ANALISIS METODE REGRESI UNTUK IMPUTASI DATA PADA SURVEI SAMPEL

SKRIPSI

Oleh:
NUGRAHENI FITROH REZQI SYAKARNA
NIM. 09610045

JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2014

ANALISIS METODE REGRESI UNTUK IMPUTASI DATA PADA SURVEI SAMPEL

SKRIPSI

Diajukan kepada:
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Sains (S.Si)

Oleh:
NUGRAHENI FITROH REZQI SYAKARNA
NIM. 09610045

JURUSAN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2014

ANALISIS METODE REGRESI UNTUK IMPUTASI DATA PADA SURVEI SAMPEL

SKRIPSI

Oleh:
NUGRAHENI FITROH REZQI SYAKARNA
NIM. 09610045

Telah Diperiksa dan Disetujui untuk Diuji
Tanggal: 27 Desember 2013

Pembimbing I,	Pembimbing II,
Fachrur Rozi, M.Si	Abdussakir, M.Pd
NIP. 198005272008011 012	NIP. 19751006 200312 1 001

Mengetahui,
Ketua Jurusan Matematika

Abdussakir, M.Pd
NIP. 19751006 200312 1 001

ANALISIS METODE REGRESI UNTUK IMPUTASI DATA PADA SURVEI SAMPEL

SKRIPSI

Oleh:
NUGRAHENI FITROH REZQI SYAKARNA
NIM. 09610045

Telah Dipertahankan di Depan Dewan Penguji Skripsi dan Dinyatakan Diterima sebagai Salah Satu Persyaratan untuk Memperoleh Gelar Sarjana Sains (S.Si)
Tanggal: 09 Januari 2014

Penguji Utama	:	Dr. Sri Harini, M.Si NIP. 19731010 200112 2 001	______________
Ketua Penguji	:	Dr. H. Imam Sujarwo, M.Pd NIP. 19630502 198703 1 005	______________
Sekretaris Penguji	:	FachrurRozi, M.Si NIP. 19800527 200801 1 012	______________
Anggota Penguji	:	Abdussakir, M.Pd NIP.19751006 200312 1 001	______________

Mengesahkan,
Ketua Jurusan Matematika

Abdussakir, M.Pd
NIP.19751006 200312 1 001

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Nugraheni Fitroh Rezqi Syakarna
NIM	: 09610045
Jurusan	: Matematika
Fakultas	: Sains dan Teknologi

menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 23 Januari 2014
Yang membuat pernyataan,

Nugraheni Fitroh Rezqi S.
NIM. 09610045

MOTTO

حَسْبِيَ ٱللَّهُ لَآ إِلَٰهَ إِلَّا هُوَ ۖ عَلَيْهِ تَوَكَّلْتُ ۖ وَهُوَ رَبُّ ٱلْعَرْشِ ٱلْعَظِيمِ ﴿١٢٩﴾

"Cukuplah Allah bagiku; tidak ada Tuhan selain Dia. hanya kepada-Nya aku bertawakkal dan Dia adalah Tuhan yang memiliki 'Arsy yang agung"(Qs. At-Taubah:129).

PERSEMBAHAN

Penulis mempersembahkan karya ini untuk:

Ayahanda tercinta, Mahfudz yang selalu memberikan motivasi, nasehat-nasehat dan mendoakan penulis di setiap waktu.
Ibunda terkasih, Siti Ngaisah teladan kegigihan, kesabaran yang selalu memberikan motivasi dan menyebut nama penulis di setiap sholatnya,
Kakak tersayang, Willy Rabindra teladan kakak yang baik bagi penulis dan adik tersayang Tegar Ayyu yang menjadi penghibur penulis di kala sedih

YOU ALL ARE MY EVERYTHING

KATA PENGANTAR

بسم الله الرحمن الرحيم

Assalamu'alaikum Wr. Wb.

Alhamdulillah, puji syukur ke hadirat Allah SWT atas limpahan rahmat, taufik, hidayah, dan inayah-Nya, sehingga penulis dapat menyelesaikan penulisan skripsi ini dengan baik. Shalawat serta salam semoga tetap tercurahkan kepada Nabi Muhammad SAW pembimbing umat manusia, *rahmatan lil 'alamin* yang kelak diharapkan syafaatnya *fii yaumil qiyamah* Amin.

Keberhasilan penulisan skripsi ini tidak lepas dari bantuan, arahan, dan bimbingan dari berbagai pihak, baik berupa pikiran, motivasi, tenaga, dan do'a, karena itu penulis mengucapkan terima kasih kepada:

1. Prof. Dr. H. Mudjia Rahardjo, M.Si, selaku Rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. drh. Hj. Bayyinatul Muchtaromah, M.Si, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang
3. Abdussakir, M.Pd, selaku Ketua Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang dan dosen pembimbing keagamaan yang telah memberikan pencerahan dalam bidang kajian keagamaan.
4. Fachrur Rozi, M.Si, selaku dosen pembimbing skripsi yang dengan sabar telah meluangkan waktunya demi memberikan bimbingan dan arahan dalam penyelesaian skripsi ini.

5. Seluruh dosen dan staf administrasi Jurusan Matematika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
6. Kedua orang tua tercinta, kakak dan adik tersayang yang tak henti-hentinya memanjatkan do'a dan selalu memberikan semangat, motivasi untuk terus berjuang.
7. Teman-teman mahasiswa Jurusan Matematika angkatan 2009 yang telah menemani belajar selama kuliah, selama mengerjakan penelitian dan memberikan kenangan berarti dalam hidup penulis.
8. Teman-teman Kos Wisma Asri, teman-teman Jurusan Statistika Universitas Brawijaya, dan teman-teman FLP Malang terima kasih atas segala bantuannya baik berupa waktu, tenaga, motivasi, maupun pikiran.
9. Semua pihak yang tidak mugkin penulis sebut satu-persatu, atas keikhlasan bantuan, dukungan, dan do'anya.

Akhirnya, semoga skripsi ini bermanfaat bagi diri penulis dan pembaca, *Amin ya robbal 'alamin.*

Wassalamu'alaikum Wr. Wb.

Malang, Januari 2014

Penulis

DAFTAR ISI

DAFTAR SIMBOL

$\bar{y}_s$: rata-rata sampel.
$\bar{y}_r$: rata-rata data y respon.
$\hat{b}$: taksiran awal tanpa melibatkan data yang hilang.
$\bar{y}_{rat}$: estimator ratio.
y_i : data y ke-i.
$\bar{x}_n$: rata-rata data x penuh.
$\bar{x}_r$: rata-rata data x dengan tidak memasukkan data yang sejajar dengan data y respon.
x_i : data x ke-i.
$\bar{X}$: rata-rata populasi dari$X$.
$\bar{Y}$: rata-rata populasi dari Y.
$\hat{\beta}$: taksiran yang melibatkan data hilang.
$\bar{y}_{reg}$: rata-rata imputasi regresi.
s_x : simpangan baku dari x.
s_y : simpangan baku dari y.
s_{xy} : kovarian sampel x dan y.
S_x : simpangan baku dari X.
S_y : simpangan baku dari Y.
S_{xy} : kovarian populasi X dan Y.
ρ_{xy} : koefisien korelasi pada populasi Xdan Y.
C_x^2 : kesalahan relatif X.
C_y^2 : kesalahan relatif Y.
y_{imp} : $\bar{y}_r + \hat{\beta}(x_i - \bar{x}_r)$.
$\bar{y}_{reg}$: $\bar{y}_{imp}$: rata-rata dari y_{imp} .

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN

ABSTRAK

Syakarna, Nugraheni Fitroh Rezqi. 2014. **Analisis Metode Regresi untuk Imputasi Data pada Survei Sampel**. Skripsi. Jurusan Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang.
Pembimbing: (I) Fachrur Rozi, M.Si
(II) Abdussakir, M.Pd

Kata Kunci: Metode Imputasi Regresi, Data Hilang, *Mean Square Error* (MSE)

Kasus data hilang pada survei mengakibatkan pendugaan parameter menjadi tidak efisien karena ukuran data berkurang, sehingga menyebabkan kesulitan dalam menganalisis data. Metode imputasi regresi adalah salah satu metode imputasi untuk memprediksi nilai data yang hilang (y_{miss}) menggunakan pendekatan regresi. Berbagai macam uji coba, metode regresi imputasi adalah salah satu metode alternatif dari metode imputasi yang lain. Hal ini karena antara y_{miss} dan y_{imp} mempunyai kesalahan yang relatif kecil dan lebih mendekatkan pada kevalidan data.

Konsep proses imputasi adalah dengan mengambil sampel berukuran 50, 100, 200 dari populasi dengan 10 kali percobaan. Setiap 10 kali percobaan peubah y akan dihilangkan sebanyak 5%, 10%, dan 10%. Menggunakan model imputasi regresi akan dilakukan imputasi sebanyak y_{miss} tersebut. Percobaan metode regresi imputasi mempunyai hasil yang memuaskan. Antara y_{miss} dan y_{imp} mempunyai kesalahan yang relatif kecil. Hal ini bisa dilihat dari MSE setiap sampel dan setiap jumlah data yang hilang.

Data yang diambil mempunyai jumlah yang sama tapi mempunyai persentase nilai hilang semakin besar diperoleh nilai MSE semakin besar dan jika data yang diambil mempunyai jumlah semakin besar tapi mempunyai persentase nilai hilang yang sama maka MSE semakin kecil. Nilai MSE semakin besar ketika data yang hilang juga semakin besar. Hal ini dikarenakan semakin banyak data yang hilang maka data yang akan diimputkan pun juga akan semakin banyak, sehingga akan banyak muncul nilai kesalahan dari hasil pengimputan data tadi dan demikian pula sebaliknya. Nilai MSE semakin kecil ketika data yang diambil mempunyai jumlah semakin besar dengan persentase nilai hilang yang sama. Hal ini dikarenakan jumlah sampel yang besar semakin menggambarkan populasi. Penelitian selanjutnya disarankan membandingkan metode regresi imputasi dan metode *robust* imputasi terhadap outlier.

ABSTRACT

Syakarna, Nugraheni Fitroh Rezqi. 2014. Regression Analysis for Data Imputation in Sample Surveying. Thesis. Department of Mathematics, Faculty of Science and Technology, State Islamic University of Maulana Malik Ibrahim Malang.
Supervisor: (I) Fachrur Rozi, M.Si
(II) Abdussakir, M.Pd

Keyword: Regression Imputation Methods, Missing Data, Mean Square Error (MSE)

The case of missing data while surveying results in the inefficiency of parameter prediction because the size of the data decreases and that causes difficulties in data analysis. Regression imputation method is one of the imputation method to predict data value lost (y_{miss}) by using regression approach. It is one alternative among other imputation methods. It is because between y_{miss} and y_{imp} have the lowest relative error, and is closer to data validity.

The concept of imputation is the process by taking a sample size of 50, 100, 200 of the population with 10 attempts. Every 10 attempts variable y will be eliminated as much as 5%, 10%, and 10%. Using regression imputation models will do as much as y_{miss} the imputation. The experiment regression imputation methods have satisfactory results. Between y_{miss} and y_{imp} relative have small errors. It can be seen from the MSE of each sample and each amount of missing data

The data collected has the same value but the missing percentage are different, MSE value is increasing and when data collected increases but has the same percentage of missing value so MSE is decreasing. MSE value increases when missing data increasing. This is because when missing data is bigger, the data imputed also increases, results in the increasing of error value of data imputation and vice versa. MSE value gets decreasing when the data collected has the increasing amount by the same percentage of missing values. This is because the larger number of samples will show the populations. It is suggested for the next studies that they compare imputation regression method to imputation robust method toward outlier.

ملخص

شاكرنا, نوغرهيني فطرة رزقي. ٢٠١٤ . **طرق تحليل الانحدار للبيانات الإتهام في مسوحات العينة** . أطروحة.سعبة الرياضيات. العلوم والتكنولوجيا في الجامعة الإسلامية الحكومية مولانا مالك إبراهيم مالانج.
المشرف: (١) فخر الرا زي, الما جستير
(٢)عبد الشا كر, الماجستير

كلمات البحث:الانحدار الإتهامطرق، مفقودالبيانات، متوسطمربعالخط (MSE)

حالة البيانات المفقودة على نتائج المسح في تقديرات المعلمة غير فعالة لأن يتم تقليل حجم البيانات، مما يسبب صعوبة في تحليل البيانات. طريقة احتساب الانحدار هي واحدة من طريقة للتنبؤ احتساب قيمة البيانات المفقودة (y_{miss}) باستخدام نهج الانحدار. من أنواع مختلفة من الاختبار، و طريقة احتساب الانحدار هي واحدة من أساليب بدائل طرق احتساب الأخرى. وذلك لأن بين y_{miss} و y_{imp} دينا أخطاء صغيرة نسبيا و الاقتراب من صحة البيانات .

مفهوم الإسناد هو العملية التي أخذ عينة من حجم 50 ، 100، 200 من السكان مع 10 محاولات . سيتم القضاء على كل 10 محاولات المتغير y بقدر 5 ٪، 10 ٪ ، و 10 ٪ . باستخدام نماذج الانحدار احتساب سوف تفعل بقدر ما y_{miss} احتساب . من الانحدار تجربة طرق احتساب يكون لها نتائج مرضية. بين y_{miss} و y_{imp} دينا أخطاء صغيرة نسبيا. يمكن أن ينظر إليه من MSE من كل عينة و كل كمية البيانات المفقودة.

إذا اتخذت البيانات أن يكون نفس المبلغ ولكن بعد أن خسرت نسبة أكبر من قيمة من القيم التي تم الحصول عليها MSE أكبر و إذا اتخذت بيانات عدد كبير ومتزايد ولكن قد فقدت نفس النسبة من قيمة MSE أصغر. قيمة MSE أكبر عندما يتم الحصول على البيانات المفقودة أيضا أكبر. هذا هو بسبب فقدان المزيد من البيانات ثم البيانات إلى أن يتم إدخال ستكون أيضا أكثر وأكثر. وبالتالي فإن الكثير من القيمة الناشئة من الخطأ كان نتيجة ل إدخال البيانات والعكس بالعكس. MSE يحصل أصغر عندما يتم أخذ البيانات ل ديهم كمية أكبر من نفس النسبة المئوية من القيم المفقودة . وذلك لأن أكبر عدد من العينات أن السكان . لمزيد من البحث ويوصى لمقارنة طريقة احتساب و الانحدار طريقة احتساب قوية ضد القيم المتطرفة .

BAB 1
PENDAHULUAN

1.1 Latar Belakang

Kegiatan survei dilakukan untuk memperoleh informasi lebih detail dengan mengamati sebagian unit dalam suatu populasi. Dalam survei sering kali dijumpai adanya data hilang atau tidak lengkap (*missing data*). Beberapa hal yang menyebabkan *missing data* misalnya peralatan yang tidak berfungsi dengan baik, kekurangan fasilitas, penolakan responden untuk menjawab pertanyaan, dan lain sebagainya.

Adanya *missing data* mengakibatkan pendugaan parameter menjadi tidak efisien karena ukuran data berkurang sehingga menyebabkan kesulitan dalam menganalisis data. Dalam sensus atau survei sering kali ditemukan unit-unit yang tidak merespon jumlah pertanyaan yang telah diajukan. Kish (1965:67) mendefinisikan non respon di sini adalah suatu kegagalan untuk mendapatkan nilai pengamatan dari beberapa unit yang menjadi sampel. Non respon juga dapat terjadi karena kesalahan dalam menuliskan jawaban (Longford, 2005:28).

Metode analisis untuk data lengkap sering digunakan untuk mengatasi permasalahan *missing data* dengan cara menghapus unit-unit pengamatan yang mempunyai *missing data*. Prosedur tersebut tidak baik karena penghapusan unit-unit pengamatan data yang hilang akan mengurangi sampel yang sudah ditentukan awal oleh peneliti (Malahayati, 2008:01).

Di sini penulis akan memakai metode imputasi untuk menangani permasalahan *missing data* pada survei sampel. Menurut Little & Rubin (1987:56), imputasi adalah metode pengisian data untuk mengatasi *missing data* karena tidak adanya respon terhadap beberapa pertanyaan. *Missing data* karena tidak adanya respon terhadap beberapa pertanyaan dapat dianalogikan seperti dalam surat Al-Baqarah ayat 283:

وَلَا تَكْتُمُوا الشَّهَادَةَ ۚ وَمَن يَكْتُمْهَا فَإِنَّهُ آثِمٌ قَلْبُهُ ۗ وَاللَّهُ بِمَا تَعْمَلُونَ عَلِيمٌ ﴿٢٨٣﴾

Artinya: *"…….dan janganlah kamu (para saksi) menyembunyikan persaksian. Dan barangsiapa yang menyembunyikannya, maka Sesungguhnya ia adalah orang yang berdosa hatinya; dan Allah Maha mengetahui apa yang kamu kerjakan"*(Qs. Al-Baqarah:283).

Maksud dari ayat di atas adalah dilarang untuk menyembunyikan, melebih-lebihkan, dan jangan pula mengabaikan. Ibnu 'Abbas dan ulama lainnya mengatakan: "Kesaksian palsu merupakan salah satu dosa besar yang paling besar, demikian juga menyembunyikannya". Oleh karena itu, Allah berfirman: "dan barang siapa menyembunyikannya, maka sesungguhnya ia adalah orang yang berdosa hatinya" (Alu, 2007:569-571).

Maksud dari kalimat وَلاَ تَكْتُمُوا الشَّهَادَةَ yakni seorang saksi tidak boleh menyulitkan salah satu pihak yang bertransaksi dengan menutupi kesaksian. Hukum larangan ini adalah untuk diwajibkan (wajib untuk dihindari), dan salah satu tanda atau petunjuk pewajibannya adalah kalimat ancaman yang disebutkan setelahnya. Ibnu Abbas mengatakan: yang diwajibkan kepada saksi adalah untuk bersaksi sesuai dengan apa yang disaksikannya dan memberitahukan sesuai dengan keadaan yang sebenarnya.

Firman Allah SWT وَمَنْ يَكْتُمْهَا فَإِنَّهُ,ءَاثِمٌ قَلْبُهُ "dan barang siapa yang menyembunyikannya, maka sesungguhnya ia adalah orang yang berdosa hatinya." Alasan menyebut kata hati secara khusus pada ayat ini adalah karena menyembunyikan sebuah kesaksian adalah salah satu yang dilakukan oleh hati. Menyembunyikan kesaksikan menyebabkan hilangnya faktor-faktor pendukung sehatnya hati. Sehigga hati tidak dapat merespon hal-hal yang baik masuk untuk memenuhi kebutuhan jiwa. Seperti yang diriwayatkan dari Nabi SAW, yaitu bahwa hati adalah segumpal daging, yang jika baik maka seluruh tubuh menjadi baik. Oleh karena itu, kata (hati) adalah bagian dari sesuatu (tubuh), namun yang dimaksud dari penyebutan bagian tersebut adalah keseluruhannya (Al-Qurthubi, 2008a:920-922).

Beberapa contoh metode imputasi adalah metode imputasi rata-rata *(Mean Imputation)*, metode rasio, dan imputasi regresi. Singh & Deo (2002) dalam penelitiannya yang berjudul *Imputation by Power Transformation* telah membandingkan $MSE(\bar{y}_{rat})$ dan $V(\bar{y}_m)$. Hasil dari perbandingan tersebut menunjukkan metode ratio imputasi lebih baik dari metode mean imputasi jika berlaku dalam situasi yang paling praktis.

Sebuah artikel yang berjudul *"Editing and Imputation of Tax Return File-Evaluastion of Applied Methods"* memberikan kesimpulan bahwa metode imputasi rasio ini tidak dapat membaca kesalahan variabel/melokalisir kesalahan, sehingga dari kasus ini dikembangkan metode imputasi regresi. Metode imputasi regresi adalah salah satu metode imputasi pada praktek survei dengan mengganti nilai yang hilang dan memprediksi nilai tersebut menggunakan regresi pada suatu unit (Little dan Rubin, 1987:61).

Metode ini memodelkan variabel lain yang berkaitan yang terekam dalam survei untuk memprediksi *missing data* tersebut. Sebagai contoh ketika data penghasilan dari seorang responden tidak diketahui, model regresi dengan menggunakan karakteristik demografi seperti umur, jenis kelamin, pendidikan dan jabatan dari responden tersebut bisa digunakan untuk mengestimasi penghasilan (Basuki, 2010:2). Selain itu menggunakan metode imputasi regresi akan meminimal kesalahan dan lebih mendekatkan pada besarnya kevalidan data.

Singh dan Valdes (2009) pada penelitiannya yang berjudul *Optimal Method of Imputation* mencari metode optimal imputasi yang mengarah pada suatu perkiraan rata-rata populasi dengan meminimumkan MSE pada survei sampel ketika nilai data *Missing Completely at Random* (MCAR). Hasil penelitian menunjukkan bahwa gabungan dari ketiga metode imputasi yaitu metode *mean*, rasio, dan regresi menghasilkan metode optimal. MSE pada metode mean dan ratio telah terjabarkan, akan tetapi pada metode regresi ini belum terjabarkan. Dari latar belakang ini penulis tertarik untuk mengkaji metode regresi imputasi data pada survei sampel.

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah:

1. Bagaimanakah analisis tentang metode regresi untuk imputasi data pada survei sampel?

2. Bagaimanakah integrasi nilai-nilai agama dalam metode regresi untuk imputasi data pada survei sampel?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah:

1. Mengetahui analisis tentang metode regresi untuk imputasi data pada survei sampel.
2. Mengetahui integrasi nilai-nilai agama dalam metode regresi untuk imputasi data pada survei sampel.

1.4 Batasan Masalah

Agar pembahasan dalam penelitian ini tidak begitu meluas, maka peneliti hanya membahas pada metode regresi imputasi dalam jurnal Singh & Valdes (2009) berjudul *Optimal Method of Imputation in Survey Sampling*.

1.5 Manfaat Penelitian

Pada penelitian ini diharapkan dapat bermanfaat, di antaranya:

1. Sebagai suatu tambahan keilmuan dalam statistika khususnya survei sampel.
2. Metode alternatif untuk menangani permasalahan *missing data* dalam survei sampel.
3. Dapat dengan mudah mengatasi permasalahan *missing data*.

1.6 Metode Penelitian

Penelitian ini dilakukan menggunakan pendekatan penelitian perpustakaan *(library research)* dan deskriptif kuantitatif. Dimana untuk menganalisis metode regresi imputasi, terlebih dahulu dikaji mengenai pengertian imputasi dan konsep dasar regresi. Selanjutnya dilakukan analisis deskriptif tentang bentuk *missing data* dan cara mengimputasinya adalah sebagai berikut:

1. Tahap ilustrasi. Tahap ini terletak pada pembentukan ilustrasi data yang hilang dan akan dilakukan imputasi.
2. Tahap analisis metode regresi untuk imputasi. Pada tahap ini penulis akan menganalisis model regresi untuk imputasi. Tahap analisis adalah sebagai berikut:
   a. Pendefinisian model regresi untuk imputasi.
   b. Menaksir $\hat{\beta}$ model regresi untuk imputasi.
   c. Menaksir rata-rata metode regresi imputasi.
   d. Menentukan MSE
3. Melakukan simulasi.
   a. Dibangkitkan data populasi sebesar 1000 unit.
   b. Dari data populasi tersebut diambil sampel berukuran 50, 100 dan 200 dan diulang sebanyak 10 kali setiap sampelnya.
   c. Pada setiap sampel dan setiap percobaan dilakukan penghilangan data sebanyak 5%, 10% dan 15% pada peubah y, sedangkan peubah x dibiarkan lengkap.
   d. Setiap imputasi dibandingkan nilai y_{miss} dan y_{imp} dan dihitung $\bar{y}_{reg}$

e. Dihitung MSE setiap sampel dan setiap percobaan.

1.7 Sistematika Penulisan

Sistematika penulisan digunakan untuk mempermudah dalam memahami intisari dari penelitian ini, terbagi menjadi empat bagian, yaitu:

BAB I Pendahuluan

Pada bab ini akan diuraikan tentang latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, manfaat penelitian, dan metode penelitian.

BAB II Kajian Pustaka

Meliputi penjabaran materi metode imputasi, *missing data*, mekanisme data hilang, bias, dan MSE.

BAB III Pembahasan

Bab ini menguraikan keseluruhan langkah yang disebut dalam metode penelitian.

BAB IV Penutup

Pada bab ini dibahas tentang kesimpulan dari pembahasan hasil penelitian yang telah dibahas dengan dilengkapi dengan saran-saran yang berkaitan dengan penelitian ini.

BAB II
KAJIAN PUSTAKA

2.1 Pengertian dan Tujuan Survei

Menurut kamus Bahasa Indonesia survei bisa diartikan sebagai inspeksi, pemeriksaan, penilikan dan peninjauan. Sedangkan pengertian sampel adalah himpunan bagian dari populasi yang dipilih peneliti untuk diobservasi (Harini, 2008:11). Dari definisi survei dan sampel tersebut dapat disimpulkan bahwa survei sampel merupakan salah satu metode pengumpulan data melalui sebagian unit dalam populasi dan hasilnya merupakan nilai-nilai perkiraan (estimasi).

Dapat dinyatakan bahwa tujuan dari survei sampel adalah untuk menggambarkan kesimpulan tentang populasi dari suatu informasi tertentu pada suatu sample. Satu cara untuk menarik kesimpulan adalah dengan memperkirakan parameter populasi tertentu dengan memanfaatkan informasi sampel. Estimasi rata-rata populasi dinotasikan dengan μ, dan total populasi dinotasikan dengan τ (Scheaffer, dkk., 1990: 59-62).

2.2 Penarikan Sampel Acak Sederhana

Untuk memperoleh sampel acak sederhana digunakan metode yang disebut metode penarikan sampel acak sederhana *(simple random sampling)*. Cara pemilihan sampel acak sederhana dapat dilakukan dengan melalui dua cara

a. Pemilihan sampel acak sederhana tanpa pengambalian: metode pemilihan sampel di mana elemen-elemen yang sudah terpilih tidak ditempatkan kembali untuk terpilih lagi *(without replacement).*

b. Pemilihan sampel acak sederhana dengan pengambalian: metode pemilihan sampel di mana elemen-elemen yang sudah terpilh ditempatkan kembali untuk bisa dipilih kembali *(with replacement)* (Supranto, 2009: 87-88).

2.2.1 Penarikan Sampel Acak Sederhana Tanpa Pengembalian

Penarikan sampel acak sederhana tanpa pengembalian atau *simple random sampling without replacement* (SRSWOR) adalah bentuk sampling paling familiar. Jenis sampel ini disebut sederhana karena melibatkan penggambaran seluruh populasi. Misalkan $U = \{1,2,3,..,N\}$, SRSWOR adalah metode pemilihan n elemen dari U sedemikian rupa sehingga semua kemungkinan himpunan bagian dari U berukuran n mempunyai kemungkinan yang sama untuk ditarik sebagai sampel. Dalam praktiknya SRSWOR dapat melibatkan berturut-turut dalam memilih nomer acak antara 1 dan N, dan termasuk setiap keterkaitan elemen populasi pada sampel elemen ini dipilih. Jika nomer baru sudah ditarik, nomer baru ditarik secara acak (Banning, dkk., 2012:6).

2.3 Pengertian Data Hilang

Dalam sensus maupun survei, seringkali ditemukan unit-unit yang tidak merespon sejumlah pertanyaan yang diajukan (non respon) (Malahayati, 2008:01). Kish (1965:535) mendefinisikan non respon sebagai kegagalan untuk mendapatkan nilai pengamatan dari beberapa unit yang menjadi sampel.

Non respon dalam beberapa literatur sering disebut dengan data hilang umumnya dibagi menjadi dua tipe yaitu unit non respon dan item non respon. Unit non respon terjadi karena unit sampel tidak memberikan respon sama sekali dalam suatu survei. Sedangkan item non respon dapat terjadi karena beberapa item dalam kuisioner tidak direspon oleh responden. Secara umum, non respon dapat disebabkan karena responden tidak mau menjawab, tidak mampu menjawab atau tidak tahu jawabannya, atau tidak ingin melanjutkan dengan wawancara atau sesuatu yang tidak ingin diungkapkan dengan pewancara. Non respon dapat juga terjadi karena kesalahan dalam penulisan jawaban atau dalam proses input data (Longford, 2005:13).

2.4 Imputasi Data

Imputasi adalah metode yang digunakan untuk memprediksi data hilang pada kumpulan data survei karena tidak adanya respon terhadap beberapa pertanyaan. Dalam metode imputasi ada dua prosedur yaitu imputasi tunggal dan imputasi ganda.

Imputasi tunggal yaitu mengisi nilai untuk setiap data yang hilang, dan merupakan metode yang paling umum untuk mengangani item non respon pada saat praktek survei. Metode imputasi ini mempunyai kelemahan yaitu, satu nilai yang digunakan untuk menggantikan data hilang ini tidak mencerminkan keragaman penarikan sampel nilai-nilai sebenarnya saat satu model untuk non respon terbentuk. Kelemahan yang lain, tidak dapat mencerminkan ketidak pastian saat terdapat lebih dari satu model untuk non respon. Kelemahan tersebut dapat diperbaiki dengan metode imputasi ganda.

Imputasi ganda adalah setiap data hilang kita dapat memasukkan beberapa nilai. Nilai-nilai m yang diperintahkan dalam arti bahwa kumpulan nilai pertama yang diperhitungkan untuk nilai-nilai yang hilang digunakan untuk membentuk kumpulan data lengkap pertama dan sebagainya. Dengan demikian imputasi m untuk setiap data hilang membuat m data yang lengkap. Dari masing-masing gugus data tersebut diterapkan metode analisis baku untuk data lengkap kemudian hasil dari analisis itu dirata-ratakan (Rubin, 1987:11-15).

Terdapat m nilai untuk setiap data hilang dan akhirnya akan membentuk m buah gugus data yang telah dilengkapi. Dari masing-masing gugus data tersebut diterapkan metode analisis baku untuk data lengkap kemudian hasil dari analisis tersebut dirata-ratakan (Malahayati, 2008:2).

2.5 Penaksiran Metode Imputasi Rasio

Dalam metode imputasi rasio suatu variabel pendukung x_i yang berhubungan dengan y_i diperoleh untuk setiap unit di dalam sampel. Dalam praktek, x_i sering kali nilai dari y_i pada beberapa waktu yang lalu ketika sensus lengkap dilakukan. Tujuan metode ini adalah untuk memperoleh peningkatan penelitian dengan mengambil manfaat hubungan antara y_i dan x_i. Sekarang kita menganggap penarikan sampel acak sederhana.

Perkiraan rasio untuk Y, jumlah populasi y_i adalah

$$\hat{Y}_R = \frac{y}{x}X = \frac{\bar{y}}{\bar{x}}X \qquad \text{(Cochran, 2010:173).} \qquad (2.1)$$

Jika dalam kasus imputasi nilai tunggal, unit yang membutuhkan imputasi, nilai b_{x_i} diimputkan. Dimana $\hat{b} = \frac{\bar{y}_r}{\bar{x}_r}$, sehingga data setelah dilakukan imputasi mempunyai bentuk

$$y_{.i} = \begin{cases} y_i & jika \quad i \in A \\ \hat{b}_{x_i} & jika \quad i \in \bar{A} \end{cases} \tag{2.2}$$

dimana A dan $\bar{A}$ menunjukkan respon dan nonrespon suatu survei. Metode imputasi di atas disebut imputasi rasio. Kemudian diberikan rata-rata penaksir titik populasi:

$$\begin{aligned} \bar{y}_s &= \frac{1}{n} \\ &= \frac{1}{n}\sum \hat{b}x_i \\ &= \frac{1}{n}\sum \frac{\bar{y}_r}{\bar{x}_r} x_i \\ &= \frac{\sum x_i}{n}\frac{\bar{y}_r}{\bar{x}_r} = \frac{\sum x_i}{n}\frac{\bar{y}_r}{\bar{x}_r} \\ &= \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right) = \bar{y}_{rat} \end{aligned} \tag{2.3}$$

menjadi:

$$\bar{y}_{rat} = \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right) \tag{2.4}$$

dimana $\bar{x}_n = n^{-1}\sum_{i=1}^{n} x_i$, $\bar{x}_r = r^{-1}\sum_{i=1}^{r} x_i$ dan $\bar{y}_r = r^{-1}\sum_{i=1}^{r} y_i$. Akhiran rat adalah kepanjangan dari estimator rasio sedangkan akhiran s kepanjangan dari rata-rata sampel.

Berdasarkan metode rata-rata imputasi, data setelah dilakukan imputasi mengambil bentuk:

$$y_{.i} = \begin{cases} y_i & jika \quad i \in A \\ \bar{y}_r & jika \quad i \in \bar{A} \end{cases} \tag{2.5}$$

dan titik estimator (2.4) menjadi:

$$\bar{y}_m = \frac{1}{r}\sum_{i=1}^{r} y_i = \bar{y}_r \tag{2.6}$$

Berdasarkan metode imputasi regresi, data setelah dilakukan imputasi regresi mempunyai bentuk:

$$y_{.i} = \begin{cases} y_i, & jika\ i \in A \\ \bar{y}_r + \hat{\beta}(x_i - \bar{x}_r), & jika\ i \in \bar{A} \end{cases} \tag{2.7}$$

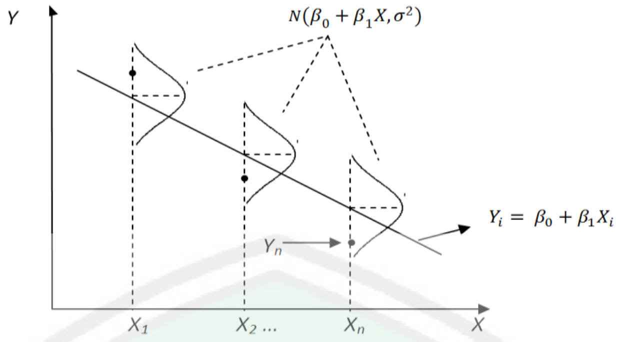
dimana $\hat{\beta} = \frac{s_{xy}}{s_x^2}$, dengan $s_{xy} = (r-1)^{-1}\sum_{i=1}^{r}(x_i - \bar{x}_r)(y_i - \bar{y}_r)$, $s_x^2 = (r-1)^{-1}\sum_{i=1}^{r}(x_i - \bar{x}_r)^2$ dan titik penaksir (2.4) menjadi:

$$\bar{y}_{reg} = \bar{y}_r + \hat{\beta}(\bar{x}_n - \bar{x}_r) \tag{2.8}$$

dimana akhiran reg kepanjangan dari penaksir regresi (Singh dan Valdes, 2009:1729-1730).

2.6 Analisis Regresi

Menurut Sumarningsih (2010:04) analisis regresi adalah analisis yang digunakan untuk mengertahui dan mempelajari suatu model hubungan fungsional linier antara peubah respon (Y) dan peubah penjelas (X). Peubah respon adalah peubah yang nlai-nilainya ditentukan berdasarkan nilai-nilai dari satu atau lebih peubah penjelas. Peubah penjelas adalah peubah yang nilai-nilainya dapat ditentukan, diatur dan yang nilainya dapat diamati. Asumsi yang melandasi model regresi adalah $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, dengan $i = 1, 2, \dots, n$ adalah $\varepsilon_i \sim NID\ (0, \sigma^2)$.



Gambar 2.1 Grafik Model Regresi
(sumber: bahan ajar perkuliahan regresi Universitas Brawijaya)

Bentuk umum persamaan linier sederhana yang menunjukkan hubungan antara dua variabel, yaitu variabel x sebagai variabel independent dan variabel Y sebagai variabel dependent adalah $Y = a + bX$. Y adalah variabel dependent, a adalah intersep titik potong kurva terhadap sumbu Y, b adalah kemiringan *(slope)* kurva linier, dan X adalah variabel independent. Persamaan $Y = a + bX$ dapat digunakan untuk menaksir nilai Y jika nilai a, b dan X diketahui. Nilai a merupakan nilai Y yang dipotong oleh kurva linier pada sumbu vertikal Y. atau dengan kata lain, a adalah nilai Y jika $X = 0$. Nilai b adalah kemiringan *(slope)* kurva linier yang menunjukkan besarnya perubahan bilai Y sebagai akibat dari perubahan setiap unit nilai X (Algifari, 2000:9).

Regresi $\hat{Y} = a + bX$ yang diperoleh menggunakan n pasang data sampel (X_i, Y_i) diharapkan bisa “mengambil alih” peran regresi dalam populasi yang memiliki persamaan berbentuk $\hat{Y} = \alpha + bX$ dengan harga-harga α dan β tidak diketahui dan masing-masing ditaksir oleh a dan b; koefisien α ditaksir oleh a dan koefisien regresi atau *bobot regresi* β ditaksir oleh b. Ada pengalihan lain mengenai kegunaan dan hubungan antara bobot regresi b dan bobot regresi β.

Sementara kita tahu bahwa b dihitung menggunakan $\frac{n\sum XY-(\sum X)(\sum Y)}{n\sum X^2-(\sum X)^2}$ maka bobot regresi β didefinisikan oleh $\beta = b\left(\frac{S_x}{S_y}\right)$ dengan S_x = simpangan baku untuk X dan S_y = simpangan baku untuk Y (Sudjana, 1992:6-12).

2.6.1 Penaksiran Regresi

Seperti pada penaksiran rasio, penaksiran regresi linear dibuat untuk meningkatkan ketelitian dengan menggunakan variabel tambahan x_i yang berkolerasi dengan y_i. Bila hubungan antara y_i dan x_i diuji, mungkin ditemukan bahwa walaupun hubungan mendekati linier, garisnya tidak melalui titik origin. Hasil ini menyarankan suatu perkiraan yang didasarkan pada regresi linear dari y_i pada x_i lebih baik daripada rasio dua variabel.

Kita misalkan bahwa y_i dan x_i masing-masing diperoleh untuk setiap unit dalam sampel dan rata-rata populasi $\bar{X}$ dari x_i diketahui. Penaksiran regresi linear $\bar{Y}$, rata-rata populasi y_i, adalah

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x}) \quad (2.9)$$

Dimana notasi lr menyatakan regresi linear dan b adalah koefisien perkiraan dari perubahan dalam y bila x meningkat. Alasan utama dari penaksiran ini adalah jika $\bar{x}$ di bawah rata-rata, kita harus mengira $\bar{y}$ juga dibawah rata-rata dari suatu jumlah $b(\bar{X} - \bar{x})$ karena regresi dari y_i pada x_i.

Meskipun dalam banyak aplikasi, b diperkirakan dari hasil sampel, kadang-kadang beralasan juga untuk memilih nilai b lebih dulu. Pada survei-survei yang dilakukan berulang, perhitungan-perhitungan sebelumnya mungkin dapat menunujukkan bahwa nilai sampel b tetap konstan atau bila x adalah nilai y pada sensus terbaru, pengetahuan umum tentang populasi dapat menyarankan bahwa b tidak jauh dari satu, sehingga $b = 1$ dipilih (Cochran, 2010:218).

2.7 *Mean Square Error* (MSE)

Rata-rata kesalahan kuadrat atau sering disebut dengan *Mean Square Error* (MSE) merupakan suatu *estimator* $\hat{\theta}$ dari sebuah parameter θ adalah fungsi dari θ yang telah didefinisikan dengan $E\left(\hat{\theta}-\theta\right)^2$, dilambangkan sebagai $\text{MSE}_{\hat{\theta}}$.

MSE mengukur selisih rata-rata kuadrat antara estimator $\hat{\theta}$ dan parameter θ, suatu ukuran yang sedikit pantas dari kinerja untuk suatu *estimator*. Menurut Songfeng Zheng, pada umumnya untuk peningkatan fungsi jarak absolute $|\hat{\theta} - \theta|$akan berfungsi mengukur kebaikan dari *estimator* (rata-rata kesalahan mutlak, $E\left(|\hat{\theta}-\theta|\right)$ adalah suatu alternatif yang masuk akal). MSE memiliki dua keunggulan dibanding ukuran jarak lain: pertama, cara analitik yang mudah dikerjakan dan kedua, mempunyai tafsiran.

$$MSE_{\hat{\theta}} = E\left(\hat{\theta}-\theta\right)^2 = Var\left(\hat{\theta}\right) + \left(E\left(\hat{\theta}\right)-\theta\right)^2 = Var\left(\hat{\theta}\right) + \left(Bias\ dari\ \hat{\theta}\right)^2$$

$$\begin{aligned} E\left(\hat{\theta}-\theta\right)^2 &= E\left(\hat{\theta}^2\right) + E\left(\hat{\theta}\right) - 2\theta E\left(\hat{\theta}\right) \\ &= \operatorname{var}\left(\hat{\theta}\right) + \left[E\left(\theta\right)\right]^2 + \theta^2 - 2\theta E\left(\hat{\theta}\right) \\ &= \operatorname{var}\left(\hat{\theta}\right) + \left[E\left(\hat{\theta}\right)-\theta\right]^2 \end{aligned} \quad (2.10)$$

2.8 Sebaran Binomial

Suatu percobaan statistik disebut percobaan Binomial atau Bernoulli jika percobaan statistik tersebut mempunyai ciri-ciri sebagai berikut:

1. Percobaan diulang sebanyak n kali,
2. Setiap hasil pecobaan dibedakan menjadi dua, yaitu kejadian sukses (S) dan kejadian gagal (G),
3. Probabilitas terjadi kejadian sukses (S) dan gagal (G), yaitu yaitu $P(\text{sukses}) = P(S) = p$ dan $P(\text{gagal}) = P(G) = 1 - p = q$, adalah tetap pada tiap kali percobaan diulang, dan
4. Semua hasil yang muncul saling bebas satu sama lain (Boediono dan Koster, 2004: 306).

Apabila percobaan sebanyak n kali, atau pengamatan berukuran n orang, kita mempunyai peubah acak w dengan nilai pengamatan $w_1, w_2, w_3, .., w_n$. Dimana $w_i = 1$ jika hasil sebagaimana yang dimaksud dan 0 jika hasilnya bukan yang dimaksud.

Jika semua w_i bernilai 1 atau 1, 1, 1, 1, …, 1 sebanyak n kali, maka

$$X = \sum_{i=1}^{n} w_i = n \qquad (2.11)$$

dan peluang untuk $X = n$ ini adalah

$$P(X = n) = P(1 \text{ dan } 1 \text{ dan } 1 \text{ dan } ... \text{ dan } 1)$$

$$= P(w = 1)\, P(w = 1)\, P(w = 1) .. P(w = 1)$$

$$= p \; x \; p \; x \; p \; x \; ... \; x \; p$$

$$= pn$$

$$= pn(1 - p)^o$$

$$= pn(1 - p)^{n-n}$$

Untuk $X = r(r < n)$ mempunyai (salah satunya) adalah

1 1 1 1 1…1 (sebanyak r kali)

0 0 0 0 0…0 (sebanyak $(n-r)$ kali)

dan ada sebanyak $\binom{n}{r}$ susunan yang mempunyai nilai 1 sebanyak r dan 0 sebanyak $(n-r)$ tersebut. Oleh karena itu

$$P(X=r)=\binom{n}{r}p^r\,(1-p)^{n-r} \qquad (2.12)$$

Ini merupakan fungsi peluang, atau tepatnya fungsi sebaran peluang. Karena berdasarkan atas percobaan Binomial (atau Bernoulli), maka disebut fungsi peluang Binomial atau apabila dikaitkan dengan peubah X itu sendiri disebut sebaran Binomial (Yitnusumarto, 1988:138-141).

2.9 Relatif *Error*

Dalam beberapa situasi relatif *error* berguna untuk mempertimbangkan beberapa ukuran relatif bukan ukuran mutlak variasi. Ukuran mutlak, *standard deviasi* dan *standard error*, muncul dalam unit pengukuran variabel, dan ini menyebabakan kesulitan dalam beberapa perbandingan. Ukuran relatif adalah koefisien variansi, dimana unit pengukuran dibatalkan dengan membagi dengan rata-rata. Elemen koefisien variansi diperoleh dari *standard deviasi*:

$$C_y=\frac{S_y}{\bar{Y}},\ \text{ditaksir dengan } c_y=\frac{s_y}{\bar{y}} \qquad (2.13)$$

Koefisien variasi rata-rata $(\bar{y})$ diperoleh dengan cara yang sama dari *standard error:*

$$CV(\bar{y})=\frac{SE(\bar{y})}{\bar{Y}}\text{, diestimasi dengan } cv(\bar{y})=\frac{se(\bar{y})}{\bar{y}} \qquad (2.14)$$

Kuadrat jumlah koresponden ini berturut-turut dengan variasi dari elemen dan rata-rata

$C_y^2 = \frac{S_y^2}{\bar{Y}^2}$, ditaksir dengan $c_y^2 = \frac{s_y^2}{\bar{y}^2}$ (Kish, 1965: 47). (2.15)

2.10 Harapan dan Momen

Definisi 2.1 Bagi suatu peubah acak X didefinisikan harapannya [EX atau $E(X)$] sebagai $EX = \int_{-\infty}^{\infty} x f_x(x) dx$ [$EX = \sum x_i px(x_i)$] bila X kontinu mutlak dengan fungsi padat $f_X(x)$ [bila X diskret dengan fungsi peluang $P_X(x)$], asal saja integral (jumlah) ini ada dan terhingga. Bila X suatu p.a, maka EX ada jika dan hanya jika $\int_0^{\infty} x f_X(x) dx$ dan $\int_{-\infty}^{0} x f_X dx$ terhingga bila X kontinu mutlak $\sum_{x_i<0} x_i P_X(x_i)$ berhingga bila X diskret, dalam hal itu $EX = \int_{-\infty}^{\infty} x f_X(x) dx$ [$EX = \sum x_i P_X(x_i)$] (Dudewicz dan Mishra, 1995:246-247).

Teorema 2.1 Sifat harapan bila c suatu tetapan dan $g(X), g_1(X),$ dan $g_2(X)$ fungsi dari peubah acak X yang harapannya ada, maka

1. $E[c] = c$;
2. $E[cg(X)] = cE[g(X)]$;
3. $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$ (Dudewicz dan Mishra, 1995:249).

Bukti:

Misalkan fungsi massa peluang di X adalah $P_X(x)$

1. $$E[c] = \sum_{i=1}^{n} c p_X(x_i)$$
$$= c\sum_{i=1}^{n} p_X(x_i)$$
$$= c$$

2. $$E\left[cg(X)\right]=\sum_{i=1}^{n}cg\left(x_i\right)p_X\left(x_i\right)$$
$$=c\sum_{i=1}^{n}g\left(x_i\right)p_X\left(x_i\right)$$
$$=cE\left[g(X)\right]$$

3. $$E\left[g_1(X)+g_2(X)\right]=\sum_{i=1}^{n}(g_1(x_i)+g_2(x_i))p_X(x_i)$$
$$=\sum_{i=1}^{n}g_1(x_i)p_X(x_i)+\sum_{i=1}^{n}g_2(x_i)p_X(x_i)$$
$$=E\left[g_1(X)\right]+E\left[g_2(X)\right]$$

Definisi 2.2 Tuliskanlah $\sigma^2(X)$ hanya sebagai σ^2 (variansi). Maka σ (akar positif dari σ^2) disebut simpangan baku dari X dan sering dituliskan sebagai $\sigma\ (X)$.

Teorema 2.2 $Var(X) = EX^2 - (EX)^2$

Bukti:

$$Var(X)=E(X-EX)^2$$
$$=E\{X^2-2XEX+(EX)^2\}$$
$$=EX^2-2(EX)^2+(EX)^2$$
$$=EX^2-(EX)^2$$

Definisi 2.3 Misalkan (X_1, X_2) suatu p.a bermatra 2. Untuk setiap n_1, n_2 (bilangan bulat tak negatif) didefinisikan $\mu_{n_1,n_2} = E\{(X_1 - EX_1)^{n_1}(X_2 - EX_2)^{n_2}\}$ (bila harapan ini ada). Ini disebut momen pusat gabungan ordo $(n_1 + n_2)$ dari (X_1, X_2).

Contoh: Misalkan (X_1, X_2) suatu p.a bermatra 2. Maka $\mu_{1,0} = \mu_{0,1} = 0, \mu_{2,0} = Var(X_1)$, $\mu_{0,2} = Var\ (X_2), \mu_{1,1} = E\{(X_1 - EX)(X_2 - EX_2)\}$. Perhatikan bahwa $\mu_{1,1}$ disebut kovariansi dari X_1 dan X_2, dinyatakan dengan Kov (X_1, X_2).

Teorema 2.3 Kov $(X_1, X_2) = E(X_1X_2) - EX_1EX_2$

Bukti:

$$
\begin{aligned}
Kov(X_1, X_2) &\equiv E\{(X_1 - EX_1)(X_2 - EX_2)\} \\
&= E\{X_1X_2 - X_2EX_1 - X_1EX_2 + EX_1EX_2\} \\
&= E(X_1X_2) - 2EX_2EX_1 + EX_1EX_2 \\
&= E(X_1X_2) - EX_1EX_2
\end{aligned}
$$

(Dudewicz dan Mishra, 1995: 273).

2.11 Variansi Penaksiran

Variansi y_i dalam sebuah populasi terbatas biasanya ditetapkan sebagai

$$\sigma^2 = \frac{\sum_1^N (y_i - \bar{Y})^2}{N} \tag{2.16}$$

Dengan sedikit perluasan pada notasi, pembagian N diganti menjadi $(N-1)$. diperoleh

$$S^2 = \frac{\sum_1^N (y_i - \bar{Y})^2}{N-1} \tag{2.17}$$

Perluasan ini biasanya dipakai oleh mereka yang memakai teori penarikan sampel dengan maksud menganalisis varians. Sekarang perhatikan variansi $\bar{y}$, yang dimaksud adalah $E(\bar{y} - \bar{Y})^2$ yang diperoleh untuk seluruh ${}_N C_n$ sampel.

Teorema 2.4. Variansi dari rata-rata $\bar{y}$ dari sampel acak sederhana adalah

$$V(\bar{y}) = E(\bar{y} - \bar{Y})^2 = \frac{S^2}{n}\frac{(N-n)}{N} = \frac{S^2}{n}(1-f) \tag{2.18}$$

Dimana $f = n/N$ adalah fraksi penarikan sampel

Bukti.

$$n(\bar{y} - \bar{Y}) = (y_1 - \bar{Y}) + (y_2 - \bar{Y}) + \cdots + (y_n - \bar{Y}) \tag{2.19}$$

$$E[(y_1 - \bar{Y})^2 + \cdots + (y_n - \bar{Y})^2] = \frac{n}{N}[(y_1 - \bar{Y})^2 + \cdots + (y_N - \bar{Y})^2] \tag{2.20}$$

dan juga bahwa

$E[(y_1-\bar{Y})(y_2-\bar{Y})+(y_1-\bar{Y})(y_3-\bar{Y})+\cdots+(y_{n-1}-\bar{Y})(y_n-\bar{Y})]=$

$\frac{n(n-1)}{N(N-1)}[(y_1-\bar{Y})(y_2-\bar{Y})+(y_1-\bar{Y})(y_3-\bar{Y})+\cdots+(y_{N-1}-\bar{Y})(y_N-$

$\bar{Y})]$ (2.21)

Pada (2.21) jumlahnya terdiri dari seluruh pasangan unit-unit dalam sampel dan populasi. Penjumlahan di kiri terdiri atas $\frac{n(n-1)}{2}$ suku dan di kanan terdiri atas $\frac{N(N-1)}{2}$ suku. Sekarang (2.19) dikuadratkan dan rata-ratakan seluruh sampel acak sederhana. Dengan menggunakan rumus (2.20) dan (2.21) kita peroleh

$$n^2E(\bar{y}-\bar{Y})^2=\frac{n}{N}\Big\{(y_1-\bar{Y})^2+\cdots+(y_N-\bar{Y})^2 \\ +\frac{2(n-1)}{(N-1)}[(y_1-\bar{Y})(y_2-\bar{Y})+\cdots+(y_{N-1}-\bar{Y})(y_N-\bar{Y})]\Big\}$$

Kuadrat selengkapnya atas perkalian silangnya, kita dapatkan

$$n^2E(\bar{y}-\bar{Y})^2=\frac{n}{N}\Big\{\Big(1-\frac{n-1}{N-1}\Big)[(y_1-\bar{Y})^2+\cdots+(y_N-\bar{Y})^2] \\ +\frac{(n-1)}{(N-1)}[(y_1-\bar{Y})+\cdots+(y_N-\bar{Y})]^2\Big\}$$

Suku kedua dalam tanda kurung akan hilang karena jumlah dari y_i sama dengan $N\bar{Y}$. Setelah dibagi n^2 menjadi

$V(\bar{y})=E(\bar{y}-\bar{Y})^2=\frac{N-n}{nN(N-1)}\sum_{i=1}^{N}(y_i-\bar{Y})^2=\frac{S^2}{n}\frac{(N-n)}{N}$ (Cochran, 2010:27-28).

Rumus kesalahan baku dari estimasi rata-rata populasi dan jumlah populasi digunakan terutama untuk tiga tujuan: (1) membandingkan ketelitian yang diperoleh dari penarikan sampel acak sederhana dengan metode penerikan sampel lainnya, (2) untuk memperkirakan ukuran sampel yang dibutuhkan dalam survei yang telah direncanakan, dan (3) untuk memperkirakan ketelitian sebenarnya yang didapat dalam suatu survei yang telah dilaksanakan. Rumus-rumusnya mencakup S^2, variansi populasi (Cochran, 2010:30).

2.12 Allah Menghitung Segala Sesuatu yang Dilakukan oleh Manusia

Dalam melakukan survei untuk mendapatkan hasil analisis yang valid data yang diperoleh harus lengkap. Jika ada beberapa data yang tersembunyi atau hilang maka secara otomatis akan mempengaruhi hasil dari penelitian yang dilakukan oleh surveior. Ketika melakukan survei, surveior harus mengetahui keadaan data yang diperoleh, artinya keadaan data harus selalu dihitung dan diawasi oleh para surveior. Hal ini sesuai dengan firman Allah SWT sebagai berikut. Allah juga menghitung segala sesuatu dan setiap yang dilakukan oleh manusia.

وَكُلَّ شَيْءٍ أَحْصَيْنَٰهُ فِيٓ إِمَامٍ مُّبِينٍ ١٢

Artinya: *"……dan segala sesuatu Kami kumpulkan dalam kitab Induk yang nyata (Lauh Mahfuzh)"*(Qs. Yaasin: 12).

Ayat ini menjelaskan segala sesuatu yang dilakukan oleh manusia akan dikumpulkan dalam suatu catatan yang nyata di Lauhul Mahfuzh. Segala sesuatu yang ada di dunia ini tidak luput dari penglihatanNya.

Qatadah berkata, *"Maknanya adalah menghitung setiap amal"*. Demikian juga yang dikatakan oleh Mujahid dan ibnu zaid. Ini sama dengan firman Allah SWT, عَلِمَتْ نَفْسٌ مَّا قَدَّ مَتْ وَأَخَّرَتْ *"Maka tiap-tiap jiwa akan mengetahui apa yang telah dikerjakan dan dilalaikan (Qs. Al-Infithaar: 5)"*. Dan juga firman Allah يُنَبَّؤُاالإِنْسنُ يَوْمَئِذٍ بِما قَدَّمَ وَأَخَّرَ *"Pada hari itu diberitakan kepada manusia apa yang* *telah dkerjakan dan apa yang dilalaikannya (Qs. Al-Qiyaamah: 13)".* Jadi apa yang telah dilakukan oleh seseorang di masa lalu, baik yang berupa kebaikan maupun keburukan, setiap tradisi baik maupun tradisi buruk mendapatkan balasan (Al-Qurthubi, 2008b:920-922).

Kasus *missing data* sering dijumpai ketika melakukan survei. Sehingga untuk menghindari hal ini, para surveior diharapkan untuk menghitung data dan mengetahui keadaan data. Jadi ketika terjadi kasus seperti ini akan segera diketahui dan dicari solusi untuk mengatasinya.

BAB III

PEMBAHASAN

3.1 Tahap Ilustrasi Metode Regresi untuk Imputasi

Ketika melakukan survei sering ditemukan kasus data hilang. Data hilang disebabkan ketika surveior mengajukan beberapa pertanyaan pada responden seringkali ditemukan responden yang tidak menjawab pertanyaan yang telah diajukan, sehingga mengakibatkan pendugaan parameter menjadi tidak efisien karena ukuran data berkurang dan menyebabkan kesulitan dalam menganalisis data. Kasus seperti ini dinamakan non respon. Untuk mengatasi permasalahan tersebut dilakukan imputasi data.

Beberapa metode imputasi data adalah metode imputasi mean, rasio dan regresi. Pada suatu kasus ketika melakukan survei akan ditemukan data dengan 2 variabel x dan y yang tidak memperhatikan hubungan atau memperhatikan hubungan, jenis data seperti ini dapat diatasi menggunakan ketiga metode tersebut. Pertama-tama akan dicoba diimputasi menggunakan metode imputasi mean. Data y di sini sebagai data non respon ditaksir menggunakan metode mean yaitu dengan menjumlahkan data respon (yang tidak hilang) kemudian merata-ratakannya ($\bar{y}_r$).

Akan tetapi jika data tersebut mempunyai kelipatan, kasus seperti ini dapat diatasi menggunakan metode rasio dan regresi. Data yang berkelipatan lebih diutamakan menggunakan metode imputasi rasio. Hal ini dengan alasan, jika diatasi lagi menggunakan metode *mean*, memberikan informasi variansi kurang bagus dan tidak ada unsur x yang dapat meminumumkan variansinya sedangkan jika menggunakan metode regresi, $\bar{y}$ menjadi nol sama halnya dengan rasio atau akan turun menjadi metode imputasi rasio kembali. Sehingga dari kelemahan metode imputasi mean ini dikembangkan metode rasio yang memperhatikan variabel x. Cara kerja untuk mendapatkan nilai dari data hilang dengan metode rasio adalah mengalikan nilai rata-rata dari y respon dengan rata-rata data x penuh ($\bar{x}_n$) dibagi data x respon ($\bar{x}_r$).

Selanjutnya ketika dihadapkan pada kondisi data saling berhubungan dan tidak kelipatan, jika diatasi menggunakan metode imputasi rasio kembali, nilai dari data y aslinya akan hilang dan hanya kelipatan dari nilai data x saja tidak mengambil dari data y. Dari permasalahan ini, metode imputasi regresi digunakan, selain mengatasi data yang saling berhubungan dan tidak berkelipatan, variansi yang didapat lebih bagus. Maka dari sini penulis menggunakan metode regresi untuk mengatasi permasalahan data hilang.

Dari pernyataan di atas akan diberikan ilustrasi data hilang di mana terdapat data pengamatan X dan Y yang saling berhubungan dan tidak berkelipatan pada tabel (3.1) dan akan dilakukan imputasi regresi.

Tabel 3.1 Ilustrasi Data Non Rrespon pada Data y_4, y_5, y_6, y_7, dan y_8

Responden	X	Y
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
4	x_4	…
5	x_5	…
6	x_6	…
7	x_7	…
8	x_8	…
9	x_9	y_9
10	x_{10}	y_{10}

Misalnya data Y adalah variabel tidak bebas dan data X adalah variabel bebas kemudian ada beberapa data yang hilang dari data Y maka model kasus seperti ini dapat ditaksir menggunakan metode imputasi regresi. Imputasi regresi ini berguna untuk menaksir parameter dari nilai yang hilang dengan menginputkan rata-rata nilai y respon dari data Y yang disimbolkan dengan notasi $\bar{y}_r$ dimana $r = 1, 2, 3, \dots n$, kemudian menjumlahkan taksiran yang melibatkan data hilang dimana berhubungan langsung dengan data x ke-i dikurangi dengan data x respon yang sejajar dengan data y respon.

3.2 Tahap Analisis Metode Regresi untuk Imputasi

3.2.1 Pendefinisian Model Regresi untuk Imputasi

Berdasarkan batasan penelitian ini, model regresi imputasi yang akan digunakan adalah model dalam jurnal Singh & Valdes (2009) berjudul *Optimal Method of Imputation in Survey Sampling*. Model ini merupakan pengembangan dari model (2.2), sehingga memperoleh model sebagai berikut:

$$y_{.i} = \begin{cases} y_i, & \text{jika i} \in \text{A} \\ \bar{y}_r + \hat{\beta}(x_i - \bar{x}_r), & \text{jika i} \in \bar{\text{A}} \end{cases} \tag{3.1}$$

Dimana A dan $\bar{A}$ menunjukkan respon dan nonrespon suatu survei. Bentuk data ini menggunakan model perkiraan regresi linear. Perkiraan regresi linear ini dibuat untuk meningkatkan ketelitian dengan menggunakan variabel tambahan x_i yang berkolerasi dengan y_i. Bila hubungan antara y_i dan x_i diuji, ditemukan bahwa walaupun hubungan mendekati linier, garisnya tidak melalui titik origin. Hasil ini menyarankan suatu perkiraan yang didasarkan pada regresi linear dari y_i pada x_i lebih baik daripada rasio dua variabel (Cochran, 1991:216).

3.2.2 Mentaksir $\hat{\boldsymbol{\beta}}$ Model Regresi untuk Imputasi

Taksiran $\hat{\beta}$ diperoleh dari hasil penjabaran model regresi imputasi data. Untuk menaksir data yang tidak hilang maka menggunakan r, dengan $i = 1, 2, 3, \dots, r$ dengan model duga regresi sebagai berikut

$y_i = \bar{y}_r + \beta(x_i - \bar{x}_r) + \varepsilon$ atau dapat ditulis

$\varepsilon = y_i - [\bar{y}_r - \hat{\beta}(x_i - \bar{x}_r)]$

dimisalkan$[\bar{y}_r - \hat{\beta}(x_i - \bar{x}_r)] = a$

$$S = \sum \varepsilon^2 = \sum (y_i - a)^2$$

$$= \sum (y_i^2 - 2y_i a + a^2)$$
$$= \sum y_i^2 - 2\sum y_i a + \sum a^2$$
$$= \sum y_i^2 - 2\sum y_i(\bar{y}_r - \hat{\beta}(x_i - \bar{x}_r)) + \sum (\bar{y}_r - \hat{\beta}(x_i - \bar{x}_r))^2$$
$$= \sum y_i^2 - 2\sum y_i \bar{y}_r + 2\hat{\beta}\sum y_i(x_i - \bar{x}_r) + \sum (\bar{y}_r - \hat{\beta}(x_i - \bar{x}_r))^2$$

Nilai $\hat{\beta}$ didapatkan dengan menggunakan metode kuadrat terkecil yaitu metode penduga dengan meminimumkan jumlah kuadrat galat (S):

$$\frac{dS}{d\hat{\beta}} = \frac{d\sum_{i=1}^{r}\varepsilon^2}{d\hat{\beta}} = \sum y_i(x_i - \bar{x}_r) - 2\sum_{i=1}^{r}(x_i - \bar{x}_r)\left(\bar{y}_r - \hat{\beta}(x_i - \bar{x}_r)\right)$$

S akan mempunyai nilai minimum jika turunan terhadap $\hat{\beta}$ sama dengan nol.

$$\sum y_i(x_i - \bar{x}_r) - 2\sum_{i=1}^{r}(x_i - \bar{x}_r)\left(\bar{y}_r - \hat{\beta}(x_i - \bar{x}_r)\right) = 0$$

$$\sum_{i=1}^{r}(x_i y_i - \bar{x}_r y_i) - \sum_{i=1}^{r}(x_i\bar{y}_r - \bar{x}_r\bar{y}_r) + \hat{\beta}\sum_{i=1}^{r}(x_i - \bar{x}_r)^2 = 0$$

$$\hat{\beta}\sum_{i=1}^{r}(x_i - \bar{x}_r)^2 = \sum_{i=1}^{r}(x_i y_i - \bar{x}_r y_i) + \sum_{i=1}^{r}(x_i\bar{y}_r - \bar{x}_r\bar{y}_r)$$

$$\hat{\beta}\sum_{i=1}^{r}(x_i - \bar{x}_r)^2 = \sum_{i=1}^{r}(x_i - \bar{x}_r)(y_i - \bar{y}_r)$$

$$\hat{\beta} = \frac{\sum_{i=1}^{r}(x_i - \bar{x}_r)(y_i - \bar{y}_r)}{\sum_{i=1}^{r}(x_i - \bar{x}_r)^2}$$

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

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

$$= s_{xy}\,\frac{1}{s_x^2}$$

$$= \frac{s_{xy}}{s_x^2}$$

Sehingga $\hat{\beta} = \frac{s_{xy}}{s_x^2}$, dengan $s_{xy} = (r-1)^{-1}\sum_{i=1}^{r}(x_i - \bar{x}_r)(y_i - \bar{y}_r), s_x^2 = (r-1)^{-1}\sum_{i=1}^{r}(x_i - \bar{x}_r)^2$. $\hat{\beta}$ disini untuk menaksir y_{imp}. Persamaan $\bar{y}_s$ (2.3) menjadi rata-rata metode regresi imputasi ($\bar{y}_{reg}$)dengan definisi $y_{.i}$ akan menggunakan persamaan model data setelah dilakukan imputasi regresi jika $i \in \bar{A}$ (Singh dan Valdes, 2009).

3.2.3 Rata-rata Regresi Imputasi

Selanjutnya akan dijabarkan titik estimator (2.3) untuk rata-rata metode imputasi regresi

$$
\begin{aligned}
\bar{y}_s &= \frac{1}{n}\sum_{i=1}^{n} y_{.i} \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(\bar{y}_r + \hat{\beta}\left(x_i - \bar{x}_r\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\bar{y}_r + \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}\left(x_i - \bar{x}_r\right) \\
&= \frac{n\bar{y}_r}{n} + \hat{\beta}\sum_{i}\left(\frac{x_i - \bar{x}_r}{n}\right) \\
&= \bar{y}_r + \hat{\beta}\left[\frac{\sum_i x_i - \sum_i \bar{x}_r}{n}\right] \\
&= \bar{y}_r + \hat{\beta}\left(\frac{\sum_{i=1}^{n} x_i}{n} - \frac{n\bar{x}_r}{n}\right) \\
&= \bar{y}_r + \hat{\beta}\left(\bar{x}_n - \bar{x}_r\right)
\end{aligned}
$$

Sehingga titik perkiraan (2.3) menjadi

$$\bar{y}_{reg} = \bar{y}_r + \hat{\beta}\left(\bar{x}_n - \bar{x}_r\right) \qquad (3.2)$$

Model imputasi regresi (3.2) didapat dari titik perkiraan (2.3). Model ini digunakan untuk mencari MSE dari metode imputasi regresi,sehingga dari estimator regresi ini (3.2) akan ditaksir MSE regresi.

3.2.4 Menentukan MSE dari Metode Regresi untuk Imputasi

Pada tahap ini akan diuraikan MSE dari model (3.2). Diberikan ε adalah *error* antara sampel respon y dan parameter populasi Y. δ adalah *error* antara sampel respon x dan parameter populasi X sedangkan η adalah *error* antara sampel x dan parameter populasi X.

$$\varepsilon = \frac{\bar{y}_r}{\bar{Y}} - 1, \qquad \delta = \frac{\bar{x}_r}{\bar{X}} - 1, \qquad \text{dan } \eta = \frac{\bar{x}_n}{\bar{X}} - 1$$

$$E(\varepsilon) = E(\delta) = E(\eta) = 0$$

Selanjutnya akan ditaksir nilai $E(\varepsilon^2),\ E(\delta^2),\ E(\varepsilon\delta),\ E(\eta^2),\ E(\delta\eta),$ dan $E(\varepsilon\eta)$

3.2.4.1 Menghitung nilai $E(\varepsilon^2)$:

$$\begin{aligned} Var(\varepsilon) &= E(\varepsilon - E(\varepsilon))^2 \\ &= E\{\varepsilon^2 - 2\varepsilon E\varepsilon + E(\varepsilon)^2\} \\ &= E(\varepsilon^2) - 2E(\varepsilon)^2 + E(\varepsilon)^2 \\ &= E(\varepsilon^2) - E(\varepsilon)^2 \end{aligned}$$

$$\begin{aligned} E(\varepsilon^2) &= Var(\varepsilon) + E(\varepsilon)^2 \\ &= Var(\varepsilon) + 0 \\ &= Var(\varepsilon) \end{aligned}$$

$$\begin{aligned} Var(\varepsilon) &= E(\varepsilon^2) \\ &= E\left[\left(\frac{y_r}{\bar{Y}} - 1\right)^2\right] \\ &= \frac{1}{\bar{Y}^2} Var(\bar{y}_r) \end{aligned}$$

Pandang $z_i = 1$ jika unit i termasuk dalam sampel r. $z_i = 0$ jika tidak termasuk dalam sampel r, sehingga z_i mengikuti distribusi Bernoulli, dengan

$$\begin{aligned} E(z_i) &= \sum_{z_i=0}^{1} z_i P(z_i = z_i) \\ &= 0.P(z_i = 0) + 1.P(z_i = 1) \\ &= 0.\left(\frac{N-r}{N}\right) + \frac{r}{N} \\ &= \frac{r}{N} \end{aligned}$$

Maka nilai dari $Var\ (z_i)$ adalah

$$\begin{aligned} Var(z_i) &= E(z_i^2) - (E[z_i])^2 \\ &= \frac{r}{N} - \left(\frac{r}{N}\right)^2 = \frac{r}{N}\left(1 - \frac{r}{N}\right) \end{aligned}$$

Oleh karena itu $\bar{y}_r$ dapat ditulis ulang menjadi

$$\bar{y}_r = \frac{1}{r}\sum_{i=1}^{N} y_i\, z_i$$

$$Var(\bar{y}_r) = Var\left[\frac{1}{r}\sum_{i=1}^{N} y_i z_i\right] = \frac{1}{r^2}\left[\sum_{i=1}^{N} y_i^2 \operatorname{var}(z_i) + \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j \operatorname{cov}(z_i, z_j)\right]$$

Untuk $i \neq j$

$$\begin{aligned} \operatorname{cov}(z_i, z_j) &= E(z_i, z_j) - E(z_i)E(z_j) \\ &= P(z_i = 1, z_j = 1) - \left(\frac{r}{N}\right)\left(\frac{r}{N}\right) \end{aligned}$$

Peluang dua unit spesifik berada di dalam sampel adalah:

$$\frac{\binom{N-2}{r-2}}{\binom{N}{r}} = \frac{\dfrac{(N-2)!}{(r-2)!(N-r)!}}{\dfrac{N!}{(N-r)!r!}}$$

$$= \frac{(N-2)!}{(r-2)!(N-r)!} x \frac{(N-r)!r!}{N!}$$

$$= \frac{(N-2)!r(r-1)(r-2)!}{(r-2)!N(N-1)(N-2)!}$$

$$= \frac{r(r-1)}{N(N-1)}$$

$$\begin{aligned}
\operatorname{cov}(z_i, z_j) &= \frac{r(r-1}{N(N-1)} - \left(\frac{r}{N}\right)^2 \\
&= \frac{r^2 - r}{N(N-1)} - \frac{r^2}{N^2}\frac{(N-1)}{(N-1)} \\
&= \frac{1}{N-1}\left[\frac{r^2 - r}{N} - \frac{r^2}{N^2}(N-1)\right] \\
&= \frac{1}{N-1}\left[\frac{(r^2 - r)N}{N^2} - \frac{r^2}{N^2}(N-1)\right] \\
&= \frac{1}{N-1}\left[\frac{Nr^2 - rN - r^2N + r^2}{N^2}\right] \\
&= \frac{1}{N-1}\left[\frac{-rN + r^2}{N^2}\right] \\
&= \frac{1}{N-1}\left[\frac{-r(N-r)}{N^2}\right] \\
&= \frac{-Var(z_i)}{N-1}
\end{aligned}$$

$$\begin{aligned}
var(\bar{y}_r) &= \frac{1}{r^2}\left[\sum_{i=1}^{N} y_i^2 \operatorname{var}(z_i) + \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j \operatorname{cov}(z_i, z_j)\right] \\
&= \frac{1}{r^2}\left[\sum_{i=1}^{N} y_i^2 \operatorname{var}(z_i) - \operatorname{var}(z_i)\frac{\sum_{i=1}^{N}\sum_{j\neq i} y_i y_j}{N-1}\right] \\
&= \frac{1}{r^2}\frac{\operatorname{var}(z_i)}{N-1}\left[(N-1)\sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j\right] \\
&= \frac{1}{r^2}\frac{r(N-r)}{N^2}\frac{1}{(N-1)}\left[(N-1)\sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j\right]
\end{aligned}$$

dimana

$$(N-1)\sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j = N\sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j$$

$$= N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j\right)$$

$$= N\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2$$

$$= N\sum_{i=1}^{N} y_i^2 - \left(N\bar{Y}\right)^2$$

$$= N\sum_{i=1}^{N} \left(y_i - \bar{Y}\right)^2$$

$$= \frac{(N-1)}{(N-1)} N\sum_{i=1}^{N} \left(y_i - \bar{Y}\right)^2$$

$$= N(N-1)S_y^2$$

sehingga

$$\mathrm{var}(\bar{y}_r) = \frac{r(N-r)}{N^2}\frac{1}{r^2}\frac{1}{(N-1)}N(N-1)S_y^2$$

$$= \frac{(N-r)}{Nr}S_y^2$$

$$= \left(\frac{1}{r} - \frac{1}{N}\right)S_y^2$$

$$E(\varepsilon^2) = \mathrm{var}(\varepsilon)$$

$$\mathrm{var}(\varepsilon) = \frac{1}{\bar{Y}^2}\mathrm{var}(\bar{y}_r)$$

$$= \left(\frac{1}{r} - \frac{1}{N}\right)\frac{S_y^2}{\bar{Y}^2}$$

$$= \left(\frac{1}{r} - \frac{1}{N}\right)C_y^2$$

Jadi $E(\varepsilon^2) = \left(\frac{1}{r} - \frac{1}{N}\right)C_y^2$

3.2.4.2 Menghitung $E(\delta^2)$:

$$Var(\delta) = E(\delta - E(\delta))^2$$
$$= E\{\delta^2 - 2\delta E\delta + E(\delta)^2\}$$
$$= E(\delta^2) - 2E(\delta)^2 + E(\delta)^2$$
$$= E(\delta^2) - E(\delta)^2$$

$$E(\delta^2) = Var(\delta) + E(\delta)^2$$
$$= Var(\delta) + 0$$
$$= Var(\delta)$$

$$Var(\delta) = E(\delta^2)$$
$$= E\left[\left(\frac{x_r}{\bar{X}} - 1\right)^2\right]$$
$$= \frac{1}{\bar{X}^2} Var(\bar{x}_r)$$

Pandang $z_i = 1$ jika unit i termasuk dalam sampel r. $z_i = 0$ jika tidak termasuk dalam sampel r, sehingga z_i mengikuti distribusi Bernoulli, dengan

$$E(z_i) = \sum_{z_i=0}^{1} z_i P(z_i = z_i)$$
$$= 0.P(z_i = 0) + 1.P(z_i = 1)$$
$$= 0.\left(\frac{N-r}{N}\right) + \frac{r}{N}$$
$$= \frac{r}{N}$$

Maka nilai dari $Var\ (z_i)$ adalah

$$Var(z_i) = E(z_i^2) - (E[z_i])^2$$
$$= \frac{r}{N} - \left(\frac{r}{N}\right)^2 = \frac{r}{N}\left(1 - \frac{r}{N}\right)$$

Oleh karena itu $\bar{x}_r$ dapat ditulis ulang menjadi

$$\bar{x}_r = \frac{1}{r}\sum_{i=1}^{N} x_i\, z_i$$

$$Var(\bar{x}_r) = Var\left[\frac{1}{r}\sum_{i=1}^{N} x_i z_i\right] = \frac{1}{r^2}\left[\sum_{i=1}^{N} x_i^2 \operatorname{var}(z_i) + \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j \operatorname{cov}(z_i, z_j)\right]$$

untuk $i \neq j$

$$\begin{aligned}\operatorname{cov}(z_i, z_j) &= E(z_i, z_j) - E(z_i)E(z_j)\\ &= P(z_i = 1, z_j = 1) - \left(\frac{r}{N}\right)\left(\frac{r}{N}\right)\end{aligned}$$

Peluang dua unit spesifik berada di dalam sampel adalah:

$$\begin{aligned}\frac{\binom{N-2}{r-2}}{\binom{N}{r}} &= \frac{\dfrac{(N-2)!}{(r-2)!(N-r)!}}{\dfrac{N!}{(N-r)!r!}}\\ &= \frac{(N-2)!}{(r-2)!(N-r)!}\, x\, \frac{(N-r)!r!}{N!}\\ &= \frac{(N-2)!r(r-1)(r-2)!}{(r-2)!N(N-1)(N-2)!}\\ &= \frac{r(r-1)}{N(N-1)}\end{aligned}$$

$$\begin{aligned}\operatorname{cov}(z_i, z_j) &= \frac{r(r-1}{N(N-1)} - \left(\frac{r}{N}\right)^2\\ &= \frac{r^2 - r}{N(N-1)} - \frac{r^2}{N^2}\frac{(N-1)}{(N-1)}\\ &= \frac{1}{N-1}\left[\frac{r^2 - r}{N} - \frac{r^2}{N^2}(N-1)\right]\\ &= \frac{1}{N-1}\left[\frac{(r^2 - r)N}{N^2} - \frac{r^2}{N^2}(N-1)\right]\end{aligned}$$

$$= \frac{1}{N-1}\left[\frac{Nr^2 - rN - r^2N + r^2}{N^2}\right]$$

$$= \frac{1}{N-1}\left[\frac{-rN + r^2}{N^2}\right]$$

$$= \frac{1}{N-1}\left[\frac{-r\left(N-r\right)}{N^2}\right]$$

$$= \frac{-Var(z_i)}{N-1}$$

$$var(\overline{x}_r) = \frac{1}{r^2}\left[\sum_{i=1}^{N} x_i^2 \operatorname{var}(z_i) + \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j \operatorname{cov}\left(z_i, z_j\right)\right]$$

$$= \frac{1}{r^2}\left[\sum_{i=1}^{N} x_i^2 \operatorname{var}(z_i) - \operatorname{var}(z_i)\frac{\sum_{i=1}^{N}\sum_{j\neq i} x_i x_j}{N-1}\right]$$

$$= \frac{1}{r^2}\frac{\operatorname{var}(z_i)}{N-1}\left[(N-1)\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j\right]$$

$$= \frac{r(N-r)}{N^2}\frac{1}{r^2}\frac{1}{(N-1)}\left[(N-1)\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j\right]$$

dimana

$$(N-1)\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j = N\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j$$

$$= N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i^2 + \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j\right)$$

$$= N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2$$

$$= N\sum_{i=1}^{N} x_i^2 - \left(N\overline{X}\right)^2$$

$$= N\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2$$

$$= \frac{(N-1)}{(N-1)} N\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^2$$

$$= N(N-1)S_x^2$$

sehingga

$$\operatorname{var}(\bar{x}_r) = \frac{r(N-r)}{N^2}\frac{1}{r^2}\frac{1}{(N-1)}N(N-1)S_x^2$$
$$= \frac{(N-r)}{Nr}S_x^2$$
$$= \left(\frac{1}{r}-\frac{1}{N}\right)S_x^2$$

$$E(\delta^2) = Var(\delta)$$
$$Var(\delta) = \frac{1}{\bar{X}^2}\operatorname{var}(\bar{x}_r)$$
$$= \left(\frac{1}{r}-\frac{1}{N}\right)\frac{S_x^2}{\bar{X}^2}$$
$$= \left(\frac{1}{r}-\frac{1}{N}\right)C_x^2$$

Jadi $E(\delta^2) = \left(\frac{1}{r}-\frac{1}{N}\right)C_x^2$

3.2.4.3 Menghitung $E(\eta^2)$:

$$Var(\eta) = E(\eta - E(\eta))^2$$
$$= E\{\eta^2 - 2\eta E\eta + E(\eta)^2\}$$
$$= E(\eta^2) - 2E(\eta)^2 + E(\eta)^2$$
$$= E(\eta^2) - E(\eta)^2$$

$$E(\eta^2) = Var(\eta) + E(\eta)^2$$
$$= Var(\eta) + 0$$
$$= Var(\eta)$$

$$Var(\eta) = E(\eta^2)$$
$$= E\left[\left(\frac{x_n}{\bar{X}} - 1\right)^2\right]$$
$$= \frac{1}{\bar{X}^2}Var(\bar{x}_n)$$

Pandang $z_i = 1$ jika unit i termasuk dalam sampel n. $z_i = 0$ jika tidak termasuk dalam sampel n, sehingga z_i mengikuti distribusi Bernoulli, dengan

$$\begin{aligned}
E(z_i) &= \sum_{z_i=0}^{1} z_i P(z_i = z_i) \\
&= 0.P(z_i = 0) + 1.P(z_i = 1) \\
&= 0.\left(\frac{N-n}{N}\right) + \frac{n}{N} \\
&= \frac{n}{N}
\end{aligned}$$

Maka nilai dari $Var\ (z_i)$ adalah

$$\begin{aligned}
Var(z_i) &= E(z_i^2) - (E[z_i])^2 \\
&= \frac{n}{N} - \left(\frac{n}{N}\right)^2 = \frac{n}{N}\left(1 - \frac{n}{N}\right) = \frac{n(N-n)}{N^2}
\end{aligned}$$

Oleh karena itu $\bar{x}_n$ dapat ditulis ulang menjadi

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{N} x_i\, z_i$$

$$Var(\bar{x}_n) = Var\left[\frac{1}{n}\sum_{i=1}^{N} x_i z_i\right] = \frac{1}{n^2}\left[\sum_{i=1}^{N} x_i^2 \operatorname{var}(z_i) + \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j \operatorname{cov}(z_i, z_j)\right]$$

untuk $i \neq j$

$$\begin{aligned}
\operatorname{cov}(z_i, z_j) &= E(z_i, z_j) - E(z_i)E(z_j) \\
&= P(z_i = 1, z_j = 1) - \left(\frac{n}{N}\right)\left(\frac{n}{N}\right)
\end{aligned}$$

Peluang dua unit spesifik berada di dalam sampel adalah:

$$\frac{\binom{N-2}{r-2}}{\binom{N}{r}} = \frac{\dfrac{(N-2)!}{(r-2)!(N-r)!}}{\dfrac{N!}{(N-r)!r!}}$$

$$= \frac{(N-2)!}{(r-2)!(N-r)!} x \frac{(N-r)!r!}{N!}$$

$$= \frac{(N-2)!r(r-1)(r-2)!}{(r-2)!N(N-1)(N-2)!}$$

$$= \frac{r(r-1)}{N(N-1)}$$

$$\operatorname{cov}(z_i, z_j) = \frac{n(n-1)}{N(N-1)} - \left(\frac{n}{N}\right)^2$$

$$= \frac{n^2 - n}{N(N-1)} - \frac{n^2}{N^2}\frac{(N-1)}{(N-1)}$$

$$= \frac{1}{N-1}\left[\frac{n^2-n}{N} - \frac{n^2}{N^2}(N-1)\right]$$

$$= \frac{1}{N-1}\left[\frac{(n^2-n)N}{N^2} - \frac{n^2}{N^2}(N-1)\right]$$

$$= \frac{1}{N-1}\left[\frac{Nn^2 - nN - n^2N + n^2}{N^2}\right]$$

$$= \frac{1}{N-1}\left[\frac{-nN + n^2}{N^2}\right]$$

$$= \frac{1}{N-1}\left[\frac{-n(N-n)}{N^2}\right]$$

$$= \frac{-Var(z_i)}{N-1}$$

$$var(\bar{x}_n) = \frac{1}{n^2}\left[\sum_{i=1}^{N} x_i^2 \operatorname{var}(z_i) + \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j \operatorname{cov}(z_i, z_j)\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{N} x_i^2 \operatorname{var}(z_i) - \operatorname{var}(z_i)\frac{\sum_{i=1}^{N}\sum_{j\neq i} x_i x_j}{N-1}\right]$$

$$= \frac{1}{n^2}\frac{-\operatorname{var}(z_i)}{N-1}\left[(N-1)\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j\right]$$

$$= \frac{1}{n^2}\frac{n(N-n)}{N^2}\frac{1}{(N-1)}\left[(N-1)\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j\right]$$

dimana

$$
\begin{aligned}
(N-1)\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j &= N\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j \\
&= N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i^2 + \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j\right) \\
&= N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2 \\
&= N\sum_{i=1}^{N} x_i^2 - \left(N\bar{X}\right)^2 \\
&= N\sum_{i=1}^{N} \left(x_i - \bar{X}\right)^2 \\
&= \frac{(N-1)}{(N-1)} N\sum_{i=1}^{N} \left(x_i - \bar{X}\right)^2 \\
&= N(N-1)S_x^2
\end{aligned}
$$

sehingga

$$
\begin{aligned}
\operatorname{var}(\bar{x}_n) &= \frac{1}{n^2}\frac{n(N-n)}{N^2}\frac{1}{(N-1)} N\left(N-1\right)S_x^2 \\
&= \frac{\left(N-n\right)}{Nn} S_x^2 \\
&= \left(\frac{1}{n}-\frac{1}{N}\right) S_x^2
\end{aligned}
$$

$$
\begin{aligned}
E(\eta^2) &= Var(\eta) \\
Var(\eta) &= \frac{1}{\bar{X}^2}\operatorname{var}(\bar{x}_n) \\
&= \left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_x^2}{\bar{X}^2} \\
&= \left(\frac{1}{n}-\frac{1}{N}\right) C_x^2
\end{aligned}
$$

Jadi $E(\eta^2) = \left(\frac{1}{n}-\frac{1}{N}\right) C_x^2$

3.2.4.4 Menghitung $E(\delta\eta)$:

$$E(\delta\eta) = \frac{1}{X^2} E\left[\left(\bar{x}_r - \bar{X}\right)\left(\bar{x}_n - \bar{X}\right)\right]$$
$$= \frac{1}{X^2} \operatorname{var}(\bar{x})$$

$$Var(\bar{x}) = E(\delta\eta)$$
$$= E\left[\left(\frac{\bar{x}_r}{\bar{X}} - 1\right)\left(\frac{\bar{x}_n}{\bar{X}} - 1\right)\right]$$
$$= \frac{1}{\bar{X}^2} Var(\bar{x})$$

Pandang $z_i = 1$ jika unit i termasuk dalam sampel. $z_i = 0$ jika sebaliknya, sehingga z_i mengikuti distribusi Bernoulli, dengan

$$E(z_i) = \sum_{z_i=0}^{1} z_i P(z_i = z_i)$$
$$= 0.P(z_i = 0) + 1.P(z_i = 1)$$
$$= 0.\left(\frac{N-n}{N}\right) + \frac{n}{N}$$
$$= \frac{n}{N}$$

maka nilai dari $Var\ (z_i)$ adalah

$$Var(z_i) = E(z_i^2) - (E[z_i])^2$$
$$= \frac{n}{N} - \left(\frac{n}{N}\right)^2 = \frac{n}{N}\left(1 - \frac{n}{N}\right) = \frac{n(N-n)}{N^2}$$

Oleh karena itu $\bar{x}$ dapat ditulis ulang menjadi

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i\, z_i$$

$$Var(\bar{x}) = Var\left[\frac{1}{n}\sum_{i=1}^{N} x_i z_i\right] = \frac{1}{n^2}\left[\sum_{i=1}^{N} x_i^2 \operatorname{var}\left(z_i\right) + \sum_{i=1}^{N}\sum_{j\neq i} y_i y_j \operatorname{cov}(z_i, z_j)\right]$$

untuk $i \neq j$

$$\text{cov}(z_i, z_j) = E(z_i, z_j) - E(z_i)E(z_j)$$

$$= P(z_i = 1, z_j = 1) - \left(\frac{n}{N}\right)\left(\frac{n}{N}\right)$$

Peluang dua unit spesifik berada di dalam sampel adalah:

$$\frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{\frac{(N-2)!}{(n-2)!(N-n)!}}{\frac{N!}{(N-n)!n!}}$$

$$= \frac{(N-2)!}{(n-2)!(N-n)!} x \frac{(N-n)!n!}{N!}$$

$$= \frac{(N-2)!n(n-1)(n-2)!}{(n-2)!N(N-1)(N-2)!}$$

$$= \frac{n(n-1)}{N(N-1)}$$

$$\text{cov}(z_i, z_j) = \frac{n(n-1)}{N(N-1)} - \left(\frac{n}{N}\right)^2$$

$$= \frac{n^2 - n}{N(N-1)} - \frac{n^2}{N^2}\frac{(N-1)}{(N-1)}$$

$$= \frac{1}{N-1}\left[\frac{n^2 - n}{N} - \frac{n^2}{N^2}(N-1)\right]$$

$$= \frac{1}{N-1}\left[\frac{(n^2 - n)N}{N^2} - \frac{n^2}{N^2}(N-1)\right]$$

$$= \frac{1}{N-1}\left[\frac{Nn^2 - nN - n^2N + n^2}{N^2}\right]$$

$$= \frac{1}{N-1}\left[\frac{-nN + n^2}{N^2}\right]$$

$$= \frac{1}{N-1}\left[\frac{-n(N-n)}{N^2}\right]$$

$$= \frac{-Var(z_i)}{N-1}$$

$$var(\bar{x}) = \frac{1}{n^2}\left[\sum_{i=1}^{N} x_i^2 \operatorname{var}(z_i) + \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j \operatorname{cov}\left(z_i, z_j\right)\right]$$

$$= \frac{1}{n^2}\left[\sum_{i=1}^{N} x_i^2 \operatorname{var}(z_i) - \operatorname{var}(z_i)\frac{\sum_{i=1}^{N}\sum_{j\neq i} x_i x_j}{N-1}\right]$$

$$= \frac{1}{n^2}\frac{-\operatorname{var}(z_i)}{N-1}\left[\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j\right]$$

$$= \frac{1}{n^2}\frac{n(N-n)}{N^2}\frac{1}{(N-1)}\left[(N-1)\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j\right]$$

dimana

$$(N-1)\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j = N\sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N} x_i^2 - \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j$$

$$= N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i^2 + \sum_{i=1}^{N}\sum_{j\neq i} x_i x_j\right)$$

$$= N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2$$

$$= N\sum_{i=1}^{N} x_i^2 - \left(N\bar{X}\right)^2$$

$$= N\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2$$

$$= \frac{(N-1)}{(N-1)} N\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^2$$

$$= N(N-1)S_x^2$$

sehingga,

$$\operatorname{var}(\bar{x}) = \frac{1}{n^2}\frac{n(N-n)}{N^2}\frac{1}{(N-1)} N\left(N-1\right)S_x^2$$

$$= \frac{\left(N-n\right)}{Nn} S_x^2$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_x^2$$

$$E(\delta\eta) = Var(x)$$

$$Var(x) = \frac{1}{\bar{X}^2}\,\mathrm{var}(\bar{x})$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_x^2}{\bar{X}^2}$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right)C_x^2$$

Jadi, $E(\delta\eta) = \left(\frac{1}{n} - \frac{1}{N}\right)C_x^2$

3.2.4.5 Menghitung $E(\varepsilon\eta)$:

$$E(\varepsilon\eta) = E\left[\left(\frac{\bar{y}_r}{\bar{Y}} - 1\right)\left(\frac{\bar{x}_n}{\bar{X}} - 1\right)\right]$$

$$= E\left[\left(\frac{\bar{y}_r - \bar{Y}}{\bar{Y}}\right)\left(\frac{\bar{x}_n - \bar{X}}{\bar{X}}\right)\right]$$

$$= \frac{1}{\bar{Y}\bar{X}} E\left[\left(\bar{y}_r - \bar{Y}\right)\left(\bar{x}_n - \bar{X}\right)\right]$$

$$= \frac{1}{\bar{Y}\bar{X}} \mathrm{cov}\left(\bar{y}_r, \bar{x}_n\right)$$

diberikan

$$u_i = y_i + x_i$$
$$\bar{U} = \bar{Y} + \bar{X}$$

sehingga

$$\bar{u} - \bar{U} = \frac{1}{n}\sum_{i=1}^{n} y_i - \bar{Y}$$

$$n\left(\bar{u} - \bar{U}\right) = n\left(\frac{1}{n}\sum_{i=1}^{n} u_i - \bar{U}\right)$$

$$n\left(\bar{u} - \bar{U}\right) = \left(u_1 - \bar{U}\right) + \ldots + \left(u_n - \bar{U}\right) \qquad (3.3)$$

Dimisalkan

$$E\left[\left(u_i-\bar{U}\right)^2\right]=a^2$$

$$E\left[\sum_{i=1}^{n}\left(u_i-\bar{U}\right)^2\right]=na^2$$

$$E\left[\sum_{i=1}^{N}\left(u_i-\bar{U}\right)^2\right]=Na^2$$

$$E\left[\sum_{i=1}^{n}\left(u_i-\bar{U}\right)^2\right]=\frac{n}{N}E\left[\sum_{i=1}^{N}\left(u_i-\bar{U}\right)^2\right] \qquad (3.4)$$

dan juga bahwa

$$E\left[\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\left(u_1-\bar{U}\right)\left(u_3-\bar{U}\right)+\ldots+\left(u_{n-1}-\bar{U}\right)\left(u_n-\bar{U}\right)\right]=\frac{{}_nC_2}{{}_NC_2}\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+$$
$$\left(u_1-\bar{U}\right)\left(u_3-\bar{U}\right)+\ldots+\left(u_{N-1}-\bar{U}\right)\left(u_N-\bar{U}\right)]$$

$$E\left[\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\left(u_1-\bar{U}\right)\left(u_3-\bar{U}\right)+\ldots+\left(u_{n-1}-\bar{U}\right)\left(u_n-\bar{U}\right)\right]=\frac{n(n-1)}{N(N-1)}\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+$$
$$\left(u_1-\bar{U}\right)\left(u_3-\bar{U}\right)+\ldots+\left(u_{N-1}-\bar{U}\right)\left(u_N-\bar{U}\right)] \qquad (3.5)$$

Persamaan (3.3) dikuadratkan

$$\left[nE\left(\bar{u}-\bar{U}\right)\right]^2=\left[\left(u_1-\bar{U}\right)+\left(u_2-\bar{U}\right)+\ldots+\left(u_n-\bar{U}\right)\right]^2$$
$$=\left(u_1-\bar{U}\right)^2+2\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\ldots+\left(u_n-\bar{U}\right)^2+2\left(u_{n-1}-\bar{U}\right)\left(u_n-\bar{U}\right)$$
$$=\left(u_1-\bar{U}\right)^2+\ldots+\left(u_n-\bar{U}\right)^2+2\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\ldots+2\left(u_{n-1}-\bar{U}\right)\left(u_n-\bar{U}\right)$$
$$=\frac{n}{N}\{\left(u_1-\bar{U}\right)^2+\ldots+\left(u_N-\bar{U}\right)^2+2\frac{n-1}{N-1}[\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\ldots+$$
$$\left(u_{N-1}-\bar{U}\right)\left(u_N-\bar{U}\right)]$$
$$=\frac{n}{N}\{\left(u_1-\bar{U}\right)^2+\ldots+\left(u_N-\bar{U}\right)^2+\frac{n-1}{N-1}[\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\ldots+$$
$$\left(u_{N-1}-\bar{U}\right)\left(u_N-\bar{U}\right)]\}$$

$$=\frac{n}{N}\{(u_1-\bar{U})^2+\ldots+(u_N-\bar{U})^2+\frac{n-1}{N-1}[(u_1-\bar{U})+(u_2-\bar{U})+\ldots+(u_N-\bar{U})]^2-$$
$$[(u_1-\bar{U})^2+\ldots+(u_N-\bar{U})^2]\}$$

$$=\frac{n}{N}\{1-\frac{n-1}{N-1}[(u_1-\bar{U})^2+\ldots+(u_N-\bar{U})^2]+\frac{n-1}{N-1}[(u_1-\bar{U})+(u_2-\bar{U})+\ldots+$$
$$(u_N-\bar{U})]^2\}$$

Setelah dibagi menjadi n^2 menjadi

$$\frac{1}{n^2}n^2E\left(\bar{u}-\bar{U}\right)^2=\frac{1}{n^2}[\frac{n}{N}\{(1-\frac{n-1}{N-1})[(u_1-\bar{U})^2+\ldots+(u_N-\bar{U})^2]+\frac{n-1}{N-1}[(u_1-\bar{U})+$$
$$(u_2-\bar{U})+\ldots+(u_N-\bar{U})]^2\}]$$

$$E\left(\bar{u}-\bar{U}\right)^2=\frac{1}{nN}-\frac{n-1}{nN(N-1)}[(u_1-\bar{U})^2+\ldots+(u_N-\bar{U})^2]+\frac{n-1}{nN(N-1)}[(u_1-\bar{U})+$$
$$(u_2-\bar{U})+\ldots+(u_N-\bar{U})]^2$$

$$=\frac{N-n}{nN(N-1)}[(u_1-\bar{U})^2+\ldots+(u_N-\bar{U})^2]+\frac{n-1}{nN(N-1)}[(u_1-\bar{U})+$$
$$(u_2-\bar{U})+\ldots+(u_N-\bar{U})]^2$$

Suku kedua dalam tanda kurung akan hilang karena jumlah dari y_i sama dengan $N\bar{Y}$.

$$E\left(\bar{u}-\bar{U}\right)^2=\frac{N-n}{nN(N-1)}\sum\left(u_i-\bar{U}\right)^2$$

$$E\left[\left(\bar{y}+\bar{x}\right)-\left(\bar{Y}+\bar{X}\right)\right]^2=\frac{N-n}{nN(N-1)}\sum\left[\left(y_i+x_i\right)-\left(\bar{Y}+\bar{X}\right)\right]^2$$

$$E\left[\left(\bar{y}-\bar{Y}\right)+\left(\bar{x}-\bar{X}\right)\right]^2=\frac{N-n}{nN(N-1)}\sum\left[\left(y_i-\bar{Y}\right)+\left(x_i-\bar{X}\right)\right]^2$$

$$E[(\bar{y}-\bar{Y})^2+2(\bar{y}-\bar{Y})(\bar{x}-\bar{X})+(\bar{x}-\bar{X})^2]=\frac{N-n}{nN(N-1)}\sum[(y_i-\bar{Y})^2+2(y_i-\bar{Y})(x_i-\bar{X})+$$
$$(x_i-\bar{X})^2]$$

$$E\left(\bar{y}-\bar{Y}\right)\left(\bar{x}-\bar{X}\right)=\frac{N-n}{nN(N-1)}\sum\left(y_i-\bar{Y}\right)\left(x_i-\bar{X}\right)$$

$$E\left(\bar{y}-\bar{Y}\right)\left(\bar{x}-\bar{X}\right)=\frac{N-n}{nN}\frac{1}{N-1}\sum\left(y_i-\bar{Y}\right)\left(x_i-\bar{X}\right)$$

$$=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{1}{N-1}\sum\left(y_i-\bar{Y}\right)\left(x_i-\bar{X}\right)$$

$$E(\varepsilon\delta)=\frac{1}{\bar{X}\bar{Y}}\left[\left(\frac{1}{n}-\frac{1}{N}\right)S_{xy}\right]$$

$$=\frac{1}{\bar{X}\bar{Y}}\left[\left(\frac{1}{n}-\frac{1}{N}\right)S_{xy}\frac{(N-1)^{-1}\left[\sum(x_i-\bar{X})\sum(y_i-\bar{Y})\right]}{(N-1)^{-1}\left[\sum(x_i-\bar{X})\sum(y_i-\bar{Y})\right]}\right]$$

$$=\frac{1}{\bar{X}\bar{Y}}\left[\left(\frac{1}{n}-\frac{1}{N}\right)S_{xy}\frac{S_xS_y}{S_xS_y}\right]$$

$$=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_{xy}}{S_xS_y}\frac{S_xS_y}{\bar{X}\bar{Y}}$$

$$=\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{xy}C_xC_yE(\varepsilon\delta)=\frac{1}{\bar{X}\bar{Y}}\left[\left(\frac{1}{n}-\frac{1}{N}\right)S_{xy}\right]$$

$$=\frac{1}{\bar{X}\bar{Y}}\left[\left(\frac{1}{n}-\frac{1}{N}\right)S_{xy}\frac{(N-1)^{-1}\left[\sum(x_i-\bar{X})\sum(y_i-\bar{Y})\right]}{(N-1)^{-1}\left[\sum(x_i-\bar{X})\sum(y_i-\bar{Y})\right]}\right]$$

$$=\frac{1}{\bar{X}\bar{Y}}\left[\left(\frac{1}{n}-\frac{1}{N}\right)S_{xy}\frac{S_xS_y}{S_xS_y}\right]$$

$$=\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_{xy}}{S_xS_y}\frac{S_xS_y}{\bar{X}\bar{Y}}$$

$$=\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{xy}C_xC_y$$

3.2.4.6 Menghitung $E(\varepsilon\delta)$:

$$E(\varepsilon\delta)=E\left[\left(\frac{\bar{y}_r}{\bar{Y}}-1\right)\left(\frac{\bar{x}_r}{\bar{X}}-1\right)\right]$$

$$=E\left[\left(\frac{\bar{y}_r-\bar{Y}}{\bar{Y}}\right)\left(\frac{\bar{x}_r-\bar{X}}{\bar{X}}\right)\right]$$

$$=\frac{1}{\bar{Y}\bar{X}}E\left[(\bar{y}_r-\bar{Y})(\bar{x}_r-\bar{X})\right]$$

$$=\frac{1}{\bar{Y}\bar{X}}\operatorname{cov}(\bar{y}_r,\bar{x}_r)$$

diberikan

$$u_i=y_i+x_i$$
$$\bar{U}=\bar{Y}+\bar{X}$$

sehingga

$$\bar{u}-\bar{U}=\frac{1}{r}\sum_{i=1}^{n}y_i-\bar{Y}$$

$$r\left(\bar{u}-\bar{U}\right)=r\left(\frac{1}{r}\sum_{i=1}^{n}u_i-\bar{U}\right)$$

$$r\left(\bar{u}-\bar{U}\right)=\left(u_1-\bar{U}\right)+\ldots+\left(u_r-\bar{U}\right) \tag{3.5}$$

Dimisalkan

$$E\left[\left(u_i-\bar{U}\right)^2\right]=a^2$$

$$E\left[\sum_{i=1}^{r}\left(u_i-\bar{U}\right)^2\right]=ra^2$$

$$E\left[\sum_{i=1}^{N}\left(u_i-\bar{U}\right)^2\right]=Na^2$$

$$E\left[\sum_{i=1}^{r}\left(u_i-\bar{U}\right)^2\right]=\frac{r}{N}E\left[\sum_{i=1}^{N}\left(u_i-\bar{U}\right)^2\right] \tag{3.6}$$

dan juga bahwa

$$E\left[\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\left(u_1-\bar{U}\right)\left(u_3-\bar{U}\right)+\ldots+\left(u_{r-1}-\bar{U}\right)\left(u_r-\bar{U}\right)\right]=\frac{{}_rC_2}{{}_NC_2}\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+ \left(u_1-\bar{U}\right)\left(u_3-\bar{U}\right)+\ldots+\left(u_{N-1}-\bar{U}\right)\left(u_N-\bar{U}\right)]$$

$$E\left[\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\left(u_1-\bar{U}\right)\left(u_3-\bar{U}\right)+\ldots+\left(u_{r-1}-\bar{U}\right)\left(u_r-\bar{U}\right)\right]=\frac{r(r-1)}{N(N-1)}\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+ \left(u_1-\bar{U}\right)\left(u_3-\bar{U}\right)+\ldots+\left(u_{N-1}-\bar{U}\right)\left(u_N-\bar{U}\right)] \tag{3.7}$$

Persamaan (3.5) dikuadratkan

$$\left[rE\left(\bar{u}-\bar{U}\right)\right]^2=\left[\left(u_1-\bar{U}\right)+\left(u_2-\bar{U}\right)+\ldots+\left(u_r-\bar{U}\right)\right]^2$$

$$=\left(u_1-\bar{U}\right)^2+2\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\ldots+\left(u_r-\bar{U}\right)^2+2\left(u_{r-1}-\bar{U}\right)\left(u_r-\bar{U}\right)$$

$$=\left(u_1-\bar{U}\right)^2+\ldots+\left(u_r-\bar{U}\right)^2+2\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\ldots+2\left(u_{r-1}-\bar{U}\right)\left(u_r-\bar{U}\right)$$

$$=\frac{r}{N}\{\left(u_1-\bar{U}\right)^2+\ldots+\left(u_N-\bar{U}\right)^2+2\frac{r-1}{N-1}[\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\ldots+$$
$$\left(u_{N-1}-\bar{U}\right)\left(u_N-\bar{U}\right)]$$

$$=\frac{r}{N}\{\left(u_1-\bar{U}\right)^2+\ldots+\left(u_N-\bar{U}\right)^2+\frac{r-1}{N-1}[\left(u_1-\bar{U}\right)\left(u_2-\bar{U}\right)+\ldots+$$
$$\left(u_{N-1}-\bar{U}\right)\left(u_N-\bar{U}\right)]\}$$

$$=\frac{r}{N}\{(u_1-\bar{U})^2+\ldots+(u_N-\bar{U})^2+\frac{r-1}{N-1}[(u_1-\bar{U})+(u_2-\bar{U})+\ldots+(u_N-\bar{U})]^2-$$
$$[(u_1-\bar{U})^2+\ldots+(u_N-\bar{U})^2]\}$$

$$=\frac{r}{N}\{1-\frac{n-1}{N-1}[(u_1-\bar{U})^2+\ldots+(u_N-\bar{U})^2]+\frac{r-1}{N-1}[(u_1-\bar{U})+(u_2-\bar{U})+\ldots+$$
$$(u_N-\bar{U})]^2\}$$

Setelah dibagi menjadi r^2 menjadi

$$\frac{1}{r^2}r^2E\left(\bar{u}-\bar{U}\right)^2=\frac{1}{r^2}[\frac{r}{N}\{(1-\frac{r-1}{N-1})[\left(u_1-\bar{U}\right)^2+\ldots+\left(u_N-\bar{U}\right)^2]+\frac{r-1}{N-1}[(u_1-\bar{U})+$$
$$(u_2-\bar{U})+\ldots+(u_N-\bar{U})]^2$$

$$E\left(\bar{u}-\bar{U}\right)^2=\frac{1}{rN}-\frac{r-1}{rN(N-1)}\left[\left(u_1-\bar{U}\right)^2+\ldots+\left(u_N-\bar{U}\right)^2\right]+\frac{r-1}{rN(N-1)}[(u_1-\bar{U})+$$
$$(u_2-\bar{U})+\ldots(u_N-\bar{U})]^2$$

$$=\frac{N-r}{rN(N-1)}\left[\left(u_1-\bar{U}\right)^2+\ldots+\left(u_N-\bar{U}\right)^2\right]+\frac{r-1}{rN(N-1)}[(u_1-\bar{U})+$$
$$(u_2-\bar{U})+\ldots+(u_N-\bar{U})]^2$$

Suku kedua dalam tanda kurung akan hilang karena jumlah dari y_i sama dengan $N\bar{Y}$.

$$E\left(\bar{u}-\bar{U}\right)^2=\frac{N-r}{rN(N-1)}\sum\left(u_i-\bar{U}\right)^2$$

$$E\left[\left(\bar{y}_r+\bar{x}_r\right)-\left(\bar{Y}+\bar{X}\right)\right]^2=\frac{N-r}{rN(N-1)}\sum\left[\left(y_i+x_i\right)-\left(\bar{Y}+\bar{X}\right)\right]^2$$

$$E\left[\left(\bar{y}_r-\bar{Y}\right)+\left(\bar{x}_r-\bar{X}\right)\right]^2=\frac{N-r}{rN(N-1)}\sum\left[\left(y_i-\bar{Y}\right)+\left(x_i-\bar{X}\right)\right]^2$$

$$E\left[\left(\bar{y}_r-\bar{Y}\right)^2+2\left(\bar{y}_r-\bar{Y}\right)\left(\bar{x}_r-\bar{X}\right)+\left(\bar{x}_r-\bar{X}\right)^2\right]=\frac{N-r}{rN(N-1)}\sum[(y_i-\bar{Y})^2+2(y_i-\bar{Y})(x_i-\bar{X})+(x_i-\bar{X})^2]$$

$$E\left(\bar{y}_r-\bar{Y}\right)\left(\bar{x}_r-\bar{X}\right)=\frac{N-r}{rN(N-1)}\sum\left(y_i-\bar{Y}\right)\left(x_i-\bar{X}\right)$$

$$E\left(\bar{y}_r-\bar{Y}\right)\left(\bar{x}_r-\bar{X}\right)=\frac{N-r}{rN}\frac{1}{N-1}\sum\left(y_i-\bar{Y}\right)\left(x_i-\bar{X}\right)$$

$$=\left(\frac{1}{r}-\frac{1}{N}\right)\frac{1}{N-1}\sum\left(y_i-\bar{Y}\right)\left(x_i-\bar{X}\right)$$

$$E(\varepsilon\delta)=\frac{1}{\bar{X}\bar{Y}}\left[\left(\frac{1}{r}-\frac{1}{N}\right)S_{xy}\right]$$

$$=\frac{1}{\bar{X}\bar{Y}}\left[\left(\frac{1}{r}-\frac{1}{N}\right)S_{xy}\frac{(N-1)^{-1}\left[\sum\left(x_i-\bar{X}\right)\sum\left(y_i-\bar{Y}\right)\right]}{(N-1)^{-1}\left[\sum\left(x_i-\bar{X}\right)\sum\left(y_i-\bar{Y}\right)\right]}\right]$$

$$=\frac{1}{\bar{X}\bar{Y}}\left[\left(\frac{1}{r}-\frac{1}{N}\right)S_{xy}\frac{S_xS_y}{S_xS_y}\right]$$

$$=\left(\frac{1}{r}-\frac{1}{N}\right)\frac{S_{xy}}{S_xS_y}\frac{S_xS_y}{\bar{X}\bar{Y}}$$

$$=\left(\frac{1}{r}-\frac{1}{N}\right)\rho_{xy}C_xC_y$$

3.2.4.7 Menghitung MSE

Hasil penjabaran $E(\varepsilon^2)$, $E(\delta^2)$, $E(\varepsilon\delta)$, $E(\eta^2)$, $E(\delta\eta)$, dan $E(\varepsilon\eta)$ akan dihitung MSE dari metode regresi imputasi. Terlebih dahulu akandicari definisi dari $\bar{y}_r$, $\bar{x}_r$, dan $\bar{x}_n$ dari ketiga *error* tersebut. Definisi dari $\bar{y}_r = (\varepsilon + 1)\bar{Y}$ diperoleh dari

$$\varepsilon = \frac{\bar{y}_r}{\bar{Y}} - 1$$

$$\varepsilon + 1 = \frac{\bar{y}_r}{\bar{Y}}$$

$$\bar{y}_r = (\varepsilon + 1)\bar{Y}$$

Definisi dari $\bar{x}_n = (\eta + 1)\bar{X}$ diperoleh dari

$$\eta = \frac{\bar{x}_n}{\bar{X}} - 1$$

$$\eta + 1 = \frac{\bar{x}_n}{\bar{X}}$$

$$\bar{x}_n = (\eta + 1)\bar{X}$$

Definisi dari $\bar{x}_r = (\delta + 1)\bar{X}$ diperoleh dari

$$\delta = \frac{\bar{x}_r}{\bar{X}} - 1$$

$$\delta + 1 = \frac{\bar{x}}{\bar{X}}$$

$$\bar{x}_r = (\delta + 1)\bar{X}$$

$$\begin{aligned}
MSE(\bar{y}_{reg}) &= E\left[\bar{y}_{reg} - \bar{Y}\right]^2 \\
&= E\left[\left((\varepsilon+1)\bar{Y} + \hat{\beta}(\eta\bar{X} - \delta\bar{X}) - Y\right)\right]^2 \\
&= E\left[\left((\varepsilon+1)\bar{Y} + \hat{\beta}\bar{X}(\eta - \delta) - Y\right)\right]^2 \\
&= E\left[\bar{Y}^2\varepsilon^2 + \hat{\beta}^2\bar{X}^2(\eta^2 + \delta^2 - 2\eta\delta) + 2\bar{X}\hat{\beta}\bar{Y}(\varepsilon\eta - \varepsilon\delta)\right] \\
&= \bar{Y}^2 E[\varepsilon^2] + \hat{\beta}^2\bar{X}^2 E[\eta^2 + \delta^2 - 2\eta\delta] + 2\bar{X}\hat{\beta}\bar{Y}E[\varepsilon\eta - \varepsilon\delta]
\end{aligned}$$

$$=\bar{Y}^2\left(\frac{1}{r}-\frac{1}{N}\right)\frac{S_y^2}{\bar{Y}^2}+\hat{\beta}^2\bar{X}^2\left[\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_x^2}{\bar{X}^2}+\left(\frac{1}{r}-\frac{1}{N}\right)\frac{S_x^2}{\bar{X}^2}-2\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_x^2}{\bar{X}^2}\right]+$$

$$2\bar{X}\hat{\beta}\bar{Y}\left[\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{xy}\frac{S_y}{\bar{Y}}\frac{S_x}{\bar{X}}-\left(\frac{1}{r}-\frac{1}{N}\right)\rho_{xy}\frac{S_y}{\bar{Y}}\frac{S_x}{\bar{X}}\right]$$

$$=\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2+\hat{\beta}^2\left[\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2+\left(\frac{1}{r}-\frac{1}{N}\right)S_x^2-2\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2\right]+$$

$$2\hat{\beta}\left[\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{xy}S_yS_x-\left(\frac{1}{r}-\frac{1}{N}\right)\rho_{xy}S_yS_x\right]$$

$$\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2=\left(\frac{1}{r}-\frac{1}{N}\right)S_y^2$$

$$\hat{\beta}^2\left[\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2+\left(\frac{1}{r}-\frac{1}{N}\right)S_x^2-2\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2\right]=$$

$$=\frac{S_{xy}^2}{S_x^4}\left[\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2+\left(\frac{1}{r}-\frac{1}{N}\right)S_x^2-2\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2\right]$$

$$=\frac{S_{xy}^2}{S_x^4}\left[\left(\frac{1}{n}-\frac{1}{N}+\frac{1}{r}-\frac{1}{N}-2\frac{1}{n}-\frac{1}{N}\right)S_x^2\right]$$

$$=\frac{S_{xy}^2}{S_x^2}\left(\frac{1}{r}-\frac{1}{n}\right)$$

$$2\hat{\beta}\left[\left(\frac{1}{n}-\frac{1}{N}\right)\rho_{xy}S_yS_x-\left(\frac{1}{r}-\frac{1}{N}\right)\rho_{xy}S_xS_y\right]=$$

$$=2\frac{S_{xy}}{S_x^2}\left[\left(\frac{1}{n}-\frac{1}{N}-\frac{1}{r}-\frac{1}{N}\right)\left(\rho_{xy}S_yS_x\right)\right]$$

$$=2\frac{S_{xy}}{S_x^2}\left(\frac{1}{n}-\frac{1}{r}\right)\rho_{xy}S_yS_x$$

$$=2\frac{S_{xy}}{S_x}\left(\frac{1}{n}-\frac{1}{r}\right)\rho_{xy}S_y$$

$$MSE(\bar{y}_{reg}) = \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 + \frac{S_{xy}^2}{S_x^2}\left(\frac{1}{r}-\frac{1}{n}\right) + 2\frac{S_{xy}}{S_x}\left(\frac{1}{n}-\frac{1}{r}\right)\rho_{xy}S_y$$

$$= \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)\left(\frac{S_{xy}^2}{S_x^2} - 2\frac{S_{xy}}{S_x}\rho_{xy}S_y\right)$$

$$= \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)\left(\frac{S_{xy}^2}{S_x^2}\frac{S_y^2}{S_y^2} - 2\frac{S_{xy}}{S_x}\frac{S_y}{S_y}\rho_{xy}S_y\right)$$

$$= \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)\left(\rho_{xy}^2S_y^2 - 2\rho_{xy}^2S_y^2\right)$$

$$= \left(\frac{1}{r}-\frac{1}{N}\right)S_y^2 - \left(\frac{1}{r}-\frac{1}{n}\right)\rho_{xy}^2S_y^2$$

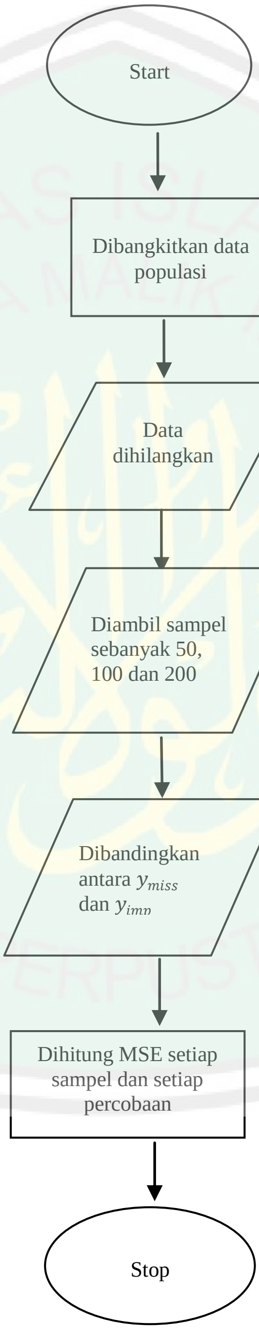
$$= \left[\left\{\left(\frac{1}{r}-\frac{1}{N}\right) + \left(\frac{1}{n}-\frac{1}{r}\right)\right\}S_y^2\right] - \left[\left(\frac{1}{r}-\frac{1}{n}\right)\rho_{xy}^2S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)S_y^2\right]$$

$$= \left(\frac{1}{n}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)S_y^2 - \rho_{xy}^2S_y^2$$

$$= \left(\frac{1}{n}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{r}-\frac{1}{n}\right)S_y^2\left(1-\rho_{xy}^2\right)$$

3.3 Simulasi Data Imputasi Regresi

Penelitian ini menggunakan data real dari penelitian Nur Malahayati (2008) yang berjudul Perbandingan *Metode Imputasi Ganda: Metode Regresi versus Metode Predictive Mean Matching untuk Mengatasi Data Hilang pada Data Survei*, kemudian dibangkitkan seperti data survei. Data ini adalah data saling berhubungan dan tidak berkelipatan. Hal ini bisa dilihat dari nilai variansi X, variansi Y, dan kovariansi XY. Data ini dibuat untuk menduga nilai tengah lingkar pinggang. Dalam membangkitan data kedua peubah berat badan (X) dan lingkar pinggang (Y) tersebut dibuat mempunyai korelasi. Diasumsikan peubah Y ini mempunyai peluang untuk nonrespon/hilang karena beberapa kendala dalam survei. Untuk mengatasi *missing data* digunakan metode regresi imputasi. Adapun langkah-langkah untuk melakukan simulasi metode ini adalah sebagai berikut:



a. Dibangkitkan data populasi sebesar 1000 unit.

b. Dari data populasi tersebut diambil sampel berukuran 50, 100 dan 200 dan diulang sebanyak 10 kali setiap sampelnya.

c. Pada setiap sampel dan setiap percobaan dilakukan penghilangan data sebanyak 5%, 10% dan 15% pada peubah y, sedangkan peubah x dibiarkan lengkap.

d. Setiap imputasi dibandingkan nilai y_{miss} dan y_{imp} kemudian dihitung $\bar{y}_{reg}$

e. Dihitung MSE setiap sampel dan setiap percobaan.

Konsep proses imputasi adalah dengan mengambil sampel berukuran 50, 100, 200 dari populasi dengan 10 kali percobaan. Setiap 10 kali percobaan peubah $y$akan dihilangkan sebanyak 5%, 10%, dan 10%. Dengan menggunakan model imputasi regresi pada (3.1) akan dilakukan imputasi sebanyak y_{miss} tersebut. Hasil pendugaan data hilang dengan metode regresi imputasi dapat dilihat pada lampiran 3

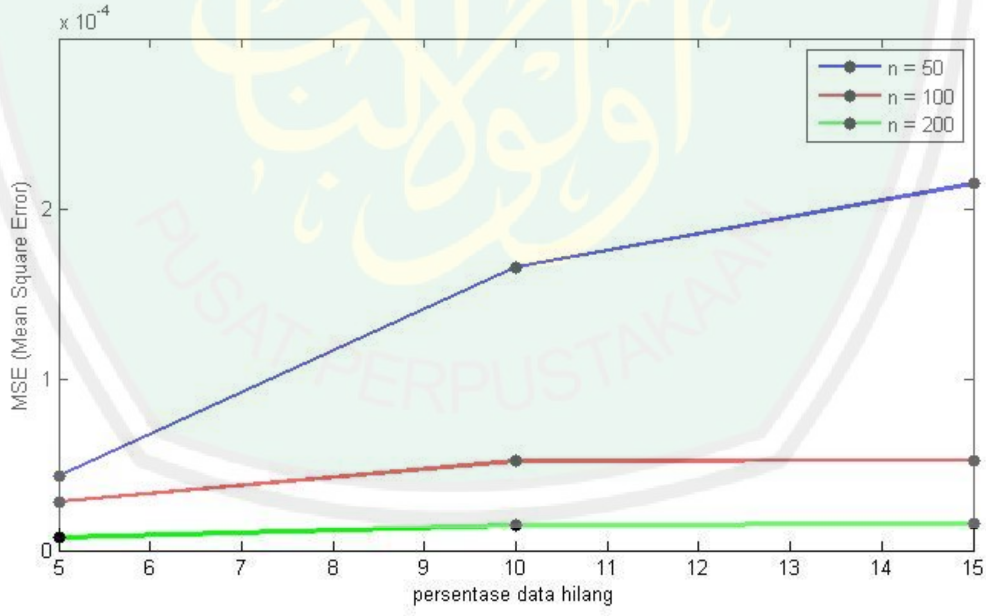
3.3.1 Analisis Metode Regresi Imputasi

Dari percobaan metode regresi imputasi mempunyai hasil yang memuaskan.Antaray_{miss} dan y_{imp} mempunyai kesalahan yang relatif kecil. Hal ini bisa dilihat dari MSE setiap sampel dan setiap jumlah data yang hilang. Untuk sampel berukuran 50 unit dengan y_{miss} 5% didapatkan $\text{MSE} = 4.36x10^{-5}$, y_{miss} 10% didapatkan $\text{MSE} = 1.66x10^{-4}$ dan y_{miss} 15% didapatkan $\text{MSE} = 2.51x10^{-4}$.

Tabel 3.2 Nilai MSE

n	y_{miss} **%**		
	5	**10**	**15**
50	$4.36x10^{-5}$	$1.66x10^{-4}$	$2.51x10^{-4}$
100	$2.87Ex10^{-5}$	$5.22091x10^{-5}$	$5.28x10^{-5}$
200	$7.46x10^{-6}$	$1.48x10^{-5}$	$1.55x10^{-5}$

Sedangkan sampel berukuran 100 unit dengan y_{miss} 5% didapatkan $\text{MSE} = 2.87Ex10^{-5}$, y_{miss} 10% didapatkan $\text{MSE} = 5.22091x10^{-5}$ dan y_{miss} 15% didapatkan $\text{MSE} = 5.28x10^{-5}$. Dan untuk sampel berukuran 200 unit dengan y_{miss} 5% didapatkan $\text{MSE} = 7.46x10^{-6}$, y_{miss} 10% didapatkan $\text{MSE} = 1.48x10^{-5}$ dan y_{miss} 15% didapatkan $\text{MSE} = 1.55x10^{-5}$.



Gambar 3.1 Grafik MSE pada Data Hilang

Dari grafik di atas bisa diambil kesimpulan bahwa jika data yang diambil mempunyai jumlah yang sama tapi mempunyai persentase nilai hilang semakin besar diperoleh nilai MSE semakin besar dan jika data yang diambil mempunyai jumlah semakin besar tapi mempunyai persentase nilai hilang yang sama maka MSE semakin kecil.

Nilai MSE semakin besar ketika data yang hilang juga semakin besar. Hal ini dikarenakan semakin banyak data yang hilang maka data yang akan diimputkan pun juga akan semakin banyak. Sehingga akan banyak muncul nilai kesalahan dari hasil pengimputan data tadi dan demikian pula sebaliknya.

Nilai MSE semakin kecil ketika data yang diambil mempunyai jumlah semakin besar dengan persentase nilai hilang yang sama. Hal ini dikarenakan jumlah sampel yang besar semakin menggambarkan populasi.

3.4 Kajian Keagamaan

Kasus data hilang ketika melakukan survei selain menyebabkan data tidak lengkap juga menyebabkan pendugaan parameter tidak efesien. Data yang diperoleh ketika melakukan survei harus teramati, tercatat, dan terkumpulkan dengan baik. Seperti yang telah dijelaskan pada bab sebelumnya, bahwa data yang valid pada suatu penelitian akan menghasilkan analisis yang memuaskan. Oleh karena itu semua data dari hasil survei harus tercatat semuanya.

Dalam Al-Qur’an dijelaskan bahwa Allah akan mencatat semua dosa, baik itu dosa kecil ataupun dosa besar. Firman Allah SWT:

لَا يُغَادِرُ صَغِيرَةً وَلَا كَبِيرَةً إِلَّآ أَحۡصَىٰهَاۚ وَوَجَدُواْ مَا عَمِلُواْ حَاضِرًاۗ وَلَا يَظۡلِمُ رَبُّكَ أَحَدًا

Artinya: *“…tidak meninggalkan yang kecil dan tidak (pula) yang besar, melainkan ia mencatat semuanya; dan mereka dapati apa yang telah mereka kerjakan ada (tertulis). Dan Tuhanmu tidak Menganiaya seorang juapun (Qs. Al-Kahfi: 49)".*

Pada ayat di atas dijelaskan bahwa segala sesuatu, baik itu dosa kecil maupun dosa besar semuanya akan dicatat olehAllah SWT. Dosa kecil: di bawah kesyirikan dan dosa besar: kesyirikan semuanya dicatat. Dalam tafsir Al-Qurthubi menceritakan bahwa Qatadah berkata,"suatu kaum mengadukan cacat jiwa namun tidak ada seorangpun yang mengadukan kezhaliman. Maka jauhilah oleh kalian dosa-dosa kecil karena semua itu dapat terhimpun pada seseorang sehingga membinasakannya".

وَوَجَدُوْا مَا عَمِلُوْا حَا ضِرًا*"Dan mereka dapati apa yang telah mereka kerjakan itu ada".* Maksudnya, mereka menemukan pencatatan semua yang telah mereka kerjakan telah ada dan mereka mendapatkan balasan atas apa-apa yang mereka lakukan.

وَ لاَ يظْلِمُ رَبُّكَ أَحَدً ا*"Dan Tuhanmu tidak menganiaya seorang juapun".* Maksudnya, Allah tidak akan menyiksa hambaNya karena dosa orang lain. Dari ayat ini bisa diambil kesimpulan, Allah akan mencatat setiap perbuatan hamba-hambaNya. Baik itu perbuatan baik, perbuatan buruk, dosa kecil maupun dosa besar. Semuanya tidak akan luput dari penglihatanNya (Al-Qurthubi, 2008c:920-922).

Bagitu pula dengan survei, untuk menghasilkan suatu penelitian dengan tingkat analisis mendekati kevalidan jumlah data harus lengkap. Surveior harus mencatat semua data yang diperoleh. Sehingga jika ada *missing data* akan mempengaruhi nilai MSE. Semakin banyak *missing data* maka semakin besar nilai MSE. Artinya akan semakin jauh dari sempurna. Demikian pula sebaliknya semakin sedikit *missing data* maka semakin kecil nilai MSE dan semakin mendekati sempurna.

BAB IV
PENUTUP

4.1 Kesimpulan

Berdasarkan pembahasan pada penelitian ini maka dapat diambil kesimpulan bahwa dari percobaan metode regresi imputasi mempunyai hasil yang memuaskan. Antara y_{miss} dan y_{imp} mempunyai kesalahan yang relatif kecil. Hal ini bisa dilihat dari MSE setiap sampel dan setiap jumlah data yang hilang. Untuk sampel berukuran 50 unit dengan y_{miss} 5% didapatkan $\text{MSE} = 4.36x10^{-5}$, y_{miss} 10% didapatkan $\text{MSE} = 1.66x10^{-4}$ dan y_{miss} 15% didapatkan $\text{MSE} = 2.51x10^{-4}$.

Sedangkan sampel berukuran 100 unit dengan y_{miss} 5% didapatkan $\text{MSE} = 2.87\text{E}x10^{-5}$, y_{miss} 10% didapatkan $\text{MSE} = 5.22091x10^{-5}$ dan y_{miss} 15% didapatkan $\text{MSE} = 5.28x10^{-5}$. Dan untuk sampel berukuran 200 unit dengan y_{miss} 5% didapatkan $\text{MSE} = 7.46x10^{-6}$, y_{miss} 10% didapatkan $\text{MSE} = 1.48x10^{-5}$ dan y_{miss} 1 5% didapatkan $\text{MSE} = 1.55x10^{-5}$.

Jika data yang diambil mempunyai jumlah yang sama tapi mempunyai persentase nilai hilang semakin besar diperoleh nilai MSE semakin besar dan jika data yang diambil mempunyai jumlah semakin besar tapi mempunyai persentase nilai hilang yang sama maka MSE semakin kecil.

Nilai MSE semakin besar ketika data yang hilang juga semakin besar. Hal ini dikarenakan semakin banyak data yang hilang maka data yang akan diimputkan pun juga akan semakin banyak. Sehingga akan banyak muncul nilai kesalahan dari hasil pengimputan data tadi dan demikian pula sebaliknya.

Nilai MSE semakin kecil ketika data yang diambil mempunyai jumlah semakin besar dengan persentase nilai hilang yang sama. Hal ini dikarenakan jumlah sampel yang besar semakin menggambarkan populasi.

Dalam Al-Qur'an dijelaskan bahwa Allah akan mencatat semua dosa, baik itu dosa kecil ataupun dosa besar. Firman Allah SWT dalam surat Al-Kahfi ayat 49:

لَا يُغَادِرُ صَغِيرَةً وَلَا كَبِيرَةً إِلَّآ أَحۡصَىٰهَاۚ وَوَجَدُواْ مَا عَمِلُواْ حَاضِرٗاۗ وَلَا يَظۡلِمُ رَبُّكَ أَحَدٗا

﴿٤٩﴾

Artinya: *"……tidak meninggalkan yang kecil dan tidak (pula) yang besar, melainkan ia mencatat semuanya; dan mereka dapati apa yang telah mereka kerjakan ada (tertulis). Dan Tuhanmu tidak Menganiaya seorang juapun (Qs. Al-Kahfi: 49)".*

Bagitu pula dengan survei, untuk menghasilkan suatu penelitian dengan tingkat analisis mendekati kevalidan jumlah data harus lengkap. Surveior harus mencatat semua data yang diperoleh. Sehingga jika ada *missing data* akan mempengaruhi nilai MSE. Semakin banyak *missing data* maka semakin besar nilai MSE. Artinya akan semakin jauh dari sempurna. Demikian pula sebaliknya semakin sedikit *missing data* maka semakin kecil nilai MSE dan semakin mendekati sempurna.

4.2 Saran

Pada penelitian selanjutnya dapat diteruskan dengan membandingkan metode regresi imputasi dan metode *robust* imputasi terhadap outlier.

DAFTAR PUSTAKA

Algifari. 2000. *Analisis Regresi Teori, Kasus, dan Solusi*. Yogyakarta: BPEE-Yogyakarta.

Alu, I.. 2007. *Tafsir Ibnu Katsir Jilid 1*. Bogor: Pustaka Imam Asy-Syafi'i.

Al-Qurthubi. 2008a. *Tafsir Al-Qurthubi Jilid 1*. Jakarta Selatan: Pustaka Azzam.

Al-Qurthubi. 2008b. *Tafsir Al-Qurthubi Jilid 10*. Jakarta Selatan: Pustaka Azzam.

Al-Qurthubi. 2008c. *Tafsir Al-Qurthubi Jilid 13*. Jakarta Selatan: Pustaka Azzam.

Banning, R., Camstar, A., Knottnerus, P.. 2012. *Sampling Theory, Sampling Design and Estimation Methods*. Netherlands: Statistic Netherlands.

Basuki, R.. 2010. Imputasi Berganda Menggunakan Metode Regresi dan Metode Predictive Mean Matching untuk Menangani Missing Data. *Tesis* Tidak Dipublikasikan Jurusan Statistika Fakultas MIPA. Surabaya: Institute Teknologi Sepuluh Nopember.

Boediono dan Koster, W.. 2004. *Teori dan Aplikasi Statistika dan Probabilitas*. Bandung: PT. Remaja Rosdakarya.

Cochran, W.G.. 2010. *Sampling Techniques*. Penj. Rudiansyah. Jakarta: UI-Press. Universitas Indonesia.

Dudewicz, E.J. dan Mishra, S. H.. 1995. *Modern Mathematical Statistics*. Penj. R.K. Sembiring. Bandung: Penerbit ITB.

Harini, S.. 2008. *Metode Statistika Pendekatan Teoritis dan Aplikatif*. Malang: UIN-Malang Press.

http://www.people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/evaluation.pdf(diunduh pada tanggal 12 Maret 2013).

Kish, L.. 1965. *Survey Sampling*. New York: Willey.

Little, R.J.A. dan Rubin, D.B.. 1987. *Statistical Analysis with Missing Data*. Los Angels: Massatchusetts.

Longford, N.T.. 2005. *Missing Data and Small-Area Estimation Modern Analytical Equipment for the Survey Statistician.* New York: Springer.

Malahayati, N.. 2008. Perbandingan Metode Imputasi Ganda: Metode Regresi Versus Metode Predective Mean Matching Untuk Mengatasi Data yang Hilang Pada Data Survei. *Skripsi* Tidak Dipublikasikan Jurusan Statistika Fakultas MIPA. Bogor: Intitute Pertanian Bogor.

Rao, J.N.K. dan Sitter, R. R.. 1995. Variance Estimation Two-Phase Sampling with Application to Imputation for Missing Data. *Biometrika*. Vol. 82 Hal. 453.

Rubin, D.B.. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: Willey.

Scheaffer, R.L., Mendenhall, W., dan Ott, L.. 1990. *Elementary Survey Sampling*. California: Duxbury Press.

Sumarningsih, E.. 2010. *Regresi*. Malang: Universitas Brawijaya.

Supranto, J. 2009. *Statistika Teori dan Aplikasi*. Jakarta: Erlangga.

Singh, S. dan Valdes, S.R.. 2009. Optimal Method of Imputation in Survey Sampling. *Applied Mathematical Sciences*. Vol. 35 Hal. 1727-1737.

Singh, S. dan Deo, B.. 2002. Imputation by Power Transformation. *Science and Management*. Vol. 44. Hal. 555-579.

Sudjana. 1992. *Teknik Analisis Regresi dan Korelasi*. Bandung: PT. Tarsito Bandung.

Yitnusumarto, S.. 1988. *Dasar-Dasar Statistika*. Malang: Universitas Brawijaya Program MIPA.

LAMPIRAN 1

Program matlab untuk membangkitkan data populasi

```
%% Membangkitkan data Populasi
clear;
Data = [49.4324 66.7163
56.899  71.045
60.8716 73.8709
57.0462 71.1986
61.9272 74.6957
60.6801 73.7195
66.6718 77.273
45.826  64.4776
50.9437 67.4766
61.7919 74.6008
49.8609 66.9404
57.1347 71.2669
59.4109 72.7755
60.2089 73.3183
54.5868 69.5767
56.7198 70.9461
58.7177 72.3472
69.7591 79.0619
45.7562 64.4586
50.8761 67.3748
54.0898 69.3025
54.1971 69.3967
59.6469 72.8951
57.1945 71.3045
56.5225 70.7828
58.3562 72.1079
53.5118 68.9023
55.9135 70.3363
59.6908 72.9341
56.9406 71.0917
59.6787 72.9249
61.2508 74.1383
52.2551 68.1591
56.4417 70.7002
69.1813 78.974
47.4949 65.6055
59.6207 72.8934
61.7111 74.5172
60.1028 73.2354
57.2394 71.3163
64.9534 76.5012
49.4265 66.7145
59.5014 72.7985
61.4507 74.2566
55.2888 69.9708
54.4116 69.4961
48.8691 66.253
61.0478 73.9923
56.6587 70.8927
46.8773 65.298
61.7837 74.5994
53.5326 68.9106
```

```
55.5621 70.1455
63.8222 75.8575
57.8411 71.7531
70.6627 80.3865
59.297  72.7088
58.2188 71.9686
55.0428 69.8403
60.1597 73.2873
53.9314 69.152
44.7839 63.7782
65.3905 76.6412
59.7752 73.0111
58.3445 72.0501
57.133  71.2655
58.557  72.2457
68.223  78.1382
47.3786 65.5552
64.1827 75.9467
59.7585 72.9709
62.2208 74.8943
64.5753 76.1953
51.687  67.8628
57.5698 71.604
63.4873 75.6435
62.8994 75.4704
61.3592 74.1686
54.6803 69.6147
63.4925 75.6624
57.5752 71.6272
65.5858 76.6875
53.1723 68.7722
62.3156 74.9539
64.2358 75.9784
53.8233 69.0765
64.5114 76.1693
59.1763 72.5925
60.022  73.2011
57.5408 71.5882
58.9646 72.4349
59.9014 73.0707
59.3503 72.724
57.0478 71.2074
57.0597 71.2202
49.9278 66.9526
56.5483 70.8065
55.6151 70.1673
55.861  70.3211
65.0771 76.5546];
X0 = Data(:,1); n=size(X0);
Y0 = Data(:,2);
X0_ = [ones(n,1) X0]
X_bar = mean(X0); Var_X = var(X0);
Y_bar = mean(Y0); Var_Y = var(Y0);
Cov_XY = cov(X0,Y0); Rho_XY = corr(X0,Y0);
Z = mvnrnd([X_bar Y_bar],Cov_XY,1000);
X = Z(:,1); Y = Z(:,2);
```

LAMPIRAN 2

Program matlab untuk menghitung beta berdasarkan data sampel, menghitung nilai imputasi, dan MSE

```
%% Menghitung Beta berdasarkan data sampel
Sampel=Data;
Xn = Sampel(:,1); n=size(Xn); Xn_bar = mean(Xn);
Yn = Sampel(:,2);
Xr =[Xn(1:2-1);Xn(2+10:100)]; r=size(Xr); Xr_bar = mean(Xr);
Yr =[Yn(1:2-1);Yn(2+10:100)]; Yr_bar=mean(Yr);
Xr_ = [ones(r,1) Xr];
Betahat = regress(Yr,Xr_); Beta = Betahat(2)
for i = 1:10,
    Yimp(i,1) = Yr_bar + Beta*(Xn(i)-Xr_bar);
end
Yimp = [Yimp;Yr];
Yreg_bar = Yr_bar + Beta*(Xn_bar-Xr_bar);
Yimp_bar = mean(Yimp);
MSE = (Yreg_bar-Y_bar)^2;
```

LAMPIRAN 3

Hasil Percobaan y_{imp}

$n = 100$ dengan y_{miss} 5%

Percobaan 1		Percobaan 2		Percobaan 3		Percobaan 4		Percobaan 5	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
68.13705	68.3669	70.29696	70.38064	79.46706	79.51019	76.024	75.98447	67.67319	67.49455
71.33015	71.47436	69.84623	69.76592	73.77797	73.70571	69.35151	69.4563	69.60563	69.67801
66.35335	66.14828	76.67029	76.62125	76.93691	77.55116	73.93381	73.55497	67.55528	67.65153
69.96358	69.98568	68.69762	68.68523	70.59593	70.56617	76.63978	76.22787	69.93335	69.89578
69.36082	69.52139	73.22134	73.01815	69.599	69.7305	74.83827	74.48122	78.67108	78.67748
$\bar{y}_{miss}$	71.56587	$\bar{y}_{miss}$	72.22573	$\bar{y}_{miss}$	72.37474	$\bar{y}_{miss}$	72.05977	$\bar{y}_{miss}$	71.77149

Percobaan 6		Percobaan 7		Percobaan 8		Percobaan 9		Percobaan 10	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
70.01752	70.0844	71.38223	71.37446	71.71859	71.71301	72.93899	72.67171	77.65207	77.71965
66.62901	66.50649	73.57919	73.8228	69.96805	69.98381	74.91252	74.62265	69.69708	69.74023
69.73274	69.86689	75.29848	74.97467	76.41085	76.4035	67.80725	67.65744	73.85828	73.88989
82.64609	82.72797	67.2292	67.08069	77.92383	77.78461	68.87168	68.83535	73.3845	72.89383
65.48298	65.38351	67.82085	67.66899	68.71491	68.6201	75.74348	75.67107	73.32673	73.27374

$\bar{y}_{miss}$	72.03292	$\bar{y}_{miss}$	71.73136	$\bar{y}_{miss}$	71.78852	$\bar{y}_{miss}$	71.8217	$\bar{y}_{miss}$	71.69711

n 100 dengan y_{miss} 10%

Percobaan 1		Percobaan 2		Percobaan 3		Percobaan 4		Percobaan 5	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
68.13705	68.36434	70.29696	70.38069	79.46706	79.52854	76.024	75.9946	67.67319	67.49675
71.33015	71.47555	69.84623	69.76634	73.77797	73.71115	69.35151	69.44516	69.60563	69.67835
66.35335	66.14304	76.67029	76.61747	76.93691	77.56516	73.93381	73.55719	67.55528	67.6536
69.96358	69.98508	68.69762	68.68632	70.59593	70.56464	76.63978	76.2388	69.93335	69.89593
69.36082	69.52023	73.22134	73.01658	69.599	69.72711	74.83827	74.48646	78.67108	78.67014
75.16866	75.38236	74.24155	74.2016	71.61677	71.67231	75.74428	75.98722	76.34721	76.12135
67.79732	67.86647	68.39286	68.18894	75.17079	75.272	70.02563	69.98746	72.81799	72.92647
67.70553	67.55673	75.73765	75.50845	77.51045	77.63303	74.85534	74.82768	70.10595	69.87709
69.6663	69.64395	72.66929	72.70598	79.94365	80.13141	65.9466	65.57281	71.22907	71.35654
75.47144	75.49875	69.63691	69.96868	71.95445	71.70292	67.20707	67.12072	65.18087	65.25732
$\bar{y}_{miss}$	71.56718	$\bar{y}_{miss}$	72.22465	$\bar{y}_{miss}$	72.37722	$\bar{y}_{miss}$	72.05712	$\bar{y}_{miss}$	71.77004

Percobaan 6		Percobaan 7		Percobaan 8		Percobaan 9		Percobaan 10	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
70.01752	70.08411	71.38223	71.3674	71.71859	71.71181	72.93899	72.67198	77.65207	77.72086
66.62901	66.49917	73.57919	73.81569	69.96805	69.98277	74.91252	74.61972	69.69708	69.73922

69.73274	69.86618	75.29848	74.96755	76.41085	76.40187	67.80725	67.66594	73.85828	73.89004
82.64609	82.75253	67.2292	67.0737	77.92383	77.78285	68.87168	68.84192	73.3845	72.8937
65.48298	65.37398	67.82085	67.66199	68.71491	68.61919	75.74348	75.66642	73.32673	73.27371
75.14348	75.43053	76.88741	76.82984	77.20953	77.20562	76.12578	75.87481	67.79809	67.78367
68.0909	68.28634	72.66421	72.59414	62.64616	62.50525	71.6469	71.60217	72.31214	72.66585
69.46489	69.34006	69.65984	69.45955	74.40212	74.36834	70.2118	70.41867	71.19362	71.04533
66.93845	66.76099	71.70837	71.6259	68.86212	69.14128	66.28549	66.34357	72.27539	72.06803
74.29062	74.45828	72.56926	72.30811	74.68849	74.4741	72.31926	72.5085	76.29141	76.26079
$\bar{y}_{miss}$	72.03646	$\bar{y}_{miss}$	71.72429	$\bar{y}_{miss}$	71.78732	$\bar{y}_{miss}$	71.82336	$\bar{y}_{miss}$	71.69664

$n = 100$ dengan y_{miss} 15%

Percobaan 1		Percobaan 2		Percobaan 3		Percobaan 4		Percobaan 5	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
68.13705	68.36378	70.29696	70.37722	79.46706	79.53182	76.024	75.99743	67.67319	67.49897
71.33015	71.47018	69.84623	69.76303	73.77797	73.70825	69.35151	69.44743	69.60563	69.6833
66.35335	66.14591	76.67029	76.61243	76.93691	77.56635	73.93381	73.55981	67.55528	67.65602
69.96358	69.98201	68.69762	68.68328	70.59593	70.55839	76.63978	76.24165	69.93335	69.90116
69.36082	69.51788	73.22134	73.01245	69.599	69.71997	74.83827	74.48916	78.67108	78.68632
75.16866	75.37095	74.24155	74.19717	71.61677	71.66724	75.74428	75.99005	76.34721	76.13434
67.79732	67.86668	68.39286	68.18603	75.17079	75.27076	70.02563	69.98977	72.81799	72.93547
67.70553	67.55742	75.73765	75.50369	77.51045	77.6343	74.85534	74.83041	70.10595	69.88229
69.6663	69.64142	72.66929	72.70193	79.94365	80.13534	65.9466	65.57474	71.22907	71.36358

75.47144	75.48715	69.63691	69.96532	71.95445	71.69788	67.20707	67.12279	65.18087	65.25675
80.34129	80.16625	74.92171	74.87632	69.51225	69.18896	72.42741	72.28555	75.00272	75.19205
70.03421	70.14075	72.96728	73.11693	72.42975	72.41593	72.67394	72.77482	69.87543	69.84967
68.08623	67.92856	75.36878	75.22218	74.31795	74.44351	76.18119	75.97545	68.21323	68.39062
71.99305	71.81828	69.66826	69.54096	72.75941	72.52238	72.89427	73.26134	75.28575	75.54282
75.95746	75.84027	73.59724	73.41201	71.41329	71.44778	75.91114	76.01469	75.26799	75.36152
$\bar{y}_{miss}$	71.56167	$\bar{y}_{miss}$	72.22071	$\bar{y}_{miss}$	72.3729	$\bar{y}_{miss}$	72.05961	$\bar{y}_{miss}$	71.7776

Percobaan 6		Perconaan 7		Percobaan 8		Percobaan 9		Percobaan 10	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
70.01752	70.08505	71.38223	71.36451	71.71859	71.71171	72.93899	72.67372	77.65207	77.72127
66.62901	66.49933	73.57919	73.81432	69.96805	69.98331	74.91252	74.62635	69.69708	69.74003
69.73274	69.86706	75.29848	74.96689	76.41085	76.40003	67.80725	67.65508	73.85828	73.89064
82.64609	82.75621	67.2292	67.06814	77.92383	77.7805	68.87168	68.83402	73.3845	72.89435
65.48298	65.37389	67.82085	67.6568	68.71491	68.62023	75.74348	75.67568	73.32673	73.27435
75.14348	75.43262	76.88741	76.83034	77.20953	77.20348	76.12578	75.8846	67.79809	67.78458
68.0909	68.28688	72.66421	72.59201	62.64616	62.50855	71.6469	71.60121	72.31214	72.66651
69.46489	69.34083	69.65984	69.45548	74.40212	74.36726	70.2118	70.41473	71.19362	71.04608
66.93845	66.7612	71.70837	71.62317	68.86212	69.14212	66.28549	66.32939	72.27539	72.06872
74.29062	74.46016	72.56926	72.3058	74.68849	74.47298	72.31926	72.50982	76.29141	76.26127
71.4653	71.58466	68.92143	68.81796	73.48341	73.48493	70.6854	70.81971	73.92173	73.84107
71.51056	71.3532	69.80595	69.90167	79.94111	79.85482	71.65661	71.99508	73.7448	73.86158
75.68597	76.06801	73.31307	73.22279	74.62503	74.29273	62.23747	61.94972	63.0439	63.11599

76.89112	76.6142	73.97506	73.9944	76.39985	76.66648	76.92687	76.8099	66.14277	66.07987
69.97696	70.03432	67.86434	67.70253	70.14471	70.28489	72.18955	72.09072	78.38103	78.40042
$\bar{y}_{miss}$	72.03782	$\bar{y}_{miss}$	71.72163	$\bar{y}_{miss}$	71.78719	$\bar{y}_{miss}$	71.82296	$\bar{y}_{miss}$	71.69736

$n = 200$ dengan y_{miss} 5%

Percobaan 1		Perconaan 2		Percobaan 3		Percobaan 4		Percobaan 5	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
68.13705	68.36695	79.46706	79.51373	67.67319	67.47328	71.38223	71.3598	72.93899	72.69118
71.33015	71.47871	73.77797	73.72259	69.60563	69.66542	73.57919	73.81417	74.91252	74.64201
66.35335	66.14527	76.93691	77.55921	67.55528	67.63088	75.29848	74.96888	67.80725	67.67721
69.96358	69.98797	70.59593	70.59028	69.93335	69.88406	67.2292	67.05545	68.87168	68.85505
69.36082	69.52304	69.599	69.75653	78.67108	78.70067	67.82085	67.6452	75.74348	75.69037
$\bar{y}_{miss}$	71.89713	$\bar{y}_{miss}$	72.22306	$\bar{y}_{miss}$	71.90168	$\bar{y}_{miss}$	71.76071	$\bar{y}_{miss}$	71.7619

Percobaan 6		Perconaan 7		Percobaan 8		Percobaan 9		Percobaan 10	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
70.29696	70.36984	76.024	76.00155	70.01752	70.10713	71.71859	71.71884	77.65207	77.66554
69.84623	69.75664	69.35151	69.45253	66.62901	66.55425	69.96805	69.98244	69.69708	69.73111
76.67029	76.59508	73.93381	73.5643	69.73274	69.89115	76.41085	76.42886	73.85828	73.85738
68.69762	68.67862	76.63978	76.24573	82.64609	82.66228	77.92383	77.81572	73.3845	72.86693
73.22134	73.00086	74.83827	74.4935	65.48298	65.43912	68.71491	68.61306	73.32673	73.2447
$\bar{y}_{miss}$	72.29645	$\bar{y}_{miss}$	71.9161	$\bar{y}_{miss}$	71.88451	$\bar{y}_{miss}$	71.80946	$\bar{y}_{miss}$	71.62897

$n = 200$ dengan y_{miss} 10%

Percobaan 1		Percobaan 2		Percobaan 3		Percobaan 4		Percobaan 5	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
68.13705	68.36576	79.46706	79.5229	67.67319	67.47361	71.38223	71.35614	72.93899	72.69177
71.33015	71.47924	73.77797	73.72558	69.60563	69.66515	73.57919	73.81054	74.91252	74.64111
66.35335	66.14285	76.93691	77.56629	67.55528	67.63117	75.29848	74.96527	67.80725	67.6816
69.96358	69.98768	70.59593	70.58992	69.93335	69.88372	67.2292	67.05172	68.87168	68.85855
69.36082	69.52249	69.599	69.75529	78.67108	78.6979	67.82085	67.64149	75.74348	75.68867
75.16866	75.38888	71.61677	71.69377	76.34721	76.1375	76.88741	76.83221	76.12578	75.89724
67.79732	67.86753	75.17079	75.28105	72.81799	72.92806	72.66421	72.58594	71.6469	71.62107
67.70553	67.55757	77.51045	77.63393	70.10595	69.8648	69.65984	69.44353	70.2118	70.4366
69.6663	69.64631	79.94365	80.12369	71.22907	71.35098	71.70837	71.61528	66.28549	66.35814
75.47144	75.50535	71.95445	71.72428	65.18087	65.22398	72.56926	72.29919	72.31926	72.52815
$\bar{y}_{miss}$	71.8979	$\bar{y}_{miss}$	72.22445	$\bar{y}_{miss}$	71.90079	$\bar{y}_{miss}$	71.75705	$\bar{y}_{miss}$	71.76319

Percobaan 6		Percobaan 7		Percobaan 8		Percobaan 9		Percobaan	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
70.29696	70.36956	76.024	76.00691	70.01752	70.10823	71.71859	71.71839	77.65207	77.66477
69.84623	69.75655	69.35151	69.44727	66.62901	66.55309	69.96805	69.98162	69.69708	69.73029
76.67029	76.59277	73.93381	73.5657	69.73274	69.89211	76.41085	76.42943	73.85828	73.85659
68.69762	68.67888	76.63978	76.25149	82.64609	82.67134	77.92383	77.81658	73.3845	72.86613
73.22134	72.99972	74.83827	74.49642	65.48298	65.43725	68.71491	68.61194	73.32673	73.2439

74.24155	74.18216	75.74428	75.99952	75.14348	75.4102	77.20953	77.23677	67.79809	67.7863
68.39286	68.18259	70.02563	69.99042	68.0909	68.3254	62.64616	62.47066	72.31214	72.63963
75.73765	75.48616	74.85534	74.83817	69.46489	69.37036	74.40212	74.38681	71.19362	71.02869
72.66929	72.6898	65.9466	65.5689	66.93845	66.81273	68.86212	69.13637	72.27539	72.04534
69.63691	69.95845	67.20707	67.11922	74.29062	74.44603	74.68849	74.49304	76.29141	76.21333
$\bar{y}_{miss}$	72.29554	$\bar{y}_{miss}$	71.91484	$\bar{y}_{miss}$	71.88673	$\bar{y}_{miss}$	71.80903	$\bar{y}_{miss}$	71.62817

$n = 200$ dengan y_{miss} 15%

Percobaan 1		Percobaan 2		Percobaan 3		Percobaan 4		Percobaan 5	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
68.13705	68.36521	79.46706	79.52471	67.67319	67.47424	71.38223	71.35444	72.93899	72.69311
71.33015	71.47704	73.77797	73.72476	69.60563	69.66721	73.57919	73.80965	74.91252	74.64449
66.35335	66.14347	76.93691	77.56721	67.55528	67.63191	75.29848	74.96476	67.80725	67.67772
69.96358	69.98627	70.59593	70.58768	69.93335	69.88592	67.2292	67.0486	68.87168	68.85589
69.36082	69.52133	69.599	69.75266	78.67108	78.70583	67.82085	67.63856	75.74348	75.69314
75.16866	75.38461	71.61677	71.69203	76.34721	76.14376	76.88741	76.83232	76.12578	75.90192
67.79732	67.86724	75.17079	75.28093	72.81799	72.93224	72.66421	72.58465	71.6469	71.6213
67.70553	67.55744	77.51045	77.63488	70.10595	69.86698	69.65984	69.4412	70.2118	70.43559
69.6663	69.64508	79.94365	80.12577	71.22907	71.35414	71.70837	71.61367	66.28549	66.35288
75.47144	75.50102	71.95445	71.72255	65.18087	65.22315	72.56926	72.2978	72.31926	72.52932
80.34129	80.18829	69.51225	69.22381	75.00272	75.19774	68.92143	68.80228	70.6854	70.8403
70.03421	70.14529	72.42975	72.43769	69.87543	69.83423	69.80595	69.88838	71.65661	72.01491

68.08623	67.92923	74.31795	74.45704	68.21323	68.36941	73.31307	73.21682	62.23747	61.97603
71.99305	71.82574	72.75941	72.5437	75.28575	75.5499	73.97506	73.99013	76.92687	76.82663
75.95746	75.85475	71.41329	71.47346	75.26799	75.36789	67.86434	67.68438	72.18955	72.11049
$\bar{y}_{miss}$	71.89548	$\bar{y}_{miss}$	72.22295	$\bar{y}_{miss}$	71.9043	$\bar{y}_{reg}$ $/\bar{y}_{miss}$	71.75548	$\bar{y}_{miss}$	71.76356

Percobaan 6		Percobaan 7		Percobaan 8		Percobaan 9		Percobaan 10	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
70.29696	70.36759	76.024	76.00872	70.01752	70.10891	71.71859	71.71894	77.65207	77.66171
69.84623	69.75468	69.35151	69.44842	66.62901	66.55393	69.96805	69.98202	69.69708	69.73107
76.67029	76.58981	73.93381	73.56727	69.73274	69.8928	76.41085	76.43037	73.85828	73.85537
68.69762	68.67718	76.63978	76.25333	82.64609	82.67144	77.92383	77.81764	73.3845	72.86539
73.22134	72.99733	74.83827	74.49808	65.48298	65.43815	68.71491	68.61223	73.32673	73.24298
74.24155	74.17959	75.74428	76.00133	75.14348	75.41064	77.20953	77.23778	67.79809	67.78802
68.39286	68.18096	70.02563	69.99162	68.0909	68.32616	62.64616	62.47043	72.31214	72.639
75.73765	75.48338	74.85534	74.83987	69.46489	69.37107	74.40212	74.38758	71.19362	71.02884
72.66929	72.68746	65.9466	65.56965	66.93845	66.81356	68.86212	69.1367	72.27539	72.045
69.63691	69.95655	67.20707	67.12013	74.29062	74.44651	74.68849	74.49382	76.29141	76.21097
74.92171	74.85732	72.42741	72.291	71.4653	71.59567	73.48341	73.50091	73.92173	73.80611
72.96728	73.10159	72.67394	72.78105	71.51056	71.36619	79.94111	79.90219	73.7448	73.82649
75.36878	75.20246	76.18119	75.98671	75.68597	76.04058	74.62503	74.31269	63.0439	63.14903
69.66826	69.53307	72.89427	73.26833	76.89112	76.58209	76.39985	76.69814	66.14277	66.09412
73.59724	73.39606	75.91114	76.02601	69.97696	70.05861	70.14471	70.28509	78.38103	78.33656

$\bar{y}_{miss}$	72.29326	$\bar{y}_{miss}$	71.91624	$\bar{y}_{miss}$	71.88733	$\bar{y}_{miss}$	71.80959	$\bar{y}_{miss}$	71.62803

$n = 50$ dengan y_{miss} 5%

Percobaan 1		Percobaan 2		Percobaan 3		Percobaan 4		Percobaan 5	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
62.58042	62.8221	65.81879	65.61137	74.92853	74.90514	67.98587	68.11207	65.98336	66.25051
65.92199	65.90003	74.6847	74.46256	75.14198	74.97075	66.46953	66.09481	67.05699	66.78943
73.6299	73.48952	72.42646	72.54677	75.22449	75.35742	71.86167	71.93787	74.08378	73.79935
70.66478	70.93839	73.20004	72.96565	75.66055	75.84696	68.22218	68.57325	74.13115	74.20341
67.85033	68.04286	71.1057	71.42092	72.24574	72.33201	78.80196	78.8922	71.37042	71.81352
$\bar{y}_{miss}$	71.42461	$\bar{y}_{miss}$	72.15005	$\bar{y}_{miss}$	72.48295	$\bar{y}_{miss}$	71.94901	$\bar{y}_{miss}$	71.8655

Percobaan 6		Percobaan 7		Percobaan 8		Percobaan 9		Percobaan 10	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
76.22247	75.99657	72.69957	72.53697	66.75275	66.67033	70.46283	70.76241	73.50633	73.52056
71.52922	71.65459	73.17054	73.58127	76.38114	76.44506	69.59459	69.17568	72.35865	72.34673
68.95429	68.90677	76.52541	77.04936	75.25299	74.95416	71.71919	71.71298	70.95326	71.02599
72.60884	72.37583	71.4748	71.14993	70.67105	70.88391	75.14895	75.18971	70.33671	70.35853
66.69229	66.6409	71.79939	71.87007	76.15678	76.35475	69.11753	68.98726	69.52368	69.18119
$\bar{y}_{miss}$	71.98731	$\bar{y}_{miss}$	72.21135	$\bar{y}_{miss}$	71.31101	$\bar{y}_{miss}$	71.69172	$\bar{y}_{miss}$	71.65523

$n = 50$ dengan y_{miss} 10%

Percobaan 1		Percobaan 2		Percobaan 3		Percobaan 4		Percobaan 5	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
62.58042	62.81739	65.81879	65.60937	74.92853	74.90938	67.98587	68.13712	65.98336	66.21167
65.92199	65.89541	74.6847	74.4675	75.14198	74.97485	66.46953	66.12574	67.05699	66.75233
73.6299	73.48511	72.42646	72.55021	75.22449	75.36071	71.86167	71.95176	74.08378	73.78478
70.66478	70.93391	73.20004	72.96942	75.66055	75.84922	68.22218	68.59695	74.13115	74.19014
67.85033	68.03829	71.1057	71.42348	72.24574	72.34166	78.80196	78.88581	71.37042	71.79257
71.39761	71.58346	71.55625	71.47763	72.61625	72.57182	74.69251	74.6972	74.87672	74.90534
74.32986	74.31375	70.52036	70.71147	74.4652	74.22033	73.08364	73.49817	70.44894	70.36769
69.58155	69.36373	72.31004	72.21471	72.64802	72.77049	63.89882	64.16231	69.05606	68.86868
75.40759	75.22257	79.24718	79.20119	74.13092	74.16515	73.95144	74.00251	72.52056	72.23483
77.92119	77.95347	77.96431	78.13688	70.56263	71.13826	73.94308	73.81509	69.72378	69.335
$\bar{y}_{miss}$	71.42013	$\bar{y}_{miss}$	72.15318	$\bar{y}_{miss}$	72.49228	$\bar{y}_{miss}$	71.96287	$\bar{y}_{miss}$	71.84471

Percobaan 6		Percobaan 7		Percobaan 8		Percobaan 9		Percobaan 10	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
76.22247	75.99174	72.69957	72.54776	66.75275	66.64745	70.46283	70.7578	73.50633	73.5182
71.52922	71.65524	73.17054	73.5928	76.38114	76.49279	69.59459	69.17339	72.35865	72.34529
68.95429	68.91091	76.52541	77.06338	75.25299	74.99112	71.71919	71.70699	70.95326	71.02558
72.60884	72.37557	71.4748	71.15973	70.67105	70.89147	75.14895	75.17865	70.33671	70.35864
66.69229	66.6479	71.79939	71.88038	76.15678	76.40182	69.11753	68.98524	69.52368	69.18223

74.12213	74.07044	69.87152	70.1526	68.22942	68.32709	68.42887	68.25817	69.9463	69.84072
78.15818	77.90552	76.19897	76.15621	70.74333	70.72998	70.59064	70.7974	74.80214	74.84302
73.69883	73.87616	77.04287	77.21476	75.34118	75.82504	70.57332	70.44117	72.27688	72.31144
76.70377	76.84375	70.4123	70.58871	73.57691	73.59854	76.97386	76.77637	68.57506	68.70388
73.07449	73.06665	68.52564	68.41028	65.48982	65.31566	70.17454	70.1961	73.7407	73.60017
$\bar{y}_{miss}$	71.98754	$\bar{y}_{miss}$	72.2219	$\bar{y}_{miss}$	71.32165	$\bar{y}_{miss}$	71.68576	$\bar{y}_{miss}$	71.65433

$n = 50$ dengan y_{miss} 15%

Percobaan 1		Percobaan 2		Percobaan 3		Percobaan 4		Percobaan 5	
y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
62.58042	62.83937	65.81879	65.61173	74.92853	74.91518	67.98587	68.15243	65.98336	66.20849
65.92199	65.91136	74.6847	74.43253	75.14198	74.98067	66.46953	66.16166	67.05699	66.74867
73.6299	73.48621	72.42646	72.52332	75.22449	75.36664	71.86167	71.92799	74.08378	73.7749
70.66478	70.94001	73.20004	72.94076	75.66055	75.8553	68.22218	68.60755	74.13115	74.1799
67.85033	68.05005	71.1057	71.40134	72.24574	72.34669	78.80196	78.791	71.37042	71.78445
71.39761	71.58829	71.55625	71.45525	72.61625	72.57692	74.69251	74.6453	74.87672	74.89447
74.32986	74.31324	70.52036	70.69233	74.4652	74.22592	73.08364	73.45855	70.44894	70.36084
69.58155	69.3729	72.31004	72.18923	72.64802	72.77565	63.89882	64.21834	69.05606	68.86315
75.40759	75.22028	79.24718	79.14626	74.13092	74.17073	73.95144	73.95773	72.52056	72.22632
77.92119	77.94584	77.96431	78.08644	70.56263	71.14293	73.94308	73.77222	69.72378	69.32906
71.39076	71.51623	76.79081	76.38687	76.13284	76.12351	80.04847	79.57289	76.57129	76.66087
72.59472	72.64918	71.6571	71.61783	74.06293	74.24485	74.03998	73.81824	67.31368	67.57213

67.64302	67.87377	74.04364	73.96442	77.74699	77.65124	75.8762	75.80661	74.69277	74.62197
73.01662	72.94973	72.98994	72.63102	78.8282	78.82967	74.80773	74.63222	70.39225	69.90415
69.7518	69.60482	71.50969	71.41271	74.64283	74.76359	74.54452	74.47252	76.94689	76.82238
$\bar{y}_{miss}$	71.42528	$\bar{y}_{miss}$	72.12796	$\bar{y}_{miss}$	72.49735	$\bar{y}_{miss}$	71.93898	$\bar{y}_{miss}$	71.83655

Percobaan 6		Percobaan 7		Percobaan 8		Percobaan 9		Percobaan 10	
y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}	y_{miss}	y_{imp}
76.22247	76.00683	72.69957	72.55063	66.75275	66.62471	70.46283	70.78005	73.50633	73.52935
71.52922	71.65035	73.17054	73.60009	76.38114	76.49062	69.59459	69.21463	72.35865	72.35785
68.95429	68.89337	76.52541	77.08533	75.25299	74.98582	71.71919	71.71786	70.95326	71.03973
72.60884	72.374	71.4748	71.15674	70.67105	70.8776	75.14895	75.14792	70.33671	70.3736
66.69229	66.61993	71.79939	71.88044	76.15678	76.39947	69.11753	69.02874	69.52368	69.1986
74.12213	74.07668	69.87152	70.14536	68.22942	68.30787	68.42887	68.31037	69.9463	69.85629
78.15818	77.92943	76.19897	76.17432	70.74333	70.71578	70.59064	70.81917	74.80214	74.85256
73.69883	73.8815	77.04287	77.23735	75.34118	75.82148	70.57332	70.46722	72.27688	72.32404
76.70377	76.86277	70.4123	70.5833	73.57691	73.59033	76.97386	76.72649	68.57506	68.72083
73.07449	73.06826	68.52564	68.39568	65.48982	65.29014	70.17454	70.22508	73.7407	73.61122
73.3638	73.22333	70.42913	70.57633	71.09281	70.76648	70.63669	70.39626	68.06875	68.27385
75.58018	75.79687	75.52409	75.6313	71.85	71.48788	70.6708	70.69812	69.9638	70.18628
71.80306	71.79893	68.87334	68.8042	71.6415	71.70758	73.13443	73.05443	73.17915	73.42175
67.25632	67.00486	66.19019	65.95852	69.72615	69.58319	68.79286	69.08696	70.45381	70.2801
71.34835	71.34038	69.35601	69.43466	72.9767	73.2107	64.73115	65.12045	69.78632	69.82453
$\bar{y}_{miss}$	71.98418	$\bar{y}_{miss}$	72.2234	$\bar{y}_{miss}$	71.30868	$\bar{y}_{miss}$	71.69689	$\bar{y}_{miss}$	71.66773

LAMPIRAN 4

MSE *(Mean Square Error)* dari sampel 50, 100 dan 200

Percobaan	100			200			50		
	10%	5%	15%	5%	10%	15%	5%	10%	15%
1	2.32601E-05	1.24E-05	4.75E-07	3.19E-06	6.49E-06	1.63E-08	1.19E-04	4.14E-05	1.34E-04
2	1.36626E-05	6.82E-06	0.998767	1.46E-05	2.72E-05	1.38E-05	2.09E-05	2.07E-06	7.11E-04
3	8.7472E-05	4.72E-05	2.53E-05	1.77E-07	1.73E-06	4.84E-06	1.78E-05	1.84E-04	3.47E-04
4	0.000182258	1.17E-04	1.21E-04	5.44E-06	3.59E-05	5.71E-05	2.89E-05	3.70E-04	2.16E-05
5	3.61E-06	1.69E-07	3.25E-05	1.29E-05	5.30E-06	3.71E-06	2.13E-05	2.62E-04	5.92E-04
6	1.72639E-05	3.71E-07	3.04E-05	2.75E-06	6.61E-06	2.35E-05	7.48E-05	7.08E-05	1.39E-04
7	0.00011994	1.51E-05	1.86E-04	2.65E-05	4.11E-05	2.51E-05	1.07E-04	4.37E-04	5.02E-04
8	1.23606E-05	5.35E-06	1.33E-05	5.30E-07	8.70E-06	1.26E-05	3.50E-06	1.57E-04	2.09E-07
9	4.21834E-05	6.65E-05	4.75E-05	7.86E-07	1.73E-06	5.79E-07	1.85E-05	1.05E-04	7.57E-07
10	2.00801E-05	1.61E-05	1.42E-05	7.65E-06	1.28E-05	1.38E-05	2.41E-05	3.38E-05	5.75E-05
Total	**0.000522091**	**2.87E-04**	**5.28E-04**	**7.46E-05**	**1.48E-04**	**1.55E-04**	**4.36E-04**	**1.66E-03**	**2.51E-03**
Rata-rata	**5.22091E-05**	**2.87E-05**	**5.28E-05**	**7.46E-06**	**1.48E-05**	**1.55E-05**	**4.36E-05**	**1.66E-04**	**2.51E-04**

LAMPIRAN 5

Grafik Persentase MSE *(Mean Square Error)*

n	y_{miss} %		
	5	**10**	**15**
50	$4.36x10^{-5}$	$1.66x10^{-4}$	$2.51x10^{-4}$
100	$2.87Ex10^{-5}$	$5.22091x10^{-5}$	$5.28x10^{-5}$
200	$7.46x10^{-6}$	$1.48x10^{-5}$	$1.55x10^{-5}$