BAB III
LANDASAN TEORI

3.1 Sejarah Tuberkulosis

Pada tanggal 24 Maret 1882, Robert Koch mendemonstrasikan *Mycobacterium tuberculosis* sebagai penyebab penyakit *tuberculosis* (TB). Sejak itu kemajuan pemahaman manusia tentang penyakit ini telah menjadi katalis utama dalam perkembangan ilmu kedokteran modern. Pada tahun 1993, Organisasi Kesehatan Dunia (WHO) menyatakan tuberkulosis sebagai penyakit darurat internasional dan merupakan suatu pernyataan yang belum pernah terjadi sebelumnya. Minat internasional dan pendanaan untuk pengendalian tuberkulosis telah meningkat sejak saat itu yang mengarah pada estabilisasi. Ketika mendekati dekade ketiga abad ke-21, ditemukannya pengetahuan tentang tes melokular otomatis genom lengkap yang memiliki sensitivitas dan sensitivitas yang belum pernah terjadi sebelumnya untuk penyakit paru-paru. Jumlah global kasus TB baru per kapita telah menurun secara perlahan sejak tahun 2002, sehingga diperkirakan ada 8,8 juta kasus baru pada tahun 2010. Meskipun jumlah absolut kasus TB baru telah menurun sejak 2006, sebagian besar dari 1,1 juta TB di dunia kematian per tahun terjadi di antara populasi miskin di negara berkembang di daerah tropis.

3.2 Penyakit Tuberkulosis

Tuberkulosis atau *Tuberculosis* disingkat TBC atau TB singkatan dari *Tubercle bacillus* merupakan penyakit menular yang umum dan banyak kasus yang bersifat mematikan. Penyakit ini disebabkan oleh berbagai strain mikobakteria, jenis bakteri yang paling sering menyerang manusia adalah *Mycobacterium tuberculosis* disingkat MTb atau MTbc. Kuman *Mycobacterium tuberculosis* ini merupakan penyebab utama penyakit *tuberculosis* di seluruh dunia sekalipun ada beberapa jenis basilus lainnya yang dapat menyebabkan terjadinya *tuberculosis*. Tuberkulosis biasanya menyerang organ pernapasan yaitu paru-paru tetapi juga bisa berdampak pada bagian tubuh lainnya. Tuberkulosis menyebar melalui udara ketika seseorang dengan infeksi *tuberculosis* aktif batuk, bersin, atau menyebarkan butiran ludah mereka melalui udara. Infeksi *tuberculosis* umumnya bersifat asimtomatik dan laten. Namun hanya satu dari sepuluh kasus infeksi *tuberculosis* laten yang dapat berkembang menjadi *tuberculosis* aktif. Bila Tuberkulosis tidak diobati maka lebih dari 50% orang yang terinfeksi bisa meninggal.

3.3 Penyebab Tuberkulosis

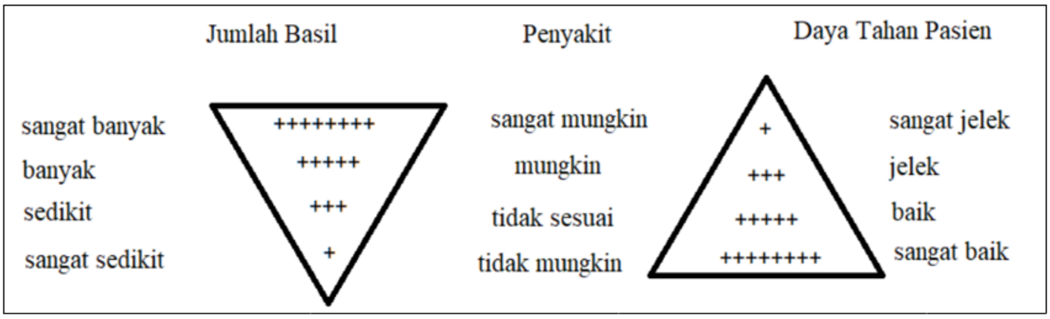
Penyebab utama penyakit tuberkulosis adalah *Mycobacterium tuberculosis*, yaitu sejenis basil aerobik kecil yang non-motil. Berbagai karakter klinis unik patogen ini disebabkan oleh tingginya kandungan lemak atau lipid yang dimilikinya. Sel-selnya membelah setiap 16 –20 jam. Kecepatan pembelahan ini termasuk lambat bila dibandingkan dengan jenis bakteri lain yang umumnya membelah setiap kurang dari satu jam. Mikobakteria memiliki lapisan ganda membran luar lipid. Bila dilakukan uji pewarnaan Gram, maka MTB akan menunjukkan pewarnaan Gram+ yang lemah atau tidak menunjukkan warna sama sekali karena kandungan lemak dan asam mikolat yang tinggi pada dinding selnya. MTB bisa tahan terhadap berbagai disinfektan lemah dan dapat bertahan hidup dalam kondisi kering selama berminggu-minggu. Di alam, bakteri hanya dapat berkembang dalam sel inang organisme tertentu, tetapi *Mycobacterium tuberculosis* bisa dilakukan di laboratorium.

Dengan menggunakan pewarnaan histologis pada sampel dahak yang diekspektorat, peneliti dapat mengidentifikasi MTB melalui mikroskop dengan pencahayaan biasa. MTB mempertahankan warna meskipun sudah diberi perlakukan larutan asam, sehingga dapat digolongkan sebagai Basil Tahan Asam (BTA). Dua jenis teknik pewarnaan asam yang paling umum yaitu teknik pewarnaan Ziehl-Neelsen (ZN) yang akan memberi warna merah terang pada bakteri BTA bila diletakkan pada latar biru dan teknik pewarnaan auramin-rhodamin lalu dilihat dengan mikroskop fluoresen.

Bakteri Kompleks *Mycobacterium tuberculosis* atau disingkat KMTB juga termasuk mikobakteria lain yang juga dapat menjadi penyebab *tuberculosis* ialah *Mycobacterium bovis*, *Mycobacterium africanum*, *Mycobacterium canetti*, dan Mycobacterium microti. Jenis *Mycobacterium africanum* tidak menyebar luas, tetapi merupakan penyebab penting tuberkulosis di sebagian wilayah Afrika. Jenis mikrobakterium ini sedikit berbeda, tipe ini resisten terhadap pengobatan tiasetazon. *Mycobacterium bovis* dapat menyebabkan infeksi pada manusia melalui susu sapi. Kasus tuberkulosis pada hewan ternak ditemukan di Eropa dan Amerika. Pemberantasan tuberculosis pada hewan ini dilakukan dengan membunuh ternak yang terinfeksi dan pasteurisasi susu sapi. Kasus infeksi pada manusia akibat kuman *Mycobacterium bovis* jarang ditemukan di Asia, karena masyarakat disana memiliki kebiasaan memasak susu sebelum diminum.

Mycobacterium non-tuberculosis merupakan spesies yang banyak ditemukan, tetapi tidak berbahaya dan jarang menimbulkan penyakit. Akan tetapi pada uji *tuberculin* kadang-kadang menunjukkan hasil positif lemah. Penyakit karena bakteri ini banyak ditemukan di Amerika dan Australia. Penyakit akibat bakteri jenis ini juga sering dijumpai pada penderita HIV positif. *Mycobacterium non-tuberculosis* resisten terhadap obat-obatan yang biasa digunakan sehingga sukar untuk disembuhkan.

Pada masa lalu saat *tuberculosis* berkembang di negara industri, pemeriksaan uji *tuberculin* dapat dikerjakan pada usia dewasa muda yang terinfeksi tetapi hanya kurang lebih 10% yang kemudian berkembang menjadi infeksi. Perkembangan penyakit ini tergantung dari virulensi kuman dan daya tahan tubuh penderita (Gambar 3.1).



Sumber : Buku Tuberkulosis Klinik Tahun 1998

Gambar 3.1. Pengaruh Jumlah Infeksi Bakteri dan Kekuatan Daya Tahan Tubuh

3.4 Proses Infeksi Bakteri Tuberkulosis

Biasanya peristiwa berkembangnya proses infeksi antara 5% dan 10% orang akan menjadi penderita TB aktif setelah terinfeksi MTB. Sekitar 3% telah mengembangbiakan bakteri dalam tahun pertama dengan mengingat bakteri berkembang dengan frekuensi yang semakin berkurang sesudahnya. Oleh karena itu, lebih dari 90% infeksi MTB tidak menyebabkan infeksi yang berkembang menjadi penyakit dalam rentang hidup manusia yang normal.

Secara klinis manifestasi TB di antara mereka yang menghasilkan penyakit TB aktif tergantung pada dua hal: keadaan sistem kekebalan dan lokasi sebagian besar multiplikasi MTB. Dalam kasus-kasus di mana penyakit ini terjadi segera setelah infeksi primer, basil berkembang biak dan menyebar dalam konteks sistem kekebalan tubuh yang lemah. Oleh karena itu, bentuk penyakit primer terjadi pada situs toraks multiplikasi awal yang umum karenanya radang selaput dada memanjang dari fokus alveolar dan kavitasi pada kelenjar getah bening hilar. Mereka juga cenderung menyebar ke banyak tempat termasuk sistem saraf pusat seperti tuberkulosis meningitis dan tuberkulosis *miliary*. Pada penyakit yang menyebar ke organ-organ tubuh lain terjadi ketika mini-granulomata (*tubercle*) berkembang di sekitar sejumlah kecil basil yang disebarkan secara luas dalam jaringan.

Dalam kasus-kasus di mana penyakit ini terjadi infeksi linier luar waktu bersamaan, baik sebagai reaktivasi infeksi laten atau sebagai hasil dari infeksi ulang dengan strain baru MTB, basil berkembang biak dalam konteks sistem kekebalan tubuh yang sensitif. Respons DTH yang terkait cenderung mengarah pada kerusakan jaringan di tempat multiplikasi sehingga terjadi luka pada lesi kavitasi di mana sejumlah besar basil multiplikasi terkandung oleh tepi sel raksasa dan granulomata yang melingkar merupakan ciri khas patologi tuberkulosis. Lesi postprimary ini paling umum di apeks paru-paru, teorinya adalah bahwa lokasi ini memberikan kombinasi ventilasi dan perfusi yang paling kondusif untuk latensi yang lebih lama dan lebih lama. Mereka juga dapat terjadi di mana pun yang diunggulkan oleh basil selama penggandaan awal sekitar waktu infeksi primer.

Dengan waktu yang cukup lama penyakit *tuberculosis* tipe primer di paru-paru cenderung menghasilkan hubungan antara patologi kavitasi dan saluran pernapasan. Basil MTB kemudian dapat mengalami aerosolisasi dalam inti tetesan dan dikeluarkan ke udara ketika individu yang sedang batuk, bersin atau berbicara. Oleh karena itu, pasien dengan penyakit paru yang mengalami kavitasi merupakan sumber utama infeksi MTB baru.

3.5 Mekanisme Penularan Bakteri Tuberkulosis

Sumber penularan adalah pasien TB BTA+ melalui sputum. Sputum adalah mukus atau dahak yang keluar saat batuk dari saluran pernapasan atas. Batuk dan meludah akan menyebabkan kuman *tuberculosis* menular pada orang lain melalui perantara udara. Ketika seseorang yang positif mengidap *tuberculosis* paru aktif sedang batuk, bersin, berbicara, bernyanyi maupun meludah, mereka sedang menyemburkan bakteri ke udara dengan diameter 0.5 hingga 5 µm. Bersin dapat melepaskan partikel kecil-kecil hingga 40,000 partikel. Setiap partikel dapat menularkan penyakit tuberkulosis karena dosis seseorang dapat terinfeksi penyakit ini sangat rendah. Seseorang yang menghirup kurang dari 10 bakteri saja bisa langsung terinfeksi *tuberculosis*.

Penderita dengan hasil pemeriksaan sputum yang positif mengandung bakteri *tuberculosis* lebih mudah menginfeksi orang lain dibandingkan dengan yang negatif hasil kulturnya. Ibu hamil yang telah terinfeksi *tuberculosis* berisiko tinggi akan melahirkan bayi yang positif terdapat bakteri *tuberculosis* dalam tubuhnya. Keluarga yang tinggal dekat bersama penderita memiliki peluang lebih tinggi untuk tertular penyakit ini. Orang-orang yang melakukan kontak dalam waktu lama dan dalam jangka frekuensi yang cukup sering, atau selalu berdekatan dengan penderita *tuberculosis* berisiko tinggi ikut terinfeksi dengan perkiraan angka infeksi sekitar 22%. Seseorang dengan tuberkulosis aktif dan tidak mendapatkan perawatan dapat menginfeksi 10-15 orang lain setiap tahunnya. Biasanya hanya mereka yang menderita *tuberculosis* aktif yang dapat menularkan penyakit ini. Penderita harus dilatih untuk menutup mulutnya dan menghadapkan wajah kea rah lain saat batuk atau dapat memakai masker penutup mulut dan hidung.

3.6 Diagnosis Tuberkulosis

Melakukan diagnosis tuberkulosis aktif jika hanya berdasarkan tanda-tanda dan gejala saja tidaklah mudah. Sulit juga mendiagnosis penyakit ini pada orang-orang dengan daya tahan tubuh yang sangat lemah. Meski demikian, orang-orang yang menunjukkan tanda-tanda bahwa mereka memiliki penyakit paru-paru atau gejala lainnya yang berlangsung lebih dari dua minggu maka bisa jadi orang tersebut tertular *tuberculosis*. Gambar sinar X dada dan pembuatan beberapa kultur sputum untuk basil tahan asam biasanya menjadi salah satu bagian evaluasi awal. Apabila dicurigai seseorang tertular penyakit *tuberculosis*, maka beberapa hal yang perlu dilakukan untuk mendapatkan diagnosis adalah:

- Anamnesa atau pemeriksaan awal oleh dokter baik terhadap pasien maupun keluarganya.
- Pemeriksaan fisik pasien.
- Pemeriksaan dengan mengambil sampel darah dan dahak untuk diuji di laboratorium.
- Pemeriksaan patologi anatomi (PA).
- Rontgen dada (*thorax photo*).
- Uji tuberkulin.

3.7 Pencegahan dan Penanganan

Tuberkulosis atau TB merupakan penyakit menular berbahaya yang dapat merusak jaringan paru-paru. Meskipun termasuk dalam kategori penyakit menular, terdapat berbagai cara untuk mencegah terjangkitnya penyakit *tuberculosis* ataupun mengobati penyakit tersebut.

a. Vaksin BCG

Vaksin *Bacillus Calmette-Guerin* (BCG) telah tersedia sejak tahun 1920-an dan merupakan vaksin untuk *tuberculosis* yang paling umum digunakan di dunia, dengan lebih dari 90% anak-anak yang mendapat vaksinasi ini. Walaupun BCG lebih efektif melawan penyakit yang menyebar pada masa kanak-kanak, masih terdapat perlindungan yang konsisten terhadap *tuberculosis* paru dan teruji efektif untuk mencegah *tuberculosis* sampai seseorang berusia 35 tahun.

b. Terapi Pencegahan *Isoniazid*

Alasan di balik terapi pencegahan *isoniazid* adalah untuk memberantas infeksi laten sebelum berkembang menjadi penyakit aktif. *Tuberculosis* laten merupakan kelompok orang yang sudah terserang bakteri *Mycobacterium tuberculosis*, namun tidak dapat dibuktikan secara klinis maupun mikrobiologis. Beberapa uji coba terkontrol plasebo pada pasien HIV- yang terinfeksi MTB menunjukkan bahwa *isoniazid* setiap hari yang diberikan selama 6-12 bulan secara substansial mengurangi risiko penyakit tuberkulosis selanjutnya. Namun terapi preventif *isoniazid* belum diakui sebagai pendekatan universal yang efektif dari segi biaya untuk pengendalian *tuberculosis*, tetapi lebih difokuskan pada orang-orang yang berisiko tinggi mengembangkan penyakit aktif. Orang-orang tersebut biasanya diidentifikasi dengan tes kulit dan merupakan kontak dari kasus indeks BTA+ yang diketahui atau orang-orang yang terpapar infeksi secara pekerjaan (seperti perawat dan dokter).

c. Strategi DOTS

Directly Observed Treatment Short-course (DOTS) adalah strategi pengawasan langsung pengobatan jangka pendek dan merupakan strategi untuk pengendalian *tuberculosis* yang memiliki tujuan untuk mengurangi mortalitas, morbiditas dan penularan penyakit dan untuk mencegah perkembangan resistansi obat. *Direct attention* adalah usaha menemukan penderita TB. *Observed* adalah pengawasan terhadap penderita dalam menelan obat. *Treatment* adalah pengobatan jangka pendek. Dalam melakukan pengobatan, pasien TB memerlukan Pengawas Menelan Obat (PMO). Pengawasan ini sangat penting mengingat jangka waktu yang lama dan jenis obat yang banyak membuat seringkali pasien TB tidak patuh menjalani pengobatan dan akibat dari putus obat ini kuman resisten dan pengobatan harus diulang. Strategi yang direkomendasikan untuk memenuhi tujuan ini adalah untuk menyediakan kemoterapi jangka pendek yang terstandarisasi di bawah pengamatan langsung selama fase awal pengobatan untuk paling tidak semua kasus TB BTA+ yang teridentifikasi (sumber infeksi). Keberhasilan strategi ini tergantung pada implementasi paket lima poin, yaitu :

1. Mikroskopi langsung untuk deteksi kasus di antara pasien simtomatik yang melaporkan diri ke layanan kesehatan.
2. Pengamatan terapi untuk pemberian kemoterapi jangka pendek terstandarisasi, untuk memastikan kepatuhan.
3. Pemantauan pengobatan melalui sistem pencatatan dan pelaporan yang terstandarisasi, memungkinkan penilaian berkelanjutan terhadap hasil perawatan.
4. Kemoterapi jangka pendek melalui sistem pasokan obat reguler dari semua obat antituberkulosis esensial, yang harus bebas untuk pasien pada titik pengiriman.
5. Komitmen pemerintah atau organisasi non-pemerintah (LSM) untuk memastikan pendekatan berkelanjutan terhadap kebijakan dan pendanaan.

Kegiatan pengendalian *tuberculosis* harus bertujuan untuk memenuhi dua target minimum: tingkat kesembuhan 85% dan tingkat deteksi kasus 70%. Kategori pelaporan hasil standar yang dijelaskan di atas memungkinkan untuk melakukan analisis kohort triwulanan terhadap hasil pengobatan dan karenanya melaporkan tingkat kesembuhan.

d. Pengobatan Menggunakan Antibiotik

Pengobatan *tuberculosis* menggunakan antibiotik dilakukan untuk membunuh mikrobakteri yang terdapat pada paru-paru yang terinfeksi *tuberculosis*. Pengobatan kasus *tuberculosis* yang efektif ternyata bukan hal yang mudah dikarenakan mikobakteri penyabab *tuberculosis* memiliki struktur dan komposisi kimia dinding sel yang tidak biasa. Dinding sel dari mikrobakteri ini mampu menahan obat masuk sehingga menyebabkan antibiotik tidak efektif. Terdapat dua jenis antibiotik yang sering digunakan adalah *isoniazid* dan *rifampicin* pengobatan dapat berlangsung selama berbulan-bulan. Penyakit *tuberculosis* aktif sebaiknya diobati dengan kombinasi beberapa antibiotik untuk menurunkan risiko berkembangnya bakteri yang resisten terhadap antibiotic yang pernah digunakan sebelumnya. Untuk pengobatan *tuberculosis* laten biasanya menggunakan antibiotik tunggal. Pasien dengan infeksi laten juga diobati untuk mencegah munculnya *tuberculosis* aktif di kemudian hari.

3.8 Statistika Deskriptif

3.8.1 Pengertian Statistika Deskriptif

Sugiyono (2007) menjelaskan statistik deskriptif adalah statistik yang digunakan untuk menggambarkan atau menganalisa suatu statistik hasil penelitian, tetapi tidak digunakan untuk membuat kesimpulan yang lebih luas (generalisasi/ inferensi). Lebih lanjut dijelaskan Sugiyono bahwa penelitian yang tidak menggunakan sampel, maka analisisnya akan menggunakan statistik deskriptif. Demikian juga dengan penelitian yang menggunakan sampel tetapi peneliti tidak bermaksud untuk membuat kesimpulan untuk populasi dari mana sampel diambil, maka statistik yang digunakan adalah statistik deskriptif. (Sugiyono, 2007)

3.8.2 Ruang Lingkup Kajian Statistik Deskriptif

Ruang lingkup kajian pada analisis statistik deskriptif dijelaskan Djarwanto dan Subagyo (1998) yaitu:

a. Distribusi frekuensi serta pengukuran nilai-nilai statistiknya seperti pengukuran nilai sentral, dispersi, skewness dan kurtosis, dan grafiknya seperti poligon, histogram dan ogive.
b. Angka indeks.
c. Time series atau deret waktu.
d. Koefisien regresi dan koefisien korelasi sederhana.

Hal senada dijelaskan Supardi (2013) mengenai ruang lingkup kajian statistik deskriptif yaitu:

a. Penyajian data dalam bentuk tabel seperti tabel tunggal, tabel kontigensi maupun tabel distribusi frekuensi.
b. Penyajian data dalam bentuk grafik seperti diagram batang, diagram garis, diagram lingkaran, diagram pencar, diagram peta, diagram simbol

maupun diagram yang disajikan dari tabel distribusi frekuensi yaitu histogram, poligon frekuensi dan ogive.

c. Ukuran nilai pusat dan letak, seperti rerata, median, modus, varian, simpangan baku, kuartil, desil, persentil.
d. Ukuran dispersi atau simpangan seperti jangkauan atau rentang, rerata simpangan, variansi, simpangan baku.
e. Model distribusi data yaitu kemencengan dan keruncingan kurva distribusi.
f. Angka indeks.
g. Times series/deret waktu atau data berkala.

3.9 Outlier

"*Outlier* adalah kasus atau data yang memiliki karakteristik unik yang terlihat sangat berbeda jauh dari observasi-observasi lainnya dan muncul dalam bentuk nilai ekstrim, baik untuk sebuah variabel tunggal maupun variabel kombinasi" (Ghozali, 2009). Menurut Hampel, Rousseeaw dan Stahel, sebagaimana dikutip oleh Olive (2006) mendefinisikan *outlier* adalah observasi yang menyimpang dari pola yang terbentuk oleh sebagian besar data.

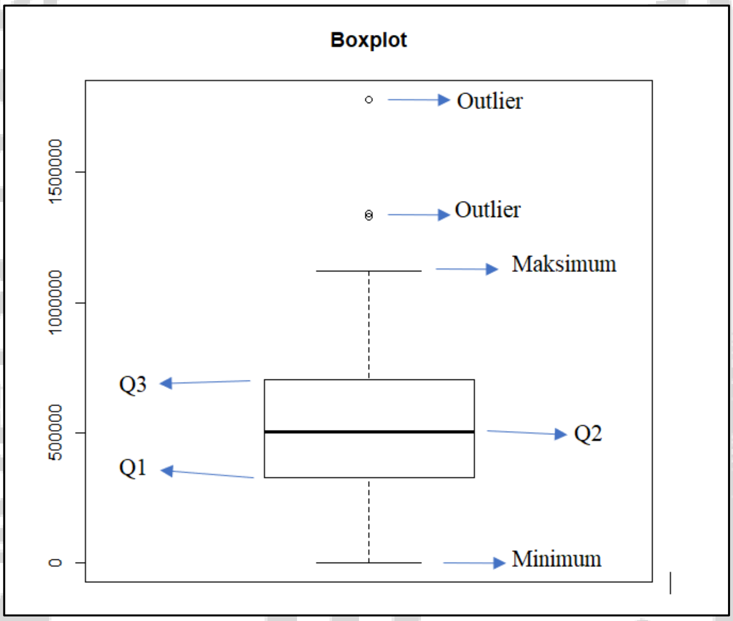
Menurut Ghozali (2009) terdapat empat penyebab timbulnya data *outlier* antara lain :

1. Kesalahan dalam memasukan data.
2. Gagal dalam menspesifikasi adanya missing value dalam program computer.
3. *Outlier* bukan merupakan anggota populasi yang di ambil sebagai sampel.
4. *Outlier* berasal dari populasi yang di ambil sebagai sampel, tetapi ditribusi dari variabel dalam populasi tersebut memiliki nilai ekstrim serta tidak terdistribusi secara normal.

3.9.1 Pengujian Outlier

Ada tidaknya *outlier* pada suatu data dapat diketahui dengan melakukan pengujian. Terdapat beberapa cara pengujian data *outlier* salah satunya dapat menggunakan *boxplot*. *Boxplot* adalah sebuah diagram berbentuk *box* atau persegi yang di dalamnya memuat ringkasan distribusi sampel yang disajikan secara grafis yang bisa menggambarkan bentuk distribusi data, ukuran pusat dan ukuran penyebaran data.

Terdapat lima ukuran statistik yang bisa dibaca dari *boxplot* yaitu nilai minimum (nilai data observasi terkecil), Q1 (kuartil terendah atau kuartil pertama), Q2 (kuartil kedua atau median), Q3 (kuartil tertinggi atau kuartil ketiga), nilai maksimum (nilai data observasi terbesar). Serta dapat menunjukkan ada tidaknya nilai *outlier* dari data observasi.

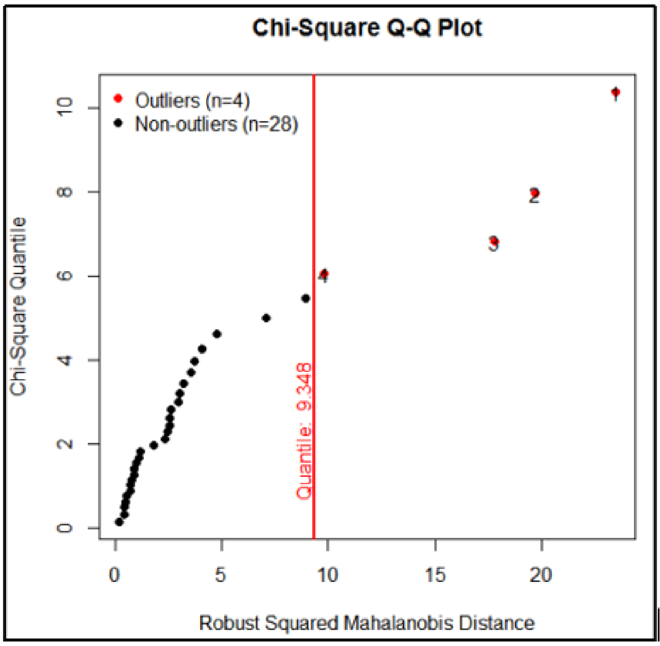


Gambar 3.2. *Boxplot*

Kuartil pertama (Q1) terletak pada garis horizontal bagian bawah *box*. Kuartil kedua (Q2) atau median terletak pada garis horizontal tengah yang melewati *box*. Sedangkan garis horizontal bagian atas *box* menyatakan letak kuatil ketiga (Q3). Badan *box* adalah bidang yang menyajikan *interquartile range* (IQR) atau bagian pertengahan dari 50% nilai observasi. IQR terkenal untuk mengukur penyebaran data. Panjang dan lebarnya box ditentukan oleh IQR. Semakin lebar bidang IQR (apabila *box* horizontal) atau semakin tinggi bidang IQR (apabila *box* vertikal) menunjukkan bahwa data semakin menyebar.

Garis yang memperpanjang *box* dinamakan *whisker*. *Whisker* dibatasi nilai maksimum pada garis diatas badan *box* dan nilai minimum pada garis dibawah badan *box*. *Whisker* menunjukkan nilai yang lebih tinggi dan lebih rendah dari data yang berada dalam IQR kecuali *outlier*. Panjang garis *whisker* bagian atas adalah ≤ Q3 + (1.5 x IQR). Sedangkan panjang garis *whisker* bagian bawah adalah ≥ Q1 – (1.5 x IQR). Nilai yang berada diatas atau dibawah garis *whisker* dinamakan nilai *outlier* atau ekstrim. Nilai *outlier* apabila nilainya lebih dari Q3 + (1.5 x IQR) dan kurang dari Q1 – (1.5 x IQR).

Selain menggunakan *boxplot* untuk mengetahui ada tidaknya nilai ekstrim, *outlier* dapat juga dilihat melalui *Q-Q plot*. Biasanya *Q-Q plot* digunakan dalam menguji kenormalan data. Namun dengan *Q-Q plot* juga bisa mengtahui berapa jumlah data *outlier* dan data bukan *outlier* dalam suatu data pengamatan. Gambar 3.3 adalah ilustrasi *Q-Q plot* yang memuat 28 data bukan *outlier* yang digambarkan titik berwarna hitam dan terdapat 4 titik berwarna merah yang merupakan nilai *outlier* .



Gambar 3.3. *Q-Q plot*

3.10 Analisis Survival

Analisis survival (*survival analysis*) atau analisis kelangsungan hidup atau analisis kesintasan bertujuan menaksir probabilitas kelangsungan hidup, kekambuhan, kematian, dan peristiwa-peristiwa lainnya sampai pada periode waktu tertentu. (Murti, 1997). Analisis survival adalah prosedur statistika untuk menganalisis data dengan waktu sampai terjadinya suatu peristiwa tertentu (*time until an event occurs*) sebagai variabel respons.

'Peristiwa tertentu' tersebut dalam analisis survival lazimnya disebut sebagai 'kegagalan' (*failure*), yang dapat berupa:

- Kematian pada penderita penyakit fatal.
- Eksaserbasi ulang pada penderita penyakit kronis dengan remisieksasebasi yang semula ada dalam fase remisi.
- Tindak kriminal ulang oleh eks-narapidana yang sedang menjalani periode hukuman percobaan.
- Kekambuhan pada eks-pecandu narkotika sehabis menjalani rehabilitasi.
- Dan lain sebagainya. (Harlan, 2018).

3.10.1 Fungsi Survival

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 event. Maka $f(t)$ menyatakan fungsi densitas probabilitasnya dan $F(t) = P(T \leq t)$ menyatakan fungsi distribusi kumulatifnya. (Harlan, 2017). 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 :

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= P(T > t) \qquad (3.1) \end{aligned}$$

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

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

$$= -S'(t) \tag{3.2}$$

Dalam praktik, karena pengamatan tidak dilakukan sejak awal terhadap risiko, melainkan sejak dimulainya penelitian, maka seluruh besaran di atas adalah besaran bersyarat, yaitu syarat bahwa subjek *survive* sampai dengan dimulainya pengamatan:

$$F(t) = F(t|T > t_0) \tag{3.3}$$

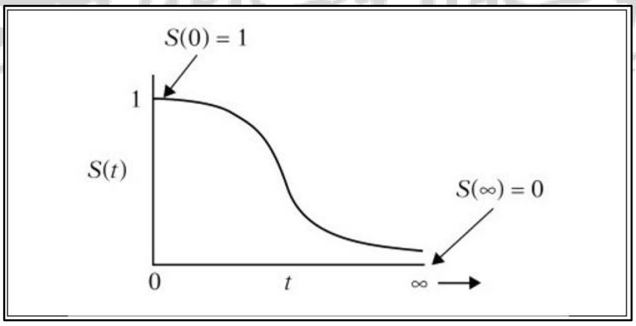
$$f(t) = f(t|T > t_0) \tag{3.4}$$

$$S(t) = S(t|T > t_0) \tag{3.5}$$

Karena syarat $T > t_0$ bersifat lazim, untuk penyerhanaan selanjutnya $S(t|T > t_0)$ hanya akan dituliskan sebagai $S(t|t_0)$, demikian pula besaran lainnya. Pengecualian yaitu untuk $h(t)$, karena *rate* (kelajuan sesaat) tidak tergantung pada waktunya. Karakteristik fungsi survival $S(t)$ antara lain adalah:

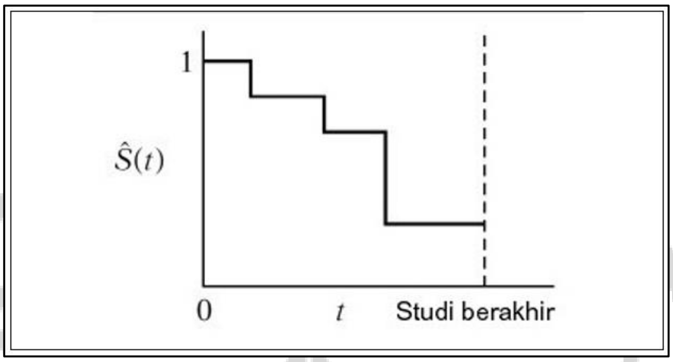
a. Tak membesar : Fungsi survival $S(t)$ mengecil sejalan dengan bertambahnya nilai t.
b. Pada waktu $t = 0$, $S(t) = S(0) = 1$ yaitu pada awal studi belum ada subjek yang mengalami event atau $P\ (T\ >\ 0)\ =\ 1$.
c. Pada waktu $t = \infty$, $S(t) = S(\infty) = 0$, yaitu jika secara teoretis periode 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 3.2 berikut:



Gambar 3.4. Grafik Fungsi Survival

Dalam praktiknya karena sifat data sampel dan cara pengumpulan data, estimasi fungsi survival $\hat{S}(t)$ merupakan fungsi bertingkat (gambar 3.3).



Gambar 3.5. Estimasi Fungsi $\hat{S}(t)$

3.11 Data Tersensor

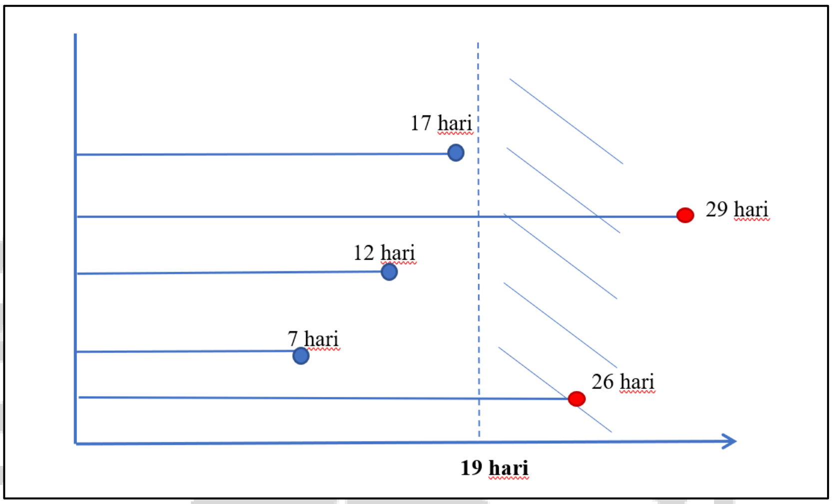
Penyensoran adalah salah satu langkah yang harus dilakukan untuk mengatasi ketidaklengkapan suatu data pengamatan. Data dikatakan tersensor apabila data tidak dapat diamati secara lengkap karena subjek penelitian hilang atau mengundurkan diri atau sampai akhir penelitian subjek tersebut belum mengalami kejadian tertentu, sedangkan data yang dapat diamati secara lengkap sampai penelitian berakhir disebut data yang tidak tersensor. (Lee & Wang, 2003)

Menurut Kleinbaum dan Klein (2005) tiga penyebab data dikatakan tersensor antara lain:

a. *Loss to follow up*, yaitu subjek menghilang selama masa pengamatan, misal subjek pindah atau menolak untuk diamati.
b. Subjek tidak mengalami kejadian selama penelitian.
c. Subjek terpaksa diberhentikan dari pengamatan karena meninggal sebelum pengamatan berakhir atau alasan lain.

3.11.1 Jenis-jenis Data Tersensor

Menurut Collet (2004) dalam analisis survival terdapat 3 tipe penyensoran yaitu sensor kanan, sensor kiri dan sensor interval. Pada kasus dari penelitian ini yaitu pasien penderita *tuberculosis* yang rawat inap diamati dari hari pertama *check-in* inap sampai dengan 19 hari dirawat. Apabila pasien tersebut belum sembuh pada waktu yang ditentukan, maka *survival time* dari pasien tersebut dianggap sebagai data tersensor. Kasus ini termasuk dalam jenis data tersensor kanan tipe I.



Gambar 3.6. Ilustrasi Data Tersensor Kanan

Sensor kanan atau (*right censoring*) adalah sensor yang terjadi dikarenakan objek pengamatan belum mengalami kejadian hingga akhir periode pengamatan, sedangkan waktu awal dari objek pengamatan dapat diamati secara penuh. Misalkan suatu individu diamati selama lima tahun dari awal pengamatan, kemudian pada tahun ketiga individu tersebut pindah ke negara lain dan tidak dapat diamati lagi (*lost to follow up*). Individu ini memiliki waktu survival dalam penelitian setidaknya dua tahun, sehingga waktu pengamatan individu tersebut dikatakan tersensor kanan. Data tersensor kanan dibagi menjadi dua tipe, yaitu :

- Tipe I

  Apabila terdapat objek yang belum mengalami *event* atau kejadian setelah penelitian selesai. Objek tersebut akan disensor.

- Tipe II

  Penelitian akan diberhentikan setelah mendapat sebanyak *k* objek yang telah mengalami *event*. Objek lain yang terjadi setelah *k* akan masuk dalam data yang disensor.

3.12 Estimator Kaplan-Meier

Estimator Kaplan-Meier adalah nonparametric estimator yang dapat digunakan untuk memperkirakan fungsi distribusi survival dari data yang disensor. Estimator ini dapat diperoleh sebagai kasus pembatas penaksir aktuarial (tabel kehidupan) klasik, dan itu tampaknya telah diusulkan oleh Bohmer pada tahun 1912. Namun hal itu dilupakan oleh para peneliti selanjutnya dan tidak diselidiki lebih lanjut sampai Kaplan & Meier (1958) muncul. Hari ini estimator tersebut dinamai estimator Kaplan-Meier. Selama hampir empat decade Estimator Kaplan-Meier telah menjadi salah satu metode statistik utama untuk menganalisis data survival yang disensor dan itu dibahas di sebagian besar buku pelajaran tentang analisis survival. Derivasi ketat dari sifat statistic dari estimator disediakan dalam buku-buku oleh Fleming & Harrington (1991) dan Andersen et al. (1993).

3.12.1 Rumus Fungsi Estimasi Kaplan-Meier

Menggunakan data uji hidup di mana kita ingin mempelajari waktu sampai mati (atau peristiwa lain) untuk populasi yang homogen dengan fungsi distribusi uji hidup $S(t)$ yang mewakili kemampuan untuk bekerja pada individu yang akan hidup pada waktu t. Asumsikan bahwa kita memiliki sampel n individu dari populasi. Pengamatan terhadap waktu bertahan hidup untuk individu-individu ini biasanya akan dikenakan sensor-kanan, artinya bagi beberapa individu kita hanya tahu bahwa waktu bertahan hidup sejati mereka melebihi waktu sensor tertentu. Penyensoran dianggap independen dalam arti bahwa pengetahuan tambahan tentang penyensoran sebelum waktu t tidak mengubah risiko kegagalan pada t. Kami menyatakan dengan $t_1 < t_2 < \cdots t_n$ saat-saat ketika kejadian diamati dan d_i adalah jumlah objek yang mengalami kejadian pada t_i.

Misalkan t_1, t_2, t_3 menunjukkan waktu sebenarnya kejadian dari n objek. Misalkan d_1, d_2, d_3 menunjukkan jumlah kejadian yang terjadi pada masing-masing waktu. Misalkan n_1, n_2, n_3 adalah jumlah objek yang tersisa. Perlu diperhatikan bahwa $n_2 = n_1 - d_1$, $n_3 = n_2 - d_2$ dan seterusnya. Kemudian $S(t_1) = P(T > t_1)$ adalah peluang bertahan sampai melampaui waktu t_1. $S(t_2) = P(T > t_2)$ adalah peluang bertahan melampaui waktu t_2 demikian juga $S(t_3) = P(T > t_3)$ adalah peluang bertahan melampaui waktu t_3. Dengan menggunakan pemahaman tersebut kita dapat mengestimasikan $\hat{S}(t)$ dari fungsi awal $S(t)$.

Untuk setiap saat $t \in [0, t_1)$, kita memiliki $S(t) = P(T > t) =$ (Probabilitas untuk bertahan sampai melampaui waktu t) adalah 1, karena belum ada kematian yang terjadi. Karena itu, untuk semua t interval ini dimisalkan $\hat{S}(t) = 1$.

Untuk waktu $t \in [t_1, t_2)$, kita memiliki persamaan :

$$\hat{S}(t) = 1 - \frac{d_1}{n_1} \tag{3.6}$$

Untuk waktu $t \in [t_2, t_3)$, kita memiliki persamaan :

$$\hat{S}(t) = \left(1 - \frac{d_1}{n_1}\right)\left(1 - \frac{d_2}{n_2}\right) \tag{3.7}$$

Sehingga untuk $t \in [t_i, t_{i+1})$ rumusnya adalah :

$$\hat{S}(t) = \left(1 - \frac{d_1}{n_1}\right)\left(1 - \frac{d_2}{n_2}\right) \dots \left(1 - \frac{d_i}{n_i}\right) \tag{3.8}$$

$$\hat{S}(t) = \prod_{i=1}^{n}\left(1 - \frac{d_i}{n_i}\right) \tag{3.9}$$

Dimana :

$\hat{S}(t)$: peluang tidak mengalami event atau peluang bertahan sampai t

d_i : banyak kejadian saat t_i

n_i : banyak objek yang belum mengalami kejadian saat sebelum t_i

3.12.2 Langkah-langkah Estimasi Fungsi Kaplan-Meier

Langkah-langkah perhitungan untuk mendapatkan estimasi fungsi hazard dan fungsi survival Kaplan-Meier :

a. Mengurutkan waktu lama rawat pasien dari waktu pasien tersingkat hingga waktu pasien terlama (dalam hari) baik yang tersensor maupun tidak tersensor.
b. Membuat waktu tahan hidup (t) yang dibentuk dari lama waktu rawat pasien.
c. Menghitung banyaknya kejadian (d_i) dalam waktu (t_i).
d. Mengestimasi fungsi hazard Kaplan-Meier.
e. Setelah mendapatkan fungsi hazard kemudian dihitung untuk mendapatkan fungsi survival Kaplan-Meier.

3.13 Regresi Linier

Analisis regresi merupakan salah satu teknik analisis data dalam statistika yang seringkali digunakan untuk mengkaji hubungan antara beberapa variabel dan meramal suatu variabel (Kutner dkk, 2004). Istilah “regresi” pertama kali dikemukakan oleh Sir Francis Galton (1822-1911), seorang antropolog dan ahli meteorologi terkenal dari Inggris. Dalam makalahnya yang berjudul “*Regression Towards Mediocrity In Hereditary Stature*”, yang dimuat dalam *Journal of the Anthropological Institute*, volume 15 tahun 1885. Galton menjelaskan bahwa biji keturunan tidak cenderung menyerupai biji induknya dalam hal besarnya, namun lebih medioker (lebih mendekati rata-rata) lebih kecil daripada induknya kalau induknya besar dan lebih besar daripada induknya kalau induknya sangat kecil. (Draper & Smith, 1998)

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*). (Harlan, 2018)

Variabel dependen pada regresi linear disebut juga sebagai respons atau kriterion, sedangkan variabel independen dikenal pula sebagai prediktor atau regresor. Kovariat adalah variabel independen yang berkorelasi dengan prediktor lainnya, juga mempengaruhi respons. Kovariat umumnya tidak diminati hubungannya dengan respons dan hanya digunakan untuk pengendalian hubungan prediktor-respons dalam model.

Respons pada regresi linear selalu berupa variabel kontinu. Sedangkan prediktor dapat berupa variabel kontinu, indikator, ataupun karegorik yang disubstitusikan menjadi variabel indikator.

3.13.1 Regresi Linier Sederhana

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

$$\boldsymbol{\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i} \tag{3.10}$$

Dimana :

$\hat{y}_i$: Variabel Response atau Variabel Akibat (Dependent)

$\hat{\beta}_0$: Konstanta intercept

$\hat{\beta}_1$: Konstanta slope

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

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

Sifat-sifat dari Model Regresi :

1. $\hat{y}_i$ merupakan jumlah dari dua komponen, yaitu suku konstan $\hat{\beta}_0 + \hat{\beta}_1 x_i$ dan suku random ε_i.
2. $E(\varepsilon_i) = 0$ maka $E(\hat{y}_i) = E\left(\hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i\right) = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Hal ini berarti distribusi dari y_i pada tingkat x dalam trial ke-i mempunyai mean $E(\hat{y}_i) = \hat{\beta}_0 + \hat{\beta}_1 x_i$.
3. Nilai pengamatan y pada trial ke-i jatuh pada jarak ε_i dari nilai fungsi regresinya $(E(\hat{y}_i))$ atau $y_i - E(\hat{y}_i) = \varepsilon_i$.
4. Eror ε_i diasumsikan mempunyai variansi konstan σ^2. Oleh karena itu, $\sigma^2(\hat{y}_i) = \sigma^2$
5. Eror diasumsikan tidak berkorelasi. Karena ε_i dan ε_j tidak berkorelasi untuk $\neq j$, maka $\hat{y}_i$ dan $\hat{y}_j$ tidak berkorelasi.

Untuk mendapatkan penaksir yang baik bagi parameter regresi β_0 dan β_1 dapat digunakan metode kuadrat terkecil (ordinary least square/OLS). Misalkan $(x_i|y_i)$, *i=1,2,...,n*, data sampel dan kita ingin menentukan koefisien regresi β_0 dan β_1 sedemikian rupa sehingga minimum.

$$J = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \tag{3.11}$$

Perlu diketahui bahwa J sama dengan bentuk :

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - y_i')^2 \tag{3.12}$$

Dalam persamaan (3.11), x_i dan $y_i$adalah bilangan yang berasal dari pengamatan. Sedangkan β_0 dan β_1 berubah apabila garis regresinya berubah. Jadi dalam hal ini, β_0 dan β_1 dianggap berubah. Dari segi kalkulus ini berarti bahwa perlu mencari turunan J terhadap β_0 dan β_1 kemudian menyamakannya dengan nol.

Jika J diturunkan terhadap β_0	Jika J diturunkan terhadap β_1
$\frac{\partial J}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta x_i) = 0$ Atau $\sum y_i - n\beta_0 - \beta_1 \sum x_i = 0$	$\frac{\partial J}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)\, x_i = 0$ Atau $\sum_{i=1}^{n} y_i x_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0$

Kemudian β_0 dan β_1 dimasukkan ke dalam suatu sistem persamaan linier yang disebut dengan persamaan normal.

$$\sum_{i=1}^{n} \beta_0 + \beta_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \quad (3.13)$$

Bila $\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$ dan $\bar{y} = \sum_{i=1}^{n} \frac{y_i}{n}$ maka persamaannya pertama :

$$\beta_0 = \sum_{i=1}^{n} \frac{y_i}{n} - \beta_1 \sum_{i=1}^{n} \frac{x_i}{n} = \bar{y} - b\bar{x} \quad (3.14)$$

Persamaan kedua :

$$\sum_{i=1}^{n} y_i x_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 = 0 \quad (3.15)$$

$$\sum y_i x_i - \left(\sum y_i / n - \beta_1 \sum x_i / n\right)\left(\sum x_i\right) - \beta_1 \sum x_i^2 = 0 \quad (3.16)$$

$$\sum y_i x_i - \sum \frac{y_i(\sum x_i)}{n} - \beta_1\{\sum x_i^2 - (\sum x_i)^2/n\} = 0 \quad (3.17)$$

Jadi,

$$\beta_1 = \frac{\sum y_i x_i - \frac{(\sum x_i)(\sum y_i)}{n}}{\frac{\sum x_i^2 - (\sum x_i)^2}{n}}$$

$$\beta_1 = \frac{\sum y_i x_i - \bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum x_i y_i - \bar{x}\sum y_i - \bar{y}\sum x_i + n\bar{x}\bar{y}}{\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2}$$

$$\beta_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (3.18)$$

Taksiran persamaan regresi dapat dituliskan dengan :

$$\hat{y} = \beta_0 + \beta_1 x_i \quad (3.19)$$

$$\hat{y} = \bar{y} - \beta_1 \bar{x} + \beta_1 x_i \quad (3.20)$$

$$\hat{y} = \bar{y} + \beta_1 (x_i - \bar{x}) \quad (3.21)$$

3.13.2 Regresi Linier Berganda

Analisis regresi linier berganda adalah pengaruh secara linier antara dua atau lebih variabel indenpenden $(\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{n})$ contoh dengan variabel dependen (Y) yaitu : Pertumbuhan penduduk (x_1) dan Inflasi (x_2) terhadap variabel terikatnya Pengangguran (Y). Analisis ini untuk mengetahui arah pengaruh antara variabel indenpenden dengan dependen apakah masing-masing variabel indenpenden berpengaruh positif atau negatif dan untuk memprediksi nilai dari variabel dependen apabila nilai variabel mengalami kenaikan atau penurunan. Data yang digunakan biasanya berskala interval atau rasio. (Sembiring, 2010)

Regresi linier berganda hampir sama dengan regresi linier sederhana, hanya saja pada regresi linier berganda variabel bebasnya lebih dari satu variabel penduga. Tujuan analisis regresi linier berganda adalah untuk mengukur intensitas hubungan antara dua variabel atau lebih dan membuat prediksi perkiraan nilai X atas Y. (Adhita, 2014)

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 indepeden $(X_1, X_2, X_3, \dots, X_n)$.

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 :

$$\boldsymbol{y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i} \tag{3.22}$$

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 ~ $\mathrm{N}(0, \sigma^2)$, $i = 1,2, \dots, n$

Penaksiran dari β_0, β_1, β_2,…,β_k dengan menggunakan metode kuadrat terkecil (MKT), penaksir tersebut dapat diperolah dengan meminmumkan bentuk kuadrat. Menurut metode kuadrat terkecil penaksir β_0, β_1, β_2,…,β_k diperoleh dengan meminimumkan bentuk kudrat dari :

$$J = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik})^2 \qquad (3.23)$$

Seperti halnya dengan regresi linier sederhana, hasil minimum diperoleh dengan mencari turunan J terhadap β_0, β_1, β_2,…,β_k. Kemudian menyamakan tiap turunan tersebut dengan nol.

$$\frac{\partial J}{\partial \beta_0} = -2 \sum (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) = 0 \qquad (3.24)$$

$$\frac{\partial J}{\partial \beta_1} = -2 \sum (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) x_{i1} = 0 \qquad (3.25)$$

$$\frac{\partial J}{\partial \beta_2} = -2 \sum (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) x_{i2} = 0 \qquad (3.26)$$

$$\vdots$$

$$\frac{\partial J}{\partial \beta_k} = -2 \sum (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) x_{ik} = 0 \qquad (3.27)$$

Kemudian sesudah disederhanakan dan mengganti koefisien regresi dengan penaksirnya. Sistem persamaan ini disebut persamaan normal dan dapat ditulis sebagai berikut :

$$n\beta_0 + \beta_1 x_{i1} + \beta_2 \sum x_{i2} + \cdots + \beta_k \sum x_{ik} = \sum y_i \qquad (3.28)$$

$$\beta_0 \sum x_{i1} + \beta_1 \sum x_{i1}^2 + \beta_2 \sum x_{i1} x_{i2} + \cdots + \beta_k \sum x_{ik} x_{i1} = \sum y_i x_{i1} \qquad (3.29)$$

$$\beta_0 \sum x_{i2} + \beta_1 \sum x_{i1} x_{i2} + \beta_2 \sum x_{i2}^2 + \cdots + \beta_k \sum x_{ik} x_{i2} = \sum y_i x_{i2} \qquad (3.30)$$

$$\vdots$$

$$\beta_0 \sum x_{ik} + \beta_1 \sum x_{i1} x_{ik} + \beta_2 \sum x_{i2} x_{ik} + \cdots + \beta_k \sum x_{ik}^2 = \sum y_i x_{ik} \qquad (3.31)$$

Menentukan $\beta_0, \beta_1, \beta_2,\ldots,\beta_k$ menggunakan metode kuadrat terkecil melalui persamaan normal dapat dicari dengan menggunakan matriks. Apabila persamaan normal disusun dalam bentuk matriks maka berbentuk :

$$(X'X)b = X'Y \qquad (3.32)$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ 1 & x_{31} & x_{32} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix}, \quad X' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ x_{12} & x_{22} & \cdots & x_{n2} \\ \vdots & \vdots & & \vdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{bmatrix}, \; Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}, \; b = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$X'X = \begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & x_{31} & \cdots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{n2} \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \\ \vdots \\ \sum x_{ik}y_i \end{bmatrix}$$

Jika $X'X$ tidak singular maka :

$$b = (X'X)^{-1}X'Y \tag{3.33}$$

3.14 Koefisien Determinasi (R-Square)

Koefisien determinasi adalah ukuran kemampuan variabel independent menjelaskan variabel dependen. Dalam bidang pemodelan, nilai koefisien determinasi (*R-Square*) dapat menjadi indikator kebaikan model. Statistik uji R sudah sangat umum digunakan dalam pemilihan model regresi. Nilai ini hanyalah salah satu dari banyak kriteria kebaikan model. Nilai *R-Square* yang tinggi adalah kriteria nilai model yang baik untuk meramalkan data. Rumus dari koefisien determinasi adalah :

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)}{\sum(y_i - \bar{y})}$$

(3.34)

Keterangan :

R^2 : koefisien determinasi

y_i : nilai observasi respon / variabel dependen ke-*i*

$\hat{y}_i$: nilai peramalan respon / variabel dependen ke-*i*

$\bar{y}$: nilai rata-rata observasi respon / variabel dependen

Nilai koefisien determinasi (*R-Square*) seringkali akan memiliki nilai dengan range 0 – 1. *R-Square* akan bernilai 1 jika model dianggap dapat menjelaskan keseluruhan variasi dari data dan jika bernilai 0 maka berlaku sebaliknya. Nilai *R-Square* seringkali akan memiliki nilai yang optimum ketika asumsi dalam pemodelan terpenuhi.

3.15 Regresi Buckley-James

Buckley dan James (1979) memperkenalkan teknik regresi yang cocok untuk variabel dependen yang disensor. Estimator mereka menggunakan persamaan estimasi kuadrat-terkecil dan mekanisme pemutakhiran yang didasarkan pada estimator non-parametrik dari distribusi residual untuk menangani sensor. Prosedur ini menarik karena penggunaan teknik kuadrat-terkecil memungkinkan untuk interpretasi hasil yang mudah dan penggunaan analisis rutin, sementara skema pembaruan cukup umum untuk mengakomodasi berbagai bentuk sensor dan pengelompokan. (Potter, 2000)

Regresi Buckley-James memiliki kelebihan dapat mengakomodasi variabel dependen yang disensor dengan memodifikasi nilai dari variabel yang tersensor tersebut. Hal ini tidak dimiliki oleh regresi linier berganda biasa karena pada regresi linier variabel dependen yang tersensor maupun tidak tersensor dianggap sama.

Regresi Buckley-James digunakan untuk membuat model regresi dengan variabel dependen mengandung data tersensor. Model regresi Buckley-James :

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \varepsilon_i^* \tag{3.35}$$

Dimana distribusi dari ε_i^* tak ditentukan, $\varepsilon_i^* \sim F$ dengan $E(\varepsilon_i^*) = 0$ dan $Var(\varepsilon_i^*) = \sigma^2 < \infty$.

Untuk mencakup data tersensor variabel dependen di modifikasi dengan mengubah *censored point* pada data tersensor dengan nilai ekspektasinya, yaitu :

$$E(y_i | y_i > t_i) \tag{3.36}$$

Nilai ekspektasi pada persamaan (3.36) tersebut akan membentuk variabel dependen baru yaitu $y_i^\$$ dengan definisi :

$$y_i^\$ = y_i \delta_i + E(y_i | y_i > t_i)(1 - \delta_i) \tag{3.37}$$

Berdasarkan persamaan (3.38) di atas, dapat dilihat bahwa $y_i^\$$ akan kembali ke bentuk y_i jika nilai $\delta_i = 1$ atau data berupa data lengkap dan $y_i^\$ = E(y_i | y_i > t_i)$ jika nilai $\delta_i = 0$ atau data berupa data tersensor. Sehingga apabila pada data tidak terdapat data tersensor, regresi Buckley-James akan memberikan estimasi seperti regresi linear klasik dan dengan mengubah *censored point* pada data tersensor ke nilai ekspektasinya $E(y_i|y_i > t_i)$ membuat model regresi linear yang diperoleh tidak bias. Oleh karena itu ekspektasi dari variabel dependen $y_i^{\$}$ dapat ditulis sebagai berikut :

$$E\left(y_i^{\$}\right) = \beta_0 + \beta_1 x_i \tag{3.38}$$

Nilai $E(y_i|y_i > t_i)$ pada persamaan (3.39) tidak diketahui karena merupakan nilai ekspektasi dari *censored point* pada data tersensor. Dengan menggunakan modifikasi sehingga nilai menjadi :

$$y_i^{\$} = b_1 x_i + [\varepsilon_i^*(b_1)\delta_i + \hat{E}_b(\varepsilon_i^*(b_1) > c_i(b_1)(1-\delta_i)] \tag{3.39}$$

Dengan $\varepsilon_i^*(b_1) = y_i - b_1 x_i$ dan $c_i(b_1) = t_i - b_1 x_i$.

Persamaan (3.40) menunjukkan untuk data lengkap maka nilai variabel dependen $y_i^{\$}$ kembali ke bantuk y_i. Untuk data tersensor dengan fungsi residual $\hat{E}_b(\varepsilon_i^*(b_i)|\varepsilon_i^*(b_1) > c_i(b_1))$ mengubah y_i ke bentuk $y_i^{\$}$.

Nilai untuk residual $\hat{E}_b(\varepsilon_i^*(b_i)|\varepsilon_i^*(b_1) > c_i(b_1))$ diperoleh dengan melakukan pembobotan untuk kombinasi linear residual. Bobot untuk kombinasi linear tersebut diperoleh dengan mengaplikasikan estimator Kaplan-Meier (*product limit estimator*) untuk residual.

Glasson (2007) menggunakan langkah pembobotan sebagai berikut :

1. Residual pengamatan $e_i(b_i) = z_i - b_i x_i$ telah diurutkan dari nilai terkecil ke nilai terbesar.
2. Kemudian melakukan pembobotan dengan persamaan :

$$\hat{E}_b\left(\in_i (b_1)\middle| \in_i (b_1) > c_i(b_1)\right) = \sum_{k=1}^{n} w_{ik}(b_1) e_k(b_1) \tag{3.40}$$

Dimana,

$$w_{ik}(b_1) = \begin{cases} \frac{d\hat{F}\left(e_k(b_1)\right)\delta_k(1-\delta_i)}{\hat{S}\left(e_i(b_1)\right)}, for\ k > 1 \\ 0 \end{cases} \tag{3.41}$$

Dengan $d\hat{F}(e_k(b_1))$ merupakan massa probabilitas yang diperoleh dari estimator Kaplan-Meier untuk residual tidak tersensor $e_k(b_1)$. Sedangkan $\hat{S}\left(e_k(b_1)\right)$ merupakan estimasi fungsi survival untuk residual $e_k(b_1)$.

Setelah langkah di atas dilakukan akan diperoleh nilai dari variabel dependen yang baru yaitu $y_i^{\$}$. Melalui estimasi *least squares* di estimasi parameter β_0 dan β_1 yang meminimalkan jumlahan kuadarat residual.

$$J = \sum_{i=1}^{n} \varepsilon_i^{*2} = \sum_{i=1}^{n} \left(y_i^{\$} - \beta_0 - \beta_1 x_i\right) 2 \tag{3.42}$$

Nilai J di atas akan minimum jika derivative parsial terhadap β_0, β_1 sama dengan nol.

$$\frac{\partial S}{\partial \beta_0} = -2 \sum \left(y_i^{\$} - \beta_0 - \beta_1 x_i\right) = 0 \tag{3.43}$$

$$\frac{\partial S}{\partial \beta_1} = -2 \sum \left(y_i^{\$} - \beta_0 - \beta_1 x_{i1} - \beta_1 x_i\right) x_i = 0 \tag{3.44}$$

Dengan melakukan proses eliminasi pada persamaan (3.44) dan (3.45) di atas diperoleh estimasi parameter β_1.

$$\sum_{i=1}^{n} \left(y_i^{\$} - \beta_0 - \beta_1 x_i\right) = 0 \tag{3.45}$$

$$\sum_{i=1}^{n} \left(y_i^{\$} x_i - \beta_0 x_i - \beta_1 x_1^2\right) = 0 \tag{3.46}$$

Untuk mengeliminasi parameter β_0, persamaan (3.46) dikali dengan $\sum x_i$ dan persamaan (3.47) akan dikalikan dengan n, maka diperoleh :

$$\sum_{i=1}^{n} (y_i^{\$} \sum_{i=1}^{n} (x_i) + n\beta_0 \sum_{i=1}^{n} (x_i) - \beta_1 (\sum_{i=1}^{n} (x_i)^2) = 0 \tag{3.47}$$

$$\sum_{i=1}^{n} (y_i^{\$} - n\beta_0 \sum_{i=1}^{n} (x_i) - n\beta_1 \sum_{i=1}^{n} (x_i^2) = 0 \tag{3.48}$$

Dengan mengeliminasi $-n\beta_0 \sum_{i=1}^{n} (x_i)$ pada persamaan (3.48) dan (3.49) selanjutnya dibawa ke bentuk sebagai berikut :

$$\sum_{i=1}^{n} (y_i^{\$} \sum_{i=1}^{n} (x_i) - \beta_1 (\sum_{i=1}^{n} (x_i)^2 = n \sum_{i=1}^{n} (y_i^{\$} - n\beta_1 \sum_{i=1}^{n} x_i^2 \tag{3.49}$$

Kemudian dilakukan pemindahan parameter β_1 ke ruas sisi kiri pada persamaan (3.50), maka diperoleh :

$$\beta_1 \left(n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2\right) - n \sum_{i=1}^{n} (y_i^{\$} x_i - \sum_{i=1}^{n} (y_i^{\$} \sum_{i=1}^{n} x_i)) \tag{3.50}$$

Dari langkah eliminasi di atas diperoleh estimasi untuk parameter β_1 sebagai berikut :

$$b_1 = \frac{n\sum_{i=1}^{n} x_i y_i^{\$}(b_1) - \sum_{i=1}^{n} x_i \sum_{i=1}^{n}(y_i^{\$}(b_1))}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{3.51}$$

Estimasi persamaan (3.52) di atas mengandung b_1 dalam $y_i^{\$}$. Estimasi parameter β_1 juga bergantung pada b_1 yang membentuk variabel dependen baru $y_i^{\$}$. Oleh karena itu, estimasi parameter β_1 memerlukan iterasi. Sehingga estimasi parameter β_1 sebagai berikut :

$$b_1^{m+1} = \frac{n\sum_{i=1}^{n} x_i y_i^{\$}(b_1^m) - \sum_{i=1}^{n} x_i \sum_{i=1}^{n}(y_i^{\$}(b_1^m))}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{3.52}$$

Sebagai inisialisasi atau nilai awal dari iterasi tersebut yaitu b_1^0 yang diperoleh dari estimasi parameter β_1 dengan regresi linear klasik atau secara sederhana inisialisasi dari iterasi diperoleh dari estimasi model regresi Buckley-James sebelum nilai dari variabel dependen diubah. Dimana nilai $m = 0,1,2,\ldots,m$ dan b_1^m adalah estimasi β_1 untuk iterasi ke-m. Iterasi dilakukan sampai konvergensi diperoleh dan $|b_1^{m+1} - b_1^m|$ cukup kecil ketika persamaan (3.53) konvergen didapatkan estimasi untuk β_1 yaitu $b_1^{\$}$, maka estimasi intersep b_0 dapat diperoleh sebagai berikut :

$$b_0 = \frac{n\sum_{i=1}^{n} y_i^{\$}}{n} - b_1^{\$}\bar{x} \tag{3.53}$$

3.15.1 Regresi Buckley-James Lebih Dari Satu Variabel Independen

Estimasi parameter Regresi Buckley-James dengan lebih dari satu variabel independen pada dasarnya sama dengan estimasi Regresi Buckley-James dengan satu variabel independen. Model dinyatakan dalam bentuk :

$$\hat{Y} = X\beta + \varepsilon_i^* \tag{3.54}$$

Dimana,

$\hat{Y}$: vektor $(n \times 1)$ variabel dependen yang mengandung data tersensor kanan

X : matriks $n \times (p+1)$ variabel independen

β : vektor parameter regresi

ε : vektor residual dengan $\varepsilon \sim F$ dengan $E(\varepsilon_i^*) = 0$ dan $var\ (\varepsilon_i^*) = \sigma^2 < \infty$

Diperhatikan bahwa vektor subjek tersensor adalah $Y = (Y_1, Y_2, \dots, Y_n)', censored\ point\ C = (C_1, C_2, \dots, C_n)'$, dan $\delta = (\delta_1, \delta_2, \dots, \delta_n)'$ merupakan indikator tersensor dengan :

$$Y = \min\,(T_i\,; C_i) \tag{3.55}$$

Dengan

$$\delta_1 = \begin{cases} 0 & ; T_i > C_i \\ 1 & ; T_i \le C_i \end{cases} \tag{3.56}$$

Dengan mengubah *censored point* ke nilai ekspektasinya, maka diperoleh variabel dependen yang baru dengan persamaan sebagai berikut :

$$Y_i^{\$}(b) = Xb + W(b)(Z - Xb) \tag{3.57}$$

Dimana $Y_i^{\$}$ merupakan variabel dependen baru yang diperoleh. Estimator Buckley-James mengubah *censored point* pada data tersensor ke nilai ekspektasinya dengan menggunakan kombinasi linear terbobot dari residual pengamatan $e(b) = (e_1(b), e_2(b), \dots, e_n(b))'$, dimana $e(b) = (Z - Xb)$. Sedangkan matriks bobot $W(b)$ merupakan matriks diagonal atas dengan segitiga atas berisi bobot dari residual dan indikator tersensor yang terdapat pada diagonal utamanya. Matriks bobot $W(b)$ didefinisikan sebagai :

$$W(b) = diag\ (\delta) + [w_{ik}(b)] \tag{3.58}$$

$$W(b) = \begin{bmatrix} \delta_1 & w_{12}(b) & w_{13}(b) & \cdots & w_{1n}(b) \\ 0 & \delta_2 & w_{23}(b) & \cdots & w_{2n}(b) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w_{(n-1)n}(b) \\ 0 & 0 & 0 & \cdots & \delta_n \end{bmatrix} \tag{3.59}$$

Dimana w_{ik} adalah :

$$w_{ik}(b) = \begin{cases} \frac{\hat{d}F(e_k(b_1))\delta_k(1-\delta_i)}{\hat{S}(e_i(b_1))} & , e_k(b) > e_i(b) \\ 0 & , lainnya \end{cases} \tag{3.60}$$

Setelah langkah diatas dilakukan akan diperoleh nilai variabel dependen yang baru yaitu $Y_i^{\$}(b)$. Melalui estimasi *least squre* diestimasi parameter regresi yang meminimalkan jumlahan kuadrat residual.

$$J = \sum_{i=1}^{n} e_i^{\$2} = e'e = (Y_i^{\$} - Xb)'(Y_i^{\$} - Xb)) \qquad (3.61)$$

$$J = Y^{\$}(b)'Y^{\$} - b'X'Y^{\$} - Y^{\$'}Xb + b'X'Xb \qquad (3.62)$$

$$J = Y^{\$'}Y^{\$} - 2b'X'Y^{\$} + b'X'Xb \qquad (3.63)$$

Nilai J diatas akan minimum jika derivative parsial terhadap b sama dengan nol.

$$\frac{\partial s}{\partial b} = -2X'Y^{\$} + 2X'X(b) = 0 \qquad (3.64)$$

Sehingga diperoleh :

$$(X'X)b = X'Y^{\$} \qquad (3.65)$$

Maka diperoleh solusi persamaan :

$$b = (X'X)^{-1}X'Y^{\$} \qquad (3.66)$$

Estimasi persamaan (3.67) diatas didalam $Y^{\$}$ mengandung b. estimasi parameter β juga bergantung pada b yang membentuk variabel dependen baru $Y^{\$}$. Oleh karena itu, estimasi parameter β memerlukan iterasi. Sehingga iterasi parameter β sebagai berikut :

$$b^{(m+1)} = (X'X)^{-1}X'Y^{\$}(b^{(m)}) \qquad (3.67)$$

Dimana $b^{(m)}$ merupakan nilai b dari iterasi ke-*m*. Sebagai inisialisasi dari iterasi tersebut, maka ditentukan estimasi slope $b^{(0)}$ yang diperoleh dari estimasi model Regresi Buckley-James sebelum nilai dari variabel dependen diubah. Selanjutnya dengan melakukan subtitusi pada persamaan (3.57) ke persamaan (3.68) diperoleh estimasi $b^{(m+1)}$.

$$b^{(m+1)} = (X'X)^{-1}X'[Xb^{(m)} + W\left(b^{(m)}\right)\left(Z - Xb^{(m)}\right)] \qquad (3.68)$$

Sedangkan untuk uji overall dan uji parsial pada Regresi Buckley-James yang digunakan untuk mengetahui model terbaik dan variabel signifikan sama halnya dengan uji overall dan uji parsial pada Regresi Linier Berganda. Perbedaan dari keduanya hanya terletak pada koefisien dari independennya.

3.16 Evaluasi Eror

Ketetapan dari suatu metode analisis merupakan kesesuaian dari suatu metode yang menunjukkan seberapa jauh model regresi tersebut mampu meramalkan variabel dependen. Pada dasarnya hasil nilai suatu model regresi tidak mungkin benar-benar akurat. Nilai peramalan variabel dependen hasil model regresi akan selalu berbeda dengan nilai observasi yang sebernarnya. Perbedaan antara nilai Nilai peramalan variabel dependen dengan nilai observasi yang sebernarnya disebut eror. Nilai eror dalam suatu analisis tidak dapat dihindari, namun untuk menghasilkan tujuan peramalan menggunakan persamaan model yang baik maka nilai error tersebut dapat diminimalisir. Dengan demikian maka diperlukan suatu konsep penilaian yang bertujuan untuk menjelaskan sejauh mana nilai eror tersebut.

3.16.1 Residual

Kesalahan (residual) ke-i adalah selisih antara nilai pengamatan y_i dengan nilai taksirannya $\hat{y}_i$, ditulis dengan notasi ε_i .

$$\varepsilon_i = y_i - \hat{y}_i = y_i - b_0 - b_1 x_i \tag{3.69}$$

Residual digunakan untuk mempelajari ketepatan model regresi untuk data sampel.

3.16.2 MAPE

Mean Absolute Percentage Error (MAPE) merupakan perhitungan perbedaan antara nilai peramalan variabel dependen hasil model regresi dengan nilai yang sebernarnya. Perbedaan tersebut diabsolutkan, kemudian dihitung ke dalam bentuk persentase terhadap data asli. Suatu model mempunyai kinerja sangat bagus jika nilai MAPE berada di bawah 10%, dan mempunyai kinerja bagus jika nilai MAPE berada di antara 10% dan 20% (Zainun & Majid, 2003). Adapun diberikan persamaan untuk menghitung MAPE yaitu :

$$MAPE = \frac{\Sigma\left|\frac{y-\hat{y}}{y}\right|}{n} \times 100\% \tag{3.70}$$

Dengan :

y : nilai observasi yang sebenarnya

$\hat{y}$: nilai peramalan variabel dependen

n : jumlah data