

# ANALISIS REGRESI

**Puji Muniarty  
Salsabila  
Lita Lovia  
Muhammad Ilham Riyadh  
Azis  
Santje Magdalena Iriyanto  
Alona Dwinata  
Rahma Faelasofi  
Bambang Suhartawan  
Suri Toding Lembang  
Rosi Ramayanti**



**GETPRESS INDONESIA**

## **ANALISIS REGRESI**

**Penulis :** Puji Muniarty

Salsabila

Lita Lovia

Muhammad Ilham Riyadh

Azis

Santje Magdalena Iriyanto

Alona Dwinata

Rahma Faelasofi

Bambang Suhartawan

Suri Toding Lembang

Rosi Ramayanti

**ISBN :**

**Editor :** Ari Yanto, M.Pd.

**Penyunting:** Yuliatr Novita, M. Hum.

**Desain Sampul dan Tata Letak :** Atyka Trianisa, S. Pd.

**Penerbit:** GET PRESS INDONESIA

Anggota IKAPI No. 033/SBA/2022

**Redaksi:**

Jl. Palarik RT 01 RW 06, Kelurahan Air Pacah  
Kecamatan Koto Tengah, Padang, Sumatera Barat

website: [www.getpress.co.id](http://www.getpress.co.id)  
email: [adm.getpress@gmail.com](mailto:adm.getpress@gmail.com)

Cetakan pertama, Agustus 2024

Hak cipta dilindungi undang-undang  
Dilarang memperbanyak karya tulis ini dalam bentuk  
dan dengan cara apapun tanpa izin tertulis dari penerbit.

## **KATA PENGANTAR**

Dengan mengucapkan puji syukur kehadiran Allah SWT, atas limpahan rahmat dan hidayahNya, maka Penulisan Buku dengan judul Analisis Regresi dapat diselesaikan dengan kerjasama tim penulis. Analisis Regresi merupakan buku yang berisikan bahasan mengenai, yaitu analisis regresi, analisis korelasi, regresi linier sederhana, regresi linear berganda, pendekatan matriks terhadap regresi linier, diagnosis dan pengujian asumsi pada model regresi, pemeriksaan sisaan model regresi linear, multikolinier, pemilihan model regresi terbaik, uji parsial, dan sekuensial, hipotesis linier umum, dan uji parsial, sekuensial dan hipotesis linier umum.

Buku ini masih banyak kekurangan dalam penyusunannya. Oleh karena itu, kami sangat mengharapkan kritik dan saran demi perbaikan dan kesempurnaan buku ini selanjutnya. Kami mengucapkan terima kasih kepada berbagai pihak yang telah membantu dalam proses penyelesaian Buku ini. Semoga Buku ini dapat menjadi sumber referensi dan literatur yang mudah dipahami.

Padang,      Juli 2024  
Penulis

## DAFTAR ISI

|                                                                         |           |
|-------------------------------------------------------------------------|-----------|
| <b>ANALISIS REGRESI.....</b>                                            | <b>1</b>  |
| <b>KATA PENGANTAR.....</b>                                              | <b>i</b>  |
| <b>DAFTAR ISI .....</b>                                                 | <b>ii</b> |
| <b>DAFTAR TABEL .....</b>                                               | <b>iv</b> |
| <b>BAB 1 ANALISIS REGRESI .....</b>                                     | <b>1</b>  |
| 1.1 Pendahuluan .....                                                   | 1         |
| 1.2 Jenis-Jenis Analisis Regresi .....                                  | 2         |
| <b>DAFTAR PUSTAKA.....</b>                                              | <b>7</b>  |
| <b>BAB 2 ANALISIS KORELASI .....</b>                                    | <b>9</b>  |
| 2.1 Koefisien Korelasi ( $r$ ) .....                                    | 10        |
| 2.2 Uji Hipotesis Analisis Korelasi .....                               | 15        |
| 2.3 Analisis Korelasi dengan SPSS .....                                 | 17        |
| <b>DAFTAR PUSTAKA.....</b>                                              | <b>19</b> |
| <b>BAB 3 REGRESI LINIER SEDERHANA.....</b>                              | <b>21</b> |
| 3.1 Pendahuluan .....                                                   | 21        |
| 3.2 Model Regresi Linier Sederhana .....                                | 21        |
| 3.3 Metode Kuadrat Terkecil .....                                       | 22        |
| 3.4 Persamaan Regresi Melewati Titik Asal.....                          | 25        |
| 3.5 Koefisien Determinasi ( $R^2$ ) .....                               | 26        |
| 3.6 Asumsi Dalam Analisis Regresi Linier Sederhana .....                | 27        |
| 3.7 Pembentukan Selang Kepercayaan Bagi $\beta_0$ dan $\beta_1$ .....   | 27        |
| 3.8 Uji keberartian Garis Regresi .....                                 | 28        |
| <b>DAFTAR PUSTAKA.....</b>                                              | <b>34</b> |
| <b>BAB 4 REGRESI LINEAR BERGANDA .....</b>                              | <b>36</b> |
| 4.1 Pendahuluan .....                                                   | 36        |
| 4.2 Model Regresi Berganda .....                                        | 37        |
| <b>DAFTAR PUSTAKA.....</b>                                              | <b>53</b> |
| <b>BAB 5 PENDEKATAN MATRIKS TERHADAP REGRESI LINIER ....</b>            | <b>54</b> |
| 5.1 Pendahuluan .....                                                   | 54        |
| 5.2 Pendekatan Matriks .....                                            | 56        |
| <b>DAFTAR PUSTAKA.....</b>                                              | <b>62</b> |
| <b>BAB 6 DIAGNOSIS DAN PENGUJIAN ASUMSI PADA MODEL<br/>REGRESI.....</b> | <b>64</b> |
| 6.1 Asumsi Linearitas .....                                             | 64        |
| 6.2 Asumsi Independensi .....                                           | 66        |

|                                                                          |            |
|--------------------------------------------------------------------------|------------|
| 6.3 Asumsi Homoskedastisitas .....                                       | 67         |
| 6.4 Asumsi Normalitas Residual.....                                      | 68         |
| 6.5 Asumsi Tidak Ada Multikolinearitas .....                             | 70         |
| 6.6 Asumsi Tidak Ada Pengaruh Leverage yang Berlebihan.....              | 71         |
| <b>DAFTAR PUSTAKA.....</b>                                               | <b>74</b>  |
| <b>BAB 7 PEMERIKSAAN SISAAN MODEL REGRESI LINEAR .....</b>               | <b>76</b>  |
| 7.1 Pendahuluan .....                                                    | 76         |
| 7.2 Penskalaan Sisaan.....                                               | 78         |
| <b>DAFTAR PUSTAKA.....</b>                                               | <b>89</b>  |
| <b>BAB 8 MULTIKOLINIER.....</b>                                          | <b>90</b>  |
| 8.1 Pendahuluan .....                                                    | 90         |
| 8.2 Pengertian Multikolinier .....                                       | 90         |
| 8.3 Efek Multikolinear .....                                             | 93         |
| 8.4 Diagnosis Multikolinear.....                                         | 97         |
| <b>DAFTAR PUSTAKA.....</b>                                               | <b>105</b> |
| <b>BAB 9 PEMILIHAN MODEL REGRESI TERBAIK.....</b>                        | <b>107</b> |
| 9.1 Pengertian Analisis Regresi .....                                    | 107        |
| 9.2 Jenis-jenis Model Regresi dalam Penelitian.....                      | 108        |
| 9.3 Pemilihan Model Regresi Terbaik .....                                | 110        |
| <b>DAFTAR PUSTAKA.....</b>                                               | <b>130</b> |
| <b>BAB 10 UJI PARSIAL, SEKUENSIAL DAN HIPOTESIS LINIER<br/>UMUM.....</b> | <b>131</b> |
| 10.1 Uji Parsial.....                                                    | 131        |
| 10.2 Sekuensial.....                                                     | 137        |
| <b>DAFTAR PUSTAKA.....</b>                                               | <b>143</b> |
| <b>BAB 11 KORELASI DAN REGRESI LINEAR SEDERHANA .....</b>                | <b>145</b> |
| 11.1 Korelasi yang Terjadi Antara Dua Variabel.....                      | 145        |
| 11.2 Koefisien Korelasi Sederhana.....                                   | 146        |
| 11.3 Interval Keeratan Korelasi Antar Variabel.....                      | 146        |
| 11.4 Koefisien Determinasi .....                                         | 147        |
| 11.5 Regresi Linier Sederhana .....                                      | 147        |
| 11.6 Model Regresi Sederhana.....                                        | 148        |
| 11.7 Kesalahan Baku Estimasi.....                                        | 148        |
| 11.8 Pengujian Hipotesis Tentang Koefisien Korelasi .....                | 154        |
| 11.9 Pengujian Hipotesis Tentang Regresi.....                            | 156        |
| <b>DAFTAR PUSTAKA.....</b>                                               | <b>157</b> |
| <b>BIODATA PENULIS .....</b>                                             | <b>159</b> |

## DAFTAR TABEL

|                                                                         |     |
|-------------------------------------------------------------------------|-----|
| Tabel 4. 1 Pengaruh Pupuk dan Tenaga Kerja Terhadap Produksi Cabai..... | 38  |
| Tabel 4. 2 Ilustrasi Mencari Nilai T Tabel.....                         | 45  |
| Tabel 4. 3 Ilustrasi penggunaan Ftabel.....                             | 47  |
| Tabel 4. 4 Hasil Olah Data Menggunakan Penggunaan Software SPSS .....   | 48  |
| Tabel 4.5 Hasil Model Summary Penggunaan Software SPSS .....            | 48  |
| Tabel 4. 6 Hasil Anova Penggunaan Software SPSS.....                    | 49  |
| Tabel 4. 7 nilai koefisien pada Penggunaan Software SPSS.....           | 50  |
| Tabel 7. 1. Data Simulasi .....                                         | 83  |
| Tabel 7. 2. Pemeriksaan pengamatan berpengaruh .....                    | 86  |
| Tabel 11. 1. Volume penjualan dan harga jual produk PT. ABC.....        | 149 |
| Tabel 11. 2. Perhitungan data 1.....                                    | 151 |
| Tabel 11. 3. Perhitungan data 2.....                                    | 154 |

# **BAB 1**

## **ANALISIS REGRESI**

**Oleh Puji Muniarty**

### **1.1 Pendahuluan**

Dalam analisis regresi statistik, hubungan antara dua variabel biasanya dibentuk dalam persamaan regresi yang menguji secara matematis hubungan antara variabel bebas (X) dan variabel terikat (Y). Persamaan ini memuat parameter parameter yang menggambarkan hubungan kuantitatif antara variabel bebas dan variabel terikat. Selanjutnya hubungan antara variabel independen dan dependen dalam regresi digunakan untuk memprediksi berbagai pertanyaan penelitian di bidang ini. Variabel adalah suatu nilai yang tidak tetap dan/atau dapat diartikan sebagai ciri dari objek yang diteliti.

Dalam analisis regresi terdapat dua variabel yang membentuk persamaan regresi merupakan variabel terikat dan variabel terikat. Ada dua jenis analisis regresi linier: analisis regresi sederhana dan analisis regresi berganda. Metode yang paling umum digunakan untuk memperkirakan parameter regresi adalah metode kuadrat terkecil.

Metode ini dikemukakan oleh matematikawan Jerman Karl Friedrich Gauss. Gauss menyatakan asumsi analisis regresi adalah normalitas residu, independensi residu (non-autokorelasi), homogenitas varians residu, dan non-multikolinearitas. Saat membangun model regresi linier, ada tiga alasan mengapa Anda tidak dapat mengesampingkan

kemungkinan bahwa model yang diperoleh pada awalnya masih kurang optimal. Penyebabnya adalah: (1) terjadi overspesifikasi yaitu terlalu banyak variabel dalam model, dan (2) model tidak optimal. variabel yang benar tidak disertakan, dan (3) model tidak memiliki hubungan yang benar secara matematis (Johnson, 2006).

Ada beberapa cara untuk memilih variabel penjelas yang sesuai untuk dimasukkan ke dalam model untuk memperoleh model terbaik. Salah satunya adalah Best Subset Regtest (Hanum, 2011). Regresi linier membuat model berdasarkan hasil analisis statistik. Hal ini dapat bermanfaat jika model yang dipilih benar dan asumsi lain yang mendasari model tersebut terpenuhi (Johnson, 2006). Model regresi linier memerlukan asumsi klasik untuk mencapai hal tersebut. Dapatkan Penaksir Kuadrat Terkecil Biasa (OLS), penduga nonringkasan linier terbaik (BIRU).

## 1.2 Jenis-Jenis Analisis Regresi

Gujarati (2003) menjelaskan bahwa analisis regresi mempelajari hubungan antara suatu variabel yang disebut variabel terikat dan satu atau dua variabel penjelas. Variabel penjelas disebut variabel respon, dan variabel penjelas biasa disebut variabel bebas. Analisis regresi digunakan: (1) mengetahui pengaruh variabel bebas terhadap variabel terikat dan (2) meramalkan nilai variabel terikat karena adanya perubahan pada variabel bebas.

### 1. Regresi Sederhana

Regresi sederhana pada dasarnya mengandung 1 (satu) variabel bebas dan 1 (satu) variabel terikat yang akan membentuk garis lurus. Adapun persamaan regresi sederhana :  $Y = a + bX$ , dengan penentuan nilai :

- Konstanta  $\rightarrow a = \frac{\sum Y - b \sum X}{n}$
- Koefisien variabel  $\rightarrow b = \frac{n \cdot \sum(XY) - \sum X \cdot \sum Y}{n \cdot \sum(X^2) - (\sum X)^2}$



- n = Jumlah sampel penelitian

Kasus : PT. ABC memiliki pengeluaran promosi (X) dan Penerimaan Penjualan (Y) sebagai berikut :

**Tabel 1**

| Tahun | Pengeluaran Promosi (X) | Penerimaan Penjualan (Y) |
|-------|-------------------------|--------------------------|
| 2010  | 10                      | 44                       |
| 2011  | 9                       | 40                       |
| 2012  | 11                      | 42                       |
| 2013  | 12                      | 46                       |
| 2014  | 13                      | 48                       |
| 2015  | 13                      | 52                       |

Misalkan seorang peneliti melakukan penelitian dengan masing-masing variabel bebas dan variabel terikat masing-masing 1 (satu) variabel yang mana peneliti tersebut dalam mengolah dan menganalisis menggunakan SPSS dengan hasil output data sebagai berikut:

**Tabel 2**

**Coefficients<sup>a</sup>**

| Model               | Unstandardized Coefficients |            | Standardized Coefficients | t     | Sig. | 95,0% Confidence Interval for B |             |
|---------------------|-----------------------------|------------|---------------------------|-------|------|---------------------------------|-------------|
|                     | B                           | Std. Error | Beta                      |       |      | Lower Bound                     | Upper Bound |
| 1 (Constant)        | 18,700                      | 6,947      |                           | 2,692 | ,055 | -,589                           | 37,989      |
| Pengeluaran Promosi | 2,350                       | ,608       | ,888                      | 3,867 | ,018 | ,663                            | 4,037       |

a. Dependent Variable: Penerimaan Penjualan

Catatan : kita asumsikan variabel bebas = pengeluaran promosi, variabel terikat = penerimaan penjualan.

Dari tabel di atas maka dibentuk persamaan regresi linier sederhana sebagai berikut :  $Y = 18,7 - 2,350X$  dengan makna :

- Konstanta (a) = 18,7 bermakna jika variabel bebas konstan atau sama dengan nol maka variabel terikat sebesar 18,7.
- Koefisien variabel (b) = 2,350 bermakna jika variabel bebas naik Rp 1 maka penerimaan penjualan akan naik sebesar 2,350 demikian sebaliknya.

## 2. Regresi Berganda

Regresi berganda digunakan untuk tujuan peramalan yang mana dalam penelitian model tersebut ada sebuah variabel terikat dan beberapa variabel bebas. Adapun persamaan regresi berganda :

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

Sehingga untuk menentukan nilai a,  $b_1$  dan  $b_2$  sampai  $b_n$  dapat ditentukan dengan formulasi sebagai berikut :

$$\Sigma Y = n \cdot a + b_1 \cdot \Sigma X_1 + b_2 \cdot \Sigma X_2 + \dots + b_n \cdot \Sigma X_n$$

$$\Sigma (X_1 Y) = a \cdot \Sigma X_1 + b_1 \cdot \Sigma X_1^2 + b_2 \cdot \Sigma (X_1 \cdot X_2) + \dots + b_n \cdot \Sigma (X_1 \cdot X_n)$$

$$\Sigma (X_2 Y) = a \cdot \Sigma X_2 + b_1 \cdot \Sigma (X_1 \cdot X_2) + b_2 \cdot \Sigma X_2^2 + \dots + b_n \cdot \Sigma (X_2 \cdot X_n)$$

$$\Sigma (X_n Y) = a \cdot \Sigma X_n + b_1 \cdot \Sigma (X_1 \cdot X_n) + b_2 \cdot \Sigma (X_2 \cdot X_n) + \dots + b_n \cdot \Sigma (X_n \cdot X_n)$$

Atau :

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

$$b_1 = \frac{(\Sigma yx_1)(\Sigma x_2^2) - (\Sigma yx_2)(\Sigma x_1x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1x_2)^2}$$

$$b_2 = \frac{(\Sigma yx_2)(\Sigma x_1^2) - (\Sigma yx_1)(\Sigma x_1x_2)}{(\Sigma x_1^2)(\Sigma x_2^2) - (\Sigma x_1x_2)^2}$$

Kasus : misalkan pada PT XYZ seorang manajer keuangan menentukan besarnya pengaruh tingkat harga ( $X_1$ ), Pengeluaran promosi ( $X_2$ ) dan pendapatan rata-rata masyarakat ( $X_3$ ) terhadap kuantitas produk perusahaan yang terjual ( $Y$ ) dapat terlihat pada tabel dibawah ini dan analisis regresi berganda dapat dilakukan dengan analisis SPSS seperti yang terlihat dalam tabel berikut ini :

**Tabel 3**

| Thn  | Tk. Harga ( $X_1$ ) | Pengeluaran Promosi ( $X_2$ ) | Pendapatan rata2 Masyarakat ( $X_3$ ) | Kuantitas produk yg Terjual ( $Y$ ) |
|------|---------------------|-------------------------------|---------------------------------------|-------------------------------------|
| 2016 | 4                   | 1                             | 5                                     | 5                                   |
| 2017 | 6                   | 7                             | 4                                     | 15                                  |
| 2018 | 9                   | 4                             | 3                                     | 20                                  |
| 2019 | 8                   | 3                             | 2                                     | 17                                  |
| 2020 | 5                   | 5                             | 1                                     | 10                                  |

**Tabel 4**  
**Coefficients<sup>a</sup>**

| Model         | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. | 95,0% Confidence Interval for B |             |
|---------------|-----------------------------|------------|---------------------------|--------|------|---------------------------------|-------------|
|               | B                           | Std. Error | Beta                      |        |      | Lower Bound                     | Upper Bound |
| 1 (Constant)  | -8,417                      | 1,284      |                           | -6,556 | ,001 | -11,559                         | -5,275      |
| Tk. Harga     | 2,783                       | ,122       | ,968                      | 22,720 | ,000 | 2,483                           | 3,083       |
| Peng. Promosi | ,798                        | ,145       | ,295                      | 5,507  | ,002 | ,444                            | 1,153       |
| Pend. rata2   | ,237                        | ,213       | ,059                      | 1,111  | ,309 | -,285                           | ,759        |

a. Dependent Variable: Kuantitas Produk

Dari tabel 4 diatas dapat dibuat persamaan regresi berganda :

$$Y = -8,417 + 2,783X_1 + 0,798X_2 + 0,237X_3$$

Dengan makna :

- Konstanta =  $a = -8,417$  yang berarti jika semua variabel bebas ( $X_1, X_2, X_3$ ) konstan atau = 0 maka kuantitas produk yang terjual sebesar 8,417.
- Koefisien variabel =  $b_1 = 2,783$  yang berarti setiap penambahan Rp 1 tingkat harga akan meningkatkan kuantitas produk yang terjual sebesar 2,783 demikian sebaliknya dengan anggapan  $X_1$  dan  $X_2$  tetap atau konstan.
- Koefisien variabel =  $b_2 = 0,798$  yang berarti setiap penambahan Rp 1 pengeluaran promosi akan meningkatkan kuantitas produk yang terjual sebesar Rp 0,798 demikian sebaliknya dengan anggapan  $X_1$  dan  $X_3$  tetap atau konstan.
- Koefisien variabel =  $b_3 = 0,237$  yang berarti setiap penambahan Rp 1 pendapatan rata-rata masyarakat akan meningkatkan kuantitas produk yang terjual sebesar Rp 0,237 demikian sebaliknya dengan anggapan  $X_1$  dan  $X_2$  tetap atau konstan.

## DAFTAR PUSTAKA

- Gujarati, D. (2003) *Ekonometrika Dasar*. Jakarta: Erlangga.
- Hanum, H. (2011) 'Perbandingan Metode Stepwise, Best Subset Regression, dan Fraksi dalam Pemilihan Model Regresi Berganda Terbaik Herlina', *jurnal penelitian Sains*, 14(2), pp. 1-6.
- Johnson, A. F. (2006) *Statistical method [6]*, *Science*. doi: 10.1126/science.152.3721.450-a.
- Muniarty, P. and Wulandari, W. (2020) *Modul Praktikum Statistika 2*. Pasuruan: Qiara.



## **BAB 2**

# **ANALISIS KORELASI**

**Oleh Salsabila**

Seringkali dalam suatu penelitian ditemui ketertarikan untuk mengetahui hubungan antara satu karakteristik objek dalam data dengan karakteristik lainnya. Karakteristik suatu objek dalam data disebut dengan variabel. Pengukuran hubungan antar variabel dapat dilakukan menggunakan koefisien korelasi. Korelasi merupakan ukuran kekuatan hubungan antara dua variabel atau lebih. Sedangkan analisis korelasi merupakan analisis statistika untuk mengetahui kebenaran hubungan antara dua variabel. Dalam bahasan bab ini, dibatasi hanya untuk korelasi antar dua variabel (korelasi bivariat).

Peneliti diharapkan mampu mengidentifikasi variabel yang akan diukur tersebut dicurigai memiliki hubungan. Sebagai contoh, variabel jam belajar dan lama tempuh lari. Secara logika, jam belajar tidak memiliki hubungan dengan lama tempuh lari. Oleh karena itu, analisis korelasi tidak perlu dilakukan. Lain halnya dengan variabel jam belajar dan IPK. Ada kemungkinan keduanya memiliki korelasi dan peneliti tertarik untuk mengukur korelasi antara keduanya.

Lebih lanjut, adanya korelasi antar dua variabel tidak berarti satu variabel mempengaruhi variabel lainnya (ada hubungan sebab-akibat). Dalam contoh sebelumnya, adanya korelasi antara variabel jam belajar dan IPK tidak berarti bahwa jam belajar mempengaruhi besarnya IPK atau sebaliknya.

Hubungan sebab-akibat (pengaruh) dapat diketahui menggunakan analisis regresi.

## 2.1 Koefisien Korelasi (r)

Terdapat beberapa jenis korelasi yang biasa digunakan dalam bidang pengetahuan. Tiga diantaranya merupakan korelasi non-parametrik (Spearman, Kendall Tau, dan Koefisien Gamma), sedangkan korelasi parametrik menggunakan koefisien korelasi *pearson* (Zaid, 2015). Nilai korelasi antara dua variabel kuantitatif dihitung berdasarkan koefisien *Pearson Product Moment*. Perhitungan koefisien korelasi tersebut dihitung menggunakan nilai kovariansi dibagi dengan akar dari variansi tiap variabelnya seperti pada rumus berikut ini (Mendenhall dkk, 2006).

$$r = \frac{S_{xy}}{S_x S_y} = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}}$$

dimana  $-1 \leq r \leq 1$

$$S_{xy} = \sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}$$

$$S_{xx} = \frac{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1}$$

$$S_{yy} = \frac{\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}}{n - 1}$$

Keterangan:

$r$  : Koefisien korelasi

$S_{xy}$  : Kovariansi x dan y

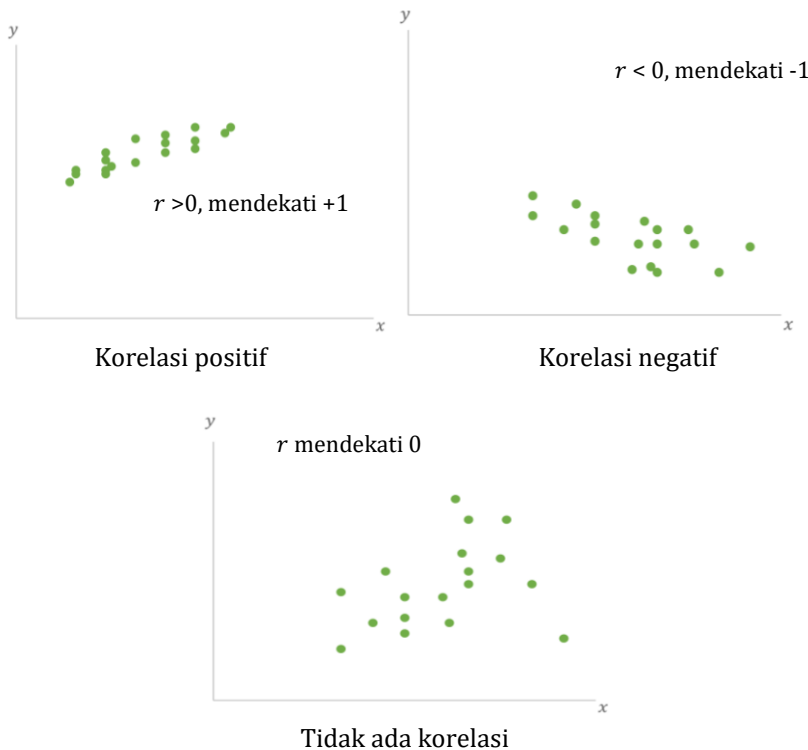
$S_{xx}$  : Variansi x

$S_{yy}$  : Variansi y



$n$  : Banyak sampel

Bentuk korelasi antara dua variabel dapat digambarkan seperti pada beberapa *scatterplot* dibawah ini.



Koefisien korelasi bernilai dari -1 hingga +1. Ketika bernilai di sekitar +1, artinya dua variabel tersebut memiliki hubungan yang sangat kuat. Hal ini dapat dilihat pula dari *scatterplot* antara dua variabel yang membentuk pola positif atau meningkat. Bentuk *scatterplot* dapat digunakan sebagai diagnosis awal terkait ada tidaknya korelasi dari kedua variabel. Analisis korelasi dapat dilakukan jika seandainya hasil diagnosis awal terbentuk pola yang mencirikan ada korelasi positif atau negatif. Dalam analisis korelasi, penentuan variabel X dan Y tidak ada kaitannya dengan variabel sebab-akibat. Variabel X tidak berarti variabel yang mempengaruhi (bebas) dan variabel Y tidak berarti variabel yang dipengaruhi (terikat). Dengan kata

lain, X dan Y hanya sebagai simbol saja untuk setiap nama variabel.

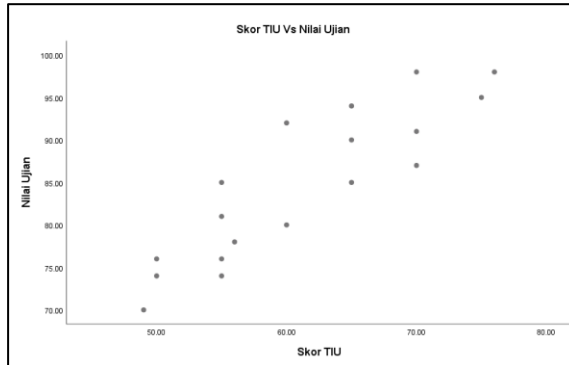
Contoh 1:

Hitunglah nilai korelasi serta jelaskan bentuk korelasi data skor Tes Intelegensia Umum (TIU) dan nilai ujian dari 18 peserta ujian dibawah ini!

| No. | Skor TIU ( $x$ ) | Nilai Ujian ( $y$ ) |
|-----|------------------|---------------------|
| 1.  | 70               | 87                  |
| 2.  | 65               | 94                  |
| 3.  | 70               | 91                  |
| 4.  | 50               | 76                  |
| 5.  | 55               | 74                  |
| 6.  | 65               | 85                  |
| 7.  | 50               | 74                  |
| 8.  | 55               | 76                  |
| 9.  | 65               | 90                  |
| 10. | 55               | 85                  |
| 11. | 70               | 98                  |
| 12. | 55               | 81                  |
| 13. | 56               | 78                  |
| 14. | 60               | 92                  |
| 15. | 49               | 70                  |
| 16. | 75               | 95                  |
| 17. | 60               | 80                  |
| 18. | 76               | 98                  |

Jawab:

Sebelum ke perhitungan, bentuk korelasi dari skor TIU dan nilai ujian dapat dilihat dari *scatterplot* berikut ini.



Penentuan variabel X dan Y didasarkan pada adanya kecurigaan bahwa skor TIU yang mempengaruhi nilai ujian, bukanlah sebaliknya. *Scatterplot* diatas menunjukkan bahwa adanya pola tertentu yaitu *trend* meningkat antara skor TIU dan nilai ujian. Hal ini menyebabkan adanya dugaan bahwa terdapat korelasi positif antara keduanya. Lebih lanjut, koefisien korelasi dapat dihitung sebagai berikut.

Berikut ini merupakan tabel data skor TIU dan nilai ujian beserta beberapa perhitungan yang dibutuhkan dalam koefisien korelasi.

| No. | Skor TIU<br>( $x$ ) | Nilai Ujian<br>( $y$ ) | $x^2$ | $y^2$ | $xy$ |
|-----|---------------------|------------------------|-------|-------|------|
| 1.  | 70                  | 87                     | 4900  | 7569  | 6090 |
| 2.  | 65                  | 94                     | 4225  | 8836  | 6110 |
| 3.  | 70                  | 91                     | 4900  | 8281  | 6370 |
| 4.  | 50                  | 76                     | 2500  | 5776  | 3800 |
| 5.  | 55                  | 74                     | 3025  | 5476  | 4070 |
| 6.  | 65                  | 85                     | 4225  | 7225  | 5525 |
| 7.  | 50                  | 74                     | 2500  | 5476  | 3700 |
| 8.  | 55                  | 76                     | 3025  | 5776  | 4180 |
| 9.  | 65                  | 90                     | 4225  | 8100  | 5850 |
| 10. | 55                  | 85                     | 3025  | 7225  | 4675 |
| 11. | 70                  | 98                     | 4900  | 9604  | 6860 |
| 12. | 55                  | 81                     | 3025  | 6561  | 4455 |
| 13. | 56                  | 78                     | 3136  | 6084  | 4368 |

| No.    | Skor TIU<br>(x) | Nilai Ujian<br>(y) | $x^2$ | $y^2$  | $xy$  |
|--------|-----------------|--------------------|-------|--------|-------|
| 14.    | 60              | 92                 | 3600  | 8464   | 5520  |
| 15.    | 49              | 70                 | 2401  | 4900   | 3430  |
| 16.    | 75              | 95                 | 5625  | 9025   | 7125  |
| 17.    | 60              | 80                 | 3600  | 6400   | 4800  |
| 18.    | 76              | 98                 | 5776  | 9604   | 7448  |
| Jumlah | 1101            | 1524               | 68613 | 130382 | 94376 |

$$S_{xy} = \sum_{i=1}^n x_i y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{n}$$

$$S_{xy} = 94376 - \frac{(1101)(1524)}{18} = 1158$$

$$S_{xx} = \frac{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1}$$

$$S_{xx} = \frac{68613 - \frac{(1101)^2}{18}}{17} = 1268.5$$

$$S_{yy} = \frac{\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}}{n - 1}$$

$$S_{yy} = \frac{130382 - \frac{(1524)^2}{18}}{17} = 1350$$

maka, nilai  $r$  dihitung sebagai berikut:

$$r = \frac{1158}{\sqrt{(1268.5)(1350)}} = 0.8849$$

Nilai  $r = 0.8849$  menunjukkan adanya korelasi yang cukup tinggi mendekati +1 antara skor TIU dan nilai ujian. Hal ini dapat diuji signifikansinya berdasarkan hasil uji hipotesis dalam analisis korelasi.

## 2.2 Uji Hipotesis Analisis Korelasi

Analisis korelasi *pearson* merupakan salah satu analisis statistika parametrik yang membutuhkan terpenuhinya beberapa asumsi, seperti sampel yang diambil secara acak, normalitas dan homogenitas. Penulis mengasumsikan dalam bahasan kali ini kedua uji prasyarat tersebut telah terpenuhi dan uji hipotesis dapat dilakukan. Pengujian hipotesis dilakukan ketika peneliti mengharapkan kesimpulan signifikan terkait angka korelasi antar variabel. Koefisien korelasi tersebut diuji dengan hipotesis sebagai berikut:

1. Uji Dua Arah

$H_0: \rho = 0$  (Tidak ada korelasi antara x dan y)

$H_1: \rho \neq 0$  (Ada korelasi antara x dan y)

2. Uji Satu arah

$H_0: \rho \leq 0$  (Tidak ada korelasi positif antara x dan y)

$H_1: \rho > 0$  (Ada korelasi positif antara x dan y)

atau

$H_0: \rho \geq 0$  (Tidak ada korelasi negatif antara x dan y)

$H_1: \rho < 0$  (Ada korelasi negatif antara x dan y)

Penentuan  $H_1$  dapat dilakukan berdasarkan bentuk *scatterplot* yang diperoleh. Gambaran pola data yang menunjukkan *trend* naik dapat menggunakan hipotesis  $H_1: \rho > 0$ , yaitu akan diuji kecurigaan adanya korelasi positif antara x dan y.

Contoh 2:

Diketahui dari contoh sebelumnya nilai  $r = 0.8849$ . Lakukanlah analisis korelasi beserta uji hipotesisnya dengan tingkat kepercayaan 95%!

Jawab:

1. Hipotesis

$H_0: \rho = 0$  (Tidak ada korelasi signifikan antara skor TIU dan nilai ujian)

$H_1: \rho \neq 0$  (Ada korelasi signifikan antara skor TIU dan nilai ujian)

2. Tingkat signifikansi

$$\alpha = 0.05$$

3. Statistik Uji

$$t_{hitung} = r \sqrt{\frac{n-2}{1-r^2}}$$

$$t_{hitung} = 10.43$$

$$t_{\left(\frac{\alpha}{2}, df=n-2\right)} = t_{(0.025, df=16)} = 2.12$$

4. Daerah Kritis

Tolak  $H_0$  jika:

Uji dua arah:  $t_{hitung} > t_{\left(\frac{\alpha}{2}, df=n-2\right)}$  atau

$$t_{hitung} < -t_{\left(\frac{\alpha}{2}, df=n-2\right)}$$

jika Uji satu arah, maka:

$$t_{hitung} > t_{(\alpha, df=n-2)} \text{ (ketika } H_1: \rho > 0 \text{ ) atau}$$

$$t_{hitung} < -t_{(\alpha, df=n-2)} \text{ (ketika } H_1: \rho < 0 \text{ )}$$

atau jika ***p-value*** <  $\alpha$  (jika menggunakan *p-value* sebagai statistik uji)

5. Keputusan

Karena  $t_{hitung} = 10.43 > t_{(0.025, df=16)}$  maka tolak  $H_0$

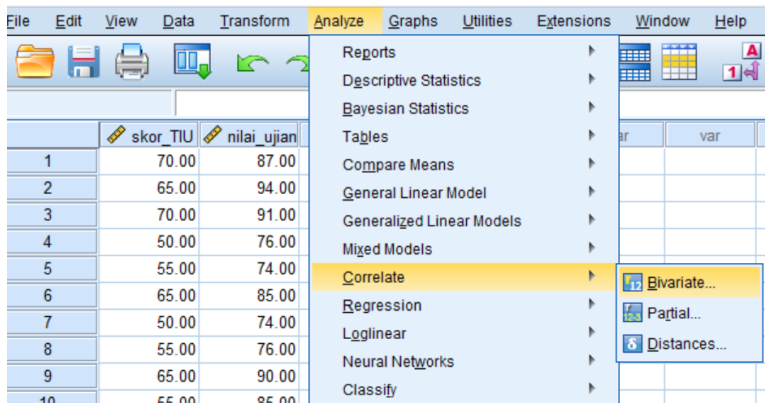
6. Kesimpulan

Berdasarkan taraf kepercayaan 95%, diperoleh keputusan tolak  $H_0$ . Artinya, cukup bukti untuk menyatakan bahwa ada korelasi signifikan antara skor TIU dan nilai ujian.

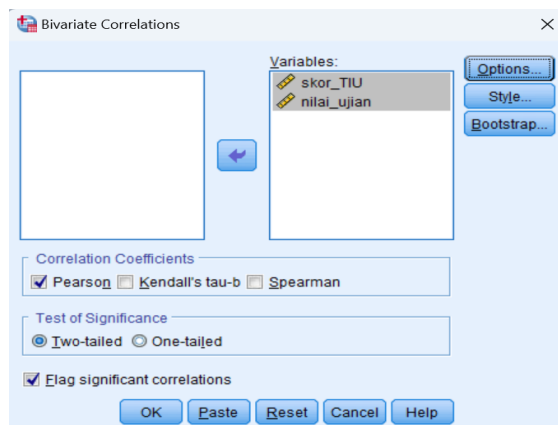
## 2.3 Analisis Korelasi dengan SPSS

SPSS merupakan *software* statistika yang sering digunakan dalam menganalisis data dikarenakan praktis dan mudah digunakan. Berdasarkan data Contoh 1, langkah-langkah analisis korelasi menggunakan SPSS sebagai berikut:

1. Masukkan data, nama variabel, dan jenis variabel: *scale*
2. Pilih *Analyze->Correlate->Bivariate*



3. Masukkan kedua variabel ke kotak *Variables*
4. Pilih *Pearson* dan sesuaikan uji signifikan yang digunakan (two-tailed atau one-tailed), Ok.



Diperoleh *Output* sebagai berikut:

| Correlations |                     |          |             |
|--------------|---------------------|----------|-------------|
|              |                     | skor_TIU | nilai_ujian |
| skor_TIU     | Pearson Correlation | 1        | .885**      |
|              | Sig. (2-tailed)     |          | .000        |
|              | N                   | 18       | 18          |
| nilai_ujian  | Pearson Correlation | .885**   | 1           |
|              | Sig. (2-tailed)     | .000     |             |
|              | N                   | 18       | 18          |

\*\*. Correlation is significant at the 0.01 level (2-tailed).

Hasil tersebut menunjukkan nilai *sig. (2-tailed)*/p-value = 0.000 yang kurang dari suatu nilai  $\alpha = 0.05$ , maka disimpulkan adanya korelasi signifikan antara skor TIU dan nilai ujian. Peneliti dapat menyusun uji hipotesis seperti pada Contoh 2 dengan statistik uji yang digunakan adalah *p-value*.



## DAFTAR PUSTAKA

- Cohen, J., Cohen P., West, S.G., & Aiken, L.S. 2002. Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Psychology Press. ISBN: 0-8058-2223-2.
- Mendenhall, W., Beaver, R.J., dan Beaver, B.M. 2006. *Introduction to Probability and Statistics. 13<sup>th</sup> Edition*. Canada: Cengage Learning, hlm. 55.
- Zaid, M.A. 2015. *Correlation and Regression Analysis*. Turkey: Statistics Department of SESRIC.



# BAB 3

## REGRESI LINIER SEDERHANA

Oleh Lita Lovia

### 3.1 Pendahuluan

Analisis regresi linier merupakan suatu teknik statistika yang digunakan untuk menarik kesimpulan mengenai hubungan linier antara dua atau lebih peubah sehingga salah satu peubah bisa diramalkan dari peubah-peubah lainnya. Analisis regresi linier yang melibatkan satu peubah tak bebas dan satu peubah bebas disebut dengan analisis regresi linier sederhana.

### 3.2 Model Regresi Linier Sederhana

Model regresi linier sederhana:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i = 1, 2, 3, \dots, n$$

dengan

$Y_i$  : Nilai variabel tak bebas dalam percobaan ke- $i$

$\beta_0, \beta_1$  : Paramater

$X_i$  : Konstanta yang diketahui nilainya, yakni nilai variabel bebas dalam percobaan ke- $i$

$\varepsilon_i$  : Galat acak

$n$  : banyak pengamatan

Setiap pengamatan  $(x_i, y_i)$  dalam sampel memenuhi hubungan:

$$y_i = b_0 + b_1 x_i + e_i$$

$$e_i = y_i - \hat{y}_i$$

$e_i$  disebut sisaan dengan  $\hat{y}_i = b_0 + b_1 x_i$

### 3.3 Metode Kuadrat Terkecil

Metode kuadrat terkecil adalah metode untuk mendapatkan garis regresi dianggap paling baik.

$$e_i = y_i - \hat{y}_i$$

$$= y_i - (b_0 + b_1 x_i)$$

$$e_i^2 = (y_i - b_0 - b_1 x_i)^2$$

Jumlah Kuadrat Sisaan (JKS)

$$JKS = \sum_i^n e_i^2$$

Pendugaan parameter model regresi dilakukan dengan meminimumkan  $JKS$ :

$$\blacksquare \frac{\partial JKS}{\partial b_0} = 0$$

$$\frac{\partial \sum_i^n (y_i - b_0 - b_1 x_i)^2}{\partial b_0} = 0$$

Sehingga diperoleh:

$$nb_0 + b_1 \sum_i^n x_i = \sum_i^n y_i$$

$$\blacksquare \frac{\partial JKS}{\partial b_1} = 0$$

$$\frac{\partial \sum_i^n (y_i - b_0 - b_1 x_i)^2}{\partial b_1} = 0$$

Sehingga diperoleh:

$$b_0 \sum_i^n x_i + b_1 \sum_i^n x_i^2 = \sum_i^n x_i y_i$$

Persamaan normal kuadrat terkecil:

$$\begin{aligned} nb_0 + b_1 \sum_i^n x_i &= \sum_i^n y_i JKS \\ &= \sum_i^n (y_i - b_0 - b_1 x_i)^2 \\ b_0 \sum_i^n x_i + b_1 \sum_i^n x_i^2 &= \sum_i^n x_i y_i \end{aligned}$$

Dengan menyelesaikan persamaan normal ini diperoleh:

$$\begin{aligned} b_1 &= \frac{\sum_i^n x_i y_i - \frac{\sum_i^n x_i \sum_i^n y_i}{n}}{\sum_i^n x_i^2 - \frac{(\sum_i^n x_i)^2}{n}} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2} \\ b_0 &= \frac{\sum_i^n y_i - b_1 \sum_i^n x_i}{n} = \bar{y} - b_1 \bar{x} \end{aligned}$$

Penulisan rumus dapat disederhanakan sebagai berikut:

$$JK_{xy} = \sum_i^n (x_i - \bar{x})(y_i - \bar{y})$$

$$JK_x = \sum_i^n (x_i - \bar{x})^2$$

$$JK_y = \sum_i^n (y_i - \bar{y})^2$$

Sehingga rumus  $b_1$  dapat ditulis:

$$b_1 = \frac{JK_{xy}}{JK_x}$$

### 3.4 Persamaan Regresi Melewati Titik Asal

Dalam beberapa kasus ditemui bahwa jika peubah ( $X$ ) bernilai nol maka peubah tak bebas ( $Y$ ) juga bernilai nol.

Contoh:

- $X$  menyatakan banyak baju yang terjual  
 $Y$  menyatakan keuntungan
- $X$  menyatakan jarak tempuh  
 $Y$  menyatakan yang kecepatan

Pada kondisi demikian maka model regresi linier sederhana yang digunakan adalah model tanpa intercept sebagai berikut:

$$Y_i = \beta_1 X_i + \varepsilon_i, i = 1, 2, 3, \dots, n$$

Dengan

$$y_i = b_1 x_i + e_i$$

$$e_i = y_i - \hat{y}_i$$

$$= y_i - b_1 x_i$$

$$e_i^2 = (y_i - b_1 x_i)^2$$

Jumlah Kuadrat Sisaan (JKS)

$$JKS = \sum_i^n e_i^2$$

$$JKS = \sum_i^n (y_i - b_1 x_i)^2$$

Pendugaan parameter model regresi dilakukan dengan meminimumkan  $JKS$

$$\blacksquare \frac{\partial JKS}{\partial b_1} = 0$$

$$\frac{\partial \sum_i^n (y_i - b_1 x_i)^2}{\partial b_1} = 0$$

Sehingga diperoleh

$$b_1 \sum_i^n x_i^2 = \sum_i^n x_i y_i$$

Atau dapat ditulis

$$b_1 = \frac{\sum_i^n x_i y_i}{\sum_i^n x_i^2}$$

### 3.5 Koefisien Determinasi ( $R^2$ )

Setelah model regresi diperoleh maka pertanyaan yang muncul adalah apakah model tersebut sudah menghasilkan ketepatan dugaan

Perhatikan bahwa

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Jumlah kuadrat Total (JKT)

$$\sum_{i=1}^n (y_i - \bar{y})^2$$



Jumlah Kuadrat Regresi (JKR)

$$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

Jumlah Kuadrat Sisaan

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$R^2 = \frac{JKR}{JKT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$R^2$  disebut dengan koefisien determinasi, makin dekat  $R^2$  dengan 1 makin baik kecocokan data dengan model, dan sebaliknya makin dekat  $R^2$  dengan 0 makin jelek kecocokan tersebut.  $R^2$  biasanya dinyatakan dalam persen.

### 3.6 Asumsi Dalam Analisis Regresi Linier Sederhana

1.  $\varepsilon_i$  adalah peubah acak dengan nilai tengah nol dan ragam homogen
2.  $\varepsilon_i$  dan  $\varepsilon_j$ ,  $i \neq j$  tidak saling berkorelasi
3.  $\varepsilon_i$  berasal dari sebaran normal
4.  $X$  bukan peubah acak

### 3.7 Pembentukan Selang Kepercayaan Bagi $\beta_0$ dan $\beta_1$

Selang kepercayaan  $(1 - \alpha)100\%$  untuk  $\beta_0$  dirumuskan sebagai,

$$b_0 \pm t_{(\alpha/2, n-2)} SE(b_0)$$

$$SE(b_0) = s \sqrt{\frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}}$$

Sedangkan Selang kepercayaan  $(1 - \alpha)100\%$  untuk  $\beta_1$  dirumuskan sebagai

$$b_1 \pm t_{(\alpha/2, n-2)} SE(b_1)$$

$$SE(b_0) = s \sqrt{\frac{n}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}}$$

Dan  $t_{(\alpha/2, n-2)}$  diperoleh dari tabel distribusi  $t$

### 3.8 Uji keberartian Garis Regresi

Prosedur analisis ragam dengan model regresi terlebih dahulu membuat tabel anova sebagai berikut.

| Sumber  | db      | Jumlah Kuadrat (JK) | Kuadrat Tengah (KT)       | F hitung  |
|---------|---------|---------------------|---------------------------|-----------|
| Regresi | 1       | JKR                 | $KTR = \frac{JKR}{1}$     | $KTR/KTS$ |
| Sisaaan | $n - 2$ | JKS                 | $KTS = \frac{JKS}{n - 2}$ |           |
| Total   | $n - 1$ | JKT                 |                           |           |

Prosedur Pengujian:

1. Tentukan hipotesis uji

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0$$

2. Tentukan taraf nyata pengujian ( $\alpha$ )

3. Hitung nilai statistik uji  $F_{hit} = \frac{KTR}{KTS}$

4. Tentukan nilai kritis ,  $F_{(\alpha,1,n-2)}$

5. Kesimpulan

$F_{hit} > F_{(\alpha,1,n-2)} : \text{tolak } H_0$

$F_{hit} < F_{(\alpha,1,n-2)} : \text{tidak tolak } H_0$

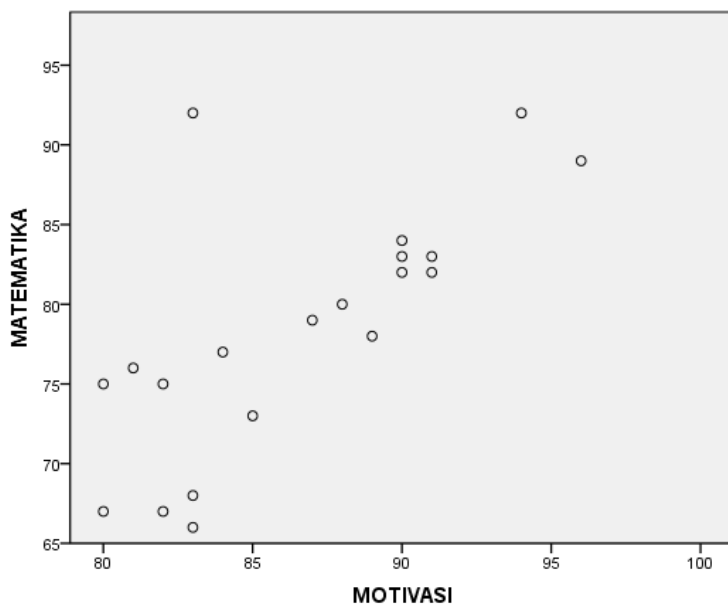
Contoh:

Berikut ini diberikan data Simulasi Motivasi dan Nilai matematika:

| NO | MOTIVASI | NILAI MATEMATIKA |
|----|----------|------------------|
| 1  | 83       | 66               |
| 2  | 87       | 79               |
| 3  | 85       | 73               |
| 4  | 91       | 82               |
| 5  | 91       | 83               |
| 6  | 81       | 76               |
| 7  | 83       | 68               |
| 8  | 80       | 67               |
| 9  | 88       | 80               |
| 10 | 90       | 84               |
| 11 | 84       | 77               |

|    |    |    |
|----|----|----|
| 12 | 82 | 67 |
| 13 | 89 | 78 |
| 14 | 90 | 83 |
| 15 | 96 | 89 |
| 16 | 94 | 92 |
| 17 | 83 | 92 |
| 18 | 82 | 75 |
| 19 | 80 | 75 |
| 20 | 90 | 82 |

1. Buatlah plot antara Motivasi dan Nilai Matematika



2. Tentukan persamaan regresi yang menggambarkan hubungan antara kedua peubah. Perhatikan bahwa

$$\hat{y}_i = b_0 + b_1 x_i$$

dan

$$b_1 = \frac{\sum_i^n x_i y_i - \frac{\sum_i^n x_i \sum_i^n y_i}{n}}{\sum_i^n x_i^2 - \frac{(\sum_i^n x_i)^2}{n}}$$

$$b_0 = \frac{\sum_i^n y_i - b_1 \sum_i^n x_i}{n}$$

| NO | X  | Y  | X <sup>2</sup> | Y <sup>2</sup> | XY   | $\hat{y}$ | $(\hat{y}_i - \bar{y})^2$ | $(y_i - \bar{y})^2$ |
|----|----|----|----------------|----------------|------|-----------|---------------------------|---------------------|
| 1  | 83 | 66 | 6889           | 4356           | 5478 | 74,376    | 16,193                    | 153,76              |
| 2  | 87 | 79 | 7569           | 6241           | 6873 | 79,084    | 0,468                     | 0,36                |
| 3  | 85 | 73 | 7225           | 5329           | 6205 | 76,73     | 2,789                     | 29,16               |
| 4  | 91 | 82 | 8281           | 6724           | 7462 | 83,792    | 29,074                    | 12,96               |
| 5  | 91 | 83 | 8281           | 6889           | 7553 | 83,792    | 29,074                    | 21,16               |
| 6  | 81 | 76 | 6561           | 5776           | 6156 | 72,022    | 40,679                    | 5,76                |
| 7  | 83 | 68 | 6889           | 4624           | 5644 | 74,376    | 16,193                    | 108,16              |
| 8  | 80 | 67 | 6400           | 4489           | 5360 | 70,845    | 57,078                    | 129,96              |
| 9  | 88 | 80 | 7744           | 6400           | 7040 | 80,261    | 3,463                     | 2,56                |
| 10 | 90 | 84 | 8100           | 7056           | 7560 | 82,615    | 17,766                    | 31,36               |
| 11 | 84 | 77 | 7056           | 5929           | 6468 | 75,553    | 8,105                     | 1,96                |
| 12 | 82 | 67 | 6724           | 4489           | 5494 | 73,199    | 27,050                    | 129,96              |
| 13 | 89 | 78 | 7921           | 6084           | 6942 | 81,438    | 9,229                     | 0,16                |
| 14 | 90 | 83 | 8100           | 6889           | 7470 | 82,615    | 17,766                    | 21,16               |
| 15 | 96 | 89 | 9216           | 7921           | 8544 | 89,677    | 127,171                   | 112,36              |

|    |    |    |      |      |      |        |        |        |
|----|----|----|------|------|------|--------|--------|--------|
| 16 | 94 | 92 | 8836 | 8464 | 8648 | 87,323 | 79,620 | 184,96 |
| 17 | 83 | 92 | 6889 | 8464 | 7636 | 74,376 | 16,193 | 184,96 |
| 18 | 82 | 75 | 6724 | 5625 | 6150 | 73,199 | 27,050 | 11,56  |
| 19 | 80 | 75 | 6400 | 5625 | 6000 | 70,845 | 57,078 | 11,56  |
| 20 | 90 | 82 | 8100 | 6724 | 7380 | 82,615 | 17,766 | 12,96  |

Selanjutnya diperoleh

$$\sum_1^{20} x_i y_i = 136063, \quad \sum_1^{20} x_i = 1729, \quad \sum_1^{20} y_i = 1568, \\ \sum_1^{20} x_i^2 = 149905, \sum_1^{20} y_i^2 = 124098, \bar{x} = 86,45$$

$$b_1 = \frac{136063 - \frac{1729 \times 1568}{20}}{149905 - \frac{(1729)^2}{20}} = 1,176579$$

$$b_0 = \frac{1568 - 1,177 \times 1729}{20} = -23,3153$$

diperoleh

$$\hat{y} = -23,3153 + 1,176579x$$

### 3. Hitung Koefisien determinasi $R^2$

$$R^2 = \frac{JKR}{JKT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \\ R^2 = \frac{599,81}{1166,8} = 0,514$$

Nilai koefisien determinasi sebesar 0,514 dapat diinterpretasikan bahwa keragaman variabel bebas sebesar 51,4% mampu menjelaskan keragaman variabel terikat, sisanya sebesar 48,6% dijelaskan oleh variabel lain yang tidak masuk dalam persamaan.

### 4. Susunlah Tabel Analisis Ragam.

$$JKR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = 599,371$$

$$JKS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = 567,451$$

$$JKT = \sum_{i=1}^n (y_i - \bar{y})^2 = 1166,822$$

$$KTR = \frac{JKR}{1} = 599,805$$

$$KTS = \frac{JKS}{n - 2} = 267,73$$

| Sumber  | db     | Jumlah Kuadrat (JK) | Kuadrat Tengah (KT) | F hitung |
|---------|--------|---------------------|---------------------|----------|
| Regresi | 1      | 599,371             | 599,805             | 0,262    |
| Sisaaan | 20 – 2 | 567,451             | 287,73              |          |
| Total   | 20 – 1 | 1166,822            |                     |          |

## DAFTAR PUSTAKA



SEMBIRING, R. (1995). *Analisis Regresi*. Penerbit ITB Bandung.  
Suyono. (2015). *Analisis Regresi Untuk Penelitian*. deepublish.

# BAB 4

## REGRESI LINEAR BERGANDA

Oleh Muhammad Ilham Riyadh

### 4.1 Pendahuluan

Pada Bab 4 ini, keterkaitan antara variable  $X$  dan  $Y$  dijelaskan dengan menggunakan Regresi Linier Sederhana yang digunakan untuk memprediksi nilai variable terikat  $Y$  apabila diketahui nilai variable bebas  $X$  ( $x = x_0$ ). Suatu persamaan regresi disebut linier berganda apabila memuat lebih dari dua variable, termasuk variable terikat  $Y$  (prediktor). Metode ini merupakan kemajuan di atas analisis regresi dasar. Dampak setiap *variable independent* terhadap *variable dependent* dalam regresi linier ganda akan dievaluasi sambil memperhitungkan pengaruh *variable independent* lainnya dalam model. Memprediksi nilai variable terikat menggunakan nilai variable bebas yang diketahui adalah tujuan utamanya. Distribusi normal merupakan salah satu syarat terjadinya regresi linier ganda. Kurangnya multikolinearitas, atau korelasi yang signifikan antara variable independen, dan distribusi kesalahan atau residu yang normal merupakan prasyarat untuk regresi linier ganda.

## 4.2 Model Regresi Berganda

Regresi Berganda dapat dilakukan apabila data bersifat linier

Model Estimasi Linier dapat ditunjukkan sebagai berikut :

$$Y = a + b_1X_1 + b_2X_2 + ..... + b_nX_n + e$$

Apabila menggunakan data non linier harus di linearakan yaitu dengan menggunakan Logaritma natural. Model Estimasi Non Linier yang di linearakan sebagai berikut :

$$LnY = a + b_1LnX_1 + b_2LnX_2 + ..... + b_nLnX_n + e$$

Dimana :

$Y$  = *Variable Dependent* atau variable tidak bebas

$a$  = *Intercept/ Konstanta*

$b_1..b_n$  = Koefisien Regresi

$X_1..X_n$  = *Variable Independent* atau variable bebas

$e$  = *Error Term*

$Ln$  = Logaritma Natural

$$\text{Rumus : } a = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

Dimana  $\bar{Y}$  =  $Y$  rata rata

$\bar{X}$  =  $X$  rata rata

$$b_1 = \frac{(\sum x_1y)(\sum x_2^2) - (\sum x_2y)(\sum x_1x_2)}{\sum x_1^2 \sum x_2^2 - (\sum x_1x_2)^2}$$

$$b_2 = \frac{(\sum x_2y)(\sum x_1^2) - (\sum x_1y)(\sum x_1x_2)}{\sum x_1^2 \sum x_2^2 - (\sum x_1x_2)^2}$$

4.2.2 Contoh Soal

Permasalahan : Apakah Pupuk dan Tenaga Kerja Berpengaruh Terhadap Produksi Cabe ?

Hipotesis : Pupuk dan Tenaga kerja berpengaruh terhadap produksi cabe

Data yang di ambil dilapangan

Dimana Y = Produksi Cabai (00Kg) ,  $X_1$  = Pupuk (0 Kg) dan  $X_2$  = Tenaga Kerja (orang), n = Jumlah Sampel,

Tabel 4. 1. Pengaruh Pupuk dan Tenaga Kerja Terhadap Produksi Cabai

| n                                              | Y       | $X_1$   | $X_2$   | y  | $y^2$       | $x_1$ | $x_1^2$    | $x_2$ | $x_2^2$   | $x_1y$     | $x_2y$     | $x_1x_2$ |
|------------------------------------------------|---------|---------|---------|----|-------------|-------|------------|-------|-----------|------------|------------|----------|
| 1                                              | 3       | 2       | 1       | -4 | 16          | -3    | 9          | -1    | 1         | 12         | 4          | 3        |
| 2                                              | 6       | 4       | 2       | 1  | 1           | -1    | 1          | 0     | 0         | 1          | 0          | 0        |
| 3                                              | 9       | 7       | 3       | 2  | 4           | 2     | 4          | 1     | 1         | 4          | 2          | 2        |
| 4                                              | 4       | 3       | 1       | -3 | 9           | -2    | 4          | -1    | 1         | 6          | 3          | 2        |
| 5                                              | 8       | 5       | 2       | 1  | 1           | 0     | 0          | 0     | 0         | 0          | 0          | 0        |
| 6                                              | 12      | 9       | 3       | 5  | 25          | 4     | 16         | 1     | 1         | 20         | 5          | 4        |
| N=6<br>Total<br>$\sum$<br>Rata<br>$\bar{\sum}$ | 42<br>7 | 30<br>5 | 12<br>2 | 0  | 56<br>35,17 | 0     | 34<br>5,67 | 0     | 4<br>0,67 | 43<br>7,16 | 14<br>9,83 | 11       |

Sumber: Data primer di olah

Angka ini dibuat untuk mempermudah menghitungnya

4.2.3 Jawaban

Hasil penjumlahan dan rata rata pada tabel diatas sebagai berikut :

$(\sum x_1y) = 43, (\sum x_2^2) = 4, (\sum x_2y) = 14, (\sum x_1 x_2) = 11 \sum x_1^2 = 34, y^2= 56$

Untuk mencari y dan x digunakan rumus berikut :  $y = Y - \bar{Y}$  dan  $x = X - \bar{X}$

Contoh :  $y = 3 - 7$  dan  $x = 2 - 5$  (lihat pada tabel )

Selanjutnya Untuk mencari nilai a = constanta terlebih dahulu menghitung rumus  $b_1$  dahulu

Yaitu sebagai berikut :

$$b_1 = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

menjadi

$$b_1 = \frac{(43 * 4) - (14 * 11)}{(34 * 4) - (11)^2} = \frac{(172) - (154)}{(136) - (121)} = \frac{18}{15} = 1,20$$

Selanjutnya menghitung rumus  $b_2$  sebagai berikut :

$(\sum x_1 y) = 43$ ,  $(\sum x_2^2) = 4$ ,  $(\sum x_2 y) = 14$ ,  $(\sum x_1 x_2) = 11$  (lihat Tabel diatas)

$$b_2 = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

Menjadi

$$b_2 = \frac{(14 * 34) - (43 * 11)}{(34 * 4) - (11)^2} = \frac{(476) - (473)}{(136) - (121)} = \frac{3}{15} = 0,20$$

Kemudian selanjutnya menghitung nilai  $a$ = konstanta sebagai berikut:

$$\begin{aligned} a &= \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 \\ &= 7 - (1,20) \times 5 - 0,20 \times 2 \\ &= 7 - 6 - 0,4 \\ &= 1 - 0,40 \\ &= 0,60 \end{aligned}$$

### ➤ Langkah pertama

Tunjukkan Model Estimasi ?

Berdasarkan model regresi linear berganda dapat di tunjukan sebagai berikut

$$Y = a + b_1X_1 + b_2X_2 + e$$

Dimana

$Y$  = Produksi Cabai (*Variable Dependent* )

$a$  = *Intercept*/ Konstanta

$b_1..b_n$  = Koefisien Regresi

$X_1$  = Pupuk (Kg) *Variable Independent*

$X_2$  = Tenaga Kerja (Orang) *Variable Independent*

$e$  = *Error Term*

### ➤ Langkah Kedua

Tunjukkan Hasil Estimasi ?

Setelah didapat nilai  $a$  konstanta kemudian nilai  $b_1$  dan  $b_2$  dapat di tunjukan hasil estimasi sebagai berikut :

$$Y = 0,60 + 1,20X_1 + 0,20X_2 + e$$

### ➤ Langkah Ketiga

Interprestasikan Hasil Estimasi

Inteprestasi nilai  $a$  = Konstanta

Dalam persamaan diatas, 0,60 adalah konstanta ( $a$ ) ini berarti bahwa ketika pupuk ( $X_1$ ) dan Tenaga Kerja ( $X_2$ ) bernilai nol, produksi ( $Y$ ) cabai yang diharapkan adalah 0,60 dengan kata lain 0,60 adalah produksi yang diharapkan mempertimbangkan pengaruh pupuk dan tenaga kerja.

- a. Inteprestasi Pengaruh Pupuk ( $X_1$  ) Terhadap produksi Cabai ( $Y$ )

Berdasarkan dari persamaan diatas menunjukkan bahwa variable Pupuk ( $X_1$ ) Berpengaruh positif terhadap Produksi Cabai (Y) sebesar 1,2 kg artinya apabila pupuk di tambahkan sebesar 1 kg maka produksi cabai meningkat sebesar 1,2 kg dengan kata lain bertambahnya pupuk dapat meningkatkan produksi cabai dengan asumsi *ceteris paribus* atau faktor faktor lain di anggap konstan.

b. Inteprestasi Pengaruh Tenaga Kerja (  $X_2$  ) Terhadap produksi Cabai (Y)

Berdasarkan dari persamaan diatas menunjukkan bahwa variable Tenaga Kerja ( $X_2$ ) Berpengaruh positif terhadap Produksi Cabai (Y) sebesar 0,2 orang artinya apabila tenaga kerja di tambahkan sebesar 1 orang maka produksi cabai meningkat sebesar 0,2 Kg dengan kata lain bertambahnya jumlah tenaga kerja dapat meningkatkan jumlah produksi cabai ketika panen dengan asumsi *ceteris paribus* atau faktor faktor lain di anggap konstan

## **Pengujian Statistik (Uji T dan Uji F)**

### **Uji T**

#### **Pengujian Uji T pada $\alpha = 5\%$**

Langkah : Hipotesis

$$H_0: b_1 = 0$$

$$H_a: b_1 \neq 0$$

Kriteria :

$H_0$  diterima apabila  $T_{hitung} < T_{tabel}$  atau  $- T_{hitung} > - T_{tabel}$  ( $\alpha = 5\%$ )  
atau  $Sig > 0,05$

$H_a$  diterima apabila  $T_{hitung} > T_{tabel}$  atau  $- T_{hitung} < - T_{tabel}$  ( $\alpha = 5\%$ )  
atau  $Sig < 0,05$

Untuk mengerjakan Thitung b1 dimulai dari langkah pertama dimana langkah pertama merupakan turunan akhir dari rumus Thitung

Dimana  $(\sum x_1 y) = 43$ ,  $(\sum x_2^2) = 4$ ,  $(\sum x_2 y) = 14$ ,  $(\sum x_1 x_2) = 11$   $\sum x_1^2 = 34$ ,  $y^2 = 56$  (lihat Tabel 1 diatas)

Untuk b1

$$\text{Langkah terakhir : } Thitung = \frac{\hat{b}_1}{Se(\hat{b}_1)} = \mathbf{3,167}$$

$$Thitung = \frac{1,20}{0,3788} = \mathbf{3,167} \text{ ( jika menggunakan software SPSS, hasil nilai nya tidak berbeda jauh)}$$

$$\text{Langkah Kelima : } Se(\hat{b}_1) = \sqrt{Variance(b_1)} =$$

$$Se(\hat{b}_1) = \sqrt{0,1435} = \mathbf{0,3788}$$

$$\text{Langkah keempat : } Variance(b_1) = \delta_u^2 \frac{\sum x_2^2}{\sum x_1^2 x_2^2 - (\sum x_1 x_2)^2} =$$

$$Variance(b_1) = 0,5376 \frac{4}{\sum 34*4 - (\sum 11)^2} =$$

$$= 0,5376 \frac{4}{136 - 121} =$$

$$= 0,5376 \frac{4}{15} = 0,5376 * 0,267$$

$$= \mathbf{0,1435}$$

$$\text{Langkah ketiga: } \delta_u^2 = \frac{\sum e^2}{n - k} =$$

Dimana k = jumlah interncept + variable atau  $k = a, X_1 \text{ dan } X_2$  yaitu  $k = 3$



$$\delta_u^2 = \frac{1,6128}{6-3} = \mathbf{0,5376}$$

Langkah kedua :  $\sum e^2 = (1 - R^2) \sum y^2 =$

$$\sum e^2 = (1 - 0,9712)56 = 0,0288 * 56 = \mathbf{1,6128}$$

Langkah pertama :  $R^2 = \frac{b_1 \sum x_1 y + b_2 \sum x_2 y}{\sum y^2} =$

$$R^2 = \frac{1,20*43+0,20*14}{56} = \frac{51,6+2,8}{56} = \frac{54,4}{56} = \mathbf{0,9714}$$

Untuk b2,

Dimana  $(\sum x_1 y) = 43$ ,  $(\sum x_2^2) = 4$ ,  $(\sum x_2 y) = 14$ ,  $(\sum x_1 x_2) = 11$   $\sum x_1^2 = 34$ ,  $y^2 = 56$  (lihat Tabel 1 diatas)

Cara pengerjaan dimulai dengan langkah pertama :

Langkah terakhir :  $Thitung = \frac{\hat{b}_2}{Se(\hat{b}_2)} =$

$$Thitung = \frac{0,20}{1,1000} = \mathbf{0,1818}$$

(jika menggunakan software SPSS, hasil nilai nya tidak berbeda jauh)

Langkah kelima :  $Se(\hat{b}_2) = \sqrt{Variance(b_2)} =$

$$Se(\hat{b}_1) = \sqrt{1,2101} = \mathbf{1,1000}$$

Langkah keempat:  $Variance(b_2) = \delta_u^2 \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} =$

$$\begin{aligned}
 \text{Variance}(b_1) &= 0,5338 \frac{34}{\sum 34*4 - (\sum 11)^2} = \\
 &= 0,5338 \frac{34}{136 - 121} = \\
 &= 0,5338 \frac{34}{15} = 0,5338 * 2,267 \\
 &= \mathbf{1,2101}
 \end{aligned}$$

Langkah ketiga  $\delta_u^2 = \frac{\sum e^2}{n - k} =$

Dimana k = jumlah *intercept* + variable atau  $k = a, X_1$  dan  $X_2$  yaitu k = 3

$$\delta_u^2 = \frac{1,6016}{6-3} = \mathbf{0,5338}$$

Langkah Kedua :  $\sum e^2 = (1 - R^2) \sum y^2 =$

$$\sum e^2 = (1 - 0,9714)56 = 0,0286 * 56 = \mathbf{1,6016}$$

Langkah Pertama :  $R^2 = \frac{b_1 \sum x_1 y + b_2 \sum x_2 y}{\sum y^2} =$

$$R^2 = \frac{1,20*43 + 0,20*14}{56} = \frac{51,6 + 2,8}{56} = \frac{54,4}{56} = \mathbf{0,971}$$

### Hasil T hitung

Thitung untuk X1 (pupuk) adalah **3,167**

Thitung untuk X2 (Tenaga Kerja) adalah **0,1818**

T Tabel ( $\alpha$ )

$$\delta f = n - k = 6 - 3 = 3$$

$$\alpha = 5\%$$

$$= 0,05/2 = 0,025$$

**Tabel 4. 2. Ilustrasi Mencari Nilai T Tabel**

| $\alpha/n$ |  |  | 0,5 | 0,025 | 0,05 | dst |
|------------|--|--|-----|-------|------|-----|
| 1          |  |  |     | ↓     |      |     |
| 2          |  |  |     |       |      |     |
| 3          |  |  |     | 3.182 |      |     |
| 4          |  |  |     |       |      |     |
| 5          |  |  |     |       |      |     |
| 6          |  |  |     |       |      |     |

Lihat Ttabel di lampiran

### Interprestasi Uji T

T hitung untuk X1 (pupuk) adalah **3,167**

Berdasarkan hasil perhitungan uji T menunjukkan bahwa Thitung untuk variable Pupuk (X1) sebesar 3,167 sedangkan Ttabel adalah 3,182 dimana  $T_{hitung} < T_{tabel}$  ( $3,167 < 3,182$ ) artinya *variable* pupuk (X1) tidak berpengaruh signifikan terhadap produksi cabai pada Tingkat kepercayaan 95 persen. T hitung untuk X2 (Tenaga Kerja) adalah **0,1818**.

Berdasarkan hasil perhitungan uji T menunjukkan bahwa Thitung untuk *variable* Tenaga Kerja (X2) sebesar 0,1818 sedangkan Ttabel adalah 3,182 dimana  $T_{hitung} < T_{tabel}$  ( $0,1818 < 3,182$ ) artinya *variable* Tenaga kerja (X2) tidak berpengaruh

signifikan terhadap produksi cabai pada Tingkat kepercayaan 95 persen . Selanjutnya adalah menghitung Fhitungnya dimana:

$$F_{hitung} = \frac{R^2 / k - 1}{(1 - R^2) / n - k} =$$

dimana

k = Jumlah *Independent variable* + *Intercept* dari suatu persamaan

R<sup>2</sup>= Koefisien Determinasi

n = Jumlah Sampel

cara menghitungnya

$$F_{hitung} = \frac{R^2 / k - 1}{(1 - R^2) / n - k}$$

$$F_{hitung} = \frac{0,971/3 - 1}{(1 - 0,971)/6 - 3}$$

$$F_{hitung} = \frac{0,971/2}{(0,029)/3}$$

$$F_{hitung} = \frac{0,4855}{0,0096} = \mathbf{50,5729} \text{ (jika menggunakan software SPSS, hasil nilai nya tidak berbeda jauh)}$$

Catatan : hasil Fhitung selalu positif

Kriteria

H<sub>0</sub> diterima apabila F<sub>hitung</sub> < F<sub>tabel</sub> atau Sig > 0,05

Ha diterima apabila F<sub>hitung</sub> > F<sub>tabel</sub> atau Sig < 0,05

Selanjutnya untuk mencari F hitung sebagai berikut :

F Tabel

$$V1 = k - 1 = 3-1 = 2$$

$$V2 = n - k = 6-3 = 3$$

**Tabel 4. 3. Ilustrasi penggunaan Ftabel**

| V1/V2 | 1 | 2 | 3      | 4 | 5 | dst |
|-------|---|---|--------|---|---|-----|
| 1     |   |   | ↓      |   |   |     |
| 2     |   |   | → 19,2 |   |   |     |
| 3     |   |   |        |   |   |     |
| 4     |   |   |        |   |   |     |
| 5     |   |   |        |   |   |     |
| 6     |   |   |        |   |   |     |

Untuk melihat Ftabel dapat dilihat pada lampiran, maka hasil nilai Ftabelnya adalah 34, 1.

### **Inteprestasikan uji F**

Berdasarkan perhitungan dengan menggunakan uji F menunjukkan bahwa nilainya sebesar 50, 5729 artinya Fhitung > Ftabel atau ( 50, 5729 > 19,2) artinya bahwa Variable Pupuk (X1) dan Tenaga Kerja (X2) secara serempak berpengaruh Signifikan terhadap Produksi Cabai (Variable Y) pada tingkat kepercayaan 95 Persen.

### **Jelaskan koefisien Determinasi (R<sup>2</sup>)**

Nilai Koefisien Determinasi dapat di lihat di perhitungan di rumus R<sup>2</sup> langkah pertama diatas yaitu sebesar 0,971 dengan interprestasinya sebagai berikut :

Berdasarkan perhitungan koefisien determinasi diatas menunjukan nilainya sebesar 0,971 artinya variable pupuk (X1) dan tenaga kerja (X2) mampu memberikan penjelasan terhadap produksi Cabai (Variable Y) sebesar 97,10 persen sedangkan sisanya sebesar 2,90 tidak dimasukan dalam model estimasi

4.2.3 Hasil Olah Data Program Software SPSS

Hasil olah data program SPSS tidak berbeda jauh dari hasil perhitungan regresi bergandara secara manual , hal ini dapat dilihat pada tabel 4.4

Tabel 4. 4. Hasil Olah Data Menggunakan Penggunaan *Software SPSS*

| Descriptive Statistics |        |                |   |
|------------------------|--------|----------------|---|
|                        | Mean   | Std. Deviation | N |
| Produksi               | 7.0000 | 3.34664        | 6 |
| Pupuk                  | 5.0000 | 2.60768        | 6 |
| Tenaga kerja           | 2.0000 | .89443         | 6 |

Pada Tabel 4.4 diatas dapat dilihat rata rata (*mean*) Produksi adalah 7, Pupuk adalah 5 dan Tenaga Kerja adalah 5 dan jumlah sampel (N) adalah 6.

Tabel 4. 5. Hasil Model *Summery Penggunaan Software SPSS*

Model Summary<sup>b</sup>

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate | Durbin-Watson |
|-------|-------------------|----------|-------------------|----------------------------|---------------|
| 1     | .986 <sup>a</sup> | .971     | .952              | .73030                     | 2.375         |

a. Predictors: (Constant), Tenaga kerja, Pupuk

b. Dependent Variable: Produksi

Pada tabel 4. 5 di atas nilai Rsquare atau Koefisien Determinasi adalah sebesar 0,971

**Tabel 4. 6. Hasil *Anova* Penggunaan *Software* SPSS**

**ANOVA<sup>b</sup>**

| Model |            | Sum of Squares | df | Mean Square | F      | Sig.              |
|-------|------------|----------------|----|-------------|--------|-------------------|
| 1     | Regression | 54.400         | 2  | 27.200      | 51.000 | .005 <sup>a</sup> |
|       | Residual   | 1.600          | 3  | .533        |        |                   |
|       | Total      | 56.000         | 5  |             |        |                   |

a. Predictors: (Constant), Tenaga kerja, Pupuk

b. Dependent Variable: Produksi

Pada Tabel 4.6 diatas menunjukkan nilai Fhitung sebesar 51 atau  $sig < 0,005$

## Inteprestasikan uji F

Berdasarkan perhitungan dengan menggunakan uji F menunjukkan bahwa nilainya *sig* 0,005 artinya *sig* < 0,05 atau ( 0,005 < 0,05) artinya bahwa Variable Pupuk (X1) dan Tenaga Kerja (X2) secara serempak berpengaruh Signifikan terhadap Produksi Cabai (Variable Y) pada tingkat kepercayaan 95 Persen

**Tabel 4. 1 nilai coefisien pada Penggunaan *Software* SPSS**  
Coefficients<sup>a</sup>

| Model        | Unstandardized Coefficients |            | Standardized Coefficients | t     | Sig. | 95% Confidence Interval for B |             |
|--------------|-----------------------------|------------|---------------------------|-------|------|-------------------------------|-------------|
|              | B                           | Std. Error | Beta                      |       |      | Lower Bound                   | Upper Bound |
| 1 (Constant) | .600                        | .811       |                           | .740  | .513 | -1.981                        | 3.181       |
| Pupuk        | 1.200                       | .377       | .935                      | 3.182 | .050 | .000                          | 2.400       |
| Tenaga kerja | .200                        | 1.099      | .053                      | .182  | .867 | -3.299                        | 3.699       |

a. Dependent Variable: Produksi

Berdasarkan dari tabel 4.7 diatas bahwa nilai koefisien regresi pada variable pupuk dan tenaga kerja nilainya sama dengan perhitungan secara manual yaitu

$$Y = 0,60 + 1,20X_1 + 0,20X_2 + e$$

Dimana

Y = Produksi Cabai

a = *Intercept*/ Konstanta

bi.bn = Koefisien Regresi

X<sub>1</sub> = Pupuk *Variable Independent* atau variabel bebas

X<sub>2</sub> = Jumlah Tenaga kerja *Variable Independent* atau variabel bebas

e = *Error Term*



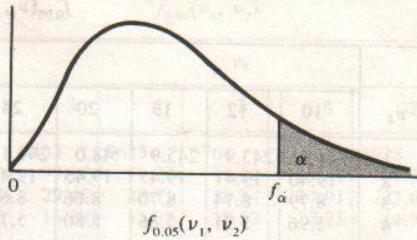
# Lampiran Tabel T dan Tabel F

TABEL A.5\*  
Nilai Kritik Sebaran t

| v    | α     |       |        |        |        |
|------|-------|-------|--------|--------|--------|
|      | 0.10  | 0.05  | 0.025  | 0.01   | 0.005  |
| 1    | 3.078 | 6.314 | 12.706 | 31.821 | 63.657 |
| 2    | 1.886 | 2.920 | 4.303  | 6.965  | 9.925  |
| 3    | 1.638 | 2.353 | 3.182  | 4.541  | 5.841  |
| 4    | 1.533 | 2.132 | 2.776  | 3.747  | 4.604  |
| 5    | 1.476 | 2.015 | 2.571  | 3.365  | 4.032  |
| 6    | 1.440 | 1.943 | 2.447  | 3.143  | 3.707  |
| 7    | 1.415 | 1.895 | 2.365  | 2.998  | 3.499  |
| 8    | 1.397 | 1.860 | 2.306  | 2.896  | 3.355  |
| 9    | 1.383 | 1.833 | 2.262  | 2.821  | 3.250  |
| 10   | 1.372 | 1.812 | 2.228  | 2.764  | 3.169  |
| 11   | 1.363 | 1.796 | 2.201  | 2.718  | 3.106  |
| 12   | 1.356 | 1.782 | 2.179  | 2.681  | 3.055  |
| 13   | 1.350 | 1.771 | 2.160  | 2.650  | 3.012  |
| 14   | 1.345 | 1.761 | 2.145  | 2.624  | 2.977  |
| 15   | 1.341 | 1.753 | 2.131  | 2.602  | 2.947  |
| 16   | 1.337 | 1.746 | 2.120  | 2.583  | 2.921  |
| 17   | 1.333 | 1.740 | 2.110  | 2.567  | 2.898  |
| 18   | 1.330 | 1.734 | 2.101  | 2.552  | 2.878  |
| 19   | 1.328 | 1.729 | 2.093  | 2.539  | 2.861  |
| 20   | 1.325 | 1.725 | 2.086  | 2.528  | 2.845  |
| 21   | 1.323 | 1.721 | 2.080  | 2.518  | 2.831  |
| 22   | 1.321 | 1.717 | 2.074  | 2.508  | 2.819  |
| 23   | 1.319 | 1.714 | 2.069  | 2.500  | 2.807  |
| 24   | 1.318 | 1.711 | 2.064  | 2.492  | 2.797  |
| 25   | 1.316 | 1.708 | 2.060  | 2.485  | 2.787  |
| 26   | 1.315 | 1.706 | 2.056  | 2.479  | 2.779  |
| 27   | 1.314 | 1.703 | 2.052  | 2.473  | 2.771  |
| 28   | 1.313 | 1.701 | 2.048  | 2.467  | 2.763  |
| 29   | 1.311 | 1.699 | 2.045  | 2.462  | 2.756  |
| inf. | 1.282 | 1.645 | 1.960  | 2.326  | 2.576  |

\*Tabel diambil dari Tabel IV R. A. Fisher, *Statistical Methods for Research Workers*, Oliver & Boyd Ltd., Edinburgh, dengan izin pengarang dan penerbit.

**TABEL A.7\***  
**Nilai Kritik Sebaran F**



| $\nu_2$  | $\nu_1$ |       |       |       |       |       |       |       |       |
|----------|---------|-------|-------|-------|-------|-------|-------|-------|-------|
|          | 1       | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     |
| 1        | 161.4   | 199.5 | 215.7 | 224.6 | 230.2 | 234.0 | 236.8 | 238.9 | 240.5 |
| 2        | 18.51   | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 |
| 3        | 10.13   | 9.55  | 9.28  | 9.12  | 9.01  | 8.94  | 8.89  | 8.85  | 8.81  |
| 4        | 7.71    | 6.94  | 6.59  | 6.39  | 6.26  | 6.16  | 6.09  | 6.04  | 6.00  |
| 5        | 6.61    | 5.79  | 5.41  | 5.19  | 5.05  | 4.95  | 4.88  | 4.82  | 4.77  |
| 6        | 5.99    | 5.14  | 4.76  | 4.53  | 4.39  | 4.28  | 4.21  | 4.15  | 4.10  |
| 7        | 5.59    | 4.74  | 4.35  | 4.12  | 3.97  | 3.87  | 3.79  | 3.73  | 3.68  |
| 8        | 5.32    | 4.46  | 4.07  | 3.84  | 3.69  | 3.58  | 3.50  | 3.44  | 3.39  |
| 9        | 5.12    | 4.26  | 3.86  | 3.63  | 3.48  | 3.37  | 3.29  | 3.23  | 3.18  |
| 10       | 4.96    | 4.10  | 3.71  | 3.48  | 3.33  | 3.22  | 3.14  | 3.07  | 3.02  |
| 11       | 4.84    | 3.98  | 3.59  | 3.36  | 3.20  | 3.09  | 3.01  | 2.95  | 2.90  |
| 12       | 4.75    | 3.89  | 3.49  | 3.26  | 3.11  | 3.00  | 2.91  | 2.85  | 2.80  |
| 13       | 4.67    | 3.81  | 3.41  | 3.18  | 3.03  | 2.92  | 2.83  | 2.77  | 2.71  |
| 14       | 4.60    | 3.74  | 3.34  | 3.11  | 2.96  | 2.85  | 2.76  | 2.70  | 2.65  |
| 15       | 4.54    | 3.68  | 3.29  | 3.06  | 2.90  | 2.79  | 2.71  | 2.64  | 2.59  |
| 16       | 4.49    | 3.63  | 3.24  | 3.01  | 2.85  | 2.74  | 2.66  | 2.59  | 2.54  |
| 17       | 4.45    | 3.59  | 3.20  | 2.96  | 2.81  | 2.70  | 2.61  | 2.55  | 2.49  |
| 18       | 4.41    | 3.55  | 3.16  | 2.93  | 2.77  | 2.66  | 2.58  | 2.51  | 2.46  |
| 19       | 4.38    | 3.52  | 3.13  | 2.90  | 2.74  | 2.63  | 2.54  | 2.48  | 2.42  |
| 20       | 4.35    | 3.49  | 3.10  | 2.87  | 2.71  | 2.60  | 2.51  | 2.45  | 2.39  |
| 21       | 4.32    | 3.47  | 3.07  | 2.84  | 2.68  | 2.57  | 2.49  | 2.42  | 2.37  |
| 22       | 4.30    | 3.44  | 3.05  | 2.82  | 2.66  | 2.55  | 2.46  | 2.40  | 2.34  |
| 23       | 4.28    | 3.42  | 3.03  | 2.80  | 2.64  | 2.53  | 2.44  | 2.37  | 2.32  |
| 24       | 4.26    | 3.40  | 3.01  | 2.78  | 2.62  | 2.51  | 2.42  | 2.36  | 2.30  |
| 25       | 4.24    | 3.39  | 2.99  | 2.76  | 2.60  | 2.49  | 2.40  | 2.34  | 2.28  |
| 26       | 4.23    | 3.37  | 2.98  | 2.74  | 2.59  | 2.47  | 2.39  | 2.32  | 2.27  |
| 27       | 4.21    | 3.35  | 2.96  | 2.73  | 2.57  | 2.46  | 2.37  | 2.31  | 2.25  |
| 28       | 4.20    | 3.34  | 2.95  | 2.71  | 2.56  | 2.45  | 2.36  | 2.29  | 2.24  |
| 29       | 4.18    | 3.33  | 2.93  | 2.70  | 2.55  | 2.43  | 2.35  | 2.28  | 2.22  |
| 30       | 4.17    | 3.32  | 2.92  | 2.69  | 2.53  | 2.42  | 2.33  | 2.27  | 2.21  |
| 40       | 4.08    | 3.23  | 2.84  | 2.61  | 2.45  | 2.34  | 2.25  | 2.18  | 2.12  |
| 60       | 4.00    | 3.15  | 2.76  | 2.53  | 2.37  | 2.25  | 2.17  | 2.10  | 2.04  |
| 120      | 3.92    | 3.07  | 2.68  | 2.45  | 2.29  | 2.17  | 2.09  | 2.02  | 1.96  |
| $\infty$ | 3.84    | 3.00  | 2.60  | 2.37  | 2.21  | 2.10  | 2.01  | 1.94  | 1.88  |

\*Direproduksi dari Tabel 18 *Biometrika Tables for Statisticians*, Vol. I, dengan izin dari E. S. Pearson dan Biometrika Trustees.

## DAFTAR PUSTAKA

- Gujarati, D. 2004. *Basic Econometrics*. 4 th Ed. *The McGraw-Hills Companies*.
- Koutsoyiannis, A. 1997. *Theory of Econometrics. An Introductory Exposition of Econometric Methods*. London: *Macmillan Press Ltd*.

# BAB 5

## PENDEKATAN MATRIKS TERHADAP REGRESI LINIER

Oleh Azis

### 5.1 Pendahuluan

Aktivitas kreatif dan hasil yang dihasilkannya dapat diukur atau diungkapkan dengan dua cara, yaitu data yang bersifat kuantitatif dan data yang bersifat kualitatif. Untuk data kuantitatif, data dapat dibedakan atas dua macam yaitu: data yang bersifat bebas dimana data yang ditentukan dengan mana suka atau bebas untuk dipilih. Data ini sering disebut dengan variabel bebas atau variabel prediktor atau *independent variable*. Sedangkan data yang tergantung atau terikat pada variabel bebas disebut dengan variabel tak bebas atau variabel tergantung atau variabel terikat atau variabel endogen atau kreterium atau *dependent variable*.

Hubungan antara variabel bebas dan variabel terikat dapat dimodelkan dalam suatu persamaan matematis yang sering juga disebut sebagai persamaan regresi. Model regresi ini setidaknya memiliki 3 dimensi yaitu untuk tujuan mendeskripsikan data atau studi kasus, untuk tujuan pengendalian dan untuk tujuan prediksi. Regresi dapat menggambarkan peristiwa data dengan memodelkan hubungan statistik. Regresi juga dapat digunakan untuk mengontrol observasi menggunakan model regresi kausal. Selain itu, model regresi dapat digunakan untuk memprediksi variabel dependen. Namun perlu diingat bahwa prediksi dalam

teori regresi hanya dapat dilakukan pada kumpulan data yang menggunakan variabel independen untuk membuat variabel.

Persamaan regresi sendiri dapat digolongkan dalam berbagai macam. Salah satunya ialah regresi linear, yaitu bentuk atau model hubungan antara variabel terikat (*dependen*; respon;  $Y$ ) dengan variabel bebas (*independen*, prediktor,  $X$ ) sebagai faktor yang berpangkat satu. Apabila banyaknya variabel bebas hanya ada satu, disebut sebagai **regresi linier sederhana**, sedangkan apabila terdapat lebih dari 1 variabel bebas, disebut sebagai **regresi linier berganda**.

Model dari regresi linear untuk populasi dapat dituliskan sebagai berikut: (P.Stevens, 2009)

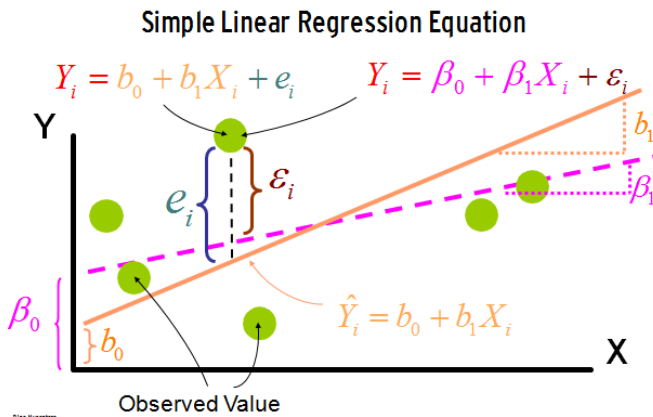
$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i,p-1} + e_i$$

dalam hal ini:

- $y_i$  = nilai variabel terikat dalam amatan ke- $i$ ;
- $\beta_0, \beta_1, \dots, \beta_{p-1}$  = parameter yang tidak diketahui;
- $X_i$  = nilai variabel bebas dari amatan ke- $i$ ;
- $e_i$  = error yang saling bebas, menyebar normal dengan rata-rata  $E\{e_i\} = 0$  dan suatu ragam tetap atau  $N(0, \sigma^2)$
- $i$  = 1, 2, 3, ...,  $n$ .

Kenyataannya seringkali kita tidak bisa mengamati seluruh anggota populasi, sehingga kita hanya mengambil sampel saja, misalnya sampel berukuran  $n$  dan dituliskan  $\{(x_i, y_i), i = 1, 2, 3, \dots, n\}$ . Persamaan yang dihasilkan merupakan estimasi dari persamaan tersebut dan dapat dituliskan:  $\hat{y}_i = b_0 + b_1 x_i$  dimana  $b_0$  adalah estimator dari  $\beta_0$  dan  $b_1$  adalah estimator dari  $\beta_1$ . (Kutner *et al.*, 2005).

Untuk variabel bebas  $x_i$  nilai pengamatan  $y_i$  tidak selalu tepat berada pada garis  $\hat{y}_i = \beta_0 + \beta_1 x_i$  (garis regresi populasi) atau  $\hat{y}_i = b_0 + b_1 x_i$  (garis regresi sampel).



## 5.2 Pendekatan Matriks

Andaikan kita telah memiliki sampel sebanyak  $n \geq k + 1$  untuk nilai-nilai  $x_1, x_2, \dots, x_k$ , dan nilai-nilai  $y_1, y_2, \dots, y_n$ . Jika  $x_{ij}$  menyatakan nilai ke- $i$  (dari responden ke- $i$ ) dari variabel independen  $x_j$  maka diperoleh sistem persamaan:

$$\hat{y}_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + e_1$$

$$\hat{y}_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + e_2$$

$\vdots$

$$\hat{y}_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_k x_{nk} + e_n$$

Sistem persamaan tersebut dapat kita tulis dalam notasi matriks berikut:

$$\hat{Y} = X\beta + \varepsilon$$

dengan

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$\hat{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Disini  $\hat{Y}$  dan  $\varepsilon$  masing-masing adalah vektor acak berukuran  $n \times 1$ ,  $\beta$  adalah vektor berukuran  $(k + 1) \times 1$ , dan  $X$  adalah matriks berukuran  $n \times (k + 1)$ . Kita asumsikan vektor acak  $\varepsilon$  berdistribusi normal dengan *mean* 0 dan variansi  $\sigma^2 I$  serta tidak berkorelasi.

Dengan Menggunakan Cara Manual Metode Matriks

$$X'X\beta = X'Y$$

$$\beta = (X'X)^{-1}X'Y$$

$$(X'X)^{-1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix}$$

$$= \begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} & \cdots & \sum X_{in} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} & \cdots & \sum X_{i1}X_{in} \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}^2 & \cdots & \sum X_{i2}X_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{in} & \sum X_{i1}X_{in} & \sum X_{i2}X_{in} & \cdots & \sum X_{in}^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{k1} & X_{k2} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_1 \\ \sum X_{i1}Y_2 \\ \sum X_{i2}Y_2 \\ \vdots \\ \sum X_{in}Y_n \end{bmatrix}$$

Jadi,

$$= \begin{bmatrix} n & \sum X_{i1} & \sum X_{i2} & \cdots & \sum X_{in} \\ \sum X_{i1} & \sum X_{i1}^2 & \sum X_{i1}X_{i2} & \cdots & \sum X_{i1}X_{in} \\ \sum X_{i2} & \sum X_{i1}X_{i2} & \sum X_{i2}^2 & \cdots & \sum X_{i2}X_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum X_{in} & \sum X_{i1}X_{in} & \sum X_{i2}X_{in} & \cdots & \sum X_{in}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \sum Y_1 \\ \sum X_{i1}Y_2 \\ \sum X_{i2}Y_2 \\ \vdots \\ \sum X_{in}Y_n \end{bmatrix}$$

Contohnya, dimisalkan data untuk memperjelas langkah-langkah perhitungan berikut:

| X <sub>1</sub> | X <sub>2</sub> | Y |
|----------------|----------------|---|
| 5              | 3              | 7 |
| 6              | 4              | 7 |
| 7              | 5              | 8 |
| 8              | 5              | 9 |
| 7              | 4              | 7 |
| 6              | 3              | 6 |
| 5              | 3              | 6 |
| 6              | 4              | 7 |
| 7              | 5              | 9 |
| 5              | 4              | 7 |



Kita mengasumsikan model setiap subjek skor  $y$  dalam fungsi linear dari  $\beta$

$$y_1 = 5 = 1.\beta_0 + 3.\beta_1 + 7.\beta_2 + \varepsilon_1$$

$$y_2 = 6 = 1.\beta_0 + 4.\beta_1 + 7.\beta_2 + \varepsilon_2$$

$$y_3 = 7 = 1.\beta_0 + 5.\beta_1 + 8.\beta_2 + \varepsilon_1$$

$$y_4 = 8 = 1.\beta_0 + 5.\beta_1 + 9.\beta_2 + \varepsilon_1$$

$$y_5 = 7 = 1.\beta_0 + 4.\beta_1 + 7.\beta_2 + \varepsilon_1$$

$$y_6 = 6 = 1.\beta_0 + 3.\beta_1 + 6.\beta_2 + \varepsilon_1$$

$$y_7 = 5 = 1.\beta_0 + 3.\beta_1 + 6.\beta_2 + \varepsilon_1$$

$$y_8 = 6 = 1.\beta_0 + 4.\beta_1 + 7.\beta_2 + \varepsilon_1$$

$$y_9 = 7 = 1.\beta_0 + 5.\beta_1 + 9.\beta_2 + \varepsilon_1$$

$$y_{10} = 5 = 1.\beta_0 + 4.\beta_1 + 7.\beta_2 + \varepsilon_1$$

Persamaan-persamaan tersebut dapat kita rangkai persamaan matriks tunggal seperti berikut.

$$Y = \begin{matrix} & \mathbf{X} & & \mathbf{\beta} & & \mathbf{\varepsilon} \\ \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \\ 7 \\ 6 \\ 5 \\ 6 \\ 7 \\ 5 \end{bmatrix} & = & \begin{bmatrix} 1 & 3 & 7 \\ 1 & 4 & 7 \\ 1 & 5 & 8 \\ 1 & 5 & 9 \\ 1 & 4 & 7 \\ 1 & 3 & 6 \\ 1 & 3 & 6 \\ 1 & 3 & 6 \\ 1 & 4 & 7 \\ 1 & 5 & 9 \\ 1 & 4 & 7 \end{bmatrix} & \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} & + & \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \\ \varepsilon_{10} \end{bmatrix} \end{matrix}$$

Dari nilai ini, jelas bahwa nilai  $y$  dan  $\varepsilon$  didefinisikan dalam vektor kolom, sedangkan tidak begitu jelas adalah bagaimana kotak di bidang dapat direpresentasikan sebagai produk dari dua matriks  $X\beta$ .

Kolom pertama matriks memberi Anda akses ke regresi konstan. Dua kolom sisanya berisi nilai subjek dan dua prediktor. Oleh karena itu, ukuran matriks untuk regresi berganda adalah:

$$Y = X\beta + \varepsilon$$

kemudian dapat dihitung dengan menggunakan kalkulus bahwa kuadrat terkecil perkiraan  $\beta$  diberikan oleh:

$$\beta = (X'X)^{-1}X'Y$$

sehingga keluar data koefisien regresi yang ditaksir menjadi:

$$\beta = \left\{ \begin{matrix} X' & X & X' & Y \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 4 & 5 & 5 & 4 & 3 & 3 & 4 & 5 & 4 \\ 7 & 7 & 8 & 9 & 7 & 6 & 6 & 7 & 9 & 7 \end{bmatrix} & \begin{bmatrix} 1 & 3 & 7 \\ 1 & 4 & 7 \\ 1 & 5 & 8 \\ 1 & 5 & 9 \\ 1 & 4 & 7 \\ 1 & 3 & 6 \\ 1 & 3 & 6 \\ 1 & 4 & 7 \\ 1 & 5 & 9 \\ 1 & 4 & 7 \end{bmatrix}^{-1} & \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 4 & 5 & 5 & 4 & 3 & 3 & 4 & 5 & 4 \\ 7 & 7 & 8 & 9 & 7 & 6 & 6 & 7 & 9 & 7 \end{bmatrix} & \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \\ 7 \\ 6 \\ 5 \\ 6 \\ 7 \\ 5 \end{bmatrix} \end{matrix} \right.$$

Langkah pertama :

$$X'X = \begin{bmatrix} 10 & 62 & 40 \\ 62 & 394 & 254 \\ 40 & 254 & 166 \end{bmatrix} \text{ dan } X'Y = \begin{bmatrix} 73 \\ 460 \\ 299 \end{bmatrix}$$

Langkah selanjutnya, menunjukkan bahwa:

$$(X'X)^{-1} = \begin{bmatrix} 4,111 & -0,611 & -0,056 \\ -0,611 & 0,278 & -0,278 \\ -0,056 & -0,278 & 0,444 \end{bmatrix}$$

Sehingga perkiraan koefisien regresi diberikan oleh:

$$\beta = \begin{bmatrix} 4,111 & -0,611 & -0,056 \\ -0,611 & 0,278 & -0,278 \\ -0,056 & -0,278 & 0,444 \end{bmatrix} \begin{bmatrix} 73 \\ 460 \\ 299 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 2,389 \\ 0,111 \\ 1,056 \end{bmatrix}$$

Sehingga, persamaan regresi (variabel bebas) adalah

$$\hat{y} = 2,389 + 0,111x_1 + 1,056x_2$$

Dari persamaan ini dapat disimpulkan bahwa nilai  $y = 2,389$  apabila nilai  $x_1 = 0$  dan  $x_2 = 0$ . Nilai  $y$  akan bertambah sesuai dengan pertambahan nilai  $x_1$  dan  $x_2$ .

## DAFTAR PUSTAKA

- Azis, A. (2021) *Belajar Statistika Dengan SPSS dan Manual*, Nashir Al-Kutub Indonesia. Mataram: Nashir Al-Kutub Indonesia.
- Kutner, M. *et al.* (2005) *Applied Statistical Linear Models*, McGraw Hill.
- P.Stevens, J. (2009) *Applied Multivariate Statistics for the Social Sciences (5th ed)*, International Statistical Review.



# **BAB 6**

## **DIAGNOSIS DAN PENGUJIAN ASUMSI PADA MODEL REGRESI**

**Oleh Santje Magdalena Iriyanto**

Diagnosis dan pengujian asumsi pada model regresi adalah langkah penting dalam analisis regresi untuk memastikan bahwa model yang digunakan sesuai dengan data dan memberikan estimasi yang valid. Berikut adalah langkah-langkah umum dalam diagnosis dan pengujian asumsi pada model regresi:

### **6.1 Asumsi Linearitas**

Asumsi linearitas dalam konteks model regresi menyatakan bahwa hubungan antara variabel independen (prediktor) dan variabel dependen (variabel yang ingin diprediksi) adalah linear. Dalam istilah yang lebih sederhana, ini berarti bahwa perubahan dalam variabel independen secara proporsional terkait dengan perubahan dalam variabel dependen.

Lebih spesifik, asumsi linearitas menyatakan bahwa efek dari setiap unit perubahan dalam variabel independen terhadap variabel dependen adalah konstan. Dengan kata lain, dalam sebuah model regresi, kita mengasumsikan bahwa garis regresi yang digunakan untuk memprediksi variabel dependen adalah garis lurus, bukan kurva atau polinomial lainnya.

Dalam menganalisis asumsi linearitas, ada beberapa hal yang dapat diperhatikan:

**1. Plot Residual vs Fitted Values:**

Plot ini adalah alat yang umum digunakan untuk memeriksa apakah hubungan antara variabel independen dan dependen adalah linear. Jika plot ini menunjukkan pola yang acak, tanpa pola tertentu (misalnya, pola U-shaped atau pola lainnya), maka asumsi linearitas mungkin terpenuhi.

**2. Transformasi Variabel:**

Kadang-kadang, hubungan antara variabel independen dan dependen tidak sepenuhnya linear. Dalam kasus seperti ini, seringkali memungkinkan untuk melakukan transformasi pada variabel agar hubungannya menjadi lebih linear. Transformasi ini bisa berupa transformasi logaritmik, kuadratik, atau jenis transformasi lainnya.

**3. Uji Statistik:**

Ada beberapa uji statistik yang dapat digunakan untuk menguji apakah hubungan antara variabel independen dan dependen adalah linear. Namun, interpretasi hasil uji statistik ini bisa agak rumit dan tergantung pada konteks kasusnya

Memastikan bahwa asumsi linearitas terpenuhi penting karena kesalahan dalam memenuhi asumsi ini dapat menyebabkan kesalahan dalam estimasi parameter model dan interpretasi hasilnya. Jika asumsi linearitas tidak terpenuhi, maka model regresi yang lebih kompleks atau transformasi variabel mungkin diperlukan.

**Diagnosis:**

- **Plot Residual vs Fitted Values:** Plot ini digunakan untuk memeriksa apakah residual tersebar secara acak di sekitar garis horizontal di sekitar nol. Jika terdapat pola yang jelas, asumsi linearitas mungkin tidak terpenuhi.

## 6.2 Asumsi Independensi

Asumsi independensi dalam konteks model regresi menyatakan bahwa observasi data adalah independen satu sama lain. Artinya, nilai-nilai dalam sampel data tidak dipengaruhi oleh nilai-nilai lainnya. Ini berarti bahwa tidak ada ketergantungan antara observasi yang berbeda dalam data.

Dalam konteks asumsi independensi, beberapa hal yang perlu diperhatikan adalah:

### 1. Data Time Series:

Jika data diambil dari waktu ke waktu (time series data), maka asumsi independensi sering kali tidak terpenuhi karena nilai-nilai observasi pada waktu tertentu dapat dipengaruhi oleh nilai-nilai pada waktu sebelumnya. Dalam hal ini, perlu dilakukan analisis khusus untuk mengatasi ketergantungan tersebut.

### 2. Autokorelasi:

Autokorelasi terjadi ketika nilai-nilai dalam deret waktu saling terkait, artinya nilai pada waktu tertentu bergantung pada nilai pada waktu sebelumnya. Untuk memeriksa autokorelasi, uji statistik seperti Durbin-Watson test atau plot residual vs. time dapat digunakan.

### 3. Pengambilan Sampel:

Penting untuk memastikan bahwa sampel data diambil secara acak dan tidak mengikuti pola atau tren tertentu. Pengambilan sampel yang tidak acak atau terpengaruh oleh faktor-faktor tertentu dapat menghasilkan observasi yang tidak independen.

### 4. Observasi Heterogenitas:

Ketika ada kumpulan observasi yang sangat berbeda dalam data, bisa jadi ada ketergantungan antarobservasi yang berasal dari kelompok-kelompok ini. Penting untuk memeriksa apakah ada ketergantungan semacam itu dan, jika ada, memperlakukan observasi dengan cara yang sesuai.



Memastikan bahwa asumsi independensi terpenuhi penting karena pelanggaran terhadap asumsi ini dapat menghasilkan estimasi parameter model yang bias atau tidak akurat. Jika asumsi independensi tidak terpenuhi, metode analisis yang lebih canggih atau penanganan khusus mungkin diperlukan untuk memperoleh hasil yang dapat diandalkan.

### **Diagnosis:**

- **Durbin-Watson Test:** Digunakan untuk mendeteksi autokorelasi di residual.
- **Plot Residual vs Time:** Untuk data time series, plot residual terhadap waktu dapat digunakan untuk melihat pola autokorelasi.

## **6.3 Asumsi Homoskedastisitas**

Asumsi homoskedastisitas dalam konteks model regresi menyatakan bahwa varians dari error (residual) adalah konstan untuk semua nilai prediktor atau nilai-nilai independen. Dengan kata lain, asumsi ini menyiratkan bahwa sebaran atau variabilitas dari kesalahan prediksi adalah sama di seluruh rentang nilai-nilai prediktor.

Dalam konteks asumsi homoskedastisitas, ada beberapa hal yang perlu diperhatikan:

### **1. Plot Residual vs. Fitted Values:**

Plot ini digunakan untuk memeriksa apakah varians residual tetap konstan sepanjang rentang nilai-nilai prediktor. Jika plot menunjukkan pola yang teratur, seperti pola yang melebar atau menyempit, ini bisa menunjukkan ketidaksesuaian dengan asumsi homoskedastisitas.

### **2. Uji Statistik:**

Beberapa uji statistik dapat digunakan untuk secara formal menguji homoskedastisitas, seperti uji Breusch-Pagan atau uji White. Jika nilai p-value dari uji ini rendah (biasanya di

bawah tingkat signifikansi yang ditetapkan), maka asumsi homoskedastisitas dianggap dilanggar.

### 3. Transformasi Variabel:

Jika ditemukan bahwa varians residual tidak konstan, kadang-kadang menerapkan transformasi pada variabel dependen atau independen dapat membantu dalam memenuhi asumsi homoskedastisitas.

### 4. Heteroskedastisitas:

Kebalikan dari homoskedastisitas adalah heteroskedastisitas, yang terjadi ketika varians residual tidak konstan. Pelanggaran terhadap asumsi homoskedastisitas dapat menyebabkan interval kepercayaan dan uji hipotesis menjadi tidak konsisten atau tidak valid.

Memastikan bahwa asumsi homoskedastisitas terpenuhi penting karena pelanggaran terhadap asumsi ini dapat menghasilkan estimasi yang bias atau tidak akurat. Jika asumsi homoskedastisitas tidak terpenuhi, perlu dipertimbangkan langkah-langkah seperti transformasi variabel atau penggunaan metode analisis yang lebih canggih untuk mengatasi masalah ini.

### Diagnosis:

- **Plot Residual vs Fitted Values:** Digunakan untuk memeriksa apakah varians residual berubah-ubah (menggebu atau menyempit) saat nilai fitted berubah.
- **Breusch-Pagan Test atau White Test:** Uji statistik untuk mendeteksi heteroskedastisitas.

## 6.4 Asumsi Normalitas Residual

Normalitas residual penting karena banyak metode inferensi statistik yang bergantung pada asumsi bahwa residual memiliki distribusi normal. Beberapa hal yang perlu diperhatikan dalam mengenali atau memeriksa asumsi normalitas residual adalah:

### 1. **Histogram:**

Histogram dari residual digunakan untuk memvisualisasikan distribusi residual. Jika histogram menyerupai kurva lonceng (bell-shaped curve) atau distribusi normal, ini menunjukkan bahwa asumsi normalitas mungkin terpenuhi.

### 2. **Q-Q Plot:**

Q-Q (quantile-quantile) plot digunakan untuk membandingkan distribusi dari residual dengan distribusi normal. Jika titik-titik dalam Q-Q plot berada pada garis diagonal, ini menunjukkan kesesuaian antara distribusi residual dan distribusi normal.

### 3. **Uji Statistik:**

Ada berbagai uji statistik yang dapat digunakan untuk secara formal menguji normalitas residual, seperti uji Shapiro-Wilk, uji Kolmogorov-Smirnov, atau Anderson-Darling test. Nilai p-value yang tinggi dari uji ini (biasanya di atas tingkat signifikansi yang ditetapkan) menunjukkan bahwa tidak ada cukup bukti untuk menolak asumsi normalitas.

### 4. **Ketidaksimetrisan:**

Ketika residual memiliki distribusi yang tidak simetris atau berat di salah satu ekor distribusi, ini dapat menunjukkan ketidaknormalan dalam residual.

### 5. **Transformasi Variabel:**

Kadang-kadang, melakukan transformasi pada variabel dependen atau independen dapat membantu dalam memenuhi asumsi normalitas residual.

Meskipun normalitas residual adalah asumsi yang penting dalam model regresi, penting juga untuk diingat bahwa asumsi ini dapat ditoleransi jika ukuran sampel cukup besar karena teorema limit sentral. Namun, memastikan normalitas residual adalah praktik yang baik karena dapat meningkatkan interpretasi dan kepercayaan hasil dari model regresi.

### **Diagnosis:**

- **Histogram dan Q-Q Plot:** Histogram residual dan Q-Q plot (quantile-quantile plot) digunakan untuk memeriksa normalitas residual.

- **Shapiro-Wilk Test atau Kolmogorov-Smirnov Test:** Uji statistik untuk memeriksa normalitas.

## **6.5 Asumsi Tidak Ada Multikolinearitas**

Asumsi tidak adanya multikolinearitas dalam konteks model regresi menyatakan bahwa tidak ada hubungan linear yang sempurna atau sangat kuat antara variabel independen yang ada dalam model. Multikolinearitas terjadi ketika dua atau lebih variabel independen dalam model regresi memiliki korelasi yang tinggi satu sama lain.

Beberapa poin penting terkait dengan asumsi tidak adanya multikolinearitas:

### **1. Korelasi Antara Variabel Independen:**

Untuk memeriksa apakah ada multikolinearitas, Anda dapat menggunakan matriks korelasi antara variabel independen. Korelasi yang tinggi antara pasangan variabel independen menunjukkan kemungkinan adanya multikolinearitas.

### **2. Variance Inflation Factor (VIF):**

VIF adalah metode yang umum digunakan untuk mengukur seberapa banyak varians dari suatu koefisien regresi yang meningkat karena multikolinearitas. Nilai VIF yang tinggi (biasanya di atas 10) menunjukkan adanya multikolinearitas yang signifikan.

### **3. Condition Index:**

Condition index adalah ukuran lain yang digunakan untuk mendeteksi multikolinearitas. Nilai condition index yang tinggi (biasanya di atas 30) menunjukkan adanya multikolinearitas yang signifikan.

### **4. Efek Tersembunyi:**

Multikolinearitas dapat menyebabkan koefisien regresi menjadi tidak stabil dan sulit untuk diinterpretasikan. Selain itu, multikolinearitas dapat menyebabkan kesalahan dalam penilaian signifikansi variabel independen dalam model.

## 5. Penanganan Multikolinearitas:

Jika multikolinearitas terdeteksi, beberapa langkah yang dapat diambil adalah menghapus variabel yang berkorelasi tinggi, menggabungkan variabel menjadi indeks baru, atau menggunakan metode regresi yang lebih tahan terhadap multikolinearitas seperti regresi ridge atau regresi lasso.

Memastikan bahwa tidak ada multikolinearitas dalam model regresi penting karena multikolinearitas dapat menghasilkan estimasi yang tidak stabil dan tidak dapat diandalkan. Dengan memeriksa dan menangani multikolinearitas, kita dapat meningkatkan keandalan dan interpretasi hasil dari model regresi.

### Diagnosis:

- **Variance Inflation Factor (VIF):** VIF yang tinggi (umumnya  $> 10$ ) menunjukkan adanya multikolinearitas yang tinggi.
- **Condition Index:** Nilai di atas 30 dapat menunjukkan multikolinearitas yang serius.

## 6.6 Asumsi Tidak Ada Pengaruh Leverage yang Berlebihan

Asumsi tidak ada pengaruh leverage yang berlebihan dalam konteks model regresi menyatakan bahwa tidak ada observasi yang memiliki pengaruh yang sangat besar atau dominan terhadap hasil dari model regresi. Pengaruh leverage merujuk pada seberapa jauh nilai observasi tertentu berada dari nilai rata-rata variabel independen.

Beberapa poin penting terkait dengan asumsi tidak ada pengaruh leverage yang berlebihan:

### 1. Leverage Values (*Hat Values*):

Leverage values adalah ukuran dari seberapa jauh observasi tertentu berada dari pusat massa variabel independen. Nilai leverage yang tinggi menunjukkan bahwa observasi

tersebut memiliki potensi untuk memiliki pengaruh yang besar terhadap hasil dari model regresi.

## 2. ***Cook's Distance*:**

*Cook's distance* adalah ukuran dari seberapa besar perubahan dalam estimasi koefisien regresi jika observasi tertentu dihapuskan dari model. Nilai *Cook's distance* yang tinggi menunjukkan bahwa observasi tersebut memiliki pengaruh yang besar terhadap hasil dari model.

## 3. **DFFITS dan DFBETAS:**

DFFITS adalah ukuran dari seberapa besar pengaruh penghapusan suatu observasi terhadap prediksi keseluruhan model, sedangkan DFBETAS adalah ukuran dari seberapa besar perubahan dalam koefisien regresi jika observasi tertentu dihapuskan dari model. Nilai DFFITS dan DFBETAS yang tinggi menunjukkan observasi tersebut memiliki pengaruh leverage yang signifikan.

## 4. **Identifikasi Observasi Berpengaruh:**

Untuk mengidentifikasi observasi yang memiliki pengaruh leverage yang berlebihan, dapat digunakan plot leverage vs. residual squared atau plot *Cook's distance*. Observasi yang terletak jauh dari garis aturan atau memiliki nilai *Cook's distance* yang signifikan dapat dianggap memiliki pengaruh leverage yang berlebihan.

## 5. **Penanganan Pengaruh Leverage:** Jika observasi dengan pengaruh leverage yang berlebihan teridentifikasi, beberapa langkah yang dapat diambil adalah mengevaluasi ulang data, mempertimbangkan untuk menghapus observasi tersebut dari analisis, atau menggunakan teknik analisis yang lebih tahan terhadap pengaruh leverage seperti regresi robust atau regresi dengan berat.

Memastikan bahwa tidak ada pengaruh leverage yang berlebihan dalam model regresi penting karena pengaruh leverage yang berlebihan dapat menyebabkan estimasi yang tidak stabil dan tidak dapat diandalkan. Dengan memeriksa dan

menangani pengaruh leverage, kita dapat meningkatkan keandalan dan interpretasi hasil dari model regresi

### **Diagnosis:**

- **Leverage Values (Hat Values):** Nilai leverage yang tinggi menunjukkan pengaruh besar pada model.
- **Cook's Distance:** Nilai Cook's Distance yang tinggi (umumnya  $> 1$ ) menunjukkan observasi yang memiliki pengaruh besar terhadap estimasi parameter model.

Melalui diagnosis dan pengujian asumsi-asumsi ini, kita dapat memastikan bahwa model regresi yang dibangun valid, estimasi parameter tidak bias, dan inferensi statistik tepat. Jika salah satu asumsi dilanggar, tindakan korektif seperti transformasi variabel, penghapusan outlier, atau penggunaan model yang lebih kompleks mungkin diperlukan. Dengan demikian, proses diagnosis dan pengujian asumsi adalah langkah krusial dalam analisis regresi untuk menghasilkan hasil yang dapat diandalkan dan akurat.

## DAFTAR PUSTAKA

- Draper, N., H. Smith. 1992. Analisis Regresi Terapan Edisi Kedua. Terjemahan oleh Bambang Sumantri. Gramedia Pustaka Utama, Jakarta.
- Herlina, H. (2011). Perbandingan Metode Stepwise, Best Subset Regression, dan Fraksi dalam Pemilihan Model Regresi Berganda Terbaik. Jurusan Matematika, F.MIPA, Universitas Sriwijaya. Tersedia di <http://statistik.studentjournal.ub.ac.id>.
- Pujilestari, S. (2016). Pemilihan Model Regresi Linear Berganda Terbaik Pada Kasus Multikolinearitas Berdasarkan Metode Principal Component Analysis (PCA) Dan Metode Stepwise. Skripsi. Jurusan Matematika, F.MIPA, Universitas Negeri Semarang
- Rencher, AC. 2002. Methods of Multivariate Analysis, Second Edition. Canada : A John Wiley & Sons, Inc. Publications.
- Samosir, dkk. (2014). Analisa Metode Backward dan Metode Forward Untuk Menentukan Persamaan Regresi Linier Berganda: Sainia Matematika, 2(4), 345-360.
- Walpole, R.E. 1982. Pengantar Statistika Edisi ke-3. PT Gramedia Pustaka Utama, Jakarta.





# BAB 7

## PEMERIKSAAN SISAAN MODEL REGRESI LINEAR

Oleh Alona Dwinata

### 7.1 Pendahuluan

Sisaan dalam analisis regresi ( $e_i$ ) merupakan selisih antara data pengamatan variabel terikat ( $y_i$ ) dengan nilai dugaan variabel terikat ( $\hat{y}_i$ ) yang diperoleh dari hasil prediksi persamaan regresi. Definisi sisaan dalam analisis regresi linear ditulis seperti persamaan (7.1).

$$e_i = y_i - \hat{y}_i \quad (7.1)$$

Variabel  $y_i$  bersifat acak karena  $e_i$  yang bersifat acak, sehingga distribusi dari variabel terikat bisa diketahui dengan memeriksa pola sebaran sisaan. Parameter model regresi linear sederhana adalah koefisien regresi yang diperoleh menggunakan Metode Kuadrat Terkecil (MKT). Nilai koefisien regresi ini dapat menggambarkan kondisi data yang sebenarnya jika memenuhi beberapa asumsi yang menjadi syarat pada MKT. Asumsi – asumsi tersebut adalah sebagai berikut (Baumer et al., 2021):

1. Linearitas – variabel bebas ( $x_i$ ) dan variabel terikat ( $y_i$ ) memiliki hubungan linear
2. Independen –  $e_i$  saling bebas
3. Homoskedastisitas –  $e_i$  memiliki variansi konstan
4. Normalitas –  $e_i$  berdistribusi Normal

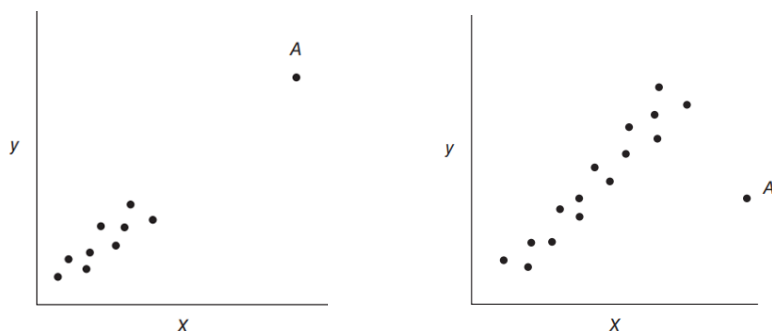
Asumsi 2, 3 dan 4 biasa ditulis dengan notasi  $e_i \sim iid N(0, \sigma^2)$  yang berarti  $e_i$  berdistribusi Normal independen dan identik dengan rata-rata nol dan variansi  $\sigma^2$ .

Analisis sisaan pada regresi linear dapat memberikan informasi tentang ukuran kecocokan model regresi dengan kondisi data pengamatan. Data pengamatan dikategorikan jadi data yang tidak biasa jika data pengamatan tersebut terpisah dari pola data secara umum. Data pengamatan yang tidak biasa ini dapat mengakibatkan perbedaan hasil analisis regresi secara substansial sehingga harus diselidiki lebih lanjut. Kondisi yang tidak biasa ini bisa terjadi karena kesalahan sistem pengukuran, kesalahan penginputan data atau bisa juga terjadi karena data tersebut memang berbeda dari biasanya. Terdapat 3 jenis data yang tidak biasa yaitu pencilan (*outlier*), *leverage point* dan amatan berpengaruh (*influence*).

Pencilan menyebabkan model regresi linear tidak sesuai dengan data pengamatan sehingga diperoleh sisaan yang sangat besar untuk beberapa titik data. Hal ini mengakibatkan model tidak menggambarkan pola sebaran data sebenarnya. *Leverage* menunjukkan besar penyimpangan pengamatan dari rata-rata suatu variabel, kondisi ini dijelaskan oleh pengamatan dengan nilai ekstrim pada suatu variabel prediktor. Suatu hasil pengamatan dikatakan *influence* jika data tersebut tidak dilibatkan saat pemodelan maka terjadi perubahan hasil dugaan koefisien regresi secara substansial. Pendekatan eksploratif mungkin dianggap kurang efisien untuk mendeteksi keberadaan data yang tidak biasa, sehingga dibutuhkan pendekatan yang lebih sistematis dengan pemeriksaan sisaan.

Data pengamatan yang tidak biasa karena kesalahan dalam pengukuran bisa diabaikan atau dibuang sebelum melakukan analisis data. Mengeluarkan data pengamatan yang tidak biasa dapat memperbaiki pola tebaran sisaan dan harus dilakukan dengan hati-hati. Namun, jika kondisi ini terjadi karena data

tersebut memang berbeda dari biasanya maka tidak bisa diabaikan begitu saja. Tindakan mempertahankan data pengamatan dapat dilakukan dengan mempertimbangkan penggunaan teknik pendugaan yang tidak terlalu terpengaruh oleh titik-titik berpengaruh seperti pada MKT. Prinsip dasar teknik pendugaan dengan menurunkan bobot pengamatan secara proporsional dengan besaran atau pengaruh sisa, sehingga pengamatan yang sangat berpengaruh akan menerima bobot lebih kecil dibandingkan dengan kuadrat terkecil. Pendekatan ini disebut analisis regresi *robust*.



**Gambar 7. 1. Contoh data yang tidak biasa**

## 7.2 Penskalaan Sisaan

Terdapat empat jenis penskalaan sisaan yaitu *Standardized Residuals*, *Studentized Residuals*, *PRESS Residuals* dan *R-student* (Douglas C. et al., 2021).

### 1. *Standardized Residuals*

Nilai harapan dari Kuadrat Tengah Sisaan (KTS) pada analisis regresi linear adalah  $\sigma^2$ , maka penduga tak bias untuk  $\sigma^2$  adalah KTS. Penskalaan sisaan terstandarisasi atau *standardized residuals* ( $d_i$ ) didefinisikan seperti Persamaan (7.2)

$$d_i = \frac{e_i}{\sqrt{KTS}} \quad i = 1, 2, \dots, n \quad (7.2)$$

*standardized residuals* memiliki nilai rata-rata nol dan variansi mendekati satu. Nilai  $|d_i| > 3$  berpotensi menunjukkan adanya pencilan.

## 2. Studentized Residuals

Sisaan akan memiliki variansi yang relatif besar jika  $x_i$  berada disekitar  $\bar{x}$  sehingga KTS sebagai penduga variansi  $e_i$  kurang tepat, sehingga diusulkan  $Var(e_i) = KTS(1 - h_{ii})$ . Penskalaan sisaan dengan membagi  $e_i$  dengan standar deviasi sisaan yang baru diusulkan ini disebut *studentized residuals*. Berbeda dengan  $d_i$ . *studentized residuals* memiliki variansi sama dengan satu ( $Var(r_i) = 1$ ). *Studentized residuals* didefinisikan seperti Persamaan (7.3)

$$r_i = \frac{e_i}{\sqrt{KTS(1-h_{ii})}} \quad ; \quad i = 1, 2, \dots, n \quad (7.3)$$

$h_{ii}$  adalah elemen diagonal ke- $i$  matrik  $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ .

Penskalaan *studentized residuals* dengan satu variabel bebas (model regresi linier sederhana) didefinisikan pada Persamaan (7.4)

$$r_i = \frac{e_i}{\sqrt{KTS \left[ 1 - \left( \frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right) \right]}} \quad ; \quad i = 1, 2, \dots, n \quad (7.4)$$

dengan  $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$

## 3. PRESS Residuals atau Deleted Residuals

*PRESS (Prediction Error Sum of Squares)* berbeda dengan dua jenis penskalaan sisaan sebelumnya yang menggunakan pendekatan sisaan dibagi dengan standar

deviasinya. *PRESS residuals* merupakan suatu metode pendeteksian pencilan menggunakan pemeriksaan kuantitas yang dihitung dari  $y_i - \hat{y}_{(i)}$ , dengan  $\hat{y}_{(i)}$  adalah nilai dugaan variabel terikat ke- $i$  dari model regresi yang dibangun tanpa melibatkan pengamatan ke- $i$  atau model regresi linear dibangun dengan  $n - 1$  pengamatan. Prediksi sisaan dengan menghapus pengamatan ke- $i$  ( $e_{(i)}$ ) ditulis seperti Persamaan (7.5)

$$e_{(i)} = y_i - \hat{y}_{(i)} \quad (7.5)$$

Jika pengamatan ke- $i$  jauh dari posisi data pada umumnya, model regresi yang diperoleh berdasarkan semua observasi sangat dipengaruhi oleh amatan ini. Akibatnya, nilai  $\hat{y}_{(i)}$  sangat mirip dengan nilai  $y_i$  sehingga sisaan ( $e_i$ ) menjadi kecil. Kondisi ini sulit untuk mendeteksi pencilan. Namun jika pengamatan ke- $i$  dihapus saat membangun model regresi linear, maka  $\hat{y}_{(i)}$  tidak dipengaruhi oleh pengamatan tersebut, sehingga sisaan yang dihasilkan memiliki kemungkinan yang lebih besar untuk menunjukkan ada atau tidaknya pencilan. Perhitungan prediksi sisaan diulang untuk setiap pengamatan. Perhitungan *PRESS residual* ke- $i$  ditulis pada Persamaan (7.6).

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2 = \sum_{i=1}^n e_{(i)}^2 = \sum_{i=1}^n \left( \frac{e_i}{1-h_{ii}} \right)^2 \quad (7.6)$$

#### 4. *R-student*

*Studentized residual* menggunakan KTS sebagai penduga  $\sigma^2$ , pendekatan ini disebut sebagai penskalaan internal dari sisaan karena penduga ini dihasilkan secara internal dengan menyesuaikan model ke semua pengamatan ( $n$ ). Pendekatan lain tidak mengikutsertakan pengamatan ke- $i$  untuk menduga  $\sigma^2$  sehingga hanya  $n - 1$  pengamatan yang

terlibat. Penduga  $\sigma^2$  yang digunakan untuk mengganti KTS adalah  $S_{(i)}^2$  seperti Persamaan (7.7).

$$S_{(i)}^2 = \frac{(n-p)KTS - \frac{e_i^2}{1-h_{ii}}}{n-p-1} \quad (7.7)$$

Pendekatan ini dikenal dengan *externally studentized residual*. Penskalaan sisaan ini biasanya disebut *R-student* ( $t_i$ ) yang dirumuskan seperti Persamaan (7.8).

$$t_i = \frac{e_i}{\sqrt{S_{(i)}^2 - (1-h_{ii})}} \quad ; i = 1, 2, \dots, n \quad (7.8)$$

## 7.2 Pemeriksaan Sisaan Model Regresi Linear

Asumsi-asumsi yang mendasari pendugaan parameter pada model regresi linear harus diperiksa untuk memastikan validitas model yang diperoleh. Pelanggaran asumsi pada model regresi linear dapat menimbulkan permasalahan yang serius seperti ketidakstabilan model. Kondisi ini dapat dilihat pada dua model yang sangat berbeda dan kesimpulannya berlawanan, padahal model ini diperoleh dari dua sampel yang berasal dari populasi yang sama. Untuk memperoleh model yang layak, sebelum melakukan pendugaan parameter model perlu dilakukan eksplorasi data dan dilanjutkan dengan diagnostik terhadap penduga model. Diagnostik penduga model dilakukan melalui analisis sisaan. Selain diagnostik penduga model melalui sisaan, diagnostik juga digunakan untuk memeriksa apakah terdapat data pencilan (*outlier*) dan amatan berpengaruh (*influence*).

Data pencilan bisa menjadi suatu masalah yang menyebabkan terjadinya pelanggaran asumsi, sehingga penduga yang diperoleh melalui MKT bersifat bias dan tidak efisien. Akibatnya model regresi yang diperoleh tidak cocok dengan data yang dimodelkan. Ada 3 cara yang dapat digunakan untuk mengidentifikasi adanya pencilan dan amatan berpengaruh yaitu *Cook's Distance*, *R-student* dan *DFITS*.

*Cook's Distance* dirumuskan oleh Cook [1977, 1979] menggunakan ukuran jarak kuadrat antara dugaan parameter dengan metode MKT yang melibatkan semua pengamatan ( $\hat{\beta}$ ) dan dugaan parameter dengan metode MKT yang diperoleh dengan tidak melibatkan pengamatan ke- $i$  ( $\hat{\beta}_{(i)}$ ). *Cook's Distance* dirumuskan seperti Persamaan (7.9).

$$D_i = \frac{r_i^2}{p} \frac{h_{ii}}{1-h_{ii}} \quad ; i = 1, 2, \dots, n \quad (7.9)$$

Pengamatan ke- $i$  dikatakan berpengaruh jika nilai  $D_i > \frac{4}{n}$ .

*R-student* ( $t_i$ ) diperoleh menggunakan Persamaan (7.8). Suatu pengamatan dicurigai sebagai pencilan jika pada taraf nyata  $\alpha$  pengamatan tersebut memiliki nilai  $|t_i| > t_{\frac{\alpha}{2}; n-p-1}$ .

*DFFITs* diusulkan oleh Belsley, Kuh, dan Welsch (Douglas C. et al., 2021) seperti Persamaan (7.10)

$$DFFITs_i = t_i \left( \sqrt{\frac{h_{ii}}{1-h_{ii}}} \right) \quad (7.10)$$

Suatu amatan merupakan *influence* jika  $|DFFITs_i| > 2 \sqrt{\frac{p}{n}}$

Pengamatan berpengaruh terjadi jika terjadi perubahan yang signifikan pada koefisien regresi jika pengamatan tersebut tidak digunakan dalam model. Secara umum, pencilan tidak selalu berarti pengamatan berpengaruh ataupun sebaliknya. Keberadaan amatan berpengaruh dalam analisis tidak hanya merubah nilai koefisien regresi tetapi juga akan merubah interpretasi atau kesimpulannya.

Contoh:

Data yang digunakan merupakan data simulasi yang digunakan pada penelitian Indra et al. tahun 2013. Data hasil simulasi dengan  $n = 30$  ditampilkan pada Tabel 7.1.

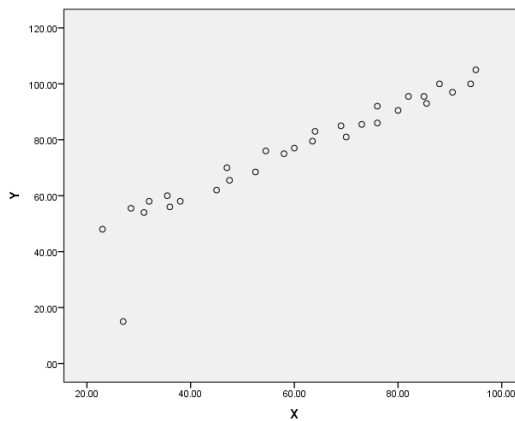


**Tabel 7. 1. Data Simulasi**

| <i>i</i>  | <i>x</i> | <i>y</i> | <i>i</i>  | <i>x</i> | <i>y</i> |
|-----------|----------|----------|-----------|----------|----------|
| <b>1</b>  | 23       | 48       | <b>16</b> | 63,5     | 79,5     |
| <b>2</b>  | 27       | 15       | <b>17</b> | 64       | 83       |
| <b>3</b>  | 28,5     | 55,5     | <b>18</b> | 69       | 85       |
| <b>4</b>  | 31       | 54       | <b>19</b> | 70       | 81       |
| <b>5</b>  | 32       | 58       | <b>20</b> | 73       | 85,5     |
| <b>6</b>  | 35,5     | 60       | <b>21</b> | 76       | 86       |
| <b>7</b>  | 36       | 56       | <b>22</b> | 76       | 92       |
| <b>8</b>  | 38       | 58       | <b>23</b> | 80       | 90,5     |
| <b>9</b>  | 45       | 62       | <b>24</b> | 82       | 95,5     |
| <b>10</b> | 47       | 70       | <b>25</b> | 85       | 95,5     |
| <b>11</b> | 47,5     | 65,5     | <b>26</b> | 85,5     | 93       |
| <b>12</b> | 52,5     | 68,5     | <b>27</b> | 88       | 100      |
| <b>13</b> | 54,5     | 76       | <b>28</b> | 90,5     | 97       |
| <b>14</b> | 58       | 75       | <b>29</b> | 94       | 100      |
| <b>15</b> | 60       | 77       | <b>30</b> | 95       | 105      |

Sumber: (Indra et al., 2013)

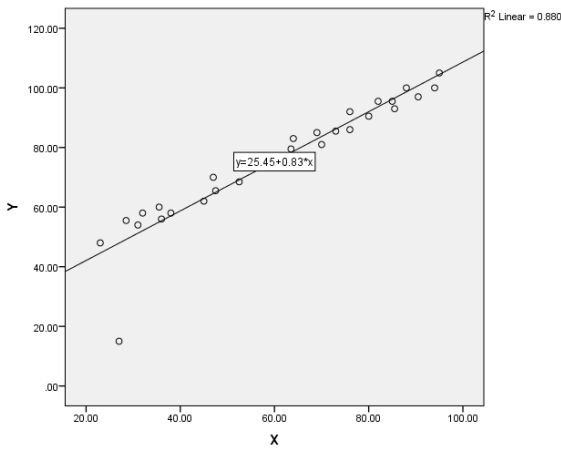
Eksplorasi data untuk melihat hubungan antara variabel *x* dan *y* dinyatakan pada Gambar 7.2.



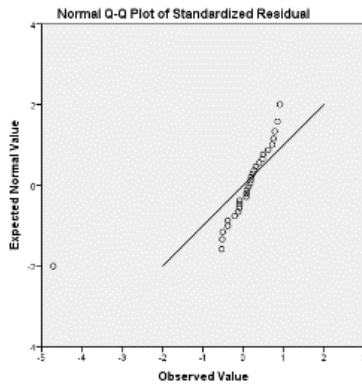
**Gambar 7. 2. Plot variabel *x* dan *y***

Gambar 7.2 menunjukkan bahwa terdapat hubungan linear antara variabel  $x$  dan  $y$ . Namun, terdapat satu titik yang terpisah jauh dari pola sebaran data secara umum. Berdasarkan hasil analisis regresi linear sederhana menggunakan SPSS dengan  $n = 30$  diperoleh persamaan regresi sebagai berikut:

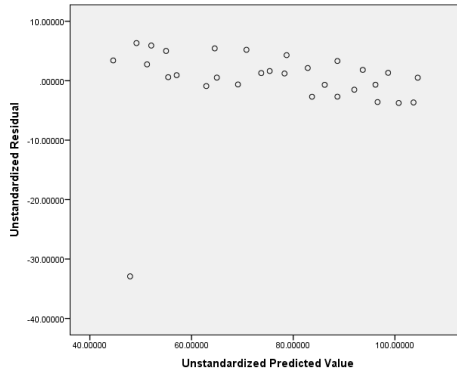
$$\hat{y} = 25,455 + 0,832x \text{ dengan } R^2 = 88\%$$



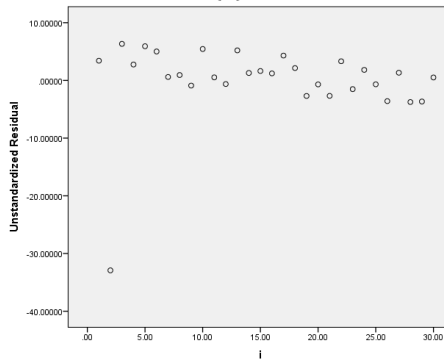
**Gambar 7. 3. Persamaan Garis Regresi Linear Sederhana  $n = 30$**



(a)



(b)



(c)

**Gambar 7. 4. Pemeriksaan Sisaan Model Regresi Linear  $n = 30$**

Pemeriksaan asumsi normalitas, homogenitas dan independen yang mendasari pendugaan parameter model regresi linear dianalisis melalui pemeriksaan sisaan. Hasil pemeriksaan sisaan terkait asumsi dapat dilihat pada Gambar 7.3 yaitu (a) hasil pemeriksaan sisaan terkait asumsi normalitas, (b) homogenitas dan (c) independen. Gambar 7.3(a) menunjukkan bahwa asumsi normalitas tidak dipenuhi karena sebaran data tidak mengikuti garis pada *Normal Q-Q Plot*. Gambar 7.3(b) menunjukkan bahwa asumsi homogenitas juga dilanggar karena terdapat satu pencaran sisaan yang jauh dari nol dan data ini juga merupakan titik amatan yang jauh dari pola sebaran data pada Gambar 7.2. Gambar 7.3(c) menunjukkan bahwa asumsi independen terpenuhi karena secara umum plot sisaan tersebar merata dan tidak membentuk suatu pola tertentu. Hasil pemeriksaan sisaan menunjukkan pelanggaran asumsi yang mendasari MKT dalam pendugaan parameter regresi. Oleh karena itu, perlu dilakukan pemeriksaan keberadaan amatan berpengaruh pada data dengan memeriksa nilai *Cook's Distance*, *R-student* dan *DFFITS*. Nilai *Cook's Distance*, *R-student* dan *DFFITS* dalam model regresi linear sederhana menggunakan SPSS dapat dilihat pada Tabel 7.2.

**Tabel 7. 2. Pemeriksaan pengamatan berpengaruh**

| <i>i</i> | <i>x</i> | <i>y</i> | <i>R-student</i> | <i>Cook's Distance</i> | <i>Standardized DFFIT</i> |
|----------|----------|----------|------------------|------------------------|---------------------------|
| 1        | 23       | 48       | 0,52             | 0,02                   | 0,20                      |
| 2        | 27       | 15       | -14,85           | 1,53                   | -5,20                     |
| 3        | 28,5     | 55,5     | 0,96             | 0,05                   | 0,32                      |
| 4        | 31       | 54       | 0,41             | 0,01                   | 0,13                      |
| 5        | 32       | 58       | 0,88             | 0,04                   | 0,28                      |
| 6        | 35,5     | 60       | 0,74             | 0,02                   | 0,21                      |
| 7        | 36       | 56       | 0,09             | 0,00                   | 0,02                      |
| 8        | 38       | 58       | 0,14             | 0,00                   | 0,04                      |
| 9        | 45       | 62       | -0,13            | 0,00                   | -0,03                     |
| 10       | 47       | 70       | 0,79             | 0,02                   | 0,17                      |
| 11       | 47,5     | 65,5     | 0,08             | 0,00                   | 0,02                      |

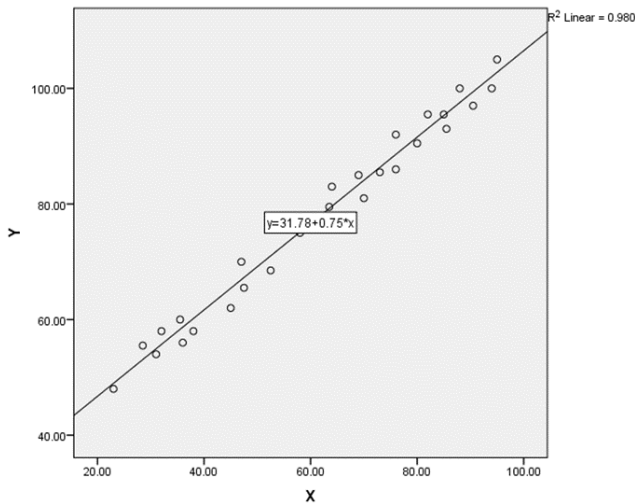
| <i>i</i> | <i>x</i> | <i>y</i> | <i>R-student</i> | <i>Cook's Distance</i> | <i>Standardized DFFIT</i> |
|----------|----------|----------|------------------|------------------------|---------------------------|
| 12       | 52,5     | 68,5     | -0,09            | 0,00                   | -0,02                     |
| 13       | 54,5     | 76       | 0,75             | 0,01                   | 0,14                      |
| 14       | 58       | 75       | 0,18             | 0,00                   | 0,03                      |
| 15       | 60       | 77       | 0,23             | 0,00                   | 0,04                      |
| 16       | 63,5     | 79,5     | 0,17             | 0,00                   | 0,03                      |
| 17       | 64       | 83       | 0,62             | 0,01                   | 0,12                      |
| 18       | 69       | 85       | 0,31             | 0,00                   | 0,06                      |
| 19       | 70       | 81       | -0,39            | 0,00                   | -0,08                     |
| 20       | 73       | 85,5     | -0,10            | 0,00                   | -0,02                     |
| 21       | 76       | 86       | -0,39            | 0,00                   | -0,09                     |
| 22       | 76       | 92       | 0,48             | 0,01                   | 0,11                      |
| 23       | 80       | 90,5     | -0,22            | 0,00                   | -0,06                     |
| 24       | 82       | 95,5     | 0,27             | 0,00                   | 0,07                      |
| 25       | 85       | 95,5     | -0,10            | 0,00                   | -0,03                     |
| 26       | 85,5     | 93       | -0,53            | 0,01                   | -0,15                     |
| 27       | 88       | 100      | 0,20             | 0,00                   | 0,06                      |
| 28       | 90,5     | 97       | -0,56            | 0,02                   | -0,18                     |
| 29       | 94       | 100      | -0,55            | 0,02                   | -0,19                     |
| 30       | 95       | 105      | 0,08             | 0,00                   | 0,03                      |

Nilai *R-student* ( $t_i$ ) pengamatan ke-2 merupakan pencilan karena memiliki nilai  $|t_i| = |-14,85| > t_{\frac{0,05}{2}, 30-2-1} = 2,052$ .

Pengamatan ke-2 merupakan pengamatan berpengaruh karena nilai  $D_i = 1,53 > \frac{4}{30} = 0,13$ . Indikasi yang sama juga ditunjukkan oleh nilai  $|DFFITS_i| = |-5,20| > 2\sqrt{\frac{p}{n}} = 0,51$ , sehingga amatan ke-2 merupakan amatan berpengaruh.

Pemodelan dilanjutkan dengan menghapus data ke-2, sehingga data yang digunakan untuk membangun persamaan regresi ada sebanyak  $n = 29$ . Persamaan regresi linear sederhana yang diperoleh adalah

$$\hat{Y} = 31,783 + 0,747X \text{ dengan } R^2 = 98\%$$



**Gambar 7. 5. Persamaan Garis Regresi Linear Sederhana  $n = 29$**

Pemeriksaan sisaan model regresi linear sederhana dengan  $n = 29$  memenuhi semua asumsi yang mendasari MKT dalam pendugaan parameter regresi linear. Berdasarkan nilai ukuran kebaikan model ditunjukkan oleh nilai  $R^2$  yang mengalami peningkatan sebesar 10%. Jadi dengan menganalisis data yang tidak mengandung pencilan dan amatan berpengaruh akan menghasilkan model regresi linear yang valid untuk menggambarkan kondisi data yang sebenarnya.

## DAFTAR PUSTAKA

- Baumer, B. S., Kaplan, D. T., & Horton, N. J. (2021). Modern Data Science with R. In *Modern Data Science with R*. <https://doi.org/10.1201/9780429200717>
- Douglas C., M., Elizabeth A., P., & G. Geoffrey, V. (2021). Introduction to Linear Regression Analysis, Sixth Edition. In *A JOHN WILEY & SONS, INC., PUBLICATION*.
- Indra, S., Vionanda, D., & Sriningsih, R. (2013). Pendeteksian Data Pencilan dan Pengamatan Berpengaruh pada Beberapa Kasus Data Menggunakan Metode Diagnostik. *Journal of Mathematics UNP*, 1(2).

# BAB 8

## MULTIKOLINIER

Oleh Rahma Faelasofi

### 8.1 Pendahuluan

Penggunaan dan interpretasi model regresi ganda sering tergantung secara eksplisit atau implisit pada perkiraan koefisien regresi individu. Beberapa contoh kesimpulan yang sering dilakukan termasuk sebagai berikut:

1. Mengidentifikasi efek relatif dari variabel regresor
2. Prediksi dan/atau perkiraan
3. Memilih variabel yang tepat untuk model

Jika tidak ada hubungan linear antara regresor, dapat dikatakan orthogonal. Namun, dalam beberapa situasi regresor mungkin sangat terkait secara linear, dan dalam kasus seperti itu inferensi berdasarkan model regresi dapat menyesatkan atau salah. Ketika ada ketergantungan hampir linear di antara regresor, masalah multikolinier dikatakan ada (Pesaran and Smith, 2019).

### 8.2 Pengertian Multikolinier

Multikolinier adalah fenomena statistik yang terjadi dalam analisis regresi ketika dua atau lebih variabel independen (predictors) dalam model berkorelasi satu sama lain. Korelasi yang kuat ini dapat menyebabkan masalah dalam menafsirkan



hasil analisis (Shrestha, 2020). Atau dapat dinyatakan bahwa multikolinear ditunjukkan dengan adanya hubungan linier yang sempurna antara beberapa atau semua variabel yang menjelaskan model regresi (Dias and Brito, 2022).

Tujuan uji multikolinier yaitu untuk mengetahui ada atau tidak korelasi antar variabel independent pada model regresi (Alita, Putra and Darwis, 2021). Atau untuk menguji apakah dalam suatu model regresi ditemukan adanya korelasi yang tinggi atau sempurna antar variabel independent .

Dimisalkan terdapat model regresi sebagai

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Dimana  $\mathbf{y}$  sebagai vektor respon berukuran  $n \times 1$ ,  $\mathbf{X}$  sebagai matriks  $n \times p$  dari variabel regresor,  $\boldsymbol{\beta}$  sebagai vektor konstan yang tidak diketahui berukuran  $p \times 1$ . Dan  $\boldsymbol{\varepsilon}$  sebagai vektor kesalahan acak berukuran  $n \times 1$ , dengan  $\varepsilon_i \sim NID(0, \sigma^2)$ . Dimisalkan ada kolom ke- $j$  pada matriks  $\mathbf{X}$  yang dinyatakan sebagai  $\mathbf{X}_j$ , sehingga  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p]$ . Jadi,  $\mathbf{X}_j$  menggunakan variabel regresor ke- $j$ . Multikolinier dapat didefinisikan secara formal dalam hal ketergantungan linier dari kolom-kolom  $\mathbf{X}$ . Vektor  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_p$  adalah linear dependent jika ada himpunan yang konstan  $t_1, t_2, \dots, t_p$ , tidak semua nol (Montgomery, Peck and Vining, 2012), terutama pada

$$\sum_{j=1}^p t_j \mathbf{X}_j = \mathbf{0} \quad (8.1)$$

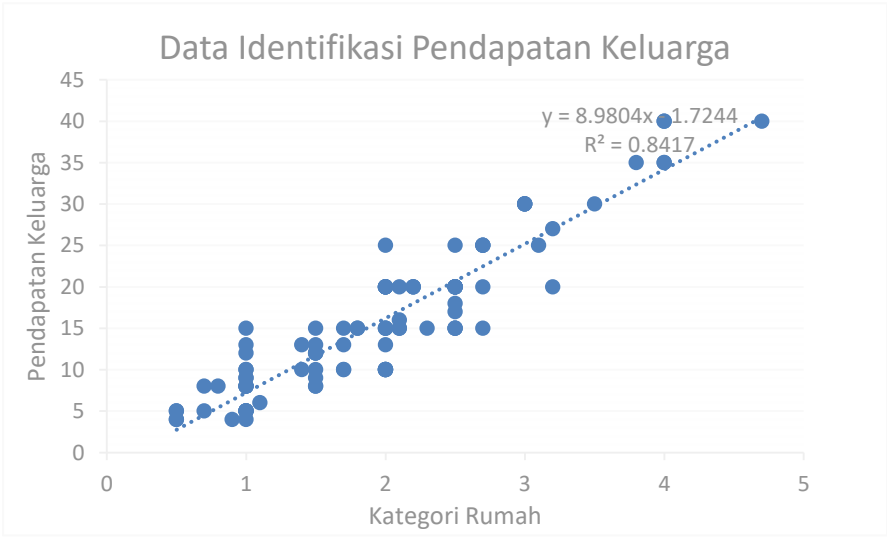
Jika (8.1) memiliki andil pada subset kolom  $\mathbf{X}$ , maka rank matriks  $\mathbf{X}'\mathbf{X}$  kurang dari  $p$  dan  $(\mathbf{X}'\mathbf{X})^{-1}$  tidak ada. Ditunjukkan bahwa persamaan (8.1) adalah hampir benar untuk beberapa subset kolom  $\mathbf{X}$ . Kemudian akan ada ketergantungan hampir linear di  $\mathbf{X}'\mathbf{X}$  dan masalah multikolinear dikatakan ada. Perhatikan bahwa multikolinear adalah bentuk kondisi tidak baik dalam matriks  $\mathbf{X}'\mathbf{X}$  (Chang and Mastrangelo, 2011).

Selanjutnya, masalah satu derajat, yaitu, setiap set data akan mengalami multikolinear sampai batas tertentu kecuali kolom  $X$  adalah orthogonal ( $X'X$  adalah matriks diagonal). Biasanya ini hanya akan terjadi dalam eksperimen yang dirancang. Seperti yang akan kita lihat, kehadiran multikolinear dapat membuat *least-square* pada model analisis regresi secara dramatis menjadi tidak memadai. Dengan demikian, ada empat sumber utama terjadinya multikolinear (Montgomery, Peck and Vining, 2012):

1. Metode pengumpulan data yang digunakan
2. Batas-batas pada model atau populasi
3. Spesifikasi model
4. Model yang terdefiniskan

Metode pengumpulan data dapat menyebabkan masalah multikolinear ketika analisis sampel hanya subsampel dari wilayah regressor (Balqis, Astutik and Solimun, 2022) yang didefinisikan pada persamaan (8.1). Multikolinear juga dapat disebabkan oleh teknik sampling tidak melekat pada model atau populasi yang diambil sampel (Vigneau and Mahévas, 2007). Selain itu, pembatasan pada model atau populasi yang diambil sampel dapat menyebabkan multikolinear (Kalatzis, Bassetto and Azzoni, 2011). Misalnya, anggaplah bahwa sebuah utilitas listrik sedang menyelidiki efek dari pendapatan keluarga ( $x_1$ ) dan ukuran kategori rumah ( $x_2$ ) pada konsumsi listrik rumah tangga. Tingkat dari dua variabel regresor yang diperoleh dalam data sampel ditunjukkan pada gambar. Perhatikan bahwa data terletak kira-kira di sepanjang garis lurus, menunjukkan masalah multikolinear potensial. Dalam contoh ini, keterbatasan fisik dalam populasi telah menyebabkan fenomena ini, yaitu, keluarga dengan pendapatan yang lebih tinggi umumnya memiliki rumah yang lebih besar daripada keluarga dengan penghasilan yang lebih rendah. Ketika keterbatasan fisik seperti ini hadir, multikolinear akan ada terlepas dari metode sampling

yang digunakan. Pembatasan sering terjadi dalam masalah yang melibatkan produksi atau proses kimia, di mana regresor adalah komponen produk, dan komponen-komponen ini menambahkan konstan.



Gambar di atas menunjukkan level pendapatan keluarga dan ukuran rumah untuk penelitian adalah garis lurus dengan intercept. Multikolinear dapat diinduksi oleh pilihan model.

Selain itu, jika rentang  $x$  kecil, menambahkan istilah  $x^2$  dapat menghasilkan multikolinear yang signifikan. Kita sering menemukan situasi seperti ini di mana dua atau lebih regresor hampir secara linear tergantung, dan mempertahankan semua regresor ini dapat berkontribusi pada multikolinear. Dalam kasus ini, beberapa subset regresor biasanya lebih disukai dari sudut pandang multikolinear.

### 8.3 Efek Multikolinear

Kehadiran multikolinear memiliki sejumlah efek yang berpotensi serius pada perkiraan kuadrat terkecil dari koefisien regresi. Beberapa efek ini dapat dengan mudah ditunjukkan.

Misalkan ada dua variabel regresor,  $x_1$  dan  $x_2$ . Model, mengasumsikan bahwa  $x_1$ ,  $x_2$ , dan  $y$  diukur menjadi unit panjang, adalah

$$y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

Persamaan normal *least-square* adalah

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

$$\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} r_{1y} \\ r_{2y} \end{bmatrix}$$

Dimana  $r_{12}$  adalah korelasi sederhana antara  $x_1$  dan  $x_2$  dan  $r_{jy}$  adalah korelasi sederhana antara  $x_j$  dan  $y$ ,  $j = 1, 2$ . Selanjutnya, invers matriks pada  $(\mathbf{X}'\mathbf{X})$  adalah

$$C = (\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} \frac{1}{1-r_{12}^2} & \frac{-r_{12}}{1-r_{12}^2} \\ \frac{-r_{12}}{1-r_{12}^2} & \frac{1}{1-r_{12}^2} \end{bmatrix} \quad (8.2)$$

Faktor-faktor yang mempengaruhi regresi adalah

$$\hat{\beta}_1 = \frac{r_{1y} - r_{12}r_{2y}}{1 - r_{12}^2}, \hat{\beta}_2 = \frac{r_{2y} - r_{12}r_{1y}}{1 - r_{12}^2}$$

Jika ada multikolinear yang kuat antara  $x_1$  dan  $x_2$ , maka koefisien korelasi  $r_{12}$  akan besar.

Persamaan (8.2) menunjukkan bahwa  $|r_{12}| \rightarrow 1, \text{Var}(\hat{\beta}_j) = C_{jj}\sigma^2 \rightarrow \infty$  dan  $\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) = C_{12}\sigma^2 \rightarrow \pm\infty$  tergantung apakah  $r_{12} \rightarrow \pm 1$ . Oleh karena itu, multikolinearitas yang kuat antara  $x_1$  dan  $x_2$  menghasilkan varian dan kovariansi yang besar untuk estimator kuadrat terkecil dari koefisien regresi. Hal ini berarti bahwa sampel yang berbeda diambil pada tingkat  $x$  yang sama dapat menyebabkan perkiraan yang sangat berbeda dari parameter model.

Ketika ada lebih dari dua variabel regresor, multikolinear menghasilkan efek yang sama. dapat ditunjukkan bahwa elemen diagonal matriks  $C = (X'X)^{-1}$  adalah

$$C_{jj} = \frac{1}{1-R_j^2}, \quad j = 1, 2, \dots, p$$

Dimana  $R_j^2$  adalah koefisien *multiple determination* dari regresi  $x_j$  pada variabel regresor  $p - 1$  yang tersisa. Jika ada multikolinear yang kuat antara  $x_j$  dan subset lain  $p - 1$ , maka nilai  $R_j^2$  akan dekat dengan kesatuan. Karena variansi pada  $\hat{\beta}_j$  adalah  $Var(\hat{\beta}_j) = C_{jj}\sigma^2 = (1 - R_j^2)^{-1}\sigma^2$ , multikolinear yang kuat menyiratkan bahwa varians dari kuadrat terkecil. Perkiraan koefisien regresi  $\beta_j$  sangat besar. Secara umum, kovarians dari  $\hat{\beta}_i$  dan  $\hat{\beta}_j$  juga akan besar jika regresor  $x_i$  dan  $x_j$  terlibat dalam hubungan multikolinier.

Multikolinier cenderung menghasilkan perkiraan paling kecil persegi yang terlalu besar dalam nilai mutlak. Pertimbangan jarak persegi dari  $\hat{\beta}$  ke vektor parameter benar  $\beta$ , misalnya,

$$L_1^2 = (\hat{\beta} - \beta)' (\hat{\beta} - \beta)$$

Jarak kuadrat yang diharapkan,  $E(L_1^2)$ , adalah

$$E(L_1^2) = E(\hat{\beta} - \beta)' (\hat{\beta} - \beta) = \sum_{j=1}^p E(\hat{\beta}_j - \beta_j)^2 =$$

$$\sum_{j=1}^p Var(\hat{\beta}_j) = \sigma^2 Tr(X'X)^{-1} \quad (8.3)$$

di mana matriks trace (Tr terkecil) hanyalah jumlah elemen diagonal utama. Ketika ada multikolinier, beberapa nilai sendiri dari  $X'X$  akan kecil. Karena jejak matriks juga merupakan kontribusi setiap regressor. Bentuk persamaan (8.3) menjadi

$$E(L_1^2) = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j} \quad (8.4)$$

Dimana  $\lambda_j > 0, j = 1, 2, \dots, p$ , sebagai nilai eigen pada  $\mathbf{X}'\mathbf{X}$ .

Jadi, jika matriks  $\mathbf{X}'\mathbf{X}$  menjadi kondisi yang tidak baik dari multikolinear, setidaknya salah satu dari  $\lambda_j$  akan kecil, dan persamaan (8.4) berakibat pada jarak dari estimasi *least-square*  $\hat{\boldsymbol{\beta}}$  pada parameter  $\boldsymbol{\beta}$  akan lebih besar. Dengan cara yang sama kita dapat menunjukkan bahwa:

$$E(L_1^2) = E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = E(\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}} - 2\hat{\boldsymbol{\beta}}'\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{\beta})$$

atau

$$E(\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}'\boldsymbol{\beta} + \sigma^2 \text{Tr}(\mathbf{X}'\mathbf{X})^{-1}$$

Bahwasannya, vektor  $\hat{\boldsymbol{\beta}}$  biasanya lebih panjang dari vektor  $\boldsymbol{\beta}$ . Ini menyiratkan bahwa metode kuadrat terkecil menghasilkan estimasi koefisien regresi yang terlalu besar dalam nilai mutlak.

Sementara metode kuadrat terkecil biasanya akan menghasilkan perkiraan buruk dari parameter model individu ketika multikolinier yang kuat hadir, ini tidak selalu berarti bahwa model yang disesuaikan adalah prediktor yang buruk (Montgomery, Peck and Vining, 2012). Jika prediksi terbatas pada daerah ruang  $x$  di mana multikolinier bertahan, model yang disesuaikan sering menghasilkan prediksi yang memuaskan. Hal ini dapat terjadi karena kombinasi linear dapat dinilai cukup baik, meskipun parameter individual  $\beta_j$  dinilai buruk.

## 8.4 Diagnosis Multikolinear

Beberapa teknik telah diusulkan untuk mendeteksi multikolinier, dalam hal ini kriteria pengambilan keputusan terkait multikolinear yaitu:

1. Nilai *Variance Inflation* (VIF), jika nilai VIF < 10 atau nilai tolerance > 0,01, maka dinyatakan tidak terjadi multikolinear
2. Nilai *Variance Inflation* (VIF), jika nilai VIF > 10 atau nilai tolerance < 0,01, maka dinyatakan terjadi multikolinear

Berdasarkan (5.2) bahwa  $C_{jj}$  elemen diagonal dari  $C$ , bisa dituliskan sebagai  $C_{jj} = \left(1 - R_j^2\right)^{-1}$ , dimana  $R_j^2$  adalah koefisien determinasi yang diperoleh ketika  $x_j$  regresi pada  $p - 1$ . Jika  $x_j$  orthogonal,  $R_j^2$  kecil dan  $C_{jj}$  dekat dengan kesatuan, sementara jika  $x_j$  mungkin secara linear tergantung pada beberapa subset regresor yang tersisa,  $R_j^2$  dekat dengan kesatuan dan  $C_{jj}$  yang besar. Karena variance ke- $j$  dengan koefisien regresi adalah  $C_{jj}\sigma^2$ , dapat dilihat  $C_{jj}$  sebagai faktor dengan variance  $\hat{\beta}_j^2$  meningkat, dengan formula  $VIF_j = C_{jj} = \left(1 - R_j^2\right)^{-1}$  (Montgomery, Peck and Vining, 2012), VIF untuk setiap istilah dalam model mengukur efek gabungan dari ketergantungan antara regresor pada varians.

Selain itu, untuk mendeteksi adanya multikolinear dapat diketahui jika nilai koefisien korelasi pada masing-masing variable bebas mencapai lebih dari 0,8 (Shrestha, 2020). Namun, jika koefisien korelasi pada masing-masing variable bebas kurang dari 0,8 maka tidak terjadi multikolinearitas.

Secara umum, data yang diambil dari  $n$  buah sampel dengan masing-masing  $p$  variabel pada setiap sampelnya dapat

dinyatakan dalam suatu matrik  $X$  berukuran  $n \times p$  sebagai berikut:

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$$

Setiap baris dari  $X$  merupakan suatu pengamatan multivariat. Rata-rata setiap variabel dapat diperoleh dari matriks rata-rata  $\bar{x}$  sebagai berikut:

$$\bar{X} = \frac{1}{n} \cdot \mathbf{1}_n' \cdot X$$

Dengan  $\mathbf{1}_n$  adalah matriks baris dengan  $n$  buah kolom, yang semua elemennya bernilai 1. Matriks simpangan terhadap rata-rata ( $T$ ) dapat ditentukan dengan rumus:

$$T = \left[ I - \frac{1}{n} \cdot \mathbf{1}_n \cdot \mathbf{1}_n' \right] X$$

dengan  $I$  adalah matriks identitas dengan orde  $n$ .

Untuk menentukan variansi dan kovariansi sampel, dapat digunakan matriks  $S$  sebagai berikut:

$$S = \frac{1}{n-1} \cdot T' T$$

Matriks tersebut dapat diperoleh dengan:

$$S = \frac{1}{n-1} (X - \mathbf{1} \cdot \bar{X})' (X - \mathbf{1} \cdot \bar{X})$$

atau

$$S = \frac{1}{n-1} X' \left( I - \frac{1}{N} \cdot \mathbf{1} \cdot \mathbf{1}' \right) \cdot X$$

Koefisien korelasi dapat diperoleh dari matrik korelasi berikut:

$$R = D \quad SD$$

$$\begin{matrix} -1/2 & -1/2 \end{matrix}$$



dengan  $\mathbf{D}^{-1/2}$  menyatakan matriks invers dari  $\mathbf{D}^{1/2}$ , sedangkan matriks  $\mathbf{D}^{1/2}$  didefinisikan sebagai matriks berikut:

$$\mathbf{D}^{1/2} = \begin{pmatrix} \sqrt{s_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sqrt{s_{pp}} \end{pmatrix}$$

Elemen baris ke- $i$  kolom ke- $j$  matriks tersebut bernilai 0 apabila  $i \neq j$  dan bernilai  $\sqrt{s_{ij}}$  apabila  $i = j$ .

Kemudian, dapat ditunjukkan bahwa:

$$\mathbf{D}^{-1/2} = \begin{pmatrix} 1/\sqrt{s_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sqrt{s_{pp}} \end{pmatrix}$$

### Contoh 1:

Berikut ini dicontohkan sebuah model regresi yang tidak memiliki masalah multikolinearitas berdasarkan diagnostik multikolinear dengan nilai VIF, dimisalkan pada contoh ini akan mendeteksi gejala multikolinearitas untuk model regresi pada pengaruh minat belajar ( $X_1$ ) dan motivasi belajar ( $X_2$ ) terhadap hasil belajar siswa ( $Y$ ). Adapun data penelitian yang kita uji dapat dilihat pada tabel berikut ini:

| No | Minat<br>( $X_1$ ) | Motivasi<br>( $X_2$ ) | Hasil<br>belajar<br>( $Y$ ) |
|----|--------------------|-----------------------|-----------------------------|
| 1  | 75                 | 75                    | 80                          |
| 2  | 70                 | 80                    | 75                          |
| 3  | 75                 | 65                    | 75                          |
| 4  | 65                 | 85                    | 90                          |
| 5  | 80                 | 75                    | 85                          |
| 6  | 75                 | 75                    | 90                          |
| 7  | 80                 | 80                    | 95                          |
| 8  | 70                 | 85                    | 80                          |
| 9  | 65                 | 75                    | 90                          |
| 10 | 80                 | 70                    | 75                          |
| 11 | 85                 | 70                    | 75                          |
| 12 | 67                 | 65                    | 80                          |
| 13 | 65                 | 60                    | 85                          |
| 14 | 80                 | 70                    | 80                          |
| 15 | 70                 | 65                    | 90                          |

Uji multikolinear akan dilakukan dengan pemrograman di R, dengan syntax sebagai berikut:

Uji multikolinearitas data minat dan motivasi belajar siswa

```
data.ujimulti=read.delim("clipboard")
```

```
data.ujimulti
```

```
View(data.ujimulti)
```

```
attach(data.ujimulti)
```

```
Modelreg<-
```

```
lm(data.ujimulti$Hasil.Belajar~data.ujimulti$Minat+data.ujimulti$Motivasi)
```

```
summary(Modelreg)
```

```
library(car)
```

vif(Modelreg)

akan diperoleh output:

```
      Minat  Motivasi  Hasil.Belajar
1       75      75      80
2       70      80      75
3       75      65      75
4       65      85      90
5       80      75      85
6       75      75      90
7       80      80      95
8       70      85      80
9       65      75      90
10      80      70      75
11      85      70      75
12      67      65      80
13      65      60      85
14      80      70      80
15      70      65      90

Residuals:
      Min       1Q   Median       3Q      Max
-10.4410  -4.7317  -0.4677   3.8096  12.5033

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    89.8188   27.2618   3.295  0.0064 **
data.ujimulti$Minat -0.2944   0.2764  -1.065  0.3078
data.ujimulti$Motivasi  0.2029   0.2416   0.840  0.4175
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.79 on 12 degrees of freedom
Multiple R-squared:  0.1356,    Adjusted R-squared:  -0.008494
F-statistic: 0.941 on 2 and 12 DF,  p-value: 0.4172

data.ujimulti$Minat data.ujimulti$Motivasi
              1.000537              1.000537
```

Dari output yang diperoleh dapat dinyatakan bahwa:

$$\hat{y} = 89,8188 - 0,2944x_1 + 0,2029x_2$$

Dengan nilai VIF  $(X_1) = 1,000537$  dan VIF  $(X_2) = 1,000537$

Karena VIF  $(X_1)$  dan VIF  $(X_2) < 10$ , maka dapat dinyatakan data tersebut tidak mengalami masalah multikolinear.

**Contoh 2:**

Berikut ini dicontohkan sebuah model regresi yang memiliki masalah multikolinearitas berdasarkan diagnostic multikolinear dengan nilai VIF, dimisalkan pada contoh ini akan mendeteksi gejala multikolinearitas untuk model regresi pada pengaruh minat belajar ( $X_1$ ) dan motivasi belajar ( $X_2$ ) terhadap hasil belajar siswa ( $Y$ ). Adapun data penelitian yang kita uji dapat dilihat pada tabel berikut ini:

| No | Minat<br>( $X_1$ ) | Motivasi<br>( $X_2$ ) | Hasil<br>belajar<br>( $Y$ ) |
|----|--------------------|-----------------------|-----------------------------|
| 1  | 25                 | 30                    | 28                          |
| 2  | 28                 | 32                    | 31                          |
| 3  | 33                 | 35                    | 35                          |
| 4  | 22                 | 24                    | 22                          |
| 5  | 22                 | 26                    | 24                          |
| 6  | 26                 | 30                    | 27                          |
| 7  | 30                 | 34                    | 32                          |
| 8  | 18                 | 20                    | 22                          |
| 9  | 15                 | 18                    | 17                          |
| 10 | 18                 | 22                    | 23                          |
| 11 | 22                 | 24                    | 25                          |
| 12 | 28                 | 30                    | 29                          |
| 13 | 33                 | 35                    | 35                          |
| 14 | 8                  | 10                    | 8                           |
| 15 | 25                 | 28                    | 27                          |
| 16 | 35                 | 38                    | 37                          |
| 17 | 20                 | 24                    | 22                          |
| 18 | 15                 | 18                    | 16                          |
| 19 | 30                 | 34                    | 31                          |

Uji multikolinear akan dilakukan dengan pemrograman di R, dengan syntax sebagai berikut:

```
Uji multikolinearitas data minat dan motivasi belajar siswa  
data.ujimulti=read.delim("clipboard")  
data.ujimulti
```

```
View(data.ujimulti)  
attach(data.ujimulti)
```

```
Modelreg<-  
lm(data.ujimulti$Hasil.Belajar~data.ujimulti$Minat+data.ujimul  
ti$Motivasi)  
summary(Modelreg)  
library(car)  
vif(Modelreg)
```

akan diperoleh output:

|    | Minat | Motivasi | Hasil.Belajar |
|----|-------|----------|---------------|
| 1  | 25    | 30       | 28            |
| 2  | 28    | 32       | 31            |
| 3  | 33    | 35       | 35            |
| 4  | 22    | 24       | 22            |
| 5  | 22    | 26       | 24            |
| 6  | 26    | 30       | 27            |
| 7  | 30    | 34       | 32            |
| 8  | 18    | 20       | 22            |
| 9  | 15    | 18       | 17            |
| 10 | 18    | 22       | 23            |
| 11 | 22    | 24       | 25            |
| 12 | 28    | 30       | 29            |
| 13 | 33    | 35       | 35            |
| 14 | 8     | 10       | 8             |
| 15 | 25    | 28       | 27            |
| 16 | 35    | 38       | 37            |
| 17 | 20    | 24       | 22            |
| 18 | 15    | 18       | 16            |
| 19 | 30    | 34       | 31            |

Residuals:

|  | Min      | 1Q       | Median  | 3Q      | Max     |
|--|----------|----------|---------|---------|---------|
|  | -1.62501 | -0.79192 | 0.01252 | 0.35687 | 2.71145 |

Coefficients:

|                         | Estimate | Std. Error | t value | Pr(> t ) |
|-------------------------|----------|------------|---------|----------|
| (Intercept)             | 0.9083   | 1.3462     | 0.675   | 0.5095   |
| data.ujimulti\$Minat    | 0.6662   | 0.3105     | 2.146   | 0.0476 * |
| data.ujimulti\$Motivasi | 0.3359   | 0.3035     | 1.107   | 0.2848   |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.274 on 16 degrees of freedom

Multiple R-squared: 0.9727, Adjusted R-squared: 0.9693

F-statistic: 284.8 on 2 and 16 DF, p-value: 3.107e-13

|                      |                         |
|----------------------|-------------------------|
| data.ujimulti\$Minat | data.ujimulti\$Motivasi |
| 54.07441             | 54.07441                |

Dari output yang diperoleh dapat dinyatakan bahwa:

$$\hat{y} = 0,9083 - 0,6662x_1 + 0,3359x_2$$

Dengan nilai VIF  $(X_1) = 54,07441$  dan VIF  $(X_2) = 54,07441$

## DAFTAR PUSTAKA

- Alita, D., Putra, A. D. and Darwis, D. (2021) 'Analysis of classic assumption test and multiple linear regression coefficient test for employee structural office recommendation', *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(3), p. 295. doi: 10.22146/ijccs.65586.
- Balqis, N. A., Astutik, S. and Solimun, S. (2022) 'Principal Component Regression Modelling with Variational Bayesian Approach to Overcome Multicollinearity at Various Levels of Missing Data Proportion', *JTAM (Jurnal Teori dan Aplikasi Matematika)*, 6(4), p. 1013. doi: 10.31764/jtam.v6i4.10223.
- Chang, Y. C. and Mastrangelo, C. (2011) 'Addressing multicollinearity in semiconductor manufacturing', *Quality and Reliability Engineering International*, 27(6), pp. 843–854. doi: 10.1002/qre.1173.
- Dias, S. and Brito, P. (2022) 'Regression Analysis with the Distribution and Symmetric Distribution Model', *Analysis of Distributional Data*, pp. 295–318. doi: 10.1201/9781315370545-13.
- Kalatzis, A. E. G., Bassetto, C. F. and Azzoni, C. R. (2011) 'Multicollinearity and financial constraint in investment decisions: A Bayesian generalized ridge regression', *Journal of Applied Statistics*, 38(2), pp. 287–299. doi: 10.1080/02664760903406462.
- Montgomery, D. C., Peck, E. A. and Vining, G. G. (2012) *Introduction to Linear Regression Analysis*. John Wiley & Sons.
- Pesaran, M. H. and Smith, R. P. (2019) 'A Bayesian analysis of linear regression models with highly collinear regressors', *Econometrics and Statistics*, 11, pp. 1–21. doi: 10.1016/j.ecosta.2018.10.001.

- Shrestha, N. (2020) 'Detecting Multicollinearity in Regression Analysis', *American Journal of Applied Mathematics and Statistics*, 8(2), pp. 39–42. doi: 10.12691/ajams-8-2-1.
- Vigneau, J. and Mahévas, S. (2007) 'Detecting sampling outliers and sampling heterogeneity when catch-at-length is estimated using the ratio estimator', *ICES Journal of Marine Science*, 64(5), pp. 1028–1032. doi: 10.1093/icesjms/fsm077.



# **BAB 9**

## **PEMILIHAN MODEL REGRESI TERBAIK**

**Oleh Bambang Suhartawan**

### **9.1 Pengertian Analisis Regresi**

Analisis regresi adalah metode statistik yang digunakan untuk menentukan hubungan antara satu atau lebih variabel independen (dikenal sebagai variabel prediktor) dan satu variabel dependen (variabel yang ingin diprediksi). Tujuan utamanya adalah untuk memahami dan menjelaskan hubungan antara variabel-variabel tersebut serta memprediksi nilai variabel dependen berdasarkan nilai variabel independen.

Dalam konteks penelitian, analisis regresi digunakan untuk mengukur kekuatan dan arah hubungan antara variabel-variabel yang diteliti. Misalnya, dalam penelitian ekonomi, analisis regresi dapat digunakan untuk menentukan sejauh mana pengeluaran konsumen dipengaruhi oleh pendapatan mereka. Dalam penelitian medis, analisis regresi dapat digunakan untuk menentukan hubungan antara asupan makanan dan risiko terkena penyakit.

Ada dua jenis analisis regresi utama: regresi linear dan regresi non-linear. Regresi linear adalah ketika hubungan antara variabel independen dan dependen dapat dijelaskan dengan garis lurus, sedangkan regresi non-linear terjadi ketika hubungan tersebut tidak dapat dijelaskan dengan garis lurus.

Analisis regresi biasanya melibatkan langkah-langkah seperti; Pengumpulan data: Mengumpulkan data tentang variabel independen dan dependen dari sampel yang relevan; Pemilihan model regresi: Memilih jenis model regresi yang sesuai dengan data dan tujuan penelitian; Estimasi parameter: Menghitung parameter model regresi untuk menentukan hubungan antara variabel-variabel; Evaluasi model: Menggunakan berbagai metrik evaluasi untuk menentukan seberapa baik model regresi memprediksi variabel dependen; Interpretasi hasil: Menafsirkan hasil analisis untuk mendapatkan pemahaman yang lebih baik tentang hubungan antara variabel-variabel yang diteliti.

Analisis regresi adalah salah satu alat yang paling umum digunakan dalam penelitian ilmiah karena fleksibilitasnya dalam memodelkan hubungan antara variabel-variabel yang kompleks.

## **9.2 Jenis-jenis Model Regresi dalam Penelitian**

Dalam penelitian, terdapat beberapa jenis model regresi yang dapat digunakan tergantung pada sifat data dan hubungan antara variabel-variabel yang diteliti. Berikut adalah beberapa jenis model regresi yang umum digunakan:

### **1. Regresi Linear Sederhana:**

Ini adalah jenis regresi di mana terdapat satu variabel independen (prediktor) dan satu variabel dependen. Model ini mencoba untuk menemukan garis lurus terbaik yang mendekati hubungan antara variabel prediktor dan variabel dependen.

### **2. Regresi Linear Berganda:**

Regresi linear berganda melibatkan lebih dari satu variabel independen untuk memprediksi satu variabel dependen. Model ini berusaha menemukan hubungan linier terbaik antara variabel-variabel prediktor dan variabel dependen.

### **3. Regresi Logistik:**

Regresi logistik digunakan ketika variabel dependen bersifat biner (yaitu memiliki dua kategori atau nilai), dan variabel independen dapat berupa kategori atau kontinu. Regresi logistik memberikan probabilitas bahwa variabel dependen termasuk dalam salah satu kategori.

### **4. Regresi Polinomial:**

Regresi polinomial digunakan ketika hubungan antara variabel independen dan dependen tidak dapat dijelaskan dengan garis lurus. Model ini menggunakan polinomial untuk memodelkan hubungan yang lebih kompleks.

### **5. Regresi Non-Linear:**

Regresi non-linear menggambarkan hubungan antara variabel independen dan dependen yang tidak linier, artinya hubungan tersebut tidak dapat dijelaskan dengan menggunakan fungsi linear.

### **6. Regresi Ridge dan Lasso:**

Regresi ridge dan lasso adalah metode regresi yang digunakan untuk mengatasi masalah multikolinieritas (korelasi tinggi antara variabel independen) dan memperbaiki keakuratan prediksi dalam regresi linear berganda dengan menerapkan penalti terhadap koefisien regresi.

### **7. Regresi Time Series:**

Regresi time series digunakan ketika variabel dependen adalah data berurutan dalam waktu dan analisis dilakukan untuk memahami pola atau tren dalam data seiring waktu.

### **8. Regresi Tobit:**

Regresi Tobit digunakan ketika data memiliki batasan atas atau bawah, dan regresi linear biasa tidak sesuai karena tidak memperhitungkan batasan tersebut.

## **9. Regresi Kuantil:**

Regresi kuantil digunakan untuk memprediksi kuantil tertentu dari distribusi variabel dependen, bukan hanya nilai rata-rata.

Pemilihan jenis model regresi yang tepat tergantung pada tujuan penelitian, sifat data yang digunakan, dan asumsi yang diinginkan tentang hubungan antara variabel-variabel yang diteliti.

### **9.3 Pemilihan Model Regresi Terbaik**

Dalam memilih model regresi terbaik dalam analisis regresi, terdapat beberapa langkah yang dapat diambil. Berikut adalah beberapa langkah umum yang dapat membantu Anda dalam memilih model regresi terbaik:

#### **1. Pilih Variabel Prediktor yang Sesuai:**

Pastikan Anda telah memilih variabel prediktor yang relevan dan memiliki hubungan yang kuat dengan variabel respon. Ini dapat dilakukan melalui analisis korelasi dan pengetahuan domain.

Pemilihan variabel prediktor yang sesuai adalah langkah kunci dalam analisis regresi karena variabel prediktor yang dipilih akan mempengaruhi kualitas dan keakuratan model regresi Anda. Berikut ini uraian detail tentang cara memilih variabel prediktor yang sesuai:

##### **a. Analisis Korelasi:**

Langkah pertama adalah melakukan analisis korelasi antara setiap variabel prediktor dan variabel respon. Korelasi adalah ukuran statistik yang mengukur sejauh mana dua variabel berkaitan satu sama lain. Anda dapat menggunakan koefisien korelasi Pearson atau koefisien korelasi lainnya, tergantung pada jenis data yang Anda miliki.

- Variabel prediktor yang memiliki korelasi yang tinggi dan signifikan dengan variabel respon cenderung menjadi kandidat yang baik untuk dimasukkan ke dalam model regresi.
- Namun, penting untuk diingat bahwa korelasi tidak selalu menunjukkan hubungan sebab-akibat. Sebuah hubungan korelasi yang kuat tidak selalu berarti bahwa satu variabel menyebabkan perubahan dalam variabel lainnya.

#### **b. Pengetahuan Domain:**

Selain melalui analisis korelasi, pengetahuan domain atau pengetahuan tentang subjek yang sedang diteliti juga sangat penting. Pengetahuan tentang domain dapat membantu Anda dalam memilih variabel prediktor yang relevan dan memahami hubungan sebab-akibat antara variabel-variabel tersebut.

- Misalnya, jika Anda melakukan studi tentang faktor-faktor yang memengaruhi kinerja akademik mahasiswa, pengetahuan tentang pendidikan, psikologi, atau faktor-faktor sosial dapat membantu Anda mengidentifikasi variabel prediktor yang penting, seperti tingkat pendidikan orangtua, waktu belajar, atau dukungan sosial.
- Pengetahuan domain juga membantu Anda dalam menghindari memasukkan variabel yang tidak relevan atau tidak berarti dalam model Anda.

#### **c. Analisis Univariate dan Multivariate:**

Selain korelasi dengan variabel respon, Anda juga dapat melakukan analisis univariate (analisis untuk setiap variabel secara individual) dan analisis multivariate (analisis yang melibatkan beberapa variabel secara bersamaan).

- Analisis univariate dapat membantu Anda mengevaluasi distribusi dan karakteristik setiap variabel prediktor secara terpisah.
- Analisis multivariate, seperti analisis regresi berganda, memungkinkan Anda untuk memeriksa hubungan antara beberapa variabel prediktor dan variabel respon secara bersamaan, memungkinkan untuk menilai pengaruh relatif dari masing-masing variabel.

Dengan menggabungkan analisis korelasi, pengetahuan domain, analisis univariate, dan analisis multivariate, Anda dapat membuat keputusan yang lebih baik dalam memilih variabel prediktor yang sesuai untuk dimasukkan ke dalam model regresi Anda. Hal ini akan membantu memastikan bahwa model regresi Anda memiliki kemampuan yang baik untuk menjelaskan variasi dalam variabel respon dan memberikan hasil yang berguna dan relevan.

## **2. Uji Asumsi Regresi:**

Sebelum memilih model, pastikan untuk menguji asumsi-asumsi dasar dari regresi, seperti keberdistribusian normalitas, homoskedastisitas, dan independensi residual.

Uji asumsi regresi penting dilakukan sebelum memilih model regresi akhir, karena dapat memberikan wawasan tentang sejauh mana model regresi Anda memenuhi asumsi-asumsi dasar yang diperlukan untuk validitas interpretasi hasil. Berikut ini uraian detail tentang asumsi-asumsi dasar dari regresi dan uji yang dapat digunakan untuk menguji asumsi tersebut:

### **a. Keberdistribusian Normalitas:**

Salah satu asumsi dasar dari regresi adalah bahwa residual (selisih antara nilai yang diamati dan nilai yang diprediksi oleh model) harus memiliki distribusi normal. Distribusi normalitas residual menunjukkan bahwa kesalahan dalam prediksi model secara acak tersebar di sekitar nilai nol.

- **Uji Shapiro-Wilk:** Uji ini adalah uji statistik yang digunakan untuk menguji apakah sebuah sampel data memiliki distribusi normal. Hipotesis nolnya adalah bahwa data memiliki distribusi normal.
- **Uji Kolmogorov-Smirnov:** Uji ini juga digunakan untuk menguji normalitas data. Ini membandingkan distribusi kumulatif empiris dari data dengan distribusi yang diharapkan (misalnya, distribusi normal).
- **Histogram dan Q-Q Plot:** Visualisasi seperti histogram atau quantile-quantile plot (Q-Q plot) dapat membantu dalam mengevaluasi seberapa dekat distribusi residual dengan distribusi normal. Jika distribusi residual cukup dekat dengan distribusi normal, maka asumsi normalitas dapat dianggap terpenuhi.

#### **b. Homoskedastisitas:**

Asumsi homoskedastisitas menyiratkan bahwa variabilitas residual adalah konstan di semua level variabel prediktor. Artinya, tidak ada pola sistematis dalam variabilitas residual seiring dengan nilai prediktor.

- **Uji Breusch-Pagan / White:** Uji ini digunakan untuk menguji homoskedastisitas dalam regresi. Hipotesis nolnya adalah bahwa variabilitas residual konstan (homoskedastisitas).
- **Uji Grafik:** Menggunakan grafik seperti scatterplot dari residu versus nilai prediktor dapat membantu dalam memeriksa apakah ada pola dalam variabilitas residual. Pola cembung atau lekukan dalam plot ini dapat menunjukkan ketidakhomoskedastisitas.

#### **c. Independensi Residual:**

Asumsi independensi residual menyiratkan bahwa tidak ada korelasi antara residual dalam model regresi. Dalam konteks regresi linear, ini berarti tidak ada pola yang

terlihat dalam plot residual versus nilai prediktor atau dalam plot residual terhadap waktu (jika data berurutan).

- **Uji Durbin-Watson:** Uji ini digunakan untuk menguji apakah ada otokorelasi (autokorelasi) dalam residu regresi. Nilai uji berada dalam kisaran 0 hingga 4, dengan nilai sekitar 2 menunjukkan tidak adanya autokorelasi.
- **Uji Grafik:** Plot residual versus nilai prediktor atau plot residual versus waktu (jika data berurutan) dapat membantu dalam mendeteksi pola autokorelasi dalam residual.

Dengan melakukan uji-asumsi regresi ini, Anda dapat mengevaluasi sejauh mana model regresi Anda memenuhi asumsi-asumsi dasar tersebut. Jika asumsi tidak terpenuhi, Anda perlu mempertimbangkan transformasi data, penggunaan model alternatif, atau menggunakan teknik analisis lainnya yang lebih sesuai dengan data Anda.

### 3. Pilih Model:

Ada beberapa jenis model regresi yang dapat dipertimbangkan, seperti regresi linear sederhana, regresi linear berganda, regresi non-linear, regresi logistik, dll. Pemilihan model akan tergantung pada sifat data dan tujuan analisis Anda.

Memilih model regresi yang sesuai adalah langkah penting dalam analisis regresi. Berikut ini uraian detail tentang beberapa jenis model regresi yang dapat dipertimbangkan, serta faktor-faktor yang perlu diperhatikan dalam pemilihan model:

#### a. Regresi Linear Sederhana:

Model regresi linear sederhana merupakan model yang paling sederhana, dengan hanya satu variabel prediktor yang digunakan untuk memprediksi variabel respon. Persamaan regresi linear sederhana adalah:  $Y = \beta_0 + \beta_1 X + \varepsilon$ . Di mana  $Y$  adalah variabel respon,  $X$  adalah variabel



prediktor,  $\beta_0$  adalah intercept,  $\beta_1$  adalah koefisien regresi, dan  $\varepsilon$  adalah kesalahan acak.

**b. Regresi Linear Berganda:**

Model regresi linear berganda melibatkan dua atau lebih variabel prediktor untuk memprediksi variabel respon. Persamaan regresi linear berganda adalah:  
$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Di mana  $Y$  adalah variabel respon,  $X_1, X_2, \dots, X_p$  adalah variabel prediktor,  $\beta_0$  adalah intercept,  $\beta_1, \beta_2, \dots, \beta_p$  adalah koefisien regresi untuk masing-masing variabel prediktor, dan  $\varepsilon$  adalah kesalahan acak.

**c. Regresi Non-linear:**

Dalam beberapa kasus, hubungan antara variabel prediktor dan variabel respon mungkin tidak linear. Dalam hal ini, regresi non-linear dapat digunakan untuk memodelkan hubungan yang lebih kompleks antara variabel prediktor dan variabel respon. Contoh regresi non-linear termasuk regresi kuadratik, regresi eksponensial, dan regresi logistik.

**d. Regresi Logistik:**

Regresi logistik digunakan ketika variabel respon adalah biner atau kategorikal. Ini adalah model yang umum digunakan dalam analisis klasifikasi dan prediksi probabilitas. Persamaan regresi logistik adalah:  
$$\text{logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$
  
Di mana  $p$  adalah probabilitas kejadian,  $X_1, X_2, \dots, X_p$  adalah variabel prediktor,  $\beta_0$  adalah intercept,  $\beta_1, \beta_2, \dots, \beta_p$  adalah koefisien regresi untuk masing-masing variabel prediktor, dan  $\text{logit}(p)$  adalah log-odds dari kejadian.

Faktor-faktor yang perlu dipertimbangkan dalam pemilihan model meliputi:

- **Jenis Data:** Apakah data Anda bersifat kontinu, kategorikal, atau biner? Jenis data ini akan mempengaruhi pilihan model yang paling sesuai.
- **Kepentingan Interpretasi:** Apakah Anda lebih tertarik pada interpretasi koefisien dan efek prediktor individual? Jika ya, regresi linear atau regresi logistik mungkin lebih cocok.
- **Kompleksitas Hubungan:** Apakah hubungan antara variabel prediktor dan variabel respon linear atau non-linear? Jika hubungan non-linear, regresi non-linear mungkin diperlukan.
- **Tujuan Analisis:** Apakah tujuan Anda adalah untuk prediksi atau penjelasan? Model yang berbeda mungkin lebih baik sesuai dengan tujuan analisis Anda.

Pemilihan model yang tepat membutuhkan pemahaman yang baik tentang data Anda, asumsi yang terlibat dalam model yang dipilih, dan tujuan analisis Anda. Evaluasi berbagai model dan memilih yang paling sesuai dengan kebutuhan Anda adalah kunci untuk analisis regresi yang efektif.

#### 4. Penyesuaian Model:

Sesuaikan model regresi Anda dengan menambahkan atau menghapus variabel prediktor, menggunakan transformasi variabel, atau menambahkan polinomial untuk meningkatkan kesesuaian model dengan data.

Penyesuaian model regresi adalah proses penting untuk meningkatkan kualitas model dan memastikan bahwa model tersebut sesuai dengan data dengan baik. Berikut ini uraian detail tentang beberapa teknik penyesuaian model regresi yang dapat digunakan:

#### **a. Menambah atau Menghapus Variabel Prediktor:**

- **Menambah Variabel Prediktor:** Jika Anda menduga bahwa terdapat variabel prediktor tambahan yang dapat meningkatkan kemampuan model dalam menjelaskan variasi dalam variabel respon, Anda dapat mencoba menambahkan variabel tersebut ke dalam model. Pastikan variabel yang ditambahkan relevan dengan masalah yang sedang Anda teliti.
- **Menghapus Variabel Prediktor:** Sebaliknya, jika terdapat variabel prediktor yang tidak relevan atau memiliki pengaruh yang kecil terhadap variabel respon, Anda dapat mempertimbangkan untuk menghapus variabel tersebut dari model. Ini dapat membantu mengurangi kompleksitas model dan mencegah overfitting.

#### **b. Transformasi Variabel:**

- **Transformasi Logaritmik:** Dalam beberapa kasus, variabel prediktor atau variabel respon mungkin memiliki distribusi yang condong atau tidak normal. Dengan melakukan transformasi logaritmik terhadap variabel tersebut, Anda dapat membantu menormalkan distribusi data dan meningkatkan kecocokan model.
- **Transformasi Kuadratik atau Pangkat Lain:** Jika hubungan antara variabel prediktor dan variabel respon tidak linear, Anda dapat mencoba menggunakan transformasi kuadratik atau pangkat lain terhadap variabel prediktor untuk menangkap hubungan non-linear tersebut.

#### **c. Menambahkan Polinomial:**

- **Polinomial Tingkat Tinggi:** Jika hubungan antara variabel prediktor dan variabel respon tidak dapat dijelaskan dengan baik oleh model linear, Anda dapat mencoba menambahkan variabel prediktor yang merupakan polinomial dari variabel asli (misalnya,  $X^2$ ,

X3, dst.). Ini memungkinkan model untuk menangkap hubungan yang lebih kompleks antara variabel prediktor dan variabel respon.

#### **d. Interaksi Antara Variabel:**

- **Interaksi:** Terkadang, efek dari satu variabel prediktor dapat bergantung pada nilai dari variabel prediktor lainnya. Dalam hal ini, Anda dapat menambahkan istilah interaksi antara variabel prediktor untuk menggambarkan hubungan yang lebih kompleks antara variabel tersebut.

#### **e. Seleksi Variabel Otomatis:**

- **Metode Seleksi Variabel:** Terdapat berbagai teknik yang dapat digunakan untuk melakukan seleksi variabel otomatis, seperti regresi berganda secara bertahap (forward selection, backward elimination, atau stepwise regression), regularisasi (misalnya, Lasso atau Ridge regression), dan seleksi variabel berbasis informasi (misalnya, Akaike Information Criterion atau Bayesian Information Criterion).

Setiap teknik penyesuaian model memiliki kelebihan dan kelemahan tertentu, dan pemilihan teknik yang tepat akan tergantung pada sifat data, tujuan analisis, dan asumsi model. Penting untuk melakukan evaluasi yang cermat terhadap setiap penyesuaian yang dilakukan untuk memastikan bahwa model yang dihasilkan sesuai dengan data dengan baik dan relevan dengan tujuan analisis Anda.

### **5. Uji Model:**

Gunakan teknik validasi silang atau pengujian lainnya untuk mengevaluasi kinerja model. Ini membantu memastikan bahwa model Anda tidak terlalu disesuaikan atau terlalu sederhana.

Uji model adalah langkah penting dalam evaluasi kinerja model regresi Anda. Ini membantu memastikan bahwa model yang dikembangkan dapat umumnya diterapkan

pada data baru dengan baik dan tidak terlalu disesuaikan atau terlalu sederhana. Berikut ini uraian detail tentang beberapa teknik uji model yang umum digunakan:

**a. Validasi Silang (Cross-Validation):**

- **K-Fold Cross-Validation:** Metode ini melibatkan pembagian data menjadi  $k$  subset yang seimbang. Setiap subset kemudian digunakan sebagai data uji satu kali sementara sisanya digunakan untuk melatih model. Proses ini diulangi  $k$  kali dengan setiap subset bergantian sebagai data uji, dan hasilnya diambil rata-rata untuk memberikan perkiraan kinerja model.
- **Leave-One-Out Cross-Validation (LOOCV):** Metode ini adalah variasi dari validasi silang di mana satu observasi digunakan sebagai data uji sementara yang lain digunakan sebagai data pelatihan. Proses ini diulangi untuk setiap observasi sehingga setiap observasi digunakan sebagai data uji satu kali.
- **Metode Cross-Validation yang Lain:** Selain K-Fold dan LOOCV, terdapat juga metode cross-validation lainnya seperti Stratified K-Fold Cross-Validation dan Time Series Cross-Validation (untuk data berurutan).

**b. Pengujian Terpisah (Holdout Testing):**

- Metode ini melibatkan pembagian data menjadi dua bagian, yaitu data pelatihan dan data pengujian. Data pelatihan digunakan untuk melatih model, sedangkan data pengujian digunakan untuk menguji kinerja model. Kinerja model kemudian dievaluasi berdasarkan data pengujian.
- Biasanya, sekitar 70-80% data digunakan sebagai data pelatihan dan sisanya sebagai data pengujian. Namun, pembagian ini dapat bervariasi tergantung pada ukuran data dan kompleksitas model.

### **c. Metrik Evaluasi Model:**

- Setelah melakukan validasi silang atau pengujian terpisah, Anda perlu menggunakan metrik evaluasi model untuk menilai kinerja model regresi Anda. Beberapa metrik evaluasi yang umum digunakan termasuk:
  - R-squared ( $R^2$ ): Mengukur seberapa baik model Anda cocok dengan data.
  - Root Mean Squared Error (RMSE): Mengukur seberapa baik model Anda memprediksi nilai sebenarnya dengan menghitung akar dari rata-rata kuadrat selisih antara nilai prediksi dan nilai sebenarnya.
  - Mean Absolute Error (MAE): Mengukur rata-rata dari selisih absolut antara nilai prediksi dan nilai sebenarnya.
  - Akaike Information Criterion (AIC) atau Bayesian Information Criterion (BIC): Mengukur kualitas relatif dari model regresi, dengan mempertimbangkan keakuratan dan kompleksitas model.

### **d. Analisis Residual:**

- Setelah menguji model, penting untuk memeriksa residual model untuk memastikan bahwa tidak ada pola yang tersisa dalam kesalahan prediksi. Plot residual vs nilai prediktor atau residual vs nilai sebenarnya dapat membantu Anda mengidentifikasi pola ini.

Dengan menggunakan teknik uji model ini, Anda dapat memastikan bahwa model regresi yang dikembangkan memiliki kinerja yang baik dan generalisasi yang kuat ke data baru. Hal ini membantu dalam memastikan bahwa model Anda dapat diandalkan dan berguna dalam

menganalisis dan memprediksi data di luar sampel pelatihan.

## **6. Evaluasi Metrik:**

Gunakan metrik evaluasi yang sesuai untuk menilai kinerja model Anda, seperti R-squared, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), atau metrik khusus domain jika ada.

Evaluasi metrik adalah langkah penting dalam menilai kinerja model regresi Anda. Berikut ini uraian detail tentang beberapa metrik evaluasi yang umum digunakan untuk mengevaluasi model regresi, beserta penjelasan tentang cara menginterpretasi hasilnya:

### **a. R-squared ( $R^2$ ):**

- R-squared adalah metrik yang mengukur seberapa baik model Anda cocok dengan data. Nilai R-squared berkisar dari 0 hingga 1, di mana semakin dekat ke 1, semakin baik modelnya.
- Interpretasi: Sebagai persentase variasi dalam variabel respon yang dapat dijelaskan oleh model Anda. Misalnya, R-squared sebesar 0,80 berarti bahwa 80% variasi dalam variabel respon dapat dijelaskan oleh model Anda, sementara sisanya dijelaskan oleh faktor-faktor lain yang tidak dimasukkan ke dalam model.

### **b. Root Mean Squared Error (RMSE):**

- RMSE mengukur rata-rata dari akar kuadrat dari selisih antara nilai prediksi dan nilai sebenarnya. Ini memberikan gagasan tentang seberapa besar kesalahan prediksi model dalam unit yang sama dengan variabel respon.
- Interpretasi: Semakin rendah nilai RMSE, semakin baik modelnya. RMSE yang lebih rendah menunjukkan bahwa model memiliki tingkat kesalahan yang lebih kecil dalam memprediksi nilai sebenarnya.

**c. Mean Absolute Error (MAE):**

- MAE mengukur rata-rata dari selisih absolut antara nilai prediksi dan nilai sebenarnya. Ini memberikan gagasan tentang seberapa besar kesalahan rata-rata dalam prediksi model.
- Interpretasi: Seperti RMSE, semakin rendah nilai MAE, semakin baik modelnya. MAE juga memberikan ukuran kesalahan prediksi dalam unit yang sama dengan variabel respon.

**d. Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC):**

- AIC dan BIC digunakan untuk membandingkan kualitas relatif dari beberapa model. Kedua metrik ini mempertimbangkan akurasi model dan kompleksitasnya.
- Interpretasi: Model dengan nilai AIC atau BIC yang lebih rendah dianggap lebih baik. Namun, perlu diingat bahwa kedua metrik ini tidak memiliki interpretasi yang langsung dalam unit yang sama dengan variabel respon.

**e. Metrik Khusus Domain:**

- Terkadang, dalam konteks domain tertentu, metrik evaluasi khusus dapat dibuat untuk menilai kinerja model. Misalnya, dalam analisis keuangan, metrik seperti Return on Investment (ROI) atau net present value (NPV) mungkin lebih relevan dalam mengevaluasi kinerja model.

Pemilihan metrik evaluasi yang tepat tergantung pada tujuan analisis Anda dan konteks spesifik masalah yang Anda hadapi. Dalam banyak kasus, penggunaan beberapa metrik evaluasi bersama-sama dapat memberikan gambaran yang lebih lengkap tentang kinerja model Anda.



## 7. Banding Variabel:

Dalam beberapa kasus, Anda mungkin ingin mempertimbangkan penggunaan variabel terikat dalam regresi, seperti menggunakan regresi logistik untuk variabel biner atau kategorikal.

Pertimbangan variabel terikat, terutama dalam konteks regresi logistik untuk variabel biner atau kategorikal, merupakan langkah penting dalam analisis regresi. Berikut ini uraian detail tentang konsep banding variabel dan implementasinya dalam regresi:

### a. Variabel Terikat atau Variabel Respon:

Variabel terikat, juga dikenal sebagai variabel respon, adalah variabel yang ingin diprediksi oleh model regresi. Dalam konteks regresi logistik, variabel terikat seringkali adalah variabel biner (dua kategori) atau variabel kategorikal dengan lebih dari dua kategori.

### b. Regresi Logistik untuk Variabel Biner:

- Dalam regresi logistik untuk variabel biner, variabel terikat hanya memiliki dua kategori yang mungkin, misalnya "Ya" atau "Tidak", "Sukses" atau "Gagal", "Positif" atau "Negatif".
- Model regresi logistik menggunakan fungsi logit untuk memprediksi probabilitas kejadian salah satu kategori. Fungsi logit adalah logaritma dari odds kejadian sukses terhadap kegagalan.
- Persamaan regresi logistik adalah:  $\text{logit}(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$ . Di mana  $p$  adalah probabilitas kejadian,  $X_1, X_2, \dots, X_p$  adalah variabel prediktor,  $\beta_0$  adalah intercept,  $\beta_1, \beta_2, \dots, \beta_p$  adalah koefisien regresi untuk masing-masing variabel prediktor, dan  $\text{logit}(p)$  adalah log-odds dari kejadian.

### **c. Regresi Logistik untuk Variabel Kategorikal:**

- Jika variabel terikat memiliki lebih dari dua kategori, regresi logistik dapat digunakan dengan modifikasi untuk menangani situasi ini.
- Misalnya, regresi logistik multinomial atau regresi logistik ordinal dapat digunakan tergantung pada sifat variabel terikat.

### **d. Interpretasi Koefisien:**

- Dalam regresi logistik, interpretasi koefisien berbeda dengan regresi linier. Koefisien regresi dalam regresi logistik menunjukkan perubahan dalam log-odds kejadian untuk setiap satu unit perubahan dalam variabel prediktor.
- Koefisien positif menunjukkan bahwa peningkatan nilai variabel prediktor berhubungan dengan peningkatan log-odds kejadian, sedangkan koefisien negatif menunjukkan hubungan sebaliknya.

### **e. Evaluasi Model:**

- Evaluasi kinerja model regresi logistik dapat dilakukan menggunakan berbagai metrik, seperti akurasi, sensitivitas, spesifisitas, dan area di bawah kurva ROC (AUC-ROC).
- Akurasi mengukur seberapa akurat model dalam memprediksi kategori yang benar.
- Sensitivitas mengukur kemampuan model dalam mendeteksi kategori positif dengan benar, sedangkan spesifisitas mengukur kemampuan model dalam mengidentifikasi kategori negatif dengan benar.
- AUC-ROC mengukur kemampuan model dalam membedakan antara dua kategori dengan mengukur area di bawah kurva ROC. Nilai AUC-ROC yang lebih tinggi menunjukkan kinerja model yang lebih baik.

Pemilihan jenis regresi dan teknik evaluasi yang tepat tergantung pada sifat data dan tujuan analisis Anda. Dengan mempertimbangkan variabel terikat dengan baik dan menerapkan metode regresi yang sesuai, Anda dapat menghasilkan model regresi yang lebih akurat dan informatif.

## **8. Validasi Model:**

Setelah memilih model, validasi dengan menggunakan data baru atau melalui teknik validasi silang untuk memastikan kinerja yang konsisten pada data yang belum terlihat sebelumnya.

Validasi model adalah langkah penting dalam pengembangan model regresi untuk memastikan bahwa model yang dikembangkan mampu melakukan prediksi dengan baik pada data yang belum terlihat sebelumnya. Berikut ini uraian detail tentang proses validasi model:

### **a. Pembagian Data:**

Langkah pertama dalam validasi model adalah membagi data menjadi set pelatihan dan set pengujian. Set pelatihan digunakan untuk melatih model, sedangkan set pengujian digunakan untuk menguji kinerja model yang dilatih. Pembagian data ini harus dilakukan secara acak untuk memastikan representasi yang seimbang dari semua kategori data.

### **b. Validasi Silang (Cross-Validation):**

- Metode ini melibatkan pembagian data menjadi k subset yang seimbang. Setiap subset kemudian digunakan sebagai data uji satu kali sementara sisanya digunakan untuk melatih model. Proses ini diulangi k kali dengan setiap subset bergantian sebagai data uji, dan hasilnya diambil rata-rata untuk memberikan perkiraan kinerja model.
- Validasi silang membantu mengurangi bias yang mungkin muncul karena pembagian acak data dan

memberikan perkiraan yang lebih stabil tentang kinerja model pada data baru.

**c. Pengujian Terpisah (Holdout Testing):**

- Metode ini melibatkan pembagian data menjadi dua bagian, yaitu data pelatihan dan data pengujian. Data pelatihan digunakan untuk melatih model, sedangkan data pengujian digunakan untuk menguji kinerja model. Kinerja model kemudian dievaluasi berdasarkan data pengujian.
- Holdout testing sering digunakan untuk evaluasi model regresi pada dataset yang cukup besar.

**d. Evaluasi Kinerja Model:**

- Setelah validasi model dilakukan, kinerja model dievaluasi menggunakan metrik evaluasi yang relevan, seperti R-squared, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), atau metrik khusus domain jika ada.
- Metrik evaluasi memberikan wawasan tentang seberapa baik model melakukan prediksi pada data pengujian yang belum terlihat sebelumnya. Semakin tinggi nilai metrik evaluasi, semakin baik kinerja model.

**e. Analisis Residual:**

- Setelah validasi model, penting untuk memeriksa residual model pada data pengujian untuk memastikan bahwa tidak ada pola yang tersisa dalam kesalahan prediksi. Plot residual vs nilai prediktor atau residual vs nilai sebenarnya dapat membantu Anda mengidentifikasi pola ini.

Dengan melakukan validasi model yang cermat, Anda dapat memastikan bahwa model regresi yang dikembangkan memiliki kinerja yang konsisten dan dapat diandalkan pada data baru yang belum terlihat sebelumnya. Hal ini

memungkinkan Anda untuk mengambil keputusan yang lebih tepat berdasarkan hasil prediksi model.

## **9. Interpretasi Model:**

Terakhir, pastikan untuk menginterpretasikan koefisien model dengan hati-hati untuk memahami kontribusi masing-masing variabel prediktor terhadap variabel respon.

Interpretasi model adalah langkah penting dalam analisis regresi yang memungkinkan pemahaman yang lebih dalam tentang bagaimana variabel prediktor berkontribusi terhadap variabel respon. Berikut ini uraian detail tentang interpretasi koefisien model:

### **a. Koefisien Regresi:**

- Setelah melatih model regresi, setiap variabel prediktor akan memiliki koefisien regresi yang menggambarkan seberapa besar perubahan dalam variabel respon yang diharapkan untuk setiap satu unit perubahan dalam variabel prediktor, dengan asumsi variabel prediktor lainnya tetap konstan.
- Misalnya, dalam regresi linear, koefisien regresi  $\beta_1$  menunjukkan perubahan rata-rata dalam variabel respon untuk setiap satu unit perubahan dalam variabel prediktor  $X_1$ .

### **b. Interpretasi Koefisien:**

- Interpretasi koefisien regresi akan bervariasi tergantung pada jenis regresi yang digunakan dan sifat variabelnya.
- Dalam regresi linear, koefisien regresi menunjukkan perubahan rata-rata dalam variabel respon untuk setiap satu unit perubahan dalam variabel prediktor, dengan asumsi variabel prediktor lainnya tetap konstan.

- Dalam regresi logistik, koefisien regresi menunjukkan perubahan dalam log-odds kejadian untuk setiap satu unit perubahan dalam variabel prediktor.
- Untuk variabel kategorikal, koefisien regresi menunjukkan perbedaan antara kelompok referensi (biasanya kelompok dengan nilai 0) dan kelompok lainnya.

### **c. Pentingnya Interpretasi Koefisien:**

- Interpretasi koefisien membantu dalam memahami hubungan antara variabel prediktor dan variabel respon dalam konteks spesifik masalah yang sedang diteliti.
- Ini membantu dalam merumuskan kesimpulan tentang variabel mana yang memiliki dampak signifikan terhadap variabel respon, serta arah dan ukuran dampaknya.

### **d. Confidence Interval (Interval Kepercayaan):**

- Selain nilai koefisien regresi, penting juga untuk memperhatikan interval kepercayaan untuk koefisien tersebut. Interval kepercayaan menggambarkan kisaran nilai yang mungkin untuk koefisien dengan tingkat kepercayaan tertentu (misalnya, 95%).
- Jika interval kepercayaan meliputi nilai 0, maka koefisien tidak signifikan secara statistik dan kontribusinya terhadap variabel respon tidak dapat dianggap signifikan.

### **e. Pengaruh Multikolinieritas:**

- Jika terdapat multikolinieritas antara variabel prediktor, interpretasi koefisien regresi dapat menjadi tidak stabil atau tidak dapat diinterpretasikan secara akurat.
- Multikolinieritas terjadi ketika dua atau lebih variabel prediktor saling berkorelasi secara kuat. Hal ini dapat

menyebabkan koefisien menjadi tidak stabil atau memiliki tanda yang bertentangan dengan apa yang diharapkan.

Dengan memperhatikan interpretasi koefisien model regresi dengan hati-hati, Anda dapat memahami kontribusi masing-masing variabel prediktor terhadap variabel respon dengan lebih baik. Ini memungkinkan Anda untuk mengambil kesimpulan yang lebih tepat dan relevan dari hasil analisis regresi Anda.

Penting untuk dicatat bahwa tidak ada satu metode atau model regresi yang "terbaik" untuk semua situasi. Pilihan model terbaik akan tergantung pada konteks data, asumsi yang memenuhi, dan tujuan analisis Anda.

## DAFTAR PUSTAKA

- Draper, N., and H. Smith, 1992, Analisis Regresi Terapan edisi kedua (Terjemahan oleh Bambang Sumantri), Gramedia, Jakarta.
- Herlina, H. (2011). Perbandingan Metode Stepwise, Best Subset Regression, dan Fraksi dalam Pemilihan Model Regresi Berganda Terbaik. Jurusan Matematika, F.MIPA, Universitas Sriwijaya. Tersedia di <http://statistik.studentjournal.ub.ac.id>.
- Pujilestari, S. (2016). Pemilihan Model Regresi Linear Berganda Terbaik Pada Kasus Multikolinearitas Berdasarkan Metode Principal Component Analysis (PCA) Dan Metode Stepwise. Skripsi. Jurusan Matematika, F.MIPA, Universitas Negeri Semarang
- Samosir, dkk. (2014). Analisa Metode Backward dan Metode Forward Untuk Menentukan Persamaan Regresi Linier Berganda: Saintia Matematika, 2(4), 345-360.



# **BAB 10**

## **UJI PARSIAL, SEKUENSIAL DAN HIPOTESIS LINIER UMUM**

**Oleh Suri Toding Lembang**

### **10.1 Uji Parsial**

Uji parsial adalah teknik statistik yang digunakan untuk menentukan pengaruh relatif dari satu variabel independen terhadap variabel dependen dalam suatu model regresi, sementara variabel independen lainnya tetap konstan. Tujuannya adalah untuk mengisolasi dampak individu dari masing-masing variabel independen terhadap variabel dependen, sambil mengontrol variabel independen lainnya. Prosesnya seringkali melibatkan penghapusan satu variabel independen dari model regresi, lalu mengukur perubahan dalam koefisien variabel lainnya. Jika ada perubahan yang signifikan dalam koefisien setelah penghapusan variabel independen tertentu, itu menunjukkan bahwa variabel tersebut memiliki pengaruh yang signifikan terhadap variabel dependen secara individual. Uji parsial memungkinkan untuk memahami kontribusi relatif dari masing-masing variabel independen terhadap variabel dependen, yang berguna untuk menyelidiki hubungan sebab-akibat antara variabel dalam model regresi. Ini adalah alat penting dalam analisis regresi dan membantu dalam membuat keputusan yang lebih tepat dalam konteks prediksi atau penjelasan fenomena yang diamati.

Ada beberapa jenis uji parsial yang umum digunakan dalam analisis regresi, tergantung pada struktur model regresi dan pertanyaan penelitian yang ingin dijawab (Bambang, 1996). Berikut adalah beberapa jenis uji parsial yang umum beserta rumusnya:

## 1. Uji T

Uji T untuk koefisien regresi tunggal adalah alat statistik yang digunakan untuk menilai signifikansi dari koefisien regresi untuk satu variabel independen dalam sebuah model regresi. Uji ini membantu dalam menentukan apakah variabel independen tersebut memiliki pengaruh yang signifikan terhadap variabel dependen secara individu. Dengan melakukan uji T untuk koefisien regresi tunggal, kita dapat menentukan apakah variabel independen tersebut memiliki pengaruh yang signifikan terhadap variabel dependen dalam model regresi. Berikut adalah langkah-langkah umum untuk melakukan uji T untuk koefisien regresi tunggal:

- a. Merumuskan Hipotesis Nol ( $H_0$ ) dan Hipotesis Alternatif ( $H_1$ )

Hipotesis nol ( $H_0$ ): Tidak ada hubungan antara variabel independen dan variabel dependen, atau koefisien regresi untuk variabel independen adalah nol.

Hipotesis alternatif ( $H_1$ ): Ada hubungan antara variabel independen dan variabel dependen, atau koefisien regresi untuk variabel independen tidak sama dengan nol.

- b. Menghitung Statistik Uji T

Statistik uji T dihitung dengan membagi koefisien regresi ( $b$ ) dengan standar error ( $SE(b)$ ) dari koefisien regresi.

Rumusnya adalah: 
$$\frac{b}{SE(b)}$$

c. Menentukan Derajat Kebebasan dan Ambil Keputusan

- Derajat kebebasan untuk uji T adalah  $n-k-1$  di mana  $n$  adalah jumlah observasi dan  $k$  adalah jumlah variabel independen dalam model.
- Selanjutnya, dengan menggunakan nilai t-statistik yang dihitung. Nilai  $p$  menunjukkan signifikansi statistik dari koefisien regresi.
- Jika nilai  $p$  lebih kecil dari tingkat signifikansi yang ditetapkan (biasanya 0.05), maka kita menolak hipotesis nol dan menyimpulkan bahwa koefisien regresi tersebut signifikan secara statistik.

## 2. Uji F

Uji F untuk variabel regresi berganda adalah alat statistik yang digunakan untuk menentukan apakah penambahan satu atau beberapa variabel independen secara bersama-sama secara signifikan meningkatkan kecocokan model regresi. Uji ini memberikan informasi tentang apakah model regresi secara keseluruhan lebih baik dalam menjelaskan variabilitas variabel dependen ketika variabel independen ditambahkan ke dalam model dengan melakukan uji F untuk variabel regresi berganda, kita dapat mengevaluasi apakah penambahan variabel independen tertentu secara bersama-sama secara signifikan meningkatkan kemampuan model regresi dalam menjelaskan variabilitas variabel dependen (Sumantri, 2010). Langkah-langkah dan konsep yang terlibat dalam uji F untuk variabel regresi berganda:

a. Merumuskan Hipotesis Nol ( $H_0$ ) dan Hipotesis Alternatif ( $H_1$ )

Hipotesis nol ( $H_0$ ): Tidak ada peningkatan signifikan dalam kecocokan model regresi setelah penambahan satu atau beberapa variabel independen.

Hipotesis alternatif (H1): Ada peningkatan signifikan dalam kecocokan model regresi setelah penambahan satu atau beberapa variabel independen.

b. Perhitungan Jumlah Kuadrat Residual (SSR)

- Statistik uji F dihitung dengan membandingkan perbedaan dalam *SSR*

*SSR* antara model regresi penuh dan model regresi yang lebih sederhana terhadap *SSR* model regresi yang lebih sederhana.

- Rumusnya adalah:

$$F = \frac{(SSR_r - SSR_{ur}) / q}{qSSR_{ur} / (n - k - 1)}$$

Dimana

*SSR<sub>r</sub>* : jumlah kuadrat residual dari model regresi penuh (dengan semua variabel independen),

*SSR<sub>ur</sub>* : jumlah kuadrat residual dari model regresi yang tidak termasuk variabel independen yang diuji,

*q* : jumlah variabel independen yang diuji,

*n* : jumlah observasi,

*k* : jumlah total variabel independen dalam model.

### 3. Menentukan Derajat Kebebasan dan pengambilan Keputusan

- a. Derajat kebebasan untuk uji F adalah *q*, yaitu jumlah variabel independen yang diuji.

- b. Dengan menggunakan nilai statistik uji F yang dihitung. Nilai  $p$  menunjukkan signifikansi statistik dari penambahan variabel independen ke dalam model regresi.
- c. Jika nilai  $p$  kurang dari tingkat signifikansi yang ditetapkan (0.05), kita menolak hipotesis nol dan menyimpulkan bahwa penambahan variabel independen tersebut signifikan secara statistik dalam meningkatkan kecocokan model regresi.

Asumsi-asumsi yang perlu dipertimbangkan ketika melakukan uji parsial, terutama uji T dan uji F dalam konteks analisis regresi (Sumantri, 2010), meliputi:

- a. Normalitas Residual

Asumsi bahwa residu (selisih antara nilai observasi aktual dan nilai yang diprediksi oleh model) didistribusikan secara normal. Ini dapat diperiksa dengan memeriksa plot Q-Q (quantile-quantile) dari residu atau dengan menggunakan tes normalitas seperti tes Kolmogorov-Smirnov atau tes Shapiro-Wilk

- b. Homoskedastisitas

Asumsi bahwa varians dari residu konstan di sepanjang seluruh rentang nilai prediksi. Ini bisa diperiksa dengan memeriksa plot residual versus nilai prediksi dan melakukan uji statistik seperti tes Breusch-Pagan atau White untuk heteroskedastisitas.

- c. Independensi Residual

Asumsi bahwa tidak ada pola yang berlebihan dalam sisa-sisa (residuals), yang menunjukkan adanya ketergantungan antara observasi. Ini bisa diperiksa dengan memeriksa plot residual terhadap variabel independen atau dengan menggunakan tes Durbin-Watson untuk mengevaluasi otonomi serial (autokorelasi).

d. Linieritas

Asumsi bahwa hubungan antara variabel independen dan variabel dependen adalah linier. Ini dapat diperiksa dengan memeriksa plot scatter dari variabel independen terhadap variabel dependen, serta dengan mempertimbangkan transformasi variabel jika diperlukan.

e. Tidak Adanya Multikolinearitas

Asumsi bahwa tidak ada hubungan linier sempurna antara dua atau lebih variabel independen. Multikolinearitas dapat menyebabkan estimasi koefisien yang tidak stabil. Hal ini dapat diperiksa dengan menghitung koefisien korelasi antara variabel independen dan dengan menggunakan faktor inflasi varians (VIF).

f. Jumlah Observasi yang Cukup

Untuk uji parsial yang menggunakan distribusi  $t$  seperti uji  $T$ , jumlah observasi harus cukup besar untuk memastikan validitas statistik dari hasil. Jumlah observasi yang kecil dapat menyebabkan perkiraan yang tidak stabil atau tidak signifikan.

Interpretasi hasil uji parsial bergantung pada jenis uji yang dilakukan, baik itu uji  $T$  untuk koefisien regresi tunggal atau uji  $F$  untuk variabel regresi berganda. Di bawah ini saya akan memberikan panduan umum untuk interpretasi hasil uji parsial (Sumantri, 2010):

**a. Uji  $T$**

Jika nilai  $t$ -statistik signifikan (nilai  $p < \text{tingkat signifikansi yang ditetapkan}$ ), maka menolak hipotesis nol. Hal ini menunjukkan bahwa koefisien regresi untuk variabel independen yang diuji signifikan secara statistik. Jika koefisien regresi positif dan signifikan, maka dapat menyimpulkan bahwa terdapat hubungan positif yang

signifikan antara variabel independen dan variabel dependen. Jika koefisien regresi negatif dan signifikan, maka hubungan antara variabel independen dan variabel dependen adalah negatif dan signifikan.

#### **b. Uji F**

Jika nilai F-statistik signifikan (nilai  $p < \text{tingkat signifikansi yang ditetapkan}$ ), maka menolak hipotesis nol. Hal ini menunjukkan bahwa penambahan variabel independen yang diuji secara bersama-sama secara signifikan meningkatkan kecocokan model regresi. Jika penambahan variabel independen secara bersama-sama meningkatkan kecocokan model, maka dapat menyimpulkan bahwa variabel independen tersebut memberikan kontribusi yang signifikan dalam menjelaskan variabilitas variabel dependen dalam model regresi.

### **10.2 Sekuensial**

Analisis regresi sekuensial adalah salah satu metode statistik yang digunakan untuk memprediksi nilai variabel dependen berdasarkan nilai variabel independen. Metode ini memiliki langkah-langkah yang sistematis dan terstruktur dalam membangun model regresi. Analisis regresi sekuensial adalah teknik statistik yang digunakan untuk mengembangkan model prediktif dengan memasukkan variabel independen secara bertahap. Tujuan utama dari analisis ini adalah untuk mengidentifikasi faktor-faktor yang paling berpengaruh terhadap variabel dependen dan membangun model regresi yang paling optimal. Berikut adalah langkah-langkah dalam melakukan analisis sekuensial (Suhartono, 2008):

## **1. Identifikasi Variabel Independen yang Akan Diuji**

Tentukan variabel-variabel independen yang ingin Anda uji satu per satu atau dalam kelompok dalam model regresi. Variabel-variabel ini dapat dipilih berdasarkan teori, literatur sebelumnya, atau analisis eksploratori awal.

## **2. Bangun Model Regresi Awal**

Bangun model regresi awal dengan menggunakan beberapa variabel independen yang sudah dipilih sebelumnya. Model ini merupakan model referensi yang akan digunakan untuk membandingkan kecocokan model setelah penambahan variabel independen baru.

## **3. Uji Kecocokan Model Awal**

Uji kecocokan model awal dengan menggunakan metrik evaluasi yang sesuai, seperti R-squared, adjusted R-squared, atau uji F untuk menguji signifikansi keseluruhan model. Hasil ini akan menjadi dasar untuk membandingkan kecocokan model setelah penambahan variabel independen.

## **4. Tambahkan Variabel Independen Pertama**

Tambahkan variabel independen pertama ke dalam model regresi awal. Lakukan perhitungan interaksi atau efek moderasi antara variabel independen yang baru ditambahkan dan variabel independen yang sudah ada dalam model.

## **5. Uji Kecocokan Model Setelah Penambahan Variabel Independen Pertama**

Uji kecocokan model baru setelah penambahan variabel independen pertama. Gunakan metrik evaluasi yang sesuai untuk menilai peningkatan signifikansi model, seperti uji F untuk mengevaluasi peningkatan signifikansi secara keseluruhan atau metrik lainnya seperti AIC



(*Akaike Information Criterion*) atau BIC (*Bayesian Information Criterion*) untuk membandingkan kecocokan model.

## **6. Analisis Hasil dan Interpretasi**

Interpretasikan hasil uji kecocokan model setelah penambahan variabel independen pertama. Tentukan apakah penambahan variabel independen tersebut secara signifikan meningkatkan kecocokan model secara keseluruhan. Jika penambahan variabel independen pertama tidak signifikan, pertimbangkan untuk menambahkan variabel independen berikutnya atau mengubah strategi analisis.

Dalam melakukan analisis regresi sekuensial, terdapat beberapa asumsi yang perlu diperhatikan untuk memastikan keandalan dan validitas hasil analisis. Memastikan bahwa asumsi-asumsi ini terpenuhi penting untuk menghasilkan hasil analisis yang dapat diandalkan dan valid. Jika ada pelanggaran asumsi, langkah-langkah koreksi atau transformasi variabel mungkin diperlukan untuk memperbaiki model regresi. Oleh karena itu, penting untuk melakukan diagnosis yang tepat sebelum dan setelah melakukan analisis regresi sekuensial (Suhartono, 2008). Berikut adalah asumsi-asumsi yang perlu dipertimbangkan:

### **1. Normalitas Residual**

- Asumsi bahwa residu (selisih antara nilai observasi aktual dan nilai yang diprediksi oleh model) didistribusikan secara normal.
- Diperiksa dengan memeriksa plot Q-Q (quantile-quantile) dari residu atau dengan menggunakan tes normalitas seperti tes Kolmogorov-Smirnov atau tes Shapiro-Wilk.

## **2. Homoskedastisitas**

- Asumsi bahwa varians dari residu konstan di sepanjang seluruh rentang nilai prediksi.
- Diperiksa dengan memeriksa plot residual versus nilai prediksi dan dengan melakukan uji statistik seperti tes Breusch-Pagan atau White untuk heteroskedastisitas.

## **3. Independensi Residual**

- Asumsi bahwa tidak ada pola yang berlebihan dalam sisa-sisa (residuals), yang menunjukkan adanya ketergantungan antara observasi.
- Diperiksa dengan memeriksa plot residual terhadap variabel independen atau dengan menggunakan tes Durbin-Watson untuk mengevaluasi otonomi serial (autokorelasi).

## **4. Linieritas**

- Asumsi bahwa hubungan antara variabel independen dan variabel dependen adalah linier.
- Diperiksa dengan memeriksa plot scatter dari variabel independen terhadap variabel dependen, serta dengan mempertimbangkan transformasi variabel jika diperlukan.

Interpretasi hasil analisis regresi sekuensial melibatkan evaluasi penambahan variabel independen satu per satu atau dalam kelompok terhadap kecocokan model regresi secara keseluruhan. Berikut adalah langkah-langkah untuk menginterpretasikan hasil analisis regresi sekuensial:

### **a. Penambahan Variabel Independen**

Interpretasi dimulai dengan memperhatikan efek penambahan setiap variabel independen ke dalam model regresi. Jika penambahan variabel independen tertentu

menghasilkan peningkatan yang signifikan dalam kecocokan model, hal ini menunjukkan bahwa variabel tersebut memberikan kontribusi yang signifikan terhadap penjelasan variabilitas dalam variabel dependen.

b. Uji Signifikansi

Evaluasi apakah peningkatan kecocokan model setelah penambahan variabel independen secara signifikan berbeda dari kecocokan model awal. Gunakan uji statistik, seperti uji F, untuk mengevaluasi signifikansi penambahan variabel independen secara bersama-sama dalam model.

c. Perbandingan Model

Bandingkan model regresi awal (tanpa variabel independen tambahan) dengan model regresi yang diperluas (termasuk variabel independen tambahan). Perhatikan peningkatan dalam metrik evaluasi model, seperti R-squared atau adjusted R-squared, untuk memahami seberapa baik model yang diperluas menjelaskan variabilitas dalam variabel dependen.

d. Interpretasi Koefisien Regresi

Perhatikan nilai koefisien regresi untuk setiap variabel independen yang dimasukkan ke dalam model regresi. Interpretasikan koefisien regresi sebagai perubahan dalam variabel dependen yang terjadi ketika variabel independen mengalami perubahan satu unit, dengan mengabaikan efek dari variabel independen lainnya dalam model.

### 7.3 Hipotesis Linier Umum

Hipotesis linier umum adalah salah satu asumsi penting dalam analisis regresi. Hipotesis ini menyatakan bahwa hubungan antara variabel independen dan variabel dependen adalah linier, yang berarti bahwa perubahan dalam variabel independen akan menyebabkan perubahan yang proporsional dalam variabel dependen. Hipotesis linier umum menunjukkan bahwa hubungan antara variabel independen dan variabel dependen dapat dimodelkan sebagai garis lurus dalam ruang dimensi yang sesuai. Hal ini memungkinkan untuk melakukan estimasi parameter regresi menggunakan metode-metode seperti metode kuadrat terkecil (least squares method). Hipotesis linier umum merupakan asumsi dasar yang harus dipertimbangkan dalam analisis regresi. Dalam kasus tersebut, teknik-teknik regresi non-linier mungkin lebih sesuai (Arikunto, 2010). Dinyatakan secara matematis, hipotesis linier umum dapat ditulis sebagai:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Keterangan:

$Y$  : variabel dependen,

$X_1, X_2, \dots, X_p$  : variabel independen,

$\beta_0$  : intercept (konstanta),

$\beta_1, \beta_2, \dots, \beta_p$  : koefisien regresi yang mengukur dampak relatif dari masing-masing variabel independen terhadap variabel dependen.

$\varepsilon$  : kesalahan acak yang tidak dapat dijelaskan oleh model.

## DAFTAR PUSTAKA

- Arikunto, S. (2010) *Statistika untuk Penelitian*. Penerbit: Rineka Cipta.
- Bambang, S. (1996) *Analisis Regres*. Penerbit : Tarsito.
- Suhartono (2008) *Dasar-dasar Analisis Regresi*. Penerbit: Andi Offset.
- Sumantri, B. (2010) *Analisis Regresi Linear Sederhana dan Bergand*. Penerbit: PT Refika Aditama.



# **BAB 11**

## **KORELASI DAN REGRESI LINEAR SEDERHANA**

**Oleh Rosi Ramayanti**

Korelasi sederhana merupakan suatu teknik statistik yang digunakan untuk mengukur kekuatan hubungan dua variabel dan juga untuk dapat mengetahui bentuk hubungan antara dua variabel tersebut dengan hasil yang sifatnya kuantitatif. Kekuatan hubungan antara dua variabel yang dimaksud disini adalah apakah hubungan tersebut ERAT, LEMAH, ataupun TIDAK ERAT sedangkan bentuk hubungannya adalah apakah bentuk korelasinya Linear Positif ataupun Linear Negatif. Dalam statistik kita mengenal hubungan antar dua variabel, yang digunakan untuk mengukur ada atau tidak hubungan antar variabel disebut Korelasi.

### **11.1 Korelasi yang Terjadi Antara Dua Variabel**

Berikut adalah jenis-jenis korelasi yang dapat terjadi antara dua variabel.

1. Korelasi Positif adalah korelasi dua variabel, apabila variabel independen (X) meningkat atau turun maka variabel dependen (Y) cenderung untuk meningkat atau turun.
2. Korelasi Negatif adalah korelasi dua variabel, apabila variabel independen (X) meningkat atau turun maka

variabel dependen (Y) cenderung untuk turun atau meningkat.

3. Tidak ada Korelasi terjadi apabila kedua variabel X dan Y tidak menunjukkan adanya hubungan.
4. Korelasi Sempurna adalah korelasi dari dua variabel yang benar-benar terjadi.

### 11.2 Koefisien Korelasi Sederhana

Untuk mengetahui hubungan antara dua variabel, maka cukup melihat nilai dari koefisien korelasi. Koefisien korelasi ( $r$ ) merupakan indeks atau bilangan yang digunakan untuk mengukur keeratan hubungan antar variabel. Berikut adalah rumus dari koefisien korelasi.

dimana

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{(n \sum X^2 - (\sum X)^2)(n \sum Y^2 - (\sum Y)^2)}}$$

$X$  = Variabel independen

$Y$  = Variabel dependen

$n$  = Banyaknya sampel

Dengan nilai dari  $r$  antara  $-1$  dan  $1$  ( $-1 \leq r \leq 1$ ).

### 11.3 Interval Keeratan Korelasi Antar Variabel

Untuk mengetahui hubungan yang terjadi antara dua variabel, apakah terjadi hubungannya sempurna, kuat, lemah, atau tidak adanya hubungan, berikut diberikan interval-interval yang menyatakan keeratan hubungan antar variabel.



1.  $r = 0$  tidak ada korelasi
2.  $0 < r \leq 0,20$  korelasi sangat lemah sekali
3.  $0,20 < r \leq 0,40$  korelasi lemah sekali
4.  $0,40 < r \leq 0,70$  korelasi yang cukup kuat
5.  $0,70 < r \leq 0,90$  korelasi yang kuat
6.  $0,90 < r < 1,00$  korelasi sangat kuat
7.  $r = 1$ , korelasi sempurna

### **11.4 Koefisien Determinasi**

Koefisien determinasi sering diartikan sebagai seberapa besar kemampuan semua variabel independen dalam menjelaskan varians dari variabel dependennya. Secara sederhana koefisien determinasi dihitung dengan mengkuadratkan koefisien korelasi ( $r$ ). Contohnya, jika nilai  $r$  adalah sebesar 0,8 maka koefisien determinasi adalah sebesar  $0,8 \times 0,8 = 0,64$ . Artinya kemampuan variabel independen dalam menjelaskan varians dari variabel dependennya adalah sebesar 64%. Berarti terdapat 36% ( $100\% - 64\%$ ) varians variabel dependen yang dijelaskan oleh faktor lain. Berdasarkan interpretasi tersebut, maka dapat disimpulkan bahwa nilai koefisien determinasi antara 0 sampai 1.

### **11.5 Regresi Linier Sederhana**

Regresi linear adalah alat statistik yang dipergunakan untuk mengetahui pengaruh antara satu atau beberapa variabel terhadap satu buah variabel. Variabel yang mempengaruhi sering disebut variabel bebas, variabel independen atau variabel penjelas. Variabel yang dipengaruhi sering disebut dengan variabel terikat atau variabel dependen. Regresi linear hanya dapat digunakan pada skala interval dan ratio. Model yang paling sederhana untuk menjelaskan pengaruh antara

variabel dependen dengan satu variabel independen merupakan regresi sederhana.

### 11.6 Model Regresi Sederhana

Persamaan regresi sederhana secara umum dituliskan sebagai berikut:

$$\bar{Y} = a + bX$$

dimana :

$Y$  = Variabel dependen

$X$  = Variabel independen

$a$  = Konstanta

$b$  = koefisien regresi

dengan

$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

### 11.7 Kesalahan Baku Estimasi

$$a = \frac{\sum Y \sum X^2 - \sum X \sum XY}{n \sum X^2 - (\sum X)^2}$$

$$n \sum X^2 - (\sum X)^2$$

Kesalahan baku atau selisih taksir standar regresi adalah nilai menyatakan seberapa jauh menyimpangnya nilai regresi tersebut terhadap nilai sebenarnya. Nilai ini digunakan untuk mengukur tingkat ketepatan suatu pendugaan dalam

menduga nilai. Jika nilai ini sama dengan nol maka penduga tersebut memiliki tingkat ketepatan 100%.

Rumus Kesalahan baku estimasi:

$$S_{x,y} = \sqrt{\frac{\sum(Y - \bar{Y}_t)^2}{n - 2}}$$

dengan

$S_{K,F}$  = Kesalahan baku

$Y$  = Variabel dependen

$\bar{Y}$  = Persamaan regresi

$n$  = banyaknya sampel

### Contoh 1

Pak Budiman, manajer pemasaran PT.ABC memiliki data harga jual dengan volume penjualan produknya selama 10 bulan, dan pak Budiman ingin mengamati hubungan, persentase variabel  $Y$  yang dapat dijelaskan oleh variabel  $X$ , pengaruh dan kesalahan baku yang terjadi antara dua variabel tersebut ?

Volume penjualan dan harga jual produk PT.ABC dinyatakan dalam Tabel 11.1 (Volume penjualan dan harga jual produk PT. ABC).

**Tabel 11. 1. Volume penjualan dan harga jual produk PT. ABC**

| Bulan | Volume penjualan<br>(Dalam ribuan) | Harga jual<br>(Dalam ribuan) |
|-------|------------------------------------|------------------------------|
| 1     | 10                                 | 1,3                          |
| 2     | 6                                  | 2,0                          |
| 3     | 5                                  | 1,7                          |
| 4     | 12                                 | 1,5                          |
| 5     | 10                                 | 1,6                          |

| Bulan | Volume penjualan<br>(Dalam ribuan) | Harga jual<br>(Dalam ribuan) |
|-------|------------------------------------|------------------------------|
| 6     | 15                                 | 1,2                          |
| 7     | 5                                  | 1,6                          |
| 8     | 12                                 | 1,4                          |
| 9     | 17                                 | 1,0                          |
| 10    | 20                                 | 1,1                          |

### Penyelesaian:

Pada contoh 1, yang ditanyakan adalah:

Kasus 1. Korelasi (hubungan antara volume penjualan dengan harga jual)

Kasus 2. Persentase variabel Y yang dapat dijelaskan oleh X (Koefisien determinasi) Kasus 3. Regresi (pengaruh) variabel independen terhadap variabel dependen Kasus 4. Kesalahan baku estimasi

Jadi, terdapat 4 kasus yang harus diselesaikan dalam contoh 1. Sebelum menyelesaikan kasus-kasus tersebut, kita harus menentukan siapa yang menjadi variabel X dan variabel Y. Dengan mengingat kembali bahwa X adalah variabel independen dan Y adalah variabel dependen. X adalah variabel yang mempengaruhi Y. Sehingga dapat ditentukan bahwa X adalah harga jual dan Y adalah volume penjualannya.

Setelah ditentukan siapa yang menjadi variabel X dan Y, langkah selanjutnya adalah menghitung nilai dari  $XY$ ,  $X^2$ , dan  $Y^2$ . Perhitungan akan lebih mudah jika disajikan dalam bentuk Tabel 2 (Perhitungan data).

**Tabel 11. 2. Perhitungan data 1**

|               | Y          | X           | XY           | X <sup>2</sup> | Y <sup>2</sup> |
|---------------|------------|-------------|--------------|----------------|----------------|
|               | 10         | 1,3         | 13           | 1,69           | 100            |
|               | 6          | 2           | 12           | 4              | 36             |
|               | 5          | 1,7         | 8,5          | 2,89           | 25             |
|               | 12         | 1,5         | 18           | 2,25           | 144            |
|               | 10         | 1,6         | 16           | 2,56           | 100            |
|               | 15         | 1,2         | 18           | 1,44           | 225            |
|               | 5          | 1,6         | 8            | 2,56           | 25             |
|               | 12         | 1,4         | 16,8         | 1,96           | 144            |
|               | 17         | 1           | 17           | 1              | 289            |
|               | 20         | 1,1         | 22           | 1,21           | 400            |
| <b>Jumlah</b> | <b>112</b> | <b>14,4</b> | <b>149,3</b> | <b>21,56</b>   | <b>1488</b>    |

### **Kasus 1. Hubungan Antara Volume Penjualan Dan Harga Jual**

Untuk melihat hubungan antara X dan Y maka dihitung nilai dari koefisien korelasi  $r$  dengan menggunakan rumus yang sudah diberikan dan melihat nilai-nilai pada Tabel 2. Contohnya,  $\sum XY$  adalah jumlah dari kolom yang menyatakan XY (Hal ini terdapat dalam kolom 4). Sehingga nilai dari  $\sum XY = 149,3$ . Diperhatikan kembali, bahwa nilai dari  $\sum X^2$  G  $(\sum X)^2$  dan  $\sum Y^2$  G  $(\sum Y)^2$ . Setelah semua data diinput, diperoleh nilai dari koefisien korelasi sebagai berikut.

$$\begin{aligned}
 r &= \frac{n \sum XY - \sum X \sum Y}{\sqrt{(n \sum X^2 - (\sum X)^2)(n \sum Y^2 - (\sum Y)^2)}} \\
 &= \frac{10 (149,3) - 14,4 (112)}{\sqrt{(10 (21,56) - 14,4^2)(10 (1488) - 112^2)}}
 \end{aligned}$$

$$= \frac{-119,8}{138,739}$$

$$= -0,863$$

Koefisien korelasi sebesar -0,87 menunjukkan hubungan linier negatif yang kuat artinya bila harga naik maka volume penjualan akan turun.

## Kasus 2 Koefisien Determinasi

Persentase variabel Y yang dapat dijelaskan variabel X, dengan menghitung koefisien determinasi yaitu dengan mengkuadratkan koefisien korelasi

$$(-0,87)^2 = 0,7569$$

Artinya kemampuan harga jual barang dalam menjelaskan varians dari volume penjualan adalah sebesar 75,69%. Berarti terdapat 24,31% (100% – 75,69%) varians volume penjualan yang dijelaskan oleh faktor lain, misalnya kualitas barang.

## Kasus 3 Pengaruh Harga Jual Terhadap Volume Penjualan

Untuk mengetahui pengaruh harga jual terhadap volume penjualan (pengaruh X terhadap Y) maka harus dilakukan pembuatan model regresi, yaitu  $\bar{Y} = a + bX$ . Sehingga terlebih dahulu harus dihitung nilai dari  $a$  dan  $b$  dengan menggunakan rumus yang sudah dijelaskan.

$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

$$= \frac{10 (149,3) - 14,4 (1612,8)}{10 (21,56) - 14,4^2}$$

$$= -14,539$$

dan

$$a = \frac{\sum Y \sum X^2 - \sum X \sum XY}{n \sum X^2 - (\sum X)^2}$$

$$= \frac{112 (21,56) - 14,4 (149,3)}{10 (21,56) - 14,4^2}$$

$$= 32,136$$

$$\sum Y \sum X^2 - \sum X \sum XY$$

Setelah nilai  $a$  dan  $b$  diperoleh maka disubstitusikan pada model regresinya, sehingga diperoleh:

$$\bar{Y} = a + bX = 32,136 - 14,539X$$

Interprestasi dari model regresi.

$$\bar{Y} = 32,136 - 14,539X$$

Nilai  $a = 32,136$  artinya jika harga sama dengan nol maka rata-rata 32.136 produk akan terjual.

Nilai  $b = -14,54$  artinya jika harga naik 1,00 (Rp.1000,00) maka volume penjualan akan turun sebesar 14,54 unit, begitu juga sebaliknya. Jika harga turun sebesar 1 (Rp. 1000,00) maka volume penjualan naik sebesar 14,54 unit. Hal ini sesuai dengan analisis mengenai korelasi antara harga jual dan volume penjualan, yang menyatakan bahwa jika harga jual naik maka volume penjualan akan turun.

Kasus 4 Kesalahan Baku Estimasi

Selanjutnya, dilakukan perhitugan mengenai kesalahan baku estimasi dengan menggunakan rumus

S\_{x,Y} = \sqrt{\frac{\sum(Y-\bar{Y}\_t)^2}{n-2}}

Sebelum data dimasukkan pada rumus ,baiknya dilakukan perhitungan dengan menggunakan Tabel 11.3 (Perhitungan data) untuk mempermudah perhitungan kesalahan baku estimasinya.

Tabel 11. 3. Perhitungan data 2

| Y      | X   | $\bar{Y}$ | $Y - \bar{Y}$ | $(Y - \bar{Y})^2$ |
|--------|-----|-----------|---------------|-------------------|
| 10     | 1,3 | 13,24     | -3,24         | 10,5              |
| 6      | 2   | 3,06      | 2,94          | 8,64              |
| 5      | 1,7 | 7,42      | -2,42         | 5,86              |
| 12     | 1,5 | 10,33     | 1,67          | 2,79              |
| 10     | 1,6 | 8,88      | 1,12          | 1,25              |
| 15     | 1,2 | 14,69     | 0,31          | 0,096             |
| 5      | 1,6 | 8,88      | -3,88         | 15,05             |
| 12     | 1,4 | 11,78     | 0,22          | 0,048             |
| 17     | 1   | 17,6      | -0,6          | 0,36              |
| 20     | 1,1 | 16,15     | 3,85          | 14,82             |
| Jumlah | 112 | 14,4      |               | 59,414            |

11.8 Pengujian Hipotesis Tentang Koefisien Korelasi

Dalam pengujian hipotesis, yang dibahas pertama dalam modul ini adalah pengujian hipotesis tentang korelasi. Perumusan hipotesis yang digunakan untuk korelasi adalah sebagai berikut.

H<sub>0</sub>: p = 0, X dan Y tidak ada hubungan

H<sub>a</sub>: p < 0, X dan Y mempunyai hubungan negatif

H<sub>a</sub>: p > 0, X dan Y mempunyai hubungan positif



$H_a: \rho \neq 0$ ,  $X$  dan  $Y$  ada hubungan

### Langkah-langkah Pengujian Hipotesis :

1. Merumuskan bentuk hipotesis :

$$H_0: \rho = 0$$

$$H_a: \rho < 0 \text{ Pengujian satu arah}$$

$$H_a: \rho > 0 \text{ Pengujian satu arah}$$

$$H_a: \rho \neq 0 \text{ Pengujian dua arah}$$

2. Menentukan nilai kesalahan =  $\alpha$ , setelah  $\alpha$  diketahui kemudian mencari  $t$  (jika satu arah) atau  $t\alpha$  (jika dua arah) dari Tabel  $t$  (Lampiran 1) dengan  $df = n - 2$ .

$df$  = derajat kebebasan

3. Menghitung nilai dari  $t$  hitung dengan rumus:

$$t_h = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

dengan

$$t_h = \text{nilai } t_{\text{hitung}}$$

$$r = \text{koefisien korelasi}$$

$$n = \text{jumlah sampel}$$

4. Keputusan

Kriteria keputusan dalam pengujian hipotesis korelasi adalah sebagai berikut. Jika nilai dari  $t_{\text{hitung}}$  lebih besar daripada nilai dari  $t_{\text{tabel}}$  maka  $H_0$  ditolak ( $H_a$  diterima). Jika nilai dari  $t_{\text{hitung}}$  lebih kecil daripada nilai dari  $t_{\text{tabel}}$  maka  $H_0$  diterima ( $H_a$  ditolak). Ringkasnya dapat dituliskan sebagai berikut.

Jika  $t_h > t_{tabel}$  maka  $H_0$  ditolak Jika  $t_h < t_{tabel}$  maka  $H_0$  diterima.

### 11.9 Pengujian Hipotesis Tentang Regresi

Pengujian hipotesis selanjutnya yang dibahas adalah pengujian hipotesis tentang regresi. Perumusan hipotesis yang digunakan untuk regresi adalah sebagai berikut.

$H_0: B = 0$ , Tidak ada pengaruh X terhadap Y

$H_a: B < 0$ , Ada pengaruh negatif X terhadap Y

$H_a: B > 0$ , Ada pengaruh positif X terhadap Y

$H_a: B \neq 0$ , Ada pengaruh X terhadap Y

#### Langkah-langkah Pengujian Hipotesis :

1. Merumuskan bentuk hipotesis :  $H_0 : B = 0$   
 $H_a : B < 0$  Pengujian satu arah  
 $H_a : B > 0$  Pengujian satu arah  
 $H_a : B \neq 0$  Pengujian dua arah
2. Menentukan nilai kesalahan =  $\alpha$ , setelah  $\alpha$  diketahui kemudian mencari  $t$  (jika satu arah) atau  $t\alpha$  (jika dua arah) dari Tabel t (Lampiran 1) dengan  $df = n - 2$ .  
 $df$  = derajat kebebasan
3. Menghitung  $t$  hitung dengan rumus
4. Keputusan  
 Kriteria keputusan dalam pengujian hipotesis korelasi adalah sebagai berikut. Jika nilai dari  $t_{hitung}$  lebih besar daripada nilai dari  $t_{tabel}$  maka  $H_0$  ditolak ( $H_a$  diterima). Jika nilai dari  $t_{hitung}$  lebih kecil daripada nilai dari  $t_{tabel}$  maka  $H_0$  diterima ( $H_a$  ditolak). Ringkasnya dapat dituliskan sebagai berikut.  
 Jika  $t_h > t_{tabel}$  maka  $H_0$  ditolak Jika  $t_h < t_{tabel}$  maka  $H_0$  diterima.

## DAFTAR PUSTAKA

- Douglas C. Montgomery, George C. Runger, 2003, *Applied Statistic and Probability for Engineer*, third edition, John Wiley and Son Inc.
- J. Supranto, M.A. ,2001, *Statistika Teori dan Aplikasi*, Erlangga, Jakarta. R Johson and D Wichern, *Applied multivariate statistics*, Prentice Hall.
- Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye, *Probabilitiy and Statistics for Engineers and Scientists*, Pearson Prentice Hall, 8th edition, 2007
- Singgih Santoso, 2014, *Panduan Lengkap SPSSversi 20*, Alex Media Komputindo. Subhash Sharma, *Applied Multivariate Techniques*, , John wiley and son



## **BIODATA PENULIS**



### **Puji Muniarty**

**Dosen Sekolah Tinggi Ilmu Ekonomi Bima  
Program Studi Manajemen**

Penulis Lahir di Nunggi pada tanggal 26 Nopember 1985. Menyelesaikan pendidikan S-1 dan S-2 Manajemen Konsentrasi Manajemen Keuangan Bisnis di Universitas Mataram. Penulis mengajar di Sekolah Tinggi Ilmu Ekonomi (STIE) Bima. Penulis pernah menjadi dosen tamu di PICOMS Internasional University College Kuala Lumpur Malaysia. Sebagai akademisi penulis juga terlibat dalam Tim Penyusunan Studi Awal Masterplan Pengembangan Ekonomi Kegiatan Koordinasi Perencanaan Pembangunan Ekonomi Kabupaten Bima, Tim Survei Pemantauan Harga (SPH) Bank Indonesia, Tim Penyusunan Naskah Akademik Pemberian Insentif Dan Kemudahan Penanaman Modal Di Kota Bima, Tim Hygiene Factor PT. PLN ULP3 Bima, Tim Konsultasi Public PT. PLN ULP3 Bima, Anggota Dewan Pengupahan Kabupaten Bima, Tim Juri Kompetisi Wirausaha Muda Tingkat Kota Bima Dan Dosen Magang Perguruan Tinggi yang ditempatkan di IPB University dan dosen yang memiliki keahlian dan kompetensi dalam bidang Keuangan ini dibuktikan dengan perolehan gelar AFA (Accredited Financial Analyst) dari American Academy of Financial Management USA. Email : puji.stiebima@gmail.com.

## **BIODATA PENULIS**



**Salsabila, S.Stat., M.Si**

Dosen Program Studi Pendidikan Matematika  
Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Raden  
Intan Lampung

Penulis lahir di Bandar Lampung pada tanggal 20 Desember 1996. Penulis adalah dosen tetap pada Program Studi Pendidikan Matematika, Fakultas Tarbiyah dan Keguruan Universitas Islam Negeri Raden Intan Lampung. Menyelesaikan pendidikan S1 pada jurusan Statistika dan melanjutkan S2 pada jurusan Statistika Terapan. Menulis merupakan sarana penulis untuk selalu dapat mengingat, memperbaharui, serta mendalami materi statistika dan penerapannya khususnya dalam bidang pendidikan. Hasil tulisan ini diharapkan dapat memberikan manfaat bagi pembacanya untuk dapat lebih memahami konsep dan mempraktekkan pengolahan data.

## **BIODATA PENULIS**



**Lita Lovia, M.Si**

Dosen Program Studi Matematika  
Universitas Dharma Andalas

Penulis lahir pada 20 April 1980 di Minang Kabau Lb. Aur. Tamat Sarjana pada Program Studi Matematika di Universitas Andalas Pada tahun 2003. Tamat Magister Program Studi Matematika di Universitas Andalas pada tahun 2011

## **BIODATA PENULIS**



**Dr. Muhammad Ilham Riyadh, SP., M.Si**

Penulis lahir di Medan 05 Februari 1979 adalah Dosen Tetap Yayasan Universitas Islam Sumatera Utara (UISU). Menyelesaikan Pendidikan Sarjana S1 Fakultas Pertanian Program Studi Sosial Ekonomi Pertanian UISU tahun 2001, dan menyelesaikan Pendidikan S2 Ilmu Ekonomi Pertanian (EPN) di Sekolah Pascasarjana Institut Pertanian Bogor (IPB) tahun 2007 dan juga menyelesaikan Pendidikan S3 program Doktor Ilmu Ekonomi USU tahun 2023, sekarang aktif Menulis dan Mengajar serta aktif di bidang Penelitian dan Pengabdian Masyarakat.



## **BIODATA PENULIS**



**Azis, S.Pd., M.Pd.**

Dosen Program Studi Pendidikan Matematika  
Fakultas Keguruan dan Ilmu Pendidikan  
Universitas Dayanu Ikhsanuddin

Penulis lahir di Kadolo Katapi tanggal 29 September 1987. Penulis adalah dosen tetap pada Program Studi Pendidikan Matematika Fakultas Keguruan dan Ilmu Pendidikan, Universitas Dayanu Ikhsanuddin. Menyelesaikan pendidikan S1 pada Program Studi Pendidikan Matematika di Universitas Dayanu Ikhsanuddin dan menyelesaikan pendidikan S2 pada Program Studi Pendidikan Matematika di Universitas Negeri Yogyakarta. Sekarang penulis sedang menempuh studi S3 pada Program Studi Pendidikan Matematika di Universitas Pendidikan Indonesia.

## **BIODATA PENULIS**



### **Santje Magdalena Iriyanto**

Dosen Teknik Sipil

Fakultas Teknik Sipil dan Perencanaan  
Universitas Sains dan Teknologi Jayapura

Santje Magdalena Iriyanto, ST., MT. adalah dosen tetap Fakultas Teknik Sipil dan Perencanaan Universitas Sains dan Teknologi Jayapura sejak tahun 1994 hingga saat ini. Lahir di Fak-Fak pada tahun 1968. Menyelesaikan Pendidikan S1 pada tahun 1994 di Universitas Kristen Indonesia Paulus Makassar. Pada tahun 2010 menyelesaikan Pendidikan Program Pasca Sarjana jurusan Universitas Tarumanagara Jakarta Beberapa Buku yang pernah dibuat berjudul Metode Penelitian dan Statistik I.

## **BIODATA PENULIS**



**Alona Dwinata, S.Si., M.Si.**

Dosen Program Studi Pendidikan Matematika  
Fakultas Keguruan dan Ilmu Pendidikan  
Universitas Maritim Raja Ali Haji

Penulis lahir di Bukittinggi tanggal 23 Februari 1988. Penulis bertugas sebagai dosen dengan bidang keahlian Statistika di Universitas Maritim Raja Ali Haji, kota Tanjungpinang – Kepulauan Riau. Penulis menempuh pendidikan S1 pada program studi Matematika di Universitas Negeri Padang pada tahun 2005 – 2009. Penulis melanjutkan pendidikan S2 pada program studi Statistika di Institut Pertanian Bogor pada tahun 2010 – 2012.

## BIODATA PENULIS



**Drs. Bambang Suhartawan, M.MT**

Dosen Jurusan Teknik Lingkungan, FTSP  
Universitas Sains dan Teknologi Jayapura

Penulis lahir di Blitar Tahun 1964. Penulis adalah dosen tetap dengan status PNS-DPK pada Jurusan Teknik Lingkungan Fakultas Teknik Sipil dan Perencanaan (FTSP) Universitas Sains dan Teknologi Jayapura. Menyelesaikan pendidikan S1 pada Program Studi Kimia Jurusan Matematika dan Ilmu Pengetahuan Alam (MIPA) Fakultas Keguruan dan Ilmu Pendidikan Universitas Cenderawasih Jayapura pada tahun 1989. Pendidikan Pasca Sarjana diperoleh dari Institut Teknologi Sepuluh Nopember (ITS) Surabaya jurusan Manajemen Teknik Lingkungan pada tahun 2002. Beberapa buku yang telah ditulis terkait bidang lingkungan antara lain : Pengelolaan Sumber Daya Air; Penyehatan Udara; Adaptasi dan Mitigasi Perubahan Iklim di Indonesia; Ekotoksikologi, Pengelolaan Limbah Industri; Sistem Lingkungan Industri; Analisis Kualitas Lingkungan; Kesehatan Lingkungan Bancana; Manajemen Penyakit Tropis Berbasis Lingkungan; Pengelolaan Limbah Padat, Limbah Industri dan B3; Lingkungan : Pemahaman Masalah dan Solusi; Kesehatan Lingkungan : Memahami Dampak Lingkungan Terhadap Kesehatan; Kimia Ramah Lingkungan; Pencemaran Udara dan Perubahan Iklim; Entomologi dalam Kesehatan

Lingkungan; Pengelolaan Limbah Cair; Kimia Lingkungan; Kesehatan Lingkungan Global; Pengenalan Pembangunan Berkelanjutan dan Memahami Konsep dan Teori Ekologi Lingkungan.

## BIODATA PENULIS



### **Suri Toding Lembang, S.Pd., M.Pd.**

Dosen Program Studi Pendidikan Matematika  
Fakultas Keguruan dan Ilmu Pendidikan Universitas Kristen  
Indonesia Toraja

Lahir di Rantepao 18 September 1990. Pendidikan yang telah ditempuh, SDN Malango Rantepao 1996, SMPN 1 Toraja Utara 2002, SMA Negeri 2 Toraja Utara 2005. Lulus S1 di Program Studi Pendidikan Matematika Universitas Kristen Indonesia Toraja (UKI Toraja) tahun 2012, kemudian melanjutkan pendidikan S2 di Universitas Negeri Makasar pada program studi pendidikan matematika. Saat ini sedang melanjutkan pendidikan S3 di Universitas Negeri Makasar pada Program studi pendidikan matematika. Aktif menulis pada berbagai jurnal ilmiah baik nasional maupun internasional. Beberapa tulisan juga telah dimuat dalam Prosiding Internasional seperti *The Analysis Concept of Integers Counting Operations in Traditional Toraja Games Si Goal and Si Patte* tahun 2021 dan *Identification Of Fractional Numbers In The Procedure For The Division Of Duku' Tedong At The Solo Sign' Event In Toraja* Tahun 2022. Penulis juga sering telah menulis buku berjudul *Aljabar Linier* Tahun 2019, *Kalkulus Differensial* 2023, *Metode Penelitian Lanjutan* 2023.