

BAB III

LANDASAN TEORI

A. Regresi

1. Pengertian Regresi

Regeresi adalah alat yang berfungsi untuk membantu memperkirakan nilai suatu varibel yang tidak diketahui dari satu atau beberapa variabel yang tidak diketahui. Analisis regresi didefinisikan sebagai kajian terhadap hubungan satu variabel yang disebut variabel yang diterangkan (*the explaind variabel*) atau sering disebut sebagai variabel tergantung, dan variabel tidak tergantung atau variabel bebas.

Metode regresi yang sering digunakan yaitu analisis regresi linier dan non linier. Jika variabel tidak bebas bersifat diskrit, analisis linier tidak layak digunakan karena bebearapa alasan, yaitu :

- a. Variabel tidak bebas di dalam metode regresi linier harus bersifat *continue*
- b. Variabel tidak bebas di dalam metode regresi linier harus dapat mengakomodasi nilai negatif

Variabel diskrit biasa juga dikatakan salam kategori dan sering juga disebut variabel nominal atau variabel kategorik.

Metode analisis regresi digunakan untuk menghasilkan hubungan antara dua variabel atau lebih dalam bentuk numerik dan untuk bagaimana dua atau lebih peubah saling berkait, dimana telah diketahui variabel lainnya dan variabel mana yang mempengaruhinya. Persamaan regresi ini merupakan persamaan garis yang paling mewakili hubungan antara dua variabel tersebut. Beberapa asumsi statistik yang diperlukan dalam melakukan analisis regresi adalah :

- a. Variabel tak bebas, yaitu fungsi linier dari variabel bebas. Jika hubungan tersebut tidak linier, data sering kali harus ditransformasikan agar menjadi linier.
- b. Variabel bebas adalah tetap atau diukur tanpa kesalahan.
- c. Tidak ada korelasi antara variabel bebas
- d. Variansi dari variabel tak bebas terhadap garis regresi adalah sama untuk seluruh nilai variabel tak bebas.
- e. Nilai variabel tak bebas harus berdistribusi normal atau mendekati normal.
- f. Nilai peubah bebas sebaiknya merupakan besaran yang relative mudah diproyeksikan.

2. Model Regresi Linier Sederhana

Dalam memperkirakan hubungan antara dua variabel terlebih dahulu membuat asumsi mengenai bentuk hubungan yang dinyatakan dalam fungsi tertentu. Dalam beberapa hal, bisa dicek asumsi tersebut setelah hubungan diperkirakan.

Regresi linier sederhana memiliki fungsi sebagai berikut :

- a. Menguji hubungan / korelasi / pengaruh satu variabel bebas terhadap satu variabel terikat.
- b. Melakukan prediksi atau estimasi variabel terikat berdasarkan variabel bebasnya.
- c. Data yang dianalisis haru berupa data yang berskala interval / rasio.

Fungsi linier, selain mudah interpretasinya, juga dapat digunakan sebagai pendekatan (*approximation*) atas hubungan yang bukan linier (non linier). Fungsi linier, mempunyai bentuk persamaan sebagai berikut :

$$Y = A + BX \dots\dots\dots 3.1$$

Dimana A dan B adalah konstanta atau parameter, yang nilainya harus diestimasi. Persamaan $Y = A + BX$ juga bisa ditulis $Y = B_0 + B_1X_1$ atau dengan symbol lainnya. Beberapa symbol yang sering digunakan dalam fungsi linier ini adalah :

Δ = delta, symbol pertambahan

ΔX = delta X, pertambahan X

ΔY = delta Y, pertambahan Y

$B = \frac{\Delta Y}{\Delta X}$ = rata – rata pertambahan Y per 1 unit (satuan) pertambahan X, atau pertambahan X 1 unit akan mengakibatkan pertambahan Y sebesar B.

$Y = 2 + 1,5X$, $A = 2, B = 1,5$ artinya kalau $X = 0$, $Y = 2$ kalau X bertambah 1 unit, Y bertambah 1,5 unit.

Hubungan di atas merupakan hubungan matematis, secara teoritis, apabila $X = 10$, Y harus $2 + 1,5 (10) = 17$. Tetapi dalam prakteknya tidak demikian, sebab yang mempengaruhi Y bukan hanya X saja melainkan masih ada faktor lain yang tidak dimasukkan dalam persamaan. Faktor – faktor tersebut secara keseluruhan disebut kesalahan pengganggu atau *disturbance error*. Kesalahan pengganggu tersebutlah yang menyebabkan suatu ramalan sering tidak tepat.

Kesalahan ramalan menyebabkan perencanaan menjadi tidak akurat, sehingga kesalahan tersebut mengakibatkan resiko, dan karenanya harus diusahakan sekecil mungkin. Dalam membuat keputusan, selalu ada resiko yang disebabkan oleh adanya kesalahan. Karena kesalahan itu tidak dapat dihilangkan sama sekali, maka resiko itu berapapun kecilnya selalu ada. Resiko hanya bisa diperkecil dengan memperkecil kesalahan (*minimized error*). Dengan

memperhitungkan kesalahan pengganggu, ε , maka bentuk persamaan fungsi linier diatas menjadi sebagai berikut :

$$Y = A + BX + \varepsilon \dots\dots\dots 3.2$$

Dimana : A dan B adalah konstanta yang harus di estimasi.

ε adalah kesalahan pengganggu (*disturbance error*)

3. Analisis Regresi Linear Berganda

Analisis yang memiliki variabel bebas lebih dari satu disebut analisis regresi linier berganda. Teknik regresi linier berganda digunakan untuk mengetahui ada tidaknya pengaruh signifikan dua atau lebih variabel bebas terhadap variabel terikat (Y). Model regresi linier berganda untuk populasi dapat ditunjukkan sebagai berikut :

$$Y = (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots K) \dots\dots\dots 3.3$$

Model regresi linier berganda untuk populasi diatas dapat ditaksir dengan model regresi linier berganda untuk sampel, yaitu : $\hat{Y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k$

dengan : \hat{Y} = nilai penduga bagi variabel

Y = dugaan bagi parameter konstanta

b_0 = dugaan bagi parameter konstanta

b_1, b_2, \dots, b_k = variabel bebas b_1, b_2, \dots, b_k

4. Hubungan Linier Antara Dua Variabel

Salah satu tujuan analisis data ialah untuk memperkirakan / memperhitungkan besarnya efek kuantitatif dari perubahan suatu kejadian terhadap kejadian lainnya. Setiap kebijakan (*policy*), baik dari pemerintah

maupun swasta, selalu dimaksudkan untuk mengadakan perubahan (*change*). Sebagai contoh, misalnya pemerintah menambah jumlah pupuk agar produksi padi meningkat, pemerintah menaikkan gaji pegawai negeri agar prestasi kerja mereka meningkat, dll.

Untuk keperluan evaluasi / penilaian suatu kebijaksanaan mungkin ingin diketahui besarnya efek kuantitatif dari perubahan suatu kejadian terhadap kejadian lainnya. Kejadian – kejadian tersebut, untuk keperluan analisis, bisa dinyatakan di dalam perubahan nilai variabel. Untuk analisis dua kejadian (events) digunakan dua variabel x dan y .

Apabila dua variabel X dan Y mempunyai hubungan (korelasi), maka perubahan nilai variabel yang satu akan memengaruhi nilai variabel lainnya. Hubungan variabel dapat dinyatakan dalam bentuk fungsi, misalnya $Y = f(X) \rightarrow Y = 2 + 1,5X$. Apabila bentuk fungsinya sudah diketahui, maka dengan mengetahui nilai dari satu variabel ($=X$), maka nilai variabel lainnya ($=Y$) dapat diperkirakan. Data hasil ramalan yang dapat menggambarkan kemampuan untuk waktu yang akan datang, sangat berguna bagi dasar perencanaan.

Untuk membuat ramalan (*forecasting*) Y dengan menggunakan nilai dari X , maka X dan Y harus mempunyai hubungan yang kuat. Kuat tidaknya hubungan X dan Y diukur dengan suatu nilai, yang disebut koefisien korelasi, sedangkan besarnya pengaruh X terhadap Y , diukur dengan koefisien regresi.

5. Regresi Dalam Variabel Terikat Data Kualitatif

Aplikasi data kualitatif sebagai variabel bebas disebut juga dengan variabel dummy. Kasus yang sering dijumpai umumnya adalah kasus data kualitatif yang dapat diterapkan dalam variabel terikat. Kasus yang bisa dijadikan contoh yaitu, kemampuan seseorang untuk memiliki sebuah kendaraan di kota yang dipengaruhi oleh jarak, pendapatan. Model yang

menggunakan variabel kualitatif atau kategori terikat dapat dibedakan dalam dua hal yaitu:

1. Regresi model probabilitas linier (*linier probability model* = LPM)
2. Regresi model logistic binar (*binary logistic regression model*)

A. Stated Preference

Stated preference adalah suatu pendekatan dengan cara menyampaikan pertanyaan pilihan (*option*) yang berupa suatu hipotesa untuk dinilai oleh responden. Selanjutnya responden ditanya mengenai pilihan apa yang mereka inginkan untuk melakukan sesuatu atau bagaimana mereka membuat ranking atau pilihan tertentu di dalam satu atau beberapa situasi dugaan. Data yang didapat dari responden selanjutnya dianalisa untuk mendapatkan suatu model berupa formulasi yang mencerminkan utilitas individu. Sifat utama dari *stated preference* adalah :

1. *Stated preference* didasarkan pada pendapat responden tentang bagaimana respon mereka terhadap beberapa alternatif.
2. Alat interview yang diberikan oleh metode ini memberikan alternatif yang dapat dimengerti oleh responden, tersusun rapi dan masuk akal.
3. Respon setiap jawaban yang diberikan oleh individu dianalisa untuk mendapatkan ukuran secara kuantitatif mengenai hal yang penting pada setiap atribut.

Metode ini digunakan secara luas dalam bidang transportasi karena metode ini dapat mengukur / memperkirakan bagaimana masyarakat memilih moda perjalanan yang belum ada atau melihat reaksi masyarakat terhadap suatu peraturan baru. *Stated preference* berarti pernyataan preferensi tentang suatu alternatif dibandingkan alternatif – alternatif yang lain. Teknik ini menggunakan pernyataan preferensi dari para responden untuk menentukan alternatif rancangan yang terbaik dari beberapa macam pilihan rancangan.

Data *stated preference* memiliki beberapa kelebihan dibandingkan dengan metode survei yang lainya diantaranya adalah :

1. Data survei yang lain rata – rata memiliki pengertian yang sesuai dengan perilaku nyata, tetapi data *stated preference* mungkin berbeda dengan perilaku ntanya.
2. Metode *stated preference* secara langsung dapat diterapkan untuk perencanaan alternatif yang baru (*non existing*)
3. Format pilihan respon dapat bervariasi misalnya memilih salah satu *rangking, rating* dan *choice*, sedangkan format pilihan untuk metode survei yang lain hanya berupa *choice*.
4. Kelebihan metode survei dengan teknik *stated preference* terletak pada kebebasannya untu melakukan desain pertanyaan untuk berbagai situasi dalam rangka memenuhi kebutuhan penelitian yang diperlukan.

B. Revealed Preference

Survei *Revealed Preference* adalah suatu bentuk kuisisioner survei yang menyatakan kepada para responden mengenai hal – hal yang sudah nyata tentang suatu yang menjadi obyek penelitian dan para responden diminta untuk memberikan tanggapannya terhadap setiap pertanyaan yang terdapat pada kuisisioner. Jawaban yang diberikan oleh para responden itu berkaitan dengan pengalaman para responden itu sendiri terhadap segala permasalahan yang terdapat pada lembar kuisisioner.

Jawaban yang diberikan oleh responden merupakan ciri dan perilaku serta pengalaman dari responden sehingga pertanyaan dalam kuisisioner harus disusun dengan cermat dan mudah untuk dimengerti. Untuk mempermudah pelaksanaan survei, penyusunan kuisisioner *revealed preference*, jawaban dari responden harus telah dikelompokkan terlebih dahulu ke dalam beberapa kelompok jawaban sehingga para responden cukup memilih dengan memberi tanda silang pada pilihan (*option*).

C. Teknik Sampling

1. Pengertian

Rancangan sampling adalah metode untuk memilih sampel yang dapat digunakan untuk menghasilkan himpunan data sampel kita. Tujuan utama dari setiap rancangan sampling adalah memberikan pedoman untuk memilih sampel yang mewakili populasi, sehingga dapat menyediakan sejumlah informasi tentang populasi dengan biaya minimum.

Menurut Amudi Pasaribu (1965), pengambilan sampel yang juga disebut sebagai penarikan sampel, bertujuan untuk memperoleh keterangan mengenai populasi dengan mengamati sebagian saja dari populasi tersebut.

Pengambilan sampel didasarkan pada anggapan – anggapan bahwa pada suatu populasi terdapat perbedaan – perbedaan atau simpangan – simpangan antara anggota – anggota populasi, yaitu perbedaan sifat – sifat anggota dan sifat umum populasi tersebut. Setiap anggota populasi dianggap berbeda dengan keadaan rata – rata dari populasi tersebut. Jika pengamatan dalam populasi itu dinyatakan dengan bilangan, maka sebagian dari anggota populasi tersebut lebih kecil dan sebagian lagi lebih besar dari harga rata – ratanya. Teori pengambilan sampel didasarkan atas adanya pengaruh saling menghilangkan diantara anggota populasi tersebut.

2. Cara Penarikan Sampel

Berkaitan dengan pengambilan sampel untuk survey pada penelitian untuk mengetahui persentase kemungkinan berpindah moda dari kendaraan pribadi ke Trans Jogja, *Ortuzar* dalam bukunya *Modelling transport* pada bab *data collection methods* memberikan ukuran sampel yang digunakan berdasarkan besarnya populasi seperti pada tabel 3.1 berikut :

Tabel 3.1 Ukuran Sampel yang Direkomendasikan Pada Survey
Tradisional

Besar Populasi	Ukuran Sampel Direkomendasikan	Minimum
< 50.000	1/5	1/10
50.000 - 150.000	1/8	1/20
150.000 - 300.000	1/10	1/35
300.000 - 500.000	1/15	1/50
500.000 – 1.000.000	1/20	1/70
>1.000.000	1/25	1/100

Sumber : Ortuzar, J.D and Willumsem, L.G (1994)

3. Kesalahan Dalam Sampling

Dalam melakukan survey, terdapat beberapa sumber kesalahan yang mungkin terjadi, sumber yang paling umum adalah variasi acak (random sampling), dalam contoh variabel pendapatan yang secara kebetulan memiliki variasi pendapatan yang sama. Sumber kesalahan yang lain dalam teknik sampling yang lain adalah spesifikasi populasi. Kesalahan yang disebabkan oleh spesifikasi populasi populasi dapat muncul dari sumber – sumber, misalnya : daftar unsur populasi yang tidak benar, anggota sampel yang keliru, kesalahan dalam pengumpulan informasi tentang sampel ataupun kesalahan dalam proses informasi sampel.

4. Teori Penarikan Sampel

Teori penarikan sampel merupakan suatu kajian tentang hubungan antara populasi dengan sampel yang diambil dari populasi tersebut. Teori ini sangat penting. Misalnya, teori ini dapat digunakan untuk menduga jumlah populasi yang tidak diketahui (seperti nilai tengah dan varians populasi, dan sebagainya yang disebut parameter populasi atau parameter, dari apa yang diketahui

mengenai kuantitas sampel yang berasal dari populasi yang bersangkutan yang disebut statistic sampel atau statistik.

Teori penarikan sampel juga berguna di dalam menentukan apakah perbedaan – perbedaan yang nampak antara dua sampel disebabkan oleh variasi secara kebetulan atau apakah memang perbedaan itu tidak terjadi secara kebetulan (signifikan).

Studi tentang inferensi yang dibuat berkenaan dengan suatu populasi dengan penggunaan sampel yang diambil dari populasi tersebut, bersama dengan indikasi kecermatan inferensi semacam itu yang menggunakan teori probabilitas disebut inferensi statistika (*statistical inference*).

5. Sampel Acak atau Bilangan Acak

Penarikan sampel secara acak merupakan satu cara dengan mana suatu sampel yang bersifat mewakili dapat diperoleh dengan melalui suatu proses penarikan sampel secara acak, dimana setiap unsur dalam populasi mempunyai probabilitas yang sama untuk dapat terpilih dalam sampel. Teknik untuk memperoleh sampel yang acak adalah dengan memberi nomor kepada setiap unsur dari populasi yang bersangkutan, menuliskan nomor – nomor ini masing – masing di atas secara kertas kecil, menempatkan kertas – kertas kecil ini dalam sebuah kotak dan setelah kertas – kertas ini diaduk – aduk, kemudian mengambil nomor – nomor tersebut.

D. Regresi Logistik Biner

Metode regresi merupakan analisis data yang mendeskripsikan hubungan kualitas antara variabel respond dan prediktor (Hosmer dan Lomeshow, 2000). Perbedaan mendasar antara regresi linier dan regresi logistik adalah type dari variabel respon. Regresi logistik merupakan salah satu metode yang dapat digunakan untuk mendapat hubungan antara variabel respon yang bersifat kategorik dengan variabel predictor (Agresti, 1990).

Menentukan nilai logistik biner dapat menggunakan persamaan sebagai berikut :

$$\text{Logit (p)} = \text{constant} - \text{variabel pengaruh} \dots\dots\dots 3.4$$

Berdasarkan persamaan diatas diperoleh nilai logit (p) yang digunakan untuk menentukan persentase probabilitas perpindahan menggunakan rumus sebagai berikut :

$$p = \frac{1}{1 + \exp^{-\text{logit (p)}}} \dots\dots\dots 3.5$$

Berdasarkan jenis skala data, regresi logistic dibedakan menjadi tiga macam, yaitu :

1. Regresi Logistik Biner (*Binary Logistic Regression*)

Regresi logistic biner ialah regresi dengan variabel respon yang mempunyai dua kategori / dua kejadian, yakni sukses atau gagal. Dengan demikian sering disebut dengan regresi logistic biner. Sedangkan jenis pada variabel prediktor dapat berupa nominal, interval, maupun ratio.

2. Regresi Logistik Multinomial (*Multinomial Logistic Regression*)

Data variabel respon pada regresi logistik multinomial yang digunakan adalah data berskala nominal dengan lebih dari 2 kategori. Sedangkan jenis data pada variabel prediktor dapat berupa nominal, ordinal , interval maupun ratio.

3. Regresi Logistik Ordinal (*Ordinal Logistic Regression*)

Data variabel respon pada regresi logistic ordinal yang digunakan adalah data berskala nominal dengan variabel dari 2 kategori. Sedangkan jenis data pada variabel prediktor dapat berupa nominal, ordinal, interval maupun ratio.