PENANGANAN OVERDISPERSI MENGGUNAKAN REGRESI M-KUANTIL DENGAN RESAMPLING BOOTSTRAP UNTUK *SMALL AREA ESTIMATION*

SKRIPSI

Oleh
Sella Aji Oktarin
NIM 111810101023

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2015

PENANGANAN OVERDISPERSI MENGGUNAKAN REGRESI M-KUANTIL DENGAN RESAMPLING BOOTSTRAP UNTUK *SMALL AREA ESTIMATION*

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat
untuk menyelesaikan Program Studi Matematika (S1)
dan mencapai gelas Sarjana Sains

Oleh
Sella Aji Oktarin
NIM 111810101023

JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2015

PERSEMBAHAN

Skripsi ini saya persembahkan untuk :

1. Allah SWT yang telah memberikan semua kemudahan dan kesempurnaan dalam kehidupan ini;
2. Ayahanda Wakijo dan Ibunda Suryati yang telah memberikan doa, cinta, kasih dan semangatnya;
3. guru-guru dan dosen sejak taman kanak-kanak hingga perguruan tinggi yang telah memberikan ilmu dan membimbing dengan penuh perhatian dan kesabaran;
4. Almamater Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

MOTTO

"Orang-orang yang sukses telah belajar membuat diri mereka melakukan hal yang harus dikerjakan ketika hal itu memang harus dikerjakan, entah mereka menyukainya atau tidak." (Aldus Huxley)*)

"Banyak kegagalan dalam hidup ini dikarenakan orang-orang tidak menyadari betapa dekatnya mereka dengan keberhasilan saat mereka menyerah." (Thomas Alva Edison)**)

*) http://www.maribelajarbk.web.id/2015/03/contoh-motto-terbaru-dalam-skripsi.html

**) http://ananda-7.blogspot.com/2012/08/kumpulan-contoh-kata-motto-skripsi.html#ixzz3eKcdo14U

PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Sella Aji Oktarin

NIM : 111810101023

menyatakan dengan sesungguhnya bahwa skripsi yang berjudul "Penanganan Overdispersi Menggunakan Regresi M-Kuantil Dengan Resampling Bootstrap Untuk *Small Area Estimation*" adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan dalam institusi manapun dan juga bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, Agustus 2015

Yang menyatakan,

Sella Aji Oktarin

NIM 111810101023

SKRIPSI

PENANGANAN OVERDISPERSI MENGGUNAKAN REGRESI M-KUANTIL DENGAN RESAMPLING BOOTSTRAP UNTUK *SMALL AREA ESTIMATION*

Oleh

Sella Aji Oktarin
NIM 111810101023

Pembimbing:

Dosen Pembimbing Utama : Dian Anggraeni, S.Si., M.Si

Dosen Pembimbing Anggota : Dr. Alfian Futuhul Hadi, S.Si., M.Si

PENGESAHAN

Skripsi berjudul “Penanganan Overdispersi Menggunakan Regresi M-Kuantil Dengan Resampling Bootstrap Untuk *Small Area Estimation*” telah diuji dan disahkan pada :

Hari, tanggal :

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember

Tim penguji,

Ketua,	Sekretaris,
Dian Anggraeni, S. Si, M. Si NIP. 198202162006042002	Dr. Alfian Futuhul Hadi, S. Si, M. Si NIP. 197407192000121001
Anggota I,	Anggota II,
Prof. Drs. I Made Tirta M. Sc., Ph. D NIP. 195912201985031002	M. Ziaul Arif, S.Si., M.Sc NIP. 198501112008121002

Mengesahkan

Dekan,

Prof. Drs. Kusno, DEA, Ph.D.
NIP 196101081986021001

RINGKASAN

Penanganan Overdispersi Menggunakan Regresi M-Kuantil Dengan Resampling Bootstrap Untuk *Small Area Estimation*; Sella Aji Oktarin; 2015; 27 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Suatu area dikatakan kecil apabila sampel yang diambil pada area tersebut tidak mencukupi untuk melakukan suatu pendugaan langsung dengan hasil dugaan yang akurat. *Small area estimation* (SAE) merupakan suatu teknik statistika untuk menduga parameter-parameter pada area kecil dengan memanfaatkan informasi dari dalam area tersebut, dari luar area, dan dari luar survei.

Penelitian ini dilakukan pada model regresi Poisson M-kuantil untuk permasalahan untuk SAE pada pendugaan langsung dan pendugaan tidak langsung. Data yang digunakan merupakan data simulasi Anggraini (2014). Data tersebut merupakan data cacahan yang mengandung overdispersi dengan area yang digunakan yaitu area 1 sampai area 20. Untuk selanjutnya dilakukan pendugaan parameter untuk kedua model yaitu Poisson dan M-kuantil. Setelah didapat pendugaan parameter, dilakukan resampling bootstrap sebanyak 1000 kali untuk mengetahui nilai rata-rata *Mean Square Error* (MSE). Rata-rata MSE tersebut digunakan untuk mengetahui seberapa besar kesalahan observasi dalam penelitian ini.

Dari data hasil analisis dapat disimpulkan bahwa model Regresi M-kuantil menghasilkan nilai pendugaan yang lebih baik daripada model Poisson pada pendugaan area kecil atau *Small Area Estimation* (SAE). Hal ini dapat ditunjukkan pada nilai rata-rata MSE resampling bootstrap yang menghasilkan model Regresi M-kuantil lebih kecil dibandingkan model Poisson serta ragam pada model regresi M-kuantil lebih konstran dari pada model Poisson. Hal ini menunjukan bahwa model regresi M-kuantil lebih bersifat kuat (*robust*) terhadap adanya pencilan (*outlier*) daripada model Poisson.

PRAKATA

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat serta hidayahNya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Penanganan Overdispersi Menggunakan Regresi M-Kuantil Dengan Resampling Bootstrap Untuk *Small Area Estimation* ". Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan strata satu (S1) di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember. Sholawat serta salam semoga selalu tercurahkan keharibaan beliau nabi Muhammad SAW yang telah menjadi pembawa rahmatan lil'alamin.

Penyusunan skripsi ini tidak terlepas dari bantuan berbagai pihak, baik secara langsung maupun tidak langsung. Penulis menyampaikan terima kasih kepada :

1. Dian Anggraeni, S.Si, M.Si selaku Dosen Pembimbing Utama dan Dr. Alfian Futuhul Hadi, S.Si, M.Si selaku Dosen Pembimbing Anggota yang telah memberikan bimbingan secara intensif dan bantuan untuk penyempurnaan skripsi ini;
2. Prof. Drs. I Made Tirta, M.Sc., Ph.D dan M. Ziaul Arif, S.Si, M.Sc. Selaku Dosen Penguji yang telah memberikan kritik dan saran yang membangun untuk penyempurnaan skripsi ini;
3. Drs. Rusli Hidayat M.Sc selaku Dosen Pembimbing Akademik yang telah membimbing selama penulis menjadi mahasiswa;
4. Ayahanda Wakijo, Ibunda Suryati serta seluruh keluarga besar yang telah memberikan doa dan dorongan semangat demi terselesaikannya skripsi ini;
5. seluruh dosen dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam yang telah memberikan ilmu serta fasilitas yang membantu selama proses perkuliahan berlangsung;
6. sahabat Upik, Eka, Kiki, kikin dan seluruh keluarga besar KRAMAT '11 yang tidak bisa saya sebut satu per satu yang selalu senantiasa menemani, memberi

dukungan, semangat perjuangan, serta saran dalam proses menyelesaikan tugas akhir;

7. keluarga Besar Niswatu, terimakasih untuk suka duka yang telah dilalui bersama dan semangat yang telah diberikan;
8. serta semua pihak yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa dalam menyusun skripsi ini masih terdapat kekurangan baik isi maupun susunannya. Penulis mengharapkan saran dan kritik demi penyempurnaan skripsi ini. Akhirnya penulis berharap semoga skripsi ini dapat memberi manfaat dan sumbangan bagi pembaca.

DAFTAR ISI

Halaman

HALAMAN SKRIPSI i
HALAMAN PERSEMBAHAN ii
HALAMAN MOTTO iii
HALAMAN PERNYATAAN iv
HALAMAN PEMBIMBINGAN v
HALAMAN PENGESAHAN vi
RINGKASAN vii
PRAKATA viii
DAFTAR ISI x
DAFTAR TABEL xii
DAFTAR GAMBAR xiii
DAFTAR LAMPIRAN xiv
BAB 1. PENDAHULUAN 1
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Batasan Masalah 3
1.4 Tujuan Penelitian 4
1.5 Manfaat 4
BAB 2. TINJAUAN PUSTAKA 5
2.1 *Small Area Estimation* (SAE) 5
2.2 Model Area Kecil 6
2.2.1 Model Level Area 6
2.2.2 Model Level Unit 7
2.3 Model Regresi Poisson 7

2.4 Overdispersi 8
2.5 Regresi M-Kuantil 9
2.4.1 Regresi M-kuantil Untuk Respon Kontinu 9
2.4.1 Regresi M-kuantil Untuk Data Cacahan 10
2.6 Pendekatan *Quasi-Likelihood* 10
2.7 Estimasi Area Kecil Oleh Regresi M-Kuantil 12
2.6.1 Estimasi Titik 12
2.6.2 *Mean Squared Error* (MSE) 13
BAB 3. METODE PENELITIAN 16
3. 1 Data Penelitian 16
3. 2 Desain Penelitian 16
3. 3 Metode Analisi Data 17
BAB 4. HASIL DAN PEMBAHASAN 19
4.1 Deskripsi Data 19
4.2 Pendugaan Parameter 20
4.2.1 Hasil Estimasi Koefisien Regresi (β) 20
4.2.2 Hasil Estimasi $\hat{y}$ Pada Model Poisson Dan Regresi M-kuantil 22
4.3 Pendugaan Model Poisson dan Regresi M-kuantil Pada *Small Area Estimation* 22
4.4 Resampling Bootstrap 24
BAB 5. PENUTUP 27
5.1 Kesimpulan 27
5.2 Saran 27
DAFTAR PUSTAKA 28
LAMPIRAN 30

DAFTAR TABEL

Halaman

4.1 Hasil *Summary* Data 18

4.2 Nilai *Standard Error* Model Poisson Dan Regresi M-Kuantil 19

4.3 Hasil Estimasi koefisien Regresi Poisson dan M-muantil 19

4.4 Model Regresi Poisson dan M-kuantil 20

4.5 Hasil Estimasi $\hat{y}$ Pada Model Poisson dan Regresi M-kuanti 20

4.6 Hasil Estimasi model Poisson dan Regresi M-kuantil Pada SAE 21

4.7 Rata-Rata MSE Resampling Bootstrap 1000 kali 22

DAFTAR GAMBAR

Halaman

3.1 Langkah-langkah Penelitian 15

4.1 Boxplot MSE Bootstrap 1000 kali 23

4.2 Boxplot Ragam MSE Bootstrap 1000 kali 23

4.3 Ragam Bootstrap Sebanyak 1000 kali 24

DAFTAR LAMPIRAN

Halaman

A. Data Penelitian 29

B. *Script* Program Model Regresi Poisson dan M-kuantil....................................30

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara dengan jumlah penduduk tertinggi keempat di dunia setelah China, India dan Amerika Serikat (Badan Pusat Statistik, 2013). Dengan jumlah penduduk yang begitu besar yaitu 248,8 juta jiwa, maka diperlukan sebuah survei atau sensus untuk menyediakan data representatif mengenai jumlah penduduk, ekonomi, sosial, wilayah, lingkungan dan pendidikan. Sensus dan survei memiliki peranan penting untuk pengambilan keputusan yang berbasis pada sebuah data. Tujuan utama dari survei ialah untuk pengumpulan data di mana yang diselidiki adalah parameter suatu populasi. Data yang diperoleh dari hasil survei merupakan data perkiraan atau *estimate value* (Supranto, 2000). Secara luas survei tidak hanya digunakan untuk menduga parameter populasi tetapi untuk menduga keragaman subpopulasi atau domain. Domain didefinisikan sebagai daerah geografik, sosio-demografi dan sebagainya. Sampel yang diperlukan untuk survei atau sensus dilakukan dengan data skala nasional dan tidak semua area (kecil) menjadi sample seperti Kabupaten/kota bahkan mungkin level kecamatan dan desa/kelurahan. Oleh karena itu, dibutuhkan suatu metode pendugaan untuk menperoleh informasi pada level area (kecil) tersebut.

Pendugaan dalam survei yaitu pendugaan langsung dan pendugaan tidak langsung. Pendugaan langsung didasarkan pada teknik penarikan sampel misalnya *simple random sampling*, *cluster sampling* dan sebagainya. Metode ini dikenal dengan *design based*. Metode pendugaan langsung memiliki kelemahan jika sampel memiliki ukuran kecil dan juga nilai varians yang besar dikarenakan sampel tidak mewakili populasi. Permasalah yang timbul dalam pendugaan langsung terhadap area kecil adalah penduga yang dihasilkan tak bias dan memiliki ragam besar dikarenakan ukuran sampel kecil dan tidak semua area kecil terwakili dalam survei maka tidak memungkinkan dilakukannya pendugaan langsung (Kurnia, 2009). Area kecil didefinisikan sebagai subpopulasi yang memiliki ukuran sampel kecil sehingga pendugaan langsung tidak dapat menghasilkan pendugaan yang teliti. Solusi yang digunakan adalah dengan pendugaan tidak langsung dengan cara menambahkan variabel-variabel pendukung dalam penduga parameter. Metode tersebut mengasumsikan bahwa area kecil memiliki karakteristik yang sama dengan area besar. Pendugaan ini bersifat meminjam kekuatan dari pengamatan sampel area yang berdekatan dengan memanfaatkan informasi tambahan yaitu dari data sensus berskala nasional (Rao, 2003). Pendugaan tidak langsung ini dikenal sebagai pendugaan area kecil (*Small Area Estimation/*SAE). SAE adalah suatu teknik statistika pendugaan parameter-parameter area kecil secara tidak langsung dengan memanfaatkan data hasil survei domain besar seperti Survei Sosial Ekonomi Nasional (SUSENAS) dan survei Pendapatan Potensi Desa (PODES) untuk menduga peubah yang menjadi perhatian pada suatu area yang relatif kecil dalam persampelan survei (*survey sampling*). Proses pendugaan tidak langsung pada suatu domain yaitu dengan cara menghubungkan informasi pada area tersebut dengan area lain melalui model yang tepat.

Permintaan untuk statistika area kecil telah sangat meningkat di seluruh dunia dalam kurun waktu beberapa tahun terakhir, menyebabkan perkembangan sejumlah model berbasis metode pendugaan area kecil (SAE). Sebagai contoh, *Empirical Best Linear Unbiased Prediction* (EBLUP) berdasarkan pada *Linear Mixed Model* (LMM) beberapa kali direkomendasikan ketika kesimpulan sasaran adalah rata-rata area kecil dari variabel respon kontinu.

Sebagian besar data survei yang didapat berbentuk biner atau dalam bentuk cacahan, data tersebut tidak cocok untuk model dasar SAE berdasarkan LMM. Untuk mengatasi hal tersebut dalam SAE terdapat metode yang digunakan untuk data biner atau data cacahan yaitu Metode *Empirical Bayes* dan *Hierarchical Bayes*. Terdapat metode alternatif untuk pendugaan area kecil pada data cacahan yaitu mengunakan metode regresi M-kuantil.

Beberapa penelitian mengenai SAE telah dilalukan oleh beberapa orang diantaranya adalah Dewi (2013) yang meneliti tentang estimasi parameter model regresi M-kuantil mengunakan metode *Iterative Reweighted Least Square* (IRLS) pada variabel respon kontinu dan penerapan model regresi M-kuantil pada pengaruh pendapatan terhadap kebutuhan pangan rumah tangga. Penelitian lain juga dilakukan oleh Anggraini (2014) yaitu meneliti penduga momen untuk parameter dispersi pada pendekatan Bayes Empirik model campuran Poisson-Gamma dalam SAE. Chambers *et al*. (2012) juga telah melakukan penelitian tentang model regresi M-kuantil untuk data biner dalam pendugaan area kecil dan diterapkan pada estimasi jumlah penganguran berusia diatas 16 tahun pada Daerah Kabupaten Inggris. Pada tahun berikutnya Chambers *et al.* (2013) juga meneliti tentang model regresi Poisson M-kuantil untuk pendugaan area kecil dan diterapkan pada kunjugan dokter di tiga Daerah administratif Italia yaitu: Liguria, Toscana, dan Umbria.

Pendugaan M-kuantil dapat dilihat sebagai alternatif semi-parametrik outlier yang kuat (*robust*) untuk lebih umum digunakan *Empirical Best Predictor Plug-in* (EBPP) yang didasarkan pada Poisson-GLMM dengan pengaruh acak Gaussian. Regresi M-kuantil juga tidak bergantung pada asumsi distribusi yang kuat maupun pada struktur hirarki data dan juga memiliki sifat tegar (*robust*) terhadap pencilan (*outliers*), oleh karena peneliti tertarik mengunakan pendekatan alternatif tersebut dalam pendugaan area kecil dengan mengunakan data Anggraini (2014).

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas dapat dirumuskan masalah bagaimana model dugaan parameter yang baik pada pendugaan area kecil dengan mengunakan pendekatan model regresi Poisson M-kuantil.

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini menggunakan model level area pada *Small Area Estimation* (SAE). Data yang digunakan merupakan data cacahan.

1.4 Tujuan

Tujuan yang ingin dicapai pada penelitian ini berdasarkan rumusan masalah diatas adalah mendapatkan model dugaan parameter yang baik pada pendugaan area kecil dengan mengunakan pendekatan model regresi Poisson dan M-kuantil.

1.5 Manfaat

Manfaat yang dapat diambil dari penelitian ini adalah dapat menambah wawasan dan pengetahuan tentang analisis statistika yaitu *Small Area Estimation* mengenai pendugaan parameter mengunakan metode alternatif yaitu model Regresi M-kuantil.

BAB 2 TINJAUAN PUSTAKA

2. 1 *Small Area Estimation* (SAE)

Suatu area dikatakan kecil apabila sampel yang diambil pada area tersebut tidak mencukupi untuk melakukan suatu pendugaan langsung dengan hasil dugaan yang akurat (Rao, 2003). *Small area estimation* (SAE) merupakan suatu teknik statistika untuk menduga parameter-parameter pada area kecil dengan memanfaatkan informasi dari dalam area dan luar area. (Longford, 2005). Permasalahan yang dapat terjadi dalam SAE adalah bagaimana menghasilkan suatu dugaan parameter yang baik untuk ukuran sampel yang kecil pada suatu domain. Dan juga bagaimana menduga *mean square error* (MSE) dari dugaan parameter tersebut. Permasalahan tersebut dapat diatasi dengan cara meminjam informasi dari dalam area, luar area maupun dari luar survei.

Pendugaan parameter pada suatu domain dalam SAE dapat dilakukan dengan mengunakan pendugaan langsung (*direct estimation*) dan tidak langsung (*indirect estimation*). Pendugaan langsung merupakan pendugaan pada suatu domain berdasarkan data sampel dari domain tersebut dengan pendekatan yang digunakan adalah pendekatan berbasis rancangan (*design-based*). Sedangkan pendugaan tidak langsung merupakan pendugaan pada suatu domain dengan cara menghubungkan informasi tambahan pada area tersebut. Metode dengan memanfaatkan informasi tambahan tersebut secara statistik memiliki sifat “meminjam kekuatan” (*borrowing strength)* informasi dari hubungan antara peubah respon dengan informasi yang ditambahkan. Dengan demikian, pendugaan tidak langsung mencakup data dari domain yang lain. Prosedur SAE pada dasarnya memanfaatkan kekuatan informasi dari area sekitarnya (*neighbouring area*) dan sumber data di luar statistiknya yang diperoleh melalui pembentukan model yang tepat untuk meningkatkan efektifitas ukuran sampel (Kurnia, 2009).

Terdapat metode pendugaan pada SAE yang yelah dikembangkan khususnya menyangkut metode berbasil model (*model-based estimation*) sebagai alternatif pendugaan langsung. Metode tersebut adalah *Empirical Best Linier Unbiased Prediction* (EBLUP), *Empirical Bayes* (EB), dan *Hierarchical Bayes* (HB). Metode EB dan HB adalah metode yang lebih umum digunakan untuk menangani data kontinu, biner maupun cacahan (Kismiantini, 2007)

2. 2 Model Area Kecil

Pada SAE terdapat dua jenis model area kecil yaitu model level area dan model level unit.

2.2.1 Model Level Area

Model level area sangat penting jika data level unit tidak tersedia. Model tersebut didasarkan pada ketersediaan data pendukung yang hanya ada pada level area tertentu, misalkan $\boldsymbol{x}_i = \left(x_1, x_2, \cdots, x_{pi}\right)^{\boldsymbol{T}}$, dan parameter yang akan digunakan adalah θ_i yang diasumsikan mempunyai hubungan dengan data pendukung $\boldsymbol{x}_i$. Data pendukung tersebut digunakan untuk membangun model linier:

$$\theta_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + b_i v_i, \qquad \text{dengan } i = 1,2,3,\dots,n \tag{2.1}$$

dengan $\boldsymbol{\beta} = \left(\boldsymbol{\beta_1}, \dots, \boldsymbol{\beta_p}\right)^{\boldsymbol{T}}$ merupakan $p \times l$ vektor regresi dari parameter dan $v_i^{iid} \sim N(0, \sigma_v^2)$ sebagai pengaruh acak yang diasumsikan *independent and identically distributed (iid)* dan menyebar normal. Kesimpulan mengenai θ_i, dapat diketahui dengan mengasumsikan bahwa model penduga langsung y_i telah tersedia, yaitu:

$$y_i = \theta_i + e_i, \qquad \text{dengan } i = 1,2,3,\dots,n \tag{2.2}$$

dan *error sampling* $e_i^{iid} \sim N(0, \sigma_e^2)$ dengan σ_e^2 diketahui. Dengan menggabungkan model (2.1) dan (2.2) maka diperoleh model campuran:

$$y_i = x_i^T \boldsymbol{\beta} + b_i v_i + e_i, \qquad \text{dengan } i = 1,2,3,\dots,n \tag{2.3}$$

dengan b_i dan v_i saling bebas. Model (2.3) ini merupakan kasus khusus dari model campuran linier dan dikenal pula sebagai model Fay-Herriot dalam literatur area kecil (Rao, 2003).

2.2.2 Model Level Unit

Model level unit merupakan suatu model dimana data-data pendukung yang tersedia bersesuaian secara individu dengan data respon. Model level unit mengasumsikan bahwa ada penyerta unit $\boldsymbol{X}_{ij} = \left(x_{1ij}, x_{2ij}, \dots, x_{pij}\right)^{\boldsymbol{T}}$ ada untuk masing-masing anggota populasi j data dalam masing-masing area kecil i namun kadang cukup dengan rata-rata populasi x_i diketahui saja. Selanjutnya peubah perhatian y_{ij} dianggap berkaitan dengan x_{ij} sehingga didapatkan suatu model regresi tersarang:

$$y_{ij} = x_{ij}^T \beta + b_i v_i + e_{ij,} \quad \text{dengan } i = 1,2, \dots, m; j = 1,2, \dots, N;$$

dengan asumsi $v_i^{iid} \sim N(0, \sigma_v^2)$ dan $e_i^{iid} \sim N(0, \sigma_e^2)$ serta v_i saling bebas (Rao, 2003).

2. 3 Model Regresi Poisson

Model regresi merupakan suatu model yang menghubungan antara variabel respons dengan variabel-variabel prediktornya. Salah satu model regresi adalah model regresi Poisson. Model regresi Poisson merupakan model regresi yang mengambarkan hubungan antara variabel respon Y yang berupa data diskrit dengan variabel prediktor X. Variabel respon dalam regresi Poisson berasal dari data cacahan yang diharapkan jarang terjadi (Sundari, 2012). Pada regresi Poisson hubungan antara variabel respon dan variabel prediktor dapat dituliskan dalam bentuk sebagai berikut

$$E(Y_i|X_i) = \mu_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

atau

$$\mu_i = x_i^T \beta.$$

Menurut Myers (1990) fungsi penghubung yang digunakan pada model regresi Poisson adalah fungsi log karena menjamin nilai variabel yang diharapkan dari variabel respon yang akan bernilai non negatif. Berikut merupakan fungsi penghubung yang digunakan pada model regresi Poisson

$$\ln E(Y_i|X_i) = \ln(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik}$$

$$\mu_i = \exp\left(x_i^T \hat{\beta}\right) = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik})$$

sehingga model regresi Poisson dapat ditulis sebagai berikut

$$E(Y_i|X_i) = t_i \exp\left(x_i^T \hat{\beta}\right) + \varepsilon_i, \qquad i = 1, 2, \ldots, n$$

dengan μ_i merupakan rata-rat jumlah kejadian dalam periode t_i dan diasumsikan μ_i tidak berubah dari titik data ke titik data secara bebas maka μ_i dapat dimodelkan sebagai fungsi dari *k* variabel prediktor.

2.4 Overdispersi

Masalah yang sering terjadi pada regresi Poisson adalah adanya overdispersi. Menurut Hilbe (2011) overdispersi pada regresi Poisson terjadi ketika ragam dari respon lebih besar daripada meannya. Jika pada data cacah terjadi overdispersi namun tetap digunakan model Poisson maka estimasi parameter koefisien regresinya tetap konsisten tetapi tidak efisien karena berdampak pada nilai *standard error* yang menjadi *underestimate* sehingga kesimpulannya menjadi tidak valid. Overdispersi pada Regresi Poisson dapat dideteksi dengan nilai *Pearson Chi- Square* (χ^2) atau devians yang dibagi dengan derajat bebasnya. Jika kedua nilai ini lebih besar dari 1 maka dikatakan terjadi overdispersi pada data.

Overdispersi memiliki dampak yang sama dengan pelanggaran asumsi homokedastisitas dalam model regresi linier dan menyebabkan nilai devians model menjadi sangat besar sehingga model yang dihasilkan menjadi kurang tepat. Salah satu cara untuk mengatasi adanya kasus overdispersi dalam Regresi Poisson adalah dengan mengganti asumsi distribusi Poisson dengan distribusi lain yang lebih fleksibel.

2.5 Regresi M-Kuantil

Regresi kuantil merupakan suatu metode pendekatan dalam analisis regresi yang menduga berbagai fungsi kuantil dari suatu distribusi Y sebagai fungsi dari X (Koenker dan Bassett, 1978). Metode regresi kuantil memiliki kelebihan yaitu dapat mengetahui perlakuan antara variabel bebas dan vatiabel tak bebas pada bagian awal, tengah, atau akhir dari data. Regresi M-kuantil merupakan bagian dari Regresi Kuantil, dimana perlakuan antara variabel bebas dan variabel tak bebas dapat dilihat pada setiap kuantil. Metode ini dapat dibentuk beberapa model untuk beberapa kuantil yang diinginkan. Sehingga diperoleh model terbaik dari beberapa model yang telah didapatkan (Dewi, 2013).

2.5.1 Regresi M-kuantil untuk Respon Kontinu

Menurut Chambers dan Tzavidis (2013) model regresi klasik meringkas perilaku rataan dari variabel acak Y pada setiap titik dalam himpunan kovariat X. Ini memberikan gambaran tidak lengkap dari suatu distribusi. Regresi Kuantil meringkas perilaku dari bagian yang berbeda dari distribusi bersyarat dari y pada setiap titik di x. Model regresi linier untuk kuantil dengan kuantil ke-q ditulis sebagai berikut

$$Q_y\left(q|x_j\right) = \boldsymbol{x}_j^T \boldsymbol{\beta}_q \tag{2.4}$$

dengan q bernilai 0 < q < 1. Estimasi dari regresi ke-q parameter β_q diperoleh dengan meminimalkan

$$\sum_{j=1}^{n} \left| y_j - \boldsymbol{x}_j^T \boldsymbol{\beta}_q \right| \left\{ (1-q) I\left(y_j - \boldsymbol{x}_j^T \boldsymbol{\beta}_q \le 0 \right) + q I\left(y_j - \boldsymbol{x}_j^T \boldsymbol{\beta}_q > 0 \right) \right\}$$

M-kuantil order q untuk kepadatan bersyarat y diberikan himpunan kovariat x, $f(y|x)$ didefinisikan sebagai solusi $MQ_y(q|x;\psi)$ dari persamaan pendugaan $\int \psi_q\{y - MQ_y(q|x;\psi)\} f(y|x)\, dy - 0$, dimana ψ_q menunjukkan suatu fungsi pengaruh asimetris, yang merupakan turunan dari fungsi kerugian asimetris ρ_q. Menurut Chambers dan Tzavidis (2006) jika M-kuantil digunakan untuk mengestimasi daerah kecil dengan unit ke-j, maka persamaan (2.4) menjadi

$$MQ_y(q|\boldsymbol{x}_j;\psi) = \boldsymbol{x}_j^T\boldsymbol{\beta}_q \tag{2.5}$$

2.5.2 Regresi M-kuantil untuk data cacahan

Pendekatan yang sering digunakan dalam pemodelan dengan hasil rataan diskrit sebagai fungsi prediktor yaitu dengan mengunakan GLM dengan mengasumsikan bahwa variabel respon mengikuti distribusi yang merupakan anggota dari keluarga distribusi eksponensial dan mengunakan fungsi link yang sesuai. Untuk data cacahan, distribusi yang tepat adalah Poisson dan fungsi link logaritma.

Pada data cacahan, ditentukan spesifikasi kontinu yang tepat (dalam q) pada $MQ_y(q|\boldsymbol{x}_j;\psi)$. Spesifikasi yang jelas untuk data cacahan adalah spesifikasi log-linier yaitu dengan menganti (2.5) oleh

$$MQ_y(q|\boldsymbol{x}_j;\psi) = t_j \exp(\boldsymbol{x}_j^T\boldsymbol{\beta}_q) \tag{2.6}$$

dengan t_j adalah syarat offset. Untuk estimasi β_q mengunakan persamaan estimasi yang didefinisikan sebagai berikut

$$\Psi(\beta) \coloneqq n^{-1}\sum_{j=1}^{n}\left\{\psi(r_j)w(\boldsymbol{x}_j)\frac{1}{\sigma(\mu_j)}\mu'_j - a(\beta)\right\} = 0 \tag{2.7}$$

dimana r_j adalah residu Pearson, $E[Y_j] = \mu_j$, turunan terhadap β adalah μ'_i, $Var[Y_j] = \sigma^2(\mu_j)$ dan $a(\beta) = n^{-1}\sum_{j=1}^{n} E[\psi(r_j)]w(\boldsymbol{x}_j)\frac{1}{\sigma(\mu_j)}\mu'_j$ untuk memastikan konsistensi Fisher dari estimator (Chambers dan Tzavidis, 2013).

2.6 Pendekatan *Quasi-Likelihood*

Pendekatan *Quasi-Likelihood* biasanya digunakan pada suatu data non-Gaussian untuk memodelkan rata-rata dan dispersinya (Pawitan *et. al*, 2006). Misalnya dalam menganalisis data cacahan melalui regresi Poisson biasa yang mengasumsikan varians sama dengan rata-ratanya. Namun, seringkali terdapat keragaman ekstra Poisson yaitu nilai variansinya lebih besar dari nilai meannya atau yang lebih dikenal dengan istilah overdispersi.

Pada persamaan (2.7) fungsi ψ dibatasi untuk mengontrol deviasi di y-ruang, mengingat pembobot $w(.)$ digunakan untuk *downweight* pengaruh titik. Menurut Cantoni dan Ronchetti (2001) ketika $w(\boldsymbol{x}_j) = 1$ digunakan pendekatan Huber *Quasi-Likelihood Estimator*. Perhatikan bahwa ketika ψ adalah fungsi identitas, maka dapat diperoleh estimator klasik Quasi-Likelohood untuk GLM. Untuk regresi M-kuantil persamaan estimasi (2.7) dapat ditulis kembali sebagai berikut

$$\Psi(\beta_q) \coloneqq \frac{1}{n}\sum_{j=1}^{n}\left\{\psi_q(r_{jq})w(\boldsymbol{x}_j)\frac{1}{\sigma\left(MQ_y(q|\boldsymbol{x}_j;\psi)\right)}MQ'_y(q|\boldsymbol{x}_j;\psi) - a(\beta_q)\right\} = 0 \qquad (2.8)$$

dengan

$$r_{jq} = \sigma\left(MQ_y(q|\boldsymbol{x}_j;\psi)\right)^{-1}\left(y_j - MQ_y(q|\boldsymbol{x}_j;\psi)\right),$$

$$\sigma\left(MQ_y(q|\boldsymbol{x}_j;\psi)\right) = MQ_y(q|\boldsymbol{x}_j;\psi)^{1/2},$$

$$MQ'_y(q|\boldsymbol{x}_j;\psi) = MQ_y(q|\boldsymbol{x}_j;\psi)\boldsymbol{x}_j$$

dan $a(\beta_q)$ Adalah syarat koreksi untuk memperoleh penduga tak bias, yang didefinisikan sebagai berikut

$$a(\beta_q) = n^{-1}\sum_{j=1}^{n} 2w_q(r_{jq})w(\boldsymbol{x}_j)\left\{cP(Y_j \geq i_2 + 1) - cP(Y_j \leq i_1) + \frac{MQ_y(q|\boldsymbol{x}_j;\psi)}{\sigma\left(MQ_y(q|\boldsymbol{x}_j;\psi)\right)}\left[P(Y_j = i_1) - P(Y_j = i_2)\right]\right\}MQ_y(q|\boldsymbol{x}_j;\psi)^{1/2}\boldsymbol{x}_j \ ,$$

dengan

- $i_1 = \left\lfloor MQ_y(q|\boldsymbol{x}_j;\psi) - c\sigma\left(MQ_y(q|\boldsymbol{x}_j;\psi)\right)\right\rfloor$,
- $i_2 = \left\lfloor MQ_y(q|\boldsymbol{x}_j;\psi) + c\sigma\left(MQ_y(q|\boldsymbol{x}_j;\psi)\right)\right\rfloor, dan$

- $w_q(r_{jq}) = [qI(r_{jq} > 0) + (1 - q)I(r_{jq} \leq 0)]$

ketika $w(\boldsymbol{x}_j) = 1, j = 1, \ldots, n$ estimator Huber Quasi-Likelihood diperoleh. (Chamber dan Tzavidis, 2013).

2.7 Estimasi Area Kecil Oleh Model Regresi M-Kuantil

2.7.1 Estimasi Titik

Konsep kunci penerapan metode M-kuantil pada data dengan struktur grup yaitu dengan mengidentifikasi keunikan 'koefisien M-kuantil' yang terkait dengan setiap data yang diamati. Koefisien tersebut kemudian dirata-rata, dalam beberapa cara yang cocok. Pengamatan yang lebih membuat grup untuk mendefinisikan level grup koefisien M-kuantil, yang dapat digunakan untuk mengkarakterisasi distribusi dari $y|x$ dalam grup.

Pada kasus y kontinu, koefisien M-kuantil untuk pengamatan j hanya didefinisikan sebagai solusi unik q_j untuk persamaaan $y_j = \widehat{MQ}_y(q_j|\boldsymbol{x}_j;\psi)$. Namun, untuk data cacahan persamaan $y_j = \widehat{MQ}_y(q_j|\boldsymbol{x}_j;\psi)$ tidak memiliki solusi ketika $y_j = 0$. Menurut Chambers *et al* (2013) untuk mengatasi masalah tersebut dengan mengunakan definisi sebagai berikut

$$\widehat{MQ}_y(q_j|\boldsymbol{x}_j;\psi) = \begin{cases} k(\boldsymbol{x}_j) & y_j = 0 \\ y_j & y_j = 1, 2, \ldots \end{cases}$$

Kemungkinanya yaitu $k(\boldsymbol{x}_j) = \widehat{MQ}_y(q_{min}|\boldsymbol{x}_j;\psi)$ dimana q_{min} menunjukkan q-value terkecil dari q-value yang digunakan untuk menentukan nilai q_j dari unit yang diamati. Namun, secara tidak langsung menyatakan bahwa $q_j = q_{min}$ setiap kali $y_j = 0$ terlepas dari nilai $\boldsymbol{x}_j$. Untuk mengatasi masalah tersebut yaitu dengan mengunakan definisi q_j untuk kasus Bernoulli. Ini berarti bahwa pengamatan dengan nilai $y_1 = 0$ sesuai dengan q-value terkecil dari pada yang lain dengan $y_2 = 0$ ketika $\widehat{MQ}_{y1}(0.5|\boldsymbol{x}_1;\psi) > \widehat{MQ}_{y2}(0.5|\boldsymbol{x}_2;\psi)$. Untuk mendefinisikan yaitu dengan cara mengatur

$$k(\boldsymbol{x}_j) = min\left\{1-\epsilon, \left[\widehat{MQ}_y(0.5|\boldsymbol{x}_j;\psi)\right]^{-1}\right\}$$

di mana $\epsilon > 0$ adalah konstanta positif kecil. Kemudian koefisien M-kuantil untuk unit j adalah q_j, dimana

$$\widehat{MQ}_y(q_j|\boldsymbol{x}_j;\psi) = \begin{cases} min\left\{1-\epsilon, \frac{1}{exp\left(x_j^T\beta_{0.5}\right)}\right\} & y_j = 0 \\ y_j & y_j = 1,2,.. \end{cases}$$

Koefisien M-kuantil di area khusus *d* yaitu $\hat{\theta}_d$ dapat didefinisikan sebagai nilai rata-rata dari koefisien sampel M-kuantil di area *d* dengan syarat terdapat sampel pengamatan di daerah kecil *d*. Berikut merupakan prediktor M-kuantil dari rata-rata jumlah $\bar{y}_d$ di area kecil *d*

$$\hat{\bar{y}}_d^{MQ} = N_d^{-1}\{\textstyle\sum_{j\in s_d} y_{dj} + \sum_{j\in r_d} \widehat{MQ}_y(\hat{\theta}_d|\boldsymbol{x}_{dj};\psi)\}$$

dimana $\widehat{MQ}_y(\hat{\theta}_d|\boldsymbol{x}_{dj};\psi) = exp\left(\boldsymbol{x}_j^T\widehat{\boldsymbol{\beta}}_{\hat{\theta}_d}\right)$ (Chambers & Tzavidis, 2006).

2.7.2 *Mean Squared Error* (MSE)

Mean Squared error (MSE) adalah metode untuk mengevaluasi metode peramalan dengan masing-masing kesalahan atau sisa dikuadratkan. Kemudian dijumlahkan dan dibagi dengan jumlah observasi. Pendekatan ini mengatur kesalahan peramalan yang besar karena kesalahan-kesalahan itu dikuadratkan. Suatu teknik yang menghasilkan kesalahan moderat mungkin lebih baik untuk salah satu yang memiliki kesalahan kecil tapi kadang-kadang menghasilkan sesuatu yang sangat besar.

MSE dari pediktor $\hat{\bar{y}}_d^{MQ}$ didefinisikan sebagai berikut

$$MSE\left(\hat{\bar{y}}_d^{MQ}\right) = E\left[\left(\hat{\bar{y}}_d^{MQ} - \bar{y}_d\right)^2\right].$$

Chambers *et al*. (2012) mengusulkan bootstrap non parametrik berbasis estimator MSE dari $\hat{\bar{y}}_d^{MQ}$ dengan membangun populasi buatan terbatas yang menyerupai populasi yang sebenarnya. Metode bootstrap pada dasarnya melakukan pengambilan sampel (*resampling*) dengan pengambilan dari sampel hasil observasi dengan replikasi B kali. Bootstrap memberikan fleksibilitas dalam hubungan jumlah replika B, khususnya pada survei dengan jumlah besar pada langkah pertama kelompok sampel. Untuk menhasilkan proses bootstrap dengan menulis prediktor linier dari model regresi Poisson M-kuantil dalam bentuk meniru bentuk model pengaruh campuran,

$$y_{dj} = \boldsymbol{x}_{dj}^{T}\beta_{0.5} + \boldsymbol{x}_{dj}^{T}\left(\beta_{\theta_d} - \beta_{0.5}\right).$$

Langkah-langkah dari prosedur Bootstrap sebagai berikut:

a. (Langkah 1) Menggunakan sampel s, fit model pada persamaan (2.6) dan mendapatkan prediktor $\hat{\bar{y}}_{d}^{MQ}$. Untuk menghitung *pseudo-random effect* setiap area kecil $\hat{u}_{d}^{MQ}$ dengan cara menghitung $E\left(\boldsymbol{x}_{dj}^{T}\left(\beta_{\theta_d} - \beta_{0.5}\right)\right)$ untuk setiap area. Hal ini sesuai untuk *re-scale* elemen-elemen $\hat{\boldsymbol{u}}^{MQ}$ sehingga mean sampel sama dengan nol.

b. (Langkah 2) Membangun vektor $\hat{\boldsymbol{u}}^{MQ*} = \left\{\hat{\boldsymbol{u}}_{1}^{MQ*}, \dots, \hat{\boldsymbol{u}}_{D}^{MQ*}\right\}^{T}$, yang elemen-elemennya diperoleh dengan mengekstraksi sampel acak sederhana dengan pengantian ukuran D dari himpunan $\left\{\hat{u}_{1}^{MQ}, \dots, \hat{u}_{D}^{MQ}\right\}^{T}$.

c. (Langkah 3) Menghasilkan populasi bootstrap $U*$ dari ukuran $N = \sum_{d=1}^{D} N_d$ dengan membangkitkan nilai dari distribusi Poisson dengan

$$\mu_{dj}^{*} = exp\left\{\boldsymbol{x}_{dj}^{T}\hat{\boldsymbol{\beta}}_{0.5} + \hat{\boldsymbol{u}}_{d}^{MQ*}\right\}, j = 1, \dots, N_d$$

dan menghitung parameter-parameter populasi bootstrap $\bar{y}_{dj}^{*}, d = 1, \dots, D.$

d. (Langkah 4) Ekstrak sampel $s*$ ukuran n dari populasi bootstrap $U*$ mengunakan asumsi *sampling design* dan menghitung SAE dengan bootstrap sampel $\hat{\bar{y}}_{d}^{MQ*}, d = 1, \dots, D.$

e. (Langkah 5) Ulangi langkah 2 – 4 B kali.

f. (Langkah 6) Penandaan oleh prediktor M-kuantil $\hat{\bar{y}}_d^{MQ*(b)}$ dalam replikasi bootstrap ke-*b* dan dengan $\bar{y}_{dj}^{*(b)}$ nilai populasi yang berhubungan dalam populasi bootstrap ke-*b*, didapatkan estimator bootstrap dari MSE adalah

$$MSE\left(\hat{\bar{y}}_d^{MQ}\right) = B^{-1}\sum_{b=1}^{B}\left(\hat{\bar{y}}_d^{MQ*(b)} - \bar{y}_{dj}^{*(b)}\right)^2$$

(Chambers dan Tzavidis, 2013).

BAB 3 METODE PENELITIAN

3.1 Data Penelitian

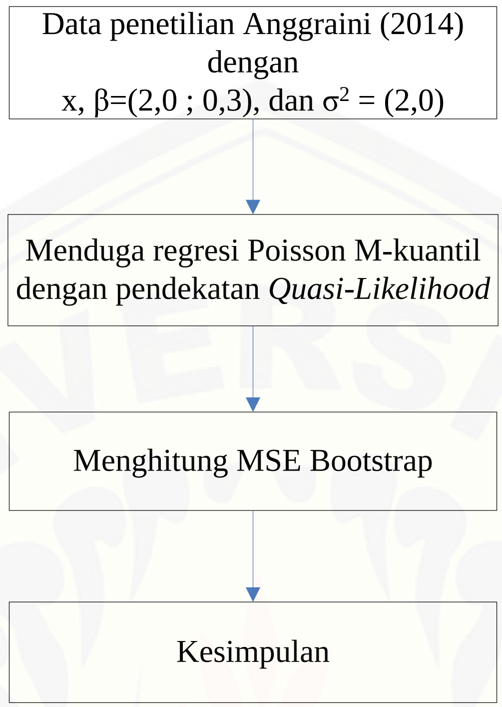
Data yang digunakan dalam penelitian ini merupakan data penelitian Anggraini (2014). Area kecil yang digunakan dalam data tersebut yaitu area 1 sampai area 20. Dengan menetapkan nilai x, $\beta = (2{,}0\,;0{,}3)$, dan $\sigma^2 = (2{,}0)$. Variabel yang digunakan adalah variabel bebas (X) dan variabel tak bebas (Y).

3.2 Desain Penelitian

Langkah-langkah dalam penelitian ini dapat dijelaskan sebagai berikut:

a. Menggunakan data penelitian Anggraini (2014) dengan $\beta = (2{,}0\,;0{,}3)$ yang merupakan vektor regresi berukuran $p \times 1$, dan ragam $\sigma^2 = (2{,}0)$. Dimana $\mu_i = x'\boldsymbol{\beta}$.
b. Melakukan pendugaan parameter untuk model regresi Poisson dan M-kuantil.
c. Menduga model regresi Poisson dan M-kuantil untuk SAE.
d. Menghitung MSE dengan metode resampling bootstrap dengan resampling sebanyak B=1000.

Dari uraian di atas, langkah-langkah yang akan dilakukan dalam penelitian ini dapat dilihat pada Gambar 3.1



Gambar 3.1 Langkah-langkah penelitian

3.3 Metode Analisis Data

Pengolahan data dalam penelitian ini mengunakan *software* Rstudio. Cara pengunaan Rstudio mirip dengan program R. Paket yang digunakan adalah `quantreg`. Fitting model mengunakan regresi sehingga fungsi R nya sebagai berikut:

```
rq.fit(x, y, tau=0.5, alpha=0.1, ci=FALSE, iid=TRUE,
        interp=TRUE, tcrit=TRUE)
```

dengan keterangan fungsi sebagai berikut:

a. `x` merupakan desain matrik.

b. `Y` merupakan variabel respon.

c. `Tau` merupakan kuantil yang di inginkan, jika `tau` diluar garis (0,1) seluruh proses diperkirakan.

d. `Alpha` merupakan kemungkinan nominal *noncoverage* untuk selang kepercayaan yaitu 1-alpha adalah kemungkinan cakupan nominal dari selang.
e. `Ci` merupakan fungsi logika jika T maka meghitung selang kepercayaan untuk parameter dengan mengunakan metode rank inversi Koenker (1994). Jika F maka kembali hanya untuk estimasi koefisien.
f. `Iid` merupakan fungsi logika jika T maka rank inversi didasarkan pada asumsi kesalahan `iid` model , jika F maka didasarkan pada asumsi kesalahan `iid`.
g. `Interp` digunakan untuk mempertimbangkan interval yang interpolasi antara nilai-nilai parameter tepat dibawah *cutoff* dan nilai diatas *cutoff* ditetapkan.
h. `Tcrit` merupakan fungsi logika jika T- nilai kritis t yang digunakan jika F maka nilai normal yang digunakan.

BAB 4 HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas estimasi parameter model Poisson dan model Regresi M-kuantil pada permasalahan pendugaan area kecil atau *Small Area Estimation* (SAE) dengan data yang digunakan merupakan data cacahan dan area kecil yang digunakan adalah area 1 sampai dengan area 20.

4.1 Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data Anggraini (2014). Data tersebut merupakan data simulasi yang overdispersi yaitu ragam lebih besar dari pada rata-rata. Nilai rata-rata dari variabel *dependent* 1,925 dan nilai ragamnya 3,923648 sehingga overdispersi karena ragam lebih besar dari rata-rata. Selain dengan nilai rata-rata dan ragam, overdispersi dapat dideteksi dengan nilai devians yang dibagi dengan derajat bebasnya. Jika kedua nilai tersebut lebih dari 1 maka dikatakan terjadi overdispersi pada data. Tabel 4.1 menunjukkan bahwa pada model regresi Poisson terjadi overdispersi karena nilai dispersinya lebih dari satu.

Tabel 4.1 Hasil uji overdispersi

Devians	db	Dispersi
28,61784	19	1,50621

Berikut merupakan hasil *summary* data dapat ditunjukkan pada Tabel 4.2 sebagai berikut.

Tabel 4.2 Hasil *summary* data

	x	y_D
Minimal	0,00	0,00
1st Quartile	1,00	0,00
median	3,00	1,00
mean	2,85	1,35
3rd Quanrtile	4,25	2,00
maksimal	6,00	5,00

4.2 Pendugaan Parameter

Pendugaan parameter dilakukan dengan mengunakan *Fitting* model yaitu model Poisson dan model M-kuantil. Pada model Poisson mengunakan fungsi `glm()` dan pada model Regresi M-kuantil mengunakan fungsi `rq.fit()` dengan paket yang digunakan adalah MASS dan `quantreg` dalam *software* Rstudio.

4.2.1 Hasil Estimasi Koefisien Regresi (β)

a. Nilai estimasi pada model Poisson dan model Regresi M-kuantil dapat ditunjukan pada Tabel 4.3 sebagi berikut.

Tabel 4.3 Hasil estimasi koefisien regresi Poisson dan M-kuantil

	koefisien regresi	Nilai estimasi	*p-value*
Poisson	β_0	0,8569	0,00313
	β_1	-0,2287	0,03180
Regresi M-kuantil	β_0	2,00000	0,03875
	β_1	-0,33333	0,21709

Pada regresi Poisson pendugaan parameter yang *underestimate* dapat menyebabkan pemodelan data yang tidak tepat. Hal tersebut dapat dilihat dalam Tabel 4.3 dimana pada regresi Poisson nilai *p-value* yang dihasilkan signifikan karena nilai dugaan *standard error* yang *uderestimate*. Hal ini dapat mengakibatkan keraguan terhadap dugaan pada regresi Poisson yang seharusnya tidak signifikan menjadi signifikan.

Berdasarkan Tabel 4.3 diatas, berikut merupakan persamaan model untuk Poisson dan Regresi M-kuantil dengan mengunakan fungsi `glm()` dan `rq.fit()` ditunjukkan Tabel 4.4.

Tabel 4.4 Model Poisson dan regresi M-kuantil

	Model
Poisson	$y = 0{,}8569 - 0{,}2287\ x_1$
Regresi M-kuantil	$y = 2{,}0000 - 0{,}3333\ x_1$

b. Nilai *standard error* pada model Poisson dan Regresi M-kuantil ditunjukkan dalam Tabel 4.5 sebagai berikut.

Tabel 4.5 Nilai *standard error* model Poisson dan regresi M-kuantil

	Variabel	*Standard error*
Poisson	*intercept*	0,2900
	x_1	0,1065
Regresi M-kuantil	*intercept*	0,89705
	x_1	0,26059

Pada data Anggraini (2014) menunjukkan adanya overdispersi sehingga nilai dugaan *standard error* dari model regresi Poisson menjadi *underestimate* seperti pada Tabel 4.5 sedangkan nilai *standard error* model regresi M-kuantil memberikan nilai yang lebih besar dikarenakan penyesuaian terhadap adanya overdispersi dalam data. Sehingga dapat dinyatakan bahwa nilai *standard error* dari kedua model tersebut merupakan nilai yang sebenarnya. Akibat dari nilai dugaan *standard error* pada model regresi Poisson yang *underestimate* tersebut, jumlah peubah yang berpengaruh nyata lebih besar dari model regresi M-kuantil dan tidak menunjukkan pengaruh yang sebenarnya. Semakin minimum *standard error* yang dihasilkan pada pendugaan yang mengalami overdispersi menyebabkan pendugaan parameter yang diperoleh tidak efisien. Pendugaan *standard error* yang terlalu rendah dapat memberikan kesimpulan yang salah untuk signifikansi parameter dalam suatu model.

4.2.2 Hasil Estimasi $\hat{y}$ Pada Model Poisson Dan Regresi M-kuantil

Tabel 4.6 Hasil estimasi $\hat{y}$ pada model Poisson dan regresi M-kuantil

Estimasi model Poisson ($\hat{y}_{pois}$)	Estimasi model Regersi M-kuantil ($\hat{y}_{MQ}$)
2,3557605	2
2,3557605	2
2,3557605	2
1,8741805	1,6666667
1,8741805	1,6666667
1,8741805	1,6666667
1,4910482	1,3333333
1,4910482	1,3333333
1,4910482	1,3333333
1,1862383	1,0000000
1,1862383	1,0000000
1,1862383	1,0000000
0,9437397	0,6666667
0,9437397	0,6666667
0,9437397	0,6666667
0,7508143	0,3333333
0,7508143	0,3333333
0,7508143	0,3333333
0,5973279	0
0,5973279	0

Setelah menduga parameter koefisien regresi (β), selanjutnya dilakukan estimasi $\hat{y}$ pada model Poisson dan M-kuantil. Tabel 4.6 merupakan Hasil estimasi $\hat{y}$ pada masing-masing model yaitu model Poisson dan Regresi M-kuantil.

4.3 Pendugaan Model Poisson dan regresi M-kuantil Pada *Small Area Estimation* (SAE)

Setelah diperoleh nilai pendugaan bagi model Poisson dan regresi M-kuantil pada SAE maka akan dibandingkan dengan nilai pendugaan langsung yang telah didapat pada penelitian Anggraini (2014). Hasil estimasi pendugaan tersebut ditunjukkan dalam Tabel 4.7 sebagai berikut.

Tabel 4.7 Hasil estimasi model Poisson dan regresi M-kuantil pada SAE

Estimasi Langsung	Estimasi model Poisson	Estimasi model Regresi M-kuantil
3	0,85686360	0,51945280
2	0,85686360	0,46945280
1	0,85686360	0,41945280
1	0,62817148	0,31472538
5	0,62817148	0,51472538
2	0,62817148	0,36472538
0	0,39947936	0,18968466
0	0,39947936	0,18968466
2	0,39947936	0,28968466
3	0,17078724	0,28591545
0	0,17078724	0,13591545
2	0,17078724	0,23591545
0	-0,05790488	0,09738800
0	-0,05790488	0,09738800
3	-0,05790488	0,24738800
0	-0,28659699	0,06978178
0	-0,28659699	0,06978178
2	-0,28659699	0,16978178
0	-0,51528911	0,05000100
1	-0,51528911	0,10000100

Tabel 4.7 diatas merupakan perbandingan hasil pendugaan langsung, pendugaan Poisson dan pendugaan Regresi M-kuantil. Rata-rata selisih pendugaan langsung dengan pendugaan Poisson adalah 1,1449089395 sedangkan untuk pendugaan regresi M-kuantil adalah 1,1084576900. Hal tersebut menunjukan bahwa model Regresi M-kuantil memberikan nilai yang lebih baik daripada model Poisson pada pendugaan area kecil atau *Small Area Estimation* (SAE).

4.4 Resampling Bootstrap

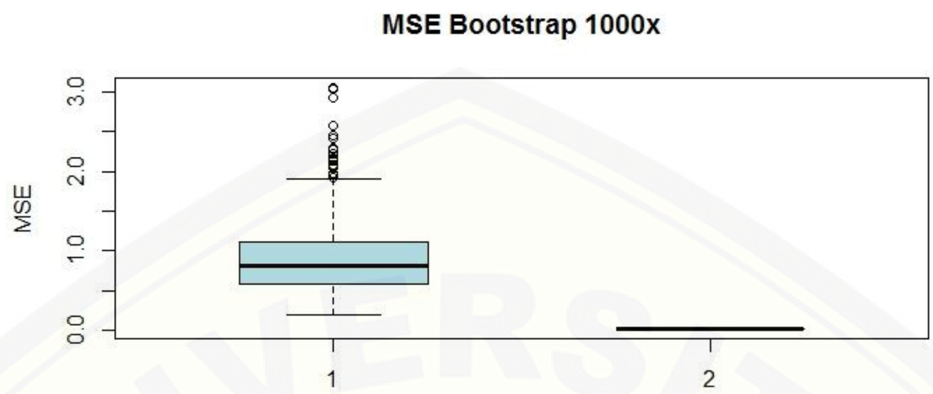
Setelah dilakukan estimasi pasa SAE tahap selanjutnya yaitu melakukan resampling Bootstrap untuk menghitung nilai *Mean Square Error* (MSE). Metode bootstrap dilakukan resampling data awal sebanyak 1000 kali dengan pengambilan sampel secara acak dengan mengunakan program Rstudio dan diperoleh data hasil sebanyak 1000 dengan jumlah data awal sama dengan jumlah data hasil resampling bootstrap. Tahap selanjutnya yaitu memodelkan data dengan fungsi model Poisson `glm()` dan fungsi model Regresi M-kuantil `rq.fit()` dengan mengambil pasangan data dari variabel bebas (X) dan variabel tak bebas (Y) . Resampling bootstrap dilakukan pada nilai estimasi kedua model tersebut dan didapatkan 1000 data hasil estimasi bootstrap.

Tahap terakhir yaitu menghitung nilai *Mean Square Error* (MSE) hasil estimasi bootstrap kedua model tersebut dengan diambil nilai rata-ratanya. Rata-rata MSE ditunjukkan pada Tabel 4.8 sebagai berikut ini.

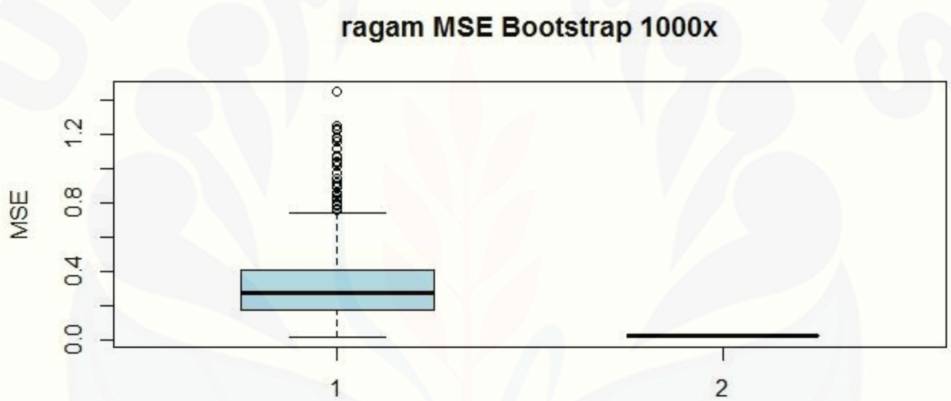
Tabel 4.8 Rata-rata mse resampling bootstrap 1000 kali

Model	Rata-Rata MSE
Poisson	0,8962059
Regresi M-kuantil	0,02497897

Dari Tabel 4.8 dapat diketahui bahwa perbandingan nilai rata-rata MSE resampling bootstrap sebanyak 1000 kali, rata-rata MSE dari model Regresi M-kuantil lebih kecil dibandingkan dengan model Poisson sehingga model terbaik adalah model Regresi M-kuantil. Resampling bootstrap untuk rata-rata MSE dilakukan sebanyak 1000 kali guna mendapatkan ketepatan nilai rata-rata MSE tersebut, sehingga semakin besar resampling bootstrap tersebut maka semakin besar pula ketepatannya. Untuk lebih jelasnya, MSE bootstrap 1000 kali ditunjukkan dalam bentuk boxplot pada Gambar 4.1 dan ragam MSE pada Gambar 4.2 berikut.



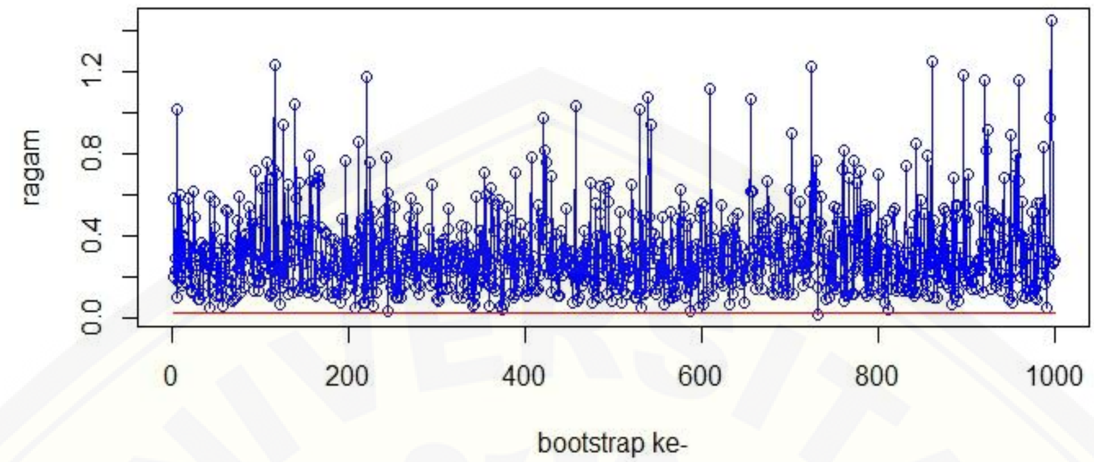
Gambar 4.1 Boxplot mse bootstrap 1000 kali



Gambar 4.2 Boxplot ragam mse bootstrap 1000 kali

Keterangan:

1. Model Poisson
2. Model Regresi M-kuantil



Gambar 4.3 Ragam bootstrap sebanyak 1000 kali

Keterangan:

: Model Poisson

: Model Regresi M-kuantil

Grafik ragam bootstrap sebanyak 1000 kali pada Gambar 4.3 menunjukkan bahwa model Regresi M-kuantil memiliki ragam yang konstan yang ditunjukkan oleh garis merah sedangkan ragam pada model Poisson lebih fluktuatif. Rata-rata ragam model regresi Poisson adalah 0,3161675 sedangkan rata-rata ragam model regresi M-kuantil 0,0231217. Hal tersebut membuktikan bahwa model regresi M-kuantil lebih bersifat kuat (*robust*) terhadap adanya pencilan (*outlier*) daripada model Poisson.

BAB 5 PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan pada bab 4 dapat diambil kesimpulan bahwa model Regresi M-kuantil menghasilkan nilai pendugaan yang lebih baik daripada model Poisson pada pendugaan area kecil atau *Small Area Estimation* (SAE). Hal ini dapat ditunjukan pada nilai rata-rata MSE resampling bootstrap yang menunjukkan model Regresi M-kuantil lebih kecil dibandingkan model Poisson. Dan juga model Regresi M-kuantil lebih bersifat kuat (*robust*) terhadap adanya pencilan (*outlier*) daripada model Poisson.

5.2 Saran

Saran untuk penelitian selanjutnya dapat dilakukan perbandingan regresi Binomial Negatif dan regresi M-kuantil untuk Pendugaan Area kecil.

DAFTAR PUSTAKA

Anggraini, A. 2014. "Penduga Momen untuk Parameter Dispersi pada Pendekatan Bayes Empirik Model Campuran Poisson-Gamma dalam Konteks *Small Area Estimation*". Tidak Diterbitkan. Skripsi. Jember: Jurusan Matematika FMIPA Universitas Jember.

Badan Pusat Statistik. 2013. Perkiraan Penduduk Beberapa Negara (Juta), 2000-2013. [on line]. http://www.bps.go.id/linkTabelStatis/view/id/1284. [17 Maret 2015].

Cantoni, E. and Ronchetti, E. 2001. Robust Inference For Generalized Linear Models. *Journal Of American Statistical Association* **96**, 1022-1030.

Chambers, R. and Tzavidis, N. 2006. M-Quantile Models For Small Area Estimation. *Biometrika*. **93**, 255-268.

Chambers, R., Salvati, N. & Tzavidis, N. 2012. M-Quantile Regression For Binary Data With Application To Small Area Estimation. *Biometrika*. 1**2**, 1-24.

Chambers, R., Dreassi, E. & Salvati, N. 2013. Disease Mapping Via Negative Binomial M-Quantile Regression. *Biostatistics.* **13**, 1-22.

Chambers, R., Dreassi, E., Rannalli, G. M., Salvati, N. & Tzavidis, N. 2013. Poisson M-Quantile Regression Small Area Estimation. *Biometrika*. **14**, 1-28.

Dewi, L. A. 2013. "Estimasi Parameter Model Regresi M-kuantil Menggunakan Metode *Iterative Reweighted Least Square*". Tidak Diterbitkan. Skripsi. Surakarta : Jurusan Matematika FMIPA Universitas Sebelas Maret.

Hilbe, J. M. 2011. *Negative Binomial Regression Second Edition.* New York: Cambridge University Press.

Koenker, R. & Bassett, G. 1978. Regression Quantiles. *Econometrika*. **46**, 33-50.

Kiamiantini. 2007. *Pendugaan Statistik Area kecil Dengan Metode Empirical Constrained Bayes*. Seminar Nasional Matematika.

Kurnia, A. 2009. "Prediksi Terbaik Empirik Untuk Model Transformasi Logaritma di Dalam Pendugaan Area Kecil Dengan Penerapan Pada Data Susenas". Tidak Diterbitkan. Disertasi. Bogor: Program Pascasarjana Institut Pertanian Bogor.

Lee, Y., Nelder, J. A. & Pawitan, Y. 2006. *Generalized Linear Models with Random Effects: Unified Analysis Via H-Likelihood*. London: Champman and Hall/CRC-Press.

Longford, N. T. 2005. *Missing Data And Small Area Estimation: Modern Analytical Equipment For Survey Statistician*. New York: Springer Science Busines Media, inc.

Myers, R. H. 1990. *Classical and Modern Regression with Applications* 2nd ed. Boston: PWS-KENT Publishing Company.

Rao, J. N. K. 2003. *Small Area Estimation*. New Jersey. John Willey & Sons.Inc

Sundari, I. 2012. “Regresi Poisson dan Penerpannya untuk Memodelkan Hubungan Usia dan Perilaku Merokok Terhadap Jumlah Kematian Penderita Penyakit Kanker Paru-Paru“. Tidak Diterbitkan. Skripsi. Padang: Jurusan Matematika FMIPA Universitas Andalas.

Supranto, J. 2000. *STATISTIK Teori dan Aplikasi Jilid 1&2*. Jakarta: Penerbit Erlangga.

LAMPIRAN

Lampiran A. Data Penelitian

Area	x_0	x_1	μ_i	β_0	β_1	σ^2	y_D
1	1	0	2.0	2.0	0.3	2.0	3
2	1	0	2.0				2
3	1	0	2.0				1
4	1	1	2.3				1
5	1	1	2.3				5
6	1	1	2.3				2
7	1	2	2.6				0
8	1	2	2.6				0
9	1	2	2.6				2
10	1	3	2.9				3
11	1	3	2.9				0
12	1	3	2.9				2
13	1	4	3.2				0
14	1	4	3.2				0
15	1	4	3.2				3
16	1	5	3.5				0
17	1	5	3.5				0
18	1	5	3.5				2
19	1	6	3.8				0
20	1	6	3.8				1

Lampiran B. *Script* Program Model Regresi Poisson dan M-kuantil

```
library(MASS)
# menetapkan X dan y
x0<-rep(1,20)
x1<-c(0,0,0,1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6)
y<-c(3,2,1,1,5,2,0,0,2,3,0,2,0,0,3,0,0,2,0,1)
X<-cbind(x0,x1)
Y<-cbind(y)
#mendiskripsikan X dan Y
summary(X)
summary(Y)
#menetapkan beta, sigma^2, hitug mu_i
beta0<-2
beta1<-0.3
beta<-as.matrix(c(beta0,beta1))
mu_i<-X%*%beta
print(mu_i)
sigma<-4
#membangkitkan teta
alpha<-mu_i^2/sigma
alphainvers<-sigma/mu_i
teta<-rgamma(alpha,alphainvers)
teta
#glm poisson untuk dugaan Y
pois<-glm(Y ~ x1, family = "poisson", data=tabel1,
          start = NULL, model = TRUE, method = "glm.fit",
          x = FALSE, y = TRUE, contrasts = NULL)
summary(pois)
Y_hat_pois<-fitted.values(pois)
print(Y_hat_pois)

#dugaan Poisson
Ydugahat_pois<- predict(pois)
Ydugahat_pois

#residual poisson
respois<-pois$residuals
respois

#estimasi dengan regresi M-kuantil
library(quantreg)
mk <- rq(Y ~ x1, tau=0.5)
```

```
summary(mk, se="iid")

#nilai MSE pada kuantil 0.5
fit<-rq.fit(X, Y, tau=0.5)
res<-fit$residuals
mse<-(sum(res^2))/20
mse

#regresi M-kuantil untuk dugaan Y
rmk<-rq.fit(X, Y, tau=0.5)
summary(rmk,se='iid')
y_hat_rmk<-fitted.values(rmk)
print(y_hat_rmk)

#koefisien M-kuantil y
epsilon <- abs(Y - X%*%beta)
epsilon

#mencari minimum koefisien dengan y
#1-epsilon
n1 <-rep(1,20)
n1
min1 <- n1-epsilon
min1
#1/exp
min2 <- 1/exp(X%*%beta)
min2
min <- c(min1,min2)
min
min(min)

#koefisien M-kuantil y hat
epsilon1 <- abs(y_hat_rmk - X%*%beta)
epsilon1
#mencari minimum koefisien dengan y
#1-epsilon
n2 <-rep(1,20)
n2
min2 <- n2-epsilon1
min2
#1/exp
min3 <- 1/exp(X%*%beta)
min3
```

```
min1 <- c(min2,min2)
min1
minkoef <- min(min1)
minkoef

#mencari nilai teta dari rata2 koefisien m-kuantil
koef_mkuantil <- c(minkoef,Y)
koef_mkuantil
teta_mkuantil <- mean(koef_mkuantil)
teta_mkuantil

#yhat_mkuantil
yhat_mkuantil <- 1/20*(Y+ (X%*%beta)*teta_mkuantil)  +
1/20 *(X%*%((beta*0.5)-(beta*teta_mkuantil)))
yhat_mkuantil
#beta hat
beta0mkuantil <- 2
beta1mkuantil <- -0.33333
betamkuantil<-as.matrix(c(beta0mkuantil,beta1mkuantil))

#mencari prediktor m-kuantil
yduga_hat <- 1/20*(Y + exp(X%*%betamkuantil))
yduga_hat

#menghitung MSE Bootstrap
bootstrap1<-function(y, x, Ydugahat_pois, B)
{
  m<-length(y)
  Ydugahat_poisB<- matrix(c(0),m,B)
  a<-cbind(y,x)
  mse <- rep(0,B)
  var <- rep(0,B)
  for (j in 1:B)
  {
    nobaris<-sample(c(1:20),20,replace=T)
    c1<-a[nobaris,]
    y<-c1[,1]
    x1<-c1[,3]
    poisB<-glm(y ~ x1, family = "poisson", data=tabel1,
               start = NULL, model = TRUE, method =
"glm.fit",
               x = FALSE, y = TRUE, contrasts = NULL)
    y_hat_poisB<-fitted.values(poisB)
```

```
    Ydugahat_poisB<- predict(poisB)
    respoisB<-poisB$residuals
    mse[j]<-(sum((respoisB)^2))/m
    var[j]<-var(Ydugahat_poisB)
  }
  return(list(teta_poisB,mse,mean(mse),var))
}


bootstrap2<-function(y, x, yduga_hat, B)
{
  m<-length(y)
  yduga_hatB<- matrix(c(0),m,B)
  b<-cbind(y,x)
  mse <- rep(0,B)
  var <- rep(0,B)
  for (k in 1:B)
  {
    nobaris<-sample(c(1:20),20,replace=T)
    c2<-b[nobaris,]
    y<-c2[,1]
    x1<-c2[,3]
    rmkB<-rq.fit(X, Y, tau=0.5)
    y_hat_rmkB<-fitted.values(rmkB)
    epsilon1B<- abs(y_hat_rmkB - X%*%beta)
    n2B <-rep(1,20)
    min2B <- n2B-epsilon1B
    min3B <- 1/exp(X%*%beta)
    min1B <- c(min2B,min3B)
    minkoefB <- min(min1B)
    koef_mkuantilB <- c(minkoefB,Y)
    teta_mkuantilB <- mean(koef_mkuantilB)
    yhat_mkuantilB <- 1/20*(Y+ (X%*%beta)*teta_mkuantilB)
+ 1/20 *(X%*%((beta*0.5)-(beta*teta_mkuantilB)))
    beta0mkuantilB <- 2
    beta1mkuantilB <- -0.33333
    betamkuantilB<-
as.matrix(c(beta0mkuantilB,beta1mkuantilB))
    yduga_hatB <- 1/20*(Y + exp(X%*%betamkuantilB))
    mse[k]<-(sum((yduga_hatB - yhat_mkuantil )^2))/m
    var[k]<-var(yduga_hatB)
  }
  return(list(yduga_hatB,mse,mean(mse),var))
```

```
}

#menghitung MSE bootstrap
hslB1<-bootstrap1(y,x,Ydugahat_pois, 1000)
hslB1[3]
hslB2<-bootstrap2(y,x,yduga_hat, 1000)
hslB2[3]

h1<-as.matrix(unlist(hslB1[2]))
h2<-as.matrix(unlist(hslB2[2]))

boot1000<-cbind(h1,h2)

boxplot(boot1000 , horizontal=F,col="lightblue",
boxwex=.5,ylab="MSE",main="MSE Bootstrap 1000x",add=F)

#menghitung ragam bootstrap
r1<-as.matrix(unlist(hslB1[4]))
r2<-as.matrix(unlist(hslB2[4]))

var1000<-cbind(r1,r2)

plot(r1,col="blue",ylab="ragam",xlab="bootstrap ke-")
lines(r1,col="blue",ylab="ragam",xlab="bootstrap ke-")
lines(r2,col="red",ylab="ragam",xlab="bootstrap ke-")

boxplot(var1000 , horizontal=F,col="lightblue",
boxwex=.5,ylab="MSE",main="ragam MSE Bootstrap
1000x",add=F)
```