

**SKRIPSI**

**PERBANDINGAN REGRESI LINEAR BERGANDA DAN  
REGRESI TOBIT UNTUK DATA DENGAN VARIABEL  
DEPENDEN YANG TERSENSOR**



**BILLY**

**NPM: 6161901032**

**PROGRAM STUDI MATEMATIKA  
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS  
UNIVERSITAS KATOLIK PARAHYANGAN  
2023**

**FINAL PROJECT**

**COMPARISON OF MULTIPLE LINEAR REGRESSION AND  
TOBIT REGRESSION FOR DATA WITH CENSORED  
DEPENDENT VARIABLE**



**BILLY**

**NPM: 6161901032**

**DEPARTMENT OF MATHEMATICS  
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES  
PARAHYANGAN CATHOLIC UNIVERSITY  
2023**

# LEMBAR PENGESAHAN

## PERBANDINGAN REGRESI LINEAR BERGANDA DAN REGRESI TOBIT UNTUK DATA DENGAN VARIABEL DEPENDEN YANG TERSENSOR

Billy

NPM: 6161901032

Bandung, 16 Agustus 2023

Menyetujui,

Pembimbing 1



Maria Anestasia, M.Si., M.Act.Sc.

Pembimbing 2



Robyn Irawan, M.Sc.

Ketua Penguji



Agus Sukmana, M.Sc.

Anggota Penguji



Dr. Ferry Jaya Permana

Mengetahui,

Ketua Program Studi



Dr. Livia Owen

## PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

### **PERBANDINGAN REGRESI LINEAR BERGANDA DAN REGRESI TOBIT UNTUK DATA DENGAN VARIABEL DEPENDEN YANG TERSENSOR**

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,  
16 Agustus 2023



Billy

NPM: 6161901032

## ABSTRAK

Penyakit *liver* sering terjadi di berbagai kalangan masyarakat, banyak masyarakat juga yang setelah divonis terkena penyakit *liver* tidak melanjutkan pengecekan medis lagi. Masalah yang terjadi setelah orang terkena penyakit *liver* adalah kemungkinan akan kematian yang disebabkan oleh penyakit tersebut. Untuk mengatasi masalah tersebut akan dilakukan pembentukan model agar dapat memberikan gambaran terbaik mengenai waktu melakukan diagnosa akan penyakit *liver*. Model regresi linear berganda merupakan model yang sangat umum digunakan untuk masalah sehari-hari dengan melibatkan variabel bebas dan terikat. Namun, ada kalanya nilai dari variabel terikat yang akan dimodelkan adalah nilai yang tersensor dan regresi linear belum dapat mengakomodasi masalah ini. Maka dari itu, digunakan regresi tobit dalam penelitian ini karena mampu mengakomodasi masalah tersebut. Variabel terikat dalam penelitian ini berupa waktu dari seorang pasien untuk melakukan diagnosa pertama atas penyakit *liver* yang dideritanya dan karena penelitian ini dilakukan hanya dalam rentang waktu tertentu, maka untuk pasien yang belum melakukan diagnosa pertamanya hingga waktu penelitian selesai, nilai variabel terikatnya dianggap tersensor. Perbedaan utama dari kedua model tersebut terdapat pada metode untuk mengestimasi parameter, di mana regresi linear berganda menggunakan *Ordinary Least Square* dan regresi tobit menggunakan *Maximum Likelihood Estimator*. Penerapan model-model regresi yang digunakan juga dilakukan pada variabel terikat dalam nilai sebenarnya dan nilai yang telah diterapkan *scaling* guna mengurangi persebaran nilai variabel terikatnya yang terlalu luas. Perbandingan hasil kedua model regresi linear berganda dan tobit akan dibandingkan dengan beberapa ukuran, seperti *Mean Square Error* untuk memperoleh model yang terbaik dalam memodelkan waktu diagnosa pertama pasien penyakit *liver*.

**Kata-kata kunci:** Data Tersensor; Penyakit *Liver*; Regresi Linear Berganda; Regresi Tobit; *Ordinary Least Square*; *Maximum Likelihood Estimator*; *Mean Square Error*.

## ABSTRACT

Liver disease is common in many communities, and many people who have been diagnosed with liver disease do not continue medical checks. The problem that occurs after people are diagnosed with liver disease is the possibility of death caused by the disease. To solve this problem, a model will be built in order to provide the best picture of the time to diagnose liver disease. Multiple linear regression models are very common models used for everyday problems involving independent and dependent variables. However, there are times when the value of the dependent variable to be modeled is censored and linear regression cannot accommodate this problem. Therefore, tobit regression is used in this study because it is able to accommodate the problem. The dependent variable in this study is the time for a patient to make the first diagnosis of his/her liver disease and since this study is conducted only within a certain time span, for patients who have not made their first diagnosis until the time of the study, the value of the dependent variable is considered censored. The main difference between the two models is in the method to estimate the parameters, where multiple linear regression uses Ordinary Least Square and tobit regression uses Maximum Likelihood Estimator. The application of the regression models used is also carried out on the dependent variable in the actual value and the value that has been applied scaling in order to reduce the spread of the dependent variable value that is too wide. Comparison of the results of both multiple linear regression and tobit models will be compared with several measures, such as the Mean Square Error to obtain the best model in modeling the first diagnosis time of liver disease patients.

**Keywords:** Censored Data; Liver Disease; Multiple Linear Regression; Tobit Regression; Ordinary Least Square; Maximum Likelihood Estimator; Mean Square Error.

## KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa, atas rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan penulisan skripsi ini yang berjudul “Perbandingan Regresi Linear Berganda dan Regresi Tobit Untuk Data Dengan Variabel Dependen Yang Tersensor” dengan baik dan lancar. Penulis menyadari bahwa dalam perjalanan penulisan skripsi ini, tidak ada yang dapat dicapai tanpa bantuan, dukungan, dan dorongan dari berbagai pihak. Oleh karena itu, penulis ingin mengucapkan rasa terima kasih yang sebesar-besarnya kepada mereka yang telah memberikan kontribusi dalam penyelesaian skripsi ini. Pada kesempatan kali ini, penulis menyampaikan terima kasih setulus-tulusnya kepada:

- Penulis ingin mengucapkan terima kasih yang tak terhingga kepada orang tua dan keluarga penulis. Kasih sayang, doa, dan semangat dari orang tua penulis senantiasa memberikan inspirasi dan motivasi dalam meniti perjalanan akademik ini. Semoga keberhasilan ini dapat menjadi buah dari pengorbanan dan upaya penuh cinta beliau.
- Ibu Maria Anastasia, M.Si., M.Act.Sc. dan Bapak Robyn Irawan, M.Sc. selaku dosen pembimbing yang selalu memberikan bantuan dan nasihat dengan sabar sehingga pembimbing dapat menyelesaikan skripsi ini.
- Bapak Agus Sukmana, M.Sc. dan Bapak Dr. Ferry Jaya Permana selaku dosen penguji yang telah memberikan saran dan kritik sehingga skripsi ini dapat menjadi lebih baik untuk dibaca dan dipelajari.
- Seluruh dosen matematika dan dosen luar matematika yang telah memberikan ilmu dan mengajarkan sopan santun kepada penulis selama kuliah.
- Penulis juga merasa beruntung memiliki sahabat-sahabat sejati yang selalu memberikan dorongan, dukungan, dan tawa dalam setiap suka dan duka. Terima kasih kepada Jason Patrick Halim, Raymond Jayawiguna, Benedict Avner, Frankie Rio, Hendry Chandra, Aloysius Joel, Karel Renaldy, Rio Jonathan, Steven Timotius, Vincentius Marcell, dan semua teman-teman yang senantiasa hadir dan berbagi cerita selama perjalanan kuliah.
- Penulis juga merasa beruntung memiliki teman-teman yang selalu hadir menemani selama masa perkuliahan. Terima kasih kepada Felix Alpha, Evan Felix, Raymond Susanto, Timothy Hoetama, Willy Zoe, Egha Hafidzal, Biqytofa, Michael Immanuel, Aditya Pradipta, Kevin Kartijaya, Luthfi Caprinanda, dan Admiral yang telah membuat masa perkuliahan menjadi lebih indah.
- Tak lupa, ucapan terima kasih juga ditujukan kepada Caroline Marlene, orang yang telah hadir dalam hidup penulis sebagai pendukung setia. Caroline Marlene telah memberikan dukungan moril dan spirit yang tak ternilai harganya, memberikan semangat dalam setiap langkah perjalanan ini.
- Semua dukungan, cinta, dan doa dari pihak-pihak istimewa ini telah menjadi pendorong kuat dalam menyelesaikan skripsi ini. Semoga hasil dari perjuangan ini dapat menjadi amal yang bermanfaat bagi kita semua.

Akhir kata, semoga skripsi ini dapat memberikan manfaat dan kontribusi dalam pengembangan ilmu pengetahuan. Mohon maaf apabila terdapat kekurangan dalam penulisan ini. Semoga Tuhan Yang Maha Esa senantiasa memberikan berkat pada segala upaya yang kita lakukan.

Bandung, 16 Agustus 2023

Penulis



# DAFTAR ISI

<b>KATA PENGANTAR</b>	<b>viii</b>
<b>DAFTAR ISI</b>	<b>x</b>
<b>DAFTAR GAMBAR</b>	<b>xii</b>
<b>DAFTAR TABEL</b>	<b>xiii</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	2
1.3 Tujuan . . . . .	2
1.4 <i>State of the Art</i> . . . . .	2
1.5 Sistematika Pembahasan . . . . .	3
<b>2 LANDASAN TEORI</b>	<b>4</b>
2.1 Data Tersensor . . . . .	4
2.2 Regresi Linear . . . . .	5
2.3 <i>Ordinary Least Square</i> . . . . .	6
2.4 Model Evaluasi . . . . .	8
2.5 Fungsi Sigmoid . . . . .	9
2.6 Algoritma <i>Stepwise</i> . . . . .	10
2.7 Regresi Tobit . . . . .	10
<b>3 REGRESI TOBIT DALAM DATA PENYAKIT <i>liver</i></b>	<b>13</b>
3.1 Deskripsi Data . . . . .	13
3.2 OLS VS Tobit . . . . .	15
3.3 Contoh Penghitungan Regresi . . . . .	16
3.3.1 Penghitungan manual Regresi Linear Berganda . . . . .	16
3.3.2 Penghitungan Manual Regresi Tobit . . . . .	18
3.4 Sistematika Pengerjaan . . . . .	20
<b>4 HASIL DAN PEMBAHASAN</b>	<b>22</b>
4.1 Analisis Data Eksploitatif . . . . .	22
4.2 Penerapan Model Regresi Linear Berganda dan Regresi Tobit . . . . .	24
4.3 Penerapan Model Regresi Linear dan Regresi Tobit dengan Proses <i>Scaling</i> . . . . .	26
4.3.1 Regresi Linear Berganda . . . . .	26
4.3.2 Regresi Tobit . . . . .	28
4.3.3 Interpretasi Hasil . . . . .	30
<b>5 KESIMPULAN DAN SARAN</b>	<b>32</b>
5.1 Kesimpulan . . . . .	32
5.2 Saran . . . . .	32



## DAFTAR GAMBAR

2.1	Data tersensor kanan (gambar kiri) dan data tersensor kiri(gambar kanan) . . . .	4
2.2	Contoh data lengkap . . . . .	5
2.3	Fungsi sigmoid . . . . .	9
3.1	Diagram batang nilai NA pada setiap variabel . . . . .	13
3.2	Diagram alir proses pengolahan data . . . . .	20
4.1	Diagram kotak garis <i>size</i> terhadap waktu diagnosis . . . . .	23
4.2	Diagram kotak garis <i>cirrhosis</i> terhadap waktu diagnosis . . . . .	23
4.3	Diagram kotak garis <i>treatment</i> terhadap waktu diagnosis . . . . .	23
4.4	Diagram sebar umur terhadap waktu diagnosis . . . . .	23
4.5	Diagram sebar pasien melakukan perawatan multidisciplinary pertama terhadap waktu diagnosis . . . . .	23
4.6	Diagram sebar pasien melakukan perawatan pertama terhadap waktu diagnosis . .	23
4.7	Diagram sebar umur dengan <i>scaling</i> terhadap waktu diagnosis . . . . .	24
4.8	Diagram sebar pasien melakukan perawatan multidisciplinary pertama dengan <i>scaling</i> terhadap waktu diagnosis . . . . .	24

## DAFTAR TABEL

3.1	Deskripsi data dan jenis data . . . . .	14
3.2	Contoh data penyakit <i>liver</i> . . . . .	16
3.3	Regresi linear berganda pada contoh data penyakit <i>liver</i> . . . . .	17
3.4	Regresi tobit pada contoh data penyakit <i>liver</i> . . . . .	20
4.1	Regresi linear berganda menggunakan algoritma <i>stepwise</i> . . . . .	24
4.2	Regresi tobit menggunakan algoritma <i>stepwise</i> . . . . .	25
4.3	Regresi linear berganda menggunakan algoritma <i>stepwise</i> dan <i>scaling</i> . . . . .	27
4.4	Regresi tobit menggunakan algoritma <i>stepwise</i> dan <i>scaling</i> . . . . .	29
4.5	Nilai MSE Pada Berbagai Model Regresi . . . . .	31



# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Penyakit *liver* merupakan penyakit peradangan pada *liver* manusia, oleh karena itu kesehatan organ *liver* sangat penting bagi tubuh manusia. *Liver* mengubah zat beracun menjadi nutrisi dan kemudian tubuh menggunakannya untuk mengendalikan hormon dalam tubuh. Penyakit *liver* bisa disebabkan oleh beberapa hal, yaitu infeksi virus, kecanduan alkohol, ada penumpukan lemak di *liver*. Gejala dari penyakit *liver* dapat berbeda-beda tergantung dari penyebab awalnya. Namun, ada beberapa ciri umum dari penyakit *liver*, sebagai contoh mual dan muntah, nyeri perut, penurunan nafsu makan, rasa lelah yang berlebihan, warna urin menjadi gelap, kulit dan mata menjadi kuning, kulit berasa gatal dan mudah memar, pembengkakan di perut, dan pembengkakan di kaki<sup>1</sup>. Setelah memahami beberapa ciri-ciri dan penyebab dari penyakit *liver*, akan dicari tahu tentang bagaimana pasien terdiagnosis pertama kali terkena penyakit *liver*.

Data penyakit *liver* dalam penelitian ini bersifat tersensor, di mana data tersensor adalah data yang waktu kejadiannya tidak diketahui nilainya dengan pasti [1]. Masalah dari data tersensor adalah ketika memodelkan data tersebut, perlu diperhatikan batas-batas nilai dari data tersebut. Batas-batas rentang nilai suatu data menjadi masalah dalam pemodelan statistik klasik, karena umumnya tidak diperhatikan sehingga model tersebut kurang merepresentasikan datanya. Berdasarkan data tersensor tersebut, akan dibandingkan hasil pemodelannya dengan dua macam model regresi, yaitu regresi linear dan regresi tobit. Regresi linear berganda menggunakan pendekatan *ordinary least square* (OLS), sedangkan untuk regresi tobit akan menggunakan pendekatan *maximum likelihood estimator* (MLE). Secara teoretis, OLS adalah metode statistik untuk mengestimasi model regresi linear dengan meminimalkan jumlah kesalahan kuadrat terkecil, sedangkan untuk MLE adalah metode statistik untuk mengestimasi model regresi dengan memaksimalkan fungsi *likelihood*.

Model regresi linear digunakan karena memudahkan penggunaannya dalam menginterpretasikan suatu model dan umumnya menggunakan pendekatan *ordinary least square* (OLS), tetapi metode ini memiliki beberapa kekurangan dikarenakan beberapa observasi memiliki data tersensor [1]. Nilai yang tersensor mengakibatkan hasil dari model regresi linear menjadi bias atau tidak sesuai dengan data yang tersedia. Maka dari itu, dari penelitian sebelumnya oleh James Tobin [2], terdapat suatu model regresi yang dapat mengakomodasi data tersensor, yaitu regresi tobit. Greene [3] mengatakan bahwa penggunaan regresi tobit akan mengurangi efek bias jika dibandingkan dengan data yang dimodelkan menggunakan regresi linear. Pada penelitian ini, penulis membandingkan

---

<sup>1</sup><https://www.alodokter.com/penyakit-liver>

model regresi linear dan regresi tobit dalam memodelkan data penyakit *liver*. Karakteristik dari pasien-pasien penyakit *liver* digunakan sebagai variabel bebas dalam pemodelan ini dengan variabel terikat berupa waktu diagnosis pertama dari pasiennya. Variabel terikat ini memiliki rentang nilai tertentu sehingga termasuk data yang tersensor dan perbandingan kedua model regresi tersebut akan dilakukan untuk memperoleh model terbaik.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, dihasilkan beberapa permasalahan yang akan dibahas dalam makalah ini, yaitu:

1. Bagaimana cara membangun model regresi linear berganda dengan menggunakan pendekatan OLS?
2. Bagaimana cara membangun model regresi tobit dengan menggunakan pendekatan MLE?
3. Bagaimana hasil perbandingan performa model regresi linear berganda dengan regresi tobit?

## 1.3 Tujuan

Setelah mengetahui rumusan masalah, ada beberapa tujuan dari pengamatan yang akan dilakukan untuk mendapat hasil yang terbaik. Berikut adalah tujuan dari skripsi :

1. Mengetahui cara memodelkan data dengan menggunakan model regresi linear berganda.
2. Mengetahui cara memodelkan data dengan menggunakan model regresi tobit.
3. Membandingkan performa model regresi linear berganda dan regresi tobit menggunakan *Mean Square Error*.

## 1.4 *State of the Art*

Pada penelitian sebelumnya tentang penyakit *liver* [4] digunakan regresi linear berganda saja tetapi pada penelitian sekarang akan digunakan regresi tobit juga untuk membandingkan hasil terbaik dari kedua regresi tersebut. Data yang dipakai dalam regresi tobit biasanya adalah data yang bersifat tersensor. Penggunaan regresi tobit menurut Greene [3] akan mengurangi efek bias jika dibandingkan dengan data yang diolah menggunakan regresi linear. Namun tidak dapat dijamin regresi tobit dengan MLE akan mengurangi bias bila dibandingkan dengan OLS sehingga data yang tersensor harus diteliti dahulu kebenarannya. Pada penelitian kali ini waktu menuju diagnosis pertamanya memiliki rentang nilai tertentu sehingga termasuk data yang tersensor dan belum pernah dilakukan pemodelan sebelumnya. Secara teoretis, regresi tobit akan memperoleh hasil yang lebih baik untuk data tersensor dibandingkan dengan regresi linear. Regresi linear berganda mengasumsikan bahwa adanya hubungan linear antara variabel independen dan dependen. Jika asumsi ini terpenuhi maka regresi linear berganda dapat memberikan hasil yang cukup baik. Di sisi lain, regresi tobit digunakan ketika terdapat variabel dependen yang memiliki batas bawah atau atas, sehingga asumsi linearitas mungkin tidak selalu terpenuhi.

## 1.5 Sistematika Pembahasan

Sistematika pembahasan dari penulisan skripsi ini terdiri dari 5 bab sebagai berikut:

1. **BAB I : Pendahuluan**

Pada bab ini berisi latar belakang, rumusan masalah, tujuan, batasan masalah, dan sistematika pembahasan dari skripsi ini.

2. **BAB II : Landasan Teori**

Pada bab ini akan dibahas dasar teori tentang data tersensor, regresi linear berganda, regresi tobit, algoritma *stepwise*, fungsi sigmoid, dan model evaluasi yang akan digunakan dalam melakukan pengujian terhadap data penyakit *liver*.

3. **BAB III : Regresi Tobit Dalam Penyakit *liver***

Pada bab ini akan dibahas metode regresi tobit dan akan dilakukan pengujian terhadap data penyakit *liver*.

4. **BAB IV : Hasil dan Pembahasan**

Pada bab ini akan dibahas penggunaan regresi tobit pada data penyakit *liver* serta pengaruh-pengaruh variabel independen menggunakan perangkat lunak R.

5. **BAB V : Kesimpulan dan Saran**

Pada bab ini akan dibahas kesimpulan dari hasil yang sudah dikerjakan dalam skripsi ini.