Pendekatan Bayes dalam Model Poisson untuk *Underreported Counts*

Reny Rian Marliana[1], Septiadi Padmadisastra[2], Achmad Zanbar Soleh[3]

Universitas Padjadjaran

Program Pendidikan Magister Program Studi Statistika Terapan, Konsentrasi Statistika Sosial

Email : [1]nhie_1jul6@yahoo.com , [2]s_padmadisastra@yahoo.com ,[3]a_zanbar_s@yahoo.com

Abstrak. Penelitian ini mengkaji mengenai pemodelan pada data penjualan produk yang mengalami *underreporting counts* akibat dari keterlambatan input data ke *sales cycle*. Tujuan analisis adalah untuk menaksir parameter model yaitu banyaknya penjualan produk yang sebenarnya. Model yang digunakan adalah hasil penggabungan antara distribusi poisson dan distribusi binomial yang dikembangkan oleh Winkelmann (1996). Penaksiran parameter model dilakukan melalui pendekatan bayes dan simulasi *Markov Chain Monte Carlo* menggunakan algoritma *gibbs sampling*. Penaksiran parameter model akan bergantung pada penentuan *burn in period* yang dilakukan sejalan dengan pemeriksaan konvergensi algoritma melalui *trace plot, autocorrelation plot* dan *ergodic mean plot*. Analisis yang dilakukan menunjukkan bahwa penentuan *burn in period* lebih mudah dilakukan melalui *ergodic mean plot.* Pada akhirnya taksiran parameter model yang diperoleh adalah rata-rata nilai sampel hasil simulasi yang dihitung dari iterasi setelah *burn in period* sampai dengan iterasi terakhir.

Kata Kunci : *Underreported counts,* Model Poisson, Bayesian, *Gibbs sampling, MCMC*

1. Pendahuluan

Misreporting counts dapat terjadi dalam setiap sistem pelaporan. Li *et al* (dalam Pararai, 2010) menyatakan bahwa dalam sebuah model, *misreporting* terjadi jika laporan seorang individu terhadap banyaknya kejadian yang diamati berbeda dari nilai yang sebenarnya. Dengan demikian, *misreporting counts* terbagi menjadi dua, yaitu *underreporting counts* dan *overreporting counts*. *Underreporting* 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 dari banyaknya kejadian yang sebenarnya.

Pada penelitian yang dilakukan, *underreported counts* terjadi sebagai akibat dari keterlambatan input data penjualan produk yang terjadi secara berulang pada sistem pelaporan yang disebut dengan *sales cycle*. Input data penjualan produk akan secara langsung mengurangi rekap *stock* produk tersebut di toko terkait, sehingga terjadinya *underreporting counts* secara tidak langsung akan menyebabkan terjadinya *overreporting counts* terhadap rekap *stock* produk tersebut di toko terkait. Hal ini merupakan faktor penyebab utama terjadinya kekeliruan dalam rencana produksi selanjutnya dan kekeliruan dalam pendistribusian produk. Untuk dapat mengurangi resiko tersebut perlu dilakukan penaksiran banyaknya penjualan produk yang sebenarnya.

Jelas bahwa, sebagai konsekuensi dari terjadinya *underreporting* adalah penjualan produk yang dilaporkan hanyalah bagian dari penjualan produk yang sebenarnya. Sehingga terdapat penjualan produk yang tidak terlaporkan (terlambat diinput ke *sales cycle*). Hal ini memberikan pengertian bahwa sebuah produk yang terjual di sebuah toko akan mendapat dua kemungkinan perlakuan dari admin yaitu diinput ke *sales cycle* atau tidak. Dalam statistika, besar kemungkinan diinputnya sebuah produk yang terjual di sebuah toko ke *sales cycle* dapat dinyatakan dalam sebuah peluang. Peluang ini memiliki nilai yang berkisar antara 0-1. Sementara diketahui pula bahwa penjualan produk yang sebenarnya di sebuah toko dalam satu bulan adalah acak dan berkisar antara 0-21 *pieces* dengan rata-rata 4 *pieces* per bulan. Dalam statistika, kedua informasi tersebut dapat dianggap sebagai sebuah informasi *prior* yang dapat digunakan dalam penaksiran penjualan produk yang sebenarnya.

Banyaknya penjualan produk (yang sebenarnya) dipengaruhi oleh kegiatan dari penjualan produk itu sendiri. Basu (dalam Sinaga, 2013) menyatakan bahwa faktor- faktor yang mempengaruhi kegiatan penjualan, salah satunya adalah kondisi pasar. Berbedanya kondisi ekonomi penduduk kota, lokasi toko berada, akan memberikan pengaruh terhadap *segmentasi* pasar dan tingkat penjualan produk di toko terkait.

Dengan demikian, untuk dapat menaksir banyaknya penjualan produk yang sebenarnya, diperlukan sebuah analisis yang dapat memodelkan hubungan antara banyaknya produk yang terjual (diperoleh dari *sales cycle*) dengan tingkat jual toko yang melibatkan kedua informasi *prior* sebelumnya.

Selama ini, pemodelan dari sebuah data cacahan (*count*) seringnya, tidak melibatkan adanya kemungkinan *underreported counts* seperti model regresi poisson dan model regresi binomial. Dengan kata lain, kedua model tersebut dapat digunakan jika data cacahan dianggap *accurately reported*. Sedangkan model untuk *underreported counts* pertama kali diperkenalkan dalam *Moran's* (1952) *characterization* dan *Rao-Rubin condition* (1964) (dalam Papadatos, 2005). Kemudian, beberapa peneliti telah mengembangkan model untuk *underreported counts*, diantaranya model regresi poisson untuk *underreported counts* yang dikembangkan oleh Winkelmann (1996), model regresi binomial negatif untuk *underreported counts* yang dikembangkan oleh Mukhopadhyay (dalam Pararai, 2010) dan *generalized poisson-poisson mixture model* yang dapat digunakan untuk *misreporting counts* (*under, over* dan *accurately reported*) dikembangkan Pararai (2010).

Dalam penaksiran parameter model, Pararai (2010) menggunakan pendekatan statistika klasik yang hanya memanfaatkan informasi data sampel melalui metode maksimum *likelihood*. Sementara, Winkelmann (1996) menggunakan pendekatan statistika Bayesian dalam penaksiran parameter model. Penaksiran parameter model dalam pendekatan statistika Bayesian, disamping memanfaatkan informasi data sampel yang direpresentatifkan sebagai fungsi *likelihood*, juga memperhitungkan adanya informasi *prior* yang direpresentatifkan sebagai distribusi *prior*.

Atas dasar inilah, penulis ingin melakukan kajian mengenai pendekatan bayes dalam model *underreported counts* untuk menaksir banyaknya penjualan produk yang sebenarnya.

2. Model Poisson untuk *Underreported Counts*

Model untuk *underreported counts* pertama kali dikenal dalam *Moran's* (1952) *characterization* distribusi poisson (dalam Papadatos, 2005). Moran (1952) (dalam Papadatos, 2005) menyatakan bahwa jika N_1 dan N_2 adalah variabel acak independen *no-degenerate* yang bernilai non-negatif dan jika distribusi bersyarat dari N_1 (N_1 + N_2 = n) adalah binomial dengan parameter $n \in N = \{0, 1, \cdots\}$ dan peluang sukses $p \in [0,1]$ untuk semua $n \in N$, $P[N_1 + N_2 = n] > 0$ maka distribusi N_1, N_2 dan $(N_1 + N_2 = n)$ adalah poisson. Dimana untuk $i \in N$, $P[N_1 = i] > 0$ dan $P[N_2 = i] > 0$.

Bentuk lain dari *Moran's* (1952) *characterization* diperkenalkan oleh Rao & Rubin (1964) (dalam Papadatos, 2005), dimana kebebasan N_1 dan N_2 disebut sebagai *Rao-Rubin condition* yang didefinisikan sebagai :

$$P[N_2 = n_2 | N_1 = 0] = P[N_2 = n_2], n_2 \in N. \qquad (1)$$

Rao & Rubin (1964) (dalam Papadatos, 2005) menyatakan bahwa jika distribusi dari $N = N_1 + N_2$ tidak terkonsentrasi pada *0* dan jika untuk semua $n \in N$ dengan $P[N = n] > 0$,

$$P[N_1 = n_1 | N = n] = \binom{n}{n_1} p^{n_1} (1-p)^{n-n_1}, n_1 = 0, 1, 2, \cdots \qquad (2)$$

untuk $p \in [0,1]$ maka *Rao-Rubin condition* pada persamaan (1) memberikan pengertian bahwa N_1 dan N_2 berdistribusi poisson dengan parameter λp dan $\lambda(1-p)$ dengan $\lambda > 0$.

Dengan kata lain, *Rao-Rubin condition* pada persamaan (1) ekuivalen dengan kebebasan N_1 dan N_2 dibawah *binomial damage model* (Papadatos, 2005). Berkaitan dengan *underreported counts*, *Rao-Rubin condition* menunjukkan N_1 sebagai observasi yang dilaporkan (*undamaged*) sedangkan $N_2 = N - N_1$ sebagai observasi yang tidak terlaporkan (*damaged*) sehingga N_1 dan N_2 adalah bagian dari variabel acak diskrit N. Dengan demikian, jelas bahwa model *underreported counts* merupakan model poisson dengan parameter λp $(\lambda > 0)$.

3. Model Regresi Poisson untuk *Underreported Counts*

Seperti Moran (1952) dan Rao & Rubin (1964) (dalam Papadatos, 2005), model regresi poisson untuk *underreported counts* yang dikembangkan oleh Winkelmann (1996) juga merupakan model hasil penggabungan antara distribusi poisson dan distribusi binomial. Anggap bahwa Y_i^* adalah banyaknya kejadian yang diamati pada waktu tertentu untuk unit pengamatan ke-*i*. Y_i^* diasumsikan bergantung pada variabel independen $\mathbf{x}_i$ dan berdistribusi poisson dengan parameter $\lambda_i = \exp[\mathbf{x}_i'\boldsymbol{\beta}]$.

$$P[Y_i^* | \lambda_i] = \frac{\{\exp[\mathbf{x}_i'\boldsymbol{\beta}]\}^{y_i^*} e^{-\exp[\mathbf{x}_i'\boldsymbol{\beta}]}}{y_i^*!}. \qquad (3)$$

untuk : $Y_i^* = 0, 1, 2, \cdots; i = 1, 2, \cdots, n$

Kemudian, asumsikan bahwa banyaknya kejadian yang diamati pada waktu tertentu untuk unit pengamatan ke-*i* yang dilaporkan adalah Y_i dan berdistribusi binomial dengan parameter Y_i^* dan *p*.

$$P\left(Y_i \middle| Y_i^*, p\right) = \binom{y_i^*}{y_i} p^{y_i} \left(1-p\right)^{y_i^* - y_i}. \tag{4}$$

untuk : $Y_i^* = 0, 1, 2, \cdots; Y_i = 0, 1, 2, \cdots; i = 1, 2, \cdots, n; 0 \le p \le 1$

p = peluang sebuah *kejadian* dilaporkan

Underreported counts terjadi jika $Y_i^* > Y_i$ sehingga fungsi distribusi peluang *marginal* Y_i dapat didefinisikan sebagai:

$$\begin{aligned} P\left(Y_i = y_i\right) &= \sum_{y^* \ge y}^{\infty} P\left(Y_i^* = y_i^* \middle| \lambda_i\right) P\left(Y_i = y_i \middle| y_i^*, p\right) \\ &= \sum_{y_i^* \ge y_i}^{\infty} \frac{\lambda_i^{y_i^*} e^{-\lambda_i}}{y_i^*!} \binom{y_i^*}{y_i} p^{y_i} \left(1-p\right)^{y_i^* - y_i} \\ &= \frac{\left(\lambda_i p\right)^{y_i} e^{-\lambda_i p}}{y_i!}. \end{aligned} \tag{5}$$

dengan : $Y_i = 0, 1, 2, \cdots; \lambda_i \ge 0; i = 1, 2, \cdots, n; 0 \le p \le 1$

p = peluang sebuah *kejadian* dilaporkan

Dengan demikian model regresi untuk *underreported counts* merupakan model poisson dengan parameter $\lambda_i p = \exp\left[\mathbf{x}_i'\boldsymbol{\beta}\right] \cdot p$.

Dalam penaksiran parameter model untuk *underreported counts*, beberapa peneliti menggunakan pendekatan statistika klasik dan beberapa lainnya menggunakan pendekatan statistika Bayesian. Dalam pendekatan statistika klasik, penaksiran parameter sepenuhnya mengandalkan proses inferensi pada data sampel yang diambil dari populasi. Sementara, pendekatan statistika Bayesian, disamping memanfaatkan data sampel yang diperoleh dari populasi, juga memperhitungkan suatu distribusi awal yang disebut dengan distribusi *prior*.

3.1. Pendekatan Statistika Klasik dalam Model Poisson untuk *Underreported Counts*

Winkelmann (1996) menyebutkan bahwa parameter p_i pada model *underreported counts* tidak dapat diperlakukan sebagai parameter *fixed* dikarenakan model yang dihasilkan adalah model singular dengan *n*+*k* parameter dengan *n* adalah ukuran data. Dengan demikian, parameter p_i diasumsikan sebagai variabel acak dan mengikuti sebuah distribusi tertentu. Oleh karena itu, penaksiran akan sulit dilakukan jika menggunakan pendekatan statistika klasik seperti metode maksimum *likelihood*.

Salah satu peneliti yang menggunakan pendekatan statistika klasik dalam penaksiran parameter model *underreported counts* adalah Olkin *et al.* (1981) (dalam Moreno, 1998). Olkin *et al.* (1981) mengunakan metode maksimum *likelihood* dan metode momen dalam menaksir parameter model *underreported counts.* Penaksir yang dihasilkan sangat tidak stabil, dimana sebuah perubahan kecil dalam data menyebabkan perubahan besar pada taksiran parameter model (Moreno, 1998). Dengan demikian diperlukan sebuah pendekatan lain dalam penaksiran parameter model, seperti pendekatan bayes.

3.2. Pendekatan Bayes dalam Model Poisson untuk *Underreported Counts*

Pendekatan statistika Bayesian memandang sebuah parameter dari sebuah model sebagai variabel acak yang memiliki distribusi yang disebut dengan distribusi *prior*. Dari distribusi *prior* selanjutnya dapat ditentukan distribusi *posterior* sehingga diperoleh taksiran Bayesian yang merupakan rata-rata dari distribusi *posterior*.

Berkaitan dengan penaksiran parameter model *underreported counts*, pemilihan distribusi *prior* yang dilakukan peneliti-peneliti adalah berbeda. Drapper & Guttman (1971) (dalam Moreno, 1998) mengasumsikan bahwa parameter N dan p pada persamaan (2) adalah variabel acak independen, dimana N berdistribusi uniform dan p berdistribusi beta. Sementara, Raftery (1988) (dalam Moreno, 1998) mengasumsikan bahwa N berdistribusi poisson dengan parameter λ dan p berdistribusi uniform. Sedangkan Winkelmann (1996) mengasumsikan bahwa N dalam persamaan (2) atau Y^* dalam persamaan (4) berdistribusi poisson dengan parameter $\lambda = \exp\left[\mathbf{x}'\boldsymbol{\beta}\right]$ dimana parameter $\boldsymbol{\beta}$ diasumsikan berdistribusi normal dengan parameter μ dan σ, sedangkan p diasumsikan berdistribusi uniform.

Sehingga dalam sebuah model regresi untuk *underreported counts,* distribusi *posterior* yang diperoleh Winkelmann (1996) didefinisikan sebagai :

$$\underbrace{P\left(y^*, p, \boldsymbol{\beta} \middle| y, \mathbf{x}\right)}_{posterior} \propto \underbrace{P\left(y \middle| y^*, p\right)}_{likelihood} \times \underbrace{P\left(y^* \middle| \boldsymbol{\beta}\right) \times f\left(\boldsymbol{\beta}\right) \times f\left(p\right)}_{prior} \tag{6}$$

Maka diperoleh :

$$\begin{aligned} P\left(Y_i^*, \boldsymbol{\beta}, p_i \middle| Y_i\right) &\propto P\left(Y_i \middle| Y_i^*, p_i\right) P\left(Y_i^* \middle| \boldsymbol{\beta}\right) f\left(\boldsymbol{\beta}\right) f\left(p_i\right) \\ &\propto \exp\left[-\frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}\right)' \Sigma^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}\right)\right] \cdot \prod_{i=1}^{n} \exp\left[\left(y_i^* \cdot \mathbf{x}_i'\boldsymbol{\beta}\right) - \exp\left[\mathbf{x}_i'\boldsymbol{\beta}\right]\right] \\ &\quad \frac{p_i^{y_i}\left(1-p_i\right)^{y_i^*-y_i}}{\left(y_i^*-y_i\right)!\, y_i\,!}. \end{aligned} \tag{7}$$

dengan : $Y_i^* = 0, 1, 2, \cdots; Y_i = 0, 1, 2, \cdots; 0 \le p_i \le 1; -\infty \le \boldsymbol{\beta} \le \infty; i = 1, 2, \cdots, n$

Jelas bahwa sulit untuk menentukan distribusi *posterior* yang diperoleh pada persamaan (7), sehingga untuk memudahkan penaksiran parameter model dilakukan simulasi *Markov Chain Monte Carlo* (MCMC). Untuk itu, perlu ditentukan fungsi peluang bersyarat dari masing-masing parameter sebagai berikut :

$$P\left(Y_i^* \middle| \boldsymbol{\beta}, p, Y_i\right) \propto \frac{\left(1-\hat{p}_i\right)^{y_i^*-y_i} \cdot \exp\left(y_i^* \cdot \mathbf{x}_i'\boldsymbol{\beta}\right)}{\left(y_i^*-y_i\right)!}. \tag{8}$$

$$P\left(p_i \middle| Y_i, Y_i^*\right) \propto p_i^{y_i}\left(1-p_i\right)^{y_i^*-y_i}. \qquad (9)$$

$$P\left(\boldsymbol{\beta} \middle| Y_i^*, Y_i\right) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu})' \Sigma^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})\right)\prod_{i=1}^{n}\exp\left[\left(y_i^* \cdot \mathbf{D}_i'\boldsymbol{\beta}\right)-\exp\left[\mathbf{D}_i'\boldsymbol{\beta}\right]\right]. \qquad (10)$$

dengan : $Y_i^* = 0, 1, 2, \cdots; Y_i = 0, 1, 2, \cdots; 0 \le p_i \le 1; -\infty \le \boldsymbol{\beta} \le \infty; i = 1, 2, \cdots, n$

Menurut Johnson & Kotz (dalam Winkelmann, 1996) fungsi peluang bersyarat parameter Y^* pada persamaan (8) merupakan sebuah *displaced poisson distribution* yaitu :

$$\left(Y_i^* - Y_i\right) \sim Poisson\left(\left(1-p_i\right)\cdot\exp\left(y_i^* \cdot \mathbf{x}_i'\boldsymbol{\beta}\right)\right) \qquad (11)$$

Sementara melalui pendekatan *conjugate* fungsi peluang bersyarat parameter p pada persamaan (9) dapat diasumsikan berdistribusi beta dengan parameter $\left(\left(y_i+1\right),\left(\left(y_i^*-y_i\right)+1\right)\right)$. Namun, *family* distribusi dari fungsi peluang bersyarat parameter $\boldsymbol{\beta}$ tidak diketahui sehingga dalam penaksiran parameter $\boldsymbol{\beta}$ menggunakan algoritma *Random-Walks Metropolis Hastings*.

3.3. *Markov Chain Monte Carlo* (MCMC)

Teknik MCMC dilakukan berdasarkan pada penyusunan rantai markov yang konvergen secara cepat (stasioner) pada distribusi *posterior*. MCMC membangkitkan data sampel parameter θ yang memiliki distribusi tertentu melalui sebuah algoritma dan dilakukan secara iterasi, dimana nilai setiap langkah bergantung pada satu langkah sebelumnya. Salah satu algoritma yang sering digunakan dalam MCMC adalah *gibbs sampling*. *Gibbs sampling* bisa diterapkan apabila distribusi bersyarat dari tiap-tiap parameter diketahui.

Oleh karena fungsi peluang bersyarat dari masing-masing parameter model *underreported counts* diketahui maka penaksiran dapat dilakukan melalui algoritma *gibbs sampling* dengan langkah-langkahnya adalah sebagai berikut :

1. Penentuan nilai awal parameter model $Y^{*(0)}$ $p^{(0)}$ $\boldsymbol{\beta}^{(0)}$
2. Tentukan ukuran iterasi T
3. Lakukan simulasi untuk t = 1, 2, …, T melalui beberapa langkah berikut:
   a. Bangkitkan kandidat nilai baru $Y_i^{*(t)}$ dari $Y_i^* \middle| \boldsymbol{\beta}^{(t-1)}, p_i^{(t-1)}, Y_i$
   b. Bangkitkan kandidat nilai baru $p_i^{(t)}$ dari $p_i \middle| Y_i, Y_i^{*(t)}$
   c. Bangkitkan kandidat baru parameter $\boldsymbol{\beta}^{(t)}$ dari $\boldsymbol{\beta} \middle| Y_i^{*(t)}, Y_i$ melalui algoritma *Random-Walks* dengan $\boldsymbol{\beta}^{(t)} = \boldsymbol{\beta}^{(t-1)} + \left[\tau \mathbf{V}_\beta^{-1/2}\mathbf{z}\right]$ dimana $\mathbf{z} \sim N_D\left(\mathbf{0}, \mathbf{I}\right)$
4. Perbaharui nilai-nilai parameter model dari nilai-nilai hasil simulasi.
5. Periksa konvergensi algoritma melalui *Autocorrelation plot, Trace plot, Ergodic mean plot*
6. Tentukan *burn-in period*
7. Dari nilai-nilai simulasi setelah *burn in* sampai dengan iterasi terakhir hitung :
   a. Nilai rata-rata dan standar deviasi dari parameter $\boldsymbol{\beta}$

b. Nilai rata-rata dan standar deviasi dari parameter p_i untuk $i = 1, 2, \cdots, n$

c. Nilai rata-rata dan standar deviasi dari parameter Y_i^* untuk $i = 1, 2, \cdots, n$

4. Data dan Hasil Penaksiran

Data yang digunakan pada penelitian ini adalah data laporan penjualan produk di 108 toko periode bulan Agustus tahun 2013 milik sebuah perusahaan garmen. *Count variabel* bernilai antara 1-26 dengan rata-rata 5,44 dan standar deviasi 4,57 sedangkan variabel independen adalah tingkat jual toko yang terdiri atas kategori sangat tinggi, tinggi, rendah dan sangat rendah. Oleh karena variabel independen terdiri atas 4 kategori maka pemodelan dilakukan dengan menggunakan 3 buah variabel *dummy*. Simulasi penaksiran parameter dalam model dilakukan dengan menggunakan bantuan *R version* 2.15.2 dengan 5000 iterasi dan algoritma yang telah dijelaskan pada pada bagian 3.3 Pemeriksaan konvergensi dilakukan menggunakan *trace plot, autocorrelation plot* dan *ergodic mean plot*.

Hasil pemeriksaan menunjukkan konvergensi algoritma telah tercapai dengan *burn-in period* 3000 iterasi pertama. Penentuan *burn in period* dilakukan sejalan dengan pemeriksaan konvergensi dan dilakukan melalui *ergodic mean plot*. *Ergodic mean plot* yang dihasilkan menunjukkan bahwa nilai *ergodic mean* dari nilai-nilai sampel $\boldsymbol{\beta}, p$ dan Y^* hasil simulasi mulai stabil setelah 3000 iterasi pertama. Sehingga taksiran parameter $\boldsymbol{\beta}, p$ dan Y^* adalah rata-rata dari nilai-nilai sampel $\boldsymbol{\beta}^{(t)}, p^{(t)}$ dan $Y^{*(t)}$ hasil simulasi yang dihitung dari $t = 3001, \cdots, 5000.$ Hasil penaksiran parameter $\boldsymbol{\beta}$ dapat dilihat pada Tabel 4.1 berikut :

Tabel 4.1 Taksiran Parameter Regresi

Parameter	Taksiran	Stdev
β_0	1,9929	0,4709
β_1	1,0721	0,4792
β_2	0,3876	0,4779
β_3	0,1774	0,4843

Dengan demikian dapat dinyatakan bahwa rata-rata penjualan produk yang sebenarnya periode bulan Agustus 2013 merupakan fungsi dari :

$$\hat{\lambda}_i = E\left[Y_i^* \middle| X_i\right] = e^{\left[1{,}9929 + 1{,}0721 D_{1i} + 0{,}3876 D_{2i} + 0{,}1774 D_{3i}\right]}.$$

Sehingga, berdasarkan kategori tingkat jual toko taksiran rata-rata penjualan produk yang sebenarnya periode Agustus 2013 dapat dilihat pada tabel berikut :

Tabel 4.2 Taksiran Rata-Rata Penjualan Produk Berdasarkan Tingkat Jual Toko

Tingkat Jual Toko	$E\left[Y_i^* \middle| X_i\right]$	**Taksiran Rata-Rata (*pcs*)**
Sangat Tinggi	$E\left[Y_i^* \middle| D_{1i} = 1, D_{2i} = 0, D_{3i} = 0\right] = e^{[1{,}9929 + 1{,}0721]}$	21

Tingkat Jual Toko	$E\left[Y_i^* \middle| X_i\right]$	**Taksiran Rata-Rata (*pcs*)**
Tinggi	$E\left[Y_i^* \middle| D_{1i}=0, D_{2i}=1, D_{3i}=0\right]=e^{[1,9929+0,3876]}$	11
Rendah	$E\left[Y_i^* \middle| D_{1i}=0, D_{2i}=0, D_{3i}=1\right]=e^{[1,9929+0,1774]}$	9
Sangat Rendah	$E\left[Y_i^* \middle| D_{1i}=0, D_{2i}=0, D_{3i}=0\right]=e^{[1,8601]}$	7

Tabel tersebut memberikan makna bahwa secara rata-rata banyaknya penjualan produk periode Agustus 2013 di toko yang memiliki tingkat jual produk yang sangat tinggi adalah 21 *pieces*, di toko yang memiliki tingkat jual produk yang tinggi adalah 11 *pieces*, di toko yang memiliki tingkat jual produk yang rendah adalah 9 *pieces* dan di toko yang memiliki tingkat jual produk yang sangat rendah adalah 7 *pieces*.

Sementara hasil penaksiran penjualan produk yang sebenarnya atau Y^* dan peluang sebuah produk yang terjual dilaporkan (diinput ke *sales cycle*) atau p berdasarkan kategori tingkat jual toko dapat dilihat pada tabel berikut :

Tabel 4.3 Kisaran Taksiran Y^* dan p Berdasarkan Tingkat Jual Toko

Tingkat Jual Toko	$\hat{y}^*$ **(*pcs*)**	$std\left(\hat{y}^*\right)$ **(*pcs*)**	$\hat{p}_i$	$std\left(\hat{p}_i\right)$	**Rata-rata** $\hat{y}_i^* \pm Stdev\left(\hat{y}_i^*\right)$ **(*pcs*)**	**Rata-rata** $\left[\hat{y}_i^* \pm Stdev\left(\hat{y}_i^*\right)\right]-y_i$ **(*pcs*)**
Sangat Tinggi	20-28	3-5	0,0925-0,8929	0,0655-0,1637	17-26	8-18
Tinggi	10-19	1-3	0,1813-0,9091	0,0797-0,1949	8-14	3-8
Rendah	8-15	1-3	0,2254-0,8830	0,0943-0,2601	6-11	2-7
Sangat Rendah	8-9	5-4	0,2634-0,6790	0,2042-0,2058	4-13	1-9

Dengan demikian dapat disimpulkan bahwa secara keseluruhan, rata-rata taksiran penjualan produk yang sebenarnya di toko dengan tingkat jual produk yang sangat tinggi berkisar antara 17-26 *pieces* dengan selisih 8-18 *pieces* dibandingkan dengan penjualan produk yang tercatat pada laporan (diperoleh dari *sales cycle*), di toko dengan tingkat jual produk yang tinggi berkisar antara 8-14 *pieces* dengan selisih 3-8 *pieces*, di toko dengan tingkat jual produk yang rendah berkisar antara 6-11 *pieces* dengan selisih 2-7 *pieces* dan di toko dengan tingkat jual produk yang sangat rendah berkisar antara 4-13 *pieces* dengan selisih 1-9 *pieces*.

5. Saran

Pada penelitian ini, pemodelan didasarkan pada data yang diasumsikan berdistribusi binomial dengan parameter N dan p yang mengikuti sebuah distribusi *prior*. Parameter N diasumsikan mengikuti sebuah distribusi poisson sehingga diasumsikan memiliki rata-rata dan varians yang sama. Namun, pada kenyataannya, parameter tersebut sering memiliki nilai rata-rata dan varians yang berbeda.

Dengan demikian, untuk pengembangan penelitian selanjutnya diharapkan menerapkan model untuk *underreported counts* yang dapat mengatasi kemungkinan terjadinya *overdispersi* atau *underdipersi* dengan kajian pemilihan distribusi *prior* yang baik pula. Dimana, model yang diterapkan tersebut dapat dianalisis secara inferensial dengan menggunakan taraf uji tertentu dan dapat digunakan untuk komponen sistematik yang multivariat.

6. Ucapan Terimakasih

Penelitian ini tidak dapat terselesaikan tanpa bimbingan, bantuan dan dorongan dari berbagai pihak, untuk itu penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada Kementrian Pendidikan dan Kebudayaan, Direktorat Jenderal Pendidikan Tinggi (DIKTI) atas kesempatan dan bantuan dana pendidikan dan penelitian melalui program Beasiswa Unggulan Calon Dosen 2012.

7. Daftar Pustaka

Agresti, Alan. 1996. *An Introduction to Categorical Data Analysis*. John Wiley &Sons, Inc., New York.

Bolstad, William M. 2007. *Introduction to Bayesian Statistics, Second Edition.* John Wiley &Sons, Inc., New Jersey.

Cameron, A. Collin & Pravin K. Trivedi. 1998. *Regression Analysis of Count Data.* Cambrige University Press, New York.

Dobson, Annette J. 1983. *Introduction to Statistical Modelling*. Chapman and Hall, Ltd., London.

Iriawan, Nur. 2006. *Bayesian : Single Parameter*. Institut Teknologi Surabaya, Surabaya.

Irwanti, Lies K. Moch. Abdul M. & Rita R. 2012. Pembangkitan Sampel Random Menggunakan Algoritma Metropolis-Hastings. Jurnal Gaussian, Volume 1, Nomor 1, Halaman 135-146.

Moreno, Elias & Javier Giron. 1998. *Estimating with Incomplete Count Data, A Bayesian Approach*. Journal of Statistical Planning and Inference 66, 147-159.

Neubauer, Gerhard. Goardana Djuras & Herwig Fried. 2011. *Models for Underreporting : A Bernaulli Sampling Approach for Reported Counts.* Austrian Journal of Statistics, Volume 40, Number 1 & 2, 85-92.

Ntzoufras, Ioannis. 2009. *Bayesian Modeling Using WinBUGS*. John Wiley & Sons, Inc. New Jersey.

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.

Scollnik, David P.M. 2006. *A Damage Generalized Poisson Model and Its Application to Reported and Unreported Accident Counts*. Astin Bulletin 36(2): 463-487.

Sinaga, Ulysa I. 2013. Faktor–Faktor yang Mempengaruhi Volume Penjualan Sepeda Motor Honda Vario pada PT. Capella Dinamik Nusantara Pekanbaru. Universitas Negeri Riau, Pekanbaru.

Winkelmann, Rainer. 1996. *Markov Chain Monte Carlo Analysis of Underreported Count Data With an Application to Worker Absenteeism*. Empirical Economis 21: 575-587.