbiased* adalah

$$v(\hat{\tau}_{MHH}) = \frac{1}{n(n-1)} \sum_1^n \left(\frac{y_i}{p_i} - \hat{\tau}_{MHH} \right)^2 \quad (16)$$

dan perkiraan *standard error*-nya adalah:

$$se(\hat{Y}) = \sqrt{v(\hat{Y})} \quad (17)$$

dengan perkiraan *relative standard error*-nya adalah:

$$RSe(\hat{Y}) = \frac{se(\hat{Y})}{\hat{Y}} \times 100\% \quad (18)$$

Product Estimator

Rasio estimator dari populasi total $\hat{Y}_R = \frac{\bar{y}}{\bar{x}} X$, Y lebih tepat daripada estimator rata-rata per unit $N\bar{y}$, asalkan koefisien korelasi ρ lebih besar dari $C_x/2C_y$ dan R positif. Jika koefisien korelasi antara peubah utama dan peubah tambahan x adalah negatif, maka kita tidak dapat menggunakan rasio estimator. Pada situasi serupa, Goodman (1960) mengusulkan tipe estimator lain untuk rata-rata \hat{Y}_p dan total Y, ditetapkan seperti

$$\hat{Y}_p = \frac{\bar{y}\bar{x}}{\bar{X}} \quad (19)$$

estimator untuk varians dari total populasi estimator produk \hat{Y}_p dapat ditulis sebagai

$$V(\hat{Y}_p) \cong Y^2 \left[\frac{V(\hat{Y})}{Y^2} + 2Cov \frac{(\hat{X}, \hat{Y})}{XY} + \frac{V(\hat{X})}{X^2} \right]$$

$$\cong V(\hat{Y}) + 2RCov(\hat{X}, \hat{Y}) + R^2V(\hat{X}) \tag{20}$$

an perkiraan *standard error*-nya adalah:

$$se(\hat{Y}) = \sqrt{v(\hat{Y})} \tag{21}$$

dengan perkiraan *relative standard error*-nya adalah:

$$RSe(\hat{Y}) = \frac{se(\hat{Y})}{\hat{Y}} \times 100\% \tag{22}$$

Relative Efficiency

Tingkat efisien juga harus diperhatikan saat menentukan sampel, karena dengan sampel yang baik akan didapatkan hasil yang optimum atau mendekati hasil yang sebenarnya. Adapun efisiensi dapat dilihat dari nilai variansi atau ragam. Nilai variansi atau ragam yang berbeda inilah yang dapat mempengaruhi nilai efisiensi sampel tersebut. Diketahui pula bahwa efisiensi inilah yang dapat memberikan kecenderungan data yang hampir sama sampel dengan populasi yang ada. Hal ini dikarenakan, penduga relatif lebih efisien dengan ragam lebih kecil dari penduga lainnya . (Miller & Miller, 1999).

Relative Efficiency antara dua metode sampling, dapat ditunjukkan dengan :

$$RE = \frac{Var(Metode A)}{Var(Metode B)} \times 100\% \tag{23}$$

Jika RE < 100%, maka metode A lebih baik dibandingkan dengan metode B

C. HASIL DAN PEMBAHASAN

Ada sebanyak 7080 data rumah tangga yang diperoleh dari pelaksanaan PKL STIS Tahun Akademik 2016/2017. Setelah dilakukan pemeriksaan ternyata hanya ada sebanyak 6523 data rumah tangga *eligible* yang digunakan pada penelitian ini. Hal ini disebabkan karena adanya *non-response unit* yaitu ketidaklengkapan isian kuesioner. Selanjutnya dari data rumah tangga yang *eligible* tersebut

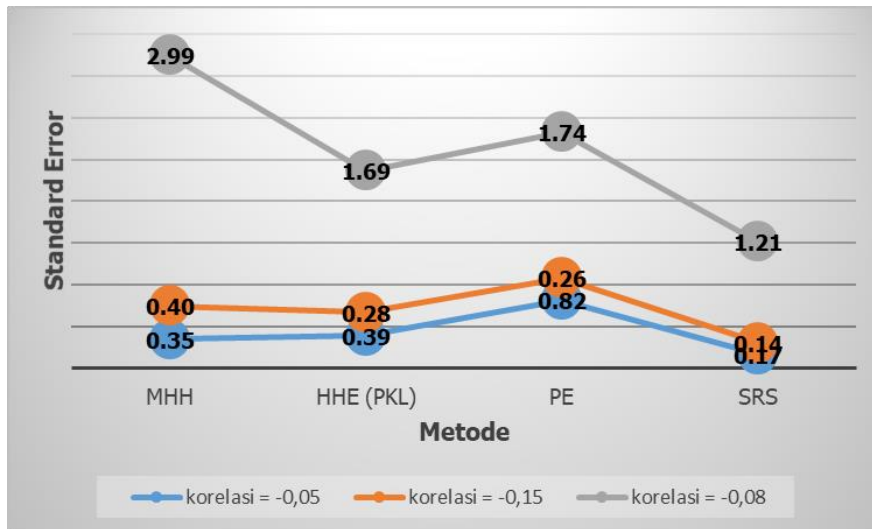
dilakukan penghitungan nilai korelasi untuk setiap variabelnya terhadap *auxiliary variable* yaitu jumlah balita yang dimiliki setiap rumah tangga dan diperoleh 9 variabel yang bernilai negatif sedang selebihnya positif. Kemudian dari 9 variabel yang bernilai negatif tersebut dipilih 3 variabel yang mewakili kondisi nilai korelasi rendah, sedang dan tinggi dari nilai korelasi yang diperoleh, yaitu variabel umur KRT, lama bekerja KRT dan pengeluaran konsumsi untuk rokok yang masing-masing memiliki nilai korelasi sebesar -0,5; -0,15 dan -0,08.

Untuk mempelajari *behaviour* dari estimator yang menggunakan metode MHH dibandingkan dengan estimator yang dihasilkan dari metode HHE (PKL), *product estimator* dan srs, maka kita mempertimbangkan nilai-nilai korelasi yang diperoleh tersebut ke dalam simulasi untuk mendapatkan nilai estimator maupun standard error yang dihasilkannya. Berdasarkan Tabel 1 di bawah ini terlihat bahwa estimasi nilai rata-rata dengan menggunakan metode *product estimator* (PE) menghasilkan nilai terendah pada setiap variabel untuk semua nilai korelasi. Sementara Metode *Modified Hansen-Hurwitz's Estimator* (MHH) menghasilkan estimasi nilai rata-rata terbesar ketika memiliki nilai korelasi terendah. Sedangkan estimasi rata-rata dari hasil PKL (menggunakan metode Hansen Hurwitz's Estimator/HHE) cenderung lebih tinggi dibanding metode lain untuk setiap nilai korelasi. Namun secara umum dapat dikatakan bahwa nilai estimasi yang dihasilkan oleh keempat metode tersebut memiliki nilai yang tidak jauh berbeda.

Tabel 1. Estimasi Nilai Rata-rata Variabel berdasarkan Korelasi dan Metode Estimasi

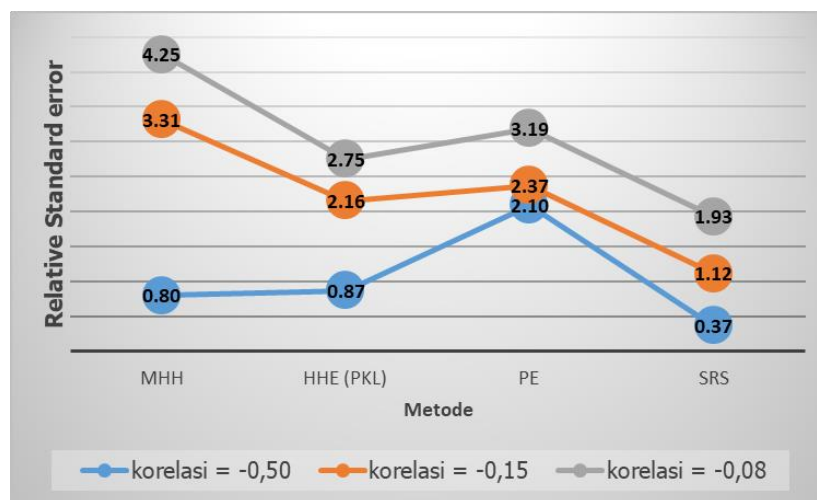
Variabel	Korelasi	MHH	HHE (PKL)	PE	SRS
Umur KRT (tahun)	-0.50	43.68	45.12	39.13	44.94
Lama Bekerja KRT (tahun)	-0.15	12.13	12.97	11.07	12.71
Pengeluaran Rokok (Rp.000)	-0.08	70.51	61.63	54.57	62.67

Sedangkan untuk mengetahui sejauh mana penyimpangan standar dari nilai estimasi yang kita peroleh dapat dilihat melalui nilai *standard error*-nya. Pada Gambar 2 terlihat bahwa untuk variable yang memiliki nilai korelasi yang rendah yaitu -0,08 menghasilkan nilai *standard error* yang rendah pula ketika menggunakan metode HHE. Sementara itu untuk variable yang memiliki nilai korelasi yang besar yaitu -0,50 menghasilkan nilai *standard error* yang kecil ketika menggunakan metode MHH. Dan apabila kita bandingkan nilai *standard error* dari hasil PKL yang menggunakan metode HHE, untuk nilai korelasi -0,50, maka hasil yang diperoleh dengan menggunakan metode MHH lebih kecil. Sehingga dapat dikatakan bahwa semakin besar nilai korelasi negatif untuk metode MHH maka akan semakin kecil nilai *standard error*-nya.



Gambar 2. Estimasi Nilai *Standard Error* Variabel berdasarkan Korelasi dan Metode Estimasi

Relative standard error adalah ukuran yang menunjukkan seberapa jauh penyimpangan nilai estimator yang dihasilkan terhadap nilai parameternya secara relatif. Nilai *relative standard error* ini biasa digunakan untuk melihat ketelitian dari estimasi nilai yang diperoleh. Semakin kecil nilai *relative standard error* maka semakin teliti estimasi nilai yang kita peroleh. Dari Gambar 3 terlihat bahwa semakin besar nilai korelasi negatif maka semakin kecil nilai *relative standard error*-nya untuk semua metode. Namun bila dibandingkan nilainya maka metode MHH memiliki nilai *relative standard error* yang lebih kecil dibandingkan dengan metode lainnya.



Gambar 3. Estimasi Nilai *Relative Standard Error* Variabel berdasarkan Korelasi dan Metode Estimasi (%)

Untuk melihat sejauh mana keefisienan dari metode sampling dibandingkan metode sampling yang lain, kita bisa mengukurnya dengan menggunakan nilai *relative efficiency*. Keefisienan secara relatif ini membandingkan varians yang diperoleh dari metode satu dengan metode yang lain. Dalam hal ini sebagai pembanding adalah metode *simple random sampling*. Apabila nilai *relative efficiency*-nya kurang dari 100 persen maka dapat dikatakan bahwa metode yang dibandingkan dengan metode SRS tersebut memiliki varians yang lebih kecil yang artinya lebih efisien. Pada Tabel 2 dapat dilihat bahwa pada metode MHH, semakin kecil nilai korelasi maka semakin kecil pula nilai efisiensi relatifnya dibandingkan hasil yang diperoleh dari data PKL yang menggunakan metode HHE.

Tabel 2. Estimasi Nilai *Relative Efficiency* Variabel berdasarkan Korelasi dan Metode Estimasi (%)

Variabel	Korelasi	MHH	HHE (PKL)	PE	SRS
Umur KRT	-0.50	4.40	5.53	24.30	100.00
Lama Bekerja KRT	-0.15	8.02	3.89	3.43	100.00
Pengeluaran Rokok	-0.08	0.00	0.00	2.07	100.00

D. KESIMPULAN DAN REKOMENDASI

Keefisienan Estimator Metode *Modified Hansen and Hurwitz (MHH)* semakin baik ketika nilai korelasi antara *auxiliary variable* dengan variabel yang diteliti semakin mendekati minus satu (-1). Hal ini sejalan dengan apa yang disarankan oleh Ikughur (2014), sehingga estimator yang dihasilkan dengan Metode *Modified Hansen and Hurwitz (MHH)* dapat dijadikan sebagai alternatif untuk mengestimasi ketika dihadapkan pada nilai *auxiliary variable* yang berkorelasi kuat dan negatif.

E. REFERENSI

- Asra, Abuzar dan Prasetyo, Achmad, 2015. Pengambilan Sampel dalam Penelitian Survei. Jakarta: Rajawali Pers.
- Bangsal, M.L. and Singh, R. 1985. *An Alternative Estimator for Multiple Characteristics in PPS Sampling*. J. Statist. Planning Infer., 11, 313-320
- Bedi, P.K. and Rao, T.J. 1996. *Efficient Utilization of Auxiliary Information at Estimation Stage*, "Biometrical Journal", 38, pp. 973-976.
- Goodman, L. A. 1960. *On The Exact Variance of Products*. JASA 55, 708-713.
- Grundy, P.M. 1954. *A Method of Sampling with Probability Exactly Proportional to Size*. Journal of the Royal Statistical Society–Series B 16(2): 236-238.

- Hansen, Morris H.; Hurwitz, William N. 1943. *On the Theory of Sampling from Finite Populations*. Ann. Math. Statist. 14, no. 4, 333--362. doi:10.1214/aoms/1177731356.
<https://projecteuclid.org/euclid.aoms/1177731356>
- Ikughur, Jonathan A. 2014. *Modification of Hansen-Hurwitz's Estimator for Negatively Correlated Variates*. CBN Journal of Applied Statistics, ISSN 2476-8472, The Central Bank of Nigeria, Abuja, Vol. 5, Iss. 1, pp. 1-14
- Lahiri, D. B. 1951. *A method for sample selection providing unbiased ratio estimates*. Bull. Int. Stat. Inst., 33, 2, 133-140.
- Miller, I., and Miller, M. (1999), *John E. Freund's Mathematical Statistics* (6th ed.), Upper Saddle River, NJ: Prentice
- Murthy, M. N. 1967. *Sampling Teory Methods*, Statistical Publishing Society
- Rao, J.T and Bedi, P.K. 1996. *PPS Method of estimation Under a Transformation*. Journal of The Indian Society of Agricultural Statistics.
- Sahoo, LN., Dalabehera M., Mangaraj, AK. 2012. *A PPS Sampling Using Harmonic Mean of Auxilliary Variable*. ARPN Journal of System and Software. Vol 2 No 7. ISSN: 2222-9833.
<http://www.scientific-journal.org>
- Singh, D. , et all. 1986. *Theory and Analysis of Sampel Survey Design*, Wiley Eastern Limited
- Singh, S. and Horn, S. 1998. *An alternative estimator for multi-character surveys*. Metrika, 48:99-107
- Yamane, Taro. 1967. *Elementary Sampling Theory*. Prentice-Hall, Inc, Englewood Cliffs, NJ.