**BAHAN AJAR ANALISIS REGRESI**

**Oleh: ENDANG LISTYANI**

**ANALISIS REGRESI**

Dalam kehidupan sehari-hari sering kali ingin diketahui hubungan antar peubah, misalnya hubungan antara : prestasi belajar dengan IQ, tingkat pendidikan ibu dengan gizi balita, dan sebagainya. Umumnya suatu peubah bersifat mempengaruhi peubah yang lainnya. Peubah yang mempengaruhi disebut peubah bebas sedangkan yang dipengaruhi disebut sebagai peubah tak bebas atau peubah terikat.

Secara kuantitatif hubungan antara peubah bebas dan peubah terikat dapat dimodelkan dalam suatu persamaan matematik, sehingga dapat diduga nilai suatu peubah terikat bila diketahui nilai peubah bebasnya. Persamaan matematik yang menggambarkan hubungan antara peubah bebas dan terikat sering disebut persamaan regresi.

Persamaan regresi dapat terdiri dari satu atau lebih peubah bebas dan satu peubah terikat. Persamaan yang terdiri dari satu peubah bebas dan satu peubah terikat disebut persamaan regresi sederhana, sedangkan yang terdiri dari satu peubah terikat dan beberapa peubah bebas disebut persamaan regresi berganda. Regresi dapat dipisahkan menjadi regresi linear dan regresi non linear

Misalkan kita mempunyai sejumlah data berpasangan {(xi , yi), i = 1, 2, 3, . . ., n} data itu dapat diplotkan atau digambarkan pada bidang Kartesius yang disebut sebagai diagram pencar atau diagram hambur. Dari diagram pencar dapat diperkirakan hubungan antara peubah-peubah itu apakah mempunyai hubungan linear atau tidak linear.

**13.1 Regresi Linear Sederhana**

Regresi linear sederhana adalah persamaan regresi yang menggambarkan hubungan antara satu peubah bebas (X) dan satu peubah tak bebas (Y), dimana hubungan keduanya dapat digambarkan sebagai suatu garis lurus. Hubungan kedua peubah tersebut dapat dituliskan dalam bentuk persamaan:

Yi = β0 + β1 + εi............... (13.1)

Y = Peubah tak bebas, X = Peubah bebas, β0 = intersep/perpotongan dengan sumbu tegak, β1 = Kemiringan/gradien, εi *error* yang saling bebas dan menyebar normal *N*(0,σ2) *i* = 1, 2, …, *n*.

Dalam kenyataan seringkali kita tidak dapat mengamati seluruh anggota populasi, sehingga hanya mengambil sampel misalkan sampel itu berukuran n dan ditulis sebagai {(xi , yi), i = 1, 2, 3, . . ., n}. Persamaan yang diperoleh adalah dugaan dari persamaan (12.1) dan dapat dituliskan sebagai:

 = b0 + b1 Xi (13.2)

b0 adalah penduga untuk β0, dan b1 adalah penduga untuk β1.

Untuk peubah bebas xi nilai pengamatan yi tidak selalu tepat berada pada garis = β0 + β1(garis regresi populasi) atau  = b0 + b1 Xi (garis regresi sampel)

Y

X

 yi  = b0 + b1 Xi

 ei

 **Gambar13.1 Garis penduga hubungan antara peubah X dan Y**

Terdapat simpangan sebesar ei (untuk sampel) atau  (untuk populasi), sehingga

Yi =  + ei atau Yi =  + 

atau

Yi = b0 + b1 Xi  + ei (model regresi sampel)

Yi =β0 + β1 +  (model regresi populasi)

Anggapan/asumsi dalam analisis regresi linear sederhana dengan model

Yi =βo + β1 +  adalah:.

1)  merupakan galat acak yang menyebar normal dengan E() = 0 dan

 Var() =  untuk semua i

2) Yi menyebar normal dengan E(Yi) = βo + β1 dan Var(Yi) =  untuk semua i

**Pendugaan Parameter β0 dan β1**

Untuk menduga nilai parameter β0 dan β1terdapat bermacam-macammetode, misalnya metode kuadrat terkecil (*least square method*), metode kemungkinan maksimum (*maximum likelihood method*), metode kuadrat terkecil terboboti (*weighted least square method)*, dsb.

 Disini metode yang digunakan adalah metode kuadrat terkecil, karena mudah dikerjakan secara manual. Prinsip dasar metode kuadrat terkecil adalah meminimumkan jumlah kuadrat simpangan atau Jumlah Kuadrat Galat

 (JKG)== 

Dengan menggunakan bantuan pelajaran kalkulus, diperoleh nilai dugaan parameter regresi sebagai berikut:

Dengan demikian dapat diperoleh hubungan;

**Contoh 13.1**

Diketahui data percobaan

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subjek i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| xi | 1,5 | 1,8 | 2,4 | 3,0 | 3,5 | 3,9 | 4,4 | 4,8 | 5,0 |
| yi | 4,8 | 5,7 | 7,0 | 8,3 | 10,9 | 12,4 | 13,1 | 13,6 | 15,3 |

Tentukan persamaan regresi dugaan

**Jawab**

Dengan menggunakan kalkulator dapat dengan mudah dihitung

= 30,3 = 91,1 = 345,09

= 115,11 = 3,3667  = 10, 1222



*b*o = 10,1222 – (2,9303)(3,3667) = 0,2568

Jadi persamaan regresi dugaan  = 0,26 + 2,93X

**Pengujian terhadap Model Regresi**

Proses selanjutnya setelah melakukan pendugaan parameter model regresi sederhana adalah pengujian terhadap model regresi apakah signifikan atau tidak, yang dapat dilakukan dengan dua cara yaitu ANAVA dengan uji F dan uji parsial dengan uji t.

Uji bagi β1=0 lawan β1≠0 melalui ANAVA

Hipotesis

 H0 : β1=0 (Tidak ada hubungan linear antara X dan Y)

 H1 : β1≠ 0 (Ada hubungan linear antara X dan Y)

Tabel 13.1. Anava untuk pengujian pada model regresi linear sederhana

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sumber Keragaman** | **db** | **JK** | **KT** | **Fhit** | **Ftabel** |
| RegresiGalat | 1n−2 | JKRJKG | KTR=JKR/1 KTG=JKG/(n −2) | Fhit=KTR/KTG | Fα(1,n−2) |
| Total | n−1 | JKT |   |  |  |

Ho  ditolak jika Fhit  > Ftabel, yang berarti model regresi signifikan atau ada hubungan liner anatara X dan Y

Keterangan

1. Uji bagi β1=0 lawan β1≠0 melalui uji t

Hipotesis

 H0 : β1=0 (Tidak ada hubungan linear antara X dan Y)

 H1 : β1≠ 0 (Ada hubungan linear antara X dan Y)

Statistik uji adalah :

dengan

Kriteria keputusan :

 H0 ditolak jika |thit|> tα/2(n−2)

1. Uji bagi β0=0 lawan β0 ≠0 melalui uji t

Hipotesis

 H0 : β0=0

 H1 : β0≠ 0

Statistik uji adalah :

dengan

Kriteria keputusan :

 H0 ditolak jika |thit|> tα/2(n−2)

Perhitungan untuk uji hipotesis menggunakan data Contoh 13.1.

Dari perhitungan sebelumnya telah diperoleh

= 30,3 = 91,1 = 345,09

= 115,11  = 3,3667  = 10, 1222

 b0 = 0,2568 b1 = 2,9303

Dengan demikian diperoleh

JKT = 1036,65 − 9. (10,1222)2 = 114,52

JKG = 1036,65 − (0,2568) 91,1 – (2,9303) 345,09 = 2,0383

JKR = 945,55 –2,0383 = 112,4813

Tabel anava untuk data tersebut disajikan dalam Tabel 13.2.

Tabel 13.2. Anava untuk data pada Contoh 13.1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sumber Keragaman** | **db** | **JK** | **KT** | **Fhit** | **Ftabel** |
| RegresiGalat | 17 | 112,48132,0383 |  KTR=112,4813 KTG=0,2911 | Fhit=386,2885 | F0,05(1,7) =5,59 |
| Total | 8 | 114,52 |   |  |  |

Berdasarkan hasil pada Tabel 13.2 diperoleh nilai F hitung lebih besar daripada nilai F tabel, sehingga H0 ditolak. Jadi ada hubungan linear antara variabel X dan Y.

Untuk uji parsial perlu dihitung terlebih dahulu nilai

dan

  0,284

Jadi untuk uji signifikansi koefisien β1

thit = 

sedangkan untuk uji signifikansi konstanta diperoleh

thit = 

Karena t tabel adalah t0,025;7 = 2,365 maka H0 ditolak untuk uji koefisien β1 dan H0 diterima untuk uji signifikansi konstanta.

Regresi Ganda

Regresi linear ganda adalah persamaan regresi yang menggambarkan hubungan antara lebih dari satu peubah bebas (X) dan satu peubah tak bebas (Y) Hubungan peubah-peubah tersebut dapat dituliskan dalam bentuk persamaan:



Y = Peubah tak bebas, X = Peubah bebas, β0 = intersep/perpotongan dengan sumbu tegak, β1, β2, ...., βp−1 = parameter model regresi, εi saling bebas dan menyebar normal *N*(0,σ2) , *i* = 1, 2, …, *n*

Persamaan regresi dugaannya adalah

Hipotesis yang harus diuji dalam analisis regresi ganda adalah

H0 : β1 = β2 = … = βp-1=0

H1 : Tidak semua βk (k=1,2,…,p −1) sama dengan nol

Untuk melakukan pendugaan parameter model regresi ganda dan menguji signifikansinya dapat dilakukan dengan program SPSS 16.

Asumsi yang harus dipenuhi dalam analisis regresi ganda adalah :

1. Tidak ada multikolinearitas (korelasi antara variabel independen)
2. Heteroskedastisitas (variansi *error* konstan)
3. Normalitas (*error* berdistribusi normal)
4. Autokorelasi (*error* bersifat acak)

Multikolinearitas

1. Multikolinearitas atau kekolinearan ganda adalah terjadinya korelasi antar peubah bebas.
2. Model regresi yang baik seharusnya tidak terjadi korelasi antar peubah bebas.
3. Metode yang banyak digunakan untuk mendeteksi adanya multikolinearitas adalah faktor inflasi ragam (*variance inflation factor/VIF*)
4. Multikolinearitas terjadi jika nilai VIF > 10

Heteroskedastisitas

1. Ragam galat diasumsikan konstan dari satu pengamatan ke pengamatan lain, hal ini disebut **homoskedastisitas**.
2. Jika ragam galat berbeda disebut **heteroskedastisitas**.
3. Model regresi yang baik adalah tidak terjadi heteroskedastisitas.
4. Untuk mendeteksi heteroskedastisitas adalah dengan membuat plot nilai dugaan yang dibakukan (*standardized predicted value*) dengan sisaan yang dibakukan (*studentized residual*).
5. Jika ada pola tertentu (bergelombang, melebar kemudian menyempit) maka **terjadi heteroskedastisitas**.
6. Jika tidak ada pola jelas, serta titik-titik (sisaan) menyebar di atas dan di bawah angka 0 pada sumbu Y, maka **tidak terjadi heteroskedastisitas**.

Normalitas (*error* berdistribusi normal)

1. Untuk mendeteksi normalitas digunakan *normal p-p plot*.
2. Jika titik-titik (sisaan) menyebar di sekitar garis diagonal dan mengikuti arah garis diagonal, maka model regresi memenuhi asumsi normalitas.
3. Jika titik-titik (sisaan) menyebar jauh dari garis diagonal dan atau tidak mengikuti arah garis diagonal, maka model regresi tidak memenuhi asumsi normalitas

Autokorelasi.

1. Bila dalam model regresi linear ganda ada korelasi antara galat pada periode t dengan galat pada periode t-1, maka dinamakan ada masalah autokorelasi.
2. Model regresi yang baik adalah model regresi yang bebas dari autokorelasi.

**Contoh 13.2**

 Misalkan dipunyai data

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Y | 10 | 6 | 5 | 12 | 10 | 15 | 5 | 12 | 17 | 20 |
| X1 | 1.3 | 2.0 | 1.7 | 1.5 | 1.6 | 1.2 | 1.6 | 1.4 | 1.0 | 1.1 |
| X2 | 9 | 7 | 5 | 14 | 15 | 12 | 6 | 10 | 15 | 21 |

Akan dilakukan pendugaan dan pengujian parameter model regresi, serta uji asumsi dengan menggunakan SPSS 16.

1. Cara memasukkan data dan melakukan analisis sama dengan pada regresi sederhana.
2. Untuk memunculkan hasil uji asumsi pada kotak dialog statistics klik juga collinearity diagnostics baru continue, sebagaimana terlihat pada gambar berikut:



1. Untuk melakukan uji asumsi pada residual klik plots, sehingga akan muncul kotak dialog :



1. Masukkan ZPRED pada kotak X dan ZRESID pada kotak Y, dan beri tanda centang (√ ) pada Normal probability plot, kemudian klik continue. Kembali ke kotak dialog awal, dan klik OK.

Hasil analisis dengan ANAVA adalah sebagai berikut:

| **ANOVAb** |
| --- |
| Model | Sum of Squares | df | Mean Square | F | Sig. |
| 1 | Regression | 217.699 | 2 | 108.849 | 47.917 | .000a |
| Residual | 15.901 | 7 | 2.272 |  |  |
| Total | 233.600 | 9 |  |  |  |
| a. Predictors: (Constant), VAR00003, VAR00002 |  |  |
| b. Dependent Variable: VAR00001 |  |  |  |

Terlihat bahwa nilai signifikansi 0,000 < 1%, sehingga H0 ditolak, yang berarti ada hubungan linear antara variabel independen X1 dan X2 dengan variabel dependen Y.

Hasil uji parsial adalah sebagai berikut :

| **Coefficientsa** |
| --- |
| Model | Unstandardized Coefficients | Standardized Coefficients | t | Sig. | Collinearity Statistics |
| B | Std. Error | Beta | Tolerance | VIF |
| 1 | (Constant) | 16.406 | 4.343 |  | 3.778 | .007 |  |  |
| X1 | -8.248 | 2.196 | -.490 | -3.756 | .007 | .572 | 1.749 |
| X2 | .585 | .134 | .571 | 4.377 | .003 | .572 | 1.749 |
| a. Dependent Variable: VAR00001 |  |  |  |  |  |

Karena nilai signifikansi 0,007 untuk konstanta dan VAR00002 dan 0,003 untuk VAR00003, sehingga H0 ditolak untuk semua uji. Jadi konstanta β0 semua dan koefisien regresi β1, dan β2 signifikan. Persamaan regresi dugaannya adalah :



Hasil uji asumsi multikolinearitas dapat dilihat pada nilai VIF, yaitu 1,749 < 10, sehingga dapat disimpulkan tidak ada multikolinearitas antara variabel X1 dan X2. Hasil uji normalitas dari error dapat dilihat pada output berikut



Karena plot mendekati garis diagonal, maka dapat disimpulkan *error* memenuhi asumsi normalitas. Uji normalitas error juga dapat dilakukan dengan uji Kolmogorov-Smirnov

Hasil plot berikut menunjukkan tidak ada pola yang jelas atau berpola acak, sehingga dapat disimpulkan tidak terjadi **heteroskedastisitas** atau ragam galat konstan dan galat bersifat acak atau tidak ada autokorelasi .



**Latihan 13.**

1. Suatu sampel acak terdiri atas 20 keluarga di suatu daerah, memberikan data sbb.:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 15 | 20 | 25 | 20 | 25 | 30 | 16 | 15 | 25 | 20 |
| Y | 10 | 15 | 20 | 16 | 22 | 25 | 15 | 14 | 10 | 18 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 16 | 18 | 20 | 25 | 30 | 25 | 19 | 10 | 20 | 20 |
| Y | 12 | 15 | 15 | 20 | 25 | 23 | 16 | 8 | 15 | 17 |

X = pendapatan keluarga perbulan dalam ratusan ribu rupiah

Y = pengeluaran keluarga perbulan dalam ratusan ribu rupiah

a) Jika diduga bahwa hubungan antara pendapatan keluarga dan pengeluaran

 keluarga linear, tentukan persamaan regresi dugaannya

b) Bila dianggap asumsi-asumsi dalam analisis regresi linear terpenuhi, ujilah apakah ada hubungan antara pendapatan keluarga perbulan dan pengeluaran keluarga perbulan. Gunakan α = 0,05.

2. Suatu penelitian dilakukan terhadap 20 mahasiswa semester satu yang diambil secara acak untuk menentukan apakah nilai mutu rata-rata (NMR) pada akhir tahun pertama (Y) dapat diprediksi dari nilai ujian masuk (X). Data yang diperoleh sbb.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 5,5 | 4,8 | 4,7 | 3,9 | 4,5 | 6,2 | 6,0 | 5,2 | 4,7 | 4,3 |
| Y | 3,1 | 2,3 | 3,0 | 1,9 | 2,5 | 3,7 | 3,4 | 2,6 | 2,8 | 1,6 |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 4,9 | 5,4 | 5,0 | 6,3 | 4,6 | 4,3 | 5,0 | 5,9 | 4,1 | 4,7 |
| Y | 2,0 | 2,9 | 2,3 | 3,2 | 1,8 | 1,4 | 2,0 | 3,8 | 2,2 | 1,5 |

1. Jika hubungan antar NMR dan nilai ujian masuk dapat dinyatakan dengan garis linear, tentukan persamaan regresi linear dugaannya.
2. Bila dianggap asumsi-asumsi dalam analisis regresi linear terpenuhi, ujilah apakah ada hubungan antara nilai ujian masuk dan nilai mutu rata-rata (NMR) pada akhir tahun pertama. Gunakan α = 0,05.
3. Tentukan nilai dugaan untuk NMR jika nilai ujian masuk 6,0
4. Bagian kepegawaian suatu perusahaan menggunakan 12 orang dalam suatu penelitian untuk menentukan hubungan antara nilai prestasi kerja (Y) dan nilai empat tes, yaitu tes kemampuan di bidang IT (X1), kemampuan berbahasa Inggris (X2), kemampuan bekerja sama (X3), dan kemampuan berkomunikasi (X4). Datanya adalah sebagai berikut

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Y | X1 | X2 | X3 | X4 |
| 11,214,517,217,819,324,521,216,914,820,013,222,5 | 56,559,569,274,581,288,078,269,058,180,558,384,0 | 71,072,576,079,584,086,280,072,068,085,071,087,2 | 38,538,242,543,547,547,444,541,842,148,137,551,0 | 43,044,849,056,360,262,058,148,146,060,347,165,2 |

* 1. Ujilah apakah ada hubungan linear antara nilai prestasi kerja (y) dan nilai empat tes, yaitu tes kemampuan di bidang IT, kemampuan berbahasa Inggris, dan kemampuan bekerja sama, kemampuan berkomunikasi. Gunakan α = 0,05.
	2. Manakah diantara empat variable yang secara signifikan berpengaruh terhadap prestasi kerja?
	3. Berdasarkan hasil b) Tentukan persamaan regresi linear dugaannya.
	4. Lakukan uji asumsi dalam analisis regresi linear dan simpulkan hasilnya.
1. Daya rentang produk fiber sintetis diperkirakan berhubungan dengan persentase bahan katun dalam fiber, waktu pengeringan fiber. Hasil percobaan terhadap 10 potong fiber yang diproduksi dalam beberapa kondisi yang berbeda diberikan pada Tabel berikut

|  |  |  |
| --- | --- | --- |
| Y | X1 | X2 |
| 213220216225235218239243233240 | 13151418192022171618 | 2,12,32,22,53,22,43,44,04,4.3 |

* 1. Lakukan analisis regresi untuk menguji apakah ada hubungan linear antara persentase bahan katun dalam fiber dan waktu pengeringan dengan daya rentang fiber sintetis.
	2. Tentukan persaman regresi dugaannya.

Lakukan uji asumsi dalam analisis regresi linear dan simpulkan hasilnya.