

## **BAB II**

### **LANDASAN TEORI DAN SURVEY LITERATUR**

#### **2.1 PENDAHULUAN**

Keselamatan dan efisiensi adalah dua tujuan utama dalam teknik transportasi. Perhatian masyarakat terhadap tingkat kinerja keselamatan jalan sangat rendah. Kecelakaan lalu lintas berdampak terhadap biaya yang besar dalam masyarakat. Terdapat dua faktor utama yang menyebabkan terjadinya kecelakaan lalu lintas. Pertama adalah berkaitan dengan pengemudi, dan kedua adalah berkaitan dengan desain jalan.

Pemilihan metodologi yang tepat harus mempertimbangkan beberapa hal dalam masalah teknis dan praktis sebagaimana telah dikemukakan oleh Ben-Akiva et al. (1999). Pertama dari sisi teknis, sekurang-kurangnya terdapat tiga masalah yang timbul yang memerlukan pemahaman dalam analisa keselamatan secara empirik : (i) fluktuasi random dalam frekuensi kecelakaan, (ii) perubahan dalam faktor eksternal selama periode penelitian misalnya perubahan dalam lalu lintas dan karakteristik operasional, dan (iii) efek dari “*regression-to-the-mean*” atau penyimpangan dalam pemilihan yang disebabkan oleh pemilihan penanganan di lokasi penelitian. Kedua, dari masalah praktis, terdapat dua masalah yang harus dipertimbangkan : (i) ketersediaan data (lebih tepatnya tidak akurasi data), dan (ii) lamanya periode penelitian. Masalah ini muncul karena terdapat penukaran. Meskipun periode penelitian sebelum dan sesudah yang lebih lama akan mengurangi efek dari fluktuasi random atau *regression-to-the-mean*, hal itu dapat meningkatkan kemungkinan dari perubahan faktor eksternal sebagaimana telah ditentukan dengan keterbatasan data yang tersedia.

#### **2.2 ANALISA STATISTIK**

Seperti yang diuraikan oleh Tjahjono (2002) peneliti-peneliti keselamatan lalu lintas membuktikan bahwa metode regresi linear kurang sesuai digunakan untuk memprediksi frekuensi kecelakaan lalu lintas. Penggunaan *generalized linear model* (GLIM) dengan sebaran Poisson menghasilkan hasil yang lebih baik. Minimal, sebaran Poisson merupakan pilihan yang sesuai dengan sifat kecelakaan

lalu lintas yang dapat digambarkan sebagai peristiwa yang sangat jarang terjadi, bersifat acak (random), diskret dan non-negatif.

Salah satu peneliti awal yang memperkenalkan model ini adalah Jovanis dan Chang pada tahun 1986. Mereka melakukan studi perbandingan antara kedua model (normal dan Poisson) dan membuktikan bahwa model Poisson lebih baik untuk merefleksikan hubungan antara kecelakaan lalu lintas dan arus lalu lintas serta faktor-faktor lingkungan jalan lainnya. Lebih lanjut, mereka menyimpulkan bahwa regresi linear tradisional mempunyai kelemahan berkaitan dengan nilai yang mungkin menjadi negative dan varians yang beragam. Apabila salah satu dasar dari proses kecelakaan lalu lintas adalah nilai rata-rata (mean) kecelakaan lalu lintas secara fungsional harus berkaitan dengan varians (dimana sebaran Poisson memiliki asumsi bahwa frekuensi rata-rata kecelakaan lalu lintas sebanding dengan variansnya), maka permasalahan dengan regresi linear bukan terhadap nilai parameter variabel-variabel penjelasnya atau variabel-variabel bebasnya, tetapi permasalahan utamanya adalah nilai limit yang diyakini menjadi bias (kesalahan standar atau nilai  $-t$  dari variabel bebas berbeda antara regresi dengan menggunakan sebaran normal dan sebaran Poisson).

Selanjutnya Tjahjono (2002) menguraikan apabila tujuan dari regresi linear adalah untuk mencari dampak signifikan dari variabel-variabel penjelasannya, maka ketidaktepatan nilai limit yang diyakini akan menghasilkan uji signifikansi parameter yang invalid. Akibatnya terdapat kemungkinan terjadinya penerimaan sesuatu variabel penjelas yang sebenarnya harus ditolak dan sebaliknya. Pembuktian lanjut dilakukan pula oleh Joshua dan Gebber (1990) dan Miaou dan Lum (1993). Mereka memperkuat argumen bahwa penggunaan sebaran Poisson merupakan pilihan yang tepat untuk mengembangkan prediksi kecelakaan lalu lintas. Penggunaan metode *Generalized linear model* (GLIM) juga dimungkinkan sejak diperkenalkan secara komprehensif oleh McCullagh dan Nelder pada tahun 1983.

### **2.3 NEGATIVE BINOMIAL**

Untuk mengatasi masalah dispersi yang berlebihan, Shankar et al pada tahun 1995 menganjurkan untuk menggunakan model negatif binomial. Demikian

pula oleh Maher dan Summersgill pada tahun 1996 memberikan alternatif untuk menggunakan model Quasi – Poisson ataupun negatif binomial. Sedangkan model negatif binomial mengasumsikan mengikuti sebaran gamma di mana menghasilkan parameter dispersi yang dapat diuji kesignifikannya sebagai layaknya uji signifikansi yang dilakukan terhadap parameter variabel-variabel bebas.

Pada tahun-tahun terakhir ini telah banyak peneliti yang meramalkan Model Kecelakaan Lalu lintas. Pada awal pekerjaan, penelitian menggunakan regresi linier untuk mendapatkan model frekwensi kecelakaan, dengan kesalahan asumsi menjadi distribusi normal. Hal tersebut telah diungkapkan dalam suatu studi yang dilaksanakan oleh Jovanis dan Pendudu ( 1986) dengan menggunakan regresi linier disetarakan dengan suatu hal negatif dalam struktur kesalahan binomial adalah lebih baik daripada menggunakan regresi linier biasa. Didasarkan pada metode negatif binomial, atau dengan istilah lain kesalahan distribusi gamma biasanya digunakan dengan mengikuti persamaan :

$$E(n_i) = \lambda_i = \exp(\beta X_i) + \varepsilon_i \dots\dots\dots(2.1)$$

di mana  $i$  adalah suatu sebaran gamma dengan suatu struktur kesalahan. Penambahan ini akan membuat perbedaan dari rata-rata mengikuti persamaan yang berikutnya :

$$\text{Var}(n_i) = E(n_i) + \theta E(n_i)^2 \dots\dots\dots(2.2)$$

Dari persamaan itu dapat dilihat bahwa jika parameter sama dengan nol, kemudian model binomial yang negatif menjadi suatu Poisson model; oleh karena itu Poisson model dapat diuraikan sebagai suatu kemutlakan model binomial yang negatif. Model binomial yang negatif diuraikan oleh persamaan berikut ini :

$$P(n_i) = \frac{\Gamma((1/\theta) + n_i)}{\Gamma(1/\theta)n_i!} \left( \frac{1/\theta}{(1/\theta) + \lambda_i} \right)^{1/\theta} \left( \frac{\lambda_i}{(1/\theta) + \lambda_i} \right)^{n_i} \dots\dots\dots(2.3)$$

Seperti dengan Poisson model, kemungkinan maksimum baku digunakan untuk perkiraan nilai  $i$ .

## 2.4 GENERALIZED LINEAR MODELLING (GLIM)

Ide yang mendasari GLIM adalah untuk membuat sebuah paket yang menyediakan sebuah perangkat yang handal untuk analisa data preliminary, seperti grafik yang fleksibel dan fasilitas tabulasi yang baik, dan untuk menggabungkan hal tersebut dengan estimator umum maksimum likelihood untuk membentuk model linear yang sesuai pada data. Algoritma yang sesuai mengestimasi nilai-nilai parameter dan standar error untuk sebuah frekuensi dengan selisih yang lebar dengan menggunakan model matematis termasuk regresi, analisis varian, tabel kontingensi, model log-linear untuk penghitungan data, model logistik untuk respon-respon binary, model-model untuk data yang mengalami peningkatan varian dengan mean, dan variasi dari model untuk menganalisis kehandalan data. Disamping itu juga terdapat fasilitas yang komprehensif untuk menghitung performa dari model dan untuk memeriksa kembali validitas dari asumsi yang telah dibuat atas struktur error.

GLIM dibentuk atas tiga komponen. Pertama, terdapat sebuah perangkat yang handal untuk pemodelan statistik. Hal ini memberikan kemampuan kepada anda untuk mencocokkan model-model statistik pada data anda, menyelesaikan goodness of fit dan menampilkan estimasi, standar error dan nilai perkiraan dari model.

Kedua, GLIM dapat digunakan untuk data eksplorasi, dalam tabulasi dan pemilihan data, dalam menggambarkan pola-pola untuk melihat kecenderungan data anda, atau untuk memeriksa kembali secara visual keberadaan outliers.

Ketiga, GLIM dapat digunakan sebagai alat hitung yang sangat handal untuk evaluasi aritmatik yang rumit, atau sebagai bahasa program untuk membentuk manipulasi data yang luas. Sebagai alat hitung, GLIM bekerja dalam scalars (angka tunggal) atau vektor-vektor (daftar angka). Hal ini dapat dikombinasikan secara umum, melibatkan pengoperasian aritmatik, relasi dan transformasi seperti penjumlahan, lebih besar daripada, logaritma atau integral probabilitas. Kemampuan untuk menggabungkan sering digunakan secara berurutan atas perintah-perintah kedalam sub-program dikenal sebagai macros,

membuat GLIM menjadi bahasa programing yang handal, sangat ideal dan cocok untuk membentuk suatu persyaratan statistik yang spesifik.

## **2.5 TEKNIK GENERALIZED LINEAR MODELLING (GLIM)**

Generalized Linear Modeling digunakan untuk menentukan koefisien model dan kualitas kesesuaian statistik terhadap data keselamatan. Teknik Generalized Linear Modeling menggunakan prinsip-prinsip maximum-likelihood pada model distribusi dari residual errors. Distribusi ini adalah bukan tipikal dari normal ataupun constant variance, sebagaimana asumsi regresi least-squares tradisional digunakan. Sebagai hasilnya, teknik regresi GLIM dapat menghasilkan koefisien variabel yang tidak menyimpang dengan standar error seminimal mungkin.

Pengembangan model berkaitan dengan kecelakaan lalu lintas dan volume telah dibahas dalam beberapa studi. Secara umum, terdapat dua pilihan utama untuk melakukan estimasi parameter-parameter dari model : pendekatan konvensional regresi linier yang menggunakan struktur error dari distribusi normal. Dan pendekatan generalized linear modeling (GLIM) yang menggunakan struktur nonnormal error (biasanya dengan metode Poisson atau negative binomial). Meskipun sebagian peneliti telah menunjukkan bahwa distribusi properti dalam model konvensional regresi linier kurang mampu menggambarkan kejadian secara random, discrete, nonnegative dan khususnya kejadian seporadis, yang merupakan karakteristik dari kecelakaan lalu lintas. GLIM memiliki keunggulan dalam mengatasi masalah yang dihadapi oleh model konvensional regresi linier.

Pendekatan GLIM : bahwa struktur error dianggap sebagai Poisson atau negative binomial. Keuntungan utama dari struktur error Poisson adalah kesederhanaan dari penghitungan (nilai mean dan varian adalah sama). Tetapi keuntungan ini juga sebuah keterbatasan. Sebagaimana dibuktikan bahwa sebagian besar data kecelakaan mungkin terlalu menyebar (nilai varian lebih besar daripada mean), yang mengindikasikan bahwa distribusi negative binomial biasanya menjadi asumsi yang lebih realistis.

Aplikasi dari GLIM dalam model kecelakaan lalu lintas pada dekade terakhir menunjukkan bahwa bentuk persamaan yang sederhana dari kumpulan variabel eksplanatori memberikan hasil yang memuaskan, sehingga bentuk persamaan yang kompleks tidak diperlukan (Taylor et al., 2000). Persamaan tersebut dapat berupa :

Persamaan pangkat .....  $X^a$

Persamaan eksponensial .....  $e^{\beta Y}$

dimana X dan Y adalah variabel eksplanatori.

Sehingga apabila FK adalah frekuensi kecelakaan lalu lintas yang akan diprediksi, serta  $X_i$  dan  $Y_j$  adalah variabel-variabel eksplanatori ( $i = 1,2,3,\dots; j = 1,2,3,\dots$ ), maka persamaan kecelakaan lalu lintas adalah sebagai berikut :

$$FK = k X_1^{\alpha_1} \times X_2^{\alpha_2} \times \dots \times \exp(\beta_1 Y_1) \times \exp(\beta_2 Y_2) \times \dots \dots \dots (2.4)$$

atau

$$\ln(FK) = \ln k + \alpha_1 \ln X_1 + \alpha_2 \ln X_2 + \dots + \beta_1 Y_1 + \beta_2 Y_2 + \dots \dots \dots (2.5)$$

Persamaan multiplikatif dapat dirangkum sebagai berikut :

$$FK = k \text{ (kecepatan) (arus) (geometri) (lingkungan) (lain-lain) (residual)}$$

dimana :

- FK = Perkiraan frekuensi kecelakaan lalu lintas
- k = Konstanta regresi
- Kecepatan = Kumpulan variabel kecepatan (antara lain : rata-rata, maksimum, minimum)
- Flow = Kumpulan variabel arus lalu lintas (antara lain : volume, kepadatan)
- Geometri = Kumpulan variabel geometri jalan (antara lain : lebar lajur, lajur dan bahu jalan)
- Lingkungan = Kumpulan variabel lingkungan (antara lain : cuaca, waktu kejadian kecelakaan lalu lintas)
- Lain-lain = Kumpulan variabel lain-lain (antara lain : karakteristik pengemudi dan kendaraan)
- Residual = Error term dari persamaan regresi.

## 2.6 ATURAN BAYES

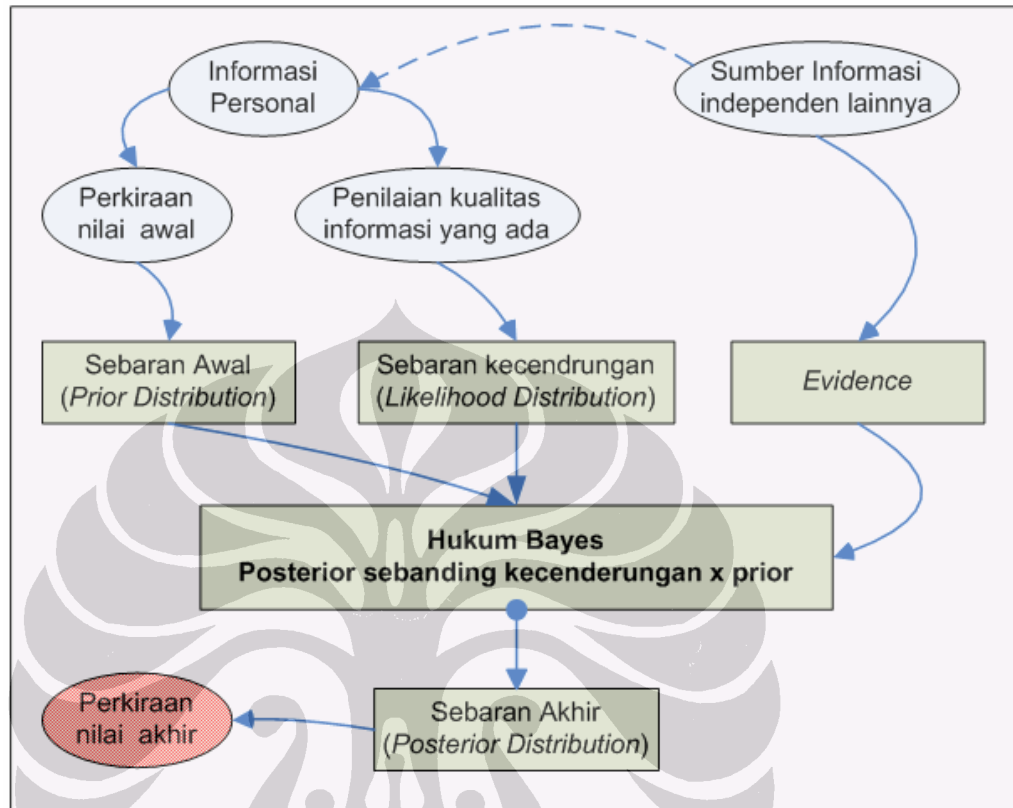
Pada hakekatnya prosedur EB berupaya memprediksi sesuatu dengan tidak menggunakan hanya satu petunjuk (*clue*) saja tetapi setidaknya-tidaknya dua petunjuk. Berikut ini sebuah contoh untuk memudahkan bagaimana suatu proses EB dapat berjalan. Sebagai contoh: sebuah persilangan sebidang antara jalan dan kereta api di sesuatu tempat (katakan lokasi  $x$ ) di Jakarta tidak terdapat sama sekali kecelakaan lalu lintas antara pengguna jalan dan kereta api dalam lima tahun terakhir. Tetapi berdasarkan data yang ada, rata-rata kecelakaan pada persilangan sebidang di Jakarta dalam lima tahun terakhir adalah 0,002 kecelakaan per tahun. Katakan bahwa Informasi kecelakaan = 0 (tidak ada kecelakaan) pada lokasi  $x$  merupakan petunjuk (*clue*) 1, sedangkan rata-rata kecelakaan = 0,002 merupakan petunjuk 2. Pertanyaan yang timbul: dapatkah kita memperkirakan jumlah kecelakaan di lokasi  $x$  adalah 0 (nol) kecelakaan per tahun? Walaupun kita mengetahui bahwa rata-rata di Jakarta sebesar 0,002 kecelakaan per tahun. Atau sebaliknya kita dapat mengatakan bahwasannya tingkat keselamatan persimpangan sebidang di Jakarta adalah 0,002 dengan menutupi kenyataan bahwasannya di lokasi  $x$  terdapat hanya 0 (nol) kecelakaan per tahun.

Perkiraan yang masuk akal tentunya harus memperhatikan kedua petunjuk (*clues*) yang ada. Artinya perkiraan jumlah rata rata kecelakaan per tahun berkisar antara 0 hingga 0,002 per tahun. Untuk mendapatkan perkiraan terbaik dibutuhkan suatu faktor pembebanan (*weighting factor*) yang dapat menghasilkan suatu nilai di antara kedua nilai petunjuk yang ada. Hal inilah yang menjadikan esensi dasar dari teori EB.

Terdapat tiga cara untuk mencari peluang dari suatu entitas sejenis atau memprediksi kecelakaan lalu lintas, yaitu dengan cara:

1. Memperkirakan secara subyektif
2. Dengan melakukan penelitian secara empirikal dari data tahun-tahun sebelumnya
3. Dengan menggunakan pendekatan matematis untuk menjabarkan secara teori sebaran (*distribution*) kecelakaan lalu lintas.

Secara skematik Jessop (1990) menguraikan proses EB seperti terlihat pada Gambar 2.1



Gambar 2.1 Perubahan dari Perkiraan Awal Sebagai Esensi Dasar EB.

## 2.7 METODE STUDI SEBELUM DAN SESUDAH

Berbagai penelitian telah dilakukan dalam rangka mencoba untuk membuktikan hubungan antara kecelakaan, volume lalu lintas dan karakteristik geometrik dari jaringan jalan. Karena tidak terpantaunya keanekaragaman antara lokasi frekuensi kecelakaan yang telah ditetapkan dari berbagai lokasi akan menunjukkan overdispersion, yang artinya bahwa data memiliki sebuah pemantauan menyimpang yang lebih besar daripada perkiraan penyimpangan yang diberikan oleh pengambilan sample model. Hal ini bertentangan dengan distribusi poisson sebagai distribusi probabilitas tunggal (single probabilistic distribution) yang menjelaskan secara memuaskan mengenai meningkatnya jumlah kecelakaan. Pada kenyataannya dapat dilihat bahwa teknik-teknik yang rumit dan advance/lanjutan sangat penting untuk mendapatkan hasil model



statistik yang lebih baik. Sehingga wajar dalam aplikasi ini bahwa generalized linear models dianggap sebagai model yang terbaik untuk tipe data seperti ini (McCullagh et al, 1989). Banyak penelitian telah berhasil menggambarkan aplikasi dari metodologi generalized linear modeling untuk mendapatkan model-model prediksi kecelakaan.

Dilandasi oleh fenomena “regression-to-the-mean”, terlihat secara empirik bahwa perhitungan aktual dari kecelakaan bukanlah estimator yang baik untuk perkiraan jumlah kecelakaan (Hauer, 1986). Sehingga sangat penting untuk menilai perlunya before-after study. Oleh sebab itu perhatian tertuju pada pengestimasian perkiraan jumlah kecelakaan. Metode empirikal bayes telah digunakan untuk memberikan estimator yang lebih baik dari perkiraan frekuensi kecelakaan. (Hauer, et al dan Mountain, et al). Metode Bayesian memberikan hasil yang baik dalam beberapa seri penelitian di akhir dekade ini, dengan aplikasi terhadap analisa kecelakaan (Tunaru).

Empat langkah pendekatan untuk membangun model evaluasi sebelum dan sesudah, sebagai berikut (Ben-Akiva, et al., 1999) :

(1) Perbandingan Klasik Sebelum dan Sesudah

Keefektifan adalah perhitungan estimasi biasa sebagai persentasi dari perbedaan antara perhitungan kecelakaan sebelum dan sesudah atau pembuktian melawan hipotesa awal yang menyatakan “frekwensi kecelakaan sebelum dan sesudah tindakan adalah sama”. Keuntungan dari pendekatan ini adalah sangat mudah dan tidak membutuhkan pertimbangan atas berbagai perubahan dalam faktor-faktor eksternal (karakteristik lalu lintas) dan penyimpangan oleh pemilihan tindakan.

(2) Estimasi Empirik Bayes

Pendekatan ini menggunakan aturan Bayesian bahwa estimasi dari ukuran kecelakaan yang diharapkan dapat diperbaiki melalui sebuah informasi tambahan. Data kecelakaan setelah tindakan kemudian dibandingkan dengan “nilai rata-rata data kecelakaan sebelum tindakan”. Metode ini mengestimasi jumlah kecelakaan yang diharapkan dengan mengatur tingkat ukuran kecelakaan yang diteliti melawan pengalaman kecelakaan dari kondisi yang mirip/serupa.

Keuntungannya adalah :

(i) Pendekatan ini secara langsung menggunakan cara dari “regression-to-the-mean”,

(ii) Pada tipe “treatment-control” eksperimental desain akan mengarah pada estimasi yang lebih baik dari jumlah kecelakaan (Hauer, 1991).

Demikian juga, pendekatan ini memiliki beberapa kelemahan seperti :

(i) Estimasi tersebut akan menentukan beberapa asumsi mengenai probabilitas distribusi dari peristiwa kecelakaan, dan

(ii) Estimasi tersebut tidak berdampak terhadap perubahan dalam faktor-faktor internal atau eksternal.

### (3) Model Regresi

Pendekatan ini sama dengan model prediksi sebelumnya yang pada dasarnya menentukan nilai yang diharapkan atas kecelakaan berdasarkan pada penjelasan variabel-variabel sebelum dan sesudah perlakuan. Sebagaimana telah dijelaskan, teknik yang paling baik adalah menggunakan generalized linear models (Atkins et al, 1989). Keuntungan dari pendekatan ini bahwa estimasi ini berdampak terhadap faktor-faktor internal dan eksternal yang mempengaruhi jumlah yang diharapkan dari kecelakaan, tetapi dapat menyebabkan penyimpangan dengan pemilihan jika estimasi ini bukan merupakan tipe “treatment-control” model.

### (4) Kombinasi dari Regresi dan Estimasi Bayesian Model

Pendekatan ini menambahkan keuntungan dari kedua pendekatan regresi dan Bayesian. Meskipun pendekatan ini membutuhkan data yang cukup untuk memastikan akurasi dari model.

Kombinasi dari estimasi regresi dan Bayesian model dipilih karena metode estimasi ini lebih unggul daripada pendekatan lainnya. Model regresi memanfaatkan hasil dari model-model prediksi untuk mendemonstrasikan penggunaan model-model untuk keperluan praktis, seperti studi sebelum dan sesudah. Tabel 1 menggambarkan typology dari pendekatan studi sebelum dan sesudah.

Hauer (1997) merekomendasikan periode tiga tahun sebelum dan tiga tahun sesudah penelitian jika tidak terdapat perubahan besar dalam faktor-faktor eksternal.

		Ketersediaan Data Kecelakaan Lalu Lintas	
		Data Kecelakaan Lalu Lintas	Jumlah Kecelakaan Lalu Lintas + Variabel Penyebabnya
Metode Statistikal	Klasikal	Komparasi sederhana Sebelum dan Sesudah	Model Regresi
	Bayesian	Estimasi Empirikal Bayes (EB)	Kombinasi Regresi dan Empirikal Bayes (EB)

Tabel 2.1. Typology dari Pendekatan Studi Sebelum dan Sesudah (Ben-Akiva et al., 1999).

Lebih jauh, Hauer (2001) menyarankan agar sangat berhati-hati dalam memperlakukan model-model yang penyebarannya terlalu luas. Dalam model statistik multivariate dari prediksi kecelakaan lalu lintas, perhitungan kecelakaan biasanya tersebar terlalu luas. Luas dari penyebaran data tersebut merupakan hasil dari estimasi. Hal ini menunjukkan bahwa sifat dari penyebaran yang terlalu luas akan berdampak terhadap estimasi kemiripan maksimum dari parameter-parameter model. Jika satu asumsi bahwa parameter over dispersion diterapkan terhadap seluruh bagian jalan dalam data base, kemudian estimasi maksimum likelihood dari parameter-parameter tidak akan berpengaruh secara tepat dengan bagian jalan yang sangat pendek dan tidak cukup mempengaruhi dengan bagian jalan yang panjang. Hal ini juga terjadi ketika estimasi maksimum likelihood diterapkan pada metode empirik bayes yang mengakibatkan tidak konsistensi yang disebabkan oleh parameter over dispersion ( $\theta$ ). Untuk menghindari masalah ini, Hauer (2001) menyarankan aplikasi pendekatan Estimasi Bayes untuk sebuah unit panjang jalan dan mengatur parameter over dispersion untuk sebuah bagian panjang jalan ( $L$ ) berdasarkan unit panjang yang tersedia ( $\theta L$ ).

Estimasi empirik bayes mempertimbangkan berbagai pengaturan berdasarkan pada informasi kemiripan dan dapat disebut sebagai “Posterior (setelah treatment)  $\alpha$  dengan kemiripan  $x$  prior (sebelum treatment)”.

Prosedur pengaturan berdasarkan pada asumsi sebagai berikut :

- (1) Kejadian kecelakaan berdasarkan proses Poisson dan,

(2) Nilai tengah dari distribusi poisson dengan beragam kondisi mengacu pada distribusi gamma (model negative binomial).

Inti dari pendekatan empirik bayes adalah kemiripan informasi harus berdasarkan kepada dua petunjuk (Hauer, 1997), yaitu :

1. Petunjuk yang berasal dari ciri-ciri suatu kondisi, untuk ciri-ciri ini dibuat mirip dengan kondisi lain dimana data keselamatan tersedia.
2. Petunjuk yang berasal dari rekaman kecelakaan dari sebuah kondisi, petunjuk ini adalah gambaran dari keselamatan atas kondisi tersebut.

