ISBN:1858-1382

J M S K
Jurnal Matematika, Statistika, & Komputasi
http://journal.unhas.ac.id/index.php/jmsk
Vol. 12, No. 1, 36-47, Juli 2015

Penaksiran Parameter Model Kalibrasi Linier yang Berdistribusi Skew-Normal dengan Algoritma-EM

Try Widyaiswara Hairil[1], Anna Islamiyati[1], Raupong[1]

Abstrak

Sebuah penelitian yang ditujukan untuk memperlihatkan perubahan pola pada kurva distribusi normal. Model kalibrasi linier yang berdistribusi skew-normal. Pada tulisan ini, yang dibahas adalah penaksiran parameter model kalibrasi linier yang berdistribusi skew-normal. Pada penaksiran parameter model kalibrasi linier yang berdistribusi skew-normal dengan metode *Maximum Likelihood Estimation* (MLE), tidak diperoleh solusi eksplisit dari taksiran parameter. Oleh karena itu, penaksiran parameternya harus dilakukan dengan metode iterasi numerik algortima-*EM* yang merupakan metode optimisasi iteratif untuk MLE yang terdiri dari dua tahap, yakni tahap ekspektasi dan tahap maksimisasi. Sebuah penerapan empirik kemudian disajikan sebagai ilustrasi dari metode yang dibahas, yaitu pada data pengukuran dimensional dan ultrasonografi pada testis dari 42 remaja.

Kata Kunci: Regresi, model kalibrasi linier, distribusi skew-normal, parameter kecondongan, *MLE*, algoritma-*EM*.

1. Pendahuluan

Analisis regresi merupakan salah satu metode statistika yang sering dijumpai dalam kehidupan sehari-hari. Dalam menggunakan analisis regresi, akan dilihat hubungan antara peubah respon dan satu atau lebih peubah bebas (penjelas). Persamaan regresi yang sering dijumpai adalah regresi linier sederhana dan regresi linier berganda. Kalibrasi statistik sangat penting dalam beberapa bidang pengetahuan manusia. Misalnya dalam bidang kesehatan, instrumen kalibrasi digunakan untuk mengukur tekanan darah, kolesterol, suhu, dan lain-lain. Percobaan kalibrasi biasanya ditandai dengan mengamati dua variabel, X dan Y, melalui fungsi yang telah diketahui. Jadi model persamaan kalibrasi statistik dapat dirumuskan ke suatu model persamaan yang biasa disebut model kalibrasi linier [5].

Distribusi normal merupakan salah satu model distribusi kontinu dalam teori statistika. Distribusi normal diterapkan dalam berbagai permasalahan. Distribusi normal memiliki kurva berbentuk lonceng yang simetris. Adapun bentuk lain dari distribusi normal, salah satu diantaranya yakni distribusi skew-normal. Distribusi skew-normal merupakan kasus khusus dimana jika parameternya bernilai positif maka kurvanya condong ke kanan, dan begitu sebaliknya. Hal yang menjadi masalah pada suatu model kalibrasi linier yakni mengenai penaksiran parameternya. Ada beberapa metode yang dapat digunakan untuk menaksir parameter suatu model kalibrasi linier, salah satu di antaranya yakni *Maximum Likelihood Estimation* (MLE).

Algoritma *Ekspektasi Maksimisasi* (Algoritma-*EM*) merupakan sebuah metode optimasi iteratif untuk MLE yang berguna dalam permasalahan data yang tidak lengkap. Dalam setiap

[1] *Jurusan Matematika FMIPA Universitas Hasanuddin, email : raupong@yahoo.com*

iterasi pada algoritma-*EM* ini terdapat dua tahap, yaitu tahap Ekspektasi atau tahap E (*E-step*) dan tahap Maksimisasi atau tahap M (*M-step*).

Adapun tujuan penulisan ini adalah untuk menaksir parameter model kalibrasi linier yang berdistribusi skew-normal dengan menggunakan algoritma-*EM* serta mengetahui penerapan model kalibrasi linier pada data pengukuran dimensional dan ultrasonografi pada testis dari 42 remaja.

2. Tinjauan Pustaka

2.1. Regresi Linier

Regresi merupakan pengukur hubungan antara dua variabel atau lebih yang dinyatakan dengan bentuk hubungan atau fungsi. Untuk menentukan bentuk hubungan (regresi) diperlukan pemisahan antara variabel bebas yang sering diberi simbol X dan variabel terikat dengan simbol Y. Pada regresi linier sederhana hanya terdapat satu variabel bebas X dan satu variabel terikat Y. Apabila hubungan antara dua variabel atau lebih bersifat kausal, maka variabel yang sebagai sebab merupakan variabel bebas atau variabel X dan akibat yang ditimbulkannya menjadi variabel terikat atau variabel Y.

Model Regresi Linier Sederhana

Bentuk hubungan yang paling sederhana antara variabel X dengan variabel Y adalah berbentuk garis lurus atau berbentuk hubungan linier yang disebut dengan regresi linier sederhana atau sering disebut regresi linier saja dengan persamaan matematikanya sebagaimana yang diungkapkan [11] adalah sebagai berikut:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \tag{1}$$

dimana: Y_i : variabel terikat, X_i : variabel bebas, α : intersep, β : koefisien regresi linier sederhana, ε_i : error atau kesalahan.

Salah satu metode dalam penaksiran parameter adalah *Maximum Likelihood Estimation* (MLE). Prinsip dari MLE adalah menentukan α dan β yang memaksimumkan fungsi *likelihood* dengan disamakan dengan nol. Misalkan Y sampel acak dengan fungsi kepadatan peluang (f.k.p) yaitu $f(Y;\alpha,\beta)$. MLE merupakan penyelesaian dari persamaan:

$$\frac{\partial}{\partial\alpha}\ln[L(\alpha,\beta)] = 0 \quad \text{dan} \quad \frac{\partial}{\partial\beta}\ln[L(\alpha,\beta)] = 0 \tag{2}$$

Jika terdapat peubah acak Y_i dengan $i = 1,2,\dots,n$ yang saling bebas dengan f.k.p dari $f(Y;\alpha,\beta)$ maka fungsi peluang bersama berbentuk sebagai berikut:

$$L(\alpha,\beta) = \prod_{i=1}^{n} f(Y_i;\alpha,\beta) \tag{3}$$

Nilai parameter α dan β dapat diperoleh dengan memaksimumkan f.k.p bersama atau disebut MLE. Hal tersebut dilakukan dengan metode turunan pertama dari fungsi *likelihood*-nya terhadap setiap parameternya sama dengan nol. Selain itu, karena biasanya sulit untuk mencari turunan fungsi *likelihood*, maka yang dilakukan adalah menentukan nilai maksimum dari logaritma natural fungsi *likelihood* tersebut atau disebut dengan fungsi *log-likelihood*. Fungsi *log-likelihood* merupakan f.k.p bersama yang diubah menjadi bentuk logaritma, tujuannya untuk mempermudah di dalam menaksir parameter. Fungsi *log-likehood* dapat ditulis dalam bentuk:

$$l = \ln L(\alpha,\beta) = \sum_{i=1}^{n} \ln\big(f(Y_i;\alpha,\beta)\big) \qquad (4)$$

Langkah-langkah untuk menentukan parameter taksiran dengan MLE yaitu (1) menentukan fungsi *likelihood*, (2) menentukan fungsi *log-likelihood*, dan (3) nilai $\hat{\alpha}$ dan $\hat{\beta}$ diperoleh dari turunan pertama disamakan dengan nol. Menyelesaikan fungsi *log-likelihood* yang diperoleh pada langkah (2) atau (3) dan mendapatkan $\hat{\alpha}$ dan $\hat{\beta}$ sebagai MLE taksirannya.

2.2. Model Kalibrasi

Model kalibrasi linier adalah model linier yang memperhitungkan nilai kalibrasi pada variabel prediktor (X) yang dapat disimbolkan dengan (X_0), sehingga diperoleh model kalibrasi linier sebagai berikut:

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \qquad i = 1, \dots, n; \qquad (5)$$
$$Y_{0j} = \alpha + \beta X_0 + \varepsilon_{0j}, \quad j = 1, \dots, k; \qquad (6)$$

dimana: Y_i : nilai-nilai variabel respon yang bersesuaian dengan X_i, Y_{0j} : nilai-nilai variabel respon yang bersesuaian dengan X_0, X_i : nilai variabel prediktor yang mengikuti pola distribusi normal, X_0 : nilai prediktor dimana mulai terjadi kemencengan (skew), α : konstanta, β : koefisien regresi, ε_i : error ke $-\,i$, ε_{0j} : error ke $-\,j$, i : banyaknya data yang mengikuti distribusi normal, j : banyaknya data yang mengalami kemencengan (skew). $\varepsilon_1, \dots, \varepsilon_n, \varepsilon_{01}, \dots, \varepsilon_{0k}$ berdistribusi identik dan saling bebas (i.i.d.) dengan rataan sama dengan nol, dan variansi sama dengan σ^2. Lalu $X_1, \dots, X_n$ diketahui konstan dan $\alpha, \beta, X_0, \sigma^2$ merupakan parameter yang tidak diketahui. Adapun kelebihan dari model kalibrasi yaitu dapat digunakan pada data yang ketika di plot terjadi kemencengan atau perubahan pola.

Perluasan dari persamaan normal, mempertimbangkan bahwa error dari model kalibrasi normal mengikuti distribusi skew-normal. Dengan asumsi baru, bahwa model kalibrasi linier sebelumnya memiliki empat parameter yakni $\alpha, \beta, X_0,$ dan σ^2, berubah menjadi lima parameter yakni $\alpha, \beta, X_0, \sigma^2,$ dan λ. Pada persamaan (5) dan (6) untuk menentukan nilai-nilai awal dari masing-masing parameternya dapat ditentukan dengan menggunakan rumus persamaan (7) sebagai berikut :

$$\alpha = \frac{\sum_{i=1}^{n} Y_i}{n} - \beta \frac{\sum_{i=1}^{n} X_i}{n} \;;\; \beta = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} {X_i}^2} \;;$$
$$\sigma_* = \sqrt{{\sigma_j}^2} \;;\; X_0 = \frac{\bar{Y}_0 - \hat{\alpha}}{\hat{\beta}} \;;\; \lambda = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)^{\frac{3}{2}}} \,.$$

dimana δ adalah koefisien korelasi antara X_j dan $|U_j|$, $-1 \le \delta \le 1$.

2.3. Distribusi Normal

Distribusi normal merupakan salah satu model distribusi kontinu yang paling penting dalam teori statistika. Dua parameter yang menentukan distribusi normal adalah rataan / ekspektasi (μ) dan standar deviasi (σ). Fungsi kepadatan peluang (f.k.p) dari distribusi normal diberikan rumus berikut:

$$f(X;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(X-\mu)^2\right] , -\infty < X < \infty,\ \mu \in \mathbb{R},\ \sigma^2 > 0 \qquad (8)$$

Distribusi Skew-Normal

Suatu variabel acak X mengikuti distribusi skew-normal standar jika memiliki f.k.p sebagai berikut :

$$f(X;\mu,\sigma^2,\lambda) = \frac{2}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(X-\mu)^2\right\}\Phi\left(\frac{\lambda}{\sigma}(X-\mu)\right) \quad (9)$$

dimana $\phi(\cdot)$ dan $\Phi(\cdot)$ merupakan fungsi kepadatan peluang dan fungsi distribusi komulatif dari distribusi normal standar.

Distribusi Truncated Normal

Misalkan $T\sim TN(\mu;\sigma)I\{a<t<b\}$ merupakan distribusi Truncated Normal yang mempunyai fungsi kepadatan peluang sebagai berikut:

$$f(t;\mu,\sigma^2) = \left\{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)\right\}^{-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\}, a<t<b \quad (10)$$

dimana a dan b merupakan batas atas dan bawah titik potong. Maka, ketika $a=0$ dan $b=\infty$ diperoleh fungsi kepadatan peluang distribusi truncated normal sebagai berikut:

$$f(t;\mu,\sigma^2) = \left\{\Phi\left(\frac{\mu}{\sigma}\right)\right\}^{-1}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\},\ 0<t<\infty \quad (11)$$

Distribusi Half Normal

Misalkan $T\sim HN(\mu;\sigma)I\{a<t<b\}$, ketika $a=\mu$ dan $b=\infty$ diperoleh sebuah fungsi kepadatan peluang suatu distribusi Half Normal sebagai berikut:

$$f(t;\mu,\sigma^2) = \sqrt{\frac{2}{\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(t-\mu)^2\right\},\ \mu<t<\infty \quad (12)$$

2.4. Algoritma-*EM*

Algoritma *Ekspektasi Maksimisasi* (Algoritma-*EM*) merupakan sebuah metode optimisasi iteratif untuk *Maximum Likelihood Estimation* (MLE) yang berguna dalam permasalahan data yang tidak lengkap. Dalam setiap iterasi pada algoritma-*EM* ini terdapat dua tahap, yaitu tahap Ekspektasi atau tahap E (*E-step*) dan tahap Maksimisasi atau tahap M (*M-step*). Langkah-langkah algoritma-*EM* diberikan sebagai berikut:

(1) *E-step* : estimasi statistik cukup (*sufficient statistic*) untuk data lengkap Y_t dengan cara menghitung nilai ekspektasinya.

(2) *M-step* : tentukan $\theta^{(t+1)}$ dengan metode *Maximum Likelihood Estimation* (MLE) dari Y_t.

(3) Iterasi sampai nilai $\theta^{(t)}$ konvergen, atau $\theta^{(t+1)}-\theta^{(t)}$ mendekati nol. Hasilnya merupakan barisan dari nilai-nilai $\theta^{(0)}\rightarrow\theta^{(1)}\rightarrow\cdots$ dimulai dari suatu nilai $\theta^{(0)}$

tertentu. Secara umum, proses iterasi pada algoritma-*EM* merupakan aturan yang berlaku untuk $\theta^{(0)}$

3. Hasil dan Pembahasan

3.1. Penaksiran Parameter Model Kalibrasi Linier

Model kalibrasi linier adalah model linier yang memperhitungkan nilai kalibrasi pada variabel prediktor (X) yang mengalami kemencengan atau berdistribusi skew-normal, menyebabkan data variabel prediktor tersebut terbagi atas dua bagian berdasarkan nilai distribusi frekuensinya. Bagian pertama pada kelompok data yang memiliki distribusi frekuensi berdistribusi normal dan bagian kedua pada kelompok data yang memiliki distribusi frekuensi yang mengalami kemencengan. Kelompok data pertama mengikuti model seperti pada persamaan (5), dan kelompok data kedua mengikuti model seperti pada persamaan (6).

Parameter-parameter yang akan ditaksir dalam model kalibrasi linier yang berdistribusi skew-normal adalah $\alpha, \beta, X_0, \sigma^2$ dan λ dengan metode yang digunakan pada penelitian ini adalah metode *Maximum Likelihood Estimation* (MLE) dan Algoritma-*EM*. Berdasarkan persamaan (5) dan (6) pada bagian sebelumnya, dimana ε_i dan ε_{0j} dari model kalibrasi linier berdistribusi skew-normal, dengan θ merupakan vektor parameter $\alpha, \beta, X_0, \sigma^2$ dan λ. Adapun fungsi kepadatan peluang (f.k.p) dari distribusi skew-normal dapat ditulis sebagai berikut:

$$f(\varepsilon_i;\theta) = \sqrt{\frac{2}{\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(\varepsilon_i)^2\right\}\Phi\left(\frac{\lambda}{\sigma}(\varepsilon_i)\right), i = 1,\ldots,n \tag{13}$$

$$f(\varepsilon_{0j};\theta) = \sqrt{\frac{2}{\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left(\varepsilon_{0j}\right)^2\right\}\Phi\left(\frac{\lambda}{\sigma}\left(\varepsilon_{0j}\right)\right), j = 1,\ldots,k \tag{14}$$

Langkah pertama adalah menentukan f.k.p bersama dari model kalibrasi linier yang berdistribusi skew-normal sebagai berikut:

$$\begin{aligned} f(\varepsilon_i, \varepsilon_{0j};\theta) = \prod_{i=1}^{n}\left[\sqrt{\frac{2}{\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(Y_i-\alpha-\beta X_i)^2\right\}\right]\prod_{\mathrm{i}=1}^{\mathrm{n}}\Phi\left(\frac{\lambda}{\sigma}(Y_i-\alpha-\beta X_i)\right)\times \\ \prod_{j=1}^{k}\left[\sqrt{\frac{2}{\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left(Y_{0j}-\alpha-\beta X_0\right)^2\right\}\right]\prod_{\mathrm{j}=1}^{\mathrm{k}}\Phi\left(\frac{\lambda}{\sigma}\left(Y_{0j}-\alpha-\beta X_0\right)\right) \end{aligned} \tag{15}$$

dimana $i = 1,\ldots,n$ dan $\mathrm{j} = 1,\ldots,\mathrm{k}$. Fungsi *likelihood* (L) dari f.k.p bersama model kalibrasi linier yang berdistribusi skew-normal sebagai berikut:

$$\begin{aligned} L(\theta;\varepsilon_i,\varepsilon_{0j}) = \left(\sqrt{\frac{2}{\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\alpha-\beta X_i)^2\right\}\prod_{i=1}^{n}\Phi\left(\frac{\lambda}{\sigma}(Y_i-\alpha-\beta X_i)\right)\times \\ \left(\sqrt{\frac{2}{\pi\sigma^2}}\right)^k \exp\left\{-\frac{1}{2\sigma^2}\sum_{j=1}^{k}\left(Y_{0j}-\alpha-\beta X_0\right)^2\right\}\prod_{j=1}^{k}\Phi\left(\frac{\lambda}{\sigma}\left(Y_{0j}-\alpha-\beta X_0\right)\right) \end{aligned} \tag{16}$$

sehingga fungsi *log-likelihood* (l) sebagai berikut:

$$l = \frac{n}{2}\ln 2 - \frac{n}{2}\ln(\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{k}(Y_i - \alpha - \beta X_i)^2 + \sum_{\mathrm{i=1}}^{\mathrm{n}}\left[\ln\left\{\Phi\left(\frac{\lambda}{\sigma}(Y_i - \alpha - \beta X_i)\right)\right\}\right] + \frac{k}{2}\ln 2 - \frac{k}{2}\ln(\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{k}\left(Y_{0j} - \alpha - \beta X_0\right)^2 + \sum_{\mathrm{j=1}}^{\mathrm{k}}\left[\ln\left\{\Phi\left(\frac{\lambda}{\sigma}\left(Y_{0j} - \alpha - \beta X_0\right)\right)\right\}\right] \quad (17)$$

Langkah selanjutnya, persamaan (17) diturunkan terhadap parameter $\alpha, \beta, X_0, \sigma, \lambda$ dan disamakan dengan nol dengan cara sebagai berikut: $\frac{\partial l}{\partial \alpha} = 0$, $\frac{\partial l}{\partial \beta} = 0$, $\frac{\partial l}{\partial X_0} = 0$, $\frac{\partial l}{\partial \sigma} = 0$, $\frac{\partial l}{\partial \lambda} = 0$. Namun karena tidak mudah memperoleh solusi eksplisit karena parameter-parameternya masih saling terkait satu sama lain, maka akan dilakukan proses maksimisasi numerik. Kesulitan dalam proses estimasi juga dipaparkan oleh Azzalini [2] dalam model kalibrasi linear. Ada beberapa metode maksimisasi numerik, seperti Newton Raphson dan Fisher, tetapi dalam proses estimasi parameter model kalibrasi terdapat kesulitan dalam mengestimasi parameter seperti yang dipaparkan oleh Liseo dan Loperfido [8]. Sehingga Sartori, memberikan suatu metode estimasi dua tahap untuk menyelesaikan masalah tersebut melalui Algoritma-*EM*.

Sebelum memulai menaksir parameter model kalibrasi linier dengan Algoritma-*EM*, terlebih dahulu mengingat beberapa hal yang berkaitan dengan distribusi truncated normal. Selanjutnya, nilai ekspektasi dari T dan T^2 pada persamaan (11) diperoleh sebagai berikut:

$$E(T) = \mu + \sigma\frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)} \qquad \text{dan} \qquad E(T^2) = \mu^2 + \sigma^2 + \mu\sigma\frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)} \quad (18)$$

Berdasarkan jurnal Henze [6] dan mempertimbangkan nilai $U_i, U_{0j}, V_i,$ dan V_{0j} yang merupakan variabel acak yang berdistribusi identik dan saling bebas $N(0,1)$, sehingga diperoleh nilai $\varepsilon_i = \sigma\delta|U_i| + \sqrt{(1-\delta^2)}V_i\sigma \sim SN(0;\ \sigma^2;\ \lambda)$, $i = 1, \dots, n$ dan $\varepsilon_{0j} = \sigma\delta\left|U_{0j}\right| + \sqrt{(1-\delta^2)}V_{0j}\sigma \sim SN(0;\ \sigma^2;\ \lambda), j = 1, \dots, k$. Berdasarkan nilai ε_i dan ε_{0j} di atas, model kalibrasi linier pada persamaan (5) dan (6) dapat ditulis kembali sebagai berikut:

$$Y_i = \alpha + \beta X_i + \sigma\delta|U_i| + \sqrt{(1-\delta^2)}V_i\sigma, \qquad i = 1, \dots, n \quad (19)$$
$$Y_{0j} = \alpha + \beta X_0 + \sigma\delta\left|U_{0j}\right| + \sqrt{(1-\delta^2)}V_{0j}\sigma, \quad j = 1, \dots, k \quad (20)$$

dengan mengganti nilai $|U_i| = t_i$, $\left|U_{0j}\right| = t_{0j}$, $\sigma\sqrt{(1-\delta^2)}V_i = r_i$, dan $\sigma\sqrt{(1-\delta^2)}V_{0j} = r_{0j}$ maka persamaan (19) dan (20) dapat ditulis kembali sebagai berikut :

$$Y_i = \alpha + \beta X_i + \sigma\delta t_i + r_i, \qquad i = 1, \dots, n \quad (21)$$
$$Y_{0j} = \alpha + \beta X_0 + \sigma\delta t_{0j} + r_{0j}, \quad j = 1, \dots, k \quad (22)$$

Diperoleh nilai r_i dan r_{0j} yang berdistribusi identik dan saling bebas (*i.i.d*) $N\left(0; \sigma^2(1-\delta^2)\right)$ serta nilai t_i dan t_{0j} yang mengikuti distribusi distribusi half normal. Oleh karena itu, dengan kondisi nilai t_i dan t_{0j} serta mengganti nilai $\sigma\delta = b$ dan $\sigma^2(1-\delta^2) = \sigma_*{}^2$ maka persamaan (21) dan (22) dapat ditulis kembali sebagai berikut:

$$Y_i = \alpha + \beta X_i + bt_i + r_i, \qquad i = 1, \dots, n \quad (23)$$
$$Y_{0j} = \alpha + \beta X_0 + bt_{0j} + r_{0j}, \quad j = 1, \dots, k \quad (24)$$

lalu menentukan kembali f.k.p dari model kalibrasi linier pada persamaan (23) dan (24) sebagai berikut:

$$f(r_i;\theta) = \frac{1}{\sqrt{\pi^2{\sigma_*}^2}}\exp\left\{-\frac{1}{2{\sigma_*}^2}(r_i)^2\right\}\exp\left\{-\frac{1}{2}(t_i)^2\right\} \quad (25)$$

$$f(r_{0j};\theta) = \frac{1}{\sqrt{\pi^2{\sigma_*}^2}}\exp\left\{-\frac{1}{2{\sigma_*}^2}\left(r_{0j}\right)^2\right\}\exp\left\{-\frac{1}{2}\left(t_{0j}\right)^2\right\} \quad (26)$$

sehingga f.k.p bersama sebagai berikut:

$$f(r_i, r_{0j};\theta) = \prod_{i=1}^{n}\left[\frac{1}{\sqrt{\pi^2{\sigma_*}^2}}\exp\left\{-\frac{1}{2{\sigma_*}^2}(Y_i-\alpha-\beta X_i-bt_i)^2\right\}\exp\left\{-\frac{1}{2}(t_i)^2\right\}\right]\times \prod_{j=1}^{k}\left[\frac{1}{\sqrt{\pi^2{\sigma_*}^2}}\exp\left\{-\frac{1}{2{\sigma_*}^2}\left(Y_{0j}-\alpha-\beta X_0-bt_{0j}\right)^2\right\}\exp\left\{-\frac{1}{2}\left(t_{0j}\right)^2\right\}\right] \quad (27)$$

dimana $i = 1, \ldots, n$ dan $\mathrm{j} = 1, \ldots, \mathrm{k}$. Sehingga diperoleh fungsi likelihood sebagai berikut :

$$L(\theta^*; r_i, r_{0j}) = \left(\frac{1}{\sqrt{\pi^2{\sigma_*}^2}}\right)^n \exp\left\{-\frac{1}{2{\sigma_*}^2}\sum_{i=1}^{n}(Y_i-\alpha-\beta X_i-bt_i)^2\right\}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(t_i)^2\right\}\times \left(\frac{1}{\sqrt{\pi^2{\sigma_*}^2}}\right)^k \exp\left\{-\frac{1}{2{\sigma_*}^2}\sum_{j=1}^{k}\left(Y_{0j}-\alpha-\beta X_0-bt_{0j}\right)^2\right\}\exp\left\{-\frac{1}{2}\sum_{j=1}^{k}\left(t_{0j}\right)^2\right\} \quad (28)$$

cara sebagai berikut:

$$l = -\frac{n}{2}\ln(\pi^2{\sigma_*}^2) - \frac{1}{2{\sigma_*}^2}\sum_{i=1}^{n}(Y_i-\alpha-\beta X_i-bt_i)^2 - \frac{1}{2}\sum_{i=1}^{n}(t_i)^2 - \frac{k}{2}\ln(\pi^2{\sigma_*}^2) - \frac{1}{2{\sigma_*}^2}\sum_{j=1}^{k}\left(Y_{0j}-\alpha-\beta X_0-bt_{0j}\right)^2 - \frac{1}{2}\sum_{j=1}^{k}\left(t_{0j}\right)^2 \quad (29)$$

Dari fungsi l di atas, maka dapat dilanjutkan dengan algoritma-*EM*. Algoritma-*EM*. Dalam setiap iterasi pada algoritma-*EM* ini terdapat dua tahap, yaitu tahap Ekspektasi atau tahap E (*E-step*) dan tahap Maksimisasi atau tahap M (*M-step*) seperti yang dikemukakan oleh Weinstein [3]. Adapun langkah-langkah algoritma-*EM* sebagai berikut.

Tahap-*E*

Tahap ini bertujuan menemukan statistik cukup dengan cara menghitung nilai ekspektasi bersyarat dari fungsi l pada persamaan (29) di atas sebagai berikut:

$$E(l) = -\frac{n}{2}\ln(\pi^2{\sigma_*}^2) - \frac{1}{2{\sigma_*}^2}\{\sum_{i=1}^{n}(Y_i-\alpha-\beta X_i)^2 - 2\sum_{i=1}^{n}(Y_i-\alpha-\beta X_i)bE(t_i|Y_i) + b^2\sum_{i=1}^{n}E({t_i}^2|Y_i)\} - \frac{1}{2}\sum_{i=1}^{n}E((t_i|Y_i)^2) - \frac{k}{2}\ln(\pi^2{\sigma_*}^2) - \frac{1}{2{\sigma_*}^2}\left\{\sum_{j=1}^{k}\left(Y_{0j}-\alpha-\beta X_0\right)^2 - 2\sum_{j=1}^{k}\left(Y_{0j}-\alpha-\beta X_0\right)bE\left(t_{0j}|Y_{0j}\right) + b^2\sum_{j=1}^{k}E\left({t_{0j}}^2|Y_{oj}\right)\right\} - \frac{1}{2}\sum_{j=1}^{k}E\left(\left(t_{0j}|Y_{oj}\right)^2\right) \quad (30)$$

Sebagai perpanjangan dari hasil yang dijelaskan Arellano *et al*. [1], bahwa $t|Y$ mengikuti distribusi truncated normal $TN\left((Y-\alpha-\beta X)\frac{b}{b^2+{\sigma_*}^2};\frac{{\sigma_*}^2}{b^2+{\sigma_*}^2}\right)$ dengan f.k.p bersyaratnya sebagai berikut:

$$f(t|Y)=\sqrt{\frac{2(b^2+{\sigma_*}^2)}{\pi{\sigma_*}^2}}\exp\left\{-\frac{(b^2+{\sigma_*}^2)}{2{\sigma_*}^2}\left[t-\frac{(Y-\alpha-\beta X)b}{(b^2+{\sigma_*}^2)}\right]^2\right\} \quad (31)$$

dengan memanfaatkan nilai ekspektasi untuk distribusi truncated normal pada persaamaan (17), diperoleh persamaan momen pertama dan kedua dari fungsi $t|Y$ yang terdiri dari persamaan momen untuk percobaan regresi untuk $i=1,\ldots,n$ sebagai berikut :

$$\widehat{w}_{1i}=\left(Y_i-\hat{\alpha}-\hat{\beta}X_i\right)\frac{\hat{b}}{\hat{b}^2+\hat{\sigma}_*^2}+\frac{\hat{\sigma}_*}{\sqrt{\hat{b}^2+\hat{\sigma}_*^2}}\frac{\phi(\cdot)}{\Phi(\cdot)} \quad (32)$$

dan

$$\widehat{w}_{2i}=\left(Y_i-\hat{\alpha}-\hat{\beta}X_i\right)^2\frac{\hat{b}^2}{\left(\hat{b}^2+\hat{\sigma}_*^2\right)^2}+\frac{\hat{\sigma}_*^2}{\hat{b}^2+\hat{\sigma}_*^2}+\left(Y_i-\hat{\alpha}-\hat{\beta}X_i\right)\frac{\hat{b}}{\hat{b}^2+\hat{\sigma}_*^2}\frac{\hat{\sigma}_*}{\sqrt{\hat{b}^2+\hat{\sigma}_*^2}}\frac{\phi(\cdot)}{\Phi(\cdot)} \quad (33)$$

dimana $(\cdot)=\frac{\hat{b}\left(Y_i-\hat{\alpha}-\hat{\beta}X_i\right)}{\hat{\sigma}_*\sqrt{\hat{b}^2+\hat{\sigma}_*^2}}$, dan diperoleh persamaan momen percobaan kalibrasi untuk $j=1,\ldots,k$ sebagai berikut:

$$\widehat{w}_{1j}=\left(Y_{0j}-\hat{\alpha}-\hat{\beta}\hat{X}_0\right)\frac{\hat{b}}{\hat{b}^2+\hat{\sigma}_*^2}+\frac{\hat{\sigma}_*}{\sqrt{\hat{b}^2+\hat{\sigma}_*^2}}\frac{\phi(\cdot\cdot)}{\Phi(\cdot\cdot)} \quad (34)$$

dan

$$\widehat{w}_{2j}=\left(Y_{0j}-\hat{\alpha}-\hat{\beta}\hat{X}_0\right)^2\frac{\hat{b}^2}{\left(\hat{b}^2+\hat{\sigma}_*^2\right)^2}+\frac{\hat{\sigma}_*^2}{\hat{b}^2+\hat{\sigma}_*^2}+\left(Y_{0j}-\hat{\alpha}-\hat{\beta}\hat{X}_0\right)\frac{\hat{b}}{\hat{b}^2+\hat{\sigma}_*^2}\frac{\hat{\sigma}_*}{\sqrt{\hat{b}^2+\hat{\sigma}_*^2}}\frac{\phi(\cdot\cdot)}{\Phi(\cdot\cdot)} \quad (35)$$

dimana $(\cdot\cdot)=\frac{\hat{b}\left(Y_{0j}-\hat{\alpha}-\hat{\beta}X_0\right)}{\hat{\sigma}_*\sqrt{\hat{b}^2+\hat{\sigma}_*^2}}$.

Setelah diperoleh persamaan momen di atas, maka nilai ekspektasi pada persamaan (18) dapat ditulis kembali seperti persamaan berikut:

$$\begin{aligned}E(l)=&-\frac{n}{2}\ln(\pi^2{\sigma_*}^2)-\frac{1}{2{\sigma_*}^2}\{\textstyle\sum_{i=1}^{n}(Y_i-\alpha-\beta X_i)^2-2\sum_{i=1}^{n}(Y_i-\alpha-\beta X_i)b\widehat{w}_{1i}+\\&b^2\textstyle\sum_{i=1}^{n}\widehat{w}_{2i}\}-\frac{1}{2}\sum_{i=1}^{n}(\widehat{w}_{1i})^2-\frac{k}{2}\ln(\pi^2{\sigma_*}^2)-\frac{1}{2{\sigma_*}^2}\left\{\sum_{j=1}^{k}\left(Y_{0j}-\alpha-\beta X_0\right)^2-\right.\\&\left.2\textstyle\sum_{j=1}^{k}\left(Y_{0j}-\alpha-\beta X_0\right)b\widehat{w}_{1j}+b^2\sum_{j=1}^{k}\widehat{w}_{2j}\right\}-\frac{1}{2}\sum_{j=1}^{k}\left(\widehat{w}_{1j}\right)^2\end{aligned} \quad (36)$$

Tahap-*M*

Tahap ini bertujuan untuk memaksimalkan hubungan dengan vektor paramater θ^* dimana $\theta^*=(\alpha,\beta,\sigma_*,b,X_0)^T$. Oleh karena itu, nilai ekspektasi pada persamaan (36) dimaksimalkan dengan cara menurunkannya terhadap parameter yang akan ditaksir dan disamakan dengan nol. Setelah itu, dilakukan beberapa manipulasi aljabar sehingga diperoleh nilai parameter sebagai berikut:

$$\hat{b} = \frac{Syw_{(i)} + Sy_0 w_{(j)} - \frac{SxySxw_{(i)}}{Sxx}}{\sum_{i=1}^{n} \widehat{w}_{2i} + \sum_{j=1}^{k} \widehat{w}_{2j} - \overline{\overline{w}}_{1(i)} \sum_{i=1}^{n} \widehat{w}_{1i} - \overline{\overline{w}}_{1(j)} \sum_{j=1}^{k} \widehat{w}_{1j} - \frac{\left(Sxw_{(i)}\right)^2}{Sxx}}$$

$$\hat{\beta} = \frac{Sxy}{Sxx} - \frac{\hat{b}Sxw_{(i)}}{Sxx}\,; \qquad \hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} - \hat{b}\overline{\overline{w}}_{1(i)}; \qquad \hat{X}_0 = \frac{\bar{Y}_0 - \hat{\alpha}}{\hat{\beta}} - \frac{\hat{b}\overline{\overline{w}}_{1(j)}}{\hat{\beta}}$$

$$\hat{\sigma}_* = \left[\frac{\left(\sum_{i=1}^{n}\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right)^2 - 2\sum_{i=1}^{n}\left(Y_i - \hat{\alpha} - \hat{\beta}X_i\right)\hat{b}\overline{\overline{w}}_{1i} + \hat{b}^2 \sum_{i=1}^{n} \widehat{w}_{2i}\right)}{(n)} + \frac{\left(\sum_{j=1}^{k}\left(Y_{0j} - \hat{\alpha} - \hat{\beta}\hat{X}_0\right)^2 - 2\sum_{j=1}^{k}\left(Y_{0j} - \hat{\alpha} - \hat{\beta}\hat{X}_0\right)\hat{b}\overline{\overline{w}}_{1j} + \hat{b}^2 \sum_{j=1}^{k} \widehat{w}_{2j}\right)}{(k)}\right]^{\frac{1}{2}}$$

dimana:

$$Syw_{(i)} = \sum_{i=1}^{n}(Y_i - \bar{Y})\left(\widehat{w}_{1i} - \overline{\overline{w}}_{1(i)}\right); \qquad Sxw_{(i)} = \sum_{i=1}^{n}(X_i - \bar{X})\left(\widehat{w}_{1i} - \overline{\overline{w}}_{1(i)}\right)$$

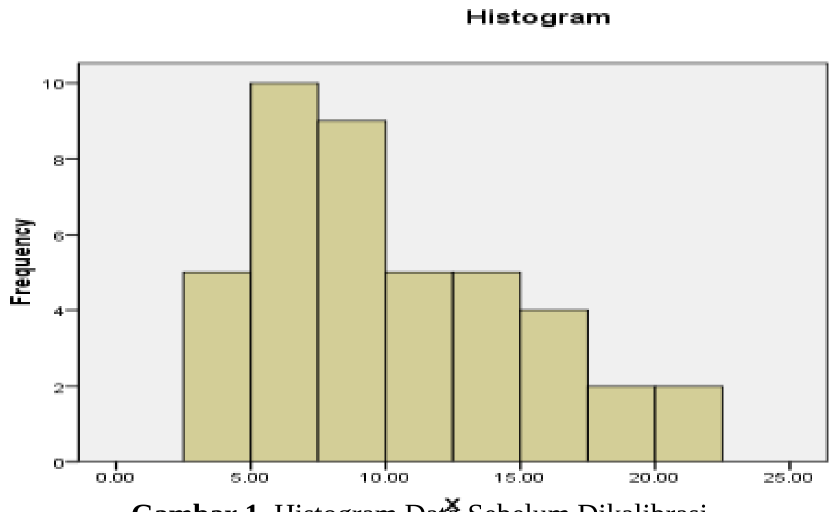
$$Sxy = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})\,; \qquad Sy_0 w_{(j)} = \sum_{j=1}^{k}\left(Y_{0j} - \bar{Y}_0\right)\left(\widehat{w}_{1j} - \overline{\overline{w}}_{1(j)}\right)$$

$$Sxx = \sum_{i=1}^{n}(X_i - \bar{X})^2\,; \qquad \overline{\overline{w}}_{1(i)} = \frac{\sum_{i=1}^{n} \widehat{w}_{1i}}{n}; \qquad \overline{\overline{w}}_{1(j)} = \frac{\sum_{j=1}^{k} \widehat{w}_{1j}}{k}$$

3.2. Aplikasi pada Data yang Diperoleh pada Jurnal [4]

Data yang digunakan dalam penelitian ini yakni data yang diperoleh pada jurnal Figueiredo [4] mengenai pengukuran dimensional dan ultrasonografi pada testis dari 42 remaja. Langkah-langkah dalam menganalisis data yaitu pertama, mem-plot frekuensi data untuk variabel X untuk mengetahui distribusi dari data tersebut. Adapun histogram datanya diberikan pada Gambar 1 berikut.

Histogram pada Gambar 1 di bawah, dapat dilihat bahwa data tersebut mengalami kemencengan atau mengikuti distribusi skew-normal. Kemencengan mulai terlihat setelah kelompok data 12,5–15. Sehingga data tersebut dapat diasumsikan mengikuti distribusi skew-normal dan dapat dimodelkan dengan pendekatan model kalibrasi. Penaksiran parameter dilakukan melalui tahapan iterasi, sehingga diperlukan bobot awal bagi setiap parameter $\alpha, \beta, X_0, \sigma^2$ dan λ yang akan ditaksir. Penentuan nilai X_0 dilakukan berdasarkan pada histogram dimana X_0 yang terpilih adalah titik awal pada X yang diasumsikan terjadinya perubahan pola data (skew/kemencengan). Berdasarkan Gambar 1, maka nilai X_0 yang dipilih adalah 16.4.



Gambar 1. Histogram Data Sebelum Dikalibrasi.

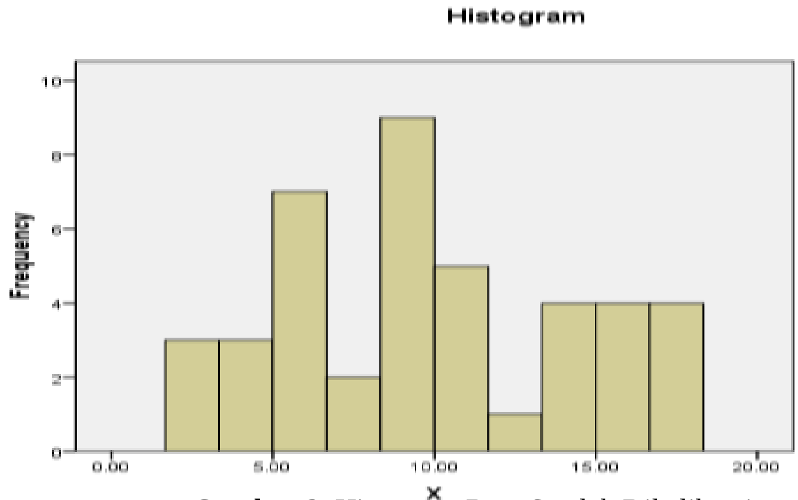
Setelah diperoleh nilai X_0, maka langkah selanjutnya menentukan nilai-nilai awal dari masing-masing parameter $\alpha, \beta, {\sigma_*}^2, \lambda$. Nilai awal untuk masing-masing parameter diperoleh dengan menggunakan rumus sesuai persamaan (7), (8), (9), (10), dan (11). Sehingga diperoleh nilai-nilai awal dari masing-masing parameter yang digunakan sebagai nilai awal pada algoritma-*EM*. Langkah selanjutnya yaitu menjalankan algoritma-*EM*. Pada penelitian ini, proses iterasi algoritma-*EM* dijalankan menggunakan Microsoft Excel 2007. Adapun hasil iterasinya sebagai berikut.

Tabel 1. Nilai Parameter Hasil Iterasi.

Iterasi	**Nilai Parameter**				
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}_*$	$\hat{X}_0$	$\hat{\lambda}$
0	$0{,}18$	$0{,}94$	$3{,}48$	$16{,}4$	$1{,}01$
1	-0.18	0.84	3.35	17.33	$1{,}57$
2	$2{,}88$	$0{,}83$	$2{,}94$	20.66	$-1{,}70$

Hasil iterasi nilai-nilai parameter pada Tabel 1 di atas, diperoleh nilai parameter awal yaitu $\hat{\alpha} = 0{,}18; \hat{\beta} = 0{,}94; \hat{\sigma}_* = 3{,}48;\ \hat{X}_0 = 16{,}4;\ \lambda = 1{,}01$. Pada proses iterasi dilakukan terlihat nilai-nilai parameter pada iterasi kedua semakin besar selisihnya dengan iterasi kedua, sehingga nilai taksiran yang digunakan adalah hasil iterasi pertama, yaitu $\hat{\alpha} = -0{,}18; \hat{\beta} = 0{,}84;\ \hat{\sigma}_* = 3{,}35;\ \hat{X}_0 = 17{,}33;\ \hat{\lambda} = 1{,}57$. Nilai taksiran X_0 sebesar $17{,}33$ berarti bahwa data variable X yang berada di atas $17{,}33$ akan dikalibrasi menjadi $17{,}33$ dengan pasangan data Y yang bersesuaian berdasarkan persamaan (6), dan diperoleh Y sebesar $14{,}35$. Dari data yang ada, menunjukkan bahwa terdapat empat data yang berada di atas $17{,}33$, yaitu $17{,}6; 20; 18{,}8$; dan $20{,}9$.

Hasil ini menunjukkan bahwa dari 42 data keseluruhan, terdapat 4 data yang dikalibrasi sehingga nilai keempat data tersebut mengikuti nilai kalibrasi yang diperoleh yaitu $X = 17{,}33$ yang berpasangan dengan nilai $Y = 14{,}35$. Selanjutnya data hasil kalibrasi tersebut diplotkan melalui histogram, dan ditunjukkan pada gambar berikut ini.



Gambar 2. Histogram Data Setelah Dikalibrasi.

Pada Gambar 1 dan Gambar 2 terlihat perubahan histogram, dimana pada Gambar 1 terlihat data yang menceng, dan setelah dikalibrasi pada Gambar 2, terlihat histogram yang sudah seimbang. Data hasil kalibrasi tersebut selanjutnya dimodelkan kembali melalui pendekatan analisis regresi, dan diperoleh hasil sebagai berikut.

Tabel 2. Perbandingan Nilai Parameter Model Regresi Liniear Kalibrasi dan Model Liniear

Parameter	Model Regresi Linear	Model Kalibrasi Linear
Alpa	0,461	0,956
Beta	0,903	0,835
Se	1,62	1,56

Berdasarkan Tabel 2 di atas, menunjukkan bahwa model yang diperoleh dari hasil kalibrasi memiliki standard error (SE) yang lebih kecil dibanding model regresi linear biasa, yaitu sebesar 1,56, artinya model kalibrasi lebih layak digunakan dalam menjelaskan hubungan antara variabel X dengan Y.

1. Kesimpulan dan Saran

Berdasarkan pembahasan hasil yang telah diselesaikan maka dapat ditarik kesimpulan sebagai berikut :

a. Penerapan pada data pengukuran dimensional dan ultrasononografi pada testis dari 42 remaja yang dikalibrasi karena error pada data tersebut berdistribusi skew-normal. Sehingga diperoleh model kalibrasi linier untuk data tersebut sebagai berikut: $\hat{Y}_i = \hat{\alpha} + \hat{\beta}X_i$ dan $\hat{Y}_{0j} = \hat{\alpha} + \hat{\beta}\hat{X}_0$ dengan nilai-nilai taksiran parameter sebagai berikut: $\hat{\alpha} = -0{,}18$; $\hat{\beta} = 0{,}84$; $\hat{\sigma}_* = 3{,}35$; $\hat{X}_0 = 17{,}33$; $\hat{\lambda} = 1{,}57$.
b. Berdasarkan Tabel 3.2, menunjukkan bahwa nilai standard error (SE) dari kalibrasi liniear lebih kecil dibanding SE dari model regresi liniear, artinya model yang diperoleh dari hasil kalibrasi lebih layak digunakan dalam menjelaskan hubungan antara variabel X dengan Y.

Tulisan ini hanya melakukan penaksiran parameter model kalibrasi linier yang berdistribusi skew-normal dengan pendekatan *MLE* dan iterasi numerik algoritma-*EM*. Disarankan bagi pembaca yang berminat lebih dalam dengan permasalahan penaksiran parameter model kalibrasi linier khususnya yang berdistribusi skew-normal untuk mencoba metode pendekatan numerik lainnya serta penggunaan algoritma-*EM* pada berbagai macam program agar menghasilkan nilai iterasi yang lebih akurat.

.

Daftar Pustaka

[1] Arellano-Valle R.B., Bolfarine H., and Lachos V.H., 2005. Skew-Normal Linear Mixed Model. *J. Data. Sci*.

[2] Azzalini A., 1985. A Class of Distribution Which Includes The Normal One. *Scand. J. Stat*.

[3] Couvreur C., 1996. *The EM Algorithm: A Guided Tour*. Belgia.

[4] Weinstein E., 2006. *Expectation-Maximization Algorithm and Aplications*. Courant Institute of Mathematical Science.

[5] Figueiredo C.C. *et al*., 2010. *On The Skew-Normal Calibration Model*. Brazil: Universitas Federal Pernambuco.

[6] Henze N., 1986. A Probabilistic Representation Of The Skew-Normal Distribution. *Scand. J. Stat*.

[7] Jilmes J.A., 1998. *A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models*. International Computer Science Institute, California.

[8] Liseo B. and Loperfido N., 2006. A Note On Reference Priors For The Scalar Skew-Normal Distribution. *J. Stat.*

[9] Mandel M. *et al*., 2003. *An EM Algorithm for Localizing Multiple Sound Sources in Reveberant Environments*. Columbia University, New York.

[10] McLachlan *et al*., 2004. *The EM Algorithm*. Center for Applied Statistics and Economics, Berlin.

[11] Walpole R.E., and Myres R.H., 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan Edisi ke-4*. Institut Teknologi Bandung, Bandung.

[12] Santuo, 2012. Penaksiran Parameter Model Regresi Inverse Gaussian Dengan Peubah Respon Kontinu Non-Negatif. *Skripsi*. Universitas Hasanuddin, Makassar.

[13] Sartori N., 2012. *Bias Prevention Of Maximum Likelihood Estimates For Scalar Skew-Normal And Skew-t Distribution*. Venezia: Universita “Ca’Foscari”http://www.fp.unud.ac.id/ind/AnalisisRegresiLinierSederhana.pdf. [Diakses pada tanggal 3 November 2012]