

**APLIKASI REGRESI SURVIVAL BUCKLEY-JAMES UNTUK
MENGATASI DATA OUTLIER YANG DISENSOR**

**(Studi Kasus : Data Rekam Medis Pasien *Tuberculosis* Rawat
Inap Tahun 2018-2019 di RSUD Soehadi Prijonegoro Sragen)**

JURNAL



Erzylia Herlin Brilliant

16611077

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN
ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2020**

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul : Aplikasi Regresi Survival Buckley-James Untuk Mengatasi Data Outlier yang Disensor

Nama Mahasiswa : Erzylia Herlin Brilliant

NIM : 16611077

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, 16 Januari 2020

Pembimbing


(Muhammad Hasan Sidiq Kurniawan, S.Si., M.Sc.)

HALAMAN PENGESAHAN

TUGAS AKHIR

**APLIKASI REGRESI SURVIVAL BUCKLEY-JAMES UNTUK
MENGATASI DATA OUTLIER YANG DISENSOR**

Nama Mahasiswa : Erzylia Herlin Brilliant

NIM : 16611077

**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL : 31 JANUARI 2020**

Nama Penguji:

1. Ir. Ali Parkhan, M.T.
2. Mujiati Dwi Kartikasari, S.Si., M.Sc.
3. Muhammad Hasan Sidiq Kurniawan, S.Si., M.Sc.

Tanda Tangan

[Handwritten signatures of the examiners]

البحث العلمي

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Muhammad Sidiq Kurniawan, S.Pd., M.Si., Ph.D.)

ABSTRACT

SURVIVAL BUCKLEY-JAMES REGRESSION APPLICATIONS TO RESOLVE CENSORED OUTLIER DATA

Erzylia Herlin Brilliant
Program Studi Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia
Email:16611077@students.uui.ac.id

Survival analysis or life test analysis is a statistical technique used to analyze data that aims to find out the results of variables that affect an initial event to the end of an event. Linear regression analysis is an analysis that aims to determine the effect and linear relationship of independent variables on the dependent variable. Ordinary linear regression models cannot be used to model survival data because of the existence of right censored data. If forced to use, ordinary linear regression will give less accurate results, because censored data is data obtained from incomplete observations. Multiple Linear Regression Analysis can be used if there are outliers but are unable to provide the right results. Outlier data will be used as censored data. To analyze the censored data Buckley-James Regression analysis was used. In this research, a case study of factors that influence the duration of recovery of a tuberculosis patient is carried out. Factors that are thought to influence the length of time of a Tuberculosis patient's recovery are gender, patient age, complications, other diseases, education, occupation, cases, and diagnosis. From the results of the analysis concluded that the Buckley-James Regression method has a smaller MAPE value compared to the Multiple Linear Regression method. So that the Buckley-James Regression can be more accurately used in survival data containing censored data.

Keywords: *Survival, Multiple Linear Regression, Buckley-James Regression, Outlier, Censored Data, Tuberculosis*

1. Pendahuluan

Ilmu statistika dapat digunakan di berbagai penelitian salah satunya ketahanan hidup pasien. Ketahanan hidup pasien adalah waktu lama hidup pasien dari sakit sampai mengalami event sembuh atau meninggal. Analisis statistika yang digunakan untuk menganalisis tahan hidup disebut analisis survival.

Analisis survival atau analisis uji hidup adalah teknik statistik yang digunakan untuk menganalisis data yang bertujuan untuk mengetahui hasil dari variabel yang mempengaruhi suatu awal kejadian sampai akhir kejadian, yaitu waktu sampai memperoleh kejadian yang dicatat dalam hari, minggu, bulan, atau tahun. Pada analisis survival terkadang terjadi data tersensor (censored data), yaitu ada informasi mengenai waktu ketahanan individu tetapi tidak diketahui secara pasti berapa lama waktu ketahanannya [19].

Tuberculosis merupakan suatu penyakit pada paru-paru yang disebabkan oleh basil *microbacterium tuberculosis* yang merupakan salah satu penyakit saluran pernafasan bagian bawah yang sebagian besar basil tuberculosis masuk ke dalam jaringan paru melalui airborne infection dan selanjutnya mengalami proses yang dikenal sebagai focus primer dari ghon [15]. Tuberculosis adalah penyakit menular yang dapat menyerang siapa saja dan dimana saja. Penyakit tuberculosis (TBC) merupakan masalah yang serius bagi dunia, karena menjadi penyebab kematian terbanyak dibanding dengan penyakit infeksi lain.

Berdasarkan data World Health Organization (WHO), jumlah kasus baru tuberculosis (TBC) pada 2015 mencapai 10,4 juta jiwa meningkat dari sebelumnya yang hanya mencapai 9,6 juta. Adapun jumlah temuan TBC terbesar adalah di India sebanyak 2,8 juta kasus, diikuti Indonesia sebanyak 1,02 juta kasus dan Tiongkok sebanyak 918 ribu kasus. Dalam laporan yang bertajuk Global Tuberculosis Report 2016 itu, angka kematian akibat TBC di Indonesia mencapai 100 ribu jiwa dalam setahun ditambah 26 ribu penderita tuberculosis yang terindikasi HIV positif. Sementara angka kematian dunia yang diakibatkan oleh bakteri *mycobacterium tuberculosis* ini mencapai 1,4 juta jiwa ditambah 390 ribu jiwa penderita yang positif terkena HIV. Sedangkan prevalensi penderita TBC di Indonesia pada 2015 sebesar 395 per 100 ribu populasi dengan angka kematian sebesar 40 per 100 ribu populasi.

Jumlah kasus baru TB di Indonesia sebanyak 420.994 kasus pada tahun 2017 (data per 17 Mei 2018). Berdasarkan jenis kelamin, jumlah kasus baru TBC tahun 2017 pada laki-laki 1,4 kali lebih besar dibandingkan pada perempuan. Bahkan berdasarkan Survei Prevalensi Tuberculosis prevalensi pada laki-laki 3 kali lebih tinggi dibandingkan pada perempuan. Begitu juga yang terjadi di negara-negara lain. Hal ini terjadi kemungkinan karena laki-laki lebih terpapar pada faktor risiko TBC misalnya merokok dan kurangnya ketidakpatuhan minum obat. Survei ini menunjukkan bahwa seluruh partisipan laki-laki yang merokok sebanyak 68,5% dan hanya 3,7% partisipan perempuan yang merokok. Berdasarkan Survei Prevalensi Tuberculosis tahun 2013-2014, prevalensi TBC dengan konfirmasi bakteriologis di Indonesia sebesar 759 per 100.000 penduduk berumur 15 tahun ke atas dan prevalensi TBC BTA positif sebesar 257 per 100.000 penduduk berumur 15 tahun keatas. Berdasarkan survey Riskesdas 2013, semakin bertambah usia prevalensinya semakin tinggi. Kemungkinan terjadi re-aktivasi TBC dan durasi

paparan TB lebih lama dibandingkan kelompok umur di bawahnya. Sebaliknya, semakin tinggi kuintil indeks kepemilikan (yang menggambarkan kemampuan sosial ekonomi) semakin rendah prevalensi TBC. Apabila dilihat menurut tingkat pendidikan menunjukkan prevalensi semakin rendah seiring dengan tingginya tingkat pendidikan. Hal ini menunjukkan bahwa kasus TBC dialami kebanyakan kelompok pasien dengan tingkat pendidikan rendah. Kuintil indeks kepemilikan menunjukkan tidak ada perbedaan antara kelompok terbawah sampai dengan menengah atas. Perbedaan hanya terjadi pada kelompok teratas. Hal ini berarti risiko TBC dapat terjadi pada hampir semua tingkatan sosial ekonomi. Kuintil adalah nilai yang menandai batas interval dari sebaran frekuensi yang berderet.

Menurut Dinas Kesehatan Kabupaten Sragen Penyakit TBC masih menjadi masalah prioritas yang harus ditangani dengan serius. Upaya yang telah dilakukan dalam penanggulangan penyakit TBC sangat berpengaruh terhadap penemuan kasus baru. Di Kabupaten Sragen tahun 2015 untuk jumlah kasus baru TB BTA+ terdapat 339 kasus dengan proporsi penderita berjenis kelamin laki-laki sebesar 65,78% sedangkan penderita berjenis kelamin perempuan sebesar 34,22%. Sedangkan CNR kasus baru BTA+ sebesar 37,54 per 100.000 penduduk. CNR seluruh kasus TB sebesar 48,51 per 100.000 penduduk. Total seluruh kasus TB adalah sebanyak 438 kasus. CNR sendiri adalah kependekan dari Case Notification Rate yaitu angka yang menunjukkan jumlah pasien baru yang ditemukan dan tercatat di antara 100.000 penduduk di suatu wilayah tertentu [28].

Berdasarkan permasalahan di atas maka perlu dilakukan penelitian yang bertujuan mengetahui tentang faktor-faktor apa saja yang mempengaruhi lama sembuh pasien TB dan bagaimana kasus ini dapat dianalisis menggunakan analisis survival. Penelitian akan dilakukan menggunakan analisis uji hidup (survival), yaitu metode Regresi Linear Berganda dan metode Regresi Buckley-James. Pada data yang akan digunakan terdapat beberapa data outlier. Analisis Regresi Linear Berganda dapat digunakan apabila terdapat outlier namun kurang mampu memberikan hasil yang tepat. Data outlier tersebut yang akan dijadikan sebagai data tersensor. Untuk menganalisis data tersensor tersebut digunakan analisis Regresi Buckley-James. Digunakannya analisis Regresi Linear Berganda sebagai metode pembandingan sekaligus untuk menunjukkan keunggulan metode Regresi Buckley-James dalam mengatasi adanya data yang tersensor. Data yang dianalisis merupakan data rekam medis pasien rawat inap yang menderita penyakit Tuberculosis tahun 2018-2019 di RSUD Prijonegoro Sragen.

2. Landasan Teori

Fungsi survival merupakan dasar untuk analisis survival karena probabilitas survival yang diperoleh untuk setiap waktu yang berbeda akan menyediakan informasi penting yang diperoleh dari data survival. Fungsi survival merupakan fungsi monoton turun terhadap waktu. Fungsi survival dinotasikan dengan $S(t)$ yaitu probabilitas satu individu akan bertahan (*survive*) lebih lama daripada waktu t . Misalkan T adalah variabel random non-negatif yang menyatakan waktu sampai dengan terjadinya kegagalan. Maka $f(t)$ menyatakan fungsi densitas probabilitasnya dan $F(t) = P(T < t)$ menyatakan fungsi distribusi kumulatifnya. Komplemen fungsi distribusi kumulatif $F(t)$ adalah fungsi survival $S(t)$. Yaitu

probabilitas bahwa subjek *survive* lebih lama daripada waktu t atau probabilitas bahwa variabel random T melebihi waktu t :

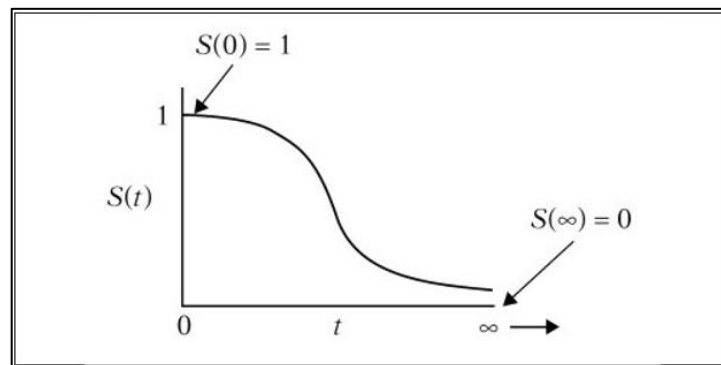
$$\begin{aligned} S(t) &= 1 - F(t) \\ &= P(T > t) \end{aligned} \quad (1)$$

Fungsi densitas $f(t)$ adalah turunan pertama $F(t)$ terhadap t , yaitu:

$$\begin{aligned} f(t) &= \frac{d F(t)}{dt} \\ &= \frac{d}{dt}\{1 - S(t)\} \\ &= -S'(t) \end{aligned} \quad (2)$$

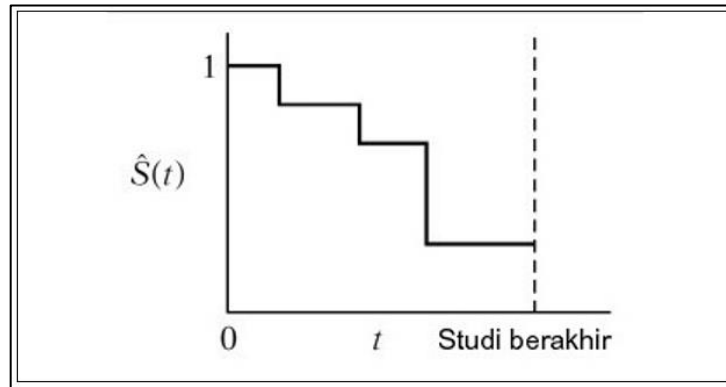
Karakteristik fungsi survival $S(t)$ antara lain adalah:

- Tak membesar : Fungsi survival $S(t)$ mengecil sejalan dengan bertambahnya nilai t .
 - Pada waktu $t = 0$, $S(t) = S(0) = 1$ yaitu pada awal studi belum ada subjek yang mengalami kegagalan atau $P(T > 0) = 1$.
 - Pada waktu $t = \infty$, $S(t) = S(\infty) = 0$, yaitu jika secara teoretis periode studi diperpanjang tanpa batas, suatu saat tidak ada lagi subjek yang *survive*.
- Grafik fungsi survival $S(t)$ terhadap waktu t secara teoretik diperlihatkan pada gambar 1 berikut:



Gambar 1. Grafik Fungsi Survival

Dalam praktiknya karena sifat data sampel dan cara pengumpulan data, estimasi fungsi survival $\hat{S}(t)$ merupakan fungsi bertingkat (gambar 2).



Gambar 2. Estimasi Fungsi $\hat{S}(t)$

Regresi linear (*linear regression*) adalah teknik yang digunakan untuk memperoleh model hubungan antara 1 variabel dependen dengan 1 atau lebih variabel independen. Jika hanya digunakan 1 variabel independen dalam model, maka teknik ini disebut sebagai regresi linear sederhana (*simple linear regression*), sedangkan jika yang digunakan adalah beberapa variabel independen, teknik ini disebut regresi linear berganda (*multiple linear regression*) [12].

Model Regresi Linear Sederhana Model regresi dengan satu variabel bebas X dapat ditulis dalam bentuk persamaan :

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (3)$$

Dimana :

y_i : Variabel Response atau Variabel Akibat (Dependent)

β_0 : Konstanta intercept

β_1 : Konstanta slope

x_i : Variabel Predictor atau Variabel Faktor Penyebab (Independent), $i = 1, 2, \dots, n$

ε_i : Error $\sim N(0, \sigma^2)$, $i = 1, 2, \dots, n$

Model regresi berganda yang paling sederhana adalah model regresi dengan tiga buah variabel, satu variabel dependen dan dua variabel independen. Model ini dikembangkan untuk mengestimasi nilai variabel dependen Y dengan menggunakan lebih dari satu variabel independen ($X_1, X_2, X_3, \dots, X_n$).

Misalnya dalam suatu persamaan regresi berganda yang mempunyai variabel dependen Y dengan dua variabel independen, yakni X_1 dan X_2 . Secara umum persamaan regresi bergandanya dapat ditulis sebagai berikut :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i \quad (4)$$

Dimana :

y_i : Nilai variabel terikat (Dependent)

β_0 : Konstanta intercept

$\beta_1, \beta_2, \dots, \beta_k$: Konstanta slope

$x_{i1}, x_{i2}, \dots, x_{ik}$: Nilai variabel bebas (Independent), $i = 1, 2, \dots, n$

ε_i : Error $\sim N(0, \sigma^2)$, $i = 1, 2, \dots, n$

Buckley dan James (1979) memperkenalkan teknik regresi yang cocok untuk variabel dependen yang disensor. Estimator mereka menggunakan persamaan estimasi kuadrat-terkecil dan mekanisme pembaruan berdasarkan non-parametrik penduga distribusi residual untuk berurusan dengan sensor. Prosedur ini menarik karena penggunaan teknik kuadrat-terkecil memungkinkan untuk interpretasi hasil yang mudah dan penggunaan analisis residual, sementara skema pembaruan cukup umum untuk mengakomodasi berbagai bentuk menyensor dan mengelompokkan.

Terdapat satu variabel dependen dan delapan variabel independen yang digunakan dalam penelitian ini. Beberapa dari variabel independen memiliki kategori.

Tabel 1. Keterangan Variabel

No	Variabel	Keterangan	Satuan Waktu	Kategori
1	Y	Lama waktu rawat inap	Hari	-
2	X1	Usia pasien	Tahun	-
3	X2	Jenis Kelamin	-	1 : Laki-laki 2 : Perempuan
4	X3	Komplikasi	-	1 : Tidak 2 : Ya
5	X4	Penyakit Lain	-	1 : Tidak Ada 2 : Ada 0 : Tidak sekolah / tidak diketahui
6	X5	Pendidikan	-	1 : SD 2 : SMP 3 : SMA 4 : Diploma / Sarjana 0 : Tidak bekerja / tidak diketahui
7	X6	Pekerjaan	-	1 : Petani 2 : Swasta, Wiraswasta, Buruh 3 : PNS, TNI, Polri, Pensiunan 4 : Lain-lain
8	X7	Kasus lama / baru	-	1 : Lama 2 : Baru A : Tuberculosis of Lung, Without Mention of Bacteriological Or Histologica B : Chronic Obstructive Pulmonary Disease, With Acute Exacerbation, Unspecifi
9	X8	Diagnosa	-	C : Tuberculosis Pleurisy, Without Confirmed by Sputum, Microscopy With Or Without

D : Tuberculosis of Lung, Confirmed By
Sputum, Microscopy With Or Without
E : Pneumonia, Unspecified
F : Another

3. Hasil dan Pembahasan

3.1 Analisis Regresi Linier Berganda

Analisis Regresi Linier Berganda biasanya digunakan untuk mengetahui pengaruh antara variabel dependen (Y) dengan variabel independent (X). Pada kasus ini akan diteliti faktor-faktor apa sajakah yang mempengaruhi lama waktu sembuh pasien *tuberculosis*. Faktor-faktor tersebut menjadi variabel independent, terdapat 8 variabel independent yang akan diteliti pengaruhnya secara signifikan. Sedangkan lama waktu sembuh pasien merupakan variabel dependen atau variabel yang dipengaruhi.

3.1.1 Uji Overall

Uji overall merupakan bagian dari uji signifikansi dari analisis regresi linier berganda. Uji ini digunakan untuk mengetahui apakah model regresi layak atau tidak layak digunakan sebagai model peramalan. Hipotesis nol (H_0) dari uji overall ini adalah model regresi tidak signifikan atau tidak layak dan hipotesis satu (H_1) adalah model regresi signifikan atau layak. Dengan menggunakan tingkat kepercayaan sebesar 95% atau $\alpha = 5\%$. Didapatkan hasil p-value sebesar 0.6489 lebih besar dari α maka keputusannya adalah gagal tolak H_0 . Jadi dengan tingkat kepercayaan sebesar 95% didapatkan kesimpulan model regresi tidak layak.

Meskipun pada perhitungan uji overall mendapatkan kesimpulan bahwa model regresi tidak layak digunakan dalam peramalan, model akan tetap digunakan sebagai pembanding model berikutnya. Uji signifikansi berikutnya adalah uji parsial.

3.1.2 Uji Parsial

Uji parsial digunakan untuk mengetahui koefisien regresi dari variabel independent berpengaruh secara signifikan terhadap model atau tidak signifikan. Masing-masing variabel independent akan diuji dan dilihat besarnya *p-value* guna menentukan apakah koefisien regresi tersebut signifikan terhadap model atau tidak.

Dengan menggunakan tingkat kepercayaan 95% didapatkan koefisien regresi yang signifikan terhadap model adalah komplikasi (X3) dan penyakit lain (X4). Sedangkan koefisien lainnya tidak signifikan terhadap model. Kemudian langkah selanjutnya yaitu eliminasi variabel-variabel yang tidak signifikan satu persatu dimulai dengan variabel yang memiliki nilai *p-value* terbesar.

Setelah melakukan proses eliminasi yang berurutan diawali dari variabel yang memiliki nilai *p-value* terbesar.

Tabel 2. Statistik Uji, Keputusan dan Kesimpulan Uji Parsial Regresi Linier Berganda Setelah Proses Eliminasi

Koef	<i>p-value</i>	Tanda	α	Keputusan	Kesimpulan
Intercept	0.0126	<	0.05	Tolak H0	Signifikan
X3 = YA	0.0165	<	0.05	Tolak H0	Signifikan
X4 = YA	0.0117	<	0.05	Tolak H0	Signifikan

Dengan menggunakan tingkat kepercayaan 95% didapatkan hasil semua koefisien telah signifikan terhadap model. Terdapat dua koefisien variabel yang signifikan, yaitu Komplikasi (X3) dan variabel Penyakit lain (X4). Sehingga dapat ditarik kesimpulan koefisien regresi signifikan terhadap model.

3.1.3 Persamaan Model Regresi Linier Berganda

a. Persamaan Model Awal

Model awal regresi semua koefisien regresi dimasukkan. Baik yang signifikan maupun yang tidak signifikan. Didapatkan persamaan model awal regresi linier berganda sebagai berikut :

$$\begin{aligned} \hat{Y} = & 5.8149 - 0.1130(X_1) + 5.3975(X_2 = P) - 16.5676(X_3 = YA) \\ & + 17.8074(X_4 = YA) + 7.1491(X_5 = 0) \\ & + 2.9527(X_5 = 1) + 11.2366(X_5 = 2) \\ & - 0.9864(X_5 = 3) - 1.1156(X_6 = 0) \\ & + 2.5537(X_6 = 1) + 6.0547(X_6 = 2) \\ & + 12.5729(X_6 = 3) - 0.4649(X_7 = L) \\ & + 1.8703(X_8 = B) - 1.9479(X_8 = C) \\ & + 0.2863(X_8 = D) - 0.3841(X_8 = E) \end{aligned}$$

b. Persamaan Model Akhir

Persamaan model akhir didapatkan setelah eliminasi variabel yang tidak signifikan. Hanya koefisien regresi yang signifikan yang dimasukkan kedalam persamaan model. Berikut persamaannya model akhir regresi linier berganda :

$$\hat{Y} = 9.422 - 16.658(X_3) + 17.955(X_4)$$

Menggunakan metode Regresi Linier Berganda dirasa kurang mampu memprediksikan Y dengan baik. Dikarenakan regresi linier biasa tidak dapat mengidentifikasi data yang terdapat outlier. Metode ini menganggap semua data tidak ada yang tersensor. Oleh karena itu, digunakan metode kedua yang dapat mendeteksi adanya data tersensor tersebut yaitu metode Regresi Buckley-James.

3.2 Analisis Regresi Buckley-James

Regresi Buckley-James merupakan model regresi linear untuk data survival. Bertujuan untuk mengetahui faktor-faktor yang secara signifikan berpengaruh pada lama waktu sembuh pasien *tuberculosis* pada kasus ini. Regresi ini dapat mendeteksi data yang mengandung data tersensor. Dari 200 data terdapat 9 diantaranya yang tersensor, maka digunakanlah analisis Regresi Buckley-James untuk mengatasi data tersensor yang terdapat pada data rawat inap pasien *tuberculosis* di RSUD Soehadi Prijonegoro Sragen tahun 2018-2019.

3.2.1 Uji Signifikansi Parameter Regresi

Uji ini digunakan untuk mengetahui variabel-variabel bebas yang berpengaruh secara signifikan didalam model atau tidak. Hipotesis nol (H_0) dari uji ini adalah koefisien variabel independent berpengaruh terhadap waktu survival. Masing-masing variabel independent akan dilihat besarnya *p-value*. Apabila *p-value* kurang α maka H_0 ditolak dan begitu pula sebaliknya.

Dengan menggunakan tingkat kepercayaan 95%, berdasarkan hasil uji signifikansi didapatkan hasil dari semua koefisien variabel independent terdapat 4 diantaranya yang berpengaruh secara signifikan terhadap waktu survival. Yaitu variabel umur (X1), komplikasi (X2), riwayat penyakit lain (X3) dan diagnosis (X8).

Kemudian melakukan proses eliminasi variabel-variabel yang tidak signifikan satu persatu dimulai dengan variabel yang memiliki nilai *p-value* terbesar hingga semua koefisien variabel signifikan. Hasil akhir dari eliminasi adalah sebagai berikut :

Tabel 3. Statistik uji, keputusan dan kesimpulan uji signifikansi parameter Regresi Buckley-James

Koefisien	<i>p-value</i>	Tanda	A	Keputusan	Kesimpulan
<i>Intercept</i>	< 0.0001	<	0.05	Tolak H0	Koefisien variabel independen berpengaruh terhadap waktu survival
X_1	0.0187	<	0.05	Tolak H0	Koefisien variabel independen berpengaruh terhadap waktu survival
$X_3 = Y_a$	0.0139	<	0.05	Tolak H0	Koefisien variabel independen berpengaruh terhadap waktu survival

Koefisien	<i>p-value</i>	Tanda	A	Keputusan	Kesimpulan
$X_4 = Y_a$	0.0042	<	0.05	Tolak H0	Koefisien variabel independent berpengaruh terhadap waktu survival
$X_6 = 0$	0.4802	>	0.05	Gagal Tolak H0	Tidak Signifikan
$X_6 = 1$	0.7338	>	0.05	Gagal Tolak H0	Tidak Signifikan
$X_6 = 2$	0.2136	>	0.05	Gagal Tolak H0	Tidak Signifikan
$X_6 = 3$	0.0285	<	0.05	Tolak H0	Koefisien variabel independent berpengaruh terhadap waktu survival
$X_7 = L$	0.0441	<	0.05	Tolak H0	Koefisien variabel independent berpengaruh terhadap waktu survival
$X_8 = B$	0.1147	>	0.05	Gagal Tolak H0	Tidak Signifikan
$X_8 = C$	0.1041	>	0.05	Gagal Tolak H0	Tidak Signifikan

Koefisien	<i>p-value</i>	Tanda	α	Keputusan	Kesimpulan
$X_8 = D$	0.4163	<	0.05	Gagal Tolak H_0	Tidak Signifikan
$X_8 = E$	0.0419	>	0.05	Tolak H_0	Koefisien variabel independen berpengaruh terhadap waktu survival

Dengan tingkat kepercayaan sebesar 95%, hasil akhir menggunakan Regresi Buckley-James terdapat 4 koefisien variabel independen yang signifikan berpengaruh terhadap waktu survival. Yaitu variabel Umur (X_1), Komplikasi (X_3), Penyakit lain (X_4), dan Diagnosis (X_8).

Pada koefisien variabel diagnosis terdapat enam kategori yaitu A, B,

C, D, E, dan F. variabel tersebut masuk dalam koefisien variabel independent yang signifikan. Hal itu diperkuat oleh adanya dua kategori dalam variabel bebas diagnosis yang memiliki nilai *p-value* kurang dari α . Maka variabel diagnosis masuk kedalam model regresi dan dianggap signifikan atau berpengaruh terhadap waktu survival.

5.1.1 Persamaan Model Regresi Buckley-James

a. Persamaan Model Awal

Model awal regresi semua koefisien regresi dimasukkan. Baik yang signifikan maupun yang tidak signifikan. Didapatkan persamaan model awal regresi Buckley-James sebagai berikut :

$$\begin{aligned} \hat{Y} = & 1.5654 + 0.0052(X_1) + 0.0206(X_2 = P) - 0.3259(X_3 = YA) \\ & + 0.4175(X_4 = YA) + 0.2457(X_5 = 0) - 0.0003(X_5 = 1) \\ & + 0.1960(X_5 = 2) - 0.0162(X_5 = 3) + 0.0513(X_6 = 0) \\ & + 0.0576(X_6 = 1) + 0.1304(X_6 = 2) + 0.5517(X_6 = 3) \\ & + 0.1735(X_7 = L) - 0.2526(X_8 = B) + 0.2193(X_8 = C) \\ & + 0.1645(X_8 = D) + 0.3758(X_8 = E) \end{aligned}$$

b. Persamaan Model Akhir

Persamaan model akhir didapatkan setelah eliminasi variabel yang tidak signifikan. Hanya koefisien regresi yang signifikan yang dimasukkan kedalam persamaan model. Berikut persamaannya model akhir regresi Buckley-James :

$$\begin{aligned} \hat{Y} = & 1.5390 + 0.0070(X_1) - 0.3627(X_3) + 0.4250(X_4) \\ & + 0.1250(X_6 = 0) + 0.0492(X_6 = 1) \\ & + 0.1671(X_6 = 2) + 0.4576(X_6 = 3) \\ & + 0.1997(X_7 = L) - 0.2623(X_8 = B) \\ & + 0.3143(X_8 = C) + 0.1367(X_8 = D) \\ & + 0.3467(X_8 = E) \end{aligned}$$

Bertambahnya satu satuan dari variabel independen akan semakin meningkatkan resiko mengalami event, jika koefisien dari variabel independen bernilai negatif akan mengakibatkan kenaikan satu satuan koefisien tersebut akan semakin mengurangi nilai dari variabel dependen dan juga sebaliknya jika koefisien dari variabel independent bernilai positif akan mengakibatkan kenaikan satu satuan koefisien tersebut akan menambah nilai variabel dependen.

5.2 Perbandingan Metode Regresi Linier Berganda dengan Regresi Buckley-James

Setelah melakukan analisis dari Regresi Linear Berganda dan Regresi Buckley-James akan dilakukan perbandingan model yang terbaik dengan mengetahui nilai MAPE dari masing-masing regresi tersebut. Diperoleh hasil MAPE sebagai berikut :

Tabel 4. Perbandingan MAPE

	Regeresi Linier Berganda	Regresi Buckley-James
MAPE	89.7%	69.8%

Dapat dilihat bahwa hasil MAPE dari Regresi Buckley-James lebih kecil dibanding MAPE Regresi Linier Berganda. Meskipun MAPE Regresi Buckley-James nilainya masih sangat besar namun Regresi Buckley-James lebih baik digunakan dalam kasus ini. Sehingga Regresi Buckley-James adalah metode yang terbaik.

4. Kesimpulan

Berdasarkan hasil dan pembahasan diatas, dapat ditarik beberapa kesimpulan sebagai berikut ini:

1. Terdapat enam faktor-faktor yang berpengaruh secara signifikan terhadap lama waktu sembuh pasien *Tuberculosis* adalah usia, komplikasi, penyakit lain yang diderita pasien, pekerjaan, jenis kasus lama atau baru dan diagnosa.
2. Pada hasil analisis Regresi Linier Berganda diperoleh dua variabel independen yang signifikan. Sedangkan hasil Regresi Buckley-James diperoleh lebih banyak variabel yang signifikan, yaitu terdapat enam variabel independen yang signifikan terhadap waktu survival.
3. Analisis Regresi Linier Berganda kurang mampu memberikan model yang baik dikarenakan hasil perhitungan MAPE regresi linier berganda lebih besar dibanding MAPE pada Regresi Buckley-James. Sehingga dapat

ditarik kesimpulan bahwa analisis regresi linier berganda tidak dapat mengatasi data *outlier* pada penelitian ini.

5. References

- [1] Arum, M., & Kartiningrum, E. (2018). Analisis Survival Penderita HIV/AIDS Di RSUD Prof. dr. Soekandar Kabupaten Mojokerto. *Jurnal Ilmiah Kesehatan Sekolah Tinggi Ilmu Kesehatan Majapahit*, 47-59.
- [2] Bachtiar, R. Y. (2018). *Analisis Lama Waktu Kesembuhan Pasien Diare Dengan Pendekatan Efron Metode Regresi Cox Proportional Hazard*. Yogyakarta: Jurusan Statistika, Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
- [3] Beeching, N., & Gill, G. (2014). *Tropical Medicine*. Liverpool, UK: Wiley Blackwell.
- [4] Buckley, J., & James, I. (1979). Linear Regression With Censored data. *Biometrics*, 66, 429-436.
- [5] Crofton, J., & Norman, H. (1998). *Tuberculosis Klinik*. Jakarta: Widya Medika.
- [6] Cui, J. (2016). Buckley-James Method For Analyzing Censored Data, With An Application To a Cardiovascular Disease And an HIV/AIDS Disease. *The Stata Journal*, 517-526.
- [7] Draper, N. &. (1998). *Applied Regression Analysis*. New York: Wiley.
- [8] Draper, N. a. (1998.). *Applied Regression Analysis, Third Edition*. . Canada: John Wiley and Sons.
- [9] Endah, N. (2015). Evaluasi Pendosisan Gentamisin Pada Paien Anak Pneumonia Berat. *Journal of Management and Pharmacy Practice*, 27-32.
- [10] Farisyi, M. A., & Khambri, D. (2018). Analisis Survival Pasien Kanker Payudara Usia Muda di RSUP DR. M. Djamil Padang Tahun 2008-2017. *Andalas Journal Of Health*, 25-28.
- [11] Feinglod, M. (1993). Choice of Prediction Estimator In Censored Regression Models. *Biometrics*, 49, 661-664.
- [12] Harlan, J. (2018). *Analisis regresi Linear*. Jakarta: Gunadarma.
- [13] Hasyim, M., & Dedy, D. (2017). Analisis Survival Laju Indeks Kinerja Dosen STKIP PGRI Tulungagung Dengan Model Regresi Cox. *Seminar Nasional Matematika dan Aplikasinya*, 289-293.
- [14] Heller, G., & J.S. Simonof. (1990). A Comparisson of Estimators for Regression With a Censored Response Variable. *Biometrics*, 77, 515-520.
- [15] Hood. (2005). *Tuberculosis Paru*. Jakarta: Gramedia.
- [16] Implementasi Cox Proportional Hazard Model Parametrik Pada Analisis Survival (Studi Kasus: Mahasiswa Universitas Internasional Batam). (2017).

- [17] Janez Stare, H. H. (2000). On The Use of Buckley-James Least Squares Regression for Survival Data. *New Approches in Applied Statistics*, 125-134.
- [18] Juridar, R. (2018). *Penerapan Regresi Survival Buckley-James Untuk Obervasi Tersensor Kanan*. Yogyakarta: Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
- [19] Kleinbaum, D. (1996). *Survival Analysis A Self-Learning Text*. Atlanta GA USA: Springer.
- [20] Kutner, M. (1990). *Applied Linear Statistical Models*. Irwin.
- [21] Lisa, N. P., & Purhadi. (2012). Analisis Survival Dengan Model Regresi Cox Study Kasus: Pasien Demam Berdarah Dengue di Rumah Sakit Haji Surabaya. *Jurnal Matematika, [S.l.]*, v. 2, n. 2., 25-32.
- [22] Mallinda, E. (2016, February 1). Profil Penderita Tuberkulosis Paru Yang Di rawat Inap Di Bagian Paru Rumah Sakit Umum Daerah Arifin Achmad Provinsi Riau Periode 01 Januari - 31 Desember 2013. *Jom FK Volume 3 No 1*, pp. 1-12.
- [23] Montgomery, D. &. (1992). *Introduction to Linear Regression Analysis*. New York: Wiley.
- [24] Nirwana, M. B. (2018). Metode Buckley-James Untuk Estimasi Model Regresi Linier Pada Data Tersensor Kanan. *Statistika Vol. 6 No. 1*.
- [25] Pangesti, S. (2016). *Model Linear Terapan*. Tangerang Selatan: Universitas Terbuka. Retrieved from <http://www.pustaka.ut.ac.id/lib/wp-content/uploads/pdfmk/SATS4312-M1.pdf>
- [26] Potter, U. (2000). A multivariate Buckley-James Estimator. *Proceedings of the 6th Tartu Conference, Tartu* (pp. 117-131). Bochum, Germany: Ruhr-Universitat Bochum.
- [27] Pramudita, K. (2017). Analisis Survival Kecepatan Kekambuhan Stroke. *Jurnal Biometrika Dan Kependudukan*, 62-69.
- [28] *Profil Kesehatan Tahun 2015*. (2015). Sragen: Dinas Kesehatan Kabupaten Sragen.
- [29] Rupert Miller, &. J. (1982). Regression With Censored Data. *Biometrika*, 521-531.
- [30] Saboe, A., & et.al. (2015). Cardiac Cachexia and Its Impact on Survival in Heart. *Jurnal Kardiologi Indonesia*, 69-74.
- [31] Sembiring, R. K. (1995). *Analisis Regresi*. Jakarta: ITB.

- [32] Setiawan, B. (2017, Oktober 30). *Teknik Hitung Manual Analisis regresi Linier Berganda Dua Variabel Bebas*. Retrieved from budisetiawan999: <http://budisetiawan999.blogspot.com>
- [33] Sutoyo, D. K. (2010). Pengamatan Pasien Tuberkulosis Paru dengan Multidrug Resistant (TB-MDR) di Poliklinik Paru RSUP Persahabatan. *J Respir Indo Vol. 30 No. 2*, 93-104.
- [34] Ulinuha, M. (2018). *Perbandingan Regresi Hazard Menggunakan Metode Cox Proporsional Hazard dan Lin dan Ying*. Yogyakarta: Fakultas Matematika Dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia.
- [35] Utami, D. T. (2015). *Analisis Data Uji Hidup Pasien Kanker Paru Di RSUP dr Kariadi Semarang Dengan Model Regresi*. Semarang: Jurusan Matematika, Fakultas Matematika Dan Ilmu Pengetahuan Alam. Universitas Negeri Semarang.
- [36] Weisberg, S. (2005.). *Applied Linear Regression, Third Edition*. New Jersey.: John Wiley and Sons.
- [37] Zhou, M. (2007). *All About the Buckley-James Estimator*. Kentucky USA: Department of Statistics, University of Kentucky.