

4.1	Uji Asumsi Klasik Regresi .....	22
4.1.1	Uji Normalitas .....	22
4.1.2	Uji Multikolinieritas .....	22
4.1.3	Uji Heteroskedastisitas .....	23
4.1.4	Uji Autokorelasi .....	24
4.2	Analisis Regresi Linier .....	24
4.2.1	Uji Hipotesis .....	26
4.2.2	Koefisien Determinasi ( $R^2$ ) .....	29
4.3	Analisis Regresi Logistik .....	30
4.3.1	Uji Kesesuaian Model Regresi Logistik .....	30
4.3.2	Uji Hipotesis .....	31
4.3.3	Koefisien Determinasi ( $R^2$ ) .....	33
4.4	Penentuan Model Terbaik .....	34
<b>BAB V KESIMPULAN DAN SARAN .....</b>		<b>35</b>
5.1	Kesimpulan .....	35
5.2	Saran .....	35
<b>DAFTAR PUSTAKA .....</b>		<b>36</b>
<b>LAMPIRAN .....</b>		<b>39</b>

## DAFTAR TABEL

Tabel 3. 1 Rincian Variabel Penelitian .....	19
Tabel 4. 1 Hasil Uji Normalitas .....	22
Tabel 4. 2 Hasil Uji Multikolinieritas .....	22
Tabel 4. 3 Hasil Uji Heteroskdastisitas .....	23
Tabel 4. 4 Hasil Uji Durbin Watson.....	24
Tabel 4. 5 Hasil Analisis Regresi Linier .....	25
Tabel 4. 6 Hasil Uji t.....	27
Tabel 4. 7 Hasil Uji F .....	29
Tabel 4. 8 Hasil Perhitungan Koefisien Determinasi.....	30
Tabel 4. 9 Hasil Uji Hipotesis .....	31
Tabel 4. 10 Hasil Perhitungan Koefisien Determinasi.....	33
Tabel 4. 11 Hasil Perhitungan Nilai AIC .....	34

## DAFTAR LAMPIRAN

Lampiran 1. Olah Data .....	39
Lampiran 2. Olah Data .....	41
Lampiran 3. Coding SPSS.....	43
Lampiran 4. Coding Program R / R Studio.....	52

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Demam berdarah atau demam *dengue* ( disingkat DBD ) adalah infeksi yang disebabkan oleh virus dengue dari genus *Flavivirus*, famili *Flaviviridae*. Virus *dengue* ditularkan atau disebarkan pada tubuh manusia melalui gigitan nyamuk *Aedes Aegypti*. Penularan sering terjadi pada musim hujan. DBD banyak ditemukan di daerah tropis dan sub-tropis. Sejak tahun 1968 hingga tahun 2009, *World Health Organization* (WHO) mencatat Negara Indonesia sebagai negara dengan kasus DBD tertinggi di Asia Tenggara. Banyaknya penderita dan luas daerah penyebarannya semakin bertambah seiring dengan meningkatnya mobilitas dan kepadatan penduduk. Penyakit ini tidak hanya sering menimbulkan Kejadian Luar Biasa (KLB) tetapi juga menimbulkan dampak buruk sosial dan ekonomi. Kerugian sosial yang terjadi diantaranya kepanikan keluarga, kematian anggota keluarga, dan berkurangnya usia harapan penduduk.

Di Sulawesi Selatan , dalam tiga tahun terakhir telah tercatat kasus DBD tertinggi terjadi pada tahun 2016 dengan 7.568 kasus, lalu pada tahun 2015 tercatat dengan 4.818 kasus, Pada tahun 2017 kasus menjadi 1.895 kasus.

Tingginya kasus DBD di Sulawesi Selatan tentu saja tidak terlepas dari beberapa faktor yang mempengaruhinya. Faktor sosial ekonomi dan beberapa faktor lainnya seperti persentase rumah/bangunan bebas jentik nyamuk *aedes*, persentase rumah sehat, persentase rumah tangga yang memiliki tempat sehat , rasio dokter umum, serta rasio banyaknya Rumah sakit diduga juga turut berpengaruh. Mengingat Provinsi Sulawesi Selatan mempunyai kasus DBD yang juga kian meningkat , maka untuk mengetahui faktor-faktor signifikan terhadap DBD perlu dibuat suatu pemodelan, sehingga diharapkan upaya untuk penanganan kasus DBD di Provinsi Sulawesi Selatan dapat lebih efisien dan terfokus.

Penelitian mengenai pemodelan DBD telah banyak dilakukan sebelumnya. Yussanti (2012) melakukan pemodelan DBD di Jawa Timur berdasarkan faktor iklim dan sosio ekonomi dengan pendekatan regresi panel semiparametrik dan diperoleh hasil bahwa terdapat dua variable sosio-ekonomi yang signifikan yaitu

pendapatan perkapita yang memberikan efek positif terhadap banyaknya kejadian DBD dan jumlah penduduk berusia 15 tahun kebawah yang memberikan efek negatif. Variabel curah hujan memberikan efek positif terhadap kejadian DBD bila intensitas curah hujan berada antara 1500 mm hingga 3670 mm. Suhu udara memberikan efek yang besar terhadap perubahan kejadian DBD pada suhu antara 22°C hingga 27°C, sedangkan kelembaban memberikan efek positif terhadap perubahan kejadian DBD ketika kelembaban berada pada interval antara 82% hingga 87%. Earnest *et al.* (2011) melakukan penelitian dengan membandingkan kedua model statistik untuk memprediksi peringatan DBD. Kedua model yang digunakan sebagai perbandingan adalah model *Autoregressive Integrated Moving Average* (ARIMA) dan Model *Knorr-Held* dua komponen. Kriteria pemilihan model terbaik didasarkan pada nilai MAPE terkecil yang dihasilkan dari kedua model. Hasil yang menunjukkan bahwa nilai MAPE yang lebih rendah diperoleh dari model *Knorr-Held*, namun *Knorr-Held* relatif lebih sulit untuk mem-*fitting* model serta membutuhkan waktu yang relatif lebih lama untuk menjalankan modelnya.

Pemodelan DBD juga ada di Surabaya dengan menggunakan tiga buah metode juga pernah dilakukan oleh Hidayanti (2014). Ketiga metode tersebut adalah *negative binomial regression*, *poisson regression*, dan *Geographically Weighted Negative Binomial Regression* (GWNBR). Dari ketiga metode tersebut dilakukan perbandingan dan diperoleh hasil bahwa metode GWNBR menghasilkan model terbaik untuk pemodelan kasus DBD di Surabaya. Untuk mendapatkan gambaran tentang daerah yang memerlukan perhatian khusus atau prioritas utama dalam penanggulangan DBD, Hidayanti menggunakan menggunakan metode *Flexibly Shaped Spatial Scan Statistic* dimana hasil yang diperoleh menunjukkan kecamatan Benowo merupakan daerah yang memiliki resiko tertinggi.

Analisis Regresi merupakan salah satu metode statistik yang digunakan untuk menyelidiki pola hubungan antara variabel prediktor dengan variabel respon. Terdapat dua jenis variabel dalam analisis regresi yakni variabel prediktor yang biasa dilambangkan dengan  $X$  dan variabel respon yang dilambangkan dengan  $Y$  dimana kedua variabel tersebut saling berkorelasi. Tipe data pada variabel  $Y$  dapat beragam, salah satunya data *count*. *Poisson regression*

merupakan salah satu metode analisis yang dipakai untuk memodelkan suatu data saat variable Y bertipe data *count*. Beberapa penelitian dengan menggunakan poisson regression pernah dilakukan oleh Yulianingsih, Sukarsa, Sucipwati (2012) yang mengaplikasikannya pada kasus banyaknya siswa SMA yang tidak lulus UN di Bali. Berikutnya dilakukan oleh Sundari (2014) mengenai aplikasi *Poisson Regression* untuk Memodelkan Hubungan Usia dan Perilaku Merokok terhadap banyaknya kematian Penderita penyakit kanker dan paru-paru.

Salah satu teknik pemodelan yang sedang berkembang saat ini adalah *Neural Networks* yaitu sebuah sistem pengolahan informasi yang secara umum memiliki kemampuan dan karakteristik menyerupai jaringan saraf biologis manusia (Fausett, 1994). *Neural network* telah berkembang secara pesat, sejak ditemukan pada tahun 1940-an, seiring dengan kecanggihan teknologi komputer saat ini. Elemen-elemen pemrosesan yang saling terhubung dengan sistem yang paralel menyebabkan pemrosesan pada *neural network* tidak membutuhkan waktu yang lama. Nilai *output* dari *neural network* diperoleh berdasarkan proses pembelajaran yang telah dilakukan sebelumnya.

Penelitian yang dilakukan Duddu & Pulugharta (2012) yang melakukan estimasi tabrakan pada zona level untuk perencanaan transportasi. Estimasi dilakukan untuk total banyaknya dari kecelakaan, kerugian kecelakaan, banyaknya kerusakan properti dengan menggunakan metode *negative binomial regression* dan *neural networks*. Metode *negative binomial regression* dipilih karena kasus overdispersi terjadi pada ketiga variabel dependen (parameter disperse  $> 0$ ). Pada penelitian Duddu & Pulugharta, arsitektur *neural networks* yang dipakai menggunakan *multi layer perceptron* dengan algoritma pembelajaran *backpropagation*. Selanjutnya kedua metode tersebut dibandingkan dengan menggunakan kriteria *Mean Absolute Deviation* (MAD), persentil error ke-50, persentil error ke 85, *Root Mean Square Error* (RMSE). Berdasarkan hasil kriteria kebaikan model dapat diketahui bahwa model *neural networks* memiliki kemampuan yang lebih baik dalam memprediksi total banyaknya dari kecelakaan, kerugian kecelakaan, dan banyaknya kerusakan properti jika dibandingkan dengan model *negative binomial regression*.

Aydin & Tiryaki (2014) melakukan pemodelan *neural networks* untuk memprediksi kekuatan kompresi dari perlakuan panas kayu dan membandingkannya dengan regresi linear berganda, dan diperoleh hasil bahwa *neural networks* memberikan hasil prediksi yang lebih baik dengan nilai  $R^2$  yang lebih besar dan MAPE yang lebih kecil daripada regresi linier berganda. Merujuk dari hasil penelitian Hidayanti (2014) bahwa faktor lokasi tidak memberikan pengaruh signifikan pada banyaknya kasus DBD di Surabaya serta penelitian Duddu & Pulugartha (2012) yang melakukan perbandingan *neural networks* dengan metode statistika konvensional, dengan ini peneliti melakukan penelitian mengenai Perbandingan Model Regresi Linear Berganda dengan Regresi Logistik pada data kasus Demam Berdarah *Dengue* (DBD) di Sulawesi Selatan .

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan, rumusan masalah pada penelitian ini.

1. Bagaimana model Regresi Linier Berganda dan Regresi Logistik pada data kasus DBD di Sulawesi Selatan pada tahun 2017 ?
2. Bagaimana model terbaik yang digunakan pada data kasus DBD di Sulawesi Selatan pada tahun 2017 ?

## **1.3 Tujuan Penelitian**

Berdasarkan rumusan masalah yang telah dipaparkan, maka penelitian ini.

1. Mengkaji model Regresi Linier Berganda dan Regresi Logistik pada data kasus DBD di Sulawesi Selatan pada tahun 2017.
2. Mengidentifikasi model terbaik yang digunakan pada data kasus DBD di Sulawesi Selatan pada tahun 2017.

#### **1.4 Manfaat Penelitian**

Hasil dari penelitian ini diharapkan dapat memberikan manfaat .

1. Memberikan alternatif model yang dapat digunakan untuk mengkaji atau evaluasi terkait kasus Demam Berdarah *Dengue* (DBD) di Sulawesi selatan.
2. Mengkaji dan menerapkan model regresi linear berganda dan regresi logistik pada data kasus DBD atau yang serupa untuk penyelesaian dan penanggulangannya.

#### **1.5 Batasan Penelitian**

Pada penelitian ini, ruang lingkup permasalahan dibatasi dengan beberapa hal.

1. Data yang digunakan pada penelitian ini adalah data banyaknya kasus DBD di Sulawesi Selatan pada tahun 2017.
2. Penelitian ini menentukan Model terbaik yang digunakan dengan melihat perhitungan AIC sebagai ukuran kebaikan model.

#### **1.6 Organisasi Skripsi**

Sistematika penulisan skripsi ini adalah :

##### **BAB I Pendahuluan**

Bab ini membahas mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta organisasi skripsi.

##### **BAB II Kajian Pustaka**

Bab ini membahas mengenai landasan teori, konsep dasar yang mendasari pokok permasalahan dalam tulisan ini, serta penelitian terkait.

##### **BAB III Metodologi Penelitian**

Bab ini berisi waktu dan tempat penelitian, tahapan penelitian, sumber data dan instrumen penelitian



#### **BAB IV Hasil dan Pembahasan**

Bab ini membahas hasil dan pembahasan dari penelitian yang telah dilakukan.

#### **BAB V Kesimpulan dan Saran**

Bab ini berisi kesimpulan dan saran yang telah didapatkan.

## BAB II

### KAJIAN PUSTAKA

#### 2.1 Demam Berdarah Dengue ( DBD )

Penyakit Demam Berdarah *Dengue* (DBD) atau *Dengue Haemorrhagic Fever* (DHF) mulai dikenal di Indonesia sejak tahun 1968 di Surabaya dan Jakarta . DBD banyak ditemukan di daerah tropis dan sub- tropis. Penyakit ini disebabkan oleh virus *Dengue* dari *genus Flavivirus*, famili *Flaviviridae*. DBD ditularkan ke manusia melalui gigitan nyamuk *Aedes* yang terinfeksi virus *Dengue*. Virus *Dengue* penyebab Demam *Dengue* (DD), Demam Berdarah *Dengue* (DBD) dan *DengShock Syndrome* (DSS) termasuk dalam kelompok B *Arthropod Virus* (Arbovirosis) yang sekarang dikenal sebagai *genus Flavivir famili Flaviviride*, dan mempunyai 4 jenis serotipe, yaitu: Den-1, Den-2, Den-3, Den-4. Cara bekerja virus *dengue* seminggu setelah digigit nyamuk yang mengandung virus, seseorang bisa jatuh sakit atau tidak sakit tetapi namun tetap sebagai sumber berpenular bagi orang lain (*carrier*).

Banyaknya penderita dan luas daerah penyebarannya semakin bertambah seiring dengan meningkatnya mobilitas, kepadatan penduduk, perkembangan wilayah perkotaan, perubahan iklim, distribusi penduduk serta faktor epidemiologi lainnya yang masih memerlukan penelitian lebih lanjut (Kementrian Kesehatan RI, 2010). Menurut Notoatmodjo (2003), faktor lingkungan juga memegang peranan penting dalam penularan penyakit, terutama lingkungan rumah yang tidak memenuhi syarat. Lingkungan rumah merupakan salah satu faktor yang memberikan pengaruh besar terhadap status kesehatan penghuninya.

Penyakit DBD masih merupakan salah satu masalah kesehatan masyarakat yang utama di Indonesia. Penyakit ini tidak hanya sering menimbulkan Kejadian Luar Biasa (KLB) tetapi juga menimbulkan dampak buruk sosial dan ekonomi. Di Indonesia kasus DBD berfluktuasi setiap tahunnya dan DBD diperkirakan akan masih cenderung meningkat dan meluas sebarannya. Hal ini karena vektor penular DBD tersebar luas baik di tempat pemukiman maupun ditempat umum. Selain itu kepadatan penduduk, mobilitas penduduk, urbanisasi yang semakin meningkat terutama sejak 3 dekade yang terakhir.

Faktor-faktor lain yang mempengaruhi penyebar luasan DBD antara lain adalah:

1. Perilaku masyarakat
2. Perubahan iklim (*climate change*) global
3. Pertumbuhan ekonomi
4. Ketersediaan air bersih

Sampai saat ini belum ada obat atau vaksin yang spesifik, tetapi bila pasien berobat dini, dan mendapat penatalaksanaan yang adekuat, umumnya kasus-kasus penyakit ini dapat diselamatkan semakin meningkat angka kesakitannya dan sebaran wilayah yang terjangkau semakin luas. Pada tahun 2016, DBD berjangkit di 463 kabupaten/kota dengan angka kesakitan sebesar 78,13 per 100.000 penduduk, namun angka kematian dapat ditekan di bawah 1 persen, yaitu 0,79 persen. KLB DBD terjadi hampir setiap tahun di tempat yang berbeda dan kejadiannya sulit diduga. Dengan demikian penyakit DBD perlu dilakukan perhatian khusus dalam upaya mencegah peningkatan penyakit ini (Kementrian Kesehatan RI, 2017).

## 2.2 Analisis Regresi Linier

Dari data yang telah dikumpulkan, maka akan diolah dengan menggunakan alat analisa regresi berganda (*Multiple Regression*) dengan menggunakan program *Statistical Product and Service Solution* (SPSS) dan *Rstudio*. Alat analisis regresi berganda digunakan untuk melihat pengaruh beberapa variabel independen terhadap satu variabel dependen. Model regresi berganda yang digunakan dalam penelitian ini disajikan dalam persamaan berikut ini :

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_kX_k + e \quad (2.1)$$

Dimana Y merupakan variabel dependen, a merupakan konstanta  $b_1$   $b_2$   $b_k$  merupakan koefisien regresi,  $X_1$   $X_2$   $X_k$  merupakan variabel independent dan e adalah *error*

### 2.2.1 Uji Asumsi Klasik

Sebelum model regresi digunakan untuk menguji hipotesis, diperlukan uji asumsi klasik untuk memastikan bahwa model telah memenuhi kriteria. Uji asumsi

klasik dilakukan untuk memastikan bahwa sampel yang diteliti terhindar dari gangguan normalitas, multikolonieritas, autokorelasi, dan heteroskedastisitas. Maka uji yang digunakan dalam penelitian ini adalah uji normalitas, uji multikolonieritas, uji autokorelasi, dan uji heteroskedastisitas.

#### 1) Uji Normalitas

Uji normalitas adalah langkah awal dalam proses analisis, walaupun hal ini tidak selalu dibutuhkan namun bila sebelum melakukan analisis terlebih dahulu dilakukan uji normalisasi, hasil analisa akan lebih baik karena variabel akan berdistribusi normal

Salah satu uji normalitas yang digunakan adalah uji Shapiro Wilk. Uji kenormalan ini dikembangkan oleh Samuel Shapiro dan Martin Wilk pada tahun 1965. Pada saat ini, uji Shapiro-Wilk menjadi uji kenormalan yang lebih disukai kerana memiliki kekuatan uji yang lebih baik dibandingkan uji-uji alternatif dari bermacam-macam range. Uji ini tergantung pada korelasi antara data yang diberikan dan kecocokan angka normalnya. Tidak seperti beberapa uji-uji yang lain, uji ini tidak perlu menentukan nilai dari rata-rata dan varian terlebih dahulu, serta valid untuk menguji kenormalan dengan ukuran sampel antara 3 sampai 50. Nilai dari uji statistik Shapiro-Wilk ini adalah positif, yaitu lebih kecil atau sama dengan satu (antara 0 dan 1) (Peng et al, 2008).

Dengan menggunakan uji statistik Shapiro-Wilk ini akan diuji hipotesis:

$H_0$  : Data ditarik dari populasi yang berdistribusi Normal

$H_1$  : Data ditarik dari populasi yang berdistribusi tidak Normal

#### 2) Uji Multikolinearitas

Menurut Ghozali (2013), uji multikolinearitas digunakan untuk mengetahui apakah dalam model regresi ditemukan ada atau tidaknya korelasi antar variabel bebas. Untuk Mendeteksi adanya multikolinearitas adalah dari besarnya nilai VIF (*Variance Inflating Factor*) dan *Tolerance* yang terdapat pada masing- masing variabel bebas. Dasar pengambilan keputusan dalam uji mulikolinearitas adalah :

- a) Jika nilai *tolerance*  $> 0,10$  atau nilai VIF  $< 10,00$  artinya tidak terjadi multikolinearitas.

- b) Jika nilai *tolerance* < 0,10 atau nilai VIF > 10,00 artinya terjadi multikoleniaritas.

### 3) Uji Autokorelasi

Uji autokorelasi bertujuan untuk menguji apakah dalam model regresi linier ada korelasi antara kesalahan pengganggu pada periode t dengan kesalahan pengganggu pada periode t-1 (sebelumnya). Jika terjadi korelasi, maka dinamakan ada problem autokorelasi. Model regresi yang baik adalah regresi bebas dari autokorelasi (Ghozali, 2013). Untuk mendeteksi gejala autokorelasi menggunakan uji Durbin-Watson (DW) dengan kriteria :

- a. Jika nilai D-W dibawah -2, berarti ada autokorelasi positif.
- b. Jika nilai D-W diantara -2 sampai +2, berarti tidak ada autokorelasi
- c. Jika nilai D-W diatas +2, berarti ada autokorelasi positif atau negatif.

### 4) Uji Heteroskedastisitas

Uji heteroskedastisitas bertujuan menguji apakah dalam model regresi terjadi ketidaksamaan variansi dari residual satu pengamatan ke pengamatan yang lain. Model regresi yang baik adalah tidak terjadi heteroskedastisitas. Cara mendeteksi ada atau tidaknya heteroskedastisitas adalah dengan melihat grafik plot antara nilai prediksi variabel dependen. Dasar analisis untuk menentukan ada atau tidaknya heteroskedastisitas yaitu:

- a. Jika ada pola tertentu, seperti titik-titik yang ada membentuk pola tertentu yang teratur (bergelombang, melebar kemudian menyempit), maka mengindikasikan telah terjadi heteroskedastisitas.
- b. Jika tidak ada pola yang jelas, serta titik-titik menyebar di atas dan dibawah angka 0 pada sumbu Y, maka tidak terjadi heteroskedastisitas.

## 2.2.2 Uji Hipotesis

Uji hipotesis digunakan untuk mengetahui besarnya kontribusi dan pengaruh variabel independen terhadap variabel dependen. Pengujian hipotesis ini adalah

### 1) Uji Statistik t

Uji t juga dikenal dengan uji parsial, yaitu untuk menguji bagaimana pengaruh masing – masing variabel independennya secara sendiri – sendiri terhadap variabel dependen. Hal ini dapat dilakukan dengan membandingkan t hitung dengan t tabel atau dengan melihat kolom signifikansi (Sig.) pada masing – masing t hitung, Pengujian dilakukan dengan menggunakan tingkat signifikan 5% ( $\alpha = 0,05$ ). Kriteria pengujian uji t adalah :

- a. Jika  $t_{hitung} > t_{tabel}$  dan nilai Sig.  $< \alpha = 0,05$  maka artinya terdapat pengaruh antara variabel independen terhadap variabel dependen.
- b. Jika  $t_{hitung} < t_{tabel}$  dan nilai Sig.  $> \alpha = 0,05$  maka artinya tidak ada pengaruh antara variabel independen terhadap variabel dependen.

### 2) Uji Simultan (Uji Statistik F)

Uji statistik F dilakukan untuk menguji apakah variabel independen secara simultan mempunyai hubungan yang signifikan atau tidak terhadap variabel dependen. Hal ini dapat dilakukan dengan membandingkan  $f_{hitung}$  dengan  $f_{tabel}$  atau dengan melihat kolom signifikansi pada masing – masing f hitung. Pengujian dilakukan dengan menggunakan *significance level* 0,05 ( $\alpha = 5\%$ ). Kriteria pengujian uji f adalah :

- a. Jika  $f_{hitung} > f_{tabel}$  dan Sig.  $\leq \alpha = 0,05$  maka artinya variabel independen secara simultan berpengaruh terhadap variabel dependen.
- b. Jika  $f_{hitung} < f_{tabel}$  dan Sig.  $\geq \alpha = 0,05$  maka artinya variabel independen secara simultan tidak berpengaruh terhadap variabel dependen.

### 3) Uji Koefisien Determinasi ( $R^2$ )

Pengujian Koefisien Determinasi ( $R^2$ ) pada intinya adalah untuk mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel dependen. Koefisien Determinasi berkisar dari nol sampai dengan satu ( $0 \leq R^2 \leq 1$ ). Hal ini berarti bila  $R^2 = 0$  menunjukkan tidak adanya pengaruh antara variabel independen terhadap variabel dependen, bila  $R^2$  semakin besar mendekati 1, menunjukkan

semakin kuatnya pengaruh variabel independen terhadap variabel dependen dan bila  $R^2$  semakin kecil mendekati nol maka dapat dikatakan semakin kecilnya pengaruh variabel independen terhadap variabel dependen. Koefisien determinasi dapat dihitung dengan menggunakan rumus :

$$D = (R^2) \times 100\% \quad (2.2)$$

Dimana D merupakan Determinasi dan R adalah Nilai Korelasi berganda 100% / Persentase Kontribusi

### 2.3 Regresi Logistik Ordinal

Regresi logistik digunakan jika variabel terikatnya (Y) berupa variabel kategori klasifikasi. Perbedaan antara regresi linear dengan regresi logistik adalah jika regresi linear merupakan suatu metode statistik yang menganalisis hubungan antara variabel respon dengan variabel prediktor. Selain itu, variabel responnya berskala interval atau rasio. Persamaan umumnya adalah

$$Y = \beta x + \varepsilon \quad (2.3)$$

Di mana errornya berdistribusi normal dengan meannya nol dan variansinya konstan. Sedangkan regresi logistik merupakan suatu metode statistik yang bertujuan menganalisis variabel respon untuk memperoleh hubungan antara variabel prediktor dengan probabilitas dari suatu kejadian yang diakibatkan oleh variabel prediktor. Variabel responnya biasanya berskala ordinal atau nominal, sedangkan variabel prediktornya bisa diskrit atau kontinyu.

Regresi logistik untuk respon berskala nominal dapat dibedakan menjadi dua, yaitu regresi logistik biner (*binary logistic regression*) dan regresi logistik multinomial (*multinomial logistic regression*). Regresi logistik biner digunakan ketika hanya ada dua kemungkinan variabel respon  $\square Y \square$ , misal membeli dan tidak membeli. Sedangkan regresi logistik multinomial digunakan ketika pada variabel respon (Y) terdapat lebih dari dua kategori.

Model logit kumulatif pertama kali diperkenalkan oleh Walker dan Duncan (1967) dan kemudian disebut model *odds proporsional (proportional odds model)* oleh McCullagh (1980). Jika variabel prediktor  $\mathbf{x} = (x_1 \ x_2 \ \dots \ x_p)^T$ , maka peluang

kumulatif logit didefinisikan (Agresti, 2002) sebagai

$$P(Y \leq j|x) = \pi_1(x) + \pi_2(x) + \dots + \pi_j(x), \quad j = 1, 2, \dots, J \quad (2.4)$$

Kumulatif logit didefinisikan sebagai

$$\begin{aligned} \text{logit}[P(Y \leq J|x)] &= \ln \left[ \frac{P(Y \leq J|x)}{1 - P(Y \leq J|x)} \right] = \ln \left[ \frac{P(Y \leq J|x)}{P(Y > J|x)} \right] \\ &= \ln \left[ \frac{\pi_1(x) + \pi_2(x) + \dots + \pi_j(x)}{\pi_{j=1}(x) + \pi_{j=2}(x) + \dots + \pi_j(x)} \right], \quad j = 1, 2, \dots, J \end{aligned} \quad (2.5)$$

Pendugaan parameter regresi logistik ordinal didapatkan dengan menurunkan fungsi log likelihood terhadap parameter yang akan diestimasi dan disamakan dengan nol. Persamaan  $\frac{\partial L(\beta)}{\partial \beta_k} = 0$  digunakan untuk menaksir intersep parameter  $\beta_k$  dimana  $k = 1, 2, \dots, p$  dan  $\frac{\partial L(\beta)}{\partial \beta_{0j}} = 0$  dipergunakan untuk menaksir intersep  $\theta_j$  dimana  $j = 1, 2, \dots, J - 1$ .

Hasil dari persamaan  $\frac{\partial L(\beta)}{\partial \beta_k} = 0$  dan  $\frac{\partial L(\beta)}{\partial \beta_{0j}} = 0$  merupakan fungsi nonlinear sehingga diperlukan metode iterasi untuk memperoleh estimasi parameternya. Metode iterasi yang dipergunakan adalah metode iterative Weighted Least Square (WLS) yaitu algoritma Newton-Raphson. Iterasi akan berhenti jika terpenuhi kondisi konvergen, yaitu selisih  $\|\beta^{(t+1)} - \beta^t\| \leq \varepsilon$ , dimana  $\varepsilon$  adalah bilangan yang sangat kecil.

Interpretasi dari variabel kontinyu secara matematis dapat dituliskan .

$$g(x) = \beta_0 + \beta_1 x \quad (2.6)$$

Dimana,

$$\beta_1 = g(x + 1) - g(x) \text{ dan } s\beta_1 = g(x + s) - g(x) \quad (2.7)$$

Berdasarkan bentuk matematis maka perubahan satu unit pada variabel prediktor akan memberikan perubahan pada  $g(x)$  sebesar  $\beta_1$ , apabila pada variabel prediktor terjadi perubahan sebesar  $s$  unit akan memberikan perubahan pada  $g(x)$  sebesar  $s\beta_1$ .

### 2.3.1 Uji Univariabel (Parsial)

Pengujian terhadap parameter model dilakukan untuk memeriksa pengaruh dari variabel prediktor terhadap respon. Pengujian dilakukan dengan menggunakan



statistik uji Wald (Hosmer dan Lemeshow, 1989:20).

Hipotesis dari pengujian ini adalah :

$$H_0 : \beta_c = 0$$

$$H_1 : \text{paling sedikit ada satu } \beta_c \neq 0, c = 1, 2, \dots, p$$

$$\text{Statistik ujinya : } W = \frac{\beta_c}{SE(\beta_c)}$$

Statistik uji Wald untuk jumlah sampel besar mengikuti distribusi normal standart dengan mean 0 dan varians 1. Jika uji Wald ini dikuadratkan, maka akan didapatkan uji  $W^2$  yang mengikuti distribusi Chi-Square dengan derajat bebas 1 untuk sampel besar juga (Hosmer dan Lemeshow, 1989:21). Daerah penolakannya adalah jika  $\|W\| > Z_{\alpha/2}$  atau jika dapat pula dilihat dari nilai p, bila nilai  $P > \alpha$  yang ditetapkan maka tolak  $H_0$ .

### 2.3.2 Kekuatan Asosiasi

Ada beberapa  $R^2$  yang dapat digunakan untuk mengukur kekuatan asosiasi antara variabel dependen dengan variabel prediktor.  $R^2$  yang dimaksud disini tidak sepenting pada regresi biasa, karena interpretasi  $R^2$  pada regresi logistik tidak kaku seperti pada regresi biasa. Pengukuran kekuatan asosiasi terdapat 3 metode, yaitu

1. Cox and Snell

$$R_{CS}^2 = 1 - \left( \frac{L(\mathbf{B}^{(0)})}{L(\mathbf{B})} \right)^{\frac{2}{n}}$$

2. Nagelkerke

$$R_N^2 = \frac{R_{CS}^2}{1 - L(\mathbf{B}^{(0)})^2}$$

3. Mc Fadden

$$R_{CS}^2 = 1 - \left( \frac{L(\mathbf{B})}{L(\mathbf{B}^{(0)})} \right)$$

Di mana ( $\mathbf{B}$ ) adalah fungsi log-likelihood model dengan estimasi parameter. ( $\mathbf{B}^{(0)}$ ) adalah fungsi log-likelihood dengan hanya memuat thresholds. n : banyaknya kasus pula dilihat dari nilai p, bila nilai  $P > \alpha$  yang ditetapkan maka tolak  $H_0$ .

## 2.4 Akaike Information Criterion (AIC)

*Akaike Information Criterion (AIC)* adalah penduga kesalahan prediksi dan dengan demikian kualitas relatif model statistik untuk kumpulan data tertentu. Mengingat kumpulan model untuk data, AIC memperkirakan kualitas setiap model, relatif terhadap masing-masing model lainnya. Dengan demikian, AIC menyediakan sarana untuk pemilihan model. AIC didasarkan pada teori informasi. Ketika model statistik digunakan untuk mewakili proses yang menghasilkan data, representasi hampir tidak akan pernah tepat; sehingga beberapa informasi akan hilang dengan menggunakan model untuk mewakili proses. AIC memperkirakan jumlah relatif informasi yang hilang oleh model tertentu, semakin sedikit informasi yang hilang dari model, semakin tinggi kualitas model itu. Dalam memperkirakan jumlah informasi yang hilang oleh model, AIC berurusan dengan *trade-off* antara kebaikan model dan kesederhanaan model. Dengan kata lain, AIC menangani risiko *overfitting* dan risiko *underfitting*.

Kriteria informasi Akaike dirumuskan oleh ahli statistik Jepang Hirotugu Akaike. Awalnya bernama "sebuah kriteria informasi". Ini pertama kali diumumkan dalam bahasa Inggris oleh Akaike pada simposium 197. Sekarang AIC menjadi dasar dari paradigma dasar statistik dan juga banyak digunakan untuk inferensi statistik (Aho, et.al., 2014). Nilai AIC dapat dihitung menggunakan rumus:

$$AIC = -2L_p + 2p \quad (2.8)$$

Dimana  $L_p$  adalah nilai *log* fungsi kemungkinan dengan  $k$  peubah penjelas,  $p = k + 1$  adalah banyaknya parameter.