# MODEL REGRESI TERSENSOR

Skripsi

Disusun untuk melengkapi syarat-syarat

guna memperoleh gelar Sarjana Sains

Christina

3125051612

PROGRAM STUDI MATEMATIKA

JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS NEGERI JAKARTA

2011

# PERSEMBAHANKU

”Kekuatan serta penghiburan diberikan Tuhan Yesus padaku... dan sesuai dengan himat Tuhan, ku diberikan apa yang perlu, suka dan derita bergantian, memperkuat imanku.”

(Kidung Jemaat No. 332)

” Rejoice always,
Pray without ceasing,
In everything give thanks, for this is the will of God in Christ Jesus for you.”

(1 Tesalonika 5 : 16 - 18)

Skripsi ini kupersembahkan kepada
Mama dan Bapa...
Ka Juni, Manis, Dita, Ivon dan Saut
Terima kasih atas
pengorbanan, dukungan doa,
cinta dan kasih sayang kalian...

# ABSTRACT

**CHRISTINA, 3125051608.** ***Censored Regression Model. Thesis. Faculty of Mathematics and Natural Sciences Jakarta State University. 2011.***

*Zero consumption is an issue that often occurs in data survey on consumption or household expenditures. It happens in which most household do not consume certain commodities, while the consumption of other household varies. The classical regression model by the method of Ordinary Least Square (OLS) will be good to be used if the assumption of homogeneity of freedom and variety are met. However, the more zero number found in observations, it will lead to heteroscedasticity. The using of OLS method will produce biased and inconsistent estimator because the underlying assumptions are not met. Censored regression model is an alternative regression model that can be used in zero consumption issue. Furthermore, this thesis analyzes the censored regression model to solve the zero consumption issue. It produces smaller mean square error (MSE) than MSE that is generated from an ordinary regression model. In addition, it results a more precise parameter than what an ordinary regression model generates.*

***Keyword :*** *Zero Consumption, Ordinary Least Square, Censored Regression Model.*

# ABSTRAK

**CHRISTINA, 3125051612. Model Regresi Tersensor. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Negeri Jakarta. 2011.**

*Zero consumption* adalah masalah yang sering terjadi pada data survey konsumsi/pengeluaran rumah tangga, dimana sebagian rumah tangga tidak mengkonsumsi jenis komoditas tertentu, sedangkan rumah tangga yang lain mengkonsumsi dalam jumlah yang sangat bervariasi. Model regresi klasik dengan metode kuadrat terkecil (*Ordinary Least Square*) akan cukup baik didekati apabila asumsi kenormalan, kebebasan dan kehomogenan ragam terpenuhi. Namun, semakin banyak pengamatan bernilai nol pada data yang diperoleh akan menyebabkan heteroskedastisitas. Penggunaan metode OLS akan menghasilkan penduga yang bias dan tidak konsisten karena asumsi yang mendasari tidak dipenuhi. Model regresi tersensor adalah model regresi alternatif yang dapat digunakan dalam masalah *zero consumption.* Pada skripsi ini dibahas model regresi tersensor untuk mengatasi masalah *zero consumption*, dimana dihasilkan MSE yang lebih kecil daripada MSE model regresi biasa, serta dihasilkan penduga parameter yang lebih tepat dibandingkan model regresi biasa.

**Kata kunci :** *Zero Consumption*,*Ordinary Least Square*, Model Regresi Tersensor.

# KATA PENGANTAR

Sgala Puji, Hormat, dan Kemuliaan layaklah penulis panjatkan kepada Tuhan atas alam semesta, Raja Sorga, Tuhan Yesus Kristus untuk segala kasih, anugerah dan pemeliharaanNya dalam hidup penulis hingga akhirnya tugas akhir skripsi yang berjudul Model Regresi Tersensor dapat diselesaikan dengan baik guna memenuhi syarat kelulusan sebagai salah satu tanggung jawab penulis kepada orang tua, dosen dan terlebih Tuhan yang menganugerahkan pendidikan tinggi ini.

Skripsi ini terselesaikan tidak lepas dari dukungan berbagai pihak, dan melalui kesempatan ini penulis berterima kasih kepada:

1. Prof. Dr. Suyono, M.Si selaku Dosen Pembimbing I dan Ir. Fariani Hermin I, M.T selaku Dosen Pembimbing II yang telah meluangkan waktu, memberi saran dan atas kesabarannya memberikan bimbingan dan motivasi dalam penyelesaian penulisan skripsi ini.

2. Seluruh dosen dan staf karyawan Jurusan Matematika, terima kasih atas ilmu dan kerja sama yang telah diberikan.

3. Keluargaku yang tercinta Mama (R. Siagian), Bapak (M.Simanjuntak) dan kakak (Yuni. Mita Atmasari) serta adik-adikku (Manis, Dita, Ivon, Saut) terima kasih atas kepercayaan, doa dan kasih sayangnya serta bantuan secara moral maupun materil selama penulis menuntut ilmu di UNJ sampai terlaksananya penulisan skripsi ini.

4. Sahabat dan Saudara seiman terbaikku, Bresna, Saudara seiman dalam Yesus, Ka Iky, Bang Roni, Hanna, Ishak, terima kasih untuk topangan doa untukku yang slalu ada dalam doa-doamu. *I hope our friendship is never end.*

5. Sahabat-sahabat kecilku yang senantiasa ada dalam kehidupan dewasaku, Tika. Sahabat SMA ku, Fera. Sahabat-sahabat kampusku, Renti, Retno, Maria, Ulfa, Putri.

6. Pemimpin Kelompok Kecilku, Ka Mega dan Teman Kelompok Kecilku, Dwi Nita. Teman-teman, kakak, abang serta adik di Persekutuan Mahasiswa Kristen (PMK UNJ). Terima kasih untuk setiap doa dan perhatian kalian.

7. Teman-teman di Matematika 2003 - 2008 dan PM05, terutama untuk Ka Dzaky, Ka Dina, Ka Ponco, Ka Jco semoga hubungan baik diantara kita dapat selalu terjalin.

8. Guru dan staff SMP Kristen Ketapang I dan SD Kristen Nasional Anglo.

9. Semua pihak yang telah membantu dalam proses penulisan skripsi ini yang mungkin tak bisa penulis sebutkan satu persatu. Terima kasih.

10. Dududbolobolo a.k.a Riko Wibawa Sitanggang. Terima kasih untuk doa, dukungan dan kelucuan yang kau berikan selama kita bersama.

Penulis menyadari adanya kekurangan dalam penyajian skripsi ini yang disebabkan oleh keterbatasan ilmu pengetahuan dan kemampuan yang penulis miliki. Penulis berharap semoga hasil dari skripsi ini dapat bermanfaat bagi pembaca dan semua pihak yang memerlukan informasi dari skripsi ini.

Jakarta, Juli 2011

Christina

# DAFTAR ISI

# DAFTAR GAMBAR

# DAFTAR TABEL

# DAFTAR LAMPIRAN

# BAB I

# PENDAHULUAN

Dalam data survey konsumsi atau pengeluaran rumah tangga sering ditemukan sebagian rumah tangga tidak mengkonsumsi jenis komoditas tertentu *(zero consumption)*, sedangkan rumah tangga yang lain mengkonsumsi dengan jumlah yang sangat bervariasi. Sebagai contoh adalah pengeluaran untuk tembakau. Nilai ini akan bernilai 0 (nol) bila obyek penelitian bukan perokok *(zero consumption)* dan bervariasi bila obyek penelitian adalah perokok. Berarti dalam hal ini ada variabel dependen (variabel tak bebas) yang terkonsentrasi pada suatu nilai, yaitu titik nol. Dalam kasus ini variabel dependen tersebut dikatakan tersensor, dimana nilai 0 (nol) adalah nilai batas bawah *(Maddala, 1983)*.

Model statistik yang dapat menggambarkan keadaan data seperti contoh di atas adalah model tobit atau dikenal sebagai model regresi tersensor *(censored regression model)*. Model regresi tersensor adalah model statistika yang variabel dependennya terkonsentrasi pada suatu nilai *(Gujarati, 1973)*.

Dalam model regresi tersensor, kelinieran hubungan antara variabel dependen dengan variabel independennya tidak terpenuhi, maka metode yang digunakan untuk menaksir parameter-parameternya adalah penaksiran maksimum likelihood. Penggunaan metode penaksiran *Least Square* untuk menaksir parameter-parameter model regresi tersensor akan menghasilkan penduga yang bias dan tidak konsisten. Karena asumsi-asumsi dasar regresi seperti asumsi kenormalan, kebebasan dan kehomogenan ragam tidak terpenuhi. Metode penaksiran *least square* akan menghasilkan penduga yang baik apabila asumsi-asumsi dasar regresi terpenuhi *(Amemiya, 1973)*.

Model regresi tersensor pertama kali ditemukan oleh Tobin pada tahun 1958 dan selanjutnya model regresi tersensor banyak mengalami perkembangan atau variasi dari model awalnya. Model regresi tersensor yang akan dibahas dalam skripsi ini adalah model regresi tersensor yang pertama kali ditemukan oleh Tobin pada tahun 1958 yang sekarang disebut sebagai model regresi tersensor standar.

## 1.1 Perumusan Masalah

Perumusan masalah yang akan dibahas dalam skripsi ini adalah :

(a) Bagaimanakah model regresi tersensor standar?

(b) Bagaimana melakukan analisis data dengan menggunakan model regresi tersensor standar?

## 1.2 Pembatasan Masalah

Skripsi ini akan membahas model regresi tersensor standar pada batas bawah yang bernilai 0 (nol).

## 1.3 Tujuan Penulisan

Tujuan yang diharapkan dari penulisan skripsi ini adalah : Mengkaji penerapan model regresi tersensor standar pada data yang mengandung banyak amatan bernilai 0 (nol).

## 1.4 Manfaat Penulisan

Penulisan ini diharapkan dapat dimanfaatkan untuk :

(a) Membantu masyarakat atau peneliti dalam menentukan model regresi untuk nilai amatan 0 (nol).

(b) Menambah pengetahuan mengenai metode-metode penentuan model regresi untuk nilai amatan 0 (nol).

# BAB II

# LANDASAN TEORI

## 2.1 Regresi Linier

### 2.1.1 Model Regresi Linier

Analisis regresi merupakan teknik statistika yang digunakan untuk menjelaskan hubungan antara sepasang atau beberapa peubah bebas dan peubah tak bebas. Dengan analisis regresi hubungan tersebut dapat dibuat model atau persamaan matematika.

Jika hubungan antara peubah bebas dengan peubah tak bebas adalah linier dan hanya terdapat satu peubah bebas, maka digunakan analisis regresi linier sederhana. Model regresi linier sederhana adalah:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

dengan $\beta_0$ adalah intersep, $\beta_1$ adalah koefisien regresi yang tidak diketahui dan $\varepsilon$ adalah galat yang diasumsikan berdistribusi normal dengan mean 0 dan variansi $\sigma^2$.

Generalisasi dari regresi linier sederhana adalah persamaan regresi linier berganda. Persamaan regresi linier berganda adalah persamaan dari satu peubah tak bebas ($Y$) dengan peubah bebas lebih dari satu ($X_1, X_2, ...., X_p$). Hubungan antara peubah-peubah tersebut dapat dirumuskan dalam bentuk persamaan:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \varepsilon$$

Jika ada amatan sebanyak $n$, maka persamaan di atas menjadi:

$$\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_{11} + \beta_2 X_{12} + ... + \beta_p X_{1p} + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_{21} + \beta_1 X_{22} + ... + \beta_p X_{2p} + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_{n1} + \beta_1 X_{n2} + ... + \beta_p X_{np} + \varepsilon_n \end{aligned}$$

Dalam notasi matriks sistem persamaan di atas dapat ditulis sebagai:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon \tag{2.1}$$

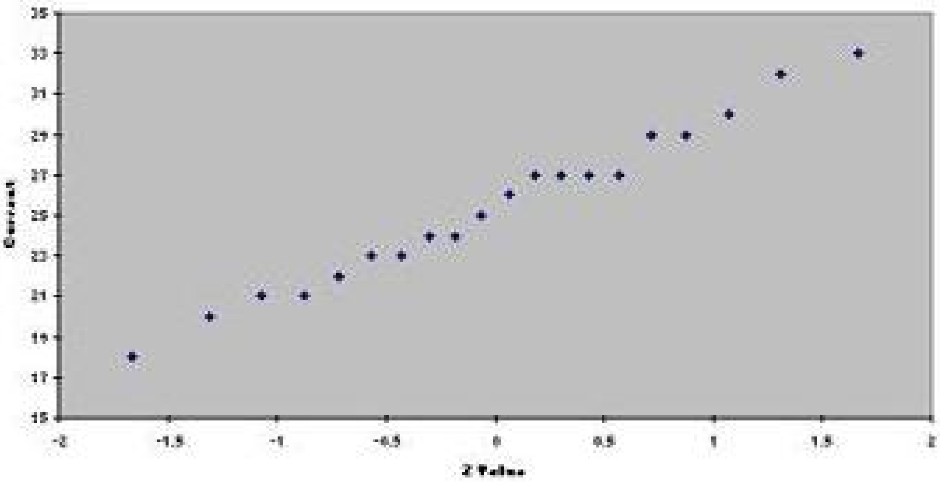
$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \text{dan } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

### 2.1.2 Asumsi-asumsi Regresi

Asumsi-asumsi yang diperlukan dalam analisis regresi adalah sebagai berikut:

(a) Asumsi Kenormalan

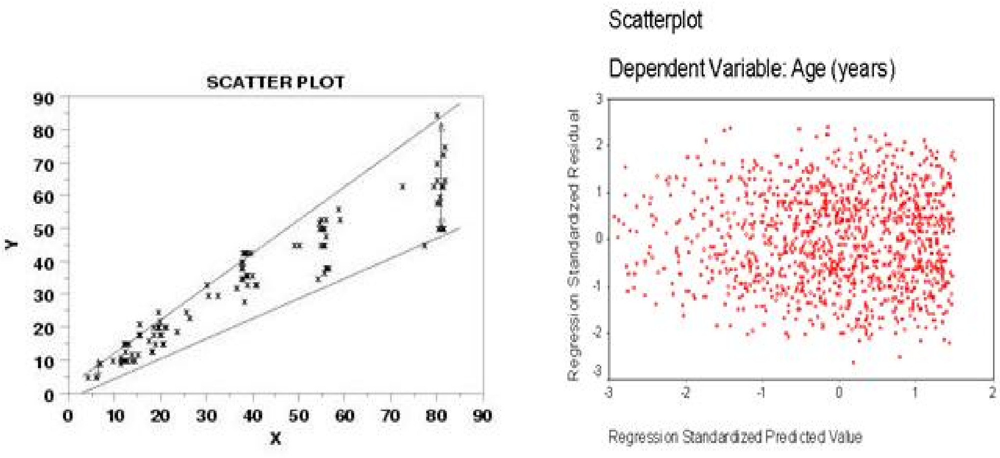
Asumsi kenormalan data merupakan asumsi yang paling mendasar dalam analisis regresi. Untuk memenuhi kriteria kenormalan data sebagai langkah awal dalam analisis regresi perlu dilakukan uji kenormalan data. Uji kenormalan data dapat dilakukan dengan uji peluang plot kenormalan data *(normal probability plot)*. Kriteria uji ini adalah melakukan perbandingan nilai distribusi kumulatif data awal dengan nilai kumulatif dari distribusi normal *(cummulative normal distribution)*. Data yang memenuhi kriteria kenormalan akan mengikuti pola garis lurus.




Gambar 2.1 Plot Data Normal

(b) Asumsi Homoskedastisitas

Homoskedastisitas adalah kondisi ketika kesalahan galat ($\varepsilon_i$) mempunyai varian yang sama ($\sigma^2$) dari pengamatan yang satu ke pengamatan yang lain. Pelanggaran asumsi ini disebut heteroskedastisitas *(heteroscedasticity)* yang merupakan kasus ketika kesalahan galat ($\varepsilon_i$) mempunyai varian yang berbeda. Konsekuensi terjadinya kasus heteroskedastisitas adalah pendugaan dari model yang diperoleh tidak lagi efisien baik untuk sampel kecil maupun untuk sampel besar. Untuk mendeteksi asumsi ini, dapat dilihat *scatter plot* antara nilai dari prediksi $Y$ terhadap residual membentuk pola acak dan menyebar di atas garis lurus, maka uji asumsi Homoskedastisitas terpenuhi.




(a) (b)

Gambar 2.2 Plot Data Heteroskedastisitas (a) dan Homoskedastisitas (b)

(c) Asumsi Non Multikolinieritas

Multikolinieritas *(multicollinearity)* merupakan adanya hubungan erat atau sangat erat yang terjadi antara beberapa variabel bebas atau terjadi korelasi yang tinggi antara variabel bebas. Salah satu cara untuk mendeteksi keberadaan multikolinieritas adalah dengan melihat besarnya nilai *Variance Inflation Factor (VIF)*. Dikatakan tidak terjadi multikolinieritas jika nilai *VIF* berbeda antara satu sampai sepuluh (Myers, 1994). Untuk mendeteksi adanya multikolinearitas digunakan korelasi Spearman dengan melihat korelasi antara variabel-variabel independen. Jika pada tingkat signifikansi $\alpha = 0.05$, variabel independen, $X_1, X_2$ tidak berkorelasi, maka tidak terdapat multikolinearitas.

### 2.1.3 Pendugaan Koefisien Regresi

Penduga bagi koefisien regresi $\beta$ dalam model regresi dengan menggunakan metode kuadrat terkecil berdasarkan model $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ yaitu dengan meminimumkan jumlah kuadrat galat (JKG) dimana JKG dirumuskan sebagai berikut:

$$JKG = \varepsilon^{'}\varepsilon = (\mathbf{Y} - \mathbf{X}\beta)^{'}(\mathbf{Y} - \mathbf{X}\beta)$$

**Teorema 2.1.1.** Jika $\mathbf{X}$ adalah matriks rancangan berukuran $n \times (p+1)$ yang bersifat *full rank* dengan $(p+1) \leq n$, maka penduga kuadrat terkecil untuk $\beta$ adalah

$$\beta = (\mathbf{X}^{'}\mathbf{X})^{-1}(\mathbf{X}^{'}Y)$$

*Bukti.* Misal $\mathbf{X}$ adalah matriks rancangan berukuran $n \times (p+1)$ dengan $(p+1) \leq n$ yang bersifat *full rank* , yaitu matriks penuh yang memiliki rank $(p+1)$. Penduga $\beta$ ditentukan dengan meminimumkan jumlah kuadrat galat. Berdasarkan $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$, penduga bagi vektor galat $\varepsilon$ dapat ditulis sebagai $\varepsilon = \mathbf{Y} - \mathbf{X}\beta$ sehingga jumlah kuadrat galat dapat dinyatakan sebagai berikut:

$$\begin{aligned}
\varepsilon^{'}\varepsilon &= (\mathbf{Y} - \mathbf{X}\beta)^{'}(\mathbf{Y} - \mathbf{X}\beta) \\
&= (\mathbf{Y}^{'} - \beta^{'}\mathbf{X}^{'})(\mathbf{Y} - \mathbf{X}\beta) \\
&= \mathbf{Y}^{'}\mathbf{Y} - \mathbf{Y}^{'}\mathbf{X}\beta - \beta^{'}\mathbf{X}^{'}\mathbf{Y} + \beta^{'}\mathbf{X}^{'}\mathbf{X}\beta
\end{aligned}$$

$\beta^{'}$ adalah matriks yang berukuran $(1 \times p)$, $\mathbf{X}^{'}$ adalah matriks yang berukuran $(p \times n)$ dan $\mathbf{Y}$ adalah matriks yang berukuran $(1 \times 1)$, maka $\beta^{'}\mathbf{X}^{'}\mathbf{Y}$ matriks berukuran $1 \times 1$. Dapat dituliskan $\beta^{'}\mathbf{X}^{'}\mathbf{Y} = (\beta^{'}\mathbf{X}^{'}\mathbf{Y})^{'} = \mathbf{Y}^{'}\mathbf{X}\beta$. Sehingga diperoleh

$$\begin{aligned}
\varepsilon^{'}\varepsilon &= \mathbf{Y}^{'}\mathbf{Y} - 2\mathbf{Y}^{'}\mathbf{X}\beta + \beta^{'}(\mathbf{X}^{'}\mathbf{X})\beta \\
&= \mathbf{Y}^{'}\mathbf{Y} - 2(\mathbf{X}^{'}\mathbf{Y})^{'}\beta + \beta^{'}(\mathbf{X}^{'}\mathbf{X})\beta \qquad (2.2)
\end{aligned}$$

Untuk memperoleh $\beta$ yang membuat $\varepsilon^{'}\varepsilon$ minimum, persamaan (2.2) diturunkan terhadap $\beta$.

$$\frac{\partial \varepsilon^{'}\varepsilon}{\partial \beta} = \frac{\partial(\mathbf{Y}^{'}\mathbf{Y})}{\partial \beta} - 2(\frac{\partial((\mathbf{X}^{'}\mathbf{Y})^{'}\beta)}{\partial \beta}) + \frac{\partial(\beta^{'}(\mathbf{X}^{'}\mathbf{X})\beta)}{\partial \beta}$$

Telah diketahui bahwa

$$\frac{\partial(\mathbf{Y}^{'}\mathbf{Y})}{\partial \beta} = 0$$

Misal: $\mathbf{Z} = a^{'}\mathbf{Y}$ dimana $a$ adalah vetor skalar. Maka $\frac{\partial Z}{\partial \mathbf{Y}} = a$. Anggap $(\mathbf{X}^{'}\mathbf{Y}) = a$, maka $\frac{\partial(a^{'}\beta)}{\partial \beta} = a$. Untuk

$$\begin{aligned}
\frac{\partial((\mathbf{X}^{'}\mathbf{Y})^{'}\beta)}{\partial \beta} &= \frac{\partial(a^{'}\beta)}{\partial \beta} \\
&= a \\
&= \mathbf{X}^{'}\mathbf{Y}
\end{aligned}$$

Dan untuk $\frac{\partial(\beta^{'}(\mathbf{X}^{'}\mathbf{X})\beta)}{\partial \beta}$, misal $(\mathbf{X}^{'}\mathbf{X}) = \mathbf{A}$. Dimana $\mathbf{A}$ adalah matriks $k \times k$, maka dapat dituliskan

$$\begin{aligned}
\frac{\partial(\beta^{'}(\mathbf{X}^{'}\mathbf{X})\beta)}{\partial \beta} &= \frac{\partial(\beta^{'}\mathbf{A}\beta)}{\partial \beta} \\
&= (\mathbf{X}^{'}\mathbf{X})\beta + (\mathbf{X}^{'}\mathbf{X})^{'}\beta.
\end{aligned}$$

Sehingga diperoleh

$$\begin{aligned}
\frac{\partial \varepsilon^{'}\varepsilon}{\partial \beta} &= \frac{\partial(\mathbf{Y}^{'}\mathbf{Y})}{\partial \beta} - 2(\frac{\partial((\mathbf{X}^{'}\mathbf{Y})^{'}\beta)}{\partial \beta}) + \frac{\partial(\beta^{'}(\mathbf{X}^{'}\mathbf{X})\beta)}{\partial \beta} \\
&= 0 - 2\mathbf{X}^{'}\mathbf{Y} + (\mathbf{X}^{'}\mathbf{X})\beta + (\mathbf{X}^{'}\mathbf{X})^{'}\beta \\
&= -2\mathbf{X}^{'}\mathbf{Y} + (\mathbf{X}^{'}\mathbf{X})\beta + (\mathbf{X}^{'}\mathbf{X})^{'}\beta
\end{aligned}$$

Setelah diturunkan terhadap $\beta$, selanjutnya hasil turunan terhadap $\beta$ di atas disamakan dengan nol, yaitu

$$-2\mathbf{X}'\mathbf{Y} + 2(\mathbf{X}'\mathbf{X})\beta = 0$$

atau

$$(\mathbf{X}'\mathbf{X})\beta = \mathbf{X}'\mathbf{Y}$$

Berdasarkan persamaan di atas maka solusi untuk $\beta$ adalah:

$$\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{2.3}$$

□

### 2.1.4 Sifat-Sifat Penaksir Kuadrat Terkecil

Penduga yang baik harus memiliki beberapa kriteria, yaitu bersifat tak bias dan memiliki varians yang minimum. Pendugaan dengan kuadrat terkecil memiliki kriteria tersebut. Dalam model umum regresi linier:

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

dianggap bahwa $\varepsilon$ bebas satu sama lain dan $E(\varepsilon) = 0$, $\mathrm{var}(\varepsilon) = \sigma^2\mathbf{I}$. Karena peubah $\mathbf{X}$ tidak mempunyai distribusi jadi diperlakukan sebagai tetapan maka $E(\mathbf{Y}) = \mathbf{X}\beta$ dan $\mathrm{var}(\mathbf{Y}) = \sigma^2\mathbf{I}$.

Penduga $\mathbf{b}$ bersifat tak bias untuk $\beta$ karena

$$\begin{aligned}
E(\mathbf{b}) &= E((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\mathbf{Y}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta \\
&= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\beta \\
&= \mathbf{I}\beta = \beta
\end{aligned}$$

Nilai variansi dari penduga kuadrat terkecil adalah:

$$\begin{aligned}
\mathrm{var}(\mathbf{b}) &= \mathrm{var}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{var}(\mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \qquad (2.4)
\end{aligned}$$

Ragam penduga $\mathbf{b}$ bernilai minimum dibandingkan dengan ragam penduga tak bias lainnya. Hal ini dinyatakan dalam teorema Gauss-Markov.

**Teorema 2.1.2.** Penduga kuadrat terkecil $\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ mempunyai variansi terkecil dalam himpunan semua penaksir linier tak bias

*Bukti.* Misalkan $\mathbf{b}^*$ penaksir linier lain dari $\beta$ yang juga tak bias. Karena $\mathbf{b}^*$ penaksir linier maka dapat dimisalkan bentuknya sebagai:

$$\mathbf{b}^* = [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + U]\mathbf{Y}$$

untuk suatu matriks $\mathbf{U}$ yang merupakan fungsi dari $\mathbf{X}$, jadi:

$$\begin{aligned} E(\mathbf{b}^*) &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + U]E(\mathbf{Y}) \\ &= [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + U](\mathbf{X}\beta) \\ &= \beta + \mathbf{XU}\beta \\ &= (\mathbf{I} + \mathbf{UX})\beta \end{aligned}$$

Agar $\mathbf{b}^*$ penaksir tak bias dari $\beta$ maka haruslah $\mathbf{UX} = 0$, jadi nilai variansi untuk $\mathbf{b}^*$ adalah:

$$\begin{aligned} \mathrm{var}(\mathbf{b}^*) &= \mathrm{var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' + U]\mathbf{Y} \\ &= [(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{U}]\mathrm{var}(\mathbf{Y})[\mathbf{U}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= \sigma^2[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{U}' + \mathbf{UU}' + (\mathbf{X}'\mathbf{X})^{-1} + \mathbf{UX}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= \sigma^2[(\mathbf{X}'\mathbf{X})^{-1} + \mathbf{U}'\mathbf{U}] \end{aligned}$$

karena diharapkan $\mathbf{UX} = 0$ dan karena $\mathbf{UX} = \mathbf{X}'\mathbf{U}' = 0$ matriks $\mathbf{UU}'$ adalah definit positif, semua unsur diagonalnya berbentuk kuadrat. Jadi terbukti bahwa variansi dari $\mathbf{b}^*$ selalu lebih besar atau paling kecil sama dengan variansi unsur $\beta$ yang sesuai. $\square$

### 2.1.5 Analisis Ragam

Analisis ragam (*analysis of variance*, ANOVA) adalah suatu metode analisis statistika yang termasuk ke dalam cabang statistika inferensi. Secara umum, analisis ragam menguji dua varians (atau ragam) berdasarkan hipotesis nol bahwa kedua varians itu sama. Supaya valid dalam menafsirkan hasilnya, analisis ragam menggantungkan diri pada empat asumsi yang harus dipenuhi dalam perancangan percobaan:

(a) Data berdistribusi normal, karena pengujiannya menggunakan uji

F

(b) Varians atau ragamnya homogen, dikenal sebagai homoskedastisitas

(c) Saling bebas

(d) Komponen-komponen dalam modelnya bersifat aditif (saling menjumlah)

Asumsi untuk pengujian hipotesis yang didasarkan pada model ANOVA faktor tunggal berhubungan dengan nilai residual atau error ($\varepsilon_{ij}$). Apabila analisis data tidak memenuhi asumsi analisis ragam, maka kesimpulan yang diambil tidak akan menggambarkan keadaan yang sebenarnya. Dengan demikian, sebelum melakukan analisis ragam, terlebih dahulu diperiksa apakah data tersebut sudah memenuhi asumsi dasar analisis ragam atau belum. Dengan demikian, asumsi-asumsi yang harus dipenuhi dalam melakukan analisis ragam adalah: normalitas, homoskedastisitas (kehomogenan ragam), independensi (kebebasan galat).

### 2.1.6 Uji Hipotesis

Dalam kasus regresi linier keragaman dari peubah tak bebas ($Y$) dapat dilihat melalui jumlah kuadrat selisih antara amatan $Y$ dengan rataannya, yang disebut jumlah kuadrat total (JKT). Jumlah kuadrat total didefinisikan sebagai hasil jumlah antara jumlah kuadrat galat (JKG) dengan jumlah kuadrat regresi (JKR).Hubungan antara JKT dengan JKG dan JKR dapat dijelaskan sebagai berikut:

$$\begin{aligned}
\sum(Y_i-\overline{Y})^2 &= \sum(Y_i-\widehat{y}+\widehat{Y}-\overline{Y})^2 \\
&= \sum[(Y_i-\widehat{Y}+\widehat{Y}-\overline{Y})]^2 \\
&= \sum(Y_i-\widehat{Y}_i)^2+\sum(\widehat{Y}_i-\overline{Y})^2+2(Y_i-\widehat{Y}_i)+(\widehat{Y}_i-\overline{Y})
\end{aligned}$$

Karena $2(Y_i - \widehat{Y}_i) + (\widehat{Y}_i - \overline{Y}) = 0$ maka:

$$\sum(\widehat{Y}_i - \overline{Y})^2 = \sum(Y_i - \widehat{Y}_i)^2 + \sum(\widehat{Y}_i - \overline{Y})^2$$

dimana

$\sum(\widehat{Y}_i - \overline{Y})^2$ adalah jumlah kuadrat terkoreksi (JKT).

$\sum(Y_i - \widehat{Y}_i)^2$ adalah jumlah kuadrat galat (JKG).

$\sum(\widehat{Y}_i - \overline{Y})^2$ adalah jumlah kuadrat regresi (JKR).

Penjabaran untuk JKT akan diperoleh sebagai berikut:

$$\begin{aligned}
\sum_{i=1}^{n}(Y_i - \overline{Y})^2 &= \sum_{i=1}^{n} Y_i^2 - 2\overline{Y}\sum_{i=1}^{n} Y_i + \sum_{i=1}^{n}\overline{Y}^2 \\
&= Y'Y - 2\frac{\sum_{i=1}^{n} Y_i}{n}\sum_{i=1}^{n} Y_i + n(\frac{\sum_{i=1}^{n} Y_i}{n})^2 \\
&= Y'Y - 2\frac{(\sum_{i=1}^{n} Y_i)^2}{n} + \frac{(\sum_{i=1}^{n} Y_i)^2}{n} \\
&= Y'Y - 2\frac{(\sum_{i=1}^{n} Y_i)^2}{n} \qquad (2.5)
\end{aligned}$$

Penjabaran untuk JKG akan diperoleh sebagai berikut:

$$\begin{aligned}
\sum(Y_i - \widehat{Y})^2 &= \sum(Y_i)^2 - 2\sum Y_i\widehat{Y}_i + \sum\widehat{Y}^2 \\
&= Y'Y - 2Y'X\widehat{\beta} + (X\beta)'(X\beta) \\
&= Y'Y - 2Y'X\widehat{\beta} + \beta'X'X\beta \\
&= Y'Y - 2Y'X\widehat{\beta} + \beta'X'X(X'X)^{-1}X'Y \\
&= Y'Y - 2Y'X\widehat{\beta} + \beta'X'X'Y \qquad (2.6)
\end{aligned}$$

Berdasarkan persamaan (2.7) dan (2.8), penjabaran untuk JKR akan diperoleh sebagai berikut:

$$\begin{aligned}\sum(Y_i-\overline{Y})^2 &= Y^{'}Y-(\frac{\sum(Y_i-\overline{Y})^2}{n})-Y^{'}Y+\beta^{'}X^{'}X^{'}Y\\ &= \widehat{\beta'}X'Y-(\frac{\sum(Y_i-\overline{Y})^2}{n})\end{aligned}$$

Rumusan hipotesis untuk menguji parameter regresi secara simultan adalah sebagai berikut:

$H_0:\ \beta=0$ ( keseluruhan koefisien regresi tidak signifikan atau tidak ada peubah bebas yang berpengaruh nyata terhadap $Y$ ),

$H_1:\ \beta\neq 0$ ( paling sedikit terdapat satu peubah bebas yang berpengaruh nyata terhadap $Y$).

Statistika uji yang dapat digunakan untuk menguji parameter regresi secara simultan adalah (R.K.Sembiring) :

$$F_{hitung}=\frac{JKR/p}{JKG/(n-p-1)}=\frac{\widehat{\beta'}X'Y-n\overline{Y}^2/p}{(Y'Y\widehat{\beta'}X'Y)/(n-p-1)} \tag{2.7}$$

Jika $F_{hitung}>F_{p,n-p-1}$, maka tolak $H_0$ yang artinya paling sedikit terdapat satu peubah bebas yang berpengaruh nyata terhadap peubah tak bebas $Y$.

# BAB III

# PEMBAHASAN

## 3.1 Variabel Acak Normal Tersensor

Anggap $Y^*$ berdistribusi normal dengan mean $\mu$ dan variansi $\sigma^2$. Misalkan $y_1^*, y_2^*, ..., y_n^*$ adalah sampel acak berukuran $n$ dari $y^*$. Untuk suatu konstanta $c$, definisikan:

$$y_i = \begin{cases} y_i^*, & \text{jika} \quad y_i^* > c, \\ c, & \text{jika} \quad y_i^* \leqslant c. \end{cases} \tag{3.1}$$

Maka sampel $y_1, y_2, ..., y_n$ dikatakan sampel tersensor dari distribusi normal.

Dari persamaan (3.1) diperoleh

$$\begin{aligned} P(y_i = c) &= P(y_i^* \leqslant c) \\ &= \int_{-\infty}^{c} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2}(\frac{y_i^* - \mu}{\sigma})^2\right] dy_i^*. \end{aligned}$$

Misalkan $\quad z = \frac{y_i^* - \mu}{\sigma} \quad$ maka $\quad \frac{1}{\sigma} dy_i^* = dz \quad$ sehingga

$$\begin{aligned} \mathrm{P}(y_i = c) &= \int_{-\infty}^{\frac{c-\mu}{\sigma}} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2} z^2\right] dz \\ &= \Phi(\frac{c - \mu}{\sigma}) \end{aligned}$$

dimana $\Phi$ adalah fungsi distribusi kumulatif dari distribusi normal standar. Dari persamaan (3.1) juga diperoleh:

$$\begin{aligned} P(y_i = y_i^*) &= P(y_i^* > c) \\ &= 1 - P(y_i^* \leq c) \\ &= 1 - \Phi(\frac{c-\mu}{\sigma}). \end{aligned}$$

Fungsi kepadatan probabilitas dari $y_i$ adalah:

$$\begin{aligned} f(y_i) &= [\Phi(\frac{c-\mu}{\sigma})]^j[\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{y_i-\mu^2}{\sigma})}]^{1-j} \\ &= [\frac{1}{\sigma}\phi(\frac{y_i-\mu}{\sigma})]^{1-j}[\Phi(\frac{c-\mu}{\sigma})]^j \end{aligned} \tag{3.2}$$

dimana $j = 1$ jika $y = c$ dan $j = 0$ untuk selainnya.

Hal ini karena jika $j = 1$ maka persamaan (3.2) menjadi

$$f(y_i) = \Phi(\frac{c-\mu}{\sigma})$$

yang merupakan fungsi kepadatan probabilitas dari $Y_i = c$ dan persamaan (3.2) menjadi

$$f(y_i) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{y_i-\mu^2}{\sigma})^2}$$

yang merupakan fungsi distribusi untuk $Y_i^*$ yang berdistribusi normal dengan mean $\mu$ dan variansi $\sigma^2$.

Dan untuk $j = 0$, persamaan (3.2) menjadi:

$$f(y_i) = \frac{1}{\sigma}\phi(\frac{y_i-\mu}{\sigma})$$

dimana $\phi(\frac{y_i-\mu}{\sigma})$ merupakan fungsi kepadatan dari normal standar.

## 3.2 Model Regresi Tersensor

Model Regresi Tersensor didefinisikan sebagai berikut:

$$\mathbf{Y}_i = \begin{cases} \beta'\mathbf{X}_i + \epsilon_i, & \text{jika} \quad \beta'\mathbf{X}_i + \epsilon_i > 0, \\ 0, & \text{jika} \quad \text{selainnya} \end{cases} \tag{3.3}$$

dimana $\beta = (\beta_1, \beta_2, ..., \beta_k)$ adalah vektor $k \times 1$ dari parameter yang tidak diketahui, $\mathbf{X}_i = (X_1, X_2, ..., X_{ik})$ adalah vektor $k \times 1$ dari konstanta yang diketahui, dan $\epsilon_i = \epsilon_{i1}, \epsilon_{i2}, ..., \epsilon_{ik})$ adalah galat yang independen dan berdistribusi normal dengan mean 0 dan variansi $\sigma^2$. Persoalan dalam regresi adalah untuk mengestimasi $\beta$ dan $\sigma^2$ didasarkan pada $N$ pengamatan untuk $\mathbf{Y}_i$ dan $\mathbf{X}_i$. Model ini pertama kali dipelajari oleh Tobin (1958). Model (3.3) dinamakan model regresi normal tersensor.

Untuk model (3.3), anggap $N_0$ merupakan banyaknya pengamatan yang mana $Y_i = 0$, dan $N_1$ banyaknya pengamatan yang mana $Y_i > 0$. Anggap bahwa $N_1$ pengamatan tidak nol dan $Y_i$ terjadi lebih dahulu. Definisikan:

$$F_i = F(\beta'\mathbf{X}_i, \sigma^2) = \int_{-\infty}^{\beta'\mathbf{X}_i} \frac{1}{\sigma\,(2\pi)^{1/2}} \exp^{-t^2/2\sigma^2} dt \tag{3.4}$$

$$f_i = f(\beta' X_i, \sigma^2) = \frac{1}{\sigma\,(2\pi)^{1/2}} \exp^{-(1/2\sigma^2)(\beta' X_i)^2} \tag{3.5}$$

Misal $t' = \frac{t}{\sigma}$ atau $t = \sigma t'$

$dt = \sigma dt'$ maka persamaan (3.4) menjadi

$$\begin{aligned}
F_i &= \int_{-\infty}^{\beta' \mathbf{X}_i} \frac{1}{\sigma\,(2\pi)^{1/2}} \exp^{-\sigma^2 (t')^2 \big/ 2\sigma^2} \sigma dt' \\
&= \int_{-\infty}^{\beta' \mathbf{X}_i} \frac{1}{(2\pi)^{1/2}} \exp^{-(t')^2 \big/ 2} dt' \\
&= \int_{-\infty}^{\beta' \mathbf{X}_i} \frac{1}{(2\pi)^{1/2}} \exp^{-t^2 \big/ 2} dt \\
&= \Phi_i(\beta' \mathbf{X}_i).
\end{aligned} \tag{3.6}$$

Bentuk baku untuk persamaan (3.4):

$$\Phi_i = F_i \ = \int_{-\infty}^{\beta' \mathbf{X}_i} \frac{1}{(2\pi)^{1/2}} \exp^{-t^2 \big/ 2} dt. \tag{3.7}$$

Dan bentuk baku untuk persamaan (3.5) yaitu:

$$\phi_i = \sigma . f_i = \frac{1}{(2\pi)^{1/2}} \exp^{-(\beta' X_i)^2 \big/ 2\sigma^2}. \tag{3.8}$$

$\phi_i$ dan $\Phi_i$ berturut-turut adalah fungsi kepadatan dan fungsi distribusi dari normal standar yang ditaksir di $\beta' X_i \big/ \sigma$. Misalkan

$$\gamma_i = \frac{\phi_i}{1 - \Phi_i} \tag{3.9}$$

$Y_1' = (Y_1, Y_2, ..., Y_{N_1})$ adalah vektor pengamatan $1 \times N_1$ yang tidak nol pada $Y_i$.

$X_1' = (X_1, X_2, ..., X_{N_1})$ adalah matriks $k \times N_1$ dari nilai $X_i$ untuk $Y_i$ yang tidak nol.

$X_0' = (Y_{N_1+1}, ..., X_N)$ adalah matriks $k \times N_0$ dari nilai $X_i$ untuk $Y_i = 0$.

$\gamma_0^{'} = (\gamma_{N_1+1}, ..., \gamma_N)$ adalah vektor $1 \times N_0$ dari nilai $\gamma_i$ untuk $Y_i = 0$.
Untuk pengamatan $Y_i = 0$, telah kita ketahui bahwa:

$$P(Y_i = 0) = P(U_i < -\beta^{'} X_i) = (1 - F_i) \tag{3.10}$$

Karena

$$\begin{aligned} P(U_i < -\beta^{'} X_i) &= \int\limits_{-\infty}^{-\beta^{'} X_i} f(u)du \\ &= \int\limits_{-\beta^{'} X_i}^{\infty} f(u)du \end{aligned}$$

$$= 1 - F(\beta^{'} X_i) = 1 - F_i \tag{3.11}$$

Untuk pengamatan $Y_i > 0$:

$$\begin{aligned} P(Y_i > 0).f(Y_i|Y_i > 0) &= F_i \frac{f(Y_i - \beta^{'} X_i, \sigma^2)}{F_i} \\ &= \frac{1}{\sigma\sqrt{2\pi}} \exp(-1/2\sigma^2)(Y_i - \beta^{'} X_i)^2 \end{aligned} \tag{3.12}$$

## 3.3 Fungsi Likelihood Model Regresi Tersensor

Dari persamaan (3.10), untuk $Y_i = 0$ atau $Y_i^* \leq 0$ didapatkan

$$\begin{aligned} Pr(Y = 0) &= Pr(Y^* \leq 0) \\ &= Pr(X_i^{'} \beta + u \leq 0) \end{aligned}$$

$$
\begin{aligned}
Pr(Y=0) &= Pr(u \leq -X_i^*\beta) \\
&= \int_{-\infty}^{-X_i'\beta} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2}(\frac{u^2}{\sigma^2})\right]
\end{aligned}
$$

Misalkan:

$u = -\sigma t$ dengan $\sigma > 0$

$du = -\sigma dt$

Perubahan batasnya:

$u = -\infty$ maka $t = \infty$

$u = -X_i'\beta$ maka $t = \frac{-X_i'\beta}{\sigma}$

Dengan mensubstitusi ini maka akan didapatkan:

$$
\begin{aligned}
\mathrm{Pr}(Y=0) &= \int\limits_{-\infty}^{-X_i'\beta} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-1}{2}(\frac{u^2}{\sigma^2})]du \\
&= \int\limits_{-\infty}^{(-X_i'\beta)/\sigma} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-1}{2}(\frac{\sigma^2 t^2}{\sigma^2})] - \sigma dt \\
&= -\int\limits_{-\infty}^{(-X_i'\beta)/\sigma} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-1}{2}(t^2)]dt \\
&= \int\limits_{(-X_i'\beta)/\sigma}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-1}{2}(t^2)]dt \\
&= 1 - \int\limits_{-\infty}^{(X_i'\beta)/\sigma} \frac{1}{\sigma\sqrt{2\pi}} \exp[\frac{-1}{2}(t^2)]dt \\
\mathrm{Pr}(Y=0) &= 1 - \Phi(X_i'\beta/\sigma)
\end{aligned}
$$

Bila pemisalan yang dilakukan adalah $u = \sigma t$, dengan $\sigma > 0$, maka akan didapatkan $Pr(Y, Y_i = 0) = \Phi(-X_i'\beta/\sigma)$

(*) Untuk $Y_i > 0$ atau $Y_i = Y_i^* > 0$ didapatkan:

$$\begin{aligned} Pr(Y, Y_i > 0 &= f(Y_i^*|Y_i^* > 0.Pr(Y^* > 0) \\ &= \frac{f(Y_i^*)}{Pr(Y* > 0}.Pr(Y^* > 0) \\ &= f(Y_i^*) \end{aligned}$$

karena $Y^* \sim N(X_i'\beta, \sigma^2)$ dan juga berlaku $Y_i = Y_i^*$ maka:

$$\begin{aligned} Pr(Y, Y_i > 0) &= \frac{1}{\sigma\sqrt{2\pi}} exp^{\frac{-1}{2}[\frac{(Y_i - X_i'\beta)}{\sigma}]^2} \\ &= \frac{1}{\sigma}\phi((Y_i - X_i'\beta)/\sigma) \\ Pr(Y, Y_i > 0) &= \sigma^{-1}\phi((Y_i - X_i'\beta)/\sigma) \end{aligned}$$

Maka didapatkan fungsi likelihood untuk model regresi tersensor:

$$\begin{aligned} L &= \prod_{y_i=0} Pr(Y, Y_i = 0) \prod_{y_i>0} Pr(Y, Y_i > 0) \\ &= \prod_{y_i=0} [1 - \Phi(X_i'\beta/\sigma)] \prod_{y_i>0} \sigma^{-1}\phi[(Y_i - X_i'\beta)/\sigma] \\ L &= \prod_{y_i=0} (1 - F_i). \prod_{y_i>0} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(Y_i - \beta' X_i)^2}{2\sigma^2}\right] \quad (3.13) \end{aligned}$$

dimana:

$\prod_{y_i=0}$ adalah perkalian dari banyaknya pengamatan dimana $Y_i = 0$ atau $Y_i^* \leq 0$.

$\prod_{y_i>0}$ adalah perkalian dari banyaknya pengamatan dimana $Y_i > 0$ atau $Y_i^* > 0$.

$\phi$, $\Phi$ masing-masing menyatakan fungsi probabilitas densitas normal standar dan fungsi distribusi dari normal standar.

Fungsi Log likelihoodnya adalah

$$\log L = \sum_{y_i=0} \log(1 - F_i) + \sum_{y_i>0} \log(\frac{1}{\sigma\sqrt{2\pi}}) - \sum_{y_i>0} (-1/2\sigma^2)(Y_i - \beta' X_i)^2$$

(3.14)

Dalam persamaan di atas, penjumlahan $\sum\limits_{y_i=0}$ adalah untuk $N_0$ pengamatan untuk setiap $Y_i = 0$ dan penjumlahan $\sum\limits_{y_i>0}$ adalah untuk $N_1$ pengamatan untuk setiap $Y_i > 0$.

## 3.4 Taksiran Parameter Model Regresi Tersensor

Akan diestimasi $\beta$ dan $\sigma^2$ berdasarkan $N$ pengamatan $Y_i$ dan $X_i$. Telah diperoleh fungsi likelihood dari persamaan (3.13)

$$L = \prod_{y_i=0}(1 - F_i).\prod_{y_i>0}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(Y_i - \beta' X_i)^2}{2\sigma^2}\right]$$

dimana:

$\prod_{y_i=0}$ perkalian dari $N_0$ pengamatan dimana $Y_i = 0$

$\prod_{y_i>0}$ perkalian dari $N_1$ pengamatan dimana $Y_i > 0$

Fungsi log-likelihoodnya adalah:

$$\log L = \sum_{y_i=0}\log(1 - F_i) + \sum_{y_i>0}\log(\frac{1}{\sigma\sqrt{2\pi}}) - \sum_{y_i>0}(Y_i - \beta' X_i)^2/2\sigma^2$$

dimana:

$\sum_{y_i=0}$ jumlah dari $N_0$ pengamatan dimana $Y_i = 0$,

$\sum_{y_i>0}$ jumlah dari $N_1$ pengamatan dimana $Y_i > 0$.

Sekarang dapat dituliskan turunan pertama log $L$ terhadap $\beta$ dan $\sigma^2$

yaitu

$$\begin{aligned}
\frac{\partial F_i}{\partial \beta} &= f_i X_i \\
\frac{\partial F_i}{\partial \sigma^2} &= -\frac{1}{2\sigma^2}\beta' X_i f_i \\
\frac{\partial f_i}{\partial \beta} &= -\frac{1}{\sigma^2}\beta' X_i f_i X_i \\
\frac{\partial f_i}{\partial \sigma^2} &= \frac{\left(\beta' X_i\right)^2 - \sigma^2}{2\sigma^4} f_i
\end{aligned} \tag{3.15}$$

Dengan menggunakan persamaan (3.15) diperoleh

$$\frac{\partial logL}{\partial \beta} = -\sum_{y_i=0} \frac{f_i X_i}{1-F_i} + \frac{1}{\sigma^2}\sum_{y_i>0}(Y_i - \beta' X_i)X_i = 0 \tag{3.16}$$

dan

$$\frac{\partial logL}{\partial \sigma^2} = 1/2\sigma^2 \sum_{y_i=0} \frac{\beta' X_i f_i}{1-F_i} - \frac{N_1}{2\sigma^2} + \frac{1}{\sigma^2}\sum_{y_i>0}(Y_i - \beta' X_i)^2 = 0 \tag{3.17}$$

Dengan mengalikan $-\beta'/2\sigma^2$ dalam persamaan (3.17) dan menambahkan hasilnya dalam persamaan (3.18) maka diperoleh:

$$\begin{aligned}
\beta'/\sigma^4 \sum_{y_i>0}(Y_i - \beta' X_i)X_i - \frac{N_1}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{y_i>0}(Y_i - \beta' X_i)^2 &= 0 \\
\beta' \sum_{y_i>0}(Y_i - \beta' X_i)X_i - N_1\sigma^2 + \sum_{y_i>0}(y_i - \beta' X_i)^2 &= 0
\end{aligned}$$

atau dapat dituliskan

$$\begin{aligned}
N_1\sigma^2 &= \beta' \sum_{y_i>0}(Y_i - \beta' X_i)X_i + \sum_{y_i>0}(Y_i - \beta' X_i)^2 \\
&= \left[\sum_{y_i>0}(Y_i - \beta' X_i)\right]\left[\beta' \sum_{y_i>0} X_i + \sum_{y_i>0}(Y_i - \beta' X_i)\right]
\end{aligned}$$

$$N_1\sigma^2 = \sum_{y_i>0}(Y_i - \beta' X_i)Y_i$$

Maka

$$\sigma^2 = \frac{Y_1'(Y_i - X_i\beta)}{N_1} \tag{3.18}$$

Dengan mengalikan persamaan (3.17) dengan $\sigma_1$ diperoleh:

$$\begin{aligned} -\sigma\sum_{y_i=0}\frac{f_i}{X_i} + \frac{1}{\sigma}\sum_{y_i>0}(Y_i - \beta' X_i) &= 0 \\ -X_0'\gamma_0 + \frac{1}{\sigma}X_1'(Y_1 - X_1\beta) &= 0 \end{aligned}$$

$$\begin{aligned} -X_0'\gamma_0 + \frac{1}{\sigma}X_1'Y_1 - \frac{1}{\sigma}X_1'X_1\beta) &= 0 \\ \frac{1}{\sigma}X_1'X_1\beta &= \frac{1}{\sigma}X_1'Y_1 - X_0'\gamma_0 \\ X_1'X_1\beta &= X_1'Y_1 - \sigma X_0'\gamma_0 \end{aligned}$$

Sebagai akibatnya

$$\begin{aligned} \beta &= (X_1'X_1)^{-1}X_1'Y_1 - \sigma(X_1'X_1)^{-1}X_0'\gamma_0 \\ &= \beta_{LS} - \sigma(X_1'X_1)^{-1}X_0'\gamma_0. \end{aligned} \tag{3.19}$$

Persamaan (3.19) memperlihatkan hubungan antara estimator maksimum likelihood untuk $\beta$ dan estimator least square yang diperoleh dari pengamatan tak nol dari $Y$. Telah dibuktikan dalam Amemiya (1973) bahwa estimator dari fungsi maksimum likelihood ini adalah estimator yang konsisten dan asimtotik normal. Solusi persamaan (3.17) dan (3.18) dapat dikerjakan dengan metode iterasi. Kriteria kecocokan model (*goodness of fit*) *least square* untuk model regresi tersensor tidak dapat dilihat secara langsung. Untuk memeriksa kecocokan model digunakan statistik uji *likelihood ratio*.

## 3.5 Uji Hipotesis

Hipotesis hubungan antara variabel dependen **Y** dengan satu atau lebih variabel independen **X** dapat diuji dengan metode Likelihood Rasio. Misalnya banyaknya parameter yang akan ditaksir seluruhnya ada $m$ parameter $\theta_1, \theta_2, \theta_3, ..., \theta_m$. Jumlah parameter yang akan diuji adalah $r$ parameter dengan $r \leq m$.

Hipotesis yang terbentuk adalah:

$$\begin{aligned} H_0 &: \theta 1 = \theta_2 = ... = \theta_r = 0 \\ H_1 &: \exists \theta_1 \neq 0. \end{aligned}$$

Statistik ujinya:

$$\chi^2 = \frac{L(0, \hat{\theta}_{r+1}, \hat{\theta}_{r+2}, ..., \hat{\theta}_m)}{L(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m))} \sim \chi^2_r \tag{3.20}$$

Dengan aturan keputusan, $H_0$ ditolak jika $\chi^2 > \chi^2_r$

## 3.6 Ilustrasi dan Simulasi Regresi Tersensor

Ilustrasi untuk model regresi tersensor adalah pengeluaran rumah tangga untuk konsumsi rokok tiap rumah tangga yang sangat bervariasi, sebagian rumah tangga tidak mengkonsumsi rokok, sedangkan rumah tangga yang lain mengkonsumsi dalam jumlah yang sangat bervariasi. Data didapatkan dari data survey konsumsi/pengeluaran rumah tangga pada SUSENAS 2010 di DKI Jakarta.

| Pemasukan(Rp) | Konsumsi Rokok (Rp) | Lama Pendidikan (Tahun) |
|---|---|---|
| 1.200.000 | 300.000 | 12 |
| 5.500.000 | 0 | 18 |
| 6.000.000 | 1.200.000 | 18 |
| 3.000.000 | 0 | 16 |
| 1.500.000 | 300.000 | 12 |
| 3.500.000 | 0 | 16 |
| 1.300.000 | 0 | 12 |
| 1.200.000 | 600.000 | 12 |
| 1.200.000 | 0 | 12 |
| 1.300.000 | 400.000 | 12 |
| 4.000.000 | 0 | 16 |
| 3.000.000 | 600.000 | 16 |
| 3.500.000 | 900.000 | 16 |
| 5.000.000 | 0 | 18 |
| 3.000.000 | 0 | 16 |

Tabel 3.1 Data Simulasi

Simulasi ini dilakukan untuk mengetahui perbedaan hasil dari pengolahan data tersensor dengan menggunakan metode kuadrat terkecil pada regresi biasa dan pengolahan data dengan menggunakan metode maksimum likelihood pada regresi tersensor. Untuk mengolah data dengan model regresi tersensor digunakan Program SAS 9.1 dan untuk model regresi biasa digunakan model regresi biasa. Hasil perbandingannya terdapat dalam tabel berikut ini.

Tabel Perbandingan Parameter Model Regresi Tersensor dengan Model Regresi Biasa

| Parameter | Regresi Tersensor | Regresi Tersensor | Regresi Biasa | Regresi Biasa |
|---|---|---|---|---|
| | Koefisien | P-Value | Koefisien | P-Value |
| $\beta_0$ | 3.306.035 | 0,2777 | 1.723.000 | 0,298 |
| $\beta_1$ | -313.025 | 0,2773 | -146.982,432 | 0,342 |
| $\beta_2$ | 0,4468 | 0,2917 | 0,251 | 0,281 |
| MSE | 828,498E8 | | 1.553E8 | |

Tabel 3.2 Perbandingan Regresi Tersensor dan Regresi Biasa

Berdasarkan tabel di atas, maka model regresi tersensornya adalah: $y =$ 3.306.035 - 313.025 $x_1$ + 0,4468 $x_2$. Terlihat juga bahwa $\hat{\beta}_0$ pada model regresi tersensor adalah 3.306.035 sedangkan pada model regresi biasa adalah 1.723.000.000. Apabila dihubungkan dengan nilai awal ($\beta_0 = 0$), hal ini menunjukkan bahwa $\hat{\beta}_0$ pada model regresi tersensor lebih mendekati nilai awal dari ($\beta_0 = 0$). Sedangkan pada model regresi biasa sangat menjauhi nilai awal dari ($\beta_0 = 0$). Hal tersebut dapat diartikan pula bahwa penduga model regresi tersensor jauh lebih tepat daripada model regresi biasa.

Penduga $\hat{\beta}_1$ dan $\hat{\beta}_2$ pada model regresi tersensor masing-masing adalah: -313.025 dan 0,4468, sedangkan pada model regresi biasa adalah = -146.982,432 dan 0,251. Apabila dihubungkan dengan nilai awal (($\beta_1 = \beta_2 = 1$)), hal tersebut menunjukkan bahwa penduga parameter dari model regresi tersensor mendekati nilai awal dari (($\beta_1 = \beta_2 = 1$)), sedangkan penduga model regresi biasa menjauhi nilai awal. Berdasarkan sifat ketidakbiasan tersebut, maka dapat dikatakan bahwa model regresi tersensor lebih mendekati nilai awal ($\beta_0 = 0$) dan (($\beta_1 = \beta_2 = 1$)) daripada model regresi biasa.

Berdasarkan tabel di atas, terlihat pula bahwa nilai MSE pada model regresi tersensor adalah $828,498 \times 10^8$ sedangkan pada model regresi biasa adalah $1.553 \times 10^8$. Jadi dapat disimpulkan bahwa model regresi tersensor lebih baik daripada model regresi biasa karena menghasilkan nilai MSE yang jauh lebih kecil.

Tabel Anova untuk pengolahan data dengan metode regresi biasa:

| Model | Sum of Squares | df | Mean Square | F | Sig |
|---|---|---|---|---|---|
| Regression | 2307E9 | 2 | 1.154E9 | 0,754 | 0,487 |
| Residual | 1.812E10 | 13 | 1.510E9 | | |
| Total | 2.042E10 | 15 | | | |

Tabel 3.3 ANOVA Regresi Biasa

Untuk menguji pengaruh masing-masing variabel digunakan uji F untuk model regresi biasa. Pengujian ini dimaksudkan untuk mengetahui pengaruh masing-masing variabel independen secara individual terhadap variabel dependen. Uji ini dilakukan dengan membandingkan nilai F-hitung dengan nilai F-tabel. Bila F-hitung lebih besar dari F-tabel, maka $H_0$ ditolak dan demikian sebaliknya.

Berdasarkan tabel diatas, nilai F-tabel pada tingkat keyakinan $\alpha = 0,05$ dan *degree of freedom* (df) dengan rumus $n-2$ sebesar 13. Untuk $\alpha = 0,05$ nilai F-tabel sebesar 4,67. nilai sig = 0,487 dan nilai $F-hitung = 0,764$, karena $F_{hit} < F_{tabel}$, maka $H_0$ diterima, yang berarti variabel-variabel independen, dalam hal ini jumlah pemasukan per bulan untuk setiap rumah tangga dan lamanya pendidikan dari kepala rumah tangga, secara bersama-sama tidak berpengaruh signifikan terhadap variabel dependen, yaitu pengeluaran untuk konsumsi rokok.

Tabel Anova untuk pengolahan data dengan metode regresi tersensor:

| Effect | DF | Wald Chi-Square | Pr $\leq ChiSq$ |
|---|---|---|---|
| Lamanya Pendidikan (Tahun) | 1 | 1.1804 | 0.2773 |
| Pendapatan per bulan | 1 | 1.1119 | 0.2917 |

Tabel 3.4 ANOVA Regresi Tersensor

Untuk menguji pengaruh masing-masing variabel digunakan uji $\chi$ untuk model regresi biasa. Pengujian ini dimaksudkan untuk mengetahui pengaruh masing-masing variabel independen secara individual terhadap variabel dependen. Uji ini dilakukan dengan membandingkan nilai $\chi_{hitung}$ dengan nilai $\chi_{tabel}$. Bila $\chi_{hitung}$ lebih besar dari $\chi_{tabel}$, maka $H_0$ ditolak dan demikian sebaliknya.

Hipotesis pertama yang diajukan adalah pendapatan per bulan berpengaruh positif dan signifikan terhadap konsumsi rokok suatu rumah tangga. Oleh karena itu, semakin tinggi pendapatan per bulan maka semakin tinggi konsumsi rokok suatu rumah tangga.

Berdasarkan hasil pengolahan data dengan analisis tobit yang tertera pada tabel diatas, nilai $\chi_{tabel}$ pada tingkat keyakinan $\alpha = 0,05$ dan *degree of freedom* (df) dengan rumus $n - k - 1$ sebesar 12, untuk $\alpha = 0,05$ nilai $\chi_{tabel} = 21,03$ dan nilai $\chi_{hitung} = 1,18$, karena $\chi_{hitung} < \chi_{tabel}$, maka $H_0$ diterima, yang berarti variabel-variabel independen, dalam hal ini jumlah pendapatan per bulan dari kepala rumah tangga tidak berpengaruh signifikan terhadap variabel dependen, yaitu pengeluaran untuk konsumsi rokok. Hal ini dikarenakan konsumsi rokok juga dapat dipengaruhi oleh kebiasaan yang tidak dapat dihilangkan tanpa memandang jumlah pendapatan per bulannya.

Hipotesis kedua yang diajukan adalah lamanya pendidikan akhir kepala rumah tangga berpengaruh positif dan signifikan terhadap konsumsi rokok suatu rumah tangga. Oleh karena itu, semakin lama pendidikan akhir kepala rumah tangga maka semakin tinggi konsumsi rokok suatu rumah tangga.

Berdasarkan hasil pengolahan data dengan analisis tobit yang tertera pada tabel diatas,untuk $\alpha = 0,05$ nilai $\chi_{tabel} = 21,03$ dan nilai $\chi_{hitung} = 1,11$, karena $\chi_{hitung} < \chi_{tabel}$, maka $H_0$ diterima, yang berarti variabel-variabel independen, dalam hal ini lamanya pendidikan dari kepala rumah tangga tidak berpengaruh signifikan terhadap variabel dependen, yaitu pengeluaran untuk konsumsi rokok. Hal ini dikarenakan konsumsi rokok juga dapat dipengaruhi oleh lingkungan kerja seseorang tanpa memandang latar belakang pendidikannya.

# BAB IV

# PENUTUP

## 4.1 Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan maka dapat disimpulkan bahwa nilai penduga parameter pada model regresi tersensor berbeda dengan model regresi biasa dan model regresi tersensor menghasilkan nilai MSE yang lebih kecil.

## 4.2 Saran

1. Pada penelitian-penelitian regresi, adanya kesalahan dalam analisis data tersensor sangatlah besar sehingga menimbulkan hasil analisis yang bias dan tidak menggambarkan data seutuhnya. Disarankan untuk menggunakan regresi tersensor, karena metode ini menggambarkan data seutuhnya.
2. Pada skripsi ini hanya dibahas estimasi parameter untuk regresi tersensor. Disarankan untuk menguji hipotesis parameter pada regresi tersensor.

# DAFTAR PUSTAKA


Amemiya, T. 1973. *Regression Analtsis When The Dependent Variable Is Truncated Normal.* Econometrica, 41, 997 - 1016.

Fair, R. 1977. *A Note on Computational of The Tobit Estimator.* Econometrica, 45, 1723 - 1727.

Greene, W.H. 1990. *Econometrics Analysis.* MacMillan Publishing Co., New York.

Maddala, G. 1983. *Limited-Dependent and Qualitative Variables in Econometrics, Cambridge.* Cambridge Unicersity Press.

Myers, R. H., & Milton, J.S. 1989. *A First Course In The Theory Of Linear Statistical Models.* PWS-KENT, Boston.

Sembiring, R.K. 1995. *Analisis Regresi.* ITB, Bandung.

Tobin, J. 1958. *Estimation of Relationships for Limited Dependent Variables.* Econometrica, 26, 24 - 36.

Weisberg, S. 1985. *Applied Linear Regression.* $2^{nd} ed. Wiley, New York.$

# LAMPIRAN-LAMPIRAN

## LAMPIRAN A
## Input SAS I

```
data subset;

input Tobacco Sme_Ed Sme_Salary @@;

if Tobacco eq 0

then lower=.;

else Lower=Tobacco;

datalines;

300.000 12 1.200.000 0 18 5.500.000 1.200.000 18 6.000.000

0 16 3.000.000 300.000 12 1.500.000 0 16 3.500.000

0 12 1.300.000 600.000 12 1.200.000 0 12 1.200.000

400.000 12 1.300.000 0 16 4.000.000 600.000 16 3.000.000

900.000 16 3.500.000 0 18 5.000.000 0 16 3.000.000
;

proc lifereg data=subset outest=OUTEST(keep=_scale_);

model (lower, tobacco) = Sme_Ed Sme_Salary / d=normal;

output out=OUT xbeta=Xbeta;
run;
```

# LAMPIRAN B
## Output SAS I

The LIFEREG Procedure

Model Information

| | |
|---|---|
| Data Set | WORK.SUBSET |
| Dependent Variable | lower |
| Dependent Variable | Tobacco |
| Number of Observations | 15 |
| Noncensored Values | 7 |
| Right Censored Values | 0 |
| Left Censored Values | 8 |
| Interval Censored Values | 0 |
| Name of Distribution | Normal |
| Log Likelihood | -108.9473344 |

| | |
|---|---|
| Number of Observations Read | 15 |
| Number of Observations Used | 15 |

Algorithm converged.

Type III Analysis of Effects

| Effect | DF | Wald Chi-Square | Pr > ChiSq |
|---|---|---|---|
| Sme_Ed | 1 | 1.1804 | 0.2773 |
| Sme_Salary | 1 | 1.1119 | 0.2917 |

Analysis of Parameter Estimates

| Parameter | DF | Estimate | Standard Error | 95% Confidence Limits | | Chi-Square | Pr > ChiSq |
|---|---|---|---|---|---|---|---|
| Intercept | 1 | 3306035 | 3045501 | -2663036 | 9275107 | 1.18 | 0.2777 |
| Sme_Ed | 1 | -313025 | 288115.3 | -877721 | 251670.2 | 1.18 | 0.2773 |
| Sme_Salary | 1 | 0.4468 | 0.4237 | -0.3836 | 1.2772 | 1.11 | 0.2917 |
| Scale | 1 | 658355.7 | 202506.1 | 360276.9 | 1203053 | | |

# LAMPIRAN C
# Input SAS II

```
data predict;

drop lambda _scale_ _prob_;

set out;

if _n_ eq 1 then set outest;

lambda = pdf('NORMAL',Xbeta/_scale_)

/ cdf ('NORMAL', Xbeta/_scale_);

Predict = cdf('NORMAL', Xbeta/_scale_)

 * (Xbeta + _scale_*lambda);

label Xbeta='P-VALUE OF UNCENSORED VARIABLE'

Predict = 'P-VALUE OF CENSORED VARIABLE';

run;

proc print data=predict noobs label;

var tobacco lower sme: xbeta predict;

run;
```

# LAMPIRAN D
## Output SAS II

| Tobacco | lower | Sme_Ed | Sme_ Salary | MEAN OF UNCENSORED VARIABLE | MEAN OF CENSORED VARIABLE |
|---|---|---|---|---|---|
| 300000 | 300000 | 12 | 1200000 | 85848.54 | 307800.01 |
| 0 | . | 18 | 5500000 | 128788.24 | 332049.49 |
| 1200000 | 1200000 | 18 | 6000000 | 352171.10 | 475437.83 |
| 0 | . | 16 | 3000000 | -362075.10 | 120357.37 |
| 300000 | 300000 | 12 | 1500000 | 219878.26 | 387098.59 |
| 0 | . | 16 | 3500000 | -138692.24 | 199106.40 |
| 0 | . | 12 | 1300000 | 130525.11 | 333053.51 |
| 600000 | 600000 | 12 | 1200000 | 85848.54 | 307800.01 |
| 0 | . | 12 | 1200000 | 85848.54 | 307800.01 |
| 400000 | 400000 | 12 | 1300000 | 130525.11 | 333053.51 |
| 0 | . | 16 | 4000000 | 84690.62 | 307161.39 |
| 600000 | 600000 | 16 | 3000000 | -362075.10 | 120357.37 |
| 900000 | 900000 | 16 | 3500000 | -138692.24 | 199106.40 |
| 0 | . | 18 | 5000000 | -94594.62 | 218055.10 |
| 0 | . | 16 | 3000000 | -362075.10 | 120357.37 |

## DAFTAR RIWAYAT HIDUP

**Christina.** Dilahirkan di Jakarta pada tanggal 17 Desember 1987. Anak kedua dari pasangan M. Simanjuntak dan R. Siagian.

**Riwayat Pendidikan.** Pendidikan formal yang yang pernah ditempuh oleh penulis adalah: TK Aliria Kelapa Dua Wetan, Ciracas, Jakarta, lulus tahun 1994. SDK IGN. Slamet Riyadi Cijantung, Jakarta, lulus tahun 1999. Melanjutkan sekolah di SMPK IGN. Slamet Riyadi Cijantung, Jakarta, lulus tahun 2002. Kemudian melanjutkan ke SMA Negeri 39 Cijantung, Jakarta, lulus tahun 2005. Dan pada tahun 2005 penulis melanjutkan ke Universitas Negeri Jakarta, Fakultas Matematika dan Ilmu Pengetahuan Alam (MIPA), Jurusan Matematika, Program Studi Matematika.

**Pengalaman Organisasi**. Penulis pernah mengikuti berbagai kepanitiaan di Persekutuan Mahasiswa Kristen (PMK) UNJ seperti Humas Perlengkapan dan Perpustakaan (Humperpus) tahun 2007 - 2008, anggota panitia Natal, Sekretaris Paskah UNJ, Panitia Dies Natalies UNJ. Tahun 2007, penulis pernah berkiprah di dunia pendidikan dengan menjadi pengajar siswa SD, SMP, dan SMA di berbagai bimbingan belajar. Sejak tahun 2008 sampai sekarang penulis tergabung menjadi anggota guru sekolah minggu HKBP Cibubur Ressort Cibubur.

Bagi para pembaca yang ingin berkomunikasi dengan penulis, dapat mengakses melalui email di `hasian_christina@yahoo.com`.