

## BAB III

### DASAR TEORI

#### 3.1 Statistika Deskriptif

Secara umum metode statistika dibagi dua, yaitu statistika deskriptif dan statistika inferensi. Statistika deskriptif adalah metode statistika yang digunakan untuk menggambarkan (*to describe*) mengenai suatu keadaan atau masalah. Jadi, metode statistika deskriptif adalah suatu ilmu yang merupakan kumpulan dari aturan-aturan tentang pengumpulan, pengolahan, penaksiran, dan penarikan kesimpulan dari data statistik untuk menguraikan suatu masalah (Rasyad dkk, 2003).

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \dots\dots\dots(3.1)$$

dengan  $\bar{x}$  merupakan rata-rata sampel,  $x_i$  merupakan data ke- $i$ ,  $n$  merupakan banyaknya data. Variansi merupakan jumlah kuadrat dari selisih nilai data observasi dari nilai rata-ratanya kemudian dibagi dengan jumlah observasinya. Rumus yang digunakan untuk variansi digunakan pada persamaan sebagai berikut

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \dots\dots\dots(3.2)$$

dengan  $s^2$  merupakan variansi,  $x_i$  merupakan data ke- $i$ ,  $\bar{x}$  merupakan rata-rata,  $n$  merupakan banyaknya data.

Statistika deskriptif bertujuan untuk memberikan gambaran tentang suatu keadaan berdasarkan hasil pengamatan yang nyata. Melalui statistika deskriptif, dilakukan penyusunan data dalam daftar atau tabel dan visualisasi dalam bentuk diagram atau grafik. Selain itu, pengolahan data serta interpretasi terhadapnya mungkin saja dilaksanakan dengan catatan bahwa analisis yang dilakukan tidak sampai pada penarikan kesimpulan yang berlaku umum (Santosa dkk, 2007).

Data statistik dapat disajikan dalam bentuk grafik statistik. Grafik statistik digunakan dalam penyusunan data khususnya data kuantitatif yaitu diagram garis, diagram batang, diagram lingkaran dan *scatter plot*. Pada diagram batang, frekuensinya digambarkan pada sumbu tegak dan kategorinya pada sumbu mendatar atau sebaliknya. Diagram batang mempunyai bentuk diagram batang tegak maupun diagram batang lingkaran. (Fauzy, 2008).

### 3.2 Regresi Logistik Ordinal

Regresi logistik merupakan metode statistika yang diterapkan untuk memodelkan variabel dependen yang bersifat kategorik (berskala nominal/ordinal) berdasarkan satu atau lebih variabel independen yang dapat merupakan data kategorik maupun kontinu (berskala interval/rasio). Apabila variabel dependen hanya terdiri dari dua kategorik maka metode regresi logistik yang diterapkan adalah regresi logistik biner, apabila variabel dependen terdiri dari lebih dari dua kategorik, maka dapat diterapkan regresi logistik multinomial dan apabila variabel dependen berskala ordinal, maka diterapkan regresi logistik ordinal. Model regresi logistik, selain bermanfaat untuk memprediksi, juga untuk melihat adanya keterkaitan antara variabel dependen dengan satu atau lebih variabel independen secara bersama-sama (Pujiati, 2010).

Regresi logistik ordinal adalah suatu analisis regresi yang digunakan untuk menggambarkan hubungan antara variabel dependen dengan sekumpulan variabel independen, dimana variabel dependen bersifat ordinal, yaitu mempunyai lebih dari dua kategori dan setiap kategori dapat diperingkat (Hosmer & Lemeshow dalam Sjahid dkk, 2010). Model yang dapat digunakan untuk regresi logistik ordinal adalah model logit, dimana sifat yang tertuang dalam peluang kumulatif sehingga *cumulative logit models* merupakan model yang dapat dibandingkan dengan peluang kumulatif yaitu peluang kurang dari atau sama dengan kategori repons ke- $j$  pada  $i$  variabel independen atau  $P(Y \leq j | x_i)$ . Peluang kumulatif didefinisikan sebagai berikut (Sjahid dkk, 2010).

$$P(Y \leq j | x_i) = \frac{\exp\left(a_j + \sum_{k=1}^p \beta_k x_{ik}\right)}{1 + \exp\left(a_j + \sum_{k=1}^p \beta_k x_{ik}\right)} \dots\dots\dots(3.3)$$

dengan  $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$  merupakan nilai pengamatan ke- $i$  ( $i = 1, 2, \dots, n$ ) dari setiap  $p$  variabel independen. Pendugaan parameter regresi dilakukan dengan menggunakan transformasi logit dari  $P(Y \leq j | x_i)$ , (Imasikah, 2013).

$$\text{Logit}P(Y \leq j | x_i) = \ln\left(\frac{P(Y \leq j | x_i)}{1 - P(Y \leq j | x_i)}\right) \dots\dots\dots(3.4)$$

### 3.3 Inferensi Regresi Logistik Ordinal

Pemeriksaan peranan variabel-variabel independen (X) dalam model, dilakukan melalui pengujian terhadap parameter model ( $\beta$ ). Pengujian secara serentak dilakukan menggunakan uji G, sedangkan secara parsial menggunakan uji *Wald*.

#### 3.3.1 Statistik Uji G

Statistik uji G adalah uji rasio kemungkinan (*likelihood ratio test*) digunakan untuk menguji peranan variabel independen di dalam model secara bersama-sama. Uji rasio kemungkinan (*likelihood ratio test*) diperoleh dengan cara membandingkan fungsi *log likelihood* dari seluruh variabel bebas dengan fungsi *log likelihood* tanpa variabel bebas (Raharjanti dan Widiyah, 2005). Uji G digunakan untuk menguji hipotesis :

$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$  (tidak ada variabel independen yang berpengaruh terhadap variabel dependen)

$H_1 : \text{Paling sedikit ada satu } j \text{ dengan } \beta_j \neq 0 \text{ } j = 1, 2, \dots, p$  (paling sedikit ada satu variabel independen yang berpengaruh terhadap variabel dependen)

Menurut Hosmer-Lemeshow (1989), statistik uji rasio *likelihood*  $G$  adalah fungsi dari  $L_0$  dan  $L_1$  yang berdistribusi  $X^2$  (*chi-square*) dengan derajat bebas  $p$  yang didefinisikan sebagai:

$$G = -2 \ln \left[ \frac{L_0}{L_1} \right] \dots \dots \dots (3.5)$$

$L_0$  = *likelihood* tanpa variabel independen

$L_1$  = *likelihood* dengan variabel independen

Jika hanya ada satu variabel independen, maka  $L_0$  adalah fungsi *likelihood* yang dihitung sebagai  $\beta_0$  dan  $L_1$  adalah fungsi kemungkinan *likelihood* yang diestimasi dengan  $\hat{\beta}$ . Statistik  $G$  akan berdistribusi *Chi Square* dengan derajat bebas  $p$ , dimana  $p$  adalah banyaknya variabel independen/prediktor (minus intersep) dalam model.  $H_0$  ditolak jika  $G > \chi^2_{(v,\alpha)}$  atau  $H_0$  ditolak jika  $p\text{-value} < \alpha$ . Statistik  $G$  akan bernilai positif dan semakin kecil nilai  $\frac{L_0}{L_1}$  maka akan meningkatkan nilai  $2 \ln \left[ \frac{L_0}{L_1} \right]$ , sehingga peneliti mempunyai bukti cukup kuat untuk menolak  $H_0$  (Agresti, 2007).

### 3.3.2 Statistik Uji Wald

Uji *Wald* digunakan untuk menguji parameter  $\hat{\beta}_j$  secara parsial. Hipotesis yang digunakan adalah:

$H_0 : \beta_j = 0$  (tidak ada pengaruh antara variabel independen ke- $j$  dengan variabel dependen)

$H_1 : \beta_j \neq 0, j = 1, 2, \dots, p$  (ada pengaruh antara variabel independen ke- $j$  dengan variabel dependen)

Statistik uji yang digunakan adalah statistik uji *Wald* (Agresti, 2007).

$$Z^2 = \left( \frac{\hat{\beta}}{SE(\hat{\beta})} \right)^2 \dots\dots\dots(3.6)$$

$Z^2$  akan berdistribusi *chi-square* dengan derajat bebas (*df*) sama dengan 1. Untuk mencari nilai *p-values* dapat menggunakan tabel *chi-square*. Daerah penolakan  $H_0$  adalah  $Z^2 > X^2_{(\alpha, v)}$ . Jika  $H_0$  ditolak berarti dapat disimpulkan bahwa variabel independen ( $X_j$ ) akan dimasukkan ke dalam model atau variabel independen ke-*j* secara parsial berpengaruh terhadap variabel dependen.

### 3.3.3 Odds Ratio

Interpretasi regresi logistik ordinal dapat dijelaskan dengan *odds ratio*. *Odds* adalah cara penyajian probabilitas yang menjelaskan probabilitas bahwa kejadian tersebut akan terjadi dibagi dengan probabilitas bahwa kejadian tersebut tidak akan terjadi (Nugraha, 2012). *Odds* adalah rasio probabilitas sukses ( $\pi$ ) terhadap probabilitas gagal ( $1-\pi$ ). Nilai *odds* bernilai positif.

$$0 < odds < \infty$$

Nilai *Odds ratio* ( $\psi$ ) dapat digunakan untuk menjelaskan kekuatan asosiasi dua variabel. Saat X dan Y adalah Independen,  $\pi_1 = \pi_2$ , sehingga nilai *Odds 1 = Odds 2* dan *odds ratio* ( $\psi$ ) adalah 1 (Nugraha, 2012). Sebagai contoh jika nilai *odds ratio* ( $\psi$ ) = 4, berarti peluang sukses empat kali untuk satu kali kegagalan. Penduga untuk *odds ratio* pada regresi logistik ordinal diberikan sebagai berikut :

$$\psi = \exp(\beta_j) \dots\dots\dots(3.7)$$

dimana  $\beta_j$  : Nilai dugaan untuk parameter  $\beta_j$ .

### 3.3.4 Tingkat Kemampuan Model (Model Summary)

Dalam regresi logistik, tidak ada  $R^2$  seperti yang terdapat dalam regresi linier. Ada beberapa langkah yang dimaksudkan untuk meniru analisis  $R^2$ . Pada analisis regresi logistik *R Square* disebut sebagai *Pseudo R Square*. Interpretasi

dari *R Square* dan *Pseudo R Square* tidaklah sama, tetapi penyimpangannya dapat dianggap sebagai ukuran seberapa buruk model cocok (kurangnya kecocokan antara nilai-nilai yang diamati dan diprediksi), (Newsom, 2015). Di SPSS terdapat tiga jenis *Pseudo R Square*, yaitu:

1. *Cox and Snell R Square*,

*Cox dan Snell* juga didasarkan pada log-likelihood tetapi memperhatikan/melihat ukuran sampel.

$$R^2_{CS} = 1 - \exp\left[-\frac{2}{n}(L_0 - L_1)\right] \dots\dots\dots(3.8)$$

Nilai *Pseudo R Square Cox dan Snell* tidak dapat mencapai nilai 1 (100%) seperti yang diinginkan oleh setiap peneliti.

2. Nagelkerke R Square,

Karena Nilai *Pseudo R Square Cox dan Snell* tidak dapat mencapai nilai 1, maka Nagelkerke memodifikasinya. Sehingga Nilai *Pseudo R Square Nagelkerke* dapat mencapai nilai 1.

$$R^2_N = \frac{R^2_{CS}}{R^2_{max}}, \text{ dimana } R^2_{MAX} = 1 - \exp\left[-\frac{2}{n}L_0\right] \dots\dots\dots(3.9)$$

3. McFadden's.

Nilai ini cenderung lebih kecil dari *R Square* dan nilai-nilai 0.2—0.4 dianggap sangat memuaskan.

$$R^2_{MCF} = 1 - \frac{L_0}{L_1} \dots\dots\dots(4.0)$$

### 3.3.5 Uji Kecocokan Model

Uji kecocokan model digunakan untuk mengevaluasi cocok tidaknya model dengan data, nilai observasi yang diperoleh sama atau mendekati dengan yang diharapkan dalam model. Hipotesis yang digunakan dalam uji kecocokan model adalah sebagai berikut.

$H_0 : \hat{\pi}_j = y_i$  (model sesuai atau tidak ada perbedaan nyata antara hasil observasi dengan kemungkinan prediksi model)

$H_1 : \hat{\pi}_i \neq y_i$  (model tidak sesuai atau ada perbedaan nyata antara hasil observasi dengan kemungkinan prediksi model)

Alat yang digunakan untuk menguji kecocokan dalam regresi logistik ordinal adalah uji *deviance*. Statistik *Deviance* dirumuskan sebagai

$$D = -2[L_m - L_s] \dots \dots \dots (3.6)$$

dimana:  $L_m$  : nilai maksimum log-likelihood untuk model  $M$

$L_s$  : nilai maksimum log-likelihood untuk model *Saturated*

Statistik *deviance* adalah *ratio likelihood* untuk membandingkan model  $M$  dengan model *Saturated*. Statistik *Deviance* menyebar menurut distribusi *Chi-square* dengan derajat bebas sama dengan jumlah observasi (penelitian) dikurangi dengan jumlah parameter yang tercantum dalam model. Semakin besar nilai  $D$  akan memperkecil nilai *p-value*, hal ini menguaktkkan dugaan bahwa model yang terbentuk tidak sesuai dengan data yang ada.

### 3.4 Kecelakaan Lalu Lintas

Kecelakaan lalu lintas dapat diartikan sebagai suatu peristiwa di jalan yang tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan/atau kerugian harta benda (UU RI No.22 Tahun 2009). Kecelakaan lalu lintas merupakan kejadian yang sulit untuk diprediksi kapan dan dimana terjadinya. Kecelakaan tidak hanya mengakibatkan trauma, cedera, ataupun kecacatan, tetapi juga dapat mengakibatkan kematian, kasus kecelakaan sulit diminimalisasi dan cenderung meningkat seiring penambahan panjang jalan dan banyaknya pergerakan dari kendaraan (Hobbs, 1995). Kecelakaan merupakan suatu kejadian yang tidak diinginkan yang

menyebabkan kerugian pada manusia, kerusakan pada properti, dan hilang atau terganggunya proses (Frank Bird, dalam Heinrich, et all 1996).

### **3.5 Dampak Kecelakaan Lalu Lintas**

Dampak yang ditimbulkan akibat kecelakaan lalu lintas dapat menimpa sekaligus atau hanya beberapa diantaranya. Berikut beberapa kondisi yang digunakan untuk mengklasifikasikan korban kecelakaan lalu lintas, yaitu: (UU RI No.22 Tahun 2009)

- a. Kecelakaan Lalu Lintas Meninggal Dunia (MD) adalah saat korban kecelakaan dipastikan meninggal dunia sebagai akibat kecelakaan lalu lintas dalam jangka waktu paling lama 30 hari setelah kecelakaan tersebut.
- b. Kecelakaan Lalu Lintas Luka Berat (LB) adalah saat korban kecelakaan yang karena luka-lukanya menderita cacat tetap atau harus dirawat inap di rumah sakit dalam jangka waktu lebih dari 30 hari sejak terjadi kecelakaan.
- c. Kecelakaan Lalu Lintas Luka Ringan (LR) adalah saat korban kecelakaan mengalami luka-luka yang tidak memerlukan yang tidak memerlukan rawat inap atau yang harus dirawat di rumah sakit kurang dari 30 hari.

### **3.6 Faktor Penyebab Kecelakaan Lalu Lintas**

Secara umum ada empat faktor penyebab kecelakaan lalu lintas; yaitu (BIN dalam *outlook* 2013 Transportasi Indonesia, 2013):

- a. Faktor pengemudi

Faktor pengemudi (manusia) merupakan faktor yang paling dominan dalam sebuah peristiwa kecelakaan lalu lintas. Sebagian besar kejadian kecelakaan diawali dengan pelanggaran rambu-rambu lalu lintas. Pelanggaran rambu-rambu lintas ini bisa terjadi dikarenakan kesengajaan, ketidaktahuan atau tidak adanya kesadaran terhadap arti yang diberlakukan dalam berkendara. Selain itu manusia sebagai pengguna jalan raya sering

lalai dalam memperhatikan keselamatan dirinya dan orang lain dalam berkendara.

b. Faktor sarana (kondisi kendaraan)

Faktor sarana atau kendaraan yang mengakibatkan sering terjadinya kecelakaan antara lain pecah ban, rem tidak berfungsi sebagaimana seharusnya (rem blong), peralatan yang sudah aus tidak diganti, dan berbagai penyebab lainnya. Keseluruhan faktor kendaraan yang berimplikasi terhadap kecelakaan sangat berhubungan dengan teknologi yang digunakan dan perawatan yang dilakukan terhadap kendaraan.

c. Faktor prasarana (jalan, marka, rambu dsb)

Kecelakaan lalu lintas pun bisa dipengaruhi oleh faktor prasarana khususnya jalan. Faktor jalan sebagai sarana untuk berlalu lintas terkait dengan kondisi permukaan jalan, pagar pengaman di daerah pegunungan, pagar pembatas di jalan raya, jarak pandang dan pencahayaan ruas jalan.

d. Faktor alam (cuaca)

Faktor cuaca hujan pun bisa mempengaruhi kinerja kendaraan. Misalnya jarak pengereman menjadi lebih jauh, jalan menjadi lebih licin dan jarak pandang berkurang. Itu semua menjadi faktor penyebab kecelakaan lalu lintas yang selanjutnya. Asap dan kabut pun dapat mengganggu jarak pandang, khususnya di daerah pegunungan.