

## PENAKSIRAN PENJUALAN PRODUK BERDASARKAN PENDEKATAN MODEL REGRESI NEGATIF BINOMIAL

**Reny Rian Marliana**

STMIK Sumedang, Jl. Angkrek Situ No.19 Sumedang 45232, 0261-207395  
e-mail: [renyrianmarliana@gmail.com](mailto:renyrianmarliana@gmail.com)

### Abstrak

Penelitian bertujuan untuk membandingkan hasil penaksiran parameter model regresi binomial negative dengan model Poisson untuk *underreported counts* pada penelitian sebelumnya. Model regresi dibentuk pada data penjualan produk yang mengalami *underreporting counts*, akibat keterlambatan input data ke aplikasi penjualan produk (*sales cycle*). Pada penelitian sebelumnya, model yang digunakan merupakan gabungan antara distribusi Binomial dan distribusi Poisson. Parameter model regresi ditaksir menggunakan pendekatan Bayes dan simulasi Markov Chain Monte Carlo melalui Algoritma *Gibbs Sampling*. Hasil penaksiran menunjukkan adanya perbedaan antara banyaknya penjualan yang dilaporkan dengan banyaknya penjualan produk yang sebenarnya. Besar perbedaan tersebut merupakan banyaknya penjualan produk yang tidak dilaporkan. Pada penelitian lanjutan ini, model yang digunakan adalah Model Regresi Negatif Binomial. Parameter regresi ditaksir menggunakan metode Iterasi Newton Rapson. Hasil penaksiran menunjukkan selisih yang cukup besar dimana model Poisson untuk *underreported counts* lebih robust sesuai dengan komponen musiman yang ada.

**Kata Kunci:** underreported, generalized poisson, negative binomial

### Abstract

The goal of study is to compare the parameters of the negative Binomial regression model and the Poisson Model for *underreported counts* in the previous study. A model is a regression model for the number of product sales that run into underreporting counts, caused by a delay on input process to the product sales applications (called sales cycle). The model used in the previous study is a mixture of the poisson and the binomial distributions developed by Winkelmann (1996). The regression parameters are estimated by a Bayesian approach and Markov Chain Monte Carlo simulation using Gibbs sampling algorithm. The results show the difference between the actual number and the reported number. This difference is the number of unreported product sales. In this study, the model used is the negative binomial regression model. The regression parameters are estimated using Newton Rapson iteration method. The results show a big gap from the previous study. It means that the Poisson Model for *underreported counts* is more robust in accordance with the seasonal components.

**Keywords:** underreported, generalized poisson, negative binomial

### PENDAHULUAN

Dalam dunia bisnis, khususnya perdagangan, mengetahui banyaknya penjualan produk yang dipasarkan sangatlah penting bagi keberlangsungan sebuah perusahaan. Dengan mengetahui banyaknya produk yang terjual, perusahaan dapat menentukan rencana produksi selanjutnya dan berbagai strategi pemasaran lainnya. Berkaitan dengan hal tersebut, keakuratan dan kevalidan data banyaknya penjualan produk yang terjual merupakan faktor utama dalam ketepatan penentuan strategi pemasaran yang harus dilakukan. Keakuratan dan kevalidan data ini, bergantung pada sistem pelaporan yang digunakan oleh perusahaan yang bersangkutan.

Pemasaran produk dari sebuah perusahaan garmen biasanya dilakukan di toko-toko yang menyebar di berbagai kota, sehingga sistem pelaporan banyaknya penjualan produk dilakukan secara bertahap. Masalah yang sering ditemui pada sistem pelaporan ini adalah *misreporting counts*.

*Misreporting counts* dapat terjadi dalam setiap sistem pelaporan. Dalam sebuah model, *misreporting* terjadi jika laporan seorang individu terhadap banyaknya kejadian yang diamati berbeda dari nilai yang sebenarnya (Li *et al* dalam Pararai, 2010). *Misreporting counts* dibedakan menjadi dua yaitu *underreporting counts* dan *overreporting counts*. *Underreporting counts* adalah sebuah masalah yang dapat terjadi saat pengumpulan data, ketika pencacahan sebuah kejadian, untuk beberapa alasan menjadi tidak lengkap (Neubauer, 2011). *Underreported counts* terjadi jika banyaknya kejadian yang diamati, dilaporkan lebih kecil dari banyaknya kejadian yang sebenarnya. Sebaliknya, *overreported counts* terjadi jika banyaknya kejadian yang sedang diamati dilaporkan melebihi banyaknya kejadian yang sebenarnya.

Pada penelitian ini, *underreported counts* terjadi sebagai akibat keterlambatan input data yang dilakukan oleh admin. Keterlambatan ini terjadi secara berulang, sehingga data penjualan produk yang diterima dalam periode waktu tertentu dapat dipastikan dalam kondisi *underreported*. Proses input data merupakan kunci utama dari sebuah sistem pelaporan sebuah aplikasi penjualan. Sistem pelaporan pada penelitian ini dapat dilihat pada Gambar 1.

*Underreporting counts* menyebabkan banyaknya penjualan produk yang terjual, yang terekam dalam aplikasi lebih kecil dibandingkan dengan banyaknya penjualan produk yang sebenarnya. Hal ini merupakan salah satu penyebab terjadinya kekeliruan dalam penentuan rencana produksi dan distribusi produk. Oleh karena itu, untuk dapat mengurangi terjadinya kekeliruan tersebut, perlu dilakukan penaksiran banyaknya penjualan produk yang sebenarnya.

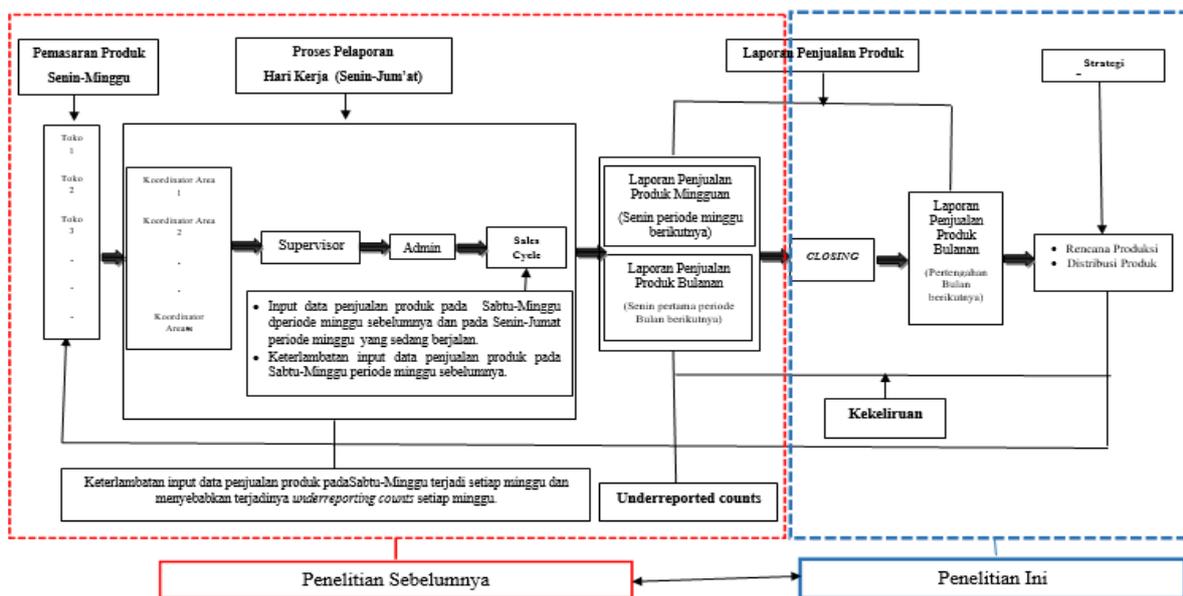
Pada kenyataannya, perusahaan garmen pada penelitian ini telah melakukan sebuah strategi untuk mengatasi terjadinya *underreporting counts*. Strategi tersebut adalah kegiatan *closing* yang bertujuan untuk melengkapi laporan data penjualan produk yang terlambat diinput. Namun, terdapat kekurangan dari strategi ini yaitu waktu pelaksanaan *closing* yang kurang dapat membantu proses penentuan strategi distribusi produk. Waktu pelaksanaan *closing* biasanya dilakukan di awal bulan berikutnya dan selesai di pertengahan bulan tersebut. Sementara, penentuan distribusi produk dilakukan dalam jangka waktu yang sama dengan waktu pelaksanaan *closing*. Dengan demikian, jelas seorang *follow up marketing* akan terpaksa menggunakan data sebelum *closing* dilakukan yang berada dalam kondisi *underreported*

counts, sehingga kekeliruan dalam penentuan strategi distribusi produk tersebut tidak dapat dihindari.

Pada penelitian sebelumnya, untuk dapat mengurangi kekeliruan dalam penentuan strategi pemasaran, baik dalam penentuan strategi distribusi produk maupun rencana produksi selanjutnya, telah dikaji mengenai penaksiran banyaknya penjualan produk yang sebenarnya berdasarkan *underreporting counts*. Penaksiran dilakukan dengan menggunakan sebuah model yang dikembangkan oleh Winkelmann (1999). Model tersebut merupakan gabungan antara distribusi Binomial dan distribusi Poisson. Parameter regresi ditaksir menggunakan pendekatan Bayes dan simulasi Markov Chain Monte Carlo melalui Algoritma *Gibbs Sampling*. Hasil penaksiran menunjukkan dengan jelas adanya perbedaan antara banyaknya penjualan yang dilaporkan dengan banyaknya penjualan produk yang sebenarnya. Besar perbedaan tersebut merupakan banyaknya penjualan produk yang tidak dilaporkan.

Sementara, pada penelitian ini, penaksiran parameter dilakukan berdasarkan laporan data hasil *closing*. Dengan demikian, data yang digunakan dalam penaksiran parameter model bukan lagi merupakan *underreported counts* melainkan *accurately reported*.

Pemodelan dari sebuah data cacahan (*count*) yang dianggap *accurately reported* di antaranya adalah model regresi poisson, model regresi negative binomial dan model regresi *generalized poisson*.



Gambar 1. Sistem Laporan Perusahaan Garmen

## METODE PENELITIAN

### Model Regresi Poisson

Asumsikan bahwa kejadian banyaknya dari satu jenis produk yang terjual di sebuah toko dalam satu bulan, yang diinput ke *sales cycle* dan telah melalui proses *closing* sebagai variabel acak  $Y = \{y_1, y_2, \dots, y_n\}$ . Jelas bahwa variabel acak  $Y$  merupakan data cacah yang mengambil nilai integer  $0, 1, 2, \dots$

Variabel acak  $Y$  dipengaruhi oleh tingkat jual toko yang dinotasikan sebagai variabel  $X$ . Sehingga jika hubungan antara kedua variabel tersebut linear maka dapat dituliskan bahwa :

$$Y_i = \beta_0 + \beta_1 X_i \quad (1)$$

dengan  $i=1, 2, \dots, n$

Bentuk hubungan linear pada persamaan di atas, hanya dapat digunakan jika  $X$  berskala interval atau rasio. Pada penelitian ini variabel independen  $X$  berskala ordinal sehingga jika dimisalkan bahwa  $X$  terdiri atas  $m$  kategori maka untuk dapat memodelkan hubungan antara variabel acak  $Y$  dan  $X$  diperlukan variabel *dummy* ( $D$ ) sebanyak  $m-1$ . Dengan demikian jika hubungan antara kedua variabel tersebut adalah linear maka dapat dituliskan bahwa:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \dots + \beta_{(m-1)} D_{(m-1)i} \quad (2)$$

Persamaan di atas memperlihatkan bahwa hubungan antara variabel acak  $Y$  dan  $X$  tidaklah linear sehingga untuk dapat merumuskan model antara kedua variabel tersebut diperlukan sebuah komponen yang dapat menggambarkan bagaimana variabel  $X$  disubstitusikan ke dalam model sebagai nilai harapan dari variabel acak  $Y$ . Dalam *generalized linear model*, komponen ini dinamakan dengan komponen sistematis. Fungsi yang menghubungkan antara komponen sistematis dengan nilai harapan dari variabel acak  $Y$  dinamakan sebagai *link function* (Dobson, 1983).

Oleh karena nilai rata-rata dari banyaknya produk yang terjual selalu positif maka *link function* yang sesuai dengan kondisi tersebut yang digunakan pada sebuah model regresi Poisson adalah *natural log* ( $\ln$ ) (Agresti (1996) & Cameron (1998)).

Dengan demikian, berdasarkan *Moran's (1952) characterization* dan *Rao-Rubin condition* (1964) (dalam Papadator, 2005), Raftery (1988) (dalam Moreno, 1998) serta Winkelmann (1996), model regresi Poisson antara variabel acak  $Y$  dan  $X$  didefinisikan sebagai :

$$\lambda = \text{Exp}[\mathbf{D}_i' \boldsymbol{\beta}] \quad (3)$$

Dalam regresi Poisson, terdapat asumsi yang harus dipenuhi yaitu nilai rata-rata dan varians dari variabel acak  $Y$  haruslah sama. Asumsi ini disebut sebagai *equidispersion*.

Pada kenyataannya, asumsi ini sangat sulit untuk dipenuhi. Seringnya, data cacah mengalami kasus *overdispersion*. Dengan kata lain nilai varians dari variabel acak  $Y$  lebih besar daripada nilai rata-ratanya.

*Overdispersion* menyebabkan nilai devians model menjadi sangat besar dan menyebabkan model yang dihasilkan menjadi kurang tepat (Utami, 2013). Untuk mengatasi masalah ini, salah satu caranya adalah menggunakan model regresi Binomial Negatif.

### Model Regresi Negatif Binomial

Model regresi Binomial Negatif dibentuk dari sebuah variabel acak berdistribusi Binomial Negatif. Jika diasumsikan kembali bahwa variabel acak  $Y$  yang dipengaruhi oleh variabel independen  $X$  dengan  $m$  kategori berdistribusi Binomial Negatif, maka model regresi Binomial Negatif yang terbentuk didefinisikan sebagai :

$$E[y_i] = \lambda_i = \text{Exp}[\mathbf{D}_i' \boldsymbol{\beta}] \quad (4)$$

dengan  $i=1,2,\dots,n$

Jelas bahwa, model regresi Binomial negatif merupakan model yang dibangun untuk sebagai pemodelan nilai harapan variabel acak  $Y$  sebagai fungsi eksponensial dari sebuah atau beberapa variabel independen  $X$ .

Model regresi Binomial negatif dapat digunakan untuk kasus *equidispersion* maupun *overdispersion*. Masalah lain yang mungkin muncul pada data riil adalah ditemukannya sebuah kasus, dimana nilai varians dari variabel acak  $Y$  lebih kecil dibandingkan dengan nilai rata-rata. Kasus ini dinamakan sebagai *underdispersion*.

Salah satu cara yang bisa digunakan untuk mengatasi kemungkinan terjadinya *underdispersion* adalah model regresi *Generalized Poisson*.

### Model Regresi Generalized Poisson

Model regresi *Generalized Poisson* merupakan perluasan dari model regresi Poisson yang bisa digunakan dalam keadaan *equidispersion*, *overdispersion* maupun *underdispersion*.

Jika diasumsikan kembali bahwa variabel acak  $Y$  yang dipengaruhi oleh variabel independen  $X$  dengan  $m$  kategori berdistribusi *generalized Poisson*, maka fungsi distribusi peluang untuk variabel acak  $Y$  dapat didefinisikan sebagai (Famoye, 2004) :

$$f_i(y_i | \lambda_i, \alpha) = \left( \frac{\lambda_i}{1 + \alpha \lambda_i} \right) \frac{(1 + \alpha y_i)^{y_i - 1}}{y_i!} \exp \left[ \frac{-\lambda_i (1 + \alpha y_i)}{1 + \alpha \lambda_i} \right] \quad (5)$$

dengan  $i=1,2,\dots,n$  dan  $\lambda_i = \text{Exp}[\mathbf{D}_i' \boldsymbol{\beta}]$ .

Rata-rata dan varians dari variabel acak  $Y$  didefinisikan sebagai :

$$E[Y_i | x_i] = \lambda_i \quad (6)$$

dan

$$V[Y_i | x_i] = \lambda_i (1 + \alpha \lambda_i)^2 \quad (7)$$

Jika  $\alpha=0$ , maka akan diperoleh nilai rata-rata dari variabel acak  $Y$  sama dengan variansnya. Dengan demikian model regresi *Generalized Poisson* merupakan model regresi Poisson. Sementara jika  $\alpha>0$ , model regresi *Generalized Poisson* yang terbentuk merupakan pemodelan untuk data cacah yang mengalami *over-dispersion*. Sebaliknya jika nilai  $\alpha<0$ , model regresi *Generalized Poisson* yang terbentuk merupakan pemodelan untuk data cacah yang mengalami *under-dispersion*.

### Data dan Langkah Analisis:

Data yang digunakan pada penelitian ini adalah data laporan penjualan produk di 108 toko dari sebuah perusahaan garmen di Bandung, Indonesia. Laporan penjualan produk yang digunakan adalah data banyaknya penjualan produk yang diinput ke *sales cycle* dan telah melalui proses *closing* periode Agustus 2013. *Count variable* bernilai antara 1-24 *pcs* dengan rata-rata 5,4167 dan standar deviasi 4,5305. Sedangkan, variabel independen yang dilibatkan adalah tingkat jual toko yang terdiri atas 4 kategori yaitu sangat tinggi, tinggi, rendah dan sangat rendah. Oleh karena variabel independen terdiri atas 4 kategori maka pemodelan dilakukan dengan menggunakan 3 buah variabel *dummy*.

Langkah-langkah analisis yang dilakukan sebagai berikut :

1. Penaksiran banyaknya penjualan produk menggunakan model regresi Poisson.
2. Penaksiran banyaknya penjualan produk menggunakan model regresi Negatif Binomial.
3. Penaksiran banyaknya penjualan produk menggunakan model regresi *Generalized Poisson*.
4. Pemilihan model terbaik menggunakan kriteria AIC (*Akaike Information Criterion*).
5. Membandingkan hasil penaksiran model terbaik dengan hasil penaksiran pada penelitian sebelumnya.

### HASIL DAN PEMBAHASAN

Hasil penaksiran parameter regresi menggunakan model regresi Poisson sebagai berikut:

**Tabel 1. Model Regresi Poisson**

| Parameter | Taksiran | Standard Error |
|-----------|----------|----------------|
| $\beta_0$ | 1,252763 | 0,3779         |
| $\beta_1$ | 0,878864 | 0,389          |
| $\beta_2$ | 0,412245 | 0,3824         |
| $\beta_3$ | 0,263584 | 0,3859         |

Hasil penaksiran parameter regresi menggunakan model regresi Negatif Binomial sebagai berikut :

**Tabel 2. Model Regresi Negatif Binomial**

| Parameter | Taksiran | Standard Error |
|-----------|----------|----------------|
| $\beta_0$ | 1,252763 | 0,5677         |
| $\beta_1$ | 0,878864 | 0,5970         |
| $\beta_2$ | 0,412245 | 0,5763         |
| $\beta_3$ | 0,263584 | 0,5817         |

Hasil penaksiran parameter regresi menggunakan model regresi *Generalized Poisson* sebagai berikut :

**Tabel 3. Model Regresi *Generalized Poisson***

| Parameter | Taksiran | Standard Error |
|-----------|----------|----------------|
| $\beta_0$ | 1,3276   | 0,5802         |
| $\beta_1$ | 0,6342   | 0,6054         |
| $\beta_2$ | 0,3595   | 0,5865         |
| $\beta_3$ | 0,2552   | 0,5915         |

Untuk menentukan model mana yang terbaik di antara ketiga hasil penaksiran parameter regresi dari ketiga model regresi tersebut, dilakukan perhitungan nilai *AIC* dimana, model yang memiliki nilai *AIC* yang paling kecil adalah model terbaik. Hasil perhitungan *AIC* untuk masing-masing model adalah sebagai berikut :

**Tabel 4. Akaike Information Criterion**

| Model Regresi              | <i>AIC</i> |
|----------------------------|------------|
| Poisson                    | 677,582    |
| Negatif Binomial           | 577,3987   |
| <i>Generalized Poisson</i> | 579,821    |

Berdasarkan tabel nilai *AIC* di atas, jelas bahwa model regresi Negatif Binomial memiliki nilai *AIC* paling kecil dibandingkan dengan dua model regresi lainnya. Dengan demikian model

regresi Negatif Binomial adalah model terbaik yang dapat digunakan untuk menaksir banyaknya penjualan produk. Dengan demikian persamaan regresi dari model tersebut dapat dituliskan sebagai berikut :

$$\lambda_i = e^{1,2528+0,8789D_1+0,4122D_1+0,2636D_1}$$

Sehingga, berdasarkan kategori tingkat jual toko, taksiran rata-rata penjualan produk periode Agustus 2013 yaitu :

**Tabel 5. Taksiran Rata-Rata**

| Tingkat Jual Toko | Taksiran Rata-Rata (Pcs) |
|-------------------|--------------------------|
| Sangat Tinggi     | 8                        |
| Tinggi            | 5                        |
| Rendah            | 4                        |
| Sangat Rendah     | 3                        |

Hasil penaksiran banyaknya penjualan produk di atas berdasarkan model regresi Negatif Binomial ini sangat jauh berbeda dengan hasil penaksiran pada penelitian sebelumnya yaitu sebagai berikut :

**Tabel 6. Taksiran Penelitian Sebelumnya**

| Tingkat Jual Toko | Taksiran Rata-Rata (Pcs) |
|-------------------|--------------------------|
| Sangat Tinggi     | 21                       |
| Tinggi            | 11                       |
| Rendah            | 9                        |
| Sangat Rendah     | 7                        |

Hasil perbandingan menunjukkan nilai selisih yang cukup besar yaitu rata-rata penjualan produk untuk toko dengan tingkat jual yang sangat tinggi memiliki selisih 13 pcs, toko dengan tingkat jual tinggi memiliki selisih 6 pcs, toko dengan tingkat jual rendah memiliki selisih 5 pcs dan toko dengan tingkat jual toko sangat rendah memiliki selisih 4 pcs.

Nilai selisih yang cukup besar ini membuktikan bahwa proses *closing* yang bertujuan untuk melengkapi banyaknya penjualan produk yang terlambat diinput kurang efektif. Hal ini didukung dengan komponen musiman yang ada. Komponen musiman yang dimaksud adalah periode Agustus merupakan periode menjelang Idul Fitri dimana tingkat penjualan selalu melonjak setiap tahunnya.

Pengalaman menunjukkan bahwa proses *closing* dalam periode menjelang Idul Fitri membutuhkan waktu berbulan-bulan untuk melengkapi banyaknya penjualan produk yang terlambat diinput. Bahkan, terkadang proses *closing* selesai dilakukan menjelang awal tahun berikutnya.

## SIMPULAN DAN SARAN

Berdasarkan hasil perbandingan antara hasil penaksiran banyakan penjualan produk dari hasil penelitian sebelumnya dan hasil penelitian ini, menunjukkan selisih yang cukup besar. Hasil penaksiran pada penelitian sebelumnya lebih robust sesuai dengan komponen musiman yang ada.

Selanjutnya, perlu dikaji kembali mengenai model regresi lain yang bisa mengatasi kemungkinan terjadinya *overdispersi* dan *underdispersi* pada data penjualan produk. Sehingga, bisa kembali dibandingkan model mana yang lebih baik dalam menaksir banyaknya penjualan produk.

## DAFTAR PUSTAKA

- Agresti, Alan (1996). *An Introduction to Categorical Data Analysis*. New York. Jhon Wiley & Sons, Inc.
- Cameron, A. Collin & Pravin K. Trivedi. (1998). *Regression Analysis of Count Data*. New York. Cambridge University Press.
- Dobson, Annette J. (1983). *Introduction to Statistical Modelling*. London. Chapman and Hall, Ltd. London.
- Moreno, Elias & Javier Giron. (1998). *Estimating with Incomplete Count Data, A Bayesian Approach*. Journal of Statistical Planning and Inference 66, 147-159.
- Neubauer, Gerhard. GoardanaDjuras& Herwig Fried. Models for Underreporting: A Bernaulli Sampling Approach for Reported Counts. *Austrian Journal of Statistics*, Volume 40, Number 1 & 2, 85-92. 2011.
- Papadatos, Nickos. (2005). *Characterizations of Discrete Distributions Using The Rao-Rubin Condition*. Journal of Statistical Planning and Inference 135, 222-228.
- Pararai, Marvis. Felix Famoye& Carl Lee. (2010). *Generalized Poisson-Poisson Mixture Model for Misreported Counts with an Application to Smoking Data*. Journal of Data Science 8, 607-617.
- Utami, Tiani Wahyu. (2013). Analisis Regresi Binomial Negatif untuk Mengatasi *Overdispersion* Regresi Poisson pada Kasus Demam Berdarah Dengue. *Statistika*, Vol.1,No.2.
- Winkelmann, Rainer. (1996). *Markov Chain Monte Carlo Analysis of Underreported Count Data with an Application to Worker Absenteeism*. *Empirical Economis* 21: 575-587.