



**PENANGANAN OVERDISPERSI PADA PEMODELAN DATA  
CACAHAN MENGGUNAKAN REGRESI BINOMIAL NEGATIF PADA  
RCIM (*ROW COLUMN INTERACTION MODEL*)**

**SKRIPSI**

Oleh

**Upik Susilowati  
NIM 1118101012**

**JURUSAN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS JEMBER  
2015**



**PENANGANAN OVERDISPERSI PADA PEMODELAN DATA  
CACAHAN MENGGUNAKAN REGRESI BINOMIAL NEGATIF PADA  
RCIM (*ROW COLUMN INTERACTION MODEL*)**

**SKRIPSI**

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

Oleh

**Upik Susilowati  
NIM 111810101012**

**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. Ayah Buyung Konok dan Ibu Siani yang telah memberikan doa, cinta, kasih dan semangatnya;
3. Adik tercinta Unik Elsa Safitri, Ummi Tri Handayani dan Vickral Geofany yang telah memberikan dukungan serta motivasi dalam pembuatan skripsi ini; serta Achmad Baichaqi Y yang telah banyak memberikan bantuan dan semangat dengan penuh kesabaran;
4. seluruh guru dan dosen sejak taman kanak-kanak hingga perguruan tinggi yang telah memberikan ilmu dan membimbing dengan penuh perhatian dan kesabaran;
5. almamater Jurusan Matematika FMIPA Universitas Jember, MAN Jember 1, SMP “Plus” Darush Sholah, SDN Blindungan 1, TK Pertiwi.

**MOTTO**

“Orang yang menuntut ilmu berarti menuntut rahmat. Orang yang menuntut ilmu berarti menjalankan rukun Islam dan pahala yang diberikan sama dengan para Nabi”  
( HR. Dailani dari Anas r.a )<sup>\*)</sup>

"Sesuatu yang belum dikerjakan, seringkali tampak mustahil. Kita baru yakin kalau kita telah berhasil melakukannya dengan baik."  
(Evelyn Underhill)<sup>\*)</sup>

---

<sup>\*)</sup> <http://www.maribelajarbk.web.id/2015/03/ccontoh-motto-terbaru-dalam-skripsi.html>

**PERNYATAAN**

Saya yang bertanda tangan di bawah ini :

Nama : Upik Susilowati

NIM : 111810101012

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “Penanganan Overdispersi pada Pemodelan Data Cacahan Menggunakan Regresi Binomial Negatif Pada RCIM (*Row Column Interaction Model*)” 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, Juni 2015

Yang menyatakan,

Upik Susilowati  
NIM 111810101012

**SKRIPSI**

**PENANGANAN OVERDISPERSI PADA PEMODELAN DATA  
CACAHAN MENGGUNAKAN REGRESI BINOMIAL NEGATIF PADA  
RCIM (*ROW COLUMN INTERACTION MODEL*)**

Oleh

**Upik Susilowati**  
**NIM 111810101012**

**Pembimbing**

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

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

**PENGESAHAN**

Skripsi berjudul “Penanganan Overdispersi pada Pemodelan Data Cacahan Menggunakan Regresi Binomial Negatif pada RCIM (*Row Column Interaction Model*)” telah diuji dan disahkan pada :

Hari, tanggal :

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas  
Jember

Tim penguji,

Ketua,

Sekretaris.

Dr. Alfian Futuhul Hadi, S. Si, M. Si  
NIP. 197407192000121001

Dian Anggraeni, S. Si, M. Si  
NIP. 198202162006042002

Anggota I,

Anggota II,

Prof. Drs. I Made Tirta M. Sc., Ph. D  
NIP. 195912201985031002

Ika Hesti Agustin, S.Si., M.Si  
NIP. 198408012008012006

Mengesahkan

Dekan,

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

## RINGKASAN

**Penanganan Overdispersi pada Pemodelan Data Cacahan Menggunakan Regresi Binomial Negatif pada RCIM (Row Column Interaction Model);** Upik Susilowati; 2015; 33 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Regresi Poisson adalah model yang biasa digunakan untuk menganalisis hubungan antara variabel Y dan X pada data cacahan. Data cacahan seringkali mengalami nilai varians yang lebih besar dari rata-ratanya yang disebut overdispersi. Overdispersi pada data cacahan ini dapat menyebabkan nilai devians dari suatu model menjadi sangat besar. Nilai dispersi pada data cacahan dapat dideteksi dengan nilai devians yang dibagi db hasilnya bernilai lebih dari 1. Salah satu penanganan yang dapat dilakukan pada data cacahan yang mengalami overdispersi adalah menggunakan Regresi Binomial Negatif. Variabel respon pada Regresi Binomial Negatif tidak mengharuskan nilai varians sama dengan rata-ratanya.

Penelitian tentang Regresi Binomial Negatif dalam menangani overdispersi pada model aditif telah dilakukan oleh beberapa peneliti namun tidak untuk model multiplikatif. Model multiplikatif merupakan model yang melibatkan dua interaksi dalam tabel dua arah. Model GAMMI (*Generalized AMMI*) merupakan model yang sering digunakan untuk model multiplikatif. Model ini merupakan model dengan interaksi dimana pengaruh utama bersifat aditif dan pengaruh interaksi dimodelkan dengan model *bilinier*. Yee dan Hadi (2014) memperkenalkan *Row Column Interaction Model (RCIM)* yang merupakan perluasan dari konsep *Reduced-Rank Vector Generalized Models (RR-VGLM)*. RCIM merupakan suatu model yang dapat diterapkan sama halnya dengan GAMMI yaitu dapat menganalisis tabel dua arah dengan matriks responnya bisa dalam bentuk cacahan (count). Kesamaan pada RCIM dan GAMMI yaitu terdapat pada pendugaannya yang sama-sama menggunakan metode pendugaan bolak-balik.



Penelitian ini bertujuan untuk mengetahui penanganan overdispersi melalui distribusi Binomial Negatif pada *Row Column Interaction Model* (RCIM) dan suku multiplikatif atau rank model dengan kompleksitas rank pada model yang terbaik. Model terbaik antara Regresi Poisson dan Binomial Negatif dalam mengatasi overdispersi pada data cacahan akan memiliki nilai MSE (*Mean Square Error*) dengan nilai terkecil

Data yang digunakan dalam penelitian ini adalah dua buah data. Data pertama adalah data percobaan naungan yang memiliki 15 galur dan 8 genotip dan mengalami overdispersi. Data kedua adalah data pertama yang mengalami penambahan *outlier* hingga hampir 20%. Data ini akan divisualisasikan interaksinya ke dalam dimensi 2 (Biplot) melalui RCIM agar terlihat stabilitas *genotype* terhadap lingkungannya.

Penelitian yang dilakukan pada dua buah data tersebut diperoleh bahwa hasil analisis devians diperoleh dari suatu nilai residual devians yang didapatkan dari nilai *log-likelihood*. Hasil analisis devians yang didapatkan dari Regresi Poisson dan Regresi Binomial Negatif menunjukkan bahwa Regresi Binomial Negatif lebih baik dalam memodelkan data cacahan yang mengalami overdispersi. Nilai devians pada Regresi Binomial Negatif bernilai lebih kecil daripada Regresi Poisson yang menunjukkan bahwa Regresi Binomial Negatif lebih baik. Model RCIM pada data yang memiliki nilai dispersi dibawah 5 untuk Regresi Poisson dan Regresi Binomial Negatif signifikan pada RCIM 2. Model RCIM pada data yang mengalami penambahan *outlier* untuk Regresi Poisson signifikan pada RCIM 3 sedangkan untuk Regresi Binomial Negatif signifikan pada RCIM 1. Hasil yang diperoleh dari biplot adalah galur 8 stabil pada semua lokasi

## PRAKATA

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat serta hidayahNya sehingga skripsi yang berjudul “Penanganan Overdispersi pada Pemodelan Data Cacahan Menggunakan Regresi Binomial Negatif pada RCIM (*Row Column Interaction Model*)“ dapat terselesaikan. Skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan strata 1 (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. Dr. Alfian Futuhul Hadi, S. Si, M. Si selaku Dosen Pembimbing Utama dan Dian Anggraeni, 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. seluruh dosen dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam yang telah memberikan ilmu serta fasilitas yang membantu selama proses perkuliahan berlangsung;
4. sahabat Dwiki, Sella, Eka, Kiki, Novi 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;
5. keluarga besar UKMS TITIK yang telah memberikan pengalaman dalam dunia seni dan cerita yang sangat berharga dan kenangan yang tidak akan terlupakan;

6. kos Pervokma yang telah membantu dan selalu memberikan dukungan dalam hal apapun;
7. 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
<b>RCIM (<i>ROW COLUMN INTERACTION MODEL</i>)</b> .....	<b>i</b>
<b>SKRIPSI</b> .....	<b>i</b>
<b>PERSEMBAHAN</b> .....	<b>ii</b>
<b>MOTTO</b> .....	<b>iii</b>
<b>PERNYATAAN</b> .....	<b>iv</b>
<b>SKRIPSI</b> .....	<b>v</b>
<b>PENGESAHAN</b> .....	<b>vi</b>
<b>RINGKASAN</b> .....	<b>vii</b>
<b>PRAKATA</b> .....	<b>ix</b>
<b>DAFTAR ISI</b> .....	<b>xi</b>
<b>DAFTAR TABEL</b> .....	<b>xiii</b>
<b>DAFTAR GAMBAR</b> .....	<b>xiv</b>
<b>DAFTAR LAMPIRAN</b> .....	<b>xv</b>
<b>BAB 1. PENDAHULUAN</b> .....	<b>1</b>
<b>1.1 Latar Belakang</b> .....	<b>1</b>
<b>1.2 Rumusan Masalah</b> .....	<b>2</b>
<b>1.3 Tujuan Penelitian</b> .....	<b>3</b>
<b>1.4 Manfaat Penelitian</b> .....	<b>3</b>
<b>BAB 2. TINJAUAN PUSTAKA</b> .....	<b>4</b>
<b>2.1 <i>Generalized Linear Model</i> (GLM)</b> .....	<b>4</b>
<b>2.2 Regresi Poisson</b> .....	<b>5</b>
<b>2.3 Overdispersi</b> .....	<b>6</b>
<b>2.4 Regresi Binomial Negatif</b> .....	<b>7</b>
<b>2.5 <i>Row Column Interaction Model</i> (RCIM)</b> .....	<b>8</b>
<b>2.6 Uji Kelayakan Model (Goodnes of Fit)</b> .....	<b>9</b>
2.6.1 Analisis Devians .....	9
2.6.2 <i>Mean Square Error</i> (MSE) .....	10
<b>2.7 Analisis Biplot</b> .....	<b>11</b>

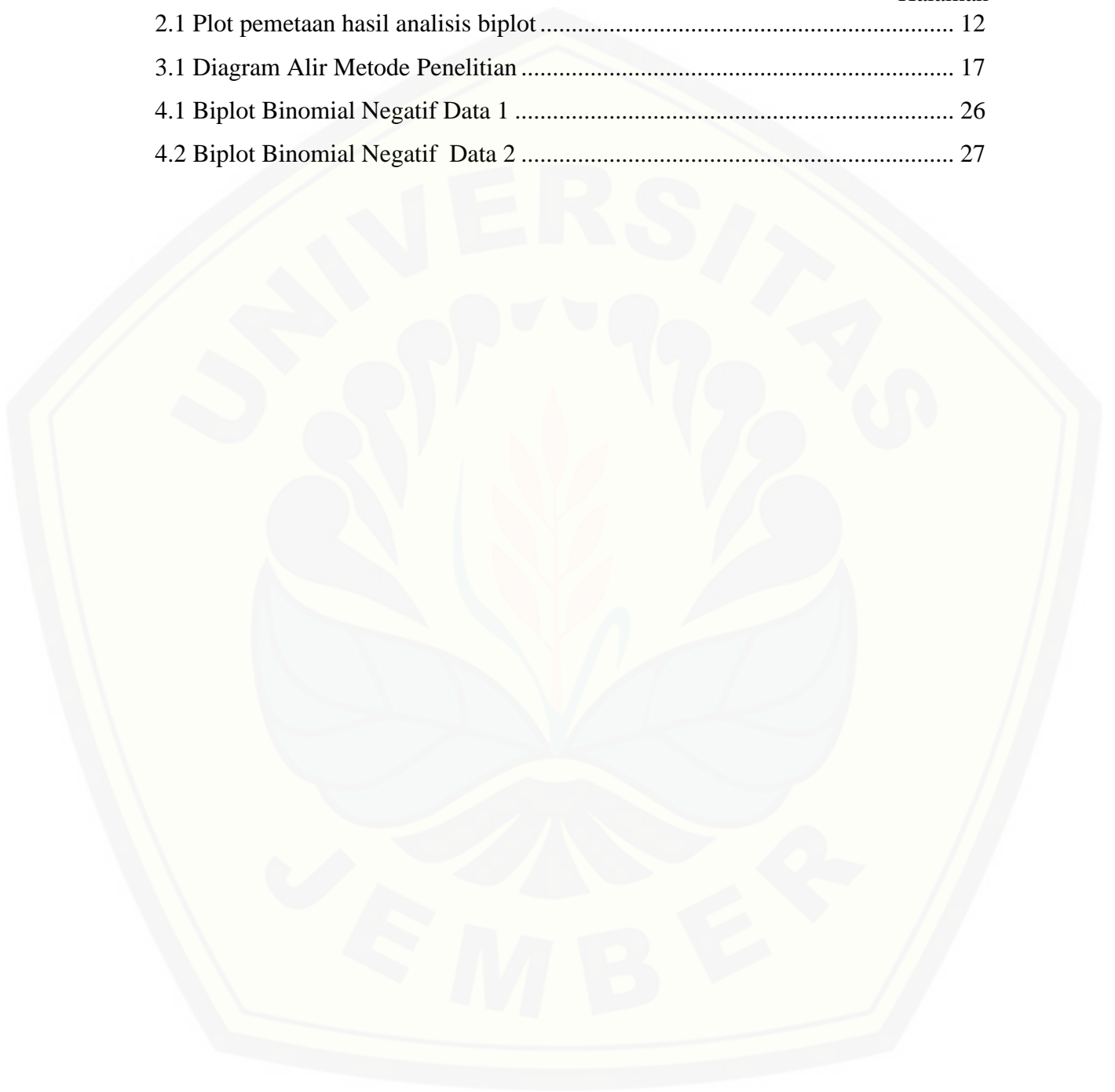
<b>BAB 3. METODE PENELITIAN.....</b>	<b>14</b>
3.1 Data dan Skenario Simulasi .....	14
3.2 Metode Analisis.....	14
<b>BAB 4. HASIL DAN PEMBAHASAN.....</b>	<b>18</b>
4.1 Deskripsi Data.....	18
4.2 Uji Overdispersi.....	19
4.3 Analisis Data Menggunakan Regresi Poisson dan Binomial Negatif ..	20
4.4 Pengaruh Penambahan <i>Outlier</i> pada Nilai Dispersi dan MSE .....	24
4.5 Interpretasi Biplot pada Regresi Binomial Negatif pada RCIM .....	25
<b>BAB 5. PENUTUP.....</b>	<b>29</b>
5.1 Kesimpulan .....	29
5.2 Saran.....	29
<b>DAFTAR PUSTAKA .....</b>	<b>30</b>
<b>LAMPIRAN.....</b>	<b>32</b>

**DAFTAR TABEL**

	Halaman
2.1 Fungsi link kanonik dalam GLM .....	5
4.1 Hasil Uji Overdispersi.....	19
4.2 Nilai Log-Likelihood Data 1 pada $rcim$ .....	20
4.3 Nilai Log-Likelihood Data 2 pada $rcim$ .....	21
4.4 Nilai Devians dan Residual Devians Data 1 pada $rcim$ .....	22
4.5 Nilai Devians dan Residual Devians Data 2 pada $rcim$ .....	23
4.6 Nilai MSE pada Data 1 dan Data 2 .....	24

**DAFTAR GAMBAR**

	Halaman
2.1 Plot pemetaan hasil analisis biplot .....	12
3.1 Diagram Alir Metode Penelitian .....	17
4.1 Biplot Binomial Negatif Data 1 .....	26
4.2 Biplot Binomial Negatif Data 2 .....	27



**DAFTAR LAMPIRAN**

	Halaman
A.1 Data Jumlah Polong Isi Kedelai .....	32
A.2 Data Simulasi Jumlah Polong Isi Kedelai .....	32
B.1 Analisis Devian Model Regresi Poisson (Data 1) .....	33
B.1 Analisis Devian Model Regresi Binomial Negatif (Data 1).....	35
B.3 Analisis Devian Model Regresi Poisson (Data 2) .....	37
B.4 Analisis Devian Model Regresi Binomial Negatif (Data 2).....	39
C.1 Grafik Pengaruh Penambahan Outlier pada Nilai Dispersi.....	41
C.2 Grafik Pengaruh Penambahan Outlier pada MSE.....	42
D.1 Script Program Regresi Poisson pada RCIM (Data 1).....	43
D.2 Script Program Regresi Binomial Negatif pada RCIM (Data 1) .....	45
D.3 Script Program Regresi Poisson pada RCIM (Data 2).....	47
D.4 Script Program Regresi Binomial Negatif pada RCIM (Data 2) .....	49
D.5 Script Biplot Regresi Binomial Negatif pada Data 1 .....	51
D.6 Script Biplot Regresi Binomial Negatif pada Data 2 .....	52
E. Hubungan Regresi Binomial Negatif Dengan Regresi Poisson .....	53



## BAB 1. PENDAHULUAN

### 1.1 Latar Belakang

Model Regresi Poisson adalah model regresi yang dapat digunakan untuk menganalisis hubungan antara variabel respon Y yang berupa data diskrit dengan variabel prediktor X. Model ini memiliki beberapa asumsi, salah satu asumsi yang harus terpenuhi adalah varians dari variabel *dependent* Y sama dengan rataannya (*equidispersion*). Penggunaan model Regresi Poisson dalam analisis data diskrit seringkali mengalami kasus nilai varians lebih besar dari meannya yang disebut overdispersi. Nilai varians yang lebih kecil dari nilai meannya disebut underdispersi (McCullagh & Nelder). Overdispersi pada Regresi Poisson menyebabkan parameter dispersi ( $\alpha$ ) bernilai lebih dari satu dan bersifat konstan. Menurut Famoye *et al.* (2004), Regresi Poisson tidak sesuai dalam memodelkan data overdispersi atau underdispersi.

Salah satu penanganan yang dapat digunakan ketika terjadi overdispersi pada Regresi Poisson adalah dengan menggunakan model Regresi Binomial Negatif. Hasil penelitian Pamungkas (2003) menyatakan bahwa saat terjadi overdispersi pada data, Regresi Binomial Negatif lebih baik digunakan dibandingkan Regresi Poisson. Simamarta (2010) dalam tugas akhirnya membahas mengenai Regresi Binomial Negatif, model regresi ini dapat mengatasi overdispersi pada Regresi Poisson karena variabel responnya diasumsikan mengikuti Regresi Binomial Negatif sehingga tidak mengharuskan nilai mean sama dengan nilai varians. Kedua peneliti tersebut hanya melibatkan satu interaksi saja dalam penelitiannya atau yang biasa disebut dengan model aditif .

Model aditif telah banyak digunakan oleh beberapa peneliti, sehingga untuk penelitian ini, peneliti menggunakan model multiplikatif pada Regresi Binomial Negatif. Model multiplikatif (bilinier) merupakan model yang melibatkan dua interaksi dan menjembatani kesenjangan antara model pengaruh utama dan model interaksi lengkap dengan sebuah parameter interaksi untuk setiap sel dalam tabel dua arah. Model ini pun memberikan visualisasi interaksi

melalui biplot sehingga pengembangan teori GLM dengan menggunakan komponen multiplikatif untuk interaksi sangat diperlukan. Van Eeuwijk (1995) memperkenalkan model *Generalized AMMI* (GAMMI) yang merupakan salah satu model yang dapat digunakan untuk menganalisa interaksi lengkap dua faktor perlakuan dengan pengaruh utama bersifat aditif dan pengaruh interaksi dengan suku multiplikatif pada data yang mempunyai respon tidak normal.

Yee dan Hastie (2003) memperkenalkan *Reduced-Rank Vector Generalized Models* (RR-VGLM) yang mengaplikasikan konsep *Reduced Rank Regression* ke dalam VGLM. *Row Column Interaction Model* (RCIM) merupakan perluasan dari konsep RR-VGLM yang diperkenalkan oleh Yee dan Hadi (2014). RCIM adalah model yang dapat menganalisis tabel dua arah dengan matriks responnya bisa dalam bentuk cacahan (*count*). Menurut Ahadiyah (2015), hasil pengepasan model GAMMI dengan menggunakan RCIM (*Row Column Interaction Model*) pada Regresi Poisson sama dengan pengepasan model GAMMI dengan menggunakan regresi bolak-balik berdasarkan Eeuwijk (1995). Dari penelitian Ahadiyah tersebut, dapat diambil kesimpulan bahwa RCIM juga dapat digunakan untuk model multiplikatif. Sehingga penelitian ini menggunakan Regresi Binomial Negatif pada RCIM untuk menganalisa data cacahan (*count*) pada data penelitian Ahadiyah (2015) yang mengalami overdispersi. Kemudian akan dibandingkan model terbaik antara Regresi Poisson dengan Regresi Binomial Negatif dengan menggunakan MSE. MSE adalah *Mean Square Error* atau dapat diartikan sebagai *error* dari masing-masing model. Model yang terbaik dalam memodelkan data overdispersi akan memiliki nilai MSE yang lebih kecil.

## 1.2 Rumusan Masalah

Sesuai dengan konsep *Reduced Rank Regression* maka penanganan overdispersi akan ditangani melalui :

- a. distribusi Binomial Negatif pada *Row Column Interaction Model* (RCIM)
- b. pembahasan suku multiplikatif atau rank model dalam hal ini akan diamati dari kompleksitas rank dengan model yang terbaik.

Setelah itu akan dievaluasi model terbaik antara Regresi Poisson dan Binomial Negatif dalam mengatasi overdispersi pada data cacahan dengan membandingkan MSE pada masing-masing model.

### 1.3 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengetahui penanganan overdispersi melalui :

- a. distribusi Binomial Negatif pada *Row Column Interaction Model* (RCIM)
- b. suku multiplikatif atau rank model dengan kompleksitas rank pada model yang terbaik.

Model terbaik antara Regresi Poisson dan Binomial Negatif dalam mengatasi overdispersi pada data cacahan dengan membandingkan MSE pada masing-masing model.

### 1.4 Manfaat Penelitian

Manfaat dari penelitian ini adalah dapat memberi pengetahuan tentang Regresi Binomial Negatif sebagai salah satu cara untuk mengatasi overdispersi. Selain itu, penelitian ini juga dapat digunakan sebagai referensi bagi peneliti lainnya tentang *Row Column Interaction Model* (RCIM).

## BAB 2. TINJAUAN PUSTAKA

### 2.1 *Generalized Linear Model* (GLM)

Model linier khususnya model regresi sudah mulai digunakan sejak awal abad 19, ditandai dengan kajian yang dilakukan oleh Francis Galton (1822-1911) tentang hubungan tinggi badan ayah dan anaknya. Dalam perkembangannya, model regresi linear dengan asumsi peubah respon  $Y_i \sim N(\mu_i, \sigma^2)$  tidaklah mampu menjawab masalah-masalah yang dihadapi dalam pemodelan statistik. Oleh karena itu perlu metode lain yang dapat memodelkan data-data yang tidak memiliki karakteristik pada model linier klasik. Nelder dan Wedderburn (1972) telah mengembangkan model linier yang dikenal dengan *Generalized Linear Model* (GLM) untuk menangani kondisi dimana respon yang ada tidak berdistribusi normal, tetapi masih saling bebas. *Generalized Linear Model* (GLM) merupakan pengembangan dari model linear “klasik” khususnya dalam mengatasi kendala peubah respon yang tidak normal. Namun demikian, asumsi yang digunakan dalam model ini diasumsikan memiliki sebaran yang termasuk dalam keluarga eksponensial yang meliputi Normal, Gamma, Poisson, dll.

Ada tiga komponen utama dalam GLM (McCullagh dan Nelder, 1989) yaitu :

- a. komponen acak, yaitu peubah respon  $Y_i = Y_1, Y_2, \dots, Y_n$  dan termasuk dalam keluarga eksponensial.
- b. komponen prediktor  $X$  dan koefisien  $\beta$ , membentuk prediktor linier dengan fungsi  $\eta_i = \sum_{j=0}^k x_{ij} \beta_j$ .
- c. fungsi *link*,  $g()$ , yang menghubungkan suatu fungsi dari nilai tengah komponen acak dan menghasilkan vektor peubah data  $g(\mu_i) = \eta_i$ .

Dibawah ini diberikan tabel fungsi link yang sering digunakan pada sebuah data.

Tabel 2.1 Fungsi link kanonik dalam GLM

Family	Nama	Link
Normal	Identitas	$\eta = y$
Poisson	Log	$\eta = \ln y$
Binomial	Logit	$\eta = \ln \frac{y}{1-y}$
Negatif Binomial	Log	$\eta = \ln y$
Gamma	Kebalikan	$\eta = y^{-1}$

Sumber: Hadi *et al.* (2010).

## 2.2 Regresi Poisson

Model Regresi Poisson merupakan model regresi yang termasuk dalam *Generalized Linier Models* (GLM). Model ini menggunakan data cacahan dimana variabel responnya mengikuti distribusi Poisson. Kondisi respon pada regresi ini tidak berdistribusi normal melainkan mengikuti keluarga eksponensial. Variabel respon dalam Regresi Poisson berasal dari data cacahan yang diharapkan jarang terjadi. Regresi Poisson merupakan analisis regresi nonlinier dari distribusi poisson, dimana analisis ini sangat cocok digunakan dalam menganalisis data diskrit (*count*). Regresi ini sering digunakan untuk memodelkan peristiwa yang memiliki peluang kejadian kecil dengan kejadiannya tergantung pada interval waktu tertentu (Osgood, 2000).

Casella & Berger (1990) menyatakan data cacah yang diasumsikan menyebar Poisson memiliki fungsi peluang sebagai berikut dengan selang waktu rata-rata  $\lambda_i$ .

$$P(y; \lambda) = \frac{e^{-\lambda} \lambda^y}{y!}; y = 0, 1, 2, \dots, \text{ dan } \lambda > 0 \quad (2.1)$$

Diperoleh nilai mean dan varian dengan parameter  $\lambda$  adalah sama yaitu  $E(Y) = \text{Var}(Y) = \lambda$ , maka persamaan (2.1) dapat dituliskan sebagai berikut :

$$\begin{aligned} f(y; \lambda) &= \exp\{y \ln(\lambda) - \lambda - \ln(y!)\} \\ &= \exp\{y\theta - b(\theta) - \ln(y!)\} \end{aligned}$$

dan ketika  $g(\lambda)$  sama dengan parameter natural  $\theta$ , maka prediktor linier  $\eta_i = \ln(\lambda)$ . Regresi Poisson memiliki fungsi  $g$  yang menghubungkan rata-rata dari variabel responnya dengan sebuah prediktor linier.

$$\ln E(y|x) = \ln(\lambda_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}$$

$$\lambda_i = \exp(x'_i \hat{\beta}) = \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik})$$

Model Regresi Poisson dapat ditulis sebagai berikut.

$$y_i = t_i \exp(x'_i \hat{\beta}) + \varepsilon_i$$

Dengan  $\lambda_i$  merupakan nilai ekspektasi dari  $y_i$  yang berdistribusi poisson dimana  $i = 1, 2, \dots, n$  (Hilbe, 2011). Jika  $\lambda_i$  adalah rata-rata jumlah kejadian dalam periode  $t_i$  dan diasumsikan  $\lambda_i$  tidak berubah dari titik data ke titik data secara bebas maka  $\lambda_i$  dapat dimodelkan sebagai fungsi dari  $k$  variabel prediktor.

### 2.3 Overdispersi

Salah satu masalah pada Regresi Poisson adalah adanya overdispersi. Hilbe (2011) menjelaskan bahwa overdispersi pada regresi Poisson terjadi ketika varians dari respon lebih besar dari meannya. Penggunaan Regresi Poisson pada data cacah yang mengalami overdispersi akan mempengaruhi nilai *standart error* yang menjadi turun atau *underestimate*, sehingga kesimpulannya menjadi tidak valid. Parameter overdispersi pada Regresi Binomial Negatif biasa disimbolkan dengan  $\alpha$ .

Berikut beberapa hal yang dapat menyebabkan overdispersi, antara lain :

- terdapat korelasi antar pengamatan.
- terdapat pelanggaran asumsi distribusi Poisson, yaitu mean tidak sama dengan varians.
- terdapat angka nol yang sangat banyak.
- terdapat *outlier* dalam data.

Overdispersi pada Regresi Poisson dapat dideteksi dengan nilai *Pearson Chi-Square* ( $\chi^2$ ) atau devians yang dibagi dengan derajat bebasnya. Jika kedua nilai ini lebih besar dari 1 maka dikatakan terjadi overdispersi pada data. Menurut Hilbe (2011), menggunakan Regresi Poisson jika nilai dispersi tidak lebih dari

1,25 untuk model yang berukuran sedang, sedangkan untuk model yang berukuran besar tidak lebih dari 1,05.

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. Hasil penelitian Pamungkas (2003) menyatakan bahwa saat terjadi overdispersi pada data, Regresi Binomial Negatif lebih baik digunakan dibandingkan Regresi Poisson.

#### 2.4 Regresi Binomial Negatif

Model Regresi Binomial Negatif memiliki kegunaan yang sama dengan model Regresi Poisson yaitu untuk menganalisis hubungan antara suatu variabel respon data *count* dengan satu atau lebih variabel acak penjelas, tetapi model Regresi Binomial Negatif lebih fleksibel dibandingkan dengan model Poisson karena mean dan variansi dari model Binomial Negatif tidak harus sama. Model Binomial Negatif merupakan salah satu solusi untuk mengatasi masalah overdispersi yang didasarkan pada model campuran Poisson-Gamma (Hardin & Hilbe 2007). Campuran Gamma menghasilkan adanya nilai overdispersi pada Regresi Poisson. Distribusi  $y$  dikondisikan pada nilai  $x$  dan  $u$  adalah means dan variansi dari Regresi Poisson yaitu  $\mu$ . Mean yang dihasilkan dari campuran Poisson-Gamma menjadi  $\lambda u$ . Distribusi bersyarat  $y$  didapatkan dari :

$$P(y|x, u) = \int_0^{\infty} \frac{e^{-(\lambda_i u_i)} (\lambda_i u_i)^{y_i}}{y_i!} g(u_i) du_i$$

Distribusi bersyarat dari  $y$  ditentukan dari definisi  $g(u)$ . Model distribusi Gamma diberikan nilai  $u = \exp(\varepsilon)$  dimana  $\ln(\mu) = x\beta + \varepsilon$  sehingga didapatkan :

$$\begin{aligned} P(y|x, u) &= \int_0^{\infty} \frac{e^{-(\lambda_i u_i)} (\lambda_i u_i)^{y_i}}{y_i!} \frac{v^v}{\Gamma(v)} u_i^{v-1} e^{-v u_i} du_i \\ &= \frac{\lambda_i^{y_i}}{\Gamma(y_i+1)} \frac{v^v}{\Gamma(v)} \int_0^{\infty} e^{-(\lambda_i+v)u_i} u_i^{(y_i+v)-1} du_i \end{aligned}$$

Nilai  $v = \frac{1}{\alpha}$  dimana  $\alpha$  adalah parameter dispersi dari Regresi Binomial Negatif dan menyamakan nilai  $\lambda = \mu$  sehingga didapatkan fungsi kepadatan peluang :

$$P(y; \mu, \alpha) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \cdot \left(\frac{1}{1 + \alpha\mu_i}\right)^{y_i} \cdot \left(\frac{1}{1 + \alpha\mu_i}\right)^{\frac{1}{\alpha}}$$

Nilai varians yang digunakan untuk model Binomial Negatif yang merupakan model campuran dari Poisson-Gamma adalah  $\mu + \alpha\mu^2$  dimana  $\mu$  adalah varians dari model Poisson dan  $\alpha\mu^2$  adalah varians dari gamma. Peubah tambahan yang dimasukkan memiliki sebaran gamma dengan asumsi nilai tengah 1 dan ragam dalam nilai rata-rata sebaran Poisson (McCullagh & Nelder 1989).

Jika  $\alpha$  menuju nol maka  $\text{var}(Y_i)$  menuju  $\mu_i$  sehingga distribusi Binomial Negatif akan mendekati suatu distribusi Poisson yang mengasumsikan mean dan varians sama.

## 2.5 Row Column Interaction Model (RCIM)

*Reduced Rank Vector Generalized Linear Models* (RR-VGLM) adalah pengaplikasian dari konsep *reduced rank* ke dalam VGLM. VGLM menyediakan prediktor linier yang lebih banyak daripada GLM. RR-VGLM pertama kali diperkenalkan oleh Yee dan Hasti pada tahun 2003. Persamaan fungsi model VGLM dengan salah satu prediktor linier ke  $-j$  dari banyaknya  $M$  prediktor linier dapat dituliskan sebagai berikut :

$$\eta_j = \eta_j(x) = \beta_j^t x = \sum_{k=1}^p \beta_{(j)k} x_k \text{ dimana } j = 1, \dots, M$$

dimana  $\eta_j$  pada VGLM diaplikasikan secara langsung pada parameter distribusi yang digunakan. Sehingga untuk RR-VGLM diberikan fungsi :

$$\eta = B_1^T x_1 + B_2^T x_2 = B_1^T x_1 + AC^T x_2 = B_1^T x_1 Av$$

dimana  $B_2$  mengalami *reduced rank regression* menjadi  $AC^T$  dengan  $C = (c_1 c_2 \dots c_R)$  berordo  $p \times r$ ,  $A = (a_1 a_2 \dots a_R)$  berordo  $M \times r$  dan  $r \leq \min(M, p)$  yang diketahui sebagai rank.

*Row Column Interaction Model* (RCIM) merupakan pengembangan konsep bagian RR-VGLM, yang dikenalkan oleh Yee dan Hadi (2014). Model ini



dapat menganalisis tabel dua arah dengan matriks responnya bisa dalam bentuk cacahan (*count*). Variabel bebas pada tabel dua arah yaitu nomer baris dan kolom sedangkan variabel terikat adalah angka-angka yang terdapat pada tabel. RCIM didefinisikan sebagai RR-VGLM yang diaplikasikan terhadap Y dengan fungsi :

$$g_1(\theta_1) \equiv \eta_{1ij} = \mu + \alpha_i + \gamma_j + \sum_{r=1}^R c_{ir} a_{jr}$$

dimana  $c$  adalah matrik dengan rank  $r \leq \min(M, p)$ . Dari fungsi tersebut dapat diambil kesimpulan bahwa parameter pertama dari model tersebut memiliki hubungan yang berkaitan dengan matrik respons. Setelah mengalami transformasi yang sesuai, maka parameter tersebut akan sebanding dengan jumlah rata-rata, efek baris dan kolom serta ditambahkan dengan pengaruh interaksi pada  $\mathbf{AC}^T$ .

Fungsi pada model RCIM memiliki parameter  $\alpha_i$  dan  $\gamma_j$  yang disebut nilai atau efek dari masing-masing baris dan kolom. Berdasarkan Yee dan Hastie (2003), RCIM dapat disesuaikan dengan RR-VGLM dengan memodifikasi variabel indikator yang tepat menjadi sebagai berikut.

$$\begin{aligned} \eta_i &= B_1^T x_{1i} + \mathbf{AC}^T x_{2i} \\ &= \left\{ \mu \mathbf{1}_M + \alpha_i \mathbf{1}_M + \sum_{j=2}^M \gamma_j e_j \right\} + \begin{pmatrix} \mathbf{0}^T \\ \mathbf{A}_{[-1]} \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{C}_{[-1]}^T \end{pmatrix} e_i \end{aligned}$$

dimana  $[-1]$  berarti elemen atau baris pertama dihapus dari vektor atau matriks dan  $e_i$  adalah komponen ke- $i$  dari  $\mathbf{1}_M$ .

## 2.6 Uji Kelayakan Model (Goodnes of Fit)

Analisis Regresi Binomial Negatif juga akan diuji terhadap ketepatan pada model yang terbaik. Pengujian ini diperlukan untuk memastikan tidak adanya kelemahan atas kesimpulan dari model yang diperoleh. Terdapat beberapa metode untuk membandingkan suatu model dengan model yang lain diantaranya.

### 2.6.1 Analisis Devians

Untuk pengujian model Regresi Poisson dan Binomial Negatif melalui pada RCIM digunakan analisis devians. Penentuan komponen multiplikatif

menggunakan uji F, dengan membandingkan rasio antara rata-rata devians komponen yang diuji rata-rata devians galat terhadap nilai F-tabel (Hadi, 2006).

Analisis devians merupakan salah satu analisis yang digunakan dalam analisis regresi pada pembentukan suatu model. Dalam suatu analisis regresi, nilai devians digunakan untuk menentukan variabel yang disertakan pada suatu model. Sedangkan pada penelitian ini, nilai devians digunakan untuk menentukan rank yang sesuai pada suatu model.

Pawitan (2001) menjelaskan bahwa devians untuk sebuah model didefinisikan sebagai *Likelihood Ratio Test* (LRT) dari *saturated* model dengan *fitting* model sebagai berikut.

$$D = 2 \log \frac{L(y; y)}{L(\mu; y)}$$

dimana  $L(y; y)$  merupakan *Likelihood* dari *saturated* model dan  $L(\mu; y)$  merupakan *Likelihood* dari *fitting* model. Devians adalah rasio *Log Likelihood* yang membandingkan model saturasi dengan model pada GLM. *P-value* digunakan untuk membantu proses analisis devians pada suatu model setelah didapatkan nilai residu devians.

### 2.6.2 Mean Square Error (MSE)

Setiap penelitian seringkali memilih metode yang berbeda untuk memilih model yang terbaik. Salah satu metode yang digunakan untuk menentukan suatu model yang paling baik sehingga memiliki nilai kesalahan yang cenderung kecil yaitu *Mean Squared Error* (MSE). MSE adalah salah satu metode yang digunakan untuk mengevaluasi suatu model. Masing-masing *error* atau *residuals* dikuadratkan yang kemudian dihitung nilai rata-ratanya.

$$MSE = \frac{1}{n} \sum (y - \hat{y})^2$$

dimana  $n$  merupakan jumlah sampel,  $y$  merupakan variabel respon dan  $\hat{y}$  merupakan variabel prediktor. Semakin kecil nilai MSE maka suatu model akan semakin baik karena kesalahan yang ada pada suatu model akan mendekati nilai 0.

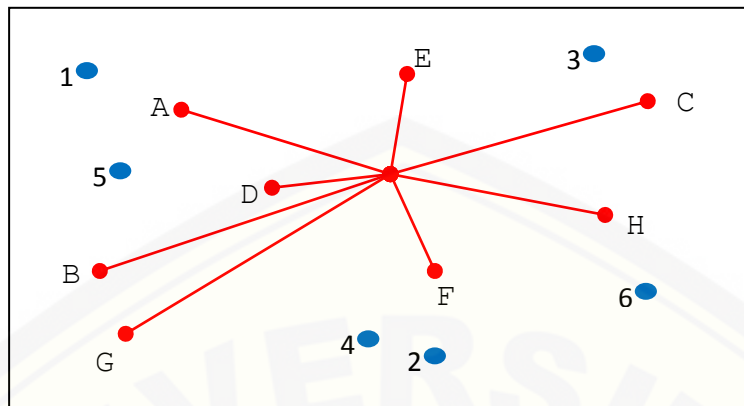
## 2.7 Analisis Biplot

Menurut Mattjik dan Sumertajaya (2011) ada banyak metode yang digunakan untuk mendeskripsikan posisi relatif beberapa objek dengan beberapa variabel atau peubah secara bersamaan. Salah satu metodenya adalah analisis biplot. Biplot merupakan upaya memberikan peragaan grafik dari matriks data  $x$  dalam suatu plot dengan menggabungkan vektor-vektor dalam ruang berdimensi rendah. Biplot pertama kali digunakan oleh Gabriel pada tahun 1971. Gabriel (1971) menjelaskan bahwa biplot adalah metode yang paling sering digunakan dalam analisis multivariat untuk menggambarkan baris dan kolom ke dalam suatu grafik. Tampilan biplot yaitu dua dimensi menyajikan visual objek dan variabel dalam satu grafik sehingga ciri-ciri peubah dan obyek pengamatan dapat dianalisis dengan melihat kemiripan karakteristik antar objek maupun keragaman peubah. Secara umum biplot adalah penguraian matriks tujuan menjadi hasil dari dua matriks yang disebut matriks kanan dan kiri  $\mathbf{S} = \mathbf{XY}'$ . Elemen dari matriks tujuan  $\mathbf{S}$  sama dengan hasil skalar antara pasangan korespondensi dari vektor pada baris  $X$  dan  $Y$  secara berturut-turut.

Menurut Sartono *et al.* (2003), analisis biplot didasarkan pada *Singular Value Decomposition* (SVD). Biplot dapat dibangun dari suatu matriks data, dengan masing-masing kolom mewakili suatu variabel dan masing-masing baris mewakili objek penelitian. Dengan penyajian seperti ini, ciri-ciri peubah dan objek pengamatan serta posisi relatif antara objek pengamatan dengan peubah dapat dianalisis (Jolliffe, 2002).

Biplot adalah upaya membuat gambar di ruang berdimensi banyak menjadi gambar di ruang dimensi dua. Pereduksian dimensi ini harus dibayar dengan menurunnya besarnya informasi yang terkandung dalam biplot. Biplot yang mampu memberikan informasi sebesar 70% dari seluruh informasi dianggap cukup (Mattjik dan Sumertajaya, 2011).

Ilustrasi biplot secara umum dapat dilihat pada gambar :



Gambar 2.1 Plot pemetaan hasil analisis biplot

Anak panah yang berupa vektor merupakan gambaran dari variabel penelitian sedangkan titik-titik merupakan gambaran dari objek penelitian. Menurut Mattjik dan Sumertajaya (2011), empat hal yang bisa didapatkan dari tampilan Biplot adalah sebagai berikut.

a. Kedekatan antar objek yang diamati

Dua objek yang posisinya berdekatan memiliki karakteristik yang sama. Sehingga pada biplot dapat diketahui objek-objek yang memiliki karakteristik yang sama. Dari Gambar dapat diketahui bahwa objek 4 dan 2 memiliki karakteristik yang sama karena posisinya yang berdekatan.

b. Keragaman variabel peubah

Pada biplot, vektor pendek diartikan sebagai variabel yang memiliki nilai keragaman yang kecil sedangkan vektor panjang diartikan sebagai variabel yang memiliki nilai keragaman yang besar. Sehingga, dapat diketahui variabel dengan keragaman yang kecil maupun besar. Variabel D memiliki nilai keragaman yang kecil karena digambarkan sebagai vektor pendek. Variabel G memiliki nilai keragaman yang besar karena digambarkan sebagai vektor panjang.

c. Korelasi antar peubah

Hubungan antar variabel juga dapat diketahui dari biplot dimana dua variabel yang memiliki nilai korelasi yang positif akan digambarkan dengan sebuah vektor dengan arah yang sama atau membentuk sudut lancip. Sedangkan

untuk vektor dengan arah yang berlawanan atau membentuk sudut tumpul diartikan memiliki nilai korelasi yang kecil. Variabel B dan G memiliki nilai korelasi yang positif karena digambarkan dengan dua vektor yang membentuk sudut lancip. Variabel A dan H memiliki nilai korelasi yang negatif karena digambarkan dengan dua vektor yang membentuk sudut tumpul.

d. Nilai peubah pada suatu objek

Objek yang terletak searah dengan vektor variabel adalah objek yang memiliki nilai di atas rata-rata. Sedangkan untuk objek yang berlawanan dengan vektor variabel adalah objek yang memiliki nilai di bawah rata-rata. Namun, untuk objek yang berada di tengah, memiliki nilai yang dekat dengan rata-rata. Dari arah vektor peubah dan objek akan didapatkan nilai peubah dari masing-masing objek.

## BAB 3. METODE PENELITIAN

Metode yang digunakan pada penelitian ini meliputi data dan analisis data yang mengalami overdispersi pada regresi Poisson menggunakan regresi Binomial Negatif pada *Row Column Interaction Model* (RCIM).

### 3.1 Data dan Skenario Simulasi

Data yang digunakan dalam penelitian ini adalah data penelitian Ahadiyah (2015) yaitu data sekunder dari Balitkabi Malang dan hasil modifikasi dari data tersebut yang kemudian diproses perhitungannya menggunakan salah satu program statistik yaitu program R Studio. Data tersebut berisi hasil percobaan terhadap 15 jenis galur kedelai yang ditanam pada 8 lokasi berbeda dengan hasil berupa rata-rata isi kedelai dalam setiap pohonnya. Interaksi yang terdapat pada data adalah antara Genotype dan Lingkungan ( $G \times L$ ).

Data kedua merupakan data cacahan yang berdistribusi Poisson dan dimodifikasi dengan menambahkan *outlier* pada masing-masing baris dan kolom hingga mencapai 20% dari data. Penambahan *outlier* dilakukan dengan menambahkan nilai maksimal dari masing-masing baris dan kolom dengan empat kali standar deviasi (Kriegel, 2010). Secara matematis dituliskan menjadi  $\max + 4\sqrt{stdev}$ .

### 3.2 Metode Analisis

Metode yang digunakan untuk menganalisis data pada penelitian ini adalah dengan menggunakan RCIM. Untuk mempermudah dalam proses perhitungan digunakan bantuan dari salah satu *software* dalam statistika yaitu Rstudio. Cara penggunaan Rstudio mirip dengan Program R. R Studio adalah Integrated Development Environment (IDE) utama bagi pengembang yang menggunakan bahasa R. Uniform Resource Locator (URL) R Studio dapat diunduh pada link <http://www.rstudio.com/ide/download/desktop>. Menurut Tirta (2009), *source* kode

dari program R dapat diakses, dimodifikasi dan dikembangkan secara bebas sesuai dengan keperluan.

Untuk menganalisis data menggunakan RCIM dibantu dengan paket yang tersedia pada Rstudio yaitu paket VGAM. Paket ini menyediakan fungsi untuk vektor GLM dan model aditif lainnya seperti VGLM dan VGAM. Fungsi yang digunakan dalam penelitian diantaranya.

- a. `vglm`, digunakan untuk memodelkan suatu data menjadi bentuk VGLM.

Bentuk dasar yang digunakan pada `vglm` yaitu :

```
vglm(formula, family, data, ... )
```

- b. `rcim`, digunakan untuk memodelkan suatu data menjadi bentuk RCIM .

Bentuk dasar yang digunakan pada `rcim` yaitu :

```
rcim(y, family, rank, ... )
```

- c. `logLik`, digunakan untuk menghitung nilai dari *Log-Likelihood* suatu model diantaranya `glm`, `lm`, dll. Bentuk dasar yang digunakan pada `logLik` yaitu :

```
logLik(object, ... )
```

- d. `deviansce`, digunakan untuk menghitung nilai devianss dari suatu model.

Bentuk dasar yang digunakan pada `deviansce` yaitu :

```
deviansce(object, ... )
```

- e. `array`, digunakan untuk membuat array atau vektor. Bentuk dasar yang digunakan pada `array` yaitu :

```
array(data, dim, ... )
```

- f. `svd`, digunakan untuk menghitung nilai dari SVD (*singular value decomposition*). Bentuk dasar yang digunakan pada `svd` yaitu :

```
svd(x, .... )
```

- g. `residuals`, digunakan untuk menghitung nilai residuals dari suatu model.

Bentuk dasar yang digunakan pada `residuals` yaitu :

```
residuals(object)
```

- h. `plot`, digunakan untuk memplotkan suatu data. Bentuk dasar yang digunakan pada `plot` yaitu :

```
plot(x, y, ...)
```

- i. `mean`, digunakan untuk mencari nilai rata – rata dari suatu data. Bentuk dasar yang digunakan pada `mean` yaitu :

`mean(object)`

- j. `pf`, digunakan untuk memodelkan suatu data menjadi bentuk VGLM. Bentuk dasar yang digunakan pada `pf` yaitu :

`pf(q, d1, d2, ... )`

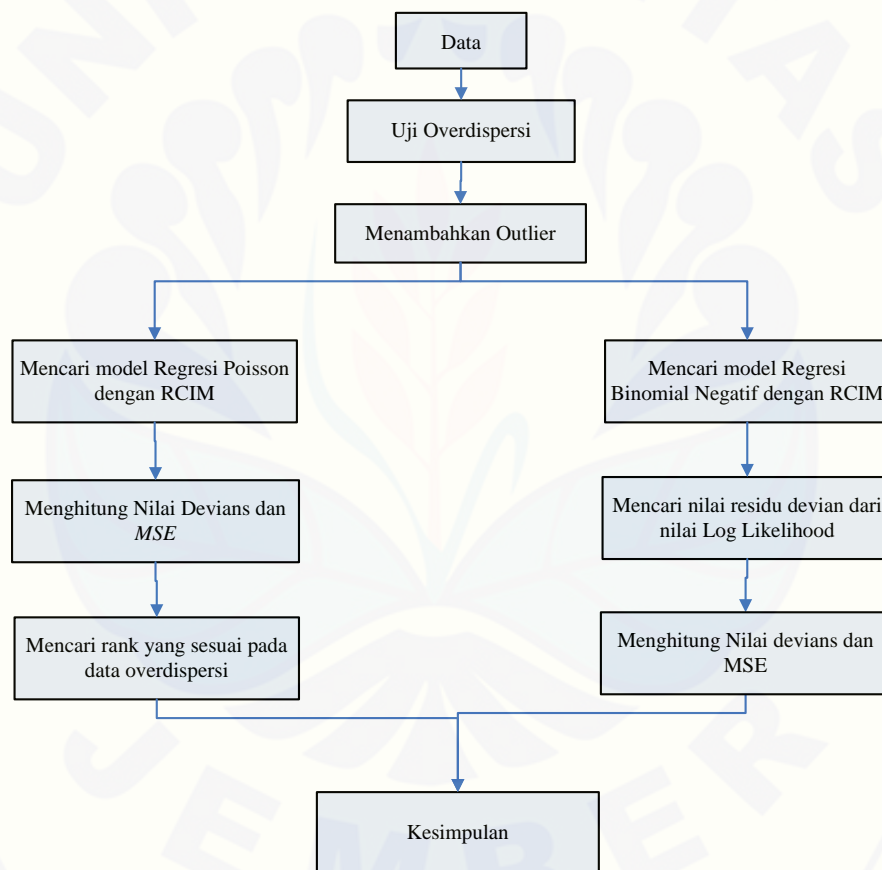
Langkah-langkah yang dilakukan di dalam pengolahan data sekunder untuk mendapatkan pola hubungan terbaik dalam analisis regresi Binomial Negatif untuk mengatasi overdispersi regresi Poisson adalah sebagai berikut.

- a. Menguji data untuk melihat adanya overdispersi pada data dengan menghitung nilai devians yang dibagi dengan db. Nilai lebih dari 1, berarti data mengalami overdispersi.
- b. Memodifikasi data yang berdistribusi Poisson dengan menambahkan *outlier*. Penambahan *outlier* terlebih dahulu dihitung nilai standar deviasi dari masing- masing baris dan kolom yang kemudian dijumlah dengan nilai maksimal pada setiap baris dan kolom. Hal ini dilakukan untuk memberi nilai *outlier* pada data yaitu dengan menambahkan nilai maksimal dengan 4 kali standar deviasi atau juga bisa lebih (Kriegel, 2010).
- c. Memodelkan data menggunakan *Row-Column Interaction Model* dengan fungsi distribusi Poisson. Fungsi yang digunakan adalah fungsi `rcim` dengan `family` yang digunakan pada *software* Rstudio adalah `poissonff`. Model multiplikatif diperoleh dengan menggunakan `Rank = 1,2, ... dst` pada fungsi tersebut.
- d. Memodelkan data menggunakan *Row-Column Interaction Model* dengan fungsi distribusi Binomial Negatif. Fungsi yang digunakan adalah fungsi `rcim` dengan `family` yang digunakan pada *software* Rstudio adalah `negbinomial`. Model multiplikatif diperoleh dengan menggunakan `Rank = 1,2, ... dst` pada fungsi tersebut.
- e. Mencari nilai *residual deviance* dan *deviance* dengan menggunakan nilai *Log-Likelihood* yang didapatkan pada point c dan d.



- f. Melihat rank yang terbaik pada Regresi Poisson yang dapat memodelkan data yang overdispersi.
- g. Melihat nilai *Log-Likelihood*, devians dan MSE pada Regresi Binomial Negatif dalam menangani data overdispersi.
- h. Analisis Biplot pada saat  $\text{Rank} = 0$  dengan memplotkan komponen utama pada matriks interaksi nilai residual dengan metode SVD
- i. Penarikan kesimpulan penanganan overdispersi.

Berikut langkah- langkah yang disajikan dalam gambar 3.1 .



Gambar 3.1 Diagram Alir Metode Penelitian

## BAB 4. HASIL DAN PEMBAHASAN

Overdispersi menyebabkan adanya pelanggaran asumsi pada Regresi Poisson. Simamarta (2010) dalam tugas akhirnya membahas mengenai Regresi Binomial Negatif model aditif, model regresi ini dapat mengatasi overdispersi pada Regresi Poisson karena variabel responnya diasumsikan mengikuti Regresi Binomial Negatif sehingga tidak mengharuskan nilai mean sama dengan nilai varians. Hubungan antara Regresi Poisson dan Regresi Binomial Negatif dapat dilihat pada lampiran E. Penelitian ini membandingkan Regresi Poisson dan Regresi Binomial Negatif model multiplikatif pada data cacahan yang mengalami overdispersi. Model yang memiliki nilai MSE yang paling kecil dianggap sebagai model yang dapat menangani overdispersi dengan baik. Penanganan overdispersi yang digunakan adalah menggunakan pada RCIM yang nantinya juga akan dibahas mengenai suku multiplikatif atau rank model yang menghasilkan model terbaik pada Regresi Binomial Negatif.

### 4.1 Deskripsi Data

Overdispersi dapat disebabkan oleh beberapa hal salah satunya yaitu terdapat *outlier* dalam data. Data yang digunakan pada penelitian ini adalah data Ahadiyah (2015) yang berisi data tentang jumlah polong isi (Lampiran A.1). Data tersebut berisi tanaman kedelai yang menggunakan 15 galur yang ditanam di 8 lokasi yang berbeda. Data tersebut ditunjukkan dengan matriks yang berukuran  $15 \times 8$ .

Data kedua adalah data hasil simulasi dari data Ahadiyah (2015) dengan cara menambahkan *outlier* pada data (Lampiran A.2) untuk meningkatkan nilai dispersi. Penambahan *outlier* pada data menggunakan *Microsoft Excel* untuk memudahkan penghitungan. Tahapan pertama, terlebih dahulu dihitung nilai standar deviasi dari masing- masing baris dan kolom yang kemudian dijumlah dengan nilai maksimal pada setiap baris dan kolom. Secara matematis dituliskan menjadi  $\max + 4\sqrt{stdev}$ .

## 4.2 Uji Overdispersi

Overdispersi pada Data 1 dan Data 2 ditunjukkan pada Tabel 4.1 dimana nilai devians yang dibagi dengan derajat bebasnya bernilai lebih dari 1. Dari nilai tersebut dapat disimpulkan bahwa telah terjadi overdispersi pada data. Data 2 yang mengalami modifikasi pada *outlier* memiliki nilai dispersi yang lebih besar daripada Data 1. Tabel 4.1 menunjukkan bahwa pada Data 1 dari rank= 1 sampai rank= 5 tidak dapat dimodelkan dengan Regresi Poisson tetapi dapat dimodelkan dengan rank= 6 pada. Data 2 menunjukkan peningkatan overdispersi pada setiap rank dibandingkan dengan Data 1. Rank= 1 sampai rank= 6 pada Data 2 tidak dapat dimodelkan dengan Regresi Poisson karena nilai dispersinya lebih dari satu atau mengalami overdispersi.

Tabel 4.1 Hasil Uji Overdispersi

Model	Data 1			Data 2		
	Devians	db	Dispersi	Devians	db	Dispersi
Rank= 1	170,615	20	8,531	201,868	20	10,093
Rank= 2	99,284	18	5,516	119,123	18	6,618
Rank= 3	61,391	16	3,837	69,151	16	4,322
Rank= 4	30,348	14	2,168	35,177	14	2,513
Rank= 5	14,192	12	1,183	16,680	12	1,390
Rank= 6	4,869	10	0,487	5,617	10	0,562

Data 1 merupakan data dengan overdispersi dan Data 2 merupakan data yang mengalami penambahan *outlier*.

Overdispersi pada Data 1 dan Data 2 menyebabkan nilai devians model menjadi sangat besar sehingga model yang dihasilkan menjadi kurang tepat. Hal tersebut dapat diatasi dengan menggunakan Regresi Binomial Negatif. Overdispersi pada Regresi Poisson tersebut akan ditangani dengan Regresi Binomial Negatif dan kompleksitas rank yang sesuai sehingga data tidak mengalami overdispersi.

### 4.3 Analisis Data Menggunakan Regresi Poisson dan Binomial Negatif

Data pada Lampiran A.1 dan Lampiran A.2 yang mengalami overdispersi akan diolah dengan bantuan program R menggunakan paket VGAM untuk memudahkan proses penghitungan. Fungsi yang digunakan pada paket VGAM adalah fungsi `rcim`. Fungsi `rcim` tersebut akan menghasilkan nilai *Log-Likelihood* dari masing-masing model Regresi Poisson dan Regresi Binomial Negatif. Pendugaan parameter pada data akan dilakukan dari rank= 0 hingga rank= 7. Nilai maksimal rank yaitu 7 didapatkan dari mencari nilai minimal baris dan kolom yang kemudian dikurangi satu. Rank= 0 menyatakan model aditif, rank= 1 menyatakan 1 komponen interaksi dan begitu seterusnya hingga rank= 7. Kompleksitas rank akan dilihat dari nilai *Log-Likelihood* dari regresi Poisson yang menghasilkan nilai yang sama dengan Regresi Binomial Negatif. Tabel 4.2 dan Tabel 4.3 merupakan nilai *Log Likelihood* dari masing-masing model pada Data 1 dan Data 2.

Tabel 4.2 Nilai Log-Likelihood Data 1 pada `rcim`

Model	Regresi Poisson	Regresi Binomial Negatif
Null	-2198,594	-631,832
Baris	-691,869	-532,965
Kolom	-2086,342	-628,163
Rank = 0	-547,002	-496,746
Rank = 1	-454,704	-451,278
<b>Rank = 2</b>	<b>-419,038</b>	<b>-419,038</b>
Rank = 3	-400,092	-400,092
Rank = 4	-384,570	-384,570
Rank = 5	-376,492	-376,492
Rank = 6	-371,831	-371,831
Rank = 7	-369,397	-369,397

Tabel 4.3 Nilai Log-Likelihood Data 2 pada rcim

Model	Regresi Poisson	Regresi Binomial Negatif
Null	-2281,542	-634,415
Baris	-739,6529	-540,196
Kolom	-2149,931	-630,240
Rank = 0	-588,291	-508,626
Rank = 1	-471,524	-462,886
Rank = 2	-430,151	-430,076
<b>Rank = 3</b>	<b>-405,165</b>	<b>-405,165</b>
Rank = 4	-388,178	-388,178
Rank = 5	-378,929	-378,930
Rank = 6	-373,398	-373,398
Rank = 7	-370,590	-370,590

Nilai dispersi yang dapat digunakan pada Regresi Poisson menggunakan RCIM adalah dari rank= 2 untuk Data 1 dan rank= 3 pada Data 2 yang dapat dilihat pada Tabel 4.2 ketika nilai *Log-Likelihood* dari Regresi Poisson dan Regresi Binomial Negatif bernilai sama. Nilai *Log-Likelihood* yang sama disebabkan pada rank tersebut Regresi Poisson dianggap dapat menangani overdispersi pada data dan dapat memodelkan data tersebut dengan baik. Semakin besar nilai dispersi dari suatu data, maka rank yang dibutuhkan oleh Regresi Poisson semakin besar untuk memodelkan data tersebut. Berbeda dengan Regresi Poisson, Regresi Binomial Negatif memiliki nilai *Log-Likelihood* yang lebih besar. Nilai *Log-Likelihood* yang lebih besar diartikan bahwa model tersebut lebih baik dalam memodelkan data.

Nilai *Log-Likelihood* yang sudah didapatkan kemudian digunakan untuk menghitung nilai devians. Cara menghitung nilai devians telah dijelaskan pada subbab 2.6.1 yaitu dua kali perbandingan nilai *Log-Likelihood* dari *saturated* model dengan model yang dicari. *Saturated Model* merupakan nilai *Log-Likelihood* pada rank= 7. Nilai devians selanjutnya akan digunakan untuk menghitung nilai residual devians yang akan digunakan untuk analisis devian.

Residual devians pada Regresi Poisson otomatis didapatkan ketika penghitungan menggunakan Rstudio, namun untuk Regresi Binomial Negatif hanya didapatkan nilai *Log-Likelihood* oleh karena itu dibutuhkan penghitungan secara manual untuk mendapatkan nilai residual devians.

Tabel 4.4 Nilai Devians dan Residual Devians Data 1 dengan Pada *rcim*

Model	Poisson		Binomial Negatif	
	Model Devians	Residual Devians	Model Devians	Residual Devians
Null	3658,392	-	524,868	-
Baris	644,942	3013,45	327,134	197,734
Kolom	3433,888	224,504	517,530	7,338
Rank = 0	355,208		254,696	
Rank = 1	170,612	184,596	163,760	90,936
Rank = 2	<b>99,280</b>	71,332	<b>99,280</b>	64,480
Rank = 3	61,388	<b>37,892</b>	61,388	<b>37,892</b>
Rank = 4	30,346	31,042	30,346	31,042
Rank = 5	14,188	16,158	14,188	16,158
Rank = 6	4,866	9,322	4,866	9,322
Rank = 7	0	4,866	0	4,866

Nilai residual devians dari model baris dan kolom didapatkan dari selisih devians Null model dengan devians model baris dan kolom. Menghitung nilai residual devians dari model rank=1 didapatkan dari selisih devians model rank=1 dengan model rank=0 dan untuk nilai residual devians dari model rank=2 didapatkan dari selisih devians model rank=2 dengan model rank=1. Cara yang sama juga dilakukan untuk mencari nilai residual devians dari model rank yang lain sehingga didapatkan masing-masing nilai devians dan residual devians pada Tabel 4.4 dan Tabel 4.5. Sama halnya dengan nilai *Log-Likelihood*, nilai devians juga bernilai sama dari rank= 3 dan residual devians dari rank= 4 pada Data 1 (lihat Tabel 4.4) dan rank= 2 untuk devians dan rank= 3 untuk residual devians pada Data 2 .

Tabel 4.5 Nilai Devians dan Residual Devians Data 2 pada rcim

Model	Poisson		Binomial Negatif	
	Model Devians	Residual Devians	Model Devians	Residual Devians
Null	3821,905	-	527,651	-
Baris	738,126	3083,778	339,212	188,440
Kolom	3558,683	263,222	519,301	8,350
Rank = 0	435,402	-	276,072	-
Rank = 1	201,868	233,534	184,592	91,479
Rank = 2	119,123	82,745	118,973	65,619
Rank = 3	<b>69,151</b>	49,972	<b>69,151</b>	49,822
Rank = 4	35,177	<b>33,974</b>	35,177	<b>33,974</b>
Rank = 5	16,680	18,498	16,680	18,498
Rank = 6	5,616	11,063	5,616	11,063
Rank = 7	0	5,616	0	5,616

Nilai devians yang didapatkan dari masing-masing model Regresi Poisson dan Regresi Binomial Negatif selanjutnya digunakan untuk analisis devians pada setiap data yang dapat dilihat pada Lampiran B. Hasil dari analisis devians tersebut selanjutnya akan digunakan untuk mencari model RCIM yang terbaik pada masing-masing model Regresi Poisson dan Binomial Negatif.

Hasil analisis Regresi Poisson pada data yang mengalami overdispersi memiliki nilai *Log-Likelihood* dan devians yang sama pada rank tertentu. Rank yang sesuai digunakan pada Regresi Poisson untuk data yang mengalami overdispersi pada setiap data berbeda-beda. Perbedaan itu didasarkan pada nilai dispersi dari setiap data. Semakin besar nilai dispersi yang ada, semakin besar rank yang dibutuhkan oleh Regresi Poisson. Nilai *Log-Likelihood* dari Regresi Poisson akan bernilai sama dengan Regresi Binomial Negatif ketika Regresi Poisson sudah bisa memodelkan data overdispersi pada rank tertentu. Tidak hanya nilai *Log-Likelihood* dan devians yang memiliki nilai yang sama, nilai MSE dari masing-masing regresi juga memiliki nilai yang hampir mendekati. Menurut Tabel 4.6, Regresi Binomial Negatif pada RCIM dapat menduga data yang

mengalami overdispersi dengan lebih baik dari Regresi Poisson. Nilai MSE pada Regresi Binomial Negatif yang lebih kecil dapat diartikan bahwa regresi ini lebih baik digunakan daripada Regresi Poisson untuk data cacahan yang mengalami overdispersi.

Tabel 4.6 Nilai MSE pada Data 1 dan Data 2

Data	Model	Poisson	Binomial Negatif
Data 1	RCIM 1	0,021562830	<b>0,020838680</b>
	RCIM 2	0,012077778	<b>0,012077740</b>
	RCIM 3	0,008437611	<b>0,008434989</b>
	RCIM 4	<b>0,005427515</b>	0,005427545
	RCIM 5	0,002552477	<b>0,002552321</b>
	RCIM 6	0,000600769	<b>0,000600608</b>
Data 2	RCIM 1	0,025835524	<b>0,024468820</b>
	RCIM 2	0,014287709	<b>0,014146090</b>
	RCIM 3	0,009145594	<b>0,009145374</b>
	RCIM 4	0,005893773	<b>0,005893766</b>
	RCIM 5	0,002428332	<b>0,002428258</b>
	RCIM 6	0,001028197	<b>0,001028055</b>

#### 4.4 Pengaruh Penambahan *Outlier* pada Nilai Dispersi dan MSE

*Outlier* pada data dapat mempengaruhi nilai dispersi model Regresi Poisson. Subbab ini akan membahas pengaruh outlier pada Regresi Poisson dan Regresi Binomial Negatif. Analisis akan dilakukan pada nilai dispersi dan MSE dari masing-masing regresi. Penambahan outlier pada data dilakukan bertahap dengan menambahkan nilai maksimal dari masing-masing baris dan kolom dengan empat kali standar deviasi hingga *outlier* mencapai 20% dari data. Proses pemilihan baris dan kolom dipilih sesuai dengan nilai yang paling maksimal dari baris dan selanjutnya kolom. Total penambahan *outlier* adalah sebanyak 23 pada data dengan persentase dimulai dari 0,8% hingga 19,2%.

Lampiran C.1 merupakan grafik nilai dispersi pada data dengan adanya penambahan *outlier*. Nilai dispersi pada rank= 1 menunjukkan bahwa pada Regresi Binomial Negatif dapat memberikan nilai dispersi yang lebih kecil dibandingkan Regresi Poisson. Kenaikan persentase *outlier* hingga 20% pada data



juga memberikan dampak terhadap nilai dispersi yang semakin meningkat. Nilai dispersi pada rank= 2 antara Regresi Binomial Negatif dan Regresi Poisson memiliki nilai yang hampir mendekati ketika mengalami penambahan *outlier* dari 0,8% hingga 9,2%. Nilai dispersi pada masing-masing regresi memiliki nilai yang sama untuk seluruh persentase pada rank= 3 hingga rank= 6 yang dapat diartikan Regresi Poisson membutuhkan rank yang lebih besar dibandingkan dengan Regresi Binomial Negatif. Nilai dispersi ketika rank= 6 memiliki nilai yang kurang dari satu atau tidak terjadi overdispersi meskipun telah terjadi penambahan *outlier*. Penambahan rank pada Regresi Poisson dapat menyebabkan nilai dispersi pada data mengalami penurunan sehingga memiliki nilai yang sama dengan Regresi Binomial Negatif.

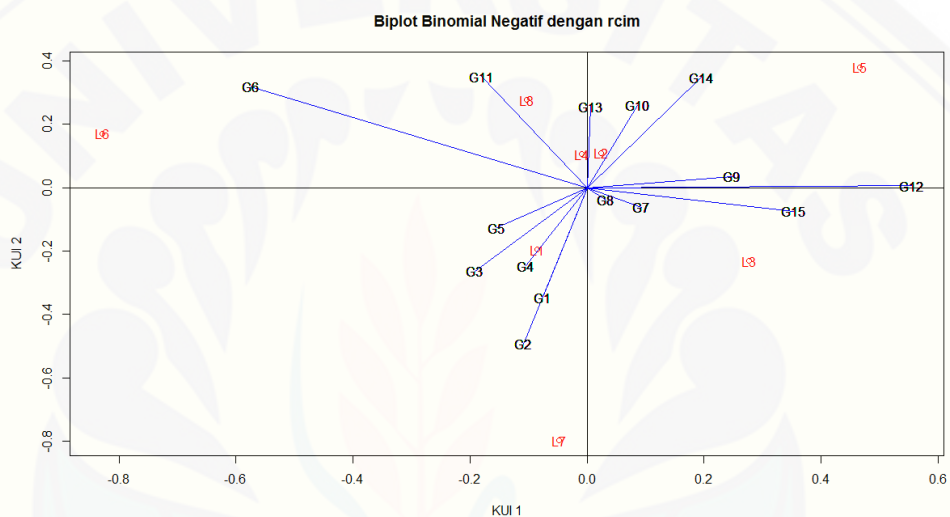
Lampiran C.2 merupakan grafik nilai MSE pada data dengan penambahan *outlier*. Sama halnya dengan nilai dispersi, pada rank= 1 nilai MSE pada Regresi Binomial Negatif memiliki nilai yang lebih kecil daripada Regresi Poisson. Nilai tersebut dapat diartikan bahwa Regresi Binomial Negatif lebih baik dalam memodelkan data yang mengalami overdispersi. Nilai MSE pada rank= 2 memiliki nilai yang hampir mendekati yaitu pada persentase *outlier* dari 0,8% hingga 7,5% pada masing-masing regresi sedangkan pada persentase yang lain masih lebih kecil nilai MSE dari Regresi Binomial Negatif. Nilai MSE pada rank= 3 hingga rank= 6 pada masing-masing regresi memiliki nilai yang hampir mendekati sehingga pada rank tersebut Regresi Poisson dapat memodelkan data sama baiknya dengan Regresi Binomial Negatif.

Grafik pengaruh penambahan *outlier* pada dispersi dan MSE yang ada pada Lampiran C dapat disimpulkan bahwa dengan adanya penambahan rank pada Regresi Poisson dapat menjadikan Regresi Poisson memodelkan data sebaik Regresi Binomial Negatif dengan nilai dispersi yang sama dan nilai MSE yang hampir mendekati meskipun masih lebih baik Regresi Binomial Negatif.

#### **4.5 Interpretasi Biplot pada Regresi Binomial Negatif pada RCIM**

Data 1 dan Data 2 selanjutnya akan diinterpretasikan ke dalam biplot. Tampilan biplot yaitu dua dimensi menyajikan visual objek dan variabel dalam

satu grafik sehingga ciri-ciri peubah dan objek pengamatan dapat dianalisis dengan melihat kemiripan karakteristik antar objek maupun keragaman peubah. Biplot digunakan untuk mengetahui karakteristik lokasi dan genotip pada masing-masing data. Biplot dapat memperlihatkan jarak lokasi dengan genotip serta sudut yang terbentuk dari vektor-vektor variabel. Pendekatan yang digunakan yaitu RCIM pada rank= 0 yaitu bentuk model aditif dari Regresi Binomial Negatif. Hasil biplot untuk Data 1 ditunjukkan pada Gambar 4.1 dengan G melambangkan genotip dan L melambangkan lokasi berikut ini:



Gambar 4.1 Biplot Binomial Negatif Data 1

Genotip yang terletak dekat dengan titik pusat biplot dikategorikan stabil pada semua lokasi. G8 merupakan genotip yang stabil untuk semua lokasi. Empat hal yang bisa didapatkan dari tampilan biplot tersebut sebagai berikut :

a. Kedekatan antar objek

Objek-objek yang memiliki karakteristik yang sama digambarkan dengan titik-titik yang posisinya berdekatan. Menurut Gambar 4.1 objek-objek yang memiliki karakteristik yang sama adalah L2 dan L4.

b. Keragaman variabel peubah

Gambar 4.1 menunjukkan variabel yang memiliki nilai keragaman yang kecil yang digambarkan dengan vektor pendek yaitu G1, G4, G5, G7, G8, G10, G13. Variabel yang memiliki nilai keragaman yang besar adalah G2, G3, G6, G9, G11, G12, G14, G15. Variabel yang memiliki nilai keragaman paling

kecil adalah G8 sedangkan yang memiliki nilai keragaman paling besar adalah G6.

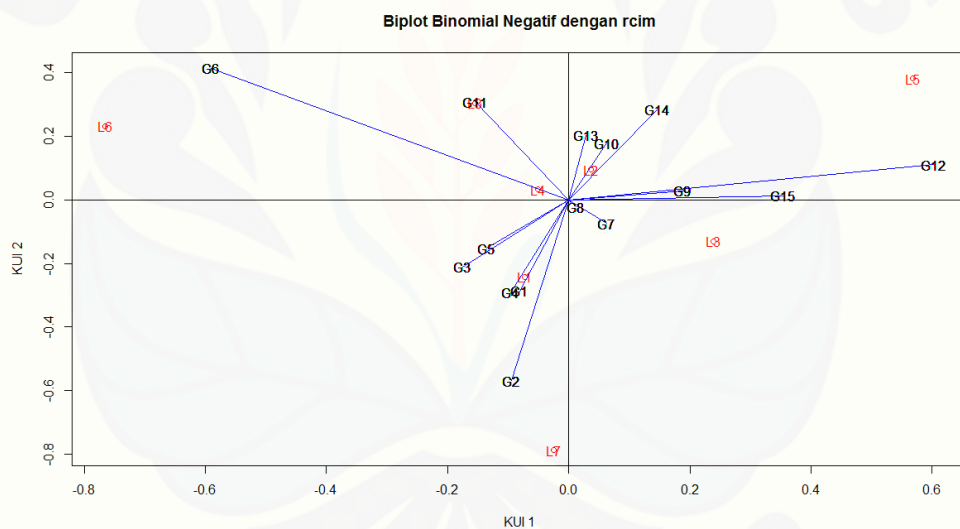
c. Korelasi antar peubah

Dua variabel yang membentuk vektor lancip memiliki korelasi yang positif diantaranya G10-G13 dan G10-G14. Dua Variabel yang membentuk vektor tumpul diartikan memiliki korelasi yang negatif diantaranya G2-G14 dan G10-G3.

d. Nilai peubah pada suatu objek

Objek L2, L4, L1 merupakan objek yang berlawanan dengan arah vektor variabel, sehingga memiliki nilai dibawah rata-rata. Objek L6, L7, L8, L5 searah dengan arah vektor variabel sehingga memiliki nilai diatas rata-rata.

Hasil biplot untuk Data 2 ditunjukkan pada Gambar 4.2 berikut ini:



Gambar 4.2 Biplot Binomial Negatif Data 2

G8 merupakan genotip yang stabil untuk semua lokasi. Sama halnya dengan biplot pada Data 1, biplot Data 2 dapat memperlihatkan empat hal berikut ini :

a. Kedekatan antar objek

Objek-objek yang memiliki karakteristik yang sama adalah L2 dan L4 pada Data 2. Dibandingkan dengan Data 1, L2 dan L4 dari data 2 memiliki karakteristik yang lebih jauh.

b. Keragaman variabel peubah

Variabel yang memiliki nilai keragaman yang kecil yaitu G1, G4, G3, G5, G7, G8, G9, G10, G13, G14. Variabel yang memiliki nilai keragaman yang besar adalah G2, G6, G11, G12, G15. Variabel yang memiliki nilai keragaman paling kecil dan paling besar pada Data 1 dan Data 2 adalah sama yaitu G8 dan G6.

c. Korelasi antar peubah

Dua variabel yang memiliki korelasi yang positif diantaranya G10-G13 dan G8-G7. Dua Variabel yang memiliki korelasi yang negatif diantaranya G6-G7 dan G2-G14.

d. Nilai peubah pada suatu objek

Objek L2, L4, L1 merupakan objek yang berlawanan dengan arah vektor variabel, sehingga memiliki nilai dibawah rata-rata. Objek L6, L7, L8, L5, L5 searah dengan arah vektor variabel sehingga memiliki nilai diatas rata-rata.

## BAB 5. PENUTUP

### 5.1 Kesimpulan

Kesimpulan yang didapatkan dari penanganan overdispersi pada pemodelan data cacahan menggunakan Regresi Binomial Negatif sebagai berikut :

- a. Hasil analisis devians yang didapatkan dari Regresi Poisson dan Regresi Binomial Negatif menunjukkan bahwa Regresi Binomial Negatif dapat menangani data cacahan yang mengalami overdispersi. Nilai devians pada Regresi Binomial Negatif bernilai lebih kecil daripada Regresi Poisson yang menunjukkan bahwa Regresi Binomial Negatif lebih baik. Model RCIM pada data yang mengalami overdispersi untuk Regresi Poisson dan Regresi Binomial Negatif signifikan pada RCIM 2. Model RCIM pada data yang mengalami penambahan *outlier* untuk Regresi Poisson signifikan pada RCIM 3 sedangkan untuk Regresi Binomial Negatif signifikan pada RCIM 1.
- b. Rank yang dibutuhkan Regresi Poisson untuk menangani overdispersi pada data cacahan berbanding lurus dengan nilai dispersi dari data cacahan tersebut. Semakin besar nilai dispersi, maka semakin besar juga rank yang dibutuhkan untuk menangani overdispersi.
- c. Nilai MSE pada Regresi Binomial Negatif lebih kecil daripada Regresi Poisson baik pada data yang mengalami overdispersi maupun pada data yang mengalami penambahan *outlier* diartikan bahwa Regresi Binomial Negatif adalah model yang dapat menangani data cacahan yang overdispersi.

### 5.2 Saran

Saran untuk penelitian selanjutnya dapat dilakukan penelitian tentang *Generalized Poisson* untuk memodelkan data cacahan yang mengalami overdispersi.

**DAFTAR PUSTAKA**

- Agresti, A. 2002. *Categorical Data Analysis Second Edition*. New York : John Wiley & Sons.
- Ahadiyah, K. 2015. “Pendekatan *Row Column Interaction Model* untuk Model *Generalized Ammi* (GAMMI) pada Data Berdistribusi Poisson dengan *Log-Likelihood* dan Analisis Devians”. Tidak Diterbitkan. Skripsi. Jember: FMIPA Universitas Jember
- Casella, G. & Berger, R. L. 1990. *Statistical Inference*. California: Brooks/Cole.
- Gabriel, K. R. 1971. The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. *Biometrics* **58**(3), p. 45
- Hadi A. F. & Mattjik, A. A., dan Sumertajaya, I. M. 2010. Generalized AMMI Models for Assessing The Endurance of Soybean to Leaf Pest. *Jurnal ILMU DASAR* **2** : 151-159.
- Hardin, W. J. & Hilbe, J. M. 2007. *Generalized Linier Models and Extensions*. Texas : Stata press.
- Hilbe, J. M. 2011. *Negative Binomial Regression Second Edition*. New York: Cambridge University Press.
- Jolliffe, I. T. 2002. *Principal Component Analysis, Second Edition*. Berlin: Springer-Verlag.
- Kismiantini. 2008. “Perbandingan Model Regresi Poisson dengan Model Regresi Binomial Negatif”. Tidak Diterbitkan. Makalah Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA. Yogyakarta: Universitas Negeri Yogyakarta
- Kriegel, H. P. 2010. “*Outlier Detection Techniques*”. Munich : University Munchen.
- Mattjik, A. A. & Sumertajaya, I. M. 2011. *Sidik Peubah Ganda dengan Menggunakan SAS*. Departemen Statistika IPB: IPB Press.
- McCullagh, P. & Nelder, J. A. 1989. *Generalized Linier Models, second edition*. London: Chapman & Hall

- Nelder, J. A. & Wedderburn. 1972. Generalized Linear Models. *J.R. Statist. Soc. A*, **370** : p.370
- Osgood. 2000. Poisson-Based Regression Analysis of Aggregate Crime Rates. *Journal of Quantitative Criminology* ,**16** : 21–43.
- Pamungkas, D. H. 2003. Kajian Pengaruh Overdispersi dalam Regresi Poisson. [Skripsi]. Tersedia : <http://repository.ipb.ac.id/pdf>. [2 Maret 2015]
- Pawitan, Y. 2001. *In All Likelihood : Statistical Modelling and Inference Using Likelihood*. Ireland : Clarendon Press. Oxford.
- Simamarta, R. T. & Ispriyanti, D. 2011. “Penanganan Overdispersi Pada Model Regresi Poisson Menggunakan Model Regresi Binomial Negatif”. Tidak Diterbitkan. Jurnal. Semarang: Universitas Diponegoro.
- Van Eeuwijk, F. A. 1995. Multiplicative Interaction in Generalized Linear Models. *Biometrics* **51** : 1017-1032.
- Yee. T. W. & Hastie, T. J. 2003. Reduced-rank vector generalized linear models. *Stat Model* **3**: 15-41.
- Yee, T. W. & Hadi, A. F. 2014. Row–column interaction models, with an R implementation. *Computational Statistics*, **29** (6), 1427-1445.

LAMPIRAN

**A. DATA**

**A.1 Data Jumlah Polong Isi Kedelai**

	L1	L2	L3	L4	L5	L6	L7	L8
galur_1	84	118	85	143	115	44	63	49
galur_2	57	118	85	163	115	49	73	35
galur_3	61	101	82	183	110	54	59	38
galur_4	86	122	53	175	130	43	62	41
galur_5	72	113	48	181	119	44	53	37
galur_6	83	159	72	187	146	105	50	67
galur_7	75	137	70	142	125	30	43	41
galur_8	56	119	75	147	139	38	51	49
galur_9	58	95	87	164	143	26	35	36
galur_10	71	113	63	196	155	35	31	37
galur_11	66	146	72	198	184	67	45	59
galur_12	53	111	79	137	225	15	46	27
galur_13	79	185	89	219	217	59	50	54
galur_14	42	99	57	152	148	25	31	46
galur_15	42	96	68	103	135	16	40	37

**A.2 Data Simulasi Jumlah Polong Isi Kedelai**

	L1	L2	L3	L4	L5	L6	L7	L8
galur_1	84	118	85	161	115	44	63	49
galur_2	57	118	85	183	115	49	87	35
galur_3	61	101	82	203	110	54	59	38
galur_4	97	122	53	196	130	43	62	41
galur_5	72	113	48	202	119	44	53	37
galur_6	83	159	72	208	146	119	50	80
galur_7	75	137	70	162	125	30	43	41
galur_8	56	119	75	167	139	38	51	49
galur_9	58	95	87	186	143	26	35	36
galur_10	71	113	63	219	155	35	31	37
galur_11	66	146	72	221	184	67	45	59
galur_12	53	111	79	153	250	15	46	27
galur_13	79	200	100	245	235	59	50	54
galur_14	42	99	57	152	148	25	31	46
galur_15	42	96	68	103	135	16	40	37



**B. ANALISIS DEVIAN****B.1 Analisis Devian Model Regresi Poisson (Data 1)**

Tabel B.1 Analisis Devian Model Regresi Poisson RCIM 6

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	184,596	9,230	15,166	0,00026	***
RCIM 2	18	71,331	3,963	6,512	0,00544	**
RCIM 3	16	37,892	2,368	3,892	0,02868	*
RCIM 4	14	31,043	2,217	3,644	0,03597	*
RCIM 5	12	16,156	1,346	2,212	0,13302	
RCIM 6	10	9,323	0,932	1,532	0,27886	
Residual	8	4,869	0,609			

Tabel B.2 Analisis Devian Model Regresi Poisson RCIM 5

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	184,596	9,230	11,707	0,000001	***
RCIM 2	18	71,331	3,963	5,026	0,005440	**
RCIM 3	16	37,892	2,368	3,004	0,028679	*
RCIM 4	14	31,043	2,217	2,812	0,035966	*
RCIM 5	12	16,156	1,346	1,708	0,133016	
Residual	18	14,192	0,788			

Tabel B.3 Analisis Devian Model Regresi Poisson RCIM 4

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	184,596	9,230	9,124	6,46E-08	***
RCIM 2	18	71,331	3,963	3,917	0,000480	***
RCIM 3	16	37,892	2,368	2,341	0,021564	*
RCIM 4	14	31,043	2,217	2,192	0,034977	*
Residual	30	30,348	1,012			

Tabel B.4 Analisis Devian Model Regresi Poisson RCIM 3

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	184,596	9,230	9,124	9,54E-08	***
RCIM 2	18	71,331	3,963	3,917	0,002441	**
RCIM 3	16	37,892	2,368	2,341	0,083514	
Residual	44	61,391	1,395			

Tabel B.5 Analisis Devian Model Regresi Poisson RCIM 2

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	184,596	9,230	5,578	1,07E-07	***
RCIM 2	18	71,331	3,963	2,395	0,006056	**
Residual	60	99,284	1,655			

Tabel B.6 Analisis Devian Model Regresi Poisson RCIM 1

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	184,596	9,230	4,220	2,21E-06	***
Residual	78	170,615	2,187			

**B.1 Analisis Devian Model Regresi Binomial Negatif (Data 1)**

Tabel B.7 Analisis Devian Model Regresi Binomial Negatif RCIM 6

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	90,936	4,547	7,471	0,00329	**
RCIM 2	18	64,480	3,582	5,886	0,00760	**
RCIM 3	16	37,892	2,368	3,892	0,02868	*
RCIM 4	14	31,043	2,217	3,644	0,03597	*
RCIM 5	12	16,156	1,346	2,212	0,13302	
RCIM 6	10	9,323	0,932	1,532	0,27886	
Residual	8	4,869	0,609			

Tabel B.8 Analisis Devian Model Regresi Binomial Negatif RCIM 5

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	90,936	4,547	5,767	0,000227	***
RCIM 2	18	64,480	3,582	4,543	0,001205	**
RCIM 3	16	37,892	2,368	3,004	0,013538	*
RCIM 4	14	31,043	2,217	2,812	0,020645	*
RCIM 5	12	16,156	1,346	1,708	0,147823	
Residual	18	14,192	0,788			

Tabel B.9 Analisis Devian Model Regresi Binomial Negatif RCIM 4

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	90,936	4,547	4,495	0,000113	***
RCIM 2	18	64,480	3,582	3,541	0,001093	***
RCIM 3	16	37,892	2,368	2,341	0,021564	*
RCIM 4	14	31,043	2,217	2,192	0,034977	*
Residual	30	30,348	1,012			

Tabel B.10 Analisis Devian Model Regresi Binomial Negatif RCIM 3

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	90,936	4,547	4,495	0,000539	***
RCIM 2	18	64,480	3,582	3,541	0,005550	**
RCIM 3	16	37,892	2,368	2,341	0,083514	
Residual	44	61,391	1,395			

Tabel B.11 Analisis Devian Model Regresi Binomial Negatif RCIM 2

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	90,936	4,547	2,748	0,001333	**
RCIM 2	18	64,480	3,582	2,165	0,013464	**
Residual	60	99,284	1,655			

Tabel B.12 Analisis Devian Model Regresi Binomial Negatif RCIM 1

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	90,936	4,547	2,166	0,008383	**
Residual	78	163,763	2,100			

**B.3 Analisis Devian Model Regresi Poisson (Data 2)**

Tabel B.13 Analisis Devian Model Regresi Poisson RCIM 6

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	233,534	11,677	16,632	0,00018	***
RCIM 2	18	82,745	4,597	6,548	0,00534	**
RCIM 3	16	49,972	3,123	4,449	0,01920	*
RCIM 4	14	33,974	2,427	3,457	0,04173	*
RCIM 5	12	18,498	1,541	2,196	0,13529	
RCIM 6	10	11,063	1,106	1,576	0,26564	
Residual	8	5,617	0,702			

Tabel B.14 Analisis Devian Model Regresi Poisson RCIM 5

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	233,534	11,677	12,601	0,000001	***
RCIM 2	18	82,745	4,597	4,961	0,000697	***
RCIM 3	16	49,972	3,123	3,370	0,007523	**
RCIM 4	14	33,974	2,427	2,619	0,028462	
RCIM 5	12	18,498	1,541	1,663	0,159622	
Residual	18	16,680	0,927			

Tabel B.15 Analisis Devian Model Regresi Poisson RCIM 4

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	233,534	11,677	9,958	2,29E-02	***
RCIM 2	18	82,745	4,597	3,920	0,000477	***
RCIM 3	16	49,972	3,123	2,664	0,009991	**
RCIM 4	14	33,974	2,427	2,070	0,046414	
Residual	30	35,177	1,173			

Tabel B.16 Analisis Devian Model Regresi Poisson RCIM 3

Model	Db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	233,534	11,677	9,958	1,73E-08	***
RCIM 2	18	82,745	4,597	3,920	0,001896	**
RCIM 3	16	49,972	3,123	2,664	0,036721	*
Residual	44	69,151	1,572			

Tabel B.17 Analisis Devian Model Regresi Poisson RCIM 2

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	233,534	11,677	5,881	4,39E-08	***
RCIM 2	18	82,745	4,597	2,315	0,007987	**
Residual	60	119,123	1,985			

Tabel B.18 Analisis Devian Model Regresi Poisson RCIM 1

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	233,534	11,677	4,512	7,27E-07	***
Residual	78	201,868	2,588			

**B.4 Analisis Devian Model Regresi Binomial Negatif (Data 2)**

Tabel B.19 Analisis Devian Model Regresi Binomial Negatif RCIM 6

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	91,479	4,574	6,515	0,00525	**
RCIM 2	18	65,619	3,646	5,193	0,01142	*
RCIM 3	16	49,822	3,114	4,435	0,01920	*
RCIM 4	14	33,974	2,427	3,457	0,04173	*
RCIM 5	12	18,498	1,541	2,196	0,13529	
RCIM 6	10	11,063	1,106	1,576	0,26564	
Residual	8	5,617	0,702			

Tabel B.20 Analisis Devian Model Regresi Binomial Negatif RCIM 5

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	91,479	4,574	4,936	0,000631	***
RCIM 2	18	65,619	3,646	3,934	0,002843	**
RCIM 3	16	49,822	3,114	3,360	0,007643	**
RCIM 4	14	33,974	2,427	2,619	0,028462	*
RCIM 5	12	18,498	1,541	1,663	0,159622	
Residual	18	16,680	0,927			

Tabel B.21 Analisis Devian Model Regresi Binomial Negatif RCIM 4

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	91,479	4,574	3,901	3,98E-04	***
RCIM 2	18	65,619	3,646	3,109	0,002940	**
RCIM 3	16	49,822	3,114	2,656	0,010181	*
RCIM 4	14	33,974	2,427	2,070	0,046414	*
Residual	30	35,177	1,173			

Tabel B.22 Analisis Devian Model Regresi Binomial Negatif RCIM 3

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	91,479	4,574	3,901	1,56E-03	***
RCIM 2	18	65,619	3,646	3,109	0,011794	*
RCIM 3	16	49,822	3,114	2,656	0,037355	*
Residual	44	69,151	1,572			

Tabel B.23 Analisis Devian Model Regresi Binomial Negatif RCIM 2

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	91,479	4,574	2,307	6,72E-03	***
RCIM 2	18	65,619	3,646	1,838	0,040946	*
Residual	60	118,973	1,983			

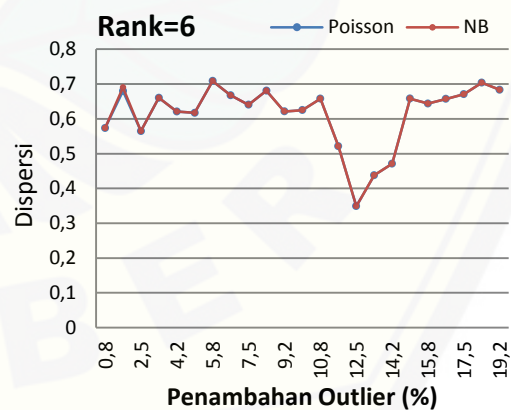
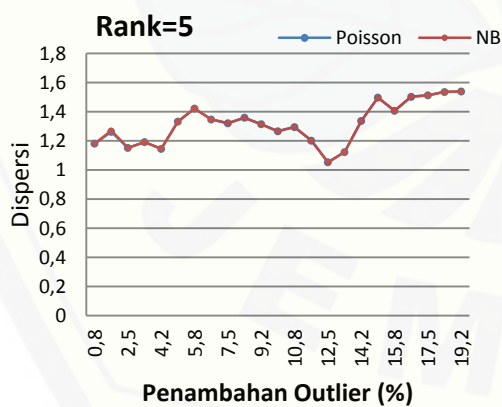
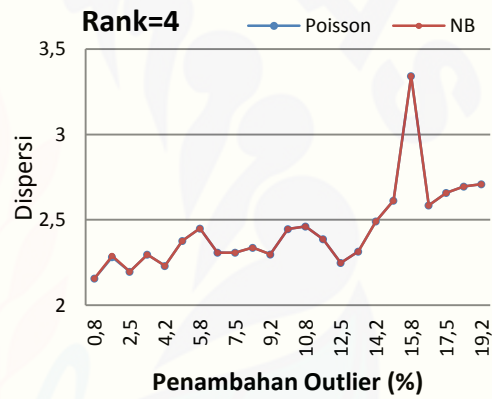
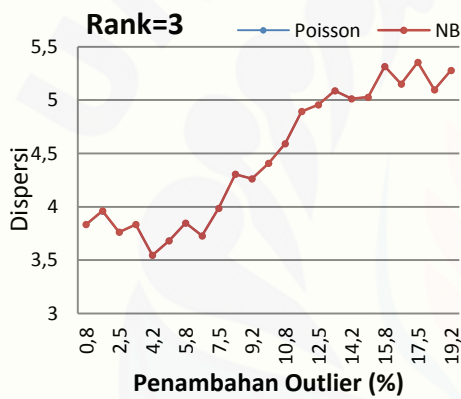
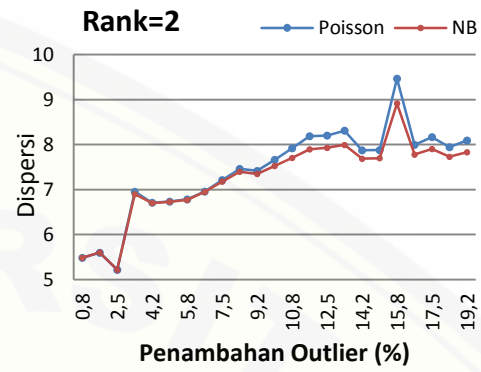
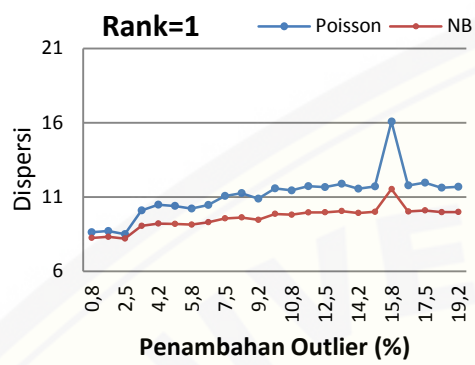
Tabel B.24 Analisis Devian Model Regresi Binomial Negatif RCIM 1

Model	db	Residual Devians	Rataan Devians	Rasio	Nilai P	
RCIM 1	20	91,479	4,574	1,933	2,11E-02	***
Residual	78	184,593	2,367			

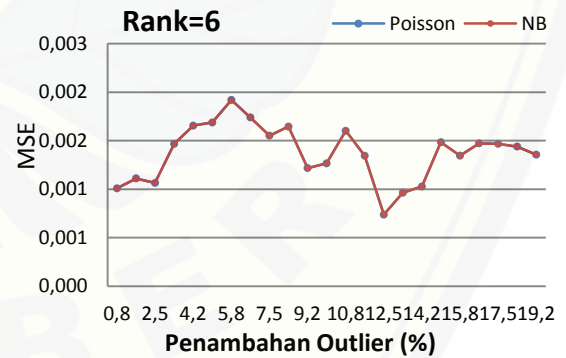
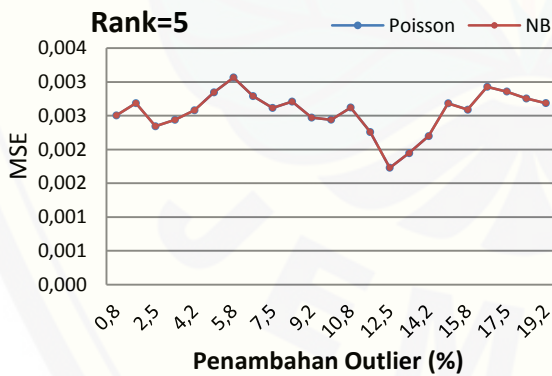
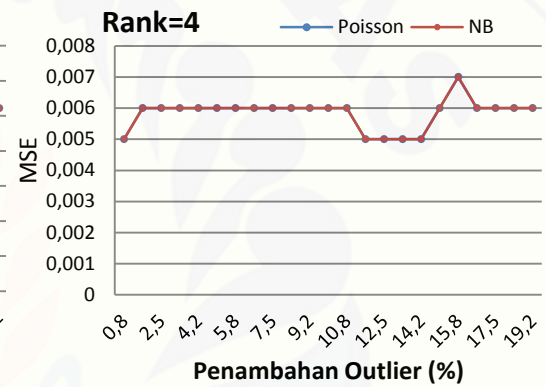
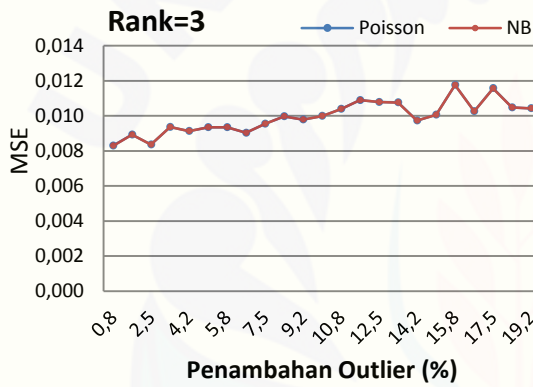
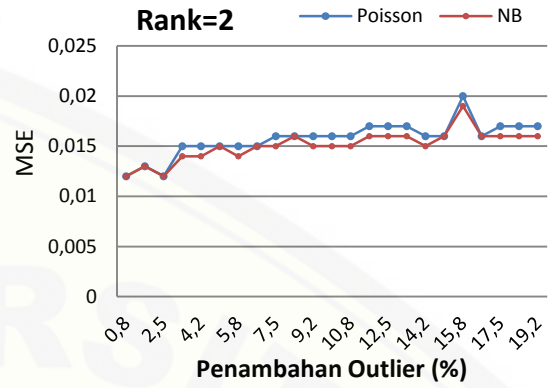
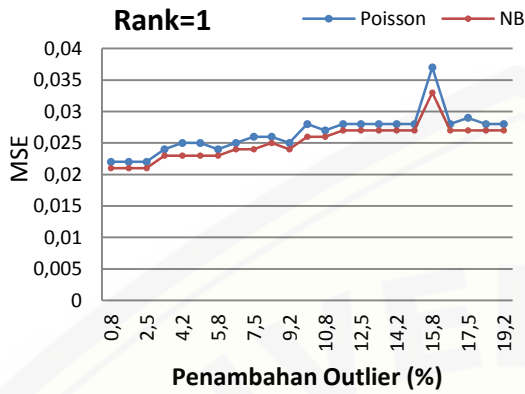


**C. Grafik Pengaruh Penambahan Outlier**

**C.1 Grafik Pengaruh Penambahan Outlier pada Nilai Dispersi**



**C.2 Grafik Pengaruh Penambahan Outlier pada MSE**



## D. SCRIPT PROGRAM R-STUDIO

### D.1 Script Program Regresi Poisson pada RCIM (Data 1)

```

#modelkolom adalah model interaksi dengan genotype
#modelkolom adalah model interaksi dengan lokasi
> library(VGAM)
> modelnull<-
  vglm(JUMLAH_POLONG_ISI~1,poissonff,data=v_polong_isi)
> logLik(modelnull);deviance(modelnull)
> modelkolom<-
  vglm(JUMLAH_POLONG_ISI~GENOTYPE,poissonff,data=v_polong_isi)
> logLik(modelkolom);deviance(modelkolom)
> modelbaris<-
  vglm(JUMLAH_POLONG_ISI~LOKASI,poissonff,data=v_polong_isi)
> logLik(modelbaris);deviance(modelbaris)
> a=rcim(polong_isi, poissonff, Svd.arg=TRUE, Rank=0)
> logLik(a);deviance(a)
> b=rcim(polong_isi, poissonff, Svd.arg=TRUE, Rank=1)
> logLik(b);deviance(b)
> c=rcim(polong_isi, poissonff, Svd.arg=TRUE, Rank=2)
> logLik(c);deviance(c)
> d=rcim(polong_isi, poissonff, Svd.arg=TRUE, Rank=3)
> logLik(d);deviance(d)
> e=rcim(polong_isi, poissonff, Svd.arg=TRUE, Rank=4)
> logLik(e);deviance(e)
> f=rcim(polong_isi, poissonff, Svd.arg=TRUE, Rank=5)
> logLik(f);deviance(f)
> g=rcim(polong_isi, poissonff, Svd.arg=TRUE, Rank=6)
> logLik(g);deviance(g)
> h=rcim(polong_isi, poissonff, Svd.arg=TRUE, Rank=7)
> logLik(h);deviance(h)

#Analisis Devian masing-masing Model
> devh<-deviance(modelnull)-deviance(modelkolom)
> devg<-deviance(modelnull)-deviance(modelbaris)
> dbham<-ncol(polong_isi)-1
> dbgen<-nrow(polong_isi)-1
> t<-min(dbham,dbgen)-1
> dev<-c(1:(t+2))
> for (h in 1:(t+2)){
  i<-(h-1)
  dev[h]<-deviance(rcim(polong_isi, poissonff,
    Svd.arg=TRUE, Rank=i))
}

```

```
> tabel<-array(0,c((1+t),5,t))
> rcim<-paste(c("RCIM"),1:t)
> rcm<-c(1:t)
> for (l in 1:t){
  rcm[l]<-paste(c("RCIM"),t-l+1)
}
> dimnames(tabel)<- list(c(rcim,"Residual"), c("db","Residual
Devians", "Rataan Devians","Ras. Rat. Devians","Nilai-
p"),rcm)
> for(k in 1:t){
  t1<-(1+t)
  t2<-(2+t)
  n<-(t-k)+1
  for (l in 1:n){
    tabel[l,1,k]<-dbham+1+dbgen+1-1-2*1
    tabel[l,2,k]<-dev[l]-dev[l+1]
    tabel[l,3,k]<-tabel[l,2,k]/tabel[l,1,k]
  }
  tabel[t1,1,k]<-dbham*dbgen-sum(tabel[1:n,1,k])
  tabel[t1,2,k]<-dev[n+1]
  tabel[t1,3,k]<- tabel[t1,2,k]/tabel[t1,1,k]
  for (p in 1:n){
    tabel[p,4,k]<-tabel[p,3,k]/tabel[t1,3,k]
    tabel[p,5,k]<-
    pf(tabel[p,4,k],tabel[p,1,k],tabel[t1,1,k],lower
.tail=FALSE)
  }
}
> tabel
```

**D.2 Script Program Regresi Binomial Negatif pada RCIM (Data 1)**

```

#modelkolom adalah model interaksi dengan genotype
#modelkolom adalah model interaksi dengan lokasi
> library(VGAM)
> modelnull<-
  vglm(JUMLAH_POLONG_ISI~1,negbinomial,data=v_polong_isi)
> logLik(modelnull);deviance(modelnull)
> modelkolom<-vglm(JUMLAH_POLONG_ISI~GENOTYPE,
  negbinomial,data=v_polong_isi)
> logLik(modelkolom);deviance(modelkolom)
> modelbaris<-vglm(JUMLAH_POLONG_ISI~LOKASI,
  negbinomial,data=v_polong_isi)
> logLik(modelbaris);deviance(modelbaris)
> a=rcim(polong_isi, negbinomial, Svd.arg=TRUE, Rank=0)
> logLik(a);deviance(a)
> b=rcim(polong_isi, negbinomial, Svd.arg=TRUE, Rank=1)
> logLik(b);deviance(b)
> c=rcim(polong_isi, negbinomial, Svd.arg=TRUE, Rank=2)
> logLik(c);deviance(c)
> d=rcim(polong_isi, negbinomial, Svd.arg=TRUE, Rank=3)
> logLik(d);deviance(d)
> e=rcim(polong_isi, negbinomial, Svd.arg=TRUE, Rank=4)
> logLik(e);deviance(e)
> f=rcim(polong_isi, negbinomial, Svd.arg=TRUE, Rank=5)
> logLik(f);deviance(f)
> g=rcim(polong_isi, negbinomial, Svd.arg=TRUE, Rank=6)
> logLik(g);deviance(g)
> h=rcim(polong_isi, negbinomial, Svd.arg=TRUE, Rank=7)
> logLik(h);deviance(h)

#Analisis Devian masing-masing Model
> devh<-deviance(modelnull)-deviance(modelkolom)
> devg<-deviance(modelnull)-deviance(modelbaris)
> dbham<-ncol(polong_isi)-1
> dbgen<-nrow(polong_isi)-1
> t<-min(dbham,dbgen)-1
> dev<-c(1:(t+2))
> for (h in 1:(t+2)){
  i<-(h-1)
  dev[h]<-deviance(rcim(polong_isi,negbinomial,
  Svd.arg=TRUE, Rank=i))
}
> tabel<-array(0,c((1+t),5,t))
> rcim<-paste(c("RCIM"),1:t)
> rcm<-c(1:t)

```

```
> for (l in 1:t){
  rcm[l]<-paste(c("RCIM"),t-l+1)
}
> dimnames(tabel)<-
  list(c(rcim,"Residual"),
  c("db","Residual Devians","Rataan Devians",
  "Ras. Rat. Devians","Nilai-p"),rcm)
> for(k in 1:t){
  tabel[1:2,1,k]<-c(dbham,dbgen)
  tabel[1:2,2,k]<-c(devh,devg)
  tabel[1:2,3,k]<-c(devh/dbham,devg/dbgen)
  t1<-(4+t)-1
  t2<-(4+t)
  n<-(t-k)+1
  for (l in 1:n){
    tabel[2+l,1,k]<-dbham+1+dbgen+1-1-2*l
    tabel[2+l,2,k]<-dev[l]-dev[l+1]
    tabel[2+l,3,k]<-tabel[2+l,2,k]/tabel[2+l,1,k]
  }
  tabel[t1,1,k]<-dbham*dbgen-sum(tabel[3:(2+n),1,k])
  tabel[t1,2,k]<-dev[n+1]
  tabel[t2,1:2,k]<-
  c(sum(tabel[1:t1,1,k]),sum(tabel[1:t1,2,k]))
  tabel[t1:t2,3,k]<-
  c(tabel[t1,2,k]/tabel[t1,1,k],sum(tabel[1:t1,2,k])/sum
  (tabel [1:t1,1,k]))
  for (p in 1:(n+2)){
    tabel[p,4,k]<-tabel[p,3,k]/tabel[t1,3,k]
    tabel[p,5,k]<-
    pf(tabel[p,4,k],tabel[p,1,k],tabel[t1,1,k],lower
    .tail=FALSE)
  }
}
> tabel
```

**D.3 Script Program Regresi Poisson pada RCIM (Data 2)**

```

#modelkolom adalah model interaksi dengan genotype
#modelkolom adalah model interaksi dengan lokasi
> library(VGAM)
> modelnull<-
  vglm(JUMLAH_POLONG_ISI~1,poissonff,data=v_polong_isi_over3)
> logLik(modelnull);deviance(modelnull)
> modelkolom<-
  vglm(JUMLAH_POLONG_ISI~GENOTYPE,poissonff,data=v_polong_isi_
  over3)
> logLik(modelkolom);deviance(modelkolom)
> modelbaris<-
  vglm(JUMLAH_POLONG_ISI~LOKASI,poissonff,data=v_polong_isi_ov
  er3)
> logLik(modelbaris);deviance(modelbaris)
> a=rcim(polong_isi_over3, poissonff, Svd.arg=TRUE, Rank=0)
> logLik(a);deviance(a)
> b=rcim(polong_isi_over3, poissonff, Svd.arg=TRUE, Rank=1)
> logLik(b);deviance(b)
> c=rcim(polong_isi_over3, poissonff, Svd.arg=TRUE, Rank=2)
> logLik(c);deviance(c)
> d=rcim(polong_isi_over3, poissonff, Svd.arg=TRUE, Rank=3)
> logLik(d);deviance(d)
> e=rcim(polong_isi_over3, poissonff, Svd.arg=TRUE, Rank=4)
> logLik(e);deviance(e)
> f=rcim(polong_isi_over3, poissonff, Svd.arg=TRUE, Rank=5)
> logLik(f);deviance(f)
> g=rcim(polong_isi_over3, poissonff, Svd.arg=TRUE, Rank=6)
> logLik(g);deviance(g)
> h=rcim(polong_isi_over3, poissonff, Svd.arg=TRUE, Rank=7)
> logLik(h);deviance(h)

#Analisis Devian masing-masing Model
> devh<-deviance(modelnull)-deviance(modelkolom)
> devg<-deviance(modelnull)-deviance(modelbaris)
> dbham<-ncol(polong_isi_over3)-1
> dbgen<-nrow(polong_isi_over3)-1
> t<-min(dbham,dbgen)-1
> dev<-c(1:(t+2))
> for (h in 1:(t+2)){
  i<-(h-1)
  dev[h]<-deviance(rcim(polong_isi_over3, poissonff,
  Svd.arg=TRUE, Rank=i))
}
> tabel<-array(0,c((1+t),5,t))

```

```

> rcim<-paste(c("RCIM"),1:t)
> rcm<-c(1:t)
> for (l in 1:t){
      rcm[l]<-paste(c("RCIM"),t-l+1)
    }
> dimnames(tabel)<- list(c(rcim,"Residual"), c("db"," Residual
      Devians", "Rataan Devians","Ras. Rat.
      Devians","Nilai-p"),rcm)

> for(k in 1:t){
  t1<-(1+t)
  t2<-(2+t)
  n<-(t-k)+1
  for (l in 1:n){
    tabel[l,1,k]<-dbham+1+dbgen+1-l-2*l
    tabel[l,2,k]<-dev[l]-dev[l+1]
    tabel[l,3,k]<-tabel[l,2,k]/tabel[l,1,k]
  }
  tabel[t1,1,k]<-dbham*dbgen-sum(tabel[1:n,1,k])
  tabel[t1,2,k]<-dev[n+1]
  tabel[t1,3,k]<- tabel[t1,2,k]/tabel[t1,1,k]
  for (p in 1:n){
    tabel[p,4,k]<-tabel[p,3,k]/tabel[t1,3,k]
    tabel[p,5,k]<-
    pf(tabel[p,4,k],tabel[p,1,k],tabel[t1,1,k],lower
      .tail=FALSE)
  }
}
tabel

```



**D.4 Script Program Regresi Binomial Negatif pada RCIM (Data 2)**

```

#modelkolom adalah model interaksi dengan genotype
#modelkolom adalah model interaksi dengan lokasi
> library(VGAM)
> modelnull<-
  vglm(JUMLAH_POLONG_ISI~1,negbinomial,data=v_polong_isi_over3
  )
> logLik(modelnull);deviance(modelnull)
> modelkolom<-vglm(JUMLAH_POLONG_ISI~GENOTYPE,
  negbinomial,data=v_polong_isi_over3)
> logLik(modelkolom);deviance(modelkolom)
> modelbaris<-vglm(JUMLAH_POLONG_ISI~LOKASI,
  negbinomial,data=v_polong_isi_over3)
> logLik(modelbaris);deviance(modelbaris)
> a=rcim(polong_isi_over3, negbinomial, Svd.arg=TRUE, Rank=0)
> logLik(a);deviance(a)
> b=rcim(polong_isi_over3, negbinomial, Svd.arg=TRUE, Rank=1)
> logLik(b);deviance(b)
> c=rcim(polong_isi_over3, negbinomial, Svd.arg=TRUE, Rank=2)
> logLik(c);deviance(c)
> d=rcim(polong_isi_over3, negbinomial, Svd.arg=TRUE, Rank=3)
> logLik(d);deviance(d)
> e=rcim(polong_isi_over3, negbinomial, Svd.arg=TRUE, Rank=4)
> logLik(e);deviance(e)
> f=rcim(polong_isi_over3, negbinomial, Svd.arg=TRUE, Rank=5)
> logLik(f);deviance(f)
> g=rcim(polong_isi_over3, negbinomial, Svd.arg=TRUE, Rank=6)
> logLik(g);deviance(g)
> h=rcim(polong_isi_over3, negbinomial, Svd.arg=TRUE, Rank=7)
> logLik(h);deviance(h)

#Analisis Devian masing-masing Model
> devh<-deviance(modelnull)-deviance(modelkolom)
> devg<-deviance(modelnull)-deviance(modelbaris)
> dbham<-ncol(polong_isi_over3)-1
> dbgen<-nrow(polong_isi_over3)-1
> t<-min(dbham,dbgen)-1
> dev<-c(1:(t+2))
> for (h in 1:(t+2)){
  i<-(h-1)
  dev[h]<-deviance(rcim(polong_isi_over3,negbinomial,
  Svd.arg=TRUE, Rank=i))
}
> tabel<-array(0,c((1+t),5,t))
> rcim<-paste(c("RCIM"),1:t)

```

```

> rcm<-c(1:t)
> for (l in 1:t){
  rcm[l]<-paste(c("RCIM"),t-l+1)
}
> dimnames(tabel)<-
  list(c(rcim,"Residual"),
  c("db"," Residual Devians","Rataan Devians",
  "Ras. Rat. Devians", "Nilai-p"),rcm)
> for(k in 1:t){
  tabel[1:2,1,k]<-c(dbham,dbgen)
  tabel[1:2,2,k]<-c(devh,devg)
  tabel[1:2,3,k]<-c(devh/dbham,devg/dbgen)
  t1<-(4+t)-1
  t2<-(4+t)
  n<-(t-k)+1
  for (l in 1:n){
    tabel[2+l,1,k]<-dbham+1+dbgen+1-1-2*l
    tabel[2+l,2,k]<-dev[l]-dev[l+1]
    tabel[2+l,3,k]<-tabel[2+l,2,k]/tabel[2+l,1,k]
  }
  tabel[t1,1,k]<-dbham*dbgen-sum(tabel[3:(2+n),1,k])
  tabel[t1,2,k]<-dev[n+1]
  tabel[t2,1:2,k]<-
  c(sum(tabel[1:t1,1,k]),sum(tabel[1:t1,2,k]))
  tabel[t1:t2,3,k]<-
  c(tabel[t1,2,k]/tabel[t1,1,k],sum(tabel[1:t1,2,k])/sum
  (tabel [1:t1,1,k]))
  for (p in 1:(n+2)){
    tabel[p,4,k]<-tabel[p,3,k]/tabel[t1,3,k]
    tabel[p,5,k]<-
    pf(tabel[p,4,k],tabel[p,1,k],tabel[t1,1,k],lower
    .tail=FALSE)
  }
}
> tabel

```

**D.5 Script Biplot Regresi Binomial Negatif pada Data 1**

```
resi<-residuals(a)
residuals<-resi[,c(1,3,5,7,9,11,13,15)]
svd.a<-svd(residuals)
svd.a
aa=svd.a$V[,c(1,2)]## Col
aa
bb=svd.a$U[,c(1,2)] ## Row
bb
cc=rbind(aa,bb)
cc
row.names(aa)=row.names(t(polong_isi))
row.names(aa)
row.names(bb)=row.names(polong_isi)
row.names(bb)
plot(1, type = "n", xlim = range(cc[,1]),
     ylim = range(cc[,2]), xlab = "KUI 1", ylab = "KUI 2",
     main = "Biplot Binomial Negatif dengan rcim")
abline(h = 0)
abline(v = 0)
points(aa[, 1], aa[, 2], type = "p", col = "red")
text(aa[, 1], aa[, 2], labels =
     c("L1", "L2", "L3", "L4", "L5", "L6", "L7", "L8"),
     adj = c(0.5, 0.5), col = "red")
for (i in 1:15) {
  x2 <- bb[, 1]
  y2 <- bb[, 2]
  x21 <- c(0, x2[i])
  y21 <- c(0, y2[i])
  points(x21, y21, type = "l", col = "blue")
  text(bb[, 1], bb[, 2], labels =
       c("G1", "G2", "G3", "G4", "G5", "G6", "G7", "G8", "G9", "G10", "G11", "G
       12", "G13", "G14", "G15"))
}
```

**D.6 Script Biplot Regresi Binomial Negatif pada Data 2**

```
resi<-residuals(a)
residuals<-resi[,c(1,3,5,7,9,11,13,15)]
svd.a<-svd(residuals)
svd.a
aa=svd.a$v[,c(1,2)]## Col
aa
bb=svd.a$u[,c(1,2)] ## Row
bb
cc=rbind(aa,bb)
cc
row.names(aa)=row.names(t(polong_isi_over3))
row.names(aa)
row.names(bb)=row.names(polong_isi_over3)
row.names(bb)
plot(1, type = "n", xlim = range(cc[,1]),
     ylim = range(cc[,2]), xlab = "KUI 1", ylab = "KUI 2",
     main = "Biplot Binomial Negatif dengan rcim")
abline(h = 0)
abline(v = 0)
points(aa[, 1], aa[, 2], type = "p", col = "red")
text(aa[, 1], aa[, 2], labels =
     c("L1", "L2", "L3", "L4", "L5", "L6", "L7", "L8"),
     adj = c(0.5, 0.5), col = "red")
for (i in 1:15) {
  x2 <- bb[, 1]
  y2 <- bb[, 2]
  x21 <- c(0, x2[i])
  y21 <- c(0, y2[i])
  points(x21, y21, type = "l", col = "blue")
  text(bb[, 1], bb[, 2], labels =
       c("G1", "G2", "G3", "G4", "G5", "G6", "G7", "G8", "G9", "G10", "G11", "G
       12", "G13", "G14", "G15"))
}
```

### E. HUBUNGAN REGRESI BINOMIAL NEGATIF DENGAN REGRESI POISSON

Regresi Poisson didasarkan dari distribusi Poisson. Regresi ini merupakan dasar yang digunakan untuk memodelkan data cacahan. Salah satu asumsi yang harus dipenuhi pada Regresi Poisson adalah nilai varians sama dengan rata-ratanya atau dapat dituliskan secara matematis menjadi :

$$Var(Y) = E(Y) = \mu$$

Masalah yang seringkali terjadi pada Regresi Poisson yaitu nilai varians yang lebih besar dari nilai rata-ratanya atau biasa disebut overdispersi. Regresi yang dapat digunakan untuk memodelkan data yang mengalami overdispersi adalah Regresi Binomial Negatif. Bentuk Regresi Binomial Negatif yang dapat digunakan yaitu NB2 yang merupakan bentuk campuran dari distribusi Poisson-Gamma. Varians dari campuran distribusi Poisson-Gamma tersebut adalah  $\mu + \frac{1}{\nu}\mu^2$  dimana  $\mu$  adalah varians dari Poisson dan  $\frac{1}{\nu}\mu^2$  adalah varians dari Gamma,  $\nu$  pada varians tersebut adalah parameter bentuk dari Gamma dan merupakan overdispersi dalam model campuran. Model Binomial Negatif ini akan menjadi Poisson ketika nilai  $\nu$  mendekati tak terhingga. Simbol standar yang digunakan untuk parameter dispersi dari Binomial Negatif adalah  $\alpha$ . Varians dan mean dari bentuk NB2 yaitu :

$$Var(Y) = \mu + \alpha\mu^2$$

$$E(Y) = \mu$$

Menurut Hilbe (2011), Poisson identik dengan Binomial Negatif NB2 ketika parameter dispersi sama dengan 0 yaitu :

$$\begin{aligned} Var(Y) &= \mu + \alpha\mu^2 \\ &= \mu + 0\mu^2 \\ &= \mu \end{aligned}$$

