ISBN 978-634-255-001-4

ANALISIS REGRESI DASAR

PENULIS :
Joni Wilson Sitopu
Dodi
Dewi Rakhmatia Nur
Yunita Widia Putri
Tedi Hartoyo
Ade Astuti Widi Rahayu
Hamdan
Patrich Papilaya

ANALISIS REGRESI DASAR

Joni Wilson Sitopu
Dodi
Dewi Rakhmatia Nur
Yunita Widia Putri
Tedi Hartoyo
Ade Astuti Widi Rahayu
Hamdan
Patrich Papilaya

GET PRESS INDONESIA

ANALISIS REGRESI DASAR

Penulis : Joni Wilson Sitopu, Dodi, Dewi Rakhmatia Nur, Yunita Widia Putri, Tedi Hartoyo, Ade Astuti Widi Rahayu, Hamdan, Patrich Papilaya
ISBN : 978-634-255-001-4
Editor : Ari Yanto M.Pd
Penyunting : Tri Putri Wahyuni., S.Pd
Desain Sampul dan Tata Letak : Atyka Trianisa, S.Pd
Penerbit : GET PRESS INDONESIA
Anggota IKAPI No. 033/SBA/2022
Redaksi : Jln. Palarik Air Pacah No 26 Kel. Air Pacah Kec. Koto Tangah Kota Padang Sumatera Barat
Website : www. getpress.co.id. Email : adm.getpress@gmail.com

Perpustakaan Nasional RI : Katalog Dalam Terbitan (KDT)	
JUDUL DAN PENANGGUNG JAWAB	Analisis Regresi Dasar; Joni Wilson Sitopu, Dodi, Dewi Rakhmatia Nur, Yunita Widia Putri, Tedi Hartoyo [dan 3 lainnya] Editor : Ari Yanto., M.Pd. Tri Putri Wahyuni., S.Pd
EDISI	Cetakan pertama, 2025
PUBLIKASI	Padang : Get Press Indonesia, 2025
DESKRIPSI FISIK	vi 132 halaman : ilustrasi ; 23 cm
IDENTIFIKASI	ISBN 978-634-255-001-4
SUBJEK	Analisis Regresi
KLASIFIKASI	519. 536 [23]
PERPUSNAS ID	https://isbn.perpusnas.go.id/bo-penerbit/penerbit/isbn/data/view-kdt/1260394

Undang Undang Republik Indonesia Nomor 19 *Tahun 2002*
Tentang Hak Cipta

Ketentuan Pidana:

Pasal 72

1. Barangsiapa dengan sengaja dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam Pasal 2 ayat (1) atau Pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling singkat 1 (satu) bulan dan/atau denda paling sedikit Rp 1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (tujuh) tahun dan/ atau denda paling banyak Rp 5.000.000.000,00 (lima miliar rupiah).
2. Barangsiapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu Ciptaan atau barang hasil pelanggaran Hak Cipta atau Hak Terkait sebagaimana dimaksud pada ayat (1) dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/ atau denda paling banyak Rp 500.000.000,00 (lima ratus juta rupiah).

KATA PENGANTAR

Segala Puji dan syukur atas kehadirat Allah SWT dalam segala kesempatan. Sholawat beriring salam dan doa kita sampaikan kepada Nabi Muhammad SAW. Alhamdulillah atas Rahmat dan Karunia-Nya penulis telah menyelesaikan Buku Analisis Regresi Dasar ini.

Buku Ini Membahas Konsep Dasar Regresi dan Korelasi, Model Regresi Linear Sederhana, Inferensi dalam Regresi Linear Sederhana, Analisis Variansi dalam Regresi, Regresi Linear Berganda, Analisis Variansi Model Regresi Ganda, Regresi dengan Peubah Boneka, Regresi Non-Linear.

Proses penulisan buku ini berhasil diselesaikan atas kerjasama tim penulis. Demi kualitas yang lebih baik dan kepuasan para pembaca, saran dan masukan yang membangun dari pembaca sangat kami harapkan.

Penulis ucapkan terima kasih kepada semua pihak yang telah mendukung dalam penyelesaian buku ini. Terutama pihak yang telah membantu terbitnya buku ini dan telah mempercayakan mendorong, dan menginisiasi terbitnya buku ini. Semoga buku ini dapat bermanfaat bagi masyarakat Indonesia.

Padang, Juni 2025

Penulis

DAFTAR ISI

KATA PENGANTAR ..... **i**
DAFTAR ISI ..... **ii**
DAFTAR TABEL ..... **vi**
BAB 1 KONSEP DASAR REGRESI DAN KORELASI **1**
1.1 Pendahuluan 1
1.2 Regresi 2
1.2.1 Analisis Regresi 3
1.2.2 Garis/Persamaan Regresi 4
1.2.3 Koefisien Regresi 6
1.3 Korelasi 8
1.3.1 Koefisien Korelasi 8
1.3.2 Analisis korelasi 10
1.3.3 Jenis-jenis Korelasi 10
DAFTAR PUSTAKA 13
BAB 2 MODEL REGRESI LINEAR SEDERHANA **15**
2.1 Pengertian 16
2.2 Sejarah Singkat Regresi 16
2.3 Komponen-Komponen Utama 17
2.4 Ilustrasi Sederhana 17
2.5 Ciri-Ciri Model Regresi Linear Sederhana 18
2.6 Kapan Model Ini Relevan? 18
2.6 Asumsi-Asumsi 18
2.7 Aplikasi Model Regresi Linear Sederhana 21
DAFTAR PUSTAKA 24
BAB 3 INFERENSI DALAM REGRESI LINEAR SEDERHANA **25**
3.1 Pendahuluan 25
3.2 Model Regresi Secara Umum 26
3.3 Estimasi Parameter 26
3.4 Asumsi dalam Regresi Linear Sederhana 27
3.5 Pengujian Hiptesis 27
3.5.1 Koefisien Regresi $\boldsymbol{\beta 1}$ 27
3.5.2 Intersep $\boldsymbol{\beta 0}$ 28
3.5.3 Uji F untuk Signifikansi Model Keseluruhan 29
3.6 Interval Kepercayaan 30

3.7 Koefisien Determinasi 30
DAFTAR PUSTAKA 36
BAB 4 ANALISIS VARIANSI DALAM REGRESI 37
4.1 Pengertian Analisis Variansi dalam Regresi 37
4.2 Tabel ANOVA untuk Regresi 38
4.3 Interpretasi Analisis Variansi dalam Regresi 44
4.4 Aplikasi untuk Analisis Variansi 46
DAFTAR PUSTAKA 48
BAB 5 REGRESI LINEAR BERGANDA 49
5.1 Pendahuluan 49
5.2 Model regresi 51
5.3 Metode LAPLACE 53
5.3.1 Contoh kasus analisis regresi ganda 54
5.3.2 Penyelesaian dengan Metode Laplace 54
5.4 Metode DOOLITTLE 58
5.4.1 Prosedur atau solusi Analisis regresi dengan metode Doolittle dimulai setelah diketauhi (X'X), (X 'Y) dan (Y'Y) 59
5.4.2 Pengujian dengan Metode Analisis Varian 61
5.4.3 Metode Doolittle pada Contoh Soal 61
5.5 Penutup 64
DAFTAR PUSTAKA 67
BAB 6 ANALISIS VARIANSI MODEL REGRESI GANDA 69
6.1 Pendahuluan 69
6.2 Konsep Umum Analisis Variansi (ANOVA) 69
6.2.1 Definisi 70
6.2.2 Jenis-jenis ANOVA 70
6.2.3 Kegunaan ANOVA 71
6.3 Statistik F dan Uji Signifikansi 72
6.3.1 Konsep Dasar Statistik F 73
6.3.2 Rumus Statistik F 73
6.3.3 Hipotesis dalam Uji F 73
6.3.4 Contoh Interpretasi 73
6.3.5 Implikasi 73
6.4 Koefisien Determinasi 74
6.5 Langkah-langkah Analisis 75
6.6 Contoh Perhitungan 76

6.7 Kesimpulan Akhir 87
DAFTAR PUSTAKA 89
BAB 7 REGRESI DENGAN PEUBAH BONEKA (DUMMY) . 91
7.1 Pendahuluan 91
7.2 Memahami Peubah Dummy 92
7.3 Kelebihan dan Kelemahan Peubah Dummy 98
7.4 Teknik Pengkodean Peubah Dummy 99
7.5 Model Regressi dengan Peubah Dummy 102
7.6 Evaluasi Model dan Uji Signifikansi 105
7.7 Studi Kasus dan Aplikasi Peubah Dummy 108
7.7 Kesimpulan 110
DAFTAR PUSTAKA 111
BAB 8 REGRESI NON-LINIER 113
8.1 Pendahuluan 113
8.2. Konsep Dasar Regresi Non-Linear 114
8.2.1 Definisi dan Karakteristik 114
8.2.2 Perbedaan dengan Regresi Linear 114
8.3 Jenis-Jenis Model Regresi Non-Linear 115
8.3.1 Model Eksponensial 115
8.3.2 Model Logaritmik 115
8.3.3 Model Polinomial 116
8.3.4 Model Logistik 116
8.4 Metode Estimasi Parameter 116
8.4.1 Metode Gauss-Newton 116
8.4.2 Metode Levenberg-Marquardt 117
8.4.3 Metode Maximum Likelihood 117
8.5 Evaluasi dan Validasi Model 118
8.5.1 Kriteria Goodness of Fit 118
8.5.2 Kriteria Informasi 118
8.5.3 Analisis Residual 118
8.6 Aplikasi dan Studi Kasus 119
8.6.1 Aplikasi dalam Biologi 119
8.6.2 Aplikasi dalam Ekonomi 119
8.6.3 Aplikasi dalam Farmakologi 119
8.7 Tantangan dan Limitasi 119
8.7.1 Masalah Konvergensi 119
8.7.2 Sensitivitas terhadap Nilai Awal 120

8.8 Perkembangan Terkini dan Masa Depan 120
8.8.1 Machine Learning dan Deep Learning..... 120
8.8.2 Regularisasi dalam Model Non-Linear 120
8.8.3 Bayesian Non-Linear Regression..... 120
DAFTAR PUSTAKA 121
BIODATA PENULIS

DAFTAR TABEL

Tabel 1.1. Perbedaan antara Korelasi dan Regresi 3
Tabel 1.2. Dua persamaan normal 6
Tabel 1.3. Rumus Koefisien Regresi 6
Tabel 4.1. Menghitung df 39
Tabel 4.2. Contoh Data 41
Tabel 4.3. Menghitung Nilai Rata-Rata X dan Y 42
Tabel 4.4. Menghitung Jumlah Kuadrat 42
Tabel 4.5. Nilai df dan Jumlah Kuadrat 44
Tabel 4.6. Menghitung Mean Square4 4
Tabel 4.7. ANOVA 45
Tabel 5.1. Analysis of Variance (ANOVA) 58
Tabel 5.2. Analisis Varian (Anova) 61

BAB 1
KONSEP DASAR REGRESI DAN KORELASI

Oleh Joni Wilson Sitopu

1.1 Pendahuluan

Dalam analisis data, statistika memainkan peran penting, dan salah satu konsep utama yang sering digunakan adalah korelasi. Korelasi adalah metode statistika untuk mengukur hubungan antara dua variabel. Hal ini berguna untuk memahami apakah ada hubungan kuat atau lemah antara kedua variabel tersebut. (dalam I Gede Purnawinadi, Etriana Meirista., dkk., (2023)). Aspek yang paling penting dalam penelitian sosial adalah memisahkan dan memahami hubungan diantara variabel independen (bebas) dengan variabel dependen (terikat). Korelasi adalah hubungan antara dua atau lebih variabel. Sejauh mana hubungan diantara variabel-variabel penelitian bisa diukur dengan menggunakan korelasi Pearson. Koefisien korelasi Pearson bisa digunakan jika data dalam bentuk rasio dan kaitan antara variabel adalah linier. (Nazariah, Noviyanti, dkk. (2022)). Analisis korelasi adalah prosedur statistik yang dengannya kita menentukan derajat hubungan atau keterkaitan antara dua variabel atau lebih. Namun korelasi tidak memprediksi apa pun tentang hubungan sebab akibat. Bahkan derajat korelasi yang tinggi tidak serta merta menyiratkan adanya hubungan sebab akibat antara kedua variabel. Studi korelasi menjadi penting hanya jika dipelajari dengan analisis regresi. Studi gabungan dari analisis korelasi dan regresi sangat menarik saat mempelajari masalah dalam ilmu sosial, penelitian pendidikan, pembuatan kebijakan dan pengambilan keputusan, dll. Koefisien korelasi dilambangkan dengan r. Karl Pearson (1967 – 1936), seorang ahli biometrika Inggris, mengembangkan rumus untuk Koefisien Korelasi. Koefisien korelasi antara dua variabel X dan Y dilambangkan dengan r(x,y) atau r_{xy}. Konsep korelasi dan regresi dapat ditinjau dan membandingkan dua

koefisien korelasi, koefisien korelasi Pearson dan Spearman, untuk mengukur hubungan linier dan nonlinier antara dua variabel kontinu.

Analisis regresi adalah analisis yang digunakan untuk mengukur ada atau tidaknya hubungan antar variabel. Regresi linier adalah regresi yang memiliki variabel bebas dengan pangkat paling tinggi adalah satu. (dalam Joni Wilson Sitopu., Yongker Baali., dkk., (2023)). Dalam kasus pengukuran hubungan linier antara prediktor dan variabel hasil, analisis regresi linier sederhana dapat dilakukan. Analisis regresi berfokus pada bentuk hubungan antar variabel, sedangkan tujuan analisis korelasi adalah untuk mendapatkan wawasan tentang kekuatan hubungan antara variabel. Analisis regresi linear merupakan suatu alat analisis dalam model statistika yang bertujuan untuk mengetahui hubungan/pengaruh antara variabel bebas dan variabel terikat. Jenis hubungan/pengaruh antara variabel bebas dan terikat dapat berbentuk hubungan/pengaruh positif maupun pengaruh yang negatif. (dalam Joni Wilson Sitopu., Akhmad Faisal Malik., dkk., (2023)).

1.2 Regresi

Analisis regresi linear sederhana merupakan suatu proses perhitungan dengan muara membentuk persamaan garis lurus yang menggambarkan keterkaitan antara tujuan penelitian (Y) dengan masalah dalam suatu penelitian (X). (dalam KMT Lasmiatun, Solehudin., dkk. (2023)).

Sebuah studi pengukuran tentang hubungan antara variabel terkait, di mana satu variabel dependen pada variabel independen lainnya, disebut Regresi. Hal ini dikembangkan oleh Sir Francis Galton pada tahun 1877 untuk mengukur hubungan tinggi badan antara orang tua dan anak-anak mereka. Analisis regresi adalah alat statistic untuk mempelajari sifat dan tingkat hubungan fungsional antara dua atau lebih variabel dan untuk memperkirakan (atau memprediksi) nilai variabel dependen yang tidak diketahui dari nilai variabel independen yang diketahui. Variabel yang menjadi dasar untuk memprediksi variabel lain dikenal sebagai Variabel Independen dan variabel yang diprediksi dikenal sebagai variabel dependen.

Misalnya, jika diketahui bahwa dua variabel berpengaruh dengan harga (X) dan permintaan (Y), maka kita dapat mengetahui nilai X yang paling mungkin untuk nilai Y tertentu atau nilai Y yang paling mungkin untuk nilai X tertentu. Demikian juga, jika diketahui dua variabel berpengaruh dengan jumlah pajak dan kenaikan harga komoditas, kita dapat mengetahui harga yang diharapkan untuk sejumlah pungutan pajak tertentu.

1.2.1 Analisis Regresi

1. Memberikan estimasi nilai variabel dependen dari nilai variabel independen.
2. Digunakan untuk memperoleh ukuran kesalahan yang terlibat dalam penggunaan garis regresi sebagai dasar estimasi.
3. Dengan bantuan analisis regresi, kita dapat memperoleh ukuran derajat hubungan atau korelasi yang ada antara kedua variabel.
4. Merupakan alat yang sangat berharga dalam penelitian ekonomi dan bisnis, karena sebagian besar masalah analisis ekonomi didasarkan pada hubungan sebab akibat.

Tabel 1.1. Perbedaan antara Korelasi dan Regresi

No.	Korelasi	Regresi
1.	Mengukur derajat dan arah hubungan antara variabel.	Mengukur sifat dan luas hubungan rata-rata antara dua atau lebih variabel dalam hal unit data asli.
2.	Korelasi adalah ukuran relatif yang menunjukkan hubungan antara variabel.	Regresi adalah ukuran absolut dari hubungan.
3.	Koefisien Korelasi tidak bergantung pada perubahan asal dan skala.	Koefisien Regresi tidak bergantung pada perubahan asal tetapi tidak pada skala.

No.	Korelasi	Regresi
4.	Koefisien Korelasi tidak bergantung pada satuan pengukuran.	Koefisien Regresi tidak independen dari satuan pengukuran.
5.	Ekspresi hubungan antar variabel berkisar antara –1 sampai +1	Ekspresi hubungan antar variabel dapat berupa salah satu bentuk berikut: $Y = a + bX$ $Y = a + bX + cX^2$
6.	Korelasi bukan alat peramalan.	Regresi adalah alat peramalan yang dapat digunakan untuk memperkirakan nilai variabel dependen dari nilai variabel independen yang diberikan.
7.	Mungkin tidak ada korelasi seperti berat badan dan pendapatan.	Tidak ada yang seperti regresi nol.

Sumber : Nazariah, Noviyanti, dkk. (2022).

1.2.2 Garis/Persamaan Regresi

Regresi merupakan suatu alat ukur yang juga digunakan untuk mengukur ada atau tidaknya korelasi antar variabelnya. Istilah regresi mengacu pada perkiraan. Analisis regresi mempelajari hubungan fungsional antara variabel-variabel yaitu ; hubungan antara satu variabel prediktor dan satu variabel kriterium disebut analisis regresi sederhana (tunggal), dan hubungan antara lebih dari satu variabel disebut analisis regresi kompleks. Persamaan yang digunakan untuk menghasilkan garis regresi pada data diagram pencar disebut persamaan regresi. Analisis regresi melihat bagaimana variabel variabel berhubungan satu sama lain. (I Gede Purnawinadi, Etriana Meirista., dkk., (2023)). Garis regresi dan persamaan regresi digunakan secara sinonim. Persamaan regresi adalah ekspresi aljabar dari garis regresi. Misalnya terdapat dua variabel : X dan Y. Jika y bergantung pada x, maka hasilnya berupa regresi sederhana. Jika kita mengambil kasus dua variabel X dan Y, kita akan memiliki dua garis regresi sebagai garis regresi X pada Y dan garis regresi Y pada X. Garis regresi Y pada X memberikan nilai Y

yang paling mungkin untuk nilai X tertentu dan garis regresi X pada Y memberikan nilai X yang paling mungkin untuk nilai Y tertentu. Jadi, kita memiliki dua garis regresi. Namun, jika ada korelasi positif sempurna atau korelasi negatif sempurna antara kedua variabel, kedua garis regresi akan bertepatan, yaitu kita akan memiliki satu garis. Jika variabelnya independen, r adalah nol dan garis regresi berada pada sudut siku-siku, yaitu sejajar dengan sumbu X dan sumbu Y. Oleh karena itu, dengan bantuan model regresi linier sederhana, kita memiliki dua garis regresi berikut ;

1. Garis regresi Y pada X : Garis ini memberikan nilai kemungkinan Y (Variabel Dependen) untuk setiap nilai X (Variabel Independen) yang diberikan.
   Garis regresi Y pada X :
   $Y - \hat{Y} = b_{xy}(X - \hat{X})$ atau $: Y = a + bX$
2. Garis regresi X pada Y: Garis ini memberikan nilai kemungkinan X (Variabel Dependen) untuk setiap nilai Y (Variabel Independen) yang diberikan.
   Garis regresi X pada Y :
   $X - \hat{X} = b_{yx}(Y - \hat{Y})$ atau $: X = a + bY$

Pada kedua garis regresi atau persamaan regresi di atas, terdapat dua parameter regresi, yaitu a dan b. Dimana, a adalah konstanta yang tidak diketahui dan b yang juga dilambangkan sebagai b_{xy} atau b_{yx}, juga merupakan konstanta lain yang tidak diketahui yang disebut sebagai koefisien regresi. Oleh karena itu, a dan b ini adalah dua konstanta yang tidak diketahui (nilai numerik tetap) yang menentukan posisi garis secara menyeluruh. Jika nilai salah satu atau keduanya diubah, garis lain akan ditentukan. Parameter a menentukan level garis yang dipasang (yaitu jarak garis tepat di atas atau di bawah titik asal). Parameter b menentukan kemiringan garis (yaitu perubahan Y untuk perubahan satuan X). Jika nilai konstanta a dan b dapat diperoleh, maka garis tersebut sepenuhnya dapat ditentukan. Namun pertanyaannya adalah bagaimana memperoleh nilai-nilai tersebut. Solusinya dengan menggunakan metode kuadrat terkecil. Dengan menggunakan aljabar dan kalkulus diferensial, dapat ditunjukkan bahwa dua

persamaan normal berikut, jika dipecahkan secara bersamaan, akan menghasilkan nilai parameter a dan b.

Tabel 1.2. Dua persamaan normal

X pada Y	Y pada X
$\sum X = Na + b\sum Y$ $\sum XY = a\sum Y + b\sum Y^2$	$\sum Y = Na + b\sum X$ $\sum XY = a\sum Y + b\sum Y^2$

Sumber : KMT Lasmiatun, Solehudin., dkk. (2023)

Metode di atas dikenal sebagai metode langsung, yang menjadi cukup rumit jika nilai X dan Y besar. Penyelesaiannya dapat disederhanakan dengan menggunakan regresi Y pada X dan X pada Y sebagai berikut;

Persamaan regresi Y pada X:

$Y = a + bX$ akan berubah menjadi $Y - \hat{Y} = b_{xy}(X - \hat{X})$

Persamaan regresi X pada Y:

$X = a + bY$ akan berubah menjadi $X - \hat{X} = b_{yx}(Y - \hat{Y})$

Dalam bentuk persamaan regresi baru ini, dapat dihitung satu parameter, yaitu b. b dilambangkan dengan b_{xy} atau b_{yx} disebut sebagai koefisien regresi.

1.2.3 Koefisien Regresi

Jumlah b dalam persamaan regresi disebut sebagai koefisien regresi atau koefisien kemiringan. Karena ada dua persamaan regresi, maka kita memiliki dua koefisien regresi.

1. Koefisien Regresi X pada Y, ditulis sebagai b_{xy}
2. Koefisien Regresi Y pada X, ditulis sebagai b_{yx}

Berbagai rumus digunakan untuk menghitung koefisien regresi :

Tabel 1.3. Rumus Koefisien Regresi

Metode	Koefisien Regresi X pada Y	Koefisien Regresi Y pada X
Menggunakan koefisien korelasi (r) dan simpangan baku (σ)	$b_{xy} = r\,\frac{\sigma_x}{\sigma_y}$	$b_{yx} = r\,\frac{\sigma_y}{\sigma_x}$
Metode Langsung: Menggunakan jumlah X dan Y	$b_{xy} = \frac{N\sum XY - \sum X\sum}{N\sum Y^2 - (\sum Y)^2}$	$b_{yx} = \frac{N\sum XY - \sum X\sum}{N\sum X^2 - (\sum X)}$
Ketika penyimpangan diambil dari rata-rata aritmatika	$b_{xy} = \frac{\sum xy}{\sum y^2}$	$b_{yx} = \frac{\sum xy}{\sum x^2}$

Sumber : KMT Lasmiatun, Solehudin., dkk. (2023)

Sifat-sifat Koefisien Regresi:

1. Koefisien korelasi adalah rata-rata geometrik dari dua koefisien regresi. Disimbolkan dengan r = b_{xy} V b_{yx}
2. Jika salah satu koefisien regresi lebih besar dari satu, yang lain harus lebih kecil dari satu, karena nilai koefisien korelasi tidak dapat melebihi satu. Misalnya jika b_{xy}= 1,2 dan b_{yx} = 1,4 , r akan menjadi = $\sqrt{1{,}2 * 1{,}4}$ = 1,29, yang tidak mungkin.
3. Kedua koefisien regresi akan memiliki tanda yang sama. Yaitu keduanya akan positif atau negatif. Dengan kata lain, tidak mungkin salah satu koefisien regresi memiliki tanda minus dan yang lainnya bertanda plus.
4. Koefisien korelasi akan memiliki tanda yang sama dengan koefisien regresi, yaitu jika koefisien regresi memiliki tanda negatif, r juga akan memiliki tanda negatif dan jika koefisien regresi memiliki tanda positif, r juga akan positif. Misalnya, jika b_{xy} = -0,2 dan b_{yx} = -0,8 maka r = - $\sqrt{0{,}2 * 0{,}8}$ = – 0,4
5. Nilai rata-rata dari kedua koefisien regresi akan lebih besar daripada nilai koefisien korelasi. Dalam simbol (b_{xy} + b_{yx}) / 2 > r. Misalnya, jika b_{xy} = 0,8 dan b_{yx} = 0,4 maka rata-rata dari

kedua nilai = (0,8 + 0,4) / 2 = 0,6 dan nilai r = r = $\sqrt{0,8*0,4}$= 0,566 yang kurang dari 0,6

6. Koefisien regresi tidak bergantung pada perubahan asal tetapi tidak pada skala.

1.3 Korelasi

Analisis korelasi memberikan wawasan yang berharga dalam berbagai disiplin ilmu, seperti ilmu sosial, ekonomi, kedokteran, dan ilmu alam. Dengan memahami sejauh mana dua variabel berkaitan, kita dapat membuat prediksi yang lebih akurat, mengidentifikasi faktor-faktor yang mempengaruhi suatu fenomena, dan mengambil keputusan yang lebih baik. (Yenny Anggreini Sarumaha, Hetty Elfina., dkk., (2023)). Analisis korelasi Person Product Moment merupakan salah satu jenis statistika parametris yang digunakan untuk mencari hubungan dan membuktikan hipotesis hubungan dua variabel bila data kedua variabel berbentuk interval atau rasion, dan sumber data dari dua variabel atau lebih adalah sama. (dalam Elfira Rahmadani, Mohan Taufiq Mashuri, Joni Wilson Sitopu., dkk. (2023)). Analisis korelasi adalah prosedur statistik yang dengannya kita menentukan derajat hubungan atau keterkaitan antara dua variabel atau lebih. Namun korelasi tidak memprediksi apa pun tentang hubungan sebab akibat. Bahkan derajat korelasi yang tinggi tidak serta merta menyiratkan adanya hubungan sebab akibat antara kedua variabel. Studi korelasi menjadi penting hanya jika dipelajari dengan analisis regresi. Studi gabungan dari analisis korelasi dan regresi sangat menarik saat mempelajari masalah dalam ilmu sosial, penelitian pendidikan, pembuatan kebijakan dan pengambilan keputusan, dll.

1.3.1 Koefisien Korelasi

Untuk menentukan korelasi Pearson sekurang kurangnya setiap subjek mempunyai dua ukuran. Misalnya kita ingin mengetahui hubungan atau korelasi diantara tingkat IQ dengan prestasi akademik. Setiap subjek harus mempunyai dua nilai, yaitu nilai IQ dan nilai prestasi akademik. Nilai korelasi diberi simbol r. Nilai r besarnya berada diantara 0 sampai dengan 1. Jika nilai r mendekati 1, korelasi semakin kuat dan jika mendekati 0 korelasinya semakin lemah. Jika r bernilai sama dengan 0 mengidentifikasi

ketiadaan hubungan diantara variabel bebas dengan variabel terikat. Nazariah, Noviyanti, dkk. (2022). Uji korelasi artinya uji untuk mencari hubungan variabel dari data kuantitatif/numerik. Korelasi atau hubungan itu ada dua. Korelasi positif dan korelasi negatif. Tujuan uji korelasi adalah untuk mengetahui keeratan/kekuatan hubungan, dan mengetahui bentuk hubungan positif atau negatif. (dalam Joni Wilson Sitopu., Akhmad Faisal Malik., dkk., (2023)). Koefisien korelasi dilambangkan dengan r. Karl Pearson (1967 – 1936), seorang ahli biometrika Inggris, mengembangkan rumus untuk Koefisien Korelasi. Koefisien korelasi antara dua variabel X dan Y dilambangkan dengan r(x,y) atau r_{xy}. Misalkan $x_1, x_2, \dots x_n$ merupakan himpunan observasi dari variabel X dan misalkan $y_1, y_2, \dots y_n$ merupakan nilai-nilai Y yang bersesuaian. Maka diperoleh :

$$r_{xy} = \frac{\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n}(x_i-\bar{x})^2}\sqrt{\frac{1}{n}(y_i-\bar{y})^2}} = \frac{Cov(X,Y)}{\sigma_x\sigma_y}$$ (dalam Nazariah, Noviyanti, dkk. (2022)).

Di mana, Cov(X,Y) disebut sebagai kovariansi antara X dan Y dengan σ_x dan σ_y masing-masing adalah deviasi standar dari X dan Y.

Dalam dunia bisnis saat ini, kita menjumpai banyak kegiatan yang saling bergantung. Dapat dilihat banyak masalah yang melibatkan penggunaan dua atau lebih variabel. Mengidentifikasi variabel-variabel ini dan ketergantungannya membantu kita dalam menyelesaikan banyak masalah. Sering kali ada masalah atau situasi di mana dua variabel tampak bergerak ke arah yang sama, seperti keduanya meningkat atau menurun. Terkadang, peningkatan satu variabel disertai dengan penurunan variabel lain. Misalnya, pendapatan dan pengeluaran keluarga, harga suatu produk dan permintaannya, pengeluaran iklan, dan volume penjualan, dll. Jika dua kuantitas bervariasi sehingga pergerakan dalam satu variabel disertai dengan pergerakan dalam variabel lain, maka kuantitas ini dikatakan berkorelasi. Artinya, Korelasi adalah teknik statistik untuk memastikan hubungan atau keterkaitan antara dua variabel atau lebih. Analisis korelasi adalah teknik statistik untuk mempelajari derajat dan arah hubungan antara dua variabel atau lebih. Koefisien korelasi adalah ukuran statistik tentang sejauh mana perubahan

pada nilai satu variabel memprediksi perubahan pada nilai variabel lain. Jika fluktuasi satu variabel secara andal memprediksi fluktuasi serupa pada variabel lain, sering kali ada kecenderungan untuk berpikir bahwa perubahan pada satu variabel menyebabkan perubahan pada variabel lain.

1.3.2 Analisis korelasi

1. Analisis korelasi membantu dalam menentukan tingkat dan arah hubungan tersebut secara tepat.
2. Efek korelasi adalah mengurangi rentang ketidakpastian prediksi kita. Prediksi yang didasarkan pada analisis korelasi akan lebih dapat diandalkan dan mendekati kenyataan.
3. Analisis korelasi berkontribusi pada pemahaman perilaku ekonomi, membantu menemukan variabel yang sangat penting yang menjadi sandaran variabel lain, dapat mengungkapkan kepada ekonom hubungan yang menyebabkan gangguan menyebar dan memberikan saran kepadanya tentang cara agar lelucon stabilisasi dapat menjadi efektif.
4. Teori ekonomi dan studi bisnis menunjukkan hubungan antara variabel seperti harga dan kuantitas yang diminta, pengeluaran iklan dan promosi penjualan, dll.
5. Ukuran koefisien korelasi adalah ukuran relatif perubahan.

1.3.3 Jenis-jenis Korelasi

Korelasi dapat dijelaskan atau diklasifikasikan dalam beberapa cara yang berbeda. Ada tiga yang paling penting yaitu : (1).Positif dan Negatif, (2).Sederhana, Parsial dan Ganda, (3).Linear dan non-linier.

1. Korelasi Positif, Negatif dan Nol:
   Apakah korelasi positif (langsung) atau negatif (sebaliknya) akan bergantung pada arah perubahan variabel?
   Korelasi Positif: Jika kedua variabel bervariasi dalam arah yang sama, korelasi dikatakan positif. Artinya jika satu variabel meningkat, yang lain secara rata-rata juga meningkat atau jika satu variabel menurun, yang lain secara rata-rata juga menurun,

maka korelasi dikatakan korelasi positif. Misalnya, korelasi antara tinggi dan berat sekelompok orang adalah korelasi positif.

Tinggi (cm): X	158	160	163	166	168	171	174	176
Berat (kg) : Y	60	62	64	65	67	69	71	72

Korelasi Negatif: Jika kedua variabel bervariasi dalam arah yang berlawanan, korelasinya dikatakan negatif. Jika berarti jika satu variabel meningkat, tetapi variabel lainnya menurun atau jika satu variabel menurun, tetapi variabel lainnya meningkat, maka korelasinya dikatakan korelasi negatif. Misalnya, korelasi antara harga suatu produk dan permintaannya adalah korelasi negatif.

Harga Produk (Rp. Per Unit) : X	6	5	4	3	2	1
Permintaan (Dalam Satuan) : Y	75	120	175	250	215	400

Korelasi Nol: Sebenarnya ini bukan jenis korelasi tetapi tetap disebut sebagai korelasi nol atau tidak ada korelasi. Jika kita tidak menemukan hubungan apa pun antara variabel, maka dikatakan korelasi nol. Artinya perubahan nilai satu variabel tidak memengaruhi atau mengubah nilai variabel lain. Misalnya, korelasi antara berat badan seseorang dan kecerdasan adalah korelasi nol atau tidak ada.

2. Korelasi Sederhana, Parsial, dan Ganda:
   Perbedaan antara korelasi sederhana, parsial, dan ganda didasarkan pada jumlah variabel yang dipelajari.
   Korelasi Sederhana: Ketika hanya dua variabel yang dipelajari, misalnya pada kasus korelasi sederhana. Misalnya, jika seseorang mempelajari hubungan antara nilai yang diperoleh oleh siswa dan kehadiran siswa di kelas, hal dapat disebut masalah korelasi sederhana.
   Korelasi Parsial: Dalam kasus korelasi parsial, seseorang mempelajari tiga variabel atau lebih tetapi menganggap hanya dua variabel yang saling memengaruhi dan pengaruh variabel lain yang memengaruhi dianggap konstan. Misalnya, dalam contoh hubungan antara nilai siswa dan kehadiran di atas, variabel lain yang memengaruhi seperti pengajaran guru yang

efektif, penggunaan alat bantu mengajar seperti komputer, papan pintar dll diasumsikan konstan.
Korelasi Ganda: jika tiga atau lebih variabel dipelajari, maka dapat disebut kasus korelasi ganda. Misalnya, dalam contoh di atas jika studi mencakup hubungan antara nilai siswa, kehadiran siswa, efektivitas guru, penggunaan alat peraga pengajaran, dll., maka hal ini dapat disebut kasus korelasi ganda.

3. Korelasi Linier dan Non-linier:
   Bergantung pada keteguhan rasio perubahan antara variabel, korelasinya bisa berupa Korelasi Linier atau Non-linier.
   Korelasi Linier: Jika jumlah perubahan dalam satu variabel memiliki rasio konstan terhadap jumlah perubahan pada variabel lain, maka korelasi dikatakan linier. Jika variabel tersebut diplot pada kertas grafik, semua titik yang diplot akan jatuh pada garis lurus. Misalnya: Jika diasumsikan bahwa, untuk memproduksi satu unit produk maka dapat membutuhkan 10 unit bahan baku, sehingga untuk memproduksi 2 unit produk maka dapat membutuhkan dua kali lipat dari satu unit tersebut.

Bahan baku : X	10	20	30	40	50	60
Produk Jadi : Y	2	4	6	8	10	12

   Korelasi Non-linier: Jika jumlah perubahan dalam satu variabel tidak memiliki rasio konstan terhadap jumlah perubahan pada variabel lain, maka korelasi dikatakan non-linier. Jika variabel tersebut diplot pada grafik, titik-titiknya akan jatuh pada kurva dan bukan pada garis lurus. Misalnya, jika kita menggandakan jumlah pengeluaran iklan, maka volume penjualan belum tentu akan berlipat ganda.

Biaya Iklan : X	10	20	30	40	50	60
Volume Penjualan : Y	2	4	6	8	10	12

DAFTAR PUSTAKA

Blumberg, B., Cooper, D. dan Schindler, P. (2014) Business Research Methods. McGraw Hill.

Elfira Rahmadani, Mohan Taufiq Mashuri, Joni Wilson Sitopu, dkk.,(2023). Statistika Pendidikan. PT GLOBAL EKSEKUTIF TEKNOLOGI. Padang Sumatera Barat.

Esa Unggul, U. (2018). Korelasi dan Regresi Linier Sederhana. Https://Lms-Paralel.Esaunggul.Ac.Id/.https://lmsparalel.esaunggul.ac.id/pluginfile.php?file=/193372/mod_resource/content/5/6_7450_esa155_042019_pdf. Gede Purnawinadi, Etriana Meirista., dkk., (2023). Biostatistika Dasar. Yayasan Kita Menulis

Joni Wilson Sitopu., Akhmad Faisal Malik., dkk., (2023). APLIKASI SPSS UNTUK ANALISIS DATA PENELITIAN KESEHATAN. Get Press Indonesia. Padang Sumatera Barat.

Joni Wilson Sitopu., Yongker Baali., dkk., (2023). STATISTIK MULTIVARIAT UNTUK EKONOMI DAN BISNIS: di Lengkapi dengan Program IBM SPSS. Get Press Indonesia. Padang Sumatera Barat.

KMT Lasmiatun, Solehudin., dkk. (2023). Manajemen Dan Analisis Data. PT GLOBAL EKSEKUTIF TEKNOLOGI. Padang Sumatera Barat.

Nazariah, Noviyanti, dkk. (2022). Statistik Dasar. PT GLOBAL EKSEKUTIF TEKNOLOGI. Padang Sumatera Barat.

Nugroho, S. (2008) Dasar – Dasar Metode Statistika. Bengkulu: PT. Gramedia Widiasarana Indonesia.

Pelliyezer karo-karo, Anita Roosmawarni., dkk., (2025). Statistik Ekonomi dan Bisnis. Get Press Indonesia. Padang Sumatera Barat.

Sujana. (2017). Metoda Statistika. edisi 7.

Supranto, J. (1988). Statistik: Teori dan Aplikasi, edisi 5.

Wapole. RE. (1975). Introduction to statistic. New York. Macmillan London. Colier Macmillan.

Wapole. RE. (1993). Pengatar Statistik. Gramedia Pustaka Utama. Indonesia

Yenny Anggreini Sarumaha, Hetty Elfina., dkk., (2023). Statistika. Get Press Indonesia. Padang Sumatera Barat.

BAB 2
MODEL REGRESI LINEAR SEDERHANA

Oleh Dodi

Dalam dunia yang kian digerakkan oleh data, kebutuhan untuk memahami dan memprediksi fenomena berdasarkan pola menjadi semakin penting. Baik dalam bidang ekonomi, kesehatan, psikologi, pendidikan, maupun ilmu data, pertanyaan-pertanyaan seperti "Seberapa besar pengaruh pendidikan terhadap pendapatan?", "Apakah tinggi badan dapat memprediksi berat badan?", hingga "Bagaimana suhu memengaruhi konsumsi energi?" merupakan pertanyaan-pertanyaan yang dapat dijawab melalui pendekatan statistik. Salah satu model statistik paling mendasar dan populer untuk menjawab pertanyaan semacam ini adalah Model Regresi Linear Sederhana.

Regresi linear sederhana merupakan teknik analisis yang digunakan untuk memahami hubungan linier antara satu variabel independen (prediktor) dan satu variabel dependen (respon). Dengan kata lain, model ini menjelaskan bagaimana suatu variabel dapat diprediksi berdasarkan variabel lain yang diasumsikan memiliki pengaruh terhadapnya. Meski terkesan sederhana karena hanya melibatkan dua variabel, kekuatan model ini sangat besar, terutama dalam tahap awal eksplorasi data dan pemodelan statistik. Ia juga menjadi pintu gerbang menuju model-model yang lebih kompleks seperti regresi berganda, regresi logistik, dan berbagai teknik machine learning.

Regresi linear sederhana bukan hanya alat prediksi, tetapi juga sarana inferensi. Dengan model ini, peneliti dapat menguji apakah pengaruh suatu variabel benar-benar signifikan terhadap variabel lain, sekaligus mengukur seberapa besar pengaruh tersebut. Dalam praktiknya, model ini banyak digunakan oleh ilmuwan sosial untuk menguji teori, oleh ekonom untuk memprediksi tren, oleh

insinyur untuk melakukan kontrol kualitas, dan oleh perusahaan teknologi untuk mengembangkan produk berbasis data.

Namun demikian, penggunaan regresi linear sederhana tidak lepas dari asumsi-asumsi statistik yang harus dipenuhi agar hasil analisis valid. Asumsi seperti normalitas residual, homoskedastisitas, independensi observasi, dan linieritas hubungan adalah syarat-syarat yang jika dilanggar, bisa menggiring pada kesimpulan yang menyesatkan. Maka dari itu, pemahaman terhadap fondasi matematis dan asumsi model ini sangat penting bagi siapa pun yang hendak menggunakannya dalam penelitian maupun praktik.

2.1 Pengertian

Regresi linear sederhana adalah sebuah model statistik yang digunakan untuk menggambarkan dan menganalisis hubungan antara satu variabel independen (X) dan satu variabel dependen (Y) dengan asumsi bahwa hubungan antara keduanya berbentuk linier. Artinya, jika kita memetakan nilai-nilai X dan Y dalam koordinat kartesius, maka pola hubungan tersebut secara ideal akan membentuk garis lurus. Secara matematis, model regresi linear sederhana dapat dituliskan dalam bentuk:

$Y=\beta_0+\beta_1X+\varepsilon$

Di mana:

1. Y adalah variabel dependen (yang diprediksi)
2. X adalah variabel independen (yang memprediksi)
3. β_0 adalah intercept (titik potong Y saat X = 0)
4. β_1 adalah koefisien regresi (kemiringan garis regresi)
5. ε adalah error atau residual (selisih antara nilai aktual dan nilai prediksi)

2.2 Sejarah Singkat Regresi

Kata “regresi” pertama kali diperkenalkan oleh Sir Francis Galton pada akhir abad ke-19 dalam studinya mengenai hereditas tinggi badan. Ia menemukan bahwa anak-anak dari orang tua yang sangat tinggi cenderung lebih pendek daripada orang tuanya, sedangkan anak dari orang tua yang sangat pendek cenderung lebih tinggi. Fenomena ini ia sebut sebagai "regression toward the mean" atau “regresi menuju rata-rata”.

Studi Galton kemudian dikembangkan oleh Karl Pearson dan rekan-rekannya dalam kerangka statistik yang lebih formal. Mereka memperkenalkan metode kuadrat terkecil (least squares method) sebagai teknik standar untuk mengestimasi garis regresi, yang hingga kini menjadi metode utama dalam analisis regresi.

2.3 Komponen-Komponen Utama

Model regresi linear sederhana hanya terdiri dari dua variabel, namun memiliki komponen penting yang harus dipahami:

1. Variabel Independen (X)
   Sering juga disebut prediktor, variabel ini diasumsikan sebagai penyebab atau faktor yang memengaruhi variabel Y. Misalnya, dalam model hubungan antara jam belajar (X) dan nilai ujian (Y), maka X adalah variabel yang bisa dikendalikan atau diobservasi lebih dulu.
2. Variabel Dependen (Y)
   Disebut juga respon, variabel ini adalah target yang ingin diprediksi atau dijelaskan berdasarkan nilai dari X.
3. Intercept (β_0)
   Merupakan nilai rata-rata Y ketika X = 0. Secara grafis, ini adalah titik di mana garis regresi memotong sumbu Y.
4. Koefisien Regresi (β_1)
   Menggambarkan seberapa besar perubahan rata-rata Y untuk setiap satu unit perubahan X. Jika $\beta_1>0$, maka hubungan X dan Y bersifat positif; jika $\beta_1<0$, maka hubungan bersifat negatif.
5. Error atau Residual (ε)
   Selalu ada faktor-faktor yang tidak dapat dijelaskan sepenuhnya oleh X. Komponen error mencerminkan deviasi antara nilai aktual Y dan nilai Y yang diprediksi oleh model.

2.4 Ilustrasi Sederhana

Bayangkan kita ingin mengetahui apakah jumlah jam belajar memengaruhi nilai ujian matematika siswa. Kita mengumpulkan data dari 30 siswa mengenai: Berapa jam mereka belajar seminggu (X) atau Nilai ujian matematika mereka (Y)

Setelah memplot data, kita perhatikan bahwa semakin banyak jam belajar, nilai ujian cenderung meningkat. Maka kita membuat model:
Nilai=β0+β1×Jam Belajar+ε
Jika hasil estimasi model menghasilkan:
Nilai=55+2.5×Jam Belajar
Interpretasinya: setiap tambahan 1 jam belajar dihubungkan dengan peningkatan rata-rata nilai sebesar 2.5 poin, dan siswa yang tidak belajar sama sekali diasumsikan memiliki nilai 55.

2.5 Ciri-Ciri Model Regresi Linear Sederhana

1. Hanya dua variabel yang dianalisis: satu independen dan satu dependen.
2. Hubungan antara X dan Y bersifat linier.
3. Koefisien regresi tetap: tidak berubah untuk setiap pasangan nilai X dan Y.
4. Asumsi normalitas dan homoskedastisitas berlaku.
5. Model sangat mudah divisualisasikan sebagai garis lurus pada scatterplot.

2.6 Kapan Model Ini Relevan?

Model regresi linear sederhana sangat cocok digunakan saat:

1. Hubungan antara dua variabel terlihat lurus secara visual.
2. Tujuan utama adalah eksplorasi awal, prediksi kasar, atau validasi hipotesis awal.
3. Tidak tersedia banyak variabel atau data yang kompleks.

Contoh penggunaan regresi linear sederhana:

1. Memprediksi pengeluaran rumah tangga dari pendapatan.
2. Mengkaji hubungan antara suhu udara dan konsumsi listrik.
3. Menilai pengaruh jumlah latihan terhadap skor kebugaran.
4. Menilai hubungan antara konsumsi kopi dan jam tidur.

2.6 Asumsi-Asumsi

Asumsi 1: Linieritas (*Linearity*)

Asumsi pertama adalah bahwa hubungan antara variabel independen (X) dan variabel dependen (Y) bersifat linier. Ini berarti

bahwa perubahan dalam X diasumsikan menghasilkan perubahan proporsional dalam Y. Hubungan ini tidak boleh bersifat eksponensial, logaritmik, atau bentuk kurva lain, kecuali sudah dilakukan transformasi data.
Cara memeriksa:

1. Gunakan scatterplot antara X dan Y. Jika pola titik-titik membentuk garis lurus atau mendekati lurus, maka asumsi ini terpenuhi.
2. Residual plot: jika residual tersebar acak di sekitar garis horizontal nol, ini juga menunjukkan linieritas.

Asumsi 2: Independensi (*Independence*)

Asumsi ini menyatakan bahwa setiap observasi data bersifat independen satu sama lain. Artinya, error dari satu pengamatan tidak boleh berkorelasi dengan error dari pengamatan lainnya.
Pelanggaran terhadap asumsi ini umum terjadi dalam data time series di mana error pada waktu t berkorelasi dengan waktu t+1.
Cara memeriksa:

1. Gunakan Durbin-Watson Test untuk mendeteksi autokorelasi.
2. Visualisasi residual terhadap waktu, jika residual membentuk pola atau tren, kemungkinan ada autokorelasi.

Asumsi 3: Homoskedastisitas (*Homoscedasticity*)

Asumsi ini menegaskan bahwa varian residual (error) bersifat konstan untuk semua nilai X. Jika variansi residual berubah tergantung pada nilai X, maka terjadi heteroskedastisitas.
Dampak jika dilanggar:

1. Standard error menjadi tidak akurat
2. Uji-t dan uji-F bisa menjadi tidak valid

Cara memeriksa:

1. Plot residual terhadap nilai prediksi ($\hat{Y}$). Jika bentuknya menyerupai corong atau pola lain, kemungkinan terjadi heteroskedastisitas.
2. Gunakan uji Breusch-Pagan atau uji White.

Asumsi 4: Normalitas Residual (*Normality of Errors*)

Asumsi ini menyatakan bahwa residual (ε) dalam model mengikuti distribusi normal. Meskipun regresi masih bisa digunakan tanpa asumsi ini, normalitas residual penting untuk:

1. Validitas uji hipotesis
2. Pembuatan interval kepercayaan
3. Pengambilan kesimpulan statistik

Cara memeriksa:

1. Plot histogram residual
2. Gunakan Q-Q plot
3. Lakukan uji Shapiro-Wilk atau Kolmogorov-Smirnov

Asumsi 5: Tidak Ada Multikolinearitas (khusus regresi ganda)

Meskipun asumsi ini tidak berlaku dalam regresi linear sederhana (karena hanya satu prediktor), asumsi ini penting jika kita nanti mengembangkan model ke regresi linear berganda. Multikolinearitas terjadi ketika dua atau lebih variabel prediktor sangat berkorelasi satu sama lain, menyebabkan ketidakstabilan dalam estimasi koefisien.

Asumsi Tambahan: Tidak Ada Outlier Ekstrem

Data outlier yang terlalu ekstrem dapat sangat memengaruhi nilai slope dan intercept. Outlier ini bisa menyebabkan model menjadi tidak representatif terhadap keseluruhan data.

Solusi:

1. Gunakan boxplot untuk mendeteksi outlier
2. Lakukan transformasi logaritmik atau winsorizing
3. Pertimbangkan untuk menggunakan model robust regression

Implikasi Pelanggaran Asumsi

Jika satu atau lebih asumsi dilanggar, model bisa menjadi:

1. Bias: nilai estimasi parameter tidak akurat
2. Tidak efisien: estimasi masih bisa benar, tapi tidak memiliki variansi minimum
3. Tidak valid: kesimpulan uji statistik bisa menyesatkan

Misalnya:

1. Jika error tidak normal: nilai-p dari uji hipotesis bisa salah
2. Jika ada heteroskedastisitas: interval kepercayaan bisa terlalu sempit/lebar.

2.7 Aplikasi Model Regresi Linear Sederhana

1. Prediksi Penjualan Berdasarkan Biaya Iklan

   Salah satu aplikasi paling klasik dan umum dari regresi linear sederhana adalah dalam memodelkan hubungan antara biaya iklan dan penjualan. Dalam hal ini:

   a. Variabel independen (X): Anggaran iklan (dalam juta rupiah).
   b. Variabel dependen (Y): Volume penjualan (dalam unit produk).

   Dengan mengumpulkan data historis dari suatu perusahaan dan menerapkan regresi linear, perusahaan dapat mengetahui:

   c. Seberapa besar dampak tambahan 1 juta rupiah anggaran iklan terhadap penjualan.
   d. Titik jenuh iklan (dengan melihat residual plot atau kurva non-linier).
   e. Prediksi penjualan jika budget iklan diubah (eksperimen simulasi).

   Contoh kasus riil ini digunakan oleh perusahaan FMCG (*Fast Moving Consumer Goods*) untuk mengoptimalkan kampanye pemasaran.

2. Analisis Harga Properti

   Dalam pasar properti, regresi linear sederhana dapat digunakan untuk memprediksi harga rumah berdasarkan satu variabel utama, misalnya luas tanah atau jumlah kamar tidur. Misalnya:

   a. X = Luas bangunan (dalam m^2).
   b. Y = Harga jual rumah (dalam juta rupiah).

   Meskipun dalam praktiknya akan lebih baik menggunakan regresi linear berganda karena banyak faktor lain yang memengaruhi harga (lokasi, umur bangunan, kondisi, dll), namun sebagai pendekatan awal, regresi sederhana sudah cukup membantu dalam:

a. Membuat model valuasi dasar properti.
b. Membandingkan rumah dengan harga wajar di pasar sekunder.

3. Hubungan Antara Indeks Massa Tubuh (IMT) dan Tekanan Darah

   Regresi linear sederhana juga digunakan dalam epidemiologi dan studi kesehatan masyarakat. Sebagai contoh:
   a. X = Indeks Massa Tubuh (BMI).
   b. Y = Tekanan darah sistolik.

   Model ini digunakan untuk melihat apakah ada hubungan linier antara obesitas dengan tekanan darah tinggi. Jika koefisien slope positif dan signifikan, hal itu mendukung hipotesis bahwa peningkatan BMI meningkatkan risiko hipertensi.

   Penggunaan model ini memungkinkan:
   a. Identifikasi pasien berisiko.
   b. Intervensi gizi dan olahraga.
   c. Edukasi kesehatan berbasis data.

4. Prediksi Nilai Ujian Berdasarkan Jam Belajar

   Dalam penelitian pendidikan, model regresi linear sederhana bisa digunakan untuk menjawab pertanyaan seperti: Apakah jumlah jam belajar siswa berkorelasi dengan nilai ujiannya?

   Variabel:
   a. X = Jam belajar per minggu.
   b. Y = Nilai ujian matematika.

   Model ini membantu guru atau institusi untuk menentukan:
   a. Efektivitas strategi belajar.
   b. Rekomendasi waktu belajar minimum.
   c. Penyesuaian kurikulum atau metode pengajaran.

5. Analisis Hubungan Antara Pendapatan dan Tingkat Kebahagiaan

   Dalam psikologi sosial, salah satu pertanyaan klasik adalah: Apakah uang benar-benar bisa membeli kebahagiaan?

   Dengan:
   a. X = Pendapatan bulanan individu.

b. Y = Skor kebahagiaan (skala 1–10).

Model regresi linear sederhana bisa membantu menjawab pertanyaan ini secara data-driven. Namun tentu, korelasi tidak selalu berarti kausalitas. Regresi sederhana memberikan titik awal untuk eksplorasi yang lebih dalam dengan model multivariat.

6. Perkiraan Kekuatan Material Berdasarkan Massa atau Volume

   Dalam ilmu teknik sipil dan material, regresi linear digunakan untuk memperkirakan kekuatan suatu bahan berdasarkan massa jenis, ukuran, atau volume. Misalnya:

   a. X = Volume beton dalam m^3.
   b. Y = Kekuatan tekan beton (MPa).

   Data hasil uji laboratorium dijadikan dasar dalam merancang material yang efisien, kuat, dan sesuai spesifikasi proyek.

   Banyak alat ukur memerlukan kalibrasi, yaitu pencocokan antara nilai aktual dan nilai yang ditunjukkan alat. Regresi linear sederhana digunakan untuk:

   a. Membuat fungsi kalibrasi.
   b. Menyesuaikan pengukuran lapangan.

   Contoh: kalibrasi sensor suhu, pH meter, atau alat ukur tekanan.

DAFTAR PUSTAKA

Efendi, Achmad, dkk. (2020). *Analisis Regresi: Teori dan Aplikasi dengan R*. Yogyakarta: ANDI.

Ghozali, Imam. (2018). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25*. Semarang: Badan Penerbit Universitas Diponegoro.

Gujarati, Damodar N. (2003). *Basic Econometrics (4th Edition).* New York: McGraw-Hill.

Maddala, G.S. (1992). *Introduction to Econometrics (2nd Edition).* New York: Macmillan Publishing Company.

Sugiyono. (2012). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.

Suyono. (2018). *Statistika untuk Penelitian*. Jakarta: Rajawali Pers.

BAB 3
INFERENSI DALAM REGRESI LINEAR SEDERHANA

Oleh Dewi Rakhmatia Nur

3.1 Pendahuluan

Dalam bidang statistik, regresi linear sederhana adalah panduan yang menjelaskan hubungan antara dua variabel, satu variabel sebagai pengubah dan variabel lain sebagai yang diubah Gareth et al., 2021). Anggaplah kita mengetahu bagaimana durasi belajar memengaruhi hasil ujian seorang pelajar, atau bagaimana biaya iklan bisa memperkirakan penjualan barang. Kita mengumpulkan informasi, menganalisis polanya, dan menggambar garis optimum yang mencerminkan hubungan tersebut.

Akan tetapi, informasi yang kita peroleh hanya merupakan sampel atau potongan kecil dari kenyataan yang jauh lebih luas. Pertanyaan penting yang akan timbul adalah dapatkah dari segmen kecil ini bisa diandalkan untuk mempresentasikan keseluruhan? Dapatkah dipercaya bahwa hubungan yang kita lihat dalam sampel berlaku untuk populasi secara keseluruhan?

Disaat inilah inferensi statistik dalam regresi linear sederhana berperan penting dan menjadikannya jembatan yang mengaitkan apa yang ada dan dipercayai. Inferensi memberikan kita kemampuan untuk melampaui sekedar deskripsi, memberika alat untuk menguji hipotesis, memperkirakan parameter populasi dengan tingkat kepercayaan tertentu, dan bahkan dapat membuat prediksi yang relevan. Tanpa penalaran, kita hanya akan memiliki gambar, bukan panduan. Dengan adanya inferensi, kita mulai menyadari jalur yang sebenernya.

3.2 Model Regresi Secara Umum

Sebelum kita memasuki materi inferensi, alakahbaiknya kita mengulas terdahulu tentang model regresi linear sederhana secara umum, yang mana bentuk umum sebagai berikut

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Dimana : Y_i : Hasil observasi pada percobaan ke-i
X_i : Konstanta yang diketahui pada percobaan ke-i
$\beta_0,\ \beta_1$: Parameter
ε_i : Random Error

Persamaan di atas bertujuan untuk menggambarkan perilaku populasi berdasarkan data sampel (Hair et al., 2024). Terdapat beberapa alasan untuk menggunakan asumsi ε_i ini, salah satunya yakni didasari pada distribusi t yang tidak peka terhadap penyimpangan normalitas yang tidak besar.

3.3 Estimasi Parameter

Dalam penerapan, langkah utama dalam inferensi adalah memperkirakan koefisien regresi populasi, yaitu β_0 dan β_1. Namun, kita tidak pernah mengetahui nilai populasi untuk β_0 dan β_1. Oleh karena itu, kita dapat menggunakan data sampel untuk mengestimasi parameter tersebut. Metode yang akan digunakan adalah Metode Kuadrat terkecil. Estimator yang akan disimbolkan dengan Estimator Intersep $(\hat{\beta}_0)$ dan Estimator Slope $(\hat{\beta}_1)$, diketaui dengan meminimalkan jumlah kuadrat terkecil dari residual (SSE) yakni

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} (Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i))^2$$

Jika β_0 dan β_1 masing masing digantikan dengan b_0 dan b_1 serta diibarakkan menjadi nol (James et al., 2021), maka akan diperoleh persamaan dalam bentuk sebagi berikut

$$b_1 = \frac{\sum_{i=n}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=n}^{n}(X_i - \bar{X})^2}$$

$$b_0 = \bar{Y} - \hat{\beta}_1 X$$

Inferensi: b_0 merupakan perkiraan terbaik kita untuk nilai rata-rata dari variabel dependen *(Y)* ketika variabel independen *(X)* sama dengan nol dalam populasi. Sementara itu, b_1 merupakan estimasi optimal kita untuk pergeseran rata-rata dalam *Y* untuk setiap satu unit perubahan dalam *X* di populasi. Perlu diingat bahwa b_0 dan b_1 merupakan estimasi, sehingga memiliki derajat ketidakpastian.

3.4 Asumsi dalam Regresi Linear Sederhana

Inferensi dalam regresi linear sederhana juga meliputi tentang asumsi-asumsi yang akan dilibatkan antara lain linearitas, independensi residual, normalitas dan homoskedatisitas (varians yang konstan)(Falk & Miller, 2021). Asumsi-asumsi tersebut sudah dijelaskan pada Bab sebelumnya .

3.5 Pengujian Hiptesis

3.5.1 Koefisien Regresi β_1

Pengujian ini merupakan salah satu langkah yang penting dengan tujuan untuk menentukan apakah variabel independen (X) memiliki hubungan linear yang signifikan dengan variabel (Y) dalam populasi. Langkah pertama yang dilakukakan dalan pengujiana hipotesis β_1 yakni merumuskan hipotesis nol (H_0) dan hipotesis Alternatif (H_1).

$(H_0): \beta_1 = 0$	pernyataan ini tidak memiliki hubungan linear antara X dan Y dalam populasi
$(H_1): \beta_1 \neq 0$ $(H_1): \beta_1 > 0$ $(H_1): \beta_1 < 0$	Pernyataan ini memiliki hubungan linear yang signifikan antara X dan Y dalam populasi

Untuk mengetahui statistik uji t $((t - Statistic)$

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$$

Yang mana $SE(\hat{\beta}_1)$ merupakan standar error dari $\hat{\beta}_1$ untuk mengukur seberapa ketepatan atau akurasi dari estimasi $\hat{\beta}_1$. s^2 merupakan estimasi tak bias dari varians $error$ (σ^2).

Dalam pengambilan keputusan dalam hipotesis (Gareth et al., 2021).

1. Membandingkan nilai $p - value$ dengan tingkat signifikan (α) yang akan digunakan.
2. Jika $p - value < \alpha$, maka ditolak H_0. Hal tersebut terbukti bahwa ada bukti yang signifikan untuk memberikan kesimpulan yang man variabel X memiliki pengaruh yang signifikan terhadap variabel Y.
3. Jika $p - value \geq \alpha$, mak a diterima H_0. Hal tersebut tidak memiliki cukup bukti untuk membuat kesimpulan bahwa variabel X memiliki pengaruh yang signifikan terhadap variabel Y.

$$SE(\hat{\beta}_1) = \sqrt{\frac{s^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

dengan

$$s^2 = \frac{SSE}{n-2}$$

3.5.2 Intersep $\boldsymbol{\beta_0}$

Pengujian hipotesis β_0 ini tidak menjadi prioritas utama, namun jika intersep memiliki arti kontekstual yang signifikan. Pengujian ini memiliki tujuan untuk menentukan apakah nilai rata-rata Y secara signifikan berbeda dari nol, ketika X adalah nol dalam populasi. Dalam merumuskan hipotesis nol (H_0) dan hipotesis Alternatif (H_1). (Hair et al., 2024)

$(H_0) : \beta_0 = 0$	Pernyataan ini menyatakan bahwa nilai rat-rata Y tidak secara signifikan berbeda dari nol ketika X bernilai di populasi
$(H_1) : \beta_0 \neq 0$	Pernyataan ini memiliki hubungan linear yang signifikan antara X dan Y dalam populasi

Statistik uji yang akan digunakan adalah uji t

$$t = \frac{\hat{\beta}_0 - 0}{SE(\hat{\beta}_0)}$$

3.5.2 Uji F untuk Signifikansi Model Keseluruhan

Uji F dalam regresi linear sederhana untuk signifikansi keseluruhan model akan memberikan nilai$-p$ yang cukup identik dengan uji $-t$ gradien (kemiringan) dari β_1. Hal ini disebabkan karena dalam model yang sederhana, kemiringan menjadi satu-satunya prediktor, sehingga signifikansi sama dengan signifikansi model. Untuk merumuskan hipotesis nol (H_0) dan hipotesis Alternatif (H_1).

$(H_0) : \beta_1 = 0$	Model tidak signifikan
$(H_1) : \beta_1 \neq 0$	Model signifikan

Statistik uji-F dapat dihitung dengan nilai rata-rata kuadrat $MSR\ (Mean\ Square\ Regression)$ dan nilai rata-rata residual $MSE\ (Mean\ Square\ Error)$ menggunakan rumus sebagai berikut

$$F_{hitung} = \frac{MSR}{MSE}$$

Dengan asumsi

$$MSR = \frac{SSR}{df_{reg}} = \frac{SSR}{k}$$

k merupakan jumlah variabel prediktor model regresi. (untuk regresi linear sederhana $k = 1$)

$$MSE = \frac{SSE}{df_{res}} = \frac{SSE}{n-k-1} \qquad SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 \qquad SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

Dengan n merupakan jumlah observasi, SSR merupakan jumlah kuadrat pada model dan SSE merupakan jumlah kuadrat residual, dan SST ialah jumlah kuadrat total variasi pada Y. Statistik F mengikuti distribusi F dengan $df_1 = k$ dan $df_2 = n-k-1$ merupakan derajat kebebasan (Falk & Miller, 2021).

Dalam pengambilan keputusan untuk uji-F antara lain

1. Menentukan taraf signifikansi (α)

2. Jika $p-(sig)$ dari $F_{hitung} < \alpha$, maka H_0 ditolak. Hal ini menyatakan bahwa secara keseluruhan untuk model regresi tidak signifikan.

Sumber variansi	**Derajat kebebasan (df)**	**Jumlah Kuadrat (SS)**	**Rata-rata Kuadrat (MS)**	**F-Hitung**	**Sig. (p-value)**
Regresi	k	SSR	$MSR = \frac{SSR}{k}$	$\frac{MSR}{MSE}$	Nilai-p
Residual	$n-k-1$	SSE	$MSE = \frac{SSE}{n-k-1}$		
Total	$n-1$	SST			

3.6 Interval Kepercayaan

Interval kepercayaan ($convidance\ intervals$) dapat memberikan nilai yang mana kemungkinan besar mengandung parameter dari populasi yang sebenarnya dengan tingkat kepercayaan tertentu. (misal 90% , 95%, 99%) (Acock, 2023).

1. Interval kepercayaan untuk koefisien regresi ($\beta_0\ dan\ \beta_1$)
   Rumus yang akan dogunakan untuk interval kepercayaan β_1
   $$\hat{\beta}_1 \pm t_{\frac{\alpha}{a},n-2} \times SE(\hat{\beta}_1)$$
   Dengan $t_{\frac{\alpha}{a},n-2}$ sebagai nilai kritis dari tabel distribusi t dengan derajat kebebasan pada tingkat taraf signifikansi dari α.
2. Interval Kepercayaan untuk rata-rata response
   $$\hat{Y}_{X_0} \pm t_{\frac{\alpha}{a},n-2} \times SE(\hat{Y}_{X_0})$$
   Dengan $SE(\hat{Y}_{X_0})$ sebagai standar error dari prediksi rata-rata.
3. Interval Prediksi untuk respons individu
   $$\hat{Y}_{X_0} \pm t_{\frac{\alpha}{a},n-2} \times SE_{pred}(\hat{Y}_{X_0})$$

Dengan $SE_{pred}(\hat{Y}_{X_0})$ sebagai standar error dari prediksi Individu.

3.7 Koefisien Determinasi

Koefisian determinasi ini digunakan untuk mengukur sebuah proporsi variansi total dalam variabel respon (Y) yang dijelaskan

oleh variabel prediktor (X) yang mana memiliki model regresinya (Wooldridge, 2013).

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Dalam nilai R^2 terdapat kisaran antara 0 sampai dan 1
$R^2 = 0$ (tidak dapat menjelaskan variabel apapun dalam Y)
$R^2 = 1$ (menjelaskan semua variasi dalam Y)

Contoh Soal

Terdapat sebuah tabel berisi data yang menyatakan nilai hasil belaja mahasiswa dan jam belajar mahasiswa.

Nomor responden	Jam Belajar (X)	Hasil Belajar Mahasiswa (Y)
1	1	50
2	1,5	40
3	2	60
4	2	65
5	2,5	75
6	3	70
7	3	80
8	3,5	85
9	4	84
10	4,5	90

Dari data di atas kita dapat menghitung secara manual atau menggunakan Microsoft Excel, sehingga diperoleh

$$\sum_{i=1}^{n} X_i = 27\,,\ \sum_{i=1}^{n} Y_i = 699\,,\ \sum_{i=1}^{n} {X_i}^2 = 84,\ \sum_{i=1}^{n} {Y_i}^2 = 51321,\ \sum_{i=1}^{n} X_i Y_i = 2036$$

Dengan menggunakan rumus pada Bab sebelumnya, maka akan diperoleh nilai

$$b_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} {X_i}^2 - (\sum_{i=1}^{n} X_i)^2}$$

$$b_0 = \frac{\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i}{n}$$

Maka nilai $b_0 = 33{,}7297$ dan $b_1 = 13{,}3963$, dan didapatkan persamaan garis regresinya adalah

$$\hat{Y} = 33{,}7297 + 13{,}3963X$$

Untuk menguji kesesuaian model regresi, Langkah pertama kita harus merumuskan hipotesis nol (H_0) dan hipotesis alternatif (H_1) terlebih dahulu

H_0: $\beta_0 = \beta_1 = 0$ (Model tidak sesuai)
H_1: Paling sedikit ada satu $\beta \neq 0$ (Model sesuai)

Langkah kedua melakukan peritungan untuk statistik uji, yang mana dapat dibantu dengan tabel sepeti di bawah

i	X	Y	$\hat{Y}$	$(Y_i - \hat{Y})^2$	$(\hat{Y}_i - \bar{Y})^2$	$(Y_i - \bar{Y})^2$
1	1	50	47,12613	8,259151	518,6493	396,01
2	1,5	40	53,82432	191,1119	258,4273	894,01
3	2	60	60,52252	0,27303	87,93708	98,01
4	2	65	60,52252	20,0478	87,93708	24,01
5	2,5	75	67,22072	60,51719	7,178537	26,01
6	3	70	73,91892	15,35793	16,15171	0,01
7	3	80	73,91892	36,97955	16,15171	102,01
8	3,5	85	80,61712	19,20966	114,8566	228,01
9	4	84	87,31532	10,99132	303,2932	198,81
10	4,5	90	94,01351	16,10829	581,4615	404,01

Dari tabel diatas diperoleh, nilai jumlah kuadrat residual yaitu

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = 378{,}8559$$

Nilai jumlah kuadrat regresi

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = 1992{,}044$$

Dan nilai jumlah kuadrat total

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = 2370{,}9$$

Dilanjutkan dengan menghitung rata-rata kuadrat regresi dengan rumus

$$MSR = \frac{SSR}{df_{reg}} = \frac{SSR}{k} = \frac{1992{,}044}{1} = 1992{,}044$$

Menghitung rata-rata kuadrat residual

$$MSE = \frac{SSE}{df_{res}} = \frac{SSE}{n-k-1} = \frac{378{,}8559}{10-1-1} = 47{,}3569$$

Menghitung statistik uji F

$$F_{hitung} = \frac{MSR}{MSE} = \frac{1992{,}044}{47{,}3569} = 42{,}0644$$

Selanjutnya Langkah ketiga untuk menentukan F tabel, diasumsikan dengan menggunakan tingkat taraf signifikan $\alpha = 5\% = 0{,}05$. Dengan derajat kebebasan $df_1 = k = 1$ dan $df_2 = n - k - 1 = 8$ maka diperoleh nilai pada F tabel = 5,32.

Titik Persentase Distribusi F untuk Probabilita = 0,05

df untuk penyebut (N2)	df untuk pembilang (N1)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161	199	216	225	230	234	237	239	241	242	243	244	245	245	246
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42	19.43
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85

Hasil perhitungan bisa disajikan dalam bentuk tabel Anava seperti di bawah

Sumber variansi	**Derajat kebebasan (df)**	**Jumlah Kuadrat (SS)**	**Rata-rata Kuadrat (MS)**	**F-Hitung**	**Sig. (p-value)**
Regresi	1	1992,044	1992,044	47,3569	5,32
Residual	8	378,8559	47,3569		
Total	9	2370,9			

Langkah keempat yaitu membuat kesimpulan, yang mana $F_{hitung} = 47{,}3569$ dan $F_{tabel} = 5{,}32$, maka diperoleh

$$F_{hitung} > F_{tabel} = 47{,}3569 > 5{,}32$$

dan H_0 ditolah dan H_1 diterima.

Sehingga model regresi linier sederhana sesuai untuk menyatakan ada hubungan antara jam belajar (X) dengan hasil belajar mahasiswa.

Untuk mencari interval kepercayaan dengan menggunakan rumus sebagai berikut:

$$\hat{\beta}_1 \pm t_{\frac{\alpha}{2},n-2} \times SE(\hat{\beta}_1)$$

Dari contoh di atas diperoleh

$$b_0 = 33{,}7297, \quad b_1 = 13{,}3963, \quad \sum_{i=1}^{n} X_i = 27, \quad \sum_{i=1}^{n} {X_i}^2 = 84,$$

$$SE(\hat{\beta}_1) = \sqrt{\frac{s^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}, \quad s^2 = \frac{SSE}{n-2}$$

$$s^2 = \frac{378{,}8559}{10-2} = 47{,}3569$$

$$SE(\hat{\beta}_1) = \sqrt{\frac{47{,}3569}{11{,}1}} = 2{,}0655$$

Dengan menggunakan $\alpha = 0{,}05$ didapat derajat bebas, $t_{\frac{0{,}05}{2},10-2} = t_{0{,}025;8} = 2{,}3060$, maka

$$13{,}3963 \pm 2{,}3060 \times 2{,}0655$$

$$13{,}3963 + 2{,}3060 \times 2{,}0655 = 18{,}1593$$

$$13{,}3963 - 2{,}3060 \times 2{,}0655 = 8{,}6332$$

$$8{,}6332 \leq \hat{\beta}_1 \leq 18{,}1593$$

Koefisien determinasi

$$R^2 = \frac{SSR}{SST} = \frac{1992{,}044}{2370{,}9} = 0{,}8402$$

Latihan

Terdapat data senyawa kimia dengan berbagai macam temperatur

X	0	15	30	45	50	75
Y	8	12	24	33	39	44

Tentukan :

a. Carilah persamaan regresi linear sederhana
b. Nilai s^2
c. Interval kepercayaan dengan tingkat signifikansi 90% dan 95%
d. Koefisien determinasi

DAFTAR PUSTAKA

Acock, A. C. (2023). *A Gentle Introduction to Stata* (7th ed.). Stata Press.

Agus, B. T., & Nano, P. (2016). Analisis Regresi Dalam Penelitian Ekonomi Dan Bisnis: Dilengkapi Aplikasi Spss Dan Eviews. Raja Grafindo Persada, Jakarta

Falk, R. F., & Miller, B. C. (2021). *A Primer for Model-Based Regression*. SAGE Publications.

Gareth, J., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2024). *Multivariate Data Analysis* (8th ed.). Cengage Learning. (Edisi terbaru)

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer

Keller, G., Rotman, J. L., Brazil Japan, A., & Singapore, M. (n.d.). *~IO pee M A N A C ie ST AT IS ABBREVIATED SOUTH-WESTERN ry o ¢ = CENGAGE Learning"*.

Wijaya, E., Indriyati, R., Rinawati, R., Utami, R. N., Negsih, T. A., Suharyanto, S., Hermawan, E., Deseria, R., Aziza, N., & Judijanto, L. (2024). Pengantar Statistika: Konsep Dasar Untuk Analisis Data. Manado: Pt. Sonpedia Publishing Indonesia.

Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). South-Western Cengage Learning.

Yordan, A., Putri, T. N., & Lamkaruna, D. H. (2019). Peramalan Penerimaan Mahasiswa Baru Universitas Samudra Menggunakan Metode Regresi Linear Sederhana. Jurnal Teknik Informatika (J-Tifa), 2(1), 21–27.

BAB 4
ANALISIS VARIANSI DALAM REGRESI

Oleh Yunita Widia Putri

4.1 Pengertian Analisis Variansi dalam Regresi

Analisis variansi atau *Analysis of Variance* (ANOVA) merupakan teknik analisis statistik yang pertama kali dikembangkan oleh Sir Ronal Aylmer Fisher pada tahun 1920 (Das, 2022). ANOVA digunakan untuk mengetahui perbedaan rata-rata antar kelompok atau populasi yang diobservasi. ANOVA merupakan perluasan uji t sehingga dapat digunakan dalam pengujian perbedaan tiga buah rata-rata populasi atau lebih sekaligus. Teknik analisis ini bertujuan untuk mengetahui perbedaan skor variabel terikat atau dependen yang disebabkan oleh adanya perbedaan nilai pada setiap variabel bebas atau independent (Blanca, 2023).

Dalam melakukan uji ANOVA terdapat beberapa syarat yang harus dipenuhi, yakni:

1. Sampel diambil secara Acak *(Random Sampling)*
   Sampel harus diambil secara acak dari populasi, sehingga setiap anggota populasi memiliki peluang yang sama untuk terpilih menjadi sampel. Dengan begitu, hasil analisis data yang dilakukan dapat digenerasilasi ke dalam populasi yang lebih luas.
2. Data Berdistribusi Normal
   Data setiap kelompok harus berdistribusi normal, dengan ciri data akan membentuk kurva lonceng saat divisualisasikan.
3. Data saling Bebas
   Data dalam setiap kelompok harus saling bebas, yang artinya pada satu pengamatan tidak boleh mempengaruhi nilai pada pengamatan lain dalam kelompok yang sama atau kelompok lain.

4. Homogenitas Variansi
   Variansi atau perbedaan antar data dalam satu kelompok pada setiap kelompok yang dibandingkan harus sama atau homogen.

Analisis Variansi dalam regresi digunakan untuk menguji signifikansi dari suatu model regresi secara keseluruhan (Das, 2022). ANOVA membandingkan variasi total dalam data dengan variasi yang dijelaskan oleh model regresi dan variasi yang tidak dapat dijelaskan atau residual. Dalam hal ini, ANOVA dapat menentukan apakah variabel bebas atau independen secara signifikan mempengaruhi variabel dependen. Selain itu, ANOVA dalam regresi dapat digunakan juga untuk menilai kontribusi variabel independen terhadap variabel dependen dan mengidentifikasi apakah model regresi yang digunakan dapat menjelaskan variasi dalam data dengan baik.

4.2 Tabel ANOVA untuk Regresi

Tabel ANOVA untuk regresi terdiri dari beberapa komponen yang memberikan utama yang memberikan informasi tentang variasi data dan bagaimana model regresi menjelaskan variasi tersebut. Komponen-komponen tersebut meliputi:

1. Sumber Variasi
   Sumber variasi dalam tabel ANOVA dibagi menjadi dua kategori utama:
   a. Varians Antara *(Between-Group Variance)*
      Varians antara adalah variasi yang dijelaskan oleh model regresi yang mencakup variasi yang disebabkan oleh variabel independen yang digunakan dalam model. Varians ini mengukur seberapa jauh rata-rata prediksi model berbeda dari rata-rata total.
   b. Varians Dalam *(Within-Group Variance)*
      Varians dalam adalah variasi yang tidak dijelaskan oleh model regresi yang mencakup kesalahan atau residual dari model. Varians ini mengukur variasi di dalam kelompok data yang sama.

2. Derajat Kebebasan atau *Degree of Freedom (df)*
   Derajat kebebasan merupakan ukuran jumlah informasi yang tersedia untuk memperkirakan varians. Dalam tabel ANOVA dalam regresi, derajat kebebasan dibagi menjadi dua bagian, yaitu:
   a. Derajat Kebebasan Regresi (df Regresi)
      Derajat kebebasan antara dihitung sebagai jumlah variabel independen dalam model. Jika ada k variabel independen, maka df antara adalah k.
   b. Derajat Kebebasan Residual/Error (df Residual)
      Derajat kebebasan residual dihitung sebagai jumlah total observasi dikurangi jumlah parameter yang diestimasi. Jika ada n total observasi dan k variabel independen, maka:
      $df\ residual = n - k - 1$ (Montgomery, 2021)
   c. Derajat Kebebasan Total (df Total)
      Derajat kebebasan total merupakan total observasi dalam data dikurangi 1. Dikurangi 1 karena menggunakan rata-rata sampel sebagai estimasi pusat data, sehingga satu derajat tersebut ($\bar{Y}$) terpakai untuk mengestimasi mean. Oleh karena itu, dengan df total dapat ditunjukan seberapa banyak data dapat bervariasi secara bebas setelah memperhitungkan rata-rata sampel. Df total dituliskan sebagai berikut:
      $df\ total = n - 1$ (Montgomery, 2021)

   Contoh:
   Model regresi untuk prediksi harga rumah berdasarkan luas tanah. dengan total observasi adalah 5.
   Penyelesaian:
   k : 1
   n : 30

Tabel 4.1. Menghitung df

Sumber Variasi	**df**	**Rumus**	**Nilai**
Regresi	df regresi	k	1
Residual	df residual	$n - k - 1$	3
Total	df total	$n - 1$	4

Verifikasi:

$df\ total = df\ regresi + df\ residual$

$df\ total = 1 + 3$

$df\ total = 4$

3. Jumlah Kuadrat/*Sum of Square* (SS)

Jumlah kuadrat merupakan total kuadrat dari deviasi antara nilai observasi dengan nilai rata-rata atau nilai prediksi (Khasanah, 2021). Dalam regresi, jumlah kuadrat digunakan untuk mengukur variasi total dalam data variabel dependen dan membaginya menjadi variasi yang dijelaskan oleh model dan variasi yang tidak dijelaskan (residual/error). Macam-macam jumlah kuadrat dalam regresi dijelaskan sebagai berikut:

a. Jumlah Kuadrat Regresi (SSR)

Jumlah kuadrat regresi digunakan untuk menunjukkan seberapa baik model regresi menjelaskan variasi data. Semakin besar nilai jumlah kuadrat regresi, maka semakin baik model regresi dalam menjelaskan variasi data. Jumlah kuadrat regresi dihasilkan oleh prediksi model terhadap nilai rata-rata yang dituliskan dengan rumus sebagai berikut:

$SSR = \sum_{i=1}^{n}\left(\widehat{Y}_i - \bar{Y}\right)^2$ (Montgomery, 2021)

Dimana,

$\widehat{Y}_i$: Nilai prediksi dari model regresi

$\bar{Y}$: Rata-rata observasi

b. Jumlah Kuadarat Residual/Error (SSE)

Jumlah kuadrat residual adalah variasi yang tidak dapat dijelaskan oleh model, yaitu selisih kuadrat antara nilai observasi dengan nilai prediksi. Jumlah kuadrat residual mencerminkan kesalahan model regresi dalam menjelaskan data. Semakin kecil nilai jumlah kuadrat residual, maka semakin akurat model. Rumus jumlah kuadrat residual dituliskan dengan rumus sebagai berikut:

$SSE = \sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2$ (Montgomery, 2021)
Dimana:
Y_i : Nilai observasi
$\widehat{Y}_i$: Nilai prediksi dari model regresi

c. Jumlah Kuadrat Total (SST)
Jumlah kuadrat total menguku total variasi dalam data variabel dependen (Y), dihitung sebagai jumlah kuadrat selisih antara setiap nilai observasi dengan rata-rata nilai observasi, yang dituliskan sebagai berikut:
$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = SSR + SSE$ (Montgomery, 2021)
Dimana:
Y_i : Nilai observasi
$\bar{Y}$: Rata-rata observasi
SSR : Jumlah kuadrat regresi
SSE : Jumlah kuadrat error

Contoh:
Tentukan jumlah kuadrat regresi, residual, dan total untuk prediksi harga rumah berdasarkan luas tanah. Total observasi adalah 5. Dengan data sebagai berikut:

Tabel 4.2. Contoh Data

No	**Luas Tanah (X) dalam m²**	**Harga Rumah (Y) dalam ratus juta**
1	10	100
2	20	175
3	30	250
4	40	325
5	50	400

Penyelesaian:

a. Mengitung Slope (b_1)

Tabel 4.3. Menghitung Nilai Rata-Rata X dan Y

X	Y	$\bar{X}$	$\bar{Y}$	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})^2$	$(X-\bar{X})(Y-\bar{Y})$
10	100	30	250	-20	-150	400	3.000
20	175			-10	-75	100	750
30	250			0	0	0	0
40	325			10	75	100	750
50	400			20	150	400	3.000
						1.000	7.500

$$b_1 = \frac{\sum(X_i - \bar{X})\,(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2}$$

$$b_1 = \frac{7.500}{1.000}$$

$$b_1 = 7{,}5$$

Maka, nilai prediksi ($\hat{Y}$) = 7,5X

b. Menghitung Jumlah Kuadrat

Tabel 4.4. Menghitung Jumlah Kuadrat

X	Y	$\bar{Y}$	$\hat{Y}$	$(\hat{Y}-\bar{Y})^2$	$(Y-\hat{Y})^2$	$(Y-\bar{Y})^2$
10	100	250	75	30.625	625	22.500
20	175		150	10.000	625	5.625
30	250		225	625	625	0
40	325		300	2.500	625	5.625
50	400		375	15.625	625	22.500
				59.375	3.125	56.250

$$SSR = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 = 56.250$$

$$SSE = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = 3.125$$

$$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = SSR + SSE = 59.375$$

4. *Mean Square (MS)*
   Mean square merupakan ukuran varians yang dihitung dengan membagi jumlah kuadrat dengan derajat kebebasan yang sesuai. Mean square dibagi menjadi dua, yaitu:
   a. *Mean Square* Regresi *atau Mean Square Regression (MSR)*
      Mean square regresi dihitung dengan membagi jumlah kuadrat *(Sum of Square)* antara dengan df antara, yang dirumuskan sebagai berikut:
      $MSR = \frac{SSR}{k}$ (Montgomery, 2021)

   b. *Mean Square* Residual *atau Mean Square Error (MSE)*
      Mean square error dihitung dengan membagi jumlah kuadrat residual/error dengan df residual, yang dirumuskan sebagai berikut:

      $MSW = \frac{SSE}{n-k-1}$(Montgomery, 2021)

5. Uji F
   Nilai F merupakan rasio antara MSR dan MSE yang digunakan untuk menguji hipotesis apakah model regresi dapat menjelaskan variasi variabel dependen secara signifikan (Blanca, 2023). F-hitung dihitung dengan rumus:

   $F = \frac{MSR}{MSE}$ (Montgomery, 2021)

   Jika nilai F-hitung lebih besar dari nilai F-tabel pada tingkat signifikansi tertentu, maka model regresi dianggap signifikan.
6. P-Value
   P-value digunakan sebagai alternatif dalam menguji hipotesis selain dengan membandingkan F-hitung dengan F-tabel. Untuk menguji hipotesis dengan menggunakan p-value, maka p-value akan dibandingkan dengan taraf signifikansi tertentu yang telah ditetapkan. Jika nilai p-value lebih kecil dari taraf signifikansi (α) yang telah ditentukan maka hipotesis nol ditolak dan hipotesis alternatif diterima

(Arraniri, 2023), sehingga ada pengaruh yang signifikan dari variabel independen terhadap variabel dependen. Pada tabel ANOVA nilai p-value ditunjukan pada kolom sig.

Contoh:

Tentukan *mean square* regresi, residual, dan total serta nilai F-hitung untuk prediksi harga rumah berdasarkan luas tanah. Total observasi adalah 5. Dengan data sebagai berikut:

Tabel 4.5. Nilai df dan Jumlah Kuadrat

Sumber Variasi	df	Nilai Jumlah Kuadrat
Regresi	1	56.250
Residual	3	3.125
Total	4	59.375

Penyelesaian:

Tabel 4.6. Menghitung Mean Square

Sumber Variasi	Rumus Mean Square	Nilai
Regresi	$\frac{SSR}{k} = \frac{56.250}{1}$	56.250
Residual	$\frac{SSE}{n-k-1} = \frac{3.125}{3}$	1.041,67

Menentukan nilai F-hitung

$$F = \frac{MSR}{MSE}$$

$$F = \frac{56.250}{1.041,67}$$

$$F = 54$$

4.3 Interpretasi Analisis Variansi dalam Regresi

Dari contoh sebelumnya, hasil analisis data dengan sau variabel bebas dapat kita ringkas dalam tabel ANOVA sebagai berikut:

Tabel 4.7. ANOVA

Sumber Variasi	**df**	***SS***	***MS***	**F-hitung**	**Sig. (estimasi)**
Regresi	1	56.250	56.250	54	< 0.05
Residual	3	3.125	1.041,67		
Total	4	59.375			

Tabel di atas akan dipakai untuk menguji hipotesis berikut:

H_0 : Tidak ada pengaruh yang signifikan luas tanah terhadap harga tanah.

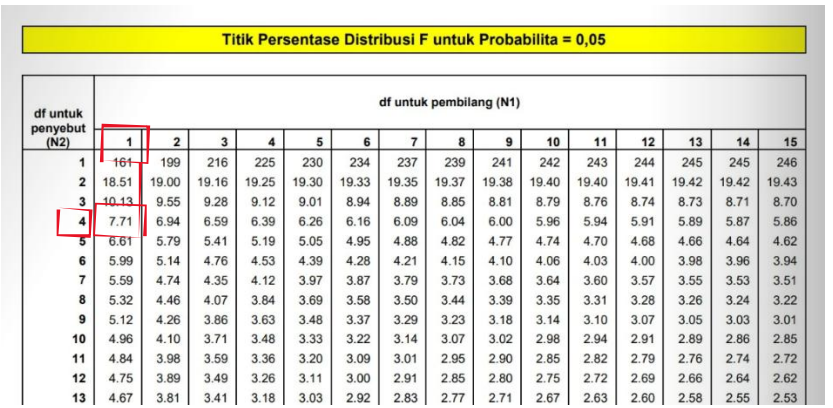
H_1 : Ada pengaruh yang signifikan luas tanah terhadap harga tanah

Dengan menggunakan tabel tersebut, dapat dilakukan pengujian hipotesis. Sebagai contoh akan digunakan taraf signifikansi (α) sebesar 5%, artinya tingkat kepercayaannya adalah 95%. Pada tabel diketahui bahwa nilai F-hitung adalah 54. Nilai tersebut selanjutnya akan dibandingkan dengan nilai F-tabel. Jika nilai F-hitung lebih besar dari nilai F-tabel pada tingkat signifikansi 5%, maka terdapat pengaruh yang signifikan luas tanah terhadap harga tanah.

Untuk mengetahui nilai F-tabel dapat dilakukan dengan mencari nilai F-tabel pada tabel F untuk tingkat signifikansi 5% yang perlu dilakukan adalah mencari nilai df untuk pembilang dan df untuk penyebut.

a. Df untuk pembilang
   Df untuk pembilang dihasilkan dari $k - 1$, karena pada contoh $k = 1$ dan bila dikurangkan 1 hasilnya 0 maka diasumsikan nilai df untuk pembilang adalah 1.
b. Df untuk penyebut
   Df untuk penyebut dihasikan dari $n - k$. Pada contoh sebelumnya diketahui bahwa nilai $k = 1$ dan $n = 5$, maka nilai df untuk penyebut adalah $n - k = 5 - 1 = 4$.

Setelah mengetahui nilai df untuk pembilang dan nilai untuk df untuk penyebut, selanjutnya memeriksa pada tabel F taraf signifikansi 5% untuk mengetahui nilai F-tabel. Hal tersebut digambarkan pada gambar berikut.



Titik Persentase Distribusi F untuk Probabilita = 0,05

df untuk penyebut (N2)	df untuk pembilang (N1)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161	199	216	225	230	234	237	239	241	242	243	244	245	245	246
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40	19.41	19.42	19.42	19.43
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53

Gambar 4.1. Cara Menentukan F-tabel
Sumber: Nasar, 2024

Pada gambar di atas diketahui bahwa nilai f-tabel sebesar 7,71. Jika kita bandingkan dengan nilai f-hitung, maka f-hitng=54 > dari f-tabel=7,71. Dengan demikian, terdapat pengaruh yang signifikan luas tanah terhadap harga tanah.

Pada kolom sig, pada Tabel 4.7. ANOVA, nilai p-value kurang dari 0.05, sehingga dengan hipotesi nol (tidak ada pengaruh signifikan luas tanah terhada harga tanah) ditolak dan hipotesis alternatif (ada pengaruh signifikan luas tanah terhadap harga tanah) diterima.

4.4 Aplikasi untuk Analisis Variansi

Pada contoh sebelumnya, sudah dilakukan pengolahan data analisis varians secara manual. Pengolahan data tersebut dapat juga dibantu dengan aplikasi yang lebih canggih sehingga memudahkan dalam melakukan analisis data dan visualisasi data (Aspriyani, 2022). Beberapa aplikasi yang dapat membantu dalam pengolahan data analisis varians dalam regresi antara lain:

1. Microsoft Excel
2. JASP
3. R & RStudio
4. Python
5. SPSS

Dari kelima aplikasi di atas JASP, R & Rstudio, dan Python dapat diakses secara gratis. Untuk R & Rstudio dan Python kita memerlukan pemrograman dasar untuk melakukan pengolahan data. Pada Python modul yang digunakan untuk menganalisis data contohnya seperti NumPy, Pandas, Statmodels, dan Matplotlip (Chandra, 2025).

DAFTAR PUSTAKA

Aspriyani, R., Hartono, B.P., Ahmad, M. and Susilowati, E., 2022. Implementasi Spss Dalam Analisis Data Bagi Mahasiswa Di Cilacap. *Jurnal Terapan Abdimas*, *7*(2), pp.230-237.

Blanca, M.J., Arnau, J., García-Castro, F.J., Alarcón, R. and Bono, R., 2023. Repeated measures ANOVA and adjusted F-tests when sphericity is violated: which procedure is best?. *Frontiers in Psychology*, *14*, p.1192453.

Candra, A.P., 2025. Analisis Data Menggunakan Python: Memperkenalkan Pandas dan NumPy. *Journal of Information System and Education Development*, *3*(1), pp.11-16.

Das, B.K., Jha, D.N., Sahu, S.K., Yadav, A.K., Raman, R.K. and Kartikeyan, M., 2022. Analysis of variance (ANOVA) and design of experiments. In *Concept Building in Fisheries Data Analysis* (pp. 119-136). Singapore: Springer Nature Singapore.

Arraniri, I., Sabtohadi, J., Suhartawan, B., Sarie, F., Abubakar, R, Maryani, S., Hikmah, Rahayu, B., and Korosand., 2023. *Pengantar Statistika*. Cendikia Mulia Mandiri.

Khasanah, U., 2021. *Analisis Regresi*. Uad Press.

Montgomery, D.C., Peck, E.A. and Vining, G.G., 2021. *Introduction to linear regression analysis*. John Wiley & Sons.

Nasar, A., Saputra, D.H., Arkaan, M.R., Ferlyando, M.B., Andriansyah, M.T. and Pangestu, P.D., 2024. Uji Prasyarat Analisis. *Jurnal Ekonomi Dan Bisnis*, *2*(6), pp.786-799.

BAB 5
REGRESI LINEAR BERGANDA

Oleh Tedi Hartoyo

5.1 Pendahuluan

Regresi adalah hubungan fungsional antara dua atau lebih variabel. Pentingnya mempelajari hubungan antara dua buah atau lebih variabel adalah untuk mengetahui (menentukan, meramal atau menilai) nilai-nilai variabel yang satu, jika diketahui nilai-nilai variabel yang lainnya. Variabel yang dihubungkan dalam regresi ini biasa dinamaka variabel tidak bebas (*dependent variable*), sedangkan variabel yang lainnya, yang digunakan untuk pendugaan disebut variabel bebas (*independent variable*).

Hubungan fungsional antara variabel tidak bebas (Y) dengan variabel bebas (X) dapat ditulis sebagai berikut Y = f(X). Beberapa contoh penerapannya antara lain:

1. Hubungan antara tinggi badan anak dengan tinggi badan ayah.
2. Menduga hubungan antara konsumsi dengan pendapatan.
3. Mencari hubungan antara jumlah ekspor salak dengan PDRB (Produk Domestik Regional Bruto).
4. Menduga hubungan antara jumlah penduduk dengan permintaan suatu produk.
5. Hubungan antara produksi (hasil panen padi) dengan faktor-faktor produksi (pupuk, tenaga kerja, luas lahan,jumlah benih, dan seterusnya).
6. Hubungan antara produktivitas tenaga kerja dengan faktor-faktor produktivitas (usia, jumlah tanggungan keluarga, jarak tempuh, jumlah insentif, dan seterusnya).

Regresi linear berganda untuk mencari hubungan antara sebuah variabel tidak bebas (Y) dengan lebih dari satu variabel bebas (X_1,X_2,X_3,…), sehingga muncul istilah *bivariable* atau *multivariable*. Selanjutnya untuk kebutuhan mencari hubungan kedua variabel tersebut harus dilakukan pengamatan terhadap sejumlah sampel

atau populasi. Pengamatan tersebut merupakan pasangan nilai dari variabel tidak bebas dan variabel bebas. Tampilan pasangan variabel antara sebuah variabel tidak bebas dengan lebih dari satu variabel bebas, dapat dilihat pada tabel di bawah ini:

Nomor Responden	**X1**	**X2**	**X3**	**. . .**	**Y**
1	X_{11}	X_{21}	X_{31}		Y_1
2	X_{12}	X_{22}	X_{32}		Y_2
3	X_{13}	X_{23}	X_{33}		Y_3
4	X_{14}	X_{24}	X_{34}		Y_4
.	.	.	.		.
.	.	.	.		.
.	.	.	.		.
n	X_{1n}	X_{2n}	X_{3n}		Y_n

Garis Regresi merupakan garis penduga hubungan antara variable tidak bebasl Y dengan variabel bebas X. Dalam bentuk matematika hubungan tersebut dapat ditampilkan balam bentuk fungsi sebagai berikut:

Regresi berganda (*Multiple regression*), terdiri dari dua atau lebih variabel X dan satu variabel Y, dengan formulasi sebagai berikut: $Y = f(X_1, X_2, ... X_n)$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + e$$

Keterangan:

β_0 = perpotongan dengan sumbu Y (intercept)

β atau $\beta_1, \beta_2, ...$ = koefisien regresi

e = error term (galat)

Dalam fungsi statistik dimana dinyatakan hubungan antara X dan Y, pada umumnya titik-titik pada diagram tidak tepat berada pada garis regresinya, berbeda dengan fungsi matematika. Jadi setiap titik mempunyai simpangan atau deviasi (*deviation*) terhadap garis regresinya. Besar simpangan pada setiap titik itu menunjukkan keeratan hubungan antara kedua variabel tersebut.

Perbedaan antara hubungan matematik (fungsi matematik) dan hubungan statistik (fungsi statistik) adalah fungsi matematik tidak mempunyai deviasi, sedangkan fungsi statistik mempunyai deviasi. Dalam fungsi statistik, tujuannya adalah mencari deviasi sekecil mungkin atau mencari jumlah kuadrat deviasi sekecil mungkin. Dengan menggunakan metode kuadrat deviasi atau sisaan terkeci, sehingga muncul formulasi untuk mencari koefisien regresi. Berdasarkan bentuk kurvanya, regresi dapat dibedakan menjadi:

1. **Regresi linear**
   a. Regresi linear sederhana
   b. Regresi linear berganda
2. **Regresi non linear**
   a. Regresi kuadratik
   b. Regresi kubik
   c. Regresi eksponensial
   d. Regresi logaritma

Model regresi berganda seperti halnya model regresi sederhana, akan tetapi model regresi ini menggambarkan hubungan antara satu variabel tidak bebas (Y) dengan dua atau lebih variabel bebas (X). Asumsi-asumsi klasik yang harus dipenuhi sama seperti yang berlaku untuk regresi sederhana. Selanjutnya untuk menggunakan regresi berganda ini dituntut untuk menguasai atau memahami tentang matrik, karena dalam proses pendugaan selalu menggunakan matriks.

5.2 Model regresi

$Y_1 = B_o + B_1X_{11} + B_2X_{21} + B_3X_{31} \ldots..B_kX_{k1} + e_1$

$B_0, B_1, \ldots\ldots B_k$ = Parameter ,koefesien regresi

Y =Variabel tidak bebas (*dependent*)

$X_1 \ldots X_k$ = Varariabel bebas (*independent*)

i = 1,2,……. n, indeks untuk sampel

Model diatas dapat dijabarkan :

$Y_1 = B_o + B_1 + X_{11} + B_0X_{21} + B_2X_{21} + \ldots\ldots\ldots\ldots\ldots B_kX_{k1} + e_1$

$Y_1 = B_o + B_1 + X_{12} + B_0X_{22} + B_2X_{22} + \ldots\ldots\ldots\ldots\ldots B_kX_{k2} + e_2$

$Y_1 = B_o + B_1 + X_{1n} + B_0X_{2n} + B_2X_{2n} + \ldots\ldots\ldots\ldots\ldots B_kX_{kn} + e_n$

Dalam bentuk matriks dapat dikemukakan Sbb :

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ . \\ . \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \ldots\ldots X_{1k} \\ 1 & X_{21} & X_{22} & \ldots\ldots X_{2k} \\ 1 & X_{31} & X_{32} & \ldots\ldots X_{3k} \\ . & & & \\ . & & & \\ . & & & \\ . & & & \\ 1 & X_{n1} & X_{n2} & \ldots\ldots X_{nk} \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \\ B_2 \\ . \\ . \\ . \\ B_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \\ . \\ . \\ . \\ e_k \end{bmatrix}$$

Secara ringkas :
[Y] = [X] [B] + [e]
X_o = 1 untuk setiap nilai ke i
1. Metode kuadrat terkecil :
(Y) = (X) (B) + (e)
(e) = (Y) – (X) (B)
$\sum e_1^2$ = [Y – (X) (B)]' [(Y) – (X) (B)]
= (Y'Y) – 2 (B) (X'Y) + (B)' (X'X) (B)
Dalam hal ini (X) (B) = (B)' (X)' dan
(B)' (X' Y) = (Y)' (X) (B)

Untuk mendapatkan hasil estimasi (B) = $b_o, b_1 \ldots\ldots b_k$, Diperlukan metode kuadrat terkecil yaitu dengan meminimumkan $\sum e_1^2 = e'e$

Deviasi e'e terhadap (B) harus = 0 untuk mendapatkan nilai minimumnya . jadi
d(e'e)/d(B) = 2 (X' Y) + 2 (X' X) (B) = 0
Akhirnya akan diperoleh :
(X' Y) =(X'X) (B) ini sebagai persamaan normal ke 1
Selanjutnya akan diperoleh :
$(B) = (X' X)^{-1} (X' Y)$ ini sebagai persamaan normal ke 2
Dalam estimasi koefesien regresi :
(1). Persamaa normal ke 1, penyelesaian akan menggunakan metode DOOLITTLE singkat.

(2). Persamaan normal ke 2, penyelesaian akan menggunakan metode LAPLACE dengan mencari matriks kebalikan (*invers*) (X 'X) atau biasa diberi simbol sebagai berikut $(X'X)^{-1}$

Dalam masalah penyelesaian regresi berganda, untuk mencari koefisien regresi dan pengujiannya akan digunakan persamaan normal ke 2 dengan menggunakan metode Laplace

5.3 Metode LAPLACE

Langkah-langkah penyelesaian dengan menggunakan Metode Laplace, adalah sebagai berikut :

1. Estimator pertama kali diperoleh adalah dengan indeks kecil , yaitu : b_o , b_1 .....terakhir b_k
2. Langkah penyelesaiaanya merupakan langkah maju
3. Pengujian signifikansi parameter, terutama menggunakan uji-t untuk setiap parameter.

Procedure atau langkah penyelesaiaan menurut metode Laplace adalah:

1. Hitung determinan dari matriks $(X' X)$
2. Tentukan elemen-elemen matrik minor dan kofaktor dan matriks adjoint $(X' X)$
3. Tentukan matrik invers (kebalikan matriks) dari matrik $(X'X)$ dengan cara $(X' X)^{-1}$ = (1/D)*matriks Adjoin (X'X)
   Dimana D adalah determinan matriks (X'X)
4. Pergunakan persamaan normal ke 2 untuk mengestimasi nilai-nilai koefesien regresi b_i
5. Dalam pengujian signifikansi kofesien regresi terlebih dahulu harus dicari :
   Matriks Varians - Covarians (b_1) , yaitu :
   Var–cov (b) = $\sigma^2 (X'X)^{-1}$
   σ^2 diduga dengan cara sebagi berikut :
   σ^2 = (JK total – JK regresi)/ (n-k-1)
   dimana:
   n = banyaknya cuplikan (sample)
   k = banyaknya variabel x
   Jk total = $(Y' Y) - n \bar{y}^2$
   Jk regresi = $(B) (X' Y) - n \bar{y}^2$

(6). Koefesien determinasi ditentukan dengan

R^2 = JK Regresi/ JK Total

= $(B)(X'Y) - n\bar{Y}^2 / (Y'Y) - n\bar{Y}^2$

5.3.1 Contoh kasus analisis regresi ganda

Data hipotesis terdapat hubungan antara Y (% tenaga kerja) dengan X_1 (pendapatan – ribu \$), X_2 (tanggunan keluarga), X_3 (tingkat pengangguran) sbb :

X_0	X_1	X_2	X_3	Y
1	1.998	2.95	4.4	64.3
1	1.114	3.40	3.4	45.4
1	1.942	3.72	1.1	26.1
1	1.998	4.43	3.1	87.5
1	2.026	3.82	7.7	71.3
1	1.853	3.90	5.0	82.4
1	1.666	3.32	6.2	26.3
1	1.434	3.80	5.4	61.6
1	1.513	3.49	12.2	52.9
1	2.008	3.85	4.8	64.7
1	1.704	4.69	2.9	64.9
1	1.525	3.89	4.8	70.5
1	1.842	3.53	3.9	87.2
1	1.725	4.96	7.2	81.2
1	1.639	3.68	3.6	67.9

5.3.2 Penyelesaian dengan Metode Laplace

1. Untuk menentukan determinan dan matriks kofaktor sulit untuk matriks berdimensi 4 x 4 oleh Karen itu terlebih dahulu dibentuk dalam (X' X) – deviasi , dengan cara :

$\sum x_1^2 = \sum X^2 - (\sum X_1)^2 / n = a_{11} - (a_{o1})(a_{o1}) / a_{oo}$

$\sum x_1 x_2 = \sum x_2 x_1 = \sum X_1 X_2 - (\sum X_1)(\sum X_2) / n = a_{12} - a_{o2} a_{1o} / a_{oo}$

$\sum x_1 x_3 = \sum x_3 x_1 = \sum X_1 X_3 - (\sum X_1)(\sum X_3) / n = a_{13} - a_{10} a_{o3} / a_{oo}$

$\sum x_2^2 = \sum X_2^2 - (\sum X_3)^2 / n = a_{22} - a_{20} a_{o2} / a_{oo}$

$\sum x_2 x_3 = \sum x_3 x_2 = \sum X_2 X_3 - (\sum X_2)(\sum X_3) / n$

$= a_{23} - a_{20} a_{o3} / a_{oo}$

$\sum x_3^2 = \sum X_3^2 - (\sum X_2)^2 / n = a_{23} - a_{30} a_{o3} / a_{oo}$

$\sum x_1 y = \sum X_1 Y - (\sum X_1)(\sum Y) / n = 1Y - a_{10} 0Y / a_{oo}$

$\sum x_2 y = \sum X_2 Y - (\sum X_2)(\sum Y)/n = 2Y - a_{20}0Y / a_{oo}$
$\sum x_3 y = \sum X_3 Y - (\sum X_3)(\sum Y)/n = 3Y - a_{30}0Y / a_{oo}$
Hasoil penyelesaiaan adalah sebagai berikut :
(D) = (X'X) deviasi

d11	d12	d13		d1y
0,950282	0,244133	-1,416593		21,047807
d21	d22	d23		d2y
0,244133	3,854973	- 1,389067		63,297934
d31	d32	d33		d3y
-1,416593	-1,389067	94,939333		28,157334

2. Menentukan determinan (X'X) – deviasi
   D = d11d22d33 + d12d23d33 + d12d32d21 - d11d32d23 + d12d21d33 + d12d32d21
   D = 332,724066

3. Menentukan matrik kofaktor , yaitu dengan ketentuan , elemen – elemennya adalah :

$K_{ij} = (-1)^{i+j} \, Mij$

Lengkapnya matrik kofaktor adalah sebagai berikut

$K_{ij} =$

363,196912	-21,209593	5,109061
-21,209593	88.210506	0,974168
5,109061	0,974168	3,595158

4. Menentukan matriks adjoint
   Matriks Adjoint adalah transformasi (putaran) daripada matriks kofaktor, tetapi karena matriks kofaktor diperoleh dari matriks X'X yang simetris, maka matriks Adjoint sama dengan Matriks Kofaktor.

5. Menentukan matriks *invers* $(X'X) = (X'X)^{-1} = (C)$

$(X'X)^{-1} = (C) = (1/D)(\text{Adjoint } X'X)$

Dengan data matriks adjoint dan determinant, maka diperoleh hasil matriks kebalikan *(invers*) sebagai berikut :

$$(C) = \begin{bmatrix} C11 & C12 & C13 \\ 1{,}09586 & -0{,}063745 & 0{,}015355 \\ C21 & C22 & C23 \\ -0{,}063745 & 0{,}265116 & 0{,}002928 \\ C31 & C32 & C33 \\ 0{,}015355 & 0{,}002928 & 0{,}010805 \end{bmatrix}$$

6. Estimasi Parameter (b) = (B) , dengan menggunakan persamaan normal ke 2 :

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

$$[b] \quad [X'X]^{-1} \quad (X'Y)$$

$$\begin{bmatrix} b1 \\ b2 \\ b3 \end{bmatrix} = \begin{bmatrix} 1{,}09586 & 0{,}063745 & 0{,}015355 \\ -0{,}063745 & 0{,}265116 & 0{,}002928 \\ 0{,}015355 & 0{,}002928 & 0{,}010805 \end{bmatrix} \begin{bmatrix} 21{,}047947 \\ 63{,}297943 \\ 28{,}157344 \end{bmatrix}$$

b1= 19,373030
b2 = 15,5220041
b3 = 0,8127670

Sedangkan $b_0 = Y - b_1X_1 - b_2X_2 - b_3X_3$

$= -36.728419$

7. Pengujian signifikansi parameter b_1

$\sigma^2 = (\text{JK total} - \text{JK regresi})/(n-k-1)$

JK Total = $Y'Y - \overline{n}\,Y^2$

$= 65893.\overline{61} - 15\,Y^2$

$= 5130.137340$

JK.Regresi $= \sum (bi\,(\sum x_i y_i))$

$= 19.373010\,(21047907) +$

$15.522041(63.297934) +$

0.812767 (28.157334) = 1413.159648

σ^2 = (JK total – JK regresi)/ (n-k-1)

= (5130,137340 – 1413,159648)/(15 – 3 – 1)

= 337,907063

Matriks varian-covarian

$$\text{Var–cov (b)} = \sigma^2 (X'X)^{-1}$$

$$= 337{,}907063 \begin{bmatrix} 1{,}09586 & 0{,}063745 & 0{,}015355 \\ -0{,}063745 & 0{,}265116 & 0{,}002928 \\ 0{,}015355 & 0{,}002928 & 0{,}010805 \end{bmatrix}$$

$$= \begin{bmatrix} 368{,}854619 & -21{,}519886 & 5{,}188563 \\ -21{,}519886 & 89{,}584569 & 0{,}989392 \\ 5{,}188563 & 0{,}989392 & 3{,}651086 \end{bmatrix}$$

Varian (b1) = 368,854619 simpangan baku (b1) = 19,205588
Varian (b2) = 89,584659 simpangan baku (b2) = 9,464912
Varian (b3) = 3,651086 simpangan baku (b3) = 1,910782

Pengujian secara parsial dengan uji t sebagai berikut:
Hipotesis $H_0 : \beta i = 0$
$H_1 : \beta i \neq 0$

t hitung = b_i / s_{bi}
diperoleh hasil sebagai berikut : $t_1 = 1{,}009$; $t_2 = 1{,}640$; $t_3 = 0{,}425$
sedangkan t-tabel pada derajat bebas 13 dan taraf nyata 0,05 yaitu sebesar 2,02
Oleh karena $t_1, t_2, t_3 < t_{025}$ (db = 13) , maka diputuskan untuk menerima H_o , yang berarti parameter – parameter bi tidak dapat dipergunakan untuk peramalan.

Untuk pengujian hipotesis yang menyatakan bahwa
H_o : $B_1 = B_2 = B_3 = 0$, dapat dilakukan juga dengan menggnakan metode analisis varian (ANOVA) sebagai berikut :

Tabel 5.1. Analysis of Variance (ANOVA)

Sumber	db	Jumlah kuadrat	Kuadrat tengah	F	
				hit	.o5
Regresi	3	1413.159698	471.053216	1.394	3.59
Deviasi	11	3716.977692	337.907063		
Total	14	5130.137340			

Karena $F_{hit} <$ F.tabel dapat diputuskan untuk menerima $H_o : B_1 = B_2 = B_3 = 0$, yang memberikan kesimpulan sama dengan metode pengujian secara poarsial di atas.

(8). Koefesien determinasi dan korelasi antar variabel

koefesien determinasi dapat dicari dari table analisis varian (ANOVA) , yaitu :

R^2 = JK Regresi/ JK Total
= 1413,159689/5130,137340
= 0,275
Selanjutnya nilai koefisien korelasi (R) merupakan akar dari koefisien determinasi (R2)
Yaitu seberar 0,524

5.4 Metode DOOLITTLE

Langkah-langkah penyelesaian dengan menggunakan Metode Doolittle, adalah sebagai berikut :

5.4.1 Prosedur atau solusi Analisis regresi dengan metode Doolittle dimulai setelah diketauhi (X'X), (X 'Y) dan (Y'Y)

Tabel berikut adalah bagan untuk mengubah (X'X) menjadi matrik segitiga bawah :

Baris	X'X					X'Y	j
0	a_{oo}	a_{o1}	a_{o2}	a_{o3}	a_{o4}	9_0	j_o
1		a_{11}	a_{12}	a_{13}	a_{14}	9_1	j_1
2			a_{22}	a_{23}	a_{24}	9_2	j_2
3				a_{33}	a_{34}	9_3	j_3
4					a_{44}	9_4	j_4
5	A_{oo}	A_{o1}	A_{o2}	A_{o3}	A_{o4}	A_{oy}	j_5
6	1	B_{oo}	B_{o2}	B_{o3}	B_{o4}	B_{oy}	j_6
7		A_{11}	A_{12}	A_{13}	A_{14}	A_{1y}	j_7
8		1	B_{12}	B_{13}	B_{14}	B_{1y}	j_8
9			A_{22}	A_{23}	A_{24}	A_{2y}	j_9
10			1	B_{23}	B_{24}	B_{2y}	j_{10}
11				A_{23}	A_{24}	A_{3y}	j_{11}
12				1	B_{24}	B_{3y}	j_{12}
13					A_{44}	A_{4y}	j_{13}
14					1	B_{4y}	j_{14}
15	b_o	b_1	b_2	b_3	b_4		

1. Baris (o) sampai dengan (4) penyajian matrik (X ' X) = (a_{ij}) dengan vector (X ' Y) dan vector (J)
2. Baris (5) s/d (14) bagan operasi untk mendapatkan nilai estimasi parameter (b)
3. $j_o = a_{00} + a_{01} + a_{02} + a_{03} + a_{04} + g_o$

   $j_1 = a_{01} + a_{11} + a_{12} + a_{13} + a_{14} + g_1$

   $j_2 = a_{02} + a_{12} + a_{22} + a_{23} + a_{24} + g_2$

   $j_3 = a_{03} + a_{13} + a_{23} + a_{33} + a_{34} + g_3$

   $j_4 = a_{04} + a_{14} + a_{24} + a_{33} + a_{34} + g_4$

Langkah-langkah operasi sebagai berikut:

4) Baris (5) : pindahan dari baris (0)

5) Baris (6) : adalah hasil bagi elemen – elemen di baris (5) dengan A_{00}

Contoh: $B_{01} = A_{01}/A_{00}$; $B_{02} = A_{02}/A_{00}$, dst.

6) Baris (7) : $A_{11} = a_{11} - B_{01}A_{01}$ $A_{14} = a_{14} - B_{01}A_{04}$

$A_{12} = a_{13} - B_{01}A_{02}$ $A_{1y} = g_1 - B_{01}A_{0y}$

$A_{13} = a_{13} - B_{01}A_{01}$ $J7 = J_1 - B_{01}J_5$

7) Baris (8) : adalah hasil bagi elemen-elemen Baris (7) dengan A_{11} Contoh : $B_{12} : \frac{A_{12}}{A_{11}}$; $B_{13} = \frac{A_{13}}{A_{11}}$; $B_{14} = \frac{A_{14}}{A_{11}}$;

$B_{1y} = A_{1y}/A_{11}$

8) Baris (9) : $A_{22} = a_{22} + B_{02}A_{o2} + B_{12}A_{12}$

$A_{23} = a_{23} + B_{02}A_{o3} + B_{12}A_{13}$

$A_{24} = a_{22} + B_{02}A_{o2} + B_{12}A_{12}$

$A_{2Y} = g_2 + B_{02}A_{oy} + B_{12}A_{1y}$

$J_9 = J_2 + B_{02}J_5 + B_{12}J_7$

9) Baris (10) : adalah hasil pembagian elemen-elemen baris (9) dengan A_{22}

Contoh: $B_{23} = A_{23}/A_{22}$ dan $B_{24} = A_{24}/A_{22}$

10) Baris (11) : $A_{33} = a_{33} - B_{03}A_{o3} - B_{13}A_{13} - B_{23}A_{23}$

$A_{34} = a_{34} - B_{03}A_{o4} - B_{13}A_{14} - B_{23}A_{24}$

$A_{3y} = g_3 - B_{03}A_{oy} - B_{13}A_{1y} - B_{23}A_{2y}$

$J_{11} = J_3 - B_{01}J_5 - B_{12}J_7 - B_{23}J_9$

11) Baris (12) : Hasil bagi elemen – elemen baris (11) dengan A_{33}

12) Baris (13) :

$A_{44} = a_{44} - B_{04}A_{o4} - B_{14}A_{14} - B_{24}A_{24} - B_{34}A_{34}$

$A_{4y} = g_4 - B_{04}A_{oy} - B_{14}A_{1y} - B_{24}A_{2y} - B_{34}A_{3y}$

$J_{13} = J_4 - B_{04}J_5 - B_{14}J_7 - B_{24}J_9 - B_{34}J_{11}$

13). Baris (14) : adalah hasil bagi elemen – elemen baris (13) dengan A_{44}

$B_{4y} = \frac{A_{4y}}{A_{44}}$; $J_{14} = \frac{J_{13}}{A_{44}}$

Untuk mendeteksi ketelitian perhitungan atau memeriksa kesalahan :

$J_6 = 1 + B_{01} + B_{02} + B_{03} + B_{04} + B_{0y}$

$J_7 = A_{11} + A_{12} + A_{13} + A_{14} + A_{1y}$

$J_8 = 1 + B_{12} + B_{13} + B_{14} + B_{1y}$

$J_9 = A_{22} + A_{23} + A_{24} + A_{2y}$

$J_{10} = 1 + B_{23} + B_{24} + B_{2y}$

J_{11} $= A_{33} + A_{34} + A_{3y}$
J_{12} $= 1 + B_{34} + B_{3y}$
J_{13} $= A_{44} + A_{4y}$
J_{14} $1 + B_{4y}$

5.4.2 Pengujian dengan Metode Analisis Varian

Anova dapat digunakan untuk menguji baik secara parsial dan simultan, dapat di lihat pada tabel berikut:

Tabel 5.2. Analisis Varian (Anova)

Sumber	Db	Jumlah kuadrat	Kuadrat tengah	F-hit
X_1 X_2 /X_1 X_3 / X_2 , X_1 X_4 / X_1 X_2 , X_3	1 1 1 1	A_{iy} x B_{1y} A_{2y} x B_{2y} A_{3y} x B_{2y} A_{4y} x B_{4y}	KT – x_1 KT – x_2 KT – x_3 KT – x_4	F_1 F_2 F_3 F_4
Regresi Deviasi	4 n-4-1	$\sum A_{iy}$ x B_{1y} JK tot - JKreg	KT –regresi KT - Deviasi	F-reg
Total	n-1	Y'Y-nY2	-	

Keterangan : KT = JK / db

$$F_{Hit} = \frac{KT.....}{KT\ deviasi};$$

Jika $F_{hit} > F_{table}$ ⟶ bi signifikan

Tabel Anova diatas masih dipergunakan untuk menghitung koefesien determinasi R^2 dengan cara seperti yang telah dibahas diatas, yaitu :

R^2 = JK Regresi/ JK Total

5.4.3 Metode Doolittle pada Contoh Soal

Hasil pengolahan diperoleh sebagai berikut:

Solusi analisis regresi dengan Metode.Doolittle ini , kita pergunakan data hubungan antara Y (% tenaga kerja) dengan X_1 (pendapatan –

ribu $), X_2 (tanggunan keluarga), X_3 (tingkat pengangguran) seperti pada contoh penggunaan metode laplace, hasil sebagai berikut :

Baris	X ′ X				X ′ Y	j
(0)	15	25.997	57.43	75.70	954.70	1128.3270
(1)	49.0065		99.7768	129.9816	1675.6703	1977.2318
(2)	223.7263		288.441		3718.526	4387.9012
(3)				476.97	4346.21	5817.1026
(4)	A_{01}	A_{01}	A_{02}	A_{03}	$A_{0ý}$	J4
	15	25.997	57.43	75.70	954.70	1128.8372
(5)	B_{oo}	B_{o1}	B_{o2}	B_{o3}	B_{oy}	J5
	1	1.7331	3.8286	5.0466	63.6466	75.2551
(6)		A_{11}	A_{12}	A_{13}	A_{1y}	J6
		0.9510	0.2459	- 1.4140	21.0797	20.8677
(7)		B_{11}	B_{12}	B_{13}	B_{oy}	J7
		1	0.2585	-1.4268	22.1658	21.9366
(8)			A_{22}	A_{23}	A_{2y}	J8
			3.7862	- 1.0185	07.324	60.6813
(9)			B_{22}	B_{23}	B_{3y}	J9
			1	- 0.2690	15.2956	16.0269
(10)				A_{33}	A_{34}	J_{10}
				92.5660	75.1407	167.7047
(11)				B_{33}	B_{3y}	J_{11}
				1	0.8117	1.8117
	b_o	b_1	b_2	b_3		
(12)	-30.4028	19.3622	15.5139	0.8117		

Proses Analisis

1. Baris (4) A_{00} s/d J_4 pindahan dari baris (o)
2. Baris (5) $B_{oo} = 15/15 = 1$; $B_{o1} = 25.997 / 15 = 1.7331$;
   $B_{o2} = 57.43 / 15 = 3.8286$; $B_{o3} = 7570 / 15 = 5.0446$
   $B_{oy} = 954.70 / 15 = 65.6446$; $J_5 = 1128.8270 / 15 = 75.2351$
3. Baris (6) $A_{11} = 46.0065 – 1.7831 (25.9970) = 0.9510$
   $A_{12} = 99.7779 – 1.7331 (57.43) = 0.2459$
   $A_{13} = 129.7816 – 1.7331 (75.70) = - 1.4140$
   $A_{iy} = 1675.6703 – 17331 (9506.70) = 21.0797$
   $J_6 = 1999.2318 - 1.7331 (1128.8270) = 20.8677$
4. Baris (7) $B_{11} = 0.9510 / 0.9510 = 1$
   $B_{12} = 0.2459 / 0.9510 = 0.2585$
   $B_{13} = 1.4140 / 0.9510 = - 1.4268$
   $B_{1y} = 21.0797 / 09510 = 22.1658$
   $J_7 = 20.8677 / 0.9510 = 21.9366$
5. Baris (8)
   $A_{22} = 223.7263 – 3.8286 (57.43) – 0.2585 (0.2459)$

= 3.7862

A_{23} = 288.441 - 3.8286 (75.70) – 0.2585 (- 14140)
= - 1.0185

A_{2y} = 3918.526 - 3.8286 (954.70) – 0.2585 (21.0797)
= 57.9125

J_8 = 4389.9012 - 3.8286 (2528.8570) – 0.2585 (20.8677)
= 60.6813

6. Baris (9) B_{22} = 3.7862 / 3.7862 = 1
   B_{23} = -1.0185 / 3.7862 = - 0.2690
   B_{2y} = 57.9124 / 3.7862 = 15.2956
   J_9 = 60.6813 / 3.9862 = 16.0369
7. Baris (10)
   A_{33} = 476.97 – 5.0466(75.70) + 1.4868 (-1.4140)
   + 0.2690 (-1.0185) = 92.5660
   A_{3y} = 4816.21 – 5.0466 (954.70) + 14388 (21.0797)
   + 0.2690 (57.9124)
   = 75.1407
   J_{10} =5817.1026 –5.0466(1128.8270)+1.4368(20.8677)
   +0.269 (60.6813) = 167.7047
8. Baris (11) B_{33} = 92.5660 / 92.5660 = 1
   B_{3y} = 75.1407 / 92.3660 = 7117
   J_{11} = 167.7047 / 92.5660 = 1.8117

9. Penentuan parameter (b_i)
   b_3 = 0.8117
   b_2 = 15.2956 - (- 2690) (0.8117) = 15.5139
   b_1 = 22.1658 – (- 1.4868) (0.8117) – 0.2585 (15.5139)
   = 19.3622
   b_o = 63.6466 – 5.0466 (0.8117) – 3.8286 (15.5139)
   – 1.7331 (19.3622) = - 35.4028

Pengujian signifikansi parameter dengan metode analisis varian

Sumber	db	Jumlah kuadrat	Kuadrat tengah	F_{hit}	$F_{0,05}$
X_1	1	467.2484	467.2414	1.38	4,84
X_2 / X_1	1	885.8049	885.8049	3.62	4,84
$X_3 / X_2, X_1$	1	60.9917	60.9917	0.18	4,84
Regresi	3	1414.0450	471.3483	1.39	3,59
Deviasi	11	3716.0923	337.8265		
Total	14	5130.1373			

Semua F_{hit} ternyata < $F_{0.05}$ diputuskan untuk menerima
$H_o : B_1 = B_2 = B_3 = 0$

Beberapa hal yang menguntungkan dari Metode Doolittle singkat dibandingkan Metode Laplace :

1. Tidak perlu mencari nilai determinan ($X'X$) yang semakin sulit dengan bertambahnya variable X
2. Dalam mencari jumlah kuadrat X_1 , tidak perlu dideviasikan terlebih dahulu
3. Tabel operasi Doolittle mempunyai fungsi ganda , yaitu penentuan JK – X_1 , parameter b_1 penentuan $(X'X)^{-1}$
4. Dilengkapi kolom control untuk memeriksa ketelitian dalam proses perhitungan

Beberapa yang perlu diperhatikan dari Metode Doolittle ini :

1. Memerlukan ketelitian dalam proses perhitungan
2. Hasil signifikasnsi lewat analisis varians untukk setiap parameter, tidak bebas karena dipengaruhi oleh parameter variable sebelumnya.

5.5 Penutup

Analisis regresi linear merupakan metode statistik yang sangat berguna baik dalam bidang pertanian, kesehatan, ekonomi dan bisnis untuk memahami dan memprediksi hubungan antar variabel. Bab ini telah membahas tenteng analisis regresi linear berganda, dan juga aplikasinya.

Beberapa poin penting yang perlu diingat dari pembahasan ini antara lain:

1. Fungsi Korelasi - Koefisien korelasi (R) mengukur keeratan hubungan antara dua variabel, dengan nilai berkisar antara -1 hingga +1. Nilai mendekati +1 atau -1 menunjukkan hubungan yang kuat, sementara nilai mendekati 0 menunjukkan hubungan yang lemah.
2. Interpretasi Koefisien Determinasi - Koefisien determinasi (R^2) menunjukkan persentase variasi dalam variabel tidak bebas yang dapat dijelaskan oleh variabel bebas.
3. Pengujian Signifikansi - Uji t dan uji F digunakan untuk menentukan apakah hubungan yang diamati antara variabel adalah signifikan secara statistik atau hanya terjadi karena kebetulan.
4. Asumsi Model Regresi - Validitas model regresi linear bergantung pada beberapa asumsi seperti linearitas, normalitas residual, dan homoskedastisitas (kesamaan varians), multikolinearitas.
5. Aplikasi Praktis mencari dan menguji koefisien regresi dalam Regresi Linear Berganda dengan mnggunakan Metode Laplace dan Metode Doolittle. Kelebihan dan kekurangan dari kedua metode tersebut, yaitu dalam Metode Doolittle tidak perlu mencari determinan mkatriks dan tidak perlu mencari matriks deviasi.

Kemampuan untuk menganalisis hubungan antara variabel-variabel dalam bidang pertanian, kesehatan, ekonomi dan bisnis melalui korelasi, determinasi dan regresi linear berganda merupakan keterampilan yang sangat berharga dalam pengambilan keputusan berdasarkan data. Namun, perlu diingat bahwa hubungan korelasional tidak selalu berarti hubungan kausal. Faktor-faktor lain yang tidak dimasukkan dalam model mungkin juga memengaruhi variabel tidak bebas.

Ada beberapa saran yang dapat di ungkapkan dalam melakukan pengujian, bahwa jika pengujian hubungan (uji F secara simultan) sudah dilakukan dan hasil nya signifikan, maka proses pencarian nilai koefisien korelasi atau koefisien determinasi dapat

dilakukan, tetapi jika sebaliknya, bahwa hasil pengujian hubungan tidak signifikan, maka tidak perlu dilakukan pencarian nilai koefisien korelasi dan determinasi.

DAFTAR PUSTAKA

Approach. Mc.Graw-Hill International Book Company. International Student Edition. Second Edition. Macmillan.

Press, Ames, Iowa, USA. Sixth Edition, Fourth Printing.

Snedecor G.W and Cochran W,G. 1971. Statistical Method. The Iowa State University

Steel R.G.D and Torrie J.H. 1981. Principles and Procedures of Statistics. A Biometrical

Sudradjat S.W. M. 1984. Mengenal Ekonometrika Pemula. Penerbit Armico, Bandung.

Walpole R.E. 1975. Introduction to Statistics. New York Macmillan London Collier

Walpole R.E. 1993. Pengantar Statistika. Gramedia Pustaka Utama. Indonesia.

BAB 6
ANALISIS VARIANSI MODEL REGRESI GANDA

Oleh Ade Astuti Widi Rahayu

6.1 Pendahuluan

Dalam dunia statistik terapan, model regresi linear ganda menjadi salah satu metode paling penting untuk menjelaskan hubungan antara satu variabel terikat (dependen) dan dua atau lebih variabel bebas (independen). Namun, penting untuk menilai apakah model tersebut signifikan secara keseluruhan, bukan hanya sebagian variabel bebasnya. Di sinilah peran Analisis Variansi (ANOVA) menjadi sentral, karena mampu menguji apakah variasi dalam data dapat dijelaskan secara signifikan oleh variabel variabel bebas (Trunfio et al. 2022). Analisis variansi (ANOVA) dalam model regresi ganda adalah teknik statistik yang digunakan untuk menguji apakah variabel-variabel independen secara bersama-sama berpengaruh signifikan terhadap variabel dependen. Regresi linier ganda sendiri merupakan pengembangan dari regresi linier sederhana, di mana terdapat lebih dari satu variabel independen yang mempengaruhi satu variabel dependen (Ghozali 2018). Regresi linear ganda memodelkan hubungan antar banyak variabel independen dan satu variabel dependen. ANOVA dalam regresi digunakan untuk menguji signifikansi keseluruhan model, apakah variabel variabel bebas secara kolektif mampu menjelaskan variasi variabel dependen (F-test) (Smith, 2020). Dengan membandingkan varians yang dijelaskan (SSR) dan yang tidak (SSE), F-statistik diturunkan untuk uji hipotesis global.

6.2 Konsep Umum Analisis Variansi (ANOVA)

Analisis Variansi atau ANOVA (*Analysis of Variance*) adalah metode statistik yang digunakan untuk membandingkan rata-rata dari dua atau lebih kelompok dan menentukan apakah terdapat

perbedaan yang signifikan secara statistik di antara kelompok tersebut. Teknik ini dikembangkan oleh Ronald A. Fisher pada awal abad ke-20 dan telah menjadi salah satu pilar utama dalam inferensi statistik modern, terutama dalam eksperimen dan analisis model regresi (Montgomery 2020; Gelman 2020).

6.2.1 Definisi

Menurut (Montgomery, 2020) ANOVA adalah prosedur yang memisahkan total variabilitas dalam data menjadi komponen-komponen yang dapat dikaitkan dengan berbagai sumber penyebab. Konsep ini mendasari pemahaman bahwa variabilitas dalam data dapat dikaitkan dengan perlakuan atau faktor-faktor yang sedang diuji dan dengan kesalahan acak (*error*).

Total Variasi = Variasi antar kelompok (antara perlakuan) + Variasi dalam kelompok (galat)

Hipotesis yang diuji dalam ANOVA:

H_0: $\mu_1 = \mu_2 = ... = \mu_k$

H_1: Paling tidak ada satu μ berbeda

6.2.2 Jenis-jenis ANOVA

Analisis Variansi (ANOVA) memiliki berbagai bentuk yang dikembangkan sesuai dengan kompleksitas desain penelitian dan jumlah faktor (variabel independen) yang terlibat. Pemilihan jenis ANOVA bergantung pada jumlah faktor, jumlah kelompok, dan apakah terdapat pengukuran berulang atau tidak (Field 2018; Heumann et al. 2016)

1. *One-Way* ANOVA (ANOVA Satu Arah)
   One-Way ANOVA digunakan ketika hanya ada satu faktor (variabel independen) dengan dua atau lebih kelompok. Tujuannya adalah untuk mengetahui apakah terdapat perbedaan rata-rata yang signifikan antar kelompok (Field 2018).
   Contoh: Menguji perbedaan nilai matematika antara tiga kelas berbeda.
2. *Two-Way* ANOVA (ANOVA Dua Arah)
   Two-Way ANOVA digunakan ketika terdapat dua faktor yang diuji secara simultan. Selain menguji efek masing-masing faktor secara

individual (*main effect*), analisis ini juga menguji apakah terdapat interaksi antara kedua faktor (Heumann et al., 2016).
Contoh: Menguji efek jenis kelamin dan metode belajar terhadap hasil ujian.

3. *Repeated Measures* ANOVA
   Digunakan ketika data dikumpulkan dari pengukuran berulang pada subjek yang sama. ANOVA ini mengontrol variabilitas dalam individu sehingga lebih sensitif terhadap perbedaan antar kondisi (Field, 2018).
   Contoh: Mengukur tekanan darah pasien yang sama sebelum, saat, dan setelah pengobatan.
4. *Mixed-Design* ANOVA
   Kombinasi antara *between-subjects factor* dan *within-subjects factor*. Digunakan dalam eksperimen kompleks yang melibatkan kedua jenis pengukuran (Field, 2018).
   Contoh: Efek jenis terapi (antara-subjek) dan waktu (dalam-subjek) terhadap tingkat stres.
5. MANOVA (*Multivariate* ANOVA)
   Perluasan dari ANOVA yang digunakan saat terdapat lebih dari satu variabel dependen. MANOVA mempertimbangkan korelasi antar variabel dependen, yang tidak diperhitungkan dalam ANOVA biasa (Tabachnick, 2019).
   Contoh: Menguji efek latihan fisik terhadap tekanan darah dan detak jantung secara simultan.
6. ANOVA dalam Regresi
   Dalam konteks regresi linear, ANOVA digunakan untuk menguraikan total variasi dalam variabel dependen menjadi bagian yang dijelaskan oleh model (SSR) dan bagian yang tidak dijelaskan (SSE). Uji F dalam regresi berperan sebagai ANOVA regresi untuk menguji signifikansi model secara keseluruhan (Kutner, 2016).

6.2.3 Kegunaan ANOVA

Analisis Variansi (ANOVA) adalah metode statistik yang sangat berguna untuk **mengidentifikasi perbedaan rata-rata antar kelompok**, serta **mengevaluasi pengaruh satu atau lebih variabel independen terhadap variabel dependen**. ANOVA menjadi alat

penting dalam penelitian eksperimental dan observasional karena kemampuannya dalam mengontrol dan membandingkan banyak kelompok secara simultan tanpa meningkatkan risiko kesalahan tipe I (Field 2018)

Berikut adalah beberapa kegunaan utama ANOVA di berbagai bidang:

1. Membandingkan rata-rata antar kelompok:
   Digunakan untuk menguji apakah rata-rata dua atau lebih kelompok berbeda secara signifikan (Field, 2018).
2. Mengukur pengaruh perlakuan:
   Bermanfaat dalam eksperimen untuk mengetahui efek suatu intervensi atau treatment (Montgomery 2020)).
3. Menganalisis interaksi antar faktor:
   Two-Way ANOVA dan MANOVA mengevaluasi pengaruh dan interaksi lebih dari satu faktor (Tabachnick, 2019).
4. Menilai efektivitas strategi atau produk:
   Dalam bisnis dan pemasaran, ANOVA membandingkan efektivitas metode promosi atau preferensi produk (C. , S. M. and S. Heumann, 2017).
5. Mengukur signifikansi model regresi:
   Digunakan dalam regresi ganda untuk menilai signifikansi keseluruhan model (Kutner, 2016).
6. Pengujian ilmiah di berbagai bidang:
   Digunakan luas dalam psikologi, pendidikan, kesehatan, ekonomi, dan ilmu sosial.

6.3 Statistik F dan Uji Signifikansi

Setelah model regresi ganda dibangun, langkah krusial berikutnya adalah mengevaluasi signifikansi model secara keseluruhan. Hal ini dilakukan dengan menggunakan statistik F, yang merupakan bagian dari kerangka Analisis Variansi (ANOVA). Uji ini digunakan untuk mengetahui apakah variabel-variabel independen yang dimasukkan ke dalam model, secara bersama-sama, memiliki pengaruh signifikan terhadap variabel dependen.

6.3.1 Konsep Dasar Statistik F

Statistik F dalam regresi linear ganda digunakan untuk menguji signifikansi model secara keseluruhan, yakni apakah semua variabel independen secara kolektif berpengaruh terhadap variabel dependen. Uji F membandingkan variabilitas yang dijelaskan oleh model (SSR) terhadap variabilitas yang tidak dijelaskan (SSE). (Kutner 2016; Field 2018).

6.3.2 Rumus Statistik F

F = (MSR / MSE) = (SSR / p) / (SSE / (n - p - 1)) (Heumann 2017)

Keterangan:

1. SSR: Regression Sum of Squares
2. SSE: Error Sum of Squares
3. MSR: Mean Square Regression = SSR / p
4. MSE: Mean Square Error = SSE / (n - p - 1)
5. p: jumlah variabel independen
6. n: jumlah observasi

6.3.3 Hipotesis dalam Uji F

H_0 (Hipotesis nol): Semua koefisien regresi sama dengan nol ($\beta_1 = \beta_2 = ... = \beta_p = 0$)

H_1 (Hipotesis alternatif): Setidaknya satu koefisien tidak nol

Jika nilai F hitung > F tabel (atau p-value < 0.05), maka H_0 ditolak: artinya model regresi signifikan secara statistik (Montgomery, 2020).

6.3.4 Contoh Interpretasi

Misal: SSR = 120, SSE = 30, p = 2, n = 10

MSR = 120 / 2 = 60

MSE = 30 / (10 - 2 - 1) = 30 / 7 ≈ 4.29

F = 60 / 4.29 ≈ 13.99

Jika F-tabel ($df_1 = 2, df_2 = 7$) = 4.74, maka F > F-tabel, model signifikan.

6.3.5 Implikasi

1. Statistik F menilai kecocokan model secara keseluruhan.
2. Tidak menunjukkan variabel mana yang signifikan — perlu uji t untuk masing-masing koefisien. (Field, 2018).

6.4 Koefisien Determinasi

Koefisien determinasi, atau R^2, dalam konteks regresi ganda dan analisis variansi (ANOVA), digunakan untuk mengukur proporsi total variasi dalam variabel dependen yang dapat dijelaskan oleh variabel-variabel independen dalam model regresi. Pendekatan ini tidak hanya menilai hubungan satu lawan satu antara variabel, tetapi juga menilai kontribusi kolektif dari semua variabel bebas terhadap model (Kutner, 2016). Secara prinsip, ANOVA dalam regresi membagi total variasi dari data menjadi dua komponen: variasi yang dijelaskan oleh model (regression sum of squares/SSR) dan variasi yang tidak dijelaskan (*error sum of squares*/SSE). Hubungan ini secara matematis dirumuskan sebagai:

$$R^2 = SSR / SST \text{ atau } R^2 = 1 - SSE / SST$$

Interpretasi dari R^2 adalah sebagai berikut: jika $R^2 = 0.80$, maka 80% dari variasi dalam variabel Y dapat dijelaskan oleh variabel-variabel X melalui model regresi tersebut. Semakin tinggi nilai R^2, semakin baik kemampuan model menjelaskan data (C. , S. M. and S. Heumann, 2017). Namun, salah satu kelemahan dari R^2 adalah bahwa nilainya selalu meningkat atau tetap ketika variabel baru ditambahkan ke dalam model, meskipun variabel tersebut tidak signifikan secara statistik. Oleh karena itu, digunakanlah Adjusted R^2, yaitu koefisien determinasi yang telah disesuaikan dengan jumlah variabel independen dan ukuran sampel, sehingga lebih akurat untuk mengevaluasi model regresi yang kompleks (Ozili, 2023). Adjusted R^2 dihitung dengan mengoreksi R^2 berdasarkan derajat kebebasan, sehingga mampu memberikan gambaran sejauh mana model menjelaskan variasi Y tanpa terjebak dalam overfitting.

Dalam praktiknya, R^2 dan Adjusted R^2 digunakan berdampingan dengan uji F untuk mengevaluasi signifikansi global model. Jika nilai R^2 tinggi namun uji F tidak signifikan, maka kemungkinan besar model mengalami overfitting atau variabel bebas tidak berkontribusi nyata. Koefisien determinasi yang diperoleh dari pendekatan ANOVA menjadikan regresi ganda tidak hanya sebagai alat prediksi, tetapi juga alat inferensial yang kuat

untuk menguji hipotesis dan memahami struktur hubungan antar variabel.

6.5 Langkah-langkah Analisis

Analisis Variansi (ANOVA) merupakan metode statistik yang digunakan untuk menguji perbedaan rata-rata antara dua kelompok atau lebih. Dalam konteks regresi maupun eksperimen, ANOVA digunakan untuk mengidentifikasi apakah variabel independen secara signifikan memengaruhi variabel dependen. Berikut adalah tahapan sistematis dalam melakukan analisis ANOVA:

1. **Menyusun Hipotesis Statistik**
   Langkah awal adalah menyatakan hipotesis nol (H_0) dan hipotesis alternatif (H_1):
   a. H_0: Tidak ada perbedaan rata-rata antar kelompok (misalnya, semua koefisien regresi = 0).
   b. H_1: Setidaknya terdapat satu perbedaan rata-rata yang signifikan. (Huitema, 2011)
2. **Menentukan Jenis ANOVA**
   Pilih jenis ANOVA yang sesuai berdasarkan desain penelitian:
   a. *One-Way* ANOVA: untuk satu faktor independen dengan ≥ 2 kategori.
   b. *Two-Way* ANOVA: untuk dua faktor independen dan interaksi antar faktor.
   c. ANOVA dalam regresi: digunakan untuk mengevaluasi signifikansi model secara keseluruhan. (Glen, 2020)
3. **Menghitung Komponen Variansi**
   Hitung:
   a. SST (*Total Sum of Squares*): total variasi dalam data.
   b. SSR (*Regression Sum of Squares*): variasi yang dijelaskan oleh model.
   c. SSE (*Error Sum of Squares*): variasi yang tidak dijelaskan oleh model.

   SST = SSR + SSE
   Menghitung Rataan Kuadrat (*Mean Squares*)
   MSR = SSR / p dan MSE = SSE / (n - p - 1)
4. **Menghitung Statistik F**
   F = MSR / MSE

Bandingkan nilai F hitung dengan nilai kritis dari distribusi F pada tingkat signifikansi tertentu (misalnya $\alpha = 0.05$) atau lihat nilai p (Ozili, 2023).

5. **Menarik Kesimpulan Statistik**
   a. Jika F > F-tabel atau p-value < α, maka H_0 ditolak.
   b. Ini berarti ada perbedaan signifikan antar kelompok atau model regresi signifikan.
6. **Melakukan Uji Lanjutan (jika perlu)**
   Jika menggunakan One-Way ANOVA dan H_0 ditolak, maka uji post-hoc seperti Tukey HSD atau Bonferroni diperlukan untuk mengetahui kelompok mana yang berbeda.
7. **Memvalidasi Asumsi**
   Pastikan asumsi ANOVA terpenuhi:
   a. Data berdistribusi normal
   b. Varians antar kelompok homogen
   c. Observasi independen (Bhattacharyya 2019)

6.6 Contoh Perhitungan

1. **ANOVA *One-Way***
   Contoh berikut menunjukkan bagaimana cara melakukan perhitungan ANOVA *One-Way* untuk menguji perbedaan rata-rata antara tiga kelompok.Peneliti ingin mengetahui apakah terdapat perbedaan rata-rata skor ujian antara tiga kelompok sekolah yang berbeda, yaitu Sekolah A, Sekolah B, dan Sekolah C. Data yang diperoleh adalah sebagai berikut:
   a. Sekolah A: 65, 68, 70, 72
   b. Sekolah B: 75, 78, 76, 77
   c. Sekolah C: 85, 83, 88, 90

   Langkah-langkah perhitungan ANOVA *One-Way* adalah sebagai berikut:

   Langkah 1: Hitung Rata-rata Tiap Kelompok dan *Grand Mean*
   a. Rata-rata Sekolah A = (65 + 68 + 70 + 72) / 4 = 68.75
   b. Rata-rata Sekolah B = (75 + 78 + 76 + 77) / 4 = 76.5
   c. Rata-rata Sekolah C = (85 + 83 + 88 + 90) / 4 = 86.5
   d. Grand Mean (GM) = (68.75 + 76.5 + 86.5) / 3 = 77.25

Langkah 2: Hitung SSB (*Sum of Squares Between Groups*)
SSB = $4 \times [(68.75 - 77.25)^2 + (76.5 - 77.25)^2 + (86.5 - 77.25)^2] = 4 \times 158.6875 = 634.75$

Langkah 3: Hitung SSW (*Sum of Squares Within Groups*)

a. Sekolah A: $(65-68.75)^2 + (68-68.75)^2 + (70-68.75)^2 + (72-68.75)^2 = 14.06$
b. Sekolah B: $(75-76.5)^2 + (78-76.5)^2 + (76-76.5)^2 + (77-76.5)^2 = 5.0$
c. Sekolah C: $(85-86.5)^2 + (83-86.5)^2 + (88-86.5)^2 + (90-86.5)^2 = 29.0$

SSW = 14.06 + 5.0 + 29.0 = 48.06

Langkah 4: Hitung Derajat Bebas dan *Mean Squares*

a. df_between = 3 - 1 = 2
b. df_within = 12 - 3 = 9
c. MSB = 634.75 / 2 = 317.38
d. MSW = 48.06 / 9 ≈ 5.34

Langkah 5: Hitung Nilai F
F = MSB / MSW = 317.38 / 5.34 ≈ 59.4
Langkah 6: Interpretasi
Dengan F hitung = 59.4 dan df1 = 2, df2 = 9, kita bandingkan dengan F_tabel pada $\alpha = 0.05$, yaitu sekitar 4.26. Karena F hitung > F tabel, maka tolak H_0. Artinya, terdapat perbedaan yang signifikan dalam nilai rata-rata antara ketiga jenis sekolah.

2. *Two-Way* ANOVA

Contoh berikut menunjukkan bagaimana cara melakukan perhitungan *Two-Way* ANOVA untuk menguji perbedaan rata-rata antara dua faktor: Metode Pengajaran (Daring dan Tatap Muka) dan Jenis Kelamin (Laki-laki dan Perempuan) pada hasil ujian matematika. Peneliti mengumpulkan data dari 6 siswa pada masing-masing kelompok, sehingga ada total 12 observasi.
Data:

a. Metode A (Daring) Laki-laki: 75, 78, 80
b. Metode A (Daring) Perempuan: 80, 85, 84

c. Metode B (Tatap Muka) Laki-laki: 78, 82, 85
d. Metode B (Tatap Muka) Perempuan: 83, 88, 87

Langkah-langkah perhitungan ANOVA *Two-Way* adalah sebagai berikut:

Langkah 1: Menyusun Hipotesis

a. Hipotesis Nol (H_0): Tidak ada perbedaan rata-rata antara metode pengajaran, jenis kelamin, dan tidak ada interaksi antara keduanya.
b. Hipotesis Alternatif (H_1): Ada perbedaan rata-rata antara kelompok atau ada interaksi antara faktor-faktor tersebut.

Langkah 2: Menghitung Rata-rata Tiap Kelompok

a. Rata-rata Metode A (Laki-laki) = (75 + 78 + 80) / 3 = 77.67
b. Rata-rata Metode A (Perempuan) = (80 + 85 + 84) / 3 = 83.00
c. Rata-rata Metode B (Laki-laki) = (78 + 82 + 85) / 3 = 81.67
d. Rata-rata Metode B (Perempuan) = (83 + 88 + 87) / 3 = 86.00
e. Grand Mean (GM) = (75 + 78 + 80 + 80 + 85 + 84 + 78 + 82 + 85 + 83 + 88 + 87) / 12 = **81.67**

Langkah 3: Menghitung *Sum of Squares* (SS)

Menghitung variasi dalam data, antara kelompok dan dalam kelompok:

SS Total (SST) = Variasi total dalam data
SS Between (SSB) = Variasi yang dijelaskan oleh faktor-faktor yang diuji
SS Within (SSW) = Variasi yang tidak dapat dijelaskan oleh model
SS Interaction (SS_AB) = Variasi yang disebabkan oleh interaksi antara faktor-faktor tersebut.

Langkah 4: Menghitung Derajat Bebas (df)

a. df_A (Metode Pengajaran) = 2 - 1 = 1
b. df_B (Jenis Kelamin) = 2 - 1 = 1
c. df_AB (Interaksi antara Metode Pengajaran dan Jenis Kelamin) = 1

d. df_E (Error) = 12 - 4 = 8

Langkah 5: Menghitung *Mean Squares* (MS)

a. MS_A = SS_A / df_A
b. MS_B = SS_B / df_B
c. MS_AB = SS_AB / df_AB
d. MS_E = SSE / df_E

Langkah 6: Menghitung Nilai F

Nilai F dihitung untuk masing-masing faktor (A, B, dan interaksi AB) sebagai berikut:
F_A = MS_A / MS_E
F_B = MS_B / MS_E
F_AB = MS_AB / MS_E

Langkah 7: Uji Hipotesis

Bandingkan nilai F-hitung dengan nilai F-tabel pada derajat kebebasan yang sesuai (df_A, df_B, df_AB, dan df_E). Jika F-hitung > F-tabel, maka tolak H_0 dan terima H_1.

Langkah 8: Kesimpulan

Jika salah satu dari nilai F yang dihitung lebih besar dari nilai F kritis, maka kita dapat menyimpulkan bahwa ada perbedaan yang signifikan antara kelompok atau interaksi antara faktor tersebut.

3. Repeated Measures ANOVA

Contoh berikut menunjukkan bagaimana cara melakukan perhitungan Repeated Measures ANOVA untuk menguji perbedaan rata-rata pada tiga waktu pengukuran: sebelum diet, selama diet, dan setelah diet, pada kelompok subjek yang sama. Peneliti mengumpulkan data dari 4 subjek.
Data:

a. Waktu Sebelum Diet: 75, 80, 70, 85
b. Waktu Selama Diet: 72, 78, 68, 83
c. Waktu Setelah Diet: 70, 76, 65, 82

Langkah-langkah perhitungan Repeated Measures ANOVA adalah sebagai berikut:

Langkah 1: Menyusun Hipotesis

a. Hipotesis Nol (H_0): Tidak ada perbedaan rata-rata berat badan antara waktu sebelum, selama, dan setelah diet.
b. Hipotesis Alternatif (H_1): Ada perbedaan rata-rata berat badan antara waktu sebelum, selama, dan setelah diet.

Langkah 2: Menghitung Rata-rata Tiap Waktu

a. Rata-rata Sebelum Diet = (75 + 80 + 70 + 85) / 4 = 77.5
b. Rata-rata Selama Diet = (72 + 78 + 68 + 83) / 4 = 75.25
c. Rata-rata Setelah Diet = (70 + 76 + 65 + 82) / 4 = 72.75
d. *Grand Mean* (GM) = (75 + 80 + 70 + 85 + 72 + 78 + 68 + 83 + 70 + 76 + 65 + 82) / 12 = 74.92

Langkah 3: Menghitung Sum of Squares (SS)

Menghitung variasi dalam data, antara waktu dan dalam waktu:
SS Total (SST) = Variasi total dalam data
SS Between (SSB) = Variasi yang dijelaskan oleh faktor waktu (Before, During, After)
SS Within (SSW) = Variasi yang tidak dapat dijelaskan oleh model
SS Error (SSE) = Variasi yang disebabkan oleh error dalam pengukuran.

Langkah 4: Menghitung Derajat Bebas (df)

a. df_time = jumlah waktu - 1 = 3 - 1 = 2
b. df_error = (jumlah subjek - 1) x jumlah waktu - 1 = (4 - 1) x 3 - 1 = 6
c. df_total = jumlah observasi - 1 = 12 - 1 = 11

Langkah 5: Menghitung Mean Squares (MS)

a. MS_time = SS_time / df_time
b. MS_error = SS_error / df_error

Langkah 6: Menghitung Nilai F

F = MS_time / MS_error

Langkah 7: Uji Hipotesis
Bandingkan F-hitung dengan nilai F-tabel pada derajat kebebasan yang sesuai (df_time dan df_error). Jika F-hitung lebih besar dari F-tabel, maka tolak H_0 dan terima H_1. Ini berarti ada perbedaan signifikan antara waktu-waktu pengukuran.

Langkah 8: Kesimpulan
Jika F-hitung > F-tabel, maka kita dapat menyimpulkan bahwa ada perbedaan signifikan antara berat badan sebelum, selama, dan setelah diet.
Seorang peneliti ingin menguji pengaruh dua faktor terhadap skor tes matematika, yaitu metode pengajaran (antara subjek) dan waktu pengukuran (dalam subjek). Peneliti memiliki dua kelompok siswa: satu kelompok yang diajar dengan metode pengajaran tradisional dan satu kelompok yang diajar dengan metode pengajaran baru. Setiap siswa diuji pada dua waktu yang berbeda, yaitu sebelum dan setelah mengikuti pelajaran.

4. *Mixed-Design* ANOVA
Faktor pertama (antara subjek): Metode pengajaran (Tradisional vs Baru)
Faktor kedua (dalam subjek): Waktu pengukuran (Sebelum vs Setelah pelajaran)
Data yang dikumpulkan:
Kelompok Tradisional:
Sebelum Pelajaran: 60, 55, 58
Setelah Pelajaran: 65, 62, 63
Kelompok Baru:
Sebelum Pelajaran: 65, 70, 68
Setelah Pelajaran: 75, 80, 78
Langkah-langkah Perhitungan Mixed-Design ANOVA:
Hipotesis:
Hipotesis nol (H_0):
a. Tidak ada perbedaan skor tes berdasarkan metode pengajaran (antara subjek).
b. Tidak ada perbedaan skor tes antara waktu pengukuran (dalam subjek).

c. Tidak ada interaksi antara metode pengajaran dan waktu pengukuran.

Hipotesis alternatif (H_1):

a. Ada perbedaan skor tes berdasarkan metode pengajaran (antara subjek).
b. Ada perbedaan skor tes antara waktu pengukuran (dalam subjek).
c. Ada interaksi antara metode pengajaran dan waktu pengukuran.

Perhitungan SS (*Sum of Squares*) untuk masing-masing faktor:

a. SS antara (Metode Pengajaran): Variasi antara kelompok metode pengajaran dihitung dengan membandingkan rata-rata kelompok dengan rata-rata total.
b. SS dalam (Waktu Pengukuran): Variasi dalam setiap kelompok berdasarkan waktu pengukuran dihitung dengan membandingkan skor dalam kelompok sebelum dan setelah pengukuran.
c. SS interaksi: Variasi yang timbul akibat interaksi antara metode pengajaran dan waktu pengukuran.
d. SS error: Variasi yang tidak dapat dijelaskan oleh model (kesalahan).

Perhitungan MS (Mean Square) untuk setiap sumber variasi:
MS = SS / df, dengan df adalah derajat kebebasan.
Untuk Metode Pengajaran: MS_antara = SS_antara / df_antara
Untuk Waktu Pengukuran: MS_dalam = SS_dalam / df_dalam
Untuk Interaksi: MS_interaksi = SS_interaksi / df_interaksi
Untuk Error: MS_error = SS_error / df_error
Perhitungan F (Rasio F):
F_antara = MS_antara / MS_error
F_dalam = MS_dalam / MS_error
F_interaksi = MS_interaksi / MS_error

Uji Signifikansi:

Tentukan p-value untuk setiap F yang dihitung. Jika p-value lebih kecil dari 0.05, maka ada perbedaan yang signifikan.

Hasil Perhitungan (Misalnya):
Metode Pengajaran: F_antara = 5.12, p_antara = 0.048
Waktu Pengukuran: F_dalam = 10.31, p_dalam = 0.008
Interaksi: F_interaksi = 2.25, p_interaksi = 0.15
Kesimpulan:

a. Ada perbedaan signifikan dalam skor tes berdasarkan metode pengajaran (p = 0.048), artinya metode pengajaran mempengaruhi skor tes.
b. Ada perbedaan signifikan dalam skor tes berdasarkan waktu pengukuran (p = 0.008), artinya skor tes berbeda antara sebelum dan setelah pelajaran.
c. Tidak ada interaksi yang signifikan antara metode pengajaran dan waktu pengukuran (p = 0.15), artinya pengaruh waktu pengukuran terhadap skor tes tidak tergantung pada metode pengajaran.

5. MANOVA

Multivariate Analysis of Variance (MANOVA) adalah teknik statistik yang digunakan untuk menguji perbedaan rata-rata antara dua atau lebih kelompok pada dua atau lebih variabel dependen secara simultan. MANOVA digunakan ketika ada lebih dari satu variabel dependen yang perlu diuji pengaruhnya terhadap faktor kategori (seperti kelompok atau perlakuan).

Pada contoh ini, kita akan menguji apakah terdapat perbedaan signifikan antara tiga kelompok diet (Diet A, Diet B, dan Diet C) terkait dengan dua variabel dependen: **berat badan** dan **kadar kolesterol**.

Data

Berikut adalah data yang digunakan dalam contoh ini:

a. **Diet A**: Berat badan = 70, 68, 72 kg; Kadar kolesterol = 180, 175, 185 mg/dL
b. **Diet B**: Berat badan = 80, 82, 78 kg; Kadar kolesterol = 210, 220, 200 mg/dL
c. **Diet C**: Berat badan = 65, 66, 63 kg; Kadar kolesterol = 170, 160, 165 mg/dL

Langkah-langkah Perhitungan MANOVA

a. Menyusun Hipotesis

- **Hipotesis nol (H_0)**: Tidak ada perbedaan yang signifikan antara kelompok diet dalam hal berat badan dan kadar kolesterol.
- **Hipotesis alternatif (H_1)**: Ada perbedaan yang signifikan antara kelompok diet dalam hal berat badan dan kadar kolesterol.

b. Menghitung Statistik Deskriptif

Rata-rata Berat Badan:

1) Diet A: (70 + 68 + 72) / 3 = 70
2) Diet B: (80 + 82 + 78) / 3 = 80
3) Diet C: (65 + 66 + 63) / 3 = 64.67

Rata-rata Kadar Kolesterol:

1) Diet A: (180 + 175 + 185) / 3 = 180
2) Diet B: (210 + 220 + 200) / 3 = 210
3) Diet C: (170 + 160 + 165) / 3 = 165

c. Menghitung Matriks Kovarians

Matriks kovarians mengukur sejauh mana dua variabel dependen (berat badan dan kadar kolesterol) saling berhubungan. Untuk setiap kelompok diet, kita menghitung kovarians antar variabel dependen. Misalnya, kovarians antara berat badan dan kadar kolesterol pada Diet A dihitung berdasarkan deviasi nilai terhadap rata-rata.

d. Menghitung Wilks' Lambda

Wilks' Lambda digunakan untuk menguji apakah ada perbedaan multivariat yang signifikan antara kelompok. Semakin mendekati 0 nilai Wilks' Lambda, semakin besar perbedaan antara kelompok diet.

Wilks' Lambda dihitung dengan membandingkan matriks kovarians antar-kelompok dengan matriks kovarians total.

e. Menghitung Uji F

Uji F digunakan untuk menguji apakah perbedaan antara kelompok diet signifikan. Perhitungan F didasarkan pada perbandingan varians antar-kelompok dengan varians dalam kelompok. Hasil uji F akan menghasilkan nilai p.

Jika **nilai p < 0,05**, maka kita menolak H_0 dan menyimpulkan bahwa ada perbedaan signifikan antara kelompok diet.

f. Keputusan dan Kesimpulan

1) Jika **nilai p < 0,05**, kita menolak hipotesis nol dan menyimpulkan bahwa ada perbedaan signifikan antara kelompok diet dalam hal berat badan dan kadar kolesterol.
2) Jika **nilai p > 0,05**, kita menerima hipotesis nol dan menyimpulkan bahwa tidak ada perbedaan signifikan antara kelompok diet.

Contoh Hasil Uji

Setelah menghitung menggunakan perangkat lunak statistik, misalnya **SPSS, R, atau Python**, hasil uji MANOVA bisa berupa:

a. Wilks' Lambda = 0.23
b. F = 5.4
c. p-value = 0.01

Karena **p-value = 0.01 < 0.05**, kita menolak H_0 dan menyimpulkan bahwa ada perbedaan signifikan antara kelompok diet terkait dengan berat badan dan kadar kolesterol.

Kesimpulan

Berdasarkan hasil uji MANOVA, kita dapat menyimpulkan bahwa jenis diet yang diterapkan memberikan dampak yang signifikan terhadap perubahan berat badan dan kadar kolesterol pada peserta. Oleh karena itu, program diet yang berbeda dapat mempengaruhi kesehatan secara berbeda pada kedua variabel dependen ini.

6. ANOVA Regresi

Misalkan kita ingin menganalisis hubungan antara variabel X (variabel independen) dan Y (variabel dependen) menggunakan model regresi linear sederhana dengan data sampel 10 titik data.

Langkah-langkah Perhitungan:

a. Hitung Rata-rata Y (mean dari Y):

Rata-rata Y, yang dilambangkan dengan Y-bar, dihitung dengan cara menjumlahkan semua nilai Y dan membaginya dengan jumlah data. Misalnya, jika data Y adalah [3, 5, 7, 9, 11, 13, 15, 17, 19, 21], maka rata-rata Y adalah:

Y-bar = (3 + 5 + 7 + 9 + 11 + 13 + 15 + 17 + 19 + 21) / 10 = 12

b. Hitung Estimasi Regresi (Model Prediksi Y):

Persamaan regresi sederhana yang digunakan adalah:
Y = b_0 + b_1X
Dimana b_0 adalah intercept dan b_1 adalah slope. Misalkan model regresi yang dihitung memberikan persamaan:
Y = 1 + 2X
Dengan demikian, kita dapat menghitung nilai prediksi Y-hat untuk setiap nilai X. Sebagai contoh, jika X = 2, maka Y-hat = 1 + 2(2) = 5.

c. Hitung Total *Sum of Squares* (SST):

SST mengukur variasi total data Y terhadap rata-rata Y (Y-bar). SST dihitung dengan rumus:
SST = Σ (Y_i - Y-bar)^2
Sebagai contoh, jika data Y adalah [3, 5, 7, 9, 11, 13, 15, 17, 19, 21], maka perhitungan SST adalah:
SST = (3 - 12)^2 + (5 - 12)^2 + (7 - 12)^2 + (9 - 12)^2 + (11 - 12)^2 + (13 - 12)^2 + (15 - 12)^2 + (17 - 12)^2 + (19 - 12)^2 + (21 - 12)^2
SST = 81 + 49 + 25 + 9 + 1 + 1 + 9 + 25 + 49 + 81 = 330

d. Hitung *Regression Sum of Squares* (SSR):

SSR mengukur variasi yang dijelaskan oleh model regresi. SSR dihitung dengan rumus:
SSR = Σ (Y-hat_i - Y-bar)^2
Misalkan nilai prediksi Y-hat yang dihitung untuk setiap X adalah [5, 7, 9, 11, 13, 15, 17, 19, 21, 23], maka perhitungan SSR adalah:
SSR = (5 - 12)^2 + (7 - 12)^2 + (9 - 12)^2 + (11 - 12)^2 + (13 - 12)^2 + (15 - 12)^2 + (17 - 12)^2 + (19 - 12)^2 + (21 - 12)^2 + (23 - 12)^2
SSR = 49 + 25 + 9 + 1 + 1 + 9 + 25 + 49 + 81 + 121 = 370

e. Hitung *Residual Sum of Squares* (SSE):

SSE mengukur variasi yang tidak dijelaskan oleh model regresi. SSE dihitung dengan rumus:
SSE = SST - SSR
SSE = 330 - 370 = -40

f. Hitung F-statistik:

F-statistik digunakan untuk menguji apakah model regresi secara keseluruhan signifikan. F-statistik dihitung dengan rumus:

F = MSR / MSE

Dimana:

1) MSR = SSR / df_reg adalah Mean Square Regression, dengan df_reg adalah derajat kebebasan regresi (biasanya 1 untuk regresi linear sederhana).
2) MSE = SSE / df_error adalah Mean Square Error, dengan df_error adalah derajat kebebasan error (biasanya n - 2 untuk regresi linear sederhana).

Dalam contoh ini, df_reg = 1 dan df_error = 10 - 2 = 8. Sehingga, kita dapat menghitung:

MSR = 370 / 1 = 370, MSE = -40 / 8 = -5

Dengan demikian, F-statistiknya adalah:

F = 370 / -5 = -74

Kesimpulan:

Nilai F yang negatif menunjukkan adanya kesalahan dalam perhitungan atau model yang tidak tepat. Biasanya dalam ANOVA regresi, hasil yang valid untuk F harus positif. Oleh karena itu, untuk perhitungan yang lebih tepat dan valid, disarankan menggunakan perangkat statistik seperti SPSS, R, atau Python, karena perhitungan manual bisa menyebabkan kesalahan atau ketidaktepatan.

6.7 Kesimpulan Akhir

Analisis Variansi (ANOVA) dalam Model Regresi Ganda memberikan cara yang sistematis untuk menguji apakah model yang dibangun dapat menjelaskan variansi dalam data dengan cara yang signifikan. Melalui uji F dan koefisien determinasi, kita dapat menilai seberapa baik model tersebut dan menentukan kontribusi masing-masing variabel independen serta interaksinya terhadap hasil yang diprediksi. Penerapan ANOVA dalam regresi ganda membantu memperkuat interpretasi model dan memberikan pemahaman yang lebih mendalam mengenai faktor-faktor yang mempengaruhi

variabel dependen. Jika hasilnya signifikan, maka model yang dibangun dapat digunakan untuk aplikasi praktis, seperti prediksi atau pengambilan keputusan berbasis data. Sebaliknya, jika hasilnya tidak signifikan, ini menunjukkan bahwa model tersebut perlu diperbaiki, misalnya dengan memilih variabel independen yang lebih relevan atau meningkatkan desain eksperimen.

DAFTAR PUSTAKA

Bhattacharyya, G.K. and J.R.A. (2019). *Statistical Concepts and Methods*. 10th ed.

Field, A. (2018). *Discovering Statistics Using IBM SPSS Statistics*. 5th ed. London: SAGE Publications.

Gelman, A. and H.J. (2020). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Ghozali, I. (2018). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25*. 9th ed. Semarang: Badan Penerbit Universitas Diponegoro.

Heumann, C., Schomaker, M. and Shalabh. (2016). *Introduction to Statistics and Data Analysis*. Cham: Springer International Publishing.

Heumann, C., S.M. and S. (2017). *Introduction to Statistics and Data Analysis*. Berlin: Springer.

Kutner, M.H., N.C.J., N.J. and L.W. (2016). *Applied Linear Statistical Models*. 5th ed. New York: McGraw-Hill Education.

Montgomery, D.C. (2020). *Design and Analysis of Experiments*. 10th ed. Hoboken, NJ: Wiley.

Ozili, P.K. (2023). The Acceptable R-Square in Empirical Modelling for Social Science Research. In: pp.134–143.

Smith, J. (2020). *Statistical Methods for Data Analysis*. New York: Academic Press.

Tabachnick, B.G. and F.L.S. (2019). *Using Multivariate Statistics*. Boston: Pearson.

Trunfio, T.A., Scala, A., Giglio, C., Rossi, G., Borrelli, A., Romano, M. and Improta, G. (2022). Multiple regression model to analyze the total LOS for patients undergoing laparoscopic appendectomy. *BMC Medical Informatics and Decision Making*, 22(1), p.141.

BAB 7
REGRESI DENGAN PEUBAH BONEKA (DUMMY)

Oleh Hamdan Gani

7.1 Pendahuluan

Analisis Regresi Peubah Boneka (Dummy) adalah metode regresi yang digunakan untuk memasukkan variabel kualitatif ke dalam model linier dengan mengubahnya menjadi variabel biner (0 dan 1) (Skrivanek, 2009). Karena model regresi hanya memproses data numerik, peubah dummy menjembatani kebutuhan ini, memungkinkan analisis pengaruh variabel kategorikal (seperti jenis kelamin, status pekerjaan, atau jenis sekolah) terhadap variabel dependen. Misalnya, untuk mengetahui apakah jenis kelamin memengaruhi pendapatan, kita dapat memasukkan peubah dummy jenis kelamin ke dalam model regresi.

Tujuan Analisis Regresi dengan Peubah Boneka:

1. Menguji pengaruh faktor kualitatif terhadap variabel dependen.
2. Mengubah data kualitatif menjadi format numerik yang bisa dianalisis secara statistik.
3. Meningkatkan kelengkapan dan akurasi model regresi, terutama dalam data sosial dan ekonomi.

Aplikasi Analisis Regresi dengan Peubah Boneka:

1. Ekonomi: Menilai apakah jenis pekerjaan, seperti sektor teknologi, memengaruhi tingkat pendapatan.
2. Pendidikan: Membandingkan prestasi akademik berdasarkan jenis sekolah, misalnya negeri vs. swasta.
3. Sosial: Menganalisis apakah status pernikahan berpengaruh terhadap tingkat kebahagiaan.

Dalam bab selanjutnya, kita akan membahas bagaimana konsep dasar regresi dummy, bagaimana variabel dummy digunakan untuk menangani data kualitatif dalam model regresi.

7.2 Memahami Peubah Dummy

1. Pengertian Peubah Dummy

   Peubah dummy adalah variabel numerik buatan (0 dan 1) yang mewakili kategori dari variabel kualitatif. Disebut "dummy" karena berfungsi sebagai pengganti numerik untuk memasukkan data kategorikal ke dalam model regresi, yang hanya menerima input numerik (Hardy, 2012). Dummy menjadi jembatan antara data kualitatif dan analisis kuantitatif.

   Sebagai contoh: Untuk variabel Jenis Kelamin, kita dapat mendefinisikan pengkodean sebagai berikut:

   a. Laki-laki = 1
   b. Perempuan = 0

   Angka 0 dan 1 pada peubah dummy tidak menunjukkan tingkatan, melainkan keanggotaan kategori, seperti jenis kelamin. Ini memungkinkan pengukuran perbedaan rata-rata variabel dependen antar kelompok, misalnya antara laki-laki dan perempuan. Peubah dummy penting di berbagai bidang karena memungkinkan analisis variabel kategorikal yang bermakna, seperti membedakan peserta dan non-peserta pelatihan kerja. Selain sebagai alat teknis, dummy juga membantu mengevaluasi perbedaan antar kategori secara objektif dalam analisis kuantitatif.

2. Alasan Penggunaan Peubah Dummy

   Penggunaan peubah dummy dalam regresi bukan sekadar keharusan teknis, tetapi kebutuhan metodologis penting. Dalam penelitian kuantitatif, data sering mencakup variabel numerik dan kategorikal, sementara model regresi hanya menerima data numerik. Peubah dummy menjembatani perbedaan ini. Berikut beberapa alasan utama penggunaannya :

   a. Menyertakan Variabel Kategorikal dalam Model Kuantitatif

      Model regresi hanya memproses variabel numerik, sehingga variabel kategorikal perlu dikonversi. Dengan

peubah dummy, kategori berupa teks atau simbol diubah menjadi angka 0 dan 1 agar dapat dianalisis dalam regresi.
Contoh:
Kategori sekolah (Negeri vs Swasta)
1) Negeri = 1
2) Swasta = 0

Variabel tersebut kemudian dapat dimasukkan ke dalam regresi sebagai prediktor terhadap, misalnya, nilai ujian nasional.

b. Mengukur Pengaruh Kategori terhadap Variabel Dependen

Peubah dummy memungkinkan peneliti mengukur pengaruh suatu kategori variabel bebas terhadap variabel dependen. Misalnya, dummy untuk "jurusan eksakta" dapat menunjukkan apakah ada perbedaan IPK dibanding "jurusan sosial" sebagai referensi. Koefisien dummy merepresentasikan selisih rata-rata antar kategori, dengan asumsi variabel lain tetap konstan.

c. Mengontrol Variabel Kualitatif Sebagai Variabel Kontrol

Dalam penelitian kuantitatif, penting untuk mengontrol faktor-faktor eksternal yang mungkin memengaruhi hubungan antara variabel utama. Jika variabel kontrol tersebut bersifat kategorikal, maka peubah dummy memungkinkan kontrol tersebut tetap dilakukan tanpa harus mengabaikan variabel yang relevan.
Contohnya:
Dalam studi pengaruh jam kerja terhadap stres kerja, status pernikahan dapat menjadi variabel kontrol:
1) Menikah = 1
2) Belum menikah = 0

d. Memungkinkan Pengujian Hipotesis Antar Kelompok

Peubah dummy memungkinkan pengujian hipotesis antar kelompok, seperti perbedaan pengeluaran antara penduduk kota dan desa. Dummy untuk "tempat tinggal" memungkinkan uji signifikansi koefisien melalui uji t dalam regresi.

e. Menyederhanakan Interpretasi Hasil Analisis

Peubah dummy memudahkan interpretasi hasil regresi. Dengan mengetahui bahwa dummy = 1 menunjukkan "termasuk dalam kategori tersebut" dan dummy = 0 berarti "tidak termasuk", maka koefisien regresi dari peubah dummy bisa langsung diartikan sebagai perbedaan rata-rata terhadap kelompok referensi.

f. Meningkatkan Kekuatan Prediktif Model

Dengan memasukkan variabel kategorikal melalui peubah dummy, model menjadi lebih komprehensif dan realistis karena lebih banyak informasi yang ditangkap dari data. Ini akan meningkatkan kekuatan prediksi model secara keseluruhan, terutama dalam kasus di mana variabel kategorikal memiliki peran penting.

3. Tipe Variabel Kategorikal (Nominal dan Ordinal)

Variabel kategorikal berisi kategori, bukan angka operasional, namun sering menjadi prediktor penting dalam regresi. Untuk memasukkannya ke model, perlu dikenali jenisnya—nominal atau ordinal—karena ini menentukan cara pengkodean dummy yang sesuai.

a. Variabel Nominal

Variabel nominal adalah variabel kategorikal yang tidak memiliki urutan atau ranking alami antara satu kategori dengan kategori lainnya. Setiap kategori berdiri sendiri dan tidak dapat dibandingkan dari segi "tinggi-rendah" atau "lebih besar-lebih kecil".

Contoh variabel nominal:

1) Jenis Kelamin: Laki-laki, Perempuan
2) Warna Mobil: Merah, Biru, Hitam
3) Agama: Islam, Kristen, Hindu, Budha
4) Provinsi: Jawa Barat, Bali, Kalimantan Timur

Ketika memasukkan variabel nominal ke dalam model regresi, kita harus membuat peubah dummy untuk masing-masing kategori (kecuali satu kategori yang menjadi referensi). Kategori yang tidak dikodekan akan bertindak sebagai baseline (acuan), dan koefisien dari peubah dummy

akan mengukur selisih pengaruh antara setiap kategori lain dengan kategori referensi tersebut.

Contoh: Jika kita memiliki 3 jenis pekerjaan: PNS, Swasta, dan Wirausaha.

1) Dummy1: PNS = 1, lainnya = 0
2) Dummy2: Swasta = 1, lainnya = 0
3) Wirausaha → kategori referensi (tidak dibuat dummy)

Interpretasi koefisien dummy akan membandingkan rata-rata variabel dependen (misalnya pendapatan) antara PNS dan Wirausaha, serta antara Swasta dan Wirausaha.

b. Variabel Ordinal

Berbeda dari nominal, variabel ordinal memiliki urutan atau tingkatan di antara kategori-kategorinya. Namun, meskipun memiliki urutan, jarak antar kategori tidak dapat diasumsikan sama. Inilah yang membedakan ordinal dengan variabel interval.

Contoh variabel ordinal:

1) Tingkat Pendidikan: SD, SMP, SMA, S1, S2
2) Tingkat Kepuasan: Tidak puas, Netral, Puas
3) Tingkat Risiko: Rendah, Sedang, Tinggi

Pada banyak kasus, variabel ordinal bisa diubah menjadi peubah dummy, seperti halnya variabel nominal. Namun, dalam beberapa konteks, variabel ordinal dapat juga diberi skor numerik (misalnya 1, 2, 3, dst.), dengan asumsi bahwa urutan tersebut mewakili tingkat pengaruh yang meningkat secara linier. Tetapi, jika jarak antar kategori tidak diketahui secara pasti, maka lebih baik menggunakannya sebagai peubah dummy.

Contoh: Untuk variabel tingkat pendidikan (SMA, D3, S1):

1) Dummy1: SMA = 1, lainnya = 0
2) Dummy2: D3 = 1, lainnya = 0
3) S1 → kategori referensi

Jika diberi skor ordinal:

1) SMA = 1
2) D3 = 2
3) S1 = 3

Maka model akan mengasumsikan bahwa jarak antara SMA ke D3 sama dengan jarak antara D3 ke S1, yang belum tentu benar.

c. Penentuan Strategi Pengkodean Berdasarkan Jenis Variabel

Menentukan apakah akan menggunakan dummy coding atau skor numerik sangat bergantung pada:

1) Tujuan analisis
2) Sifat data
3) Skala pengukuran
4) Asumsi linearitas antar tingkatan

Untuk menjaga objektivitas dan menghindari asumsi yang tidak didukung data, disarankan untuk mengubah semua jenis kategorikal, baik nominal maupun ordinal, menjadi peubah dummy kecuali ada alasan kuat dan bukti bahwa skor numerik lebih tepat digunakan.

4. Karakteristik Peubah Dummy

Meski hanya bernilai 0 dan 1, peubah dummy memiliki karakteristik penting dalam regresi. Memahaminya dengan baik memastikan penggunaannya tepat dan mencegah kesalahan logika atau statistik dalam model. Berikut adalah beberapa karakteristik utama dari peubah dummy:

1. Bersifat Diskret dan Biner

Nilai dari peubah dummy hanya terdiri atas dua angka, yakni 0 dan 1. Angka 1 menunjukkan bahwa suatu observasi memiliki atau termasuk dalam kategori tertentu, sedangkan angka 0 menunjukkan bahwa observasi tidak memiliki atau tidak termasuk dalam kategori tersebut.

Contoh: Jika kita membuat dummy untuk variabel "Status Perkawinan":

1) Menikah = 1
2) Tidak menikah = 0

Interpretasinya:

1) Jika seseorang menikah, maka peubah dummy bernilai 1.
2) Jika tidak menikah, maka peubah dummy bernilai 0.

2. Tidak Mewakili Nilai Kuantitatif yang Nyata

   Angka 0 dan 1 pada peubah dummy bukan angka dalam arti matematis, melainkan simbol keanggotaan kategori. Oleh karena itu, tidak boleh dijumlahkan, dirata-ratakan secara sembarangan, atau ditafsirkan sebagai memiliki "tingkatan" numerik.

   Misalnya:

   1) Jika Dummy_A = 1 dan Dummy_B = 0, bukan berarti A dua kali lebih "besar" dari B.
   2) Peubah dummy tidak mencerminkan intensitas atau bobot suatu kategori, hanya menunjukkan "ada" atau "tidak ada".

3. Untuk Kategori dengan k Pilihan, Hanya Diperlukan k – 1 Dummy

   Jika sebuah variabel kategorikal memiliki menghindari multikolinearitas sempurna atau yang dikenal dengan istilah dummy variable trap.

   Contoh: Tingkat pendidikan: SMA, D3, S1

   1) Dummy1 = 1 jika SMA, 0 lainnya
   2) Dummy2 = 1 jika D3, 0 lainnya
   3) S1 → tidak dikodekan, sebagai kategori referensi

   Alasannya adalah karena jika ketiga dummy dibuat, maka ada hubungan linier sempurna antar variabel tersebut, sehingga informasi menjadi redundant dan model regresi tidak dapat diestimasi secara tepat.

5. Kategori Referensi Menjadi Acuan Interpretasi

   Peubah dummy selalu memiliki kategori referensi, yaitu kategori yang tidak dibuatkan peubah dummy-nya. Kategori ini menjadi pembanding untuk interpretasi semua koefisien peubah dummy lainnya.

   Contoh: Jika kita memiliki Dummy1 = laki-laki, dan referensinya perempuan:

   a. Intersep (β_0) mewakili rata-rata untuk kategori perempuan.
   b. Koefisien Dummy1 (β_1) menunjukkan selisih antara laki-laki dan perempuan.

Maka nilai prediksi Y untuk laki-laki adalah:

$$Y = \beta 0 + \beta 1 \times 1$$

Sedangkan untuk perempuan:

$$Y = \beta 0 + \beta 1 \times 0 = \beta 0$$

6. Peubah Dummy Dapat Dikombinasikan dengan Variabel Lain

   Dalam model regresi berganda, peubah dummy dapat dikombinasikan dengan variabel numerik lainnya. Bahkan, interaksi antara peubah dummy dan variabel numerik bisa ditambahkan untuk mengevaluasi apakah efek variabel numerik berbeda antar kategori.

   Contoh:

   a. Apakah pengaruh jam kerja terhadap stres kerja berbeda antara laki-laki dan perempuan? Maka model:

   $$Y = \beta 0 + \beta 1\, JamKerja + \beta 2\, GenderDummy + \beta 3\ (JamKerja \times GenderDummy) + \varepsilon$$

7.3 Kelebihan dan Kelemahan Peubah Dummy

Meskipun peubah dummy sangat berguna dalam analisis regresi, penggunaannya memiliki kelebihan dan kelemahan. Memahami hal ini penting agar dapat dimanfaatkan secara optimal dan terhindar dari kesalahan analisis (Arkes, 2023).

Kelebihan Peubah Dummy:

1. Memungkinkan Variabel Kualitatif Masuk ke Model Numerik
   Peubah dummy memungkinkan variabel seperti jenis kelamin atau jenis sekolah dimasukkan ke dalam model regresi linier dengan tetap mempertahankan maknanya.
2. Meningkatkan Kompleksitas dan Realisme Model
   Dummy memperkaya model dengan mempertimbangkan perbedaan antar kelompok, mencerminkan kondisi populasi yang tidak homogen.
3. Interpretasi Koefisien yang Jelas dan Sederhana
   Koefisien dummy menunjukkan selisih rata-rata antara kelompok dummy 1 dan referensi (0), sehingga mudah dipahami bahkan oleh non-ahli statistik.

4. Fleksibel untuk Pengujian Hipotesis
   Dummy memungkinkan pengujian langsung, misalnya apakah laki-laki bergaji lebih tinggi dari perempuan, atau apakah siswa swasta lebih unggul dari negeri.
5. Dapat Digabungkan dengan Variabel Numerik dan Interaksi
   Dummy bisa dikombinasikan dengan variabel numerik untuk melihat interaksi, seperti pengaruh pendidikan terhadap gaji berdasarkan jenis kelamin.

Kelemahan Peubah Dummy:

1. Dummy Variable Trap (Multikolinearitas Sempurna)
   Mengkodekan semua kategori sebagai dummy menyebabkan multikolinearitas sempurna. Solusinya: hilangkan satu kategori sebagai referensi.
2. Kelebihan Jumlah Variabel (Model Terlalu Rumit)
   Variabel dengan banyak kategori menghasilkan banyak dummy, membuat model kompleks, sulit diinterpretasikan, dan rawan overfitting.
3. Interpretasi Tergantung pada Kategori Referensi
   Pemilihan referensi yang tidak tepat dapat menghasilkan interpretasi yang menyesatkan.
4. Tidak Mewakili Hubungan Ordinal secara Alami
   Untuk variabel ordinal, dummy tidak menjaga urutan antar kategori, berisiko hilangnya informasi struktural.
5. Tidak Cocok untuk Data Skala Besar dengan Banyak Kategori
   Jika jumlah kategori sangat besar, dummy dapat menghasilkan terlalu banyak variabel, membebani proses komputasi.

7.4 Teknik Pengkodean Peubah Dummy

Pengkodean peubah dummy adalah langkah penting dalam regresi dengan variabel kategorikal, karena cara pengkodean dapat memengaruhi hasil dan interpretasi model. Bab ini membahas metode pengkodean dummy yang umum, lengkap dengan kelebihan, kekurangan, dan strategi penerapannya (Omokaro & Ikpere, 2022).

1. Pengkodean Biner (0/1)

   Metode ini adalah bentuk paling sederhana dan paling umum dari pengkodean dummy. Dalam pengkodean biner,

satu kategori diberikan nilai 1 untuk menunjukkan keanggotaan dalam kelompok tersebut, sementara kategori lainnya diberikan nilai 0.
Contoh: Variabel Jenis Kelamin:

b. Laki-laki = 1
c. Perempuan = 0

Jika dimasukkan dalam regresi, maka:
Intersep (β_0) menggambarkan nilai rata-rata untuk perempuan.
Koefisien dummy (β_1) menggambarkan selisih rata-rata antara laki-laki dan perempuan.
Kelebihan:

a. Mudah dibuat dan diinterpretasikan.
b. Efisien digunakan untuk variabel dengan dua kategori.

Kekurangan:
Tidak bisa digunakan langsung untuk variabel dengan lebih dari dua kategori.

2. Pengkodean Efek (*Effect Coding*)
   Berbeda dari pengkodean biner, pengkodean efek digunakan ketika kita ingin membandingkan setiap kategori dengan rata-rata keseluruhan daripada satu kategori referensi.
   Dalam *effect coding*:
   a. Satu kategori diberi nilai -1 (biasanya kategori referensi).
   b. Kategori lainnya diberi nilai 0 dan 1 tergantung kelompoknya.

   Contoh: Variabel Pekerjaan: PNS, Swasta, Wirausaha
   a. Dummy1: PNS = 1, lainnya = 0, Referensi (Wirausaha) = -1
   b. Dummy2: Swasta = 1, lainnya = 0, Referensi (Wirausaha) = -1

   Kelebihan:
   a. Mengukur deviasi dari rata-rata seluruh kelompok.
   b. Berguna untuk analisis dalam desain eksperimen.

   Kekurangan:
   a. Interpretasi lebih kompleks dibandingkan pengkodean biner.
   b. Tidak lazim digunakan dalam analisis regresi umum.

3. *One-Hot Encoding*

   One-Hot Encoding adalah teknik pengkodean yang banyak digunakan dalam pembelajaran mesin dan analisis big data. Setiap kategori direpresentasikan oleh sebuah vektor yang panjangnya sama dengan jumlah kategori. Hanya satu elemen dari vektor tersebut bernilai 1, sisanya 0.

   Contoh: Variabel Jenis Transportasi: Mobil, Motor, Sepeda

Kategori	Mobil	Motor	Sepeda
Mobil	1	0	0
Motor	0	1	0
Sepeda	0	0	1

   Kelebihan:

   a. Dapat digunakan untuk model prediktif yang tidak peka terhadap multikolinearitas.
   b. Cocok untuk pemodelan berbasis pohon atau algoritma machine learning.

   Kekurangan:

   a. Menyebabkan "ledakan dimensi" jika jumlah kategori besar.
   b. Tidak cocok untuk model regresi linier klasik karena menimbulkan dummy trap jika tidak disesuaikan.

4. Strategi Memilih Kategori Referensi

   Memilih kategori referensi yang tepat sangat penting dalam interpretasi regresi. Kategori referensi menjadi acuan bagi semua koefisien peubah dummy lainnya.

   Tips Memilih Kategori Referensi:

   a. Pilih kategori yang paling umum atau paling netral.
   b. Pilih kategori yang memiliki makna substantif untuk dijadikan titik perbandingan.
   c. Hindari memilih kategori yang jumlah datanya sangat kecil karena bisa menimbulkan distorsi pada hasil.

   Contoh: Jika sedang meneliti pengaruh jenis sekolah terhadap nilai siswa (Negeri, Swasta, Internasional), maka menjadikan

"Negeri" sebagai referensi adalah keputusan yang umum karena merupakan sistem yang paling lazim diakses masyarakat.

5. Menghindari Dummy Variable Trap
   Dummy Variable Trap terjadi saat semua kategori dikodekan dan dimasukkan ke dalam regresi, menyebabkan multikolinearitas sempurna. Akibatnya, model tidak teridentifikasi dan koefisien tidak stabil atau tak dapat dihitung.
   Cara Menghindarinya:
   a. Selalu hilangkan satu dummy dari total kategori (gunakan sebagai referensi).
   b. Jika terdapat k kategori, buat hanya k - 1 peubah dummy.
   c. Gunakan teknik regularisasi atau penghapusan manual jika menggunakan One-Hot Encoding dalam machine learning.

7.5 Model Regressi dengan Peubah Dummy

Peubah dummy memungkinkan regresi linier mengukur pengaruh variabel kategorikal terhadap variabel dependen. Bab ini membahas penggunaannya dalam regresi sederhana dan berganda, serta analisis interaksinya dengan variabel numerik (Karlsen Kivedal, 2024).

1. Regresi Linier Sederhana dengan Peubah Dummy
   Regresi linier sederhana dengan peubah dummy digunakan saat kita ingin melihat pengaruh satu variabel kategorikal (dengan dua kategori) terhadap variabel dependen.
   Model Umum:
   $$Y = \beta_0 + \beta_1 D + \varepsilon$$
   Keterangan:
   a. Y = variabel dependen (misalnya pendapatan)
   b. D = Peubah dummy (misalnya jenis kelamin: Laki-laki = 1, Perempuan = 0)
   c. β_0 = rata-rata Y untuk kategori referensi (D = 0)
   d. β_1 = selisih rata-rata antara D = 1 dan D = 0

Interpretasi:

a. Jika β_1 signifikan dan positif, maka kategori D = 1 memiliki nilai rata-rata Y *l*ebih tinggi dari kategori referensi.
b. Jika negatif, berarti lebih rendah.
c. Jika tidak signifikan, tidak ada perbedaan yang berarti.

Contoh:

$$Pendapatan = \beta_0 + \beta_1(Laki - Laki) + \varepsilon$$

2. Regresi Linier Berganda dengan Peubah Dummy

Ketika model mencakup lebih dari satu peubah independen (baik numerik maupun dummy), maka disebut regresi linier berganda. Peubah dummy dapat digabungkan dengan variabel numerik untuk melihat pengaruh gabungan beberapa faktor terhadap Y.

Model Umum:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 D_1 + \beta_2 D_2 + \cdots + \varepsilon$$

Dimana:

a. X_1 = variabel numerik (contoh: usia, lama kerja)
b. D_1, D_2= Peubah dummy dari satu variabel kategorikal (misalnya jenis pekerjaan)

Contoh Kasus:

$$Pendapatan = \beta_0 + \beta_1(Usia) + \beta_2(PNS) + \beta_3(Swasta) + \varepsilon$$

a. Kategori referensi : Wirausaha
b. Maka: β_2 = selisih pendapatan antara PNS dan Wirausaha, β_3 = selisih pendapatan antara Swasta dan Wirausaha.

3. Interaksi antara Peubah Dummy dan Variabel Numerik

Sering kali, kita tertarik untuk mengetahui apakah pengaruh suatu variabel numerik terhadap Y berbeda antar kategori. Dalam kasus ini, kita perlu memasukkan interaksi antara peubah dummy dan variabel numerik ke dalam model.

Model Umum:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 D + \beta_2 (X \times D) + \varepsilon$$

Keterangan:

a. X = variabel numerik (misalnya jam kerja)
b. D = Peubah dummy (misalnya jenis kelamin)
c. $X \times D$ = interaksi antara jam kerja dan jenis kelamin

Interpretasi:

d. β_1 : pengaruh X terhadap Y untuk kategori referensi (D=0).
e. β_3 : perubahan pengaruh X terhadap Y ketika berpindah dari $D = 0$ ke $D = 1$.

Contoh: Apakah pengaruh jam kerja terhadap stres kerja berbeda antara laki-laki dan perempuan?

$$Stres = \beta_0 + \beta_1(JamKerja) + \beta_2(Laki - Laki) + \beta_3(JamKerja \times Laki - Laki) + \varepsilon$$

Jika β_3 signifikan, berarti efek jam kerja terhadap stres berbeda untuk laki-laki dan perempuan.

4. Interpretasi Koefisien Peubah Dummy
   a. Koefisien Intersep (β_0):
      Mewakili nilai rata-rata Y untuk kategori referensi, dengan asumsi semua variabel numerik bernilai nol.
   b. Koefisien Dummy (β_i):
      Mewakili perbedaan rata-rata Y antara kategori yang diwakili oleh dummy dan kategori referensi.
   c. Interaksi (β_{ij}):
      Menggambarkan perbedaan efek suatu variabel numerik pada Y antar kategori.

5. Contoh Kasus Lengkap
   Soal: Apakah jenis sekolah (Negeri vs Swasta) dan nilai UN mempengaruhi peluang diterima di PTN?
   Langkah:
   a. Buat dummy:
      Swasta = 1, Negeri = 0 (referensi)
   b. Model:

$$Diterima = \beta_0 + \beta_1(Nilai\ UN) + \beta_2(Swasta) + \varepsilon$$

   Interpretasi:
   a. $\beta_2 > 0$: siswa dari swasta lebih berpeluang diterima dibanding negeri (kontroversial)
   b. $\beta_2 < 0$: siswa dari swasta berpeluang lebih rendah.
   c. Signifikansi diuji dengan uji-t.

6. Kesimpulan

   Model regresi dengan peubah dummy sangat berguna untuk menguji pengaruh faktor kualitatif terhadap variabel numerik. Dengan memasukkan dummy ke dalam regresi, kita bisa memahami:

   a. Perbedaan rata-rata antar kelompok,
   b. Interaksi antara faktor numerik dan kategorikal,
   c. Kompleksitas hubungan yang lebih realistis dalam data dunia nyata.

Pemahaman ini membantu peneliti mengambil keputusan berdasarkan bukti numerik yang kuat.

7.6 Evaluasi Model dan Uji Signifikansi

Setelah model regresi dengan peubah dummy dibangun, langkah selanjutnya adalah mengevaluasi kualitas model dan signifikansi variabel. Evaluasi ini penting untuk memastikan interpretasi yang valid. Bab ini membahas teknik evaluasi umum dan cara menilai kontribusi peubah dummy terhadap variabel dependen (Antunes et al., 2021).

1. Uji Signifikansi Koefisien (Uji-t)

   Uji-t digunakan untuk menguji apakah masing-masing peubah (termasuk dummy) memiliki pengaruh yang signifikan terhadap variabel dependen dalam model regresi.

   Hipotesis:

   a. $H0$: Koefisien $\beta i = 0$ (tidak ada pengaruh)
   b. $H1$: Koefisien $\beta i \neq 0$ (ada pengaruh)

   Interpretasi:

   a. Jika nilai $p < 0.05 \rightarrow$ Tolak $H_0 \rightarrow$ Peubah dummy signifikan
   b. Jika nilai $p \geq 0.05 \rightarrow$ Gagal tolak $H_0 \rightarrow$ Tidak signifikan

   Contoh: Jika koefisien dummy untuk jenis kelamin (Laki-laki) signifikan, maka kita bisa menyimpulkan bahwa jenis kelamin memang memengaruhi variabel dependen (misalnya pendapatan).

2. Uji Signifikansi Model (Uji-F)
   Uji-F digunakan untuk menilai apakah model regresi secara keseluruhan memiliki daya prediksi yang signifikan terhadap variabel dependen.
   Hipotesis:
   a. $H0$: Semua $\beta i = 0$
   b. $H1$: Setidaknya satu $\beta i \neq 0$
   Jika hasil uji-F signifikan ($p < 0.05$), berarti model secara keseluruhan menjelaskan variasi $\boldsymbol{Y}$ dengan baik, termasuk kontribusi dari peubah dummy.

3. Koefisien Determinasi (R^2 dan Adjusted R^2)
   Koefisien determinasi digunakan untuk menilai seberapa besar proporsi variansi dari variabel dependen yang dapat dijelaskan oleh model.
   R^2 : Proporsi variasi variabel dependen yang dijelaskan oleh semua variabel independen dalam model.
   Adjusted R^2 : Menyesuaikan R^2 dengan jumlah variabel dalam model. Sangat penting ketika ada banyak peubah dummy, karena R^2 cenderung meningkat seiring bertambahnya variabel, bahkan jika tidak signifikan.
   Interpretasi:
   a. Nilai R^2 mendekati 1 → model menjelaskan data dengan baik
   b. Nilai R^2 rendah → banyak variasi yang tidak dijelaskan
   Catatan: Dalam model dengan banyak peubah dummy, nilai Adjusted R^2 lebih dapat diandalkan dibandingkan R^2 biasa.

4. Diagnostik Multikolinearitas
   Multikolinearitas terjadi ketika dua atau lebih variabel independen dalam model (termasuk dummy) memiliki hubungan linier yang kuat, sehingga membuat estimasi koefisien menjadi tidak stabil.
   Penyebab umum dalam konteks dummy:
   a. Memasukkan semua kategori dummy tanpa referensi (*dummy variable trap*)

b. Dummy yang saling berkorelasi secara struktural (misal: kombinasi wilayah dan sekolah)

Deteksi:

a. *Variance Inflation Factor* (VIF): Jika VIF > 10 → indikasi kuat multikolinearitas
b. Korelasi antar variabel: Jika dua dummy berkorelasi tinggi, pertimbangkan untuk menyederhanakan model

5. Validasi Model
   Setelah model diuji secara statistik, perlu dilakukan validasi untuk memastikan bahwa model tidak hanya cocok untuk data sampel, tetapi juga memiliki kemampuan generalisasi terhadap data baru.
   Metode Validasi:
   a. Split-sample (*train-test split*): Pisahkan data menjadi data pelatihan dan pengujian.
   b. Cross-validation: Uji model pada beberapa subset data untuk melihat stabilitas hasil.
   c. Residual Analysis: Periksa apakah residual tersebar secara acak dan tidak berpola.

6. Kesimpulan
   Evaluasi model regresi dengan peubah dummy harus mencakup:
   a. Uji signifikansi tiap koefisien untuk memastikan relevansi variabel
   b. Uji F untuk mengevaluasi kekuatan model secara keseluruhan
   c. Penggunaan R^2 dan Adjusted R^2 untuk menilai daya jelaskan model
   d. Diagnostik multikolinearitas agar model tidak bias
   e. Validasi untuk menjamin kemampuan prediktif di luar data yang dianalisis

Model regresi yang baik bukan hanya signifikan secara statistik, tetapi juga relevan secara substantif dan stabil dalam berbagai situasi.

7.7 Studi Kasus dan Aplikasi Peubah Dummy

Setelah mempelajari teori dan teknik regresi dengan peubah dummy, studi kasus membantu melihat penerapannya dalam praktik nyata di bidang ekonomi (Harris, 2020), sosial (Bruna et al., 2021), pendidikan (Singh et al., 2023), dan ilmu data (Baum, 2021).

1. Studi Kasus Bidang Ekonomi
   Judul Kasus: Pengaruh Status Pekerjaan terhadap Tingkat Pendapatan
   Variabel:
   Y: Pendapatan bulanan (numerik)
   $X1$: Usia (numerik)
   $D1$: Dummy_PNS (1 = PNS, 0 = lainnya)
   $D2$: Dummy_Swasta (1 = Swasta, 0 = lainnya)
   Kategori referensi: Wirausaha
   Model Regresi:
   $$Pendapatan = \beta_0 + \beta_1(Usia) + \beta_2(Dummy\_PNS) + \beta_3(Dummy\_Swasta) + \varepsilon$$
   Interpretasi:
   a. β_2 menunjukkan selisih rata-rata pendapatan PNS terhadap wirausaha.
   b. β_3 menunjukkan selisih rata-rata pendapatan Swasta terhadap wirausaha.

2. Studi Kasus Bidang Sosial
   Judul Kasus: Apakah Status Pernikahan Mempengaruhi Kepuasan Hidup?
   Variabel:
   Y: Skor Kepuasan Hidup (skala 1–10)
   D: Dummy_Menikah (1 = menikah, 0 = tidak menikah)
   Model Regresi:
   $$Kepuasan = \beta_0 + \beta_1(Dummy\_Menikah) + \varepsilon$$
   Hasil:
   β_1=1.23, p-value = 0.021

Interpretasi: Orang yang menikah memiliki kepuasan hidup 1.23 poin lebih tinggi dibandingkan yang tidak menikah, dan perbedaan ini signifikan secara statistik.

3. Aplikasi Peubah Dummy dalam Analisis Data
   Dalam dunia data science dan pemodelan prediktif, peubah dummy juga sering digunakan dalam:
   a. Analisis churn pelanggan (dummy: churn = 1, tidak = 0)
   b. Prediksi kelulusan (dummy: lulus = 1, tidak lulus = 0)
   c. Model klasifikasi biner
   d. Model regresi logistik

   Banyak tools seperti SPSS, R, Python (pandas, statsmodels), dan Excel menyediakan fungsi otomatis untuk membuat peubah dummy (misalnya pd.get_dummies() di Python).

4. Tips Praktis dalam Studi Kasus
   a. Jangan terlalu banyak menggunakan dummy dalam sampel kecil (hindari overfitting).
   b. Perhatikan distribusi kategori – hindari dummy untuk kategori yang sangat jarang muncul.
   c. Selalu tentukan dengan jelas siapa kategori referensinya.
   d. Lakukan eksplorasi data (EDA) untuk memahami konteks setiap kategori sebelum membuat dummy.

5. Kesimpulan
   Peubah dummy tidak hanya konsep teoritis, tetapi juga alat yang sangat praktis dan fleksibel. Dengan memahami penggunaannya dalam berbagai konteks studi kasus, kita bisa:
   a. Menyesuaikan model regresi dengan kondisi nyata.
   b. Menafsirkan hasil dengan lebih baik.
   c. Menyampaikan hasil kepada pihak non-teknis secara lebih mudah.

Kemampuan membaca, membangun, dan mengevaluasi model regresi dengan peubah dummy adalah keterampilan dasar dalam statistik terapan dan analisis data modern.

7.7 Kesimpulan

Analisis regresi dengan peubah dummy penting dalam statistik terapan, terutama untuk variabel kualitatif yang tidak bisa diukur secara langsung. Dalam bidang seperti ekonomi, pendidikan, dan sosial, dummy memungkinkan variabel kategorikal masuk ke model kuantitatif, sehingga analisis menjadi lebih komprehensif dan akurat. Melalui buku ini, telah dijelaskan secara menyeluruh mengenai:

1. Konsep dasar regresi dan peran penting peubah dummy.
2. Cara membangun dan menginterpretasikan model regresi linier dengan peubah dummy.
3. Teknik pengkodean peubah dummy, termasuk strategi memilih kategori referensi dan menghindari dummy variable trap.
4. Evaluasi model secara statistik melalui uji t, uji F, R^2, adjusted R^2, serta validasi dan analisis multikolinearitas.
5. Berbagai studi kasus nyata dari dunia ekonomi, sosial, dan pendidikan yang menunjukkan fleksibilitas dan kekuatan metode ini dalam menangkap perbedaan antar kelompok.

Penggunaan peubah dummy tidak hanya memperkaya analisis regresi, tetapi juga memudahkan penyampaian hasil yang bermakna, bahkan bagi non-statistik. Teknik ini mendorong analisis yang lebih inklusif dengan mempertimbangkan keragaman kategori. Namun, penggunaannya memerlukan kehati-hatian dalam jumlah dummy, interpretasi, dan validasi model. Dengan teori yang kuat dan teknik yang tepat, dummy menjadi alat analisis yang andal dan relevan. Buku ini diharapkan menjadi referensi penting bagi mahasiswa, peneliti, dan praktisi dalam memahami dan menerapkan analisis regresi dengan variabel dummy secara efektif.

DAFTAR PUSTAKA

Antunes, A. R., Buga, C. B., Costa, D. C., Grilo, J., Braga, A. C., & Costa, L. A. (2021). A Multivariate Analysis Approach to Diamonds' Pricing Using Dummy Variables in SPSS. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12952 LNCS*, 609–623. https://doi.org/10.1007/978-3-030-86973-1_43

Arkes, J. (2023). *Regression analysis: a practical introduction*. Routledge.

Baum, D. (2021). Cloud Data Science for Dummies. In *John Wiley & Sons, Inc.* John Wiley & Sons. https://revistas.ufrj.br/index.php/rce/article/download/1659/1508%0Ahttp://hipatiapress.com/hpjournals/index.php/qre/article/view/1348%5Cnhttp://www.tandfonline.com/doi/abs/10.1080/09500799708666915%5Cnhttps://mckinseyonsociety.com/downloads/reports/Educa

Bruna, M. G., Đặng, R., Ammari, A., & Houanti, L. H. (2021). The effect of board gender diversity on corporate social performance: An instrumental variable quantile regression approach. *Finance Research Letters*, *40*, 101734. https://doi.org/10.1016/j.frl.2020.101734

Hardy, M. (2012). Regression with Dummy Variables. *Regression with Dummy Variables*, 128–132. https://doi.org/10.4135/9781412985628

Harris, J. (2020). Economic Policy Responses to the COVID-19 Pandemic. *Journal of Economic Perspectives*, *34*(4), 35–60.

Karlsen Kivedal, B. (2024). Regression Using Dummy Variables. In *Applied Statistics and Econometrics* (pp. 125–155). Springer. https://doi.org/10.1007/978-3-031-53142-2_7

Omokaro, B., & Ikpere, C. (2022). View of AN EVALUATION OF THE APPLICATION OF DUMMY VARIABLES IN REGRESSION ANALYSIS. *Innovative Journal of Science (ISSN: 2714-3309)*, *3*(3), 174–189. https://journals.rasetass.org/index.php/ijs/article/view/166/158

Singh, H., Das, A., Dey, S., Narsimhaiah, L., Pandit, P., Sinha, K., Sahu, P. K., & Mishra, P. (2023). A Study on Academic Attainment of Agriculture Students and its Correlates: A Dummy Regression Approach. *Annals of Data Science*, *10*(1), 129–152. https://doi.org/10.1007/s40745-020-00275-z

Skrivanek, S. (2009). The Use of Dummy Variables in Regression Analysis. *MoreSteam.Com LLC*, *6*, 131–142.

BAB 8
REGRESI NON-LINIER

Oleh Patrich Phill Edrich Papilaya

8.1 Pendahuluan

Analisis regresi merupakan salah satu teknik statistik yang paling fundamental dalam penelitian kuantitatif, memungkinkan peneliti untuk memahami hubungan antara variabel dependen dan satu atau lebih variabel independen (Montgomery et al., 2021). Sementara regresi linear telah menjadi landasan utama dalam analisis statistik selama bertahun-tahun, banyak fenomena di dunia nyata menunjukkan hubungan yang tidak dapat dijelaskan secara memadai melalui model linear sederhana.

Regresi non-linear muncul sebagai jawaban atas keterbatasan model linear dalam menangkap kompleksitas hubungan antar variabel yang seringkali bersifat kurvilinear, eksponensial, logaritmik, atau mengikuti pola matematika yang lebih kompleks (Seber & Wild, 2003). Berbeda dengan regresi linear yang mengasumsikan hubungan proporsional konstan antara variabel independen dan dependen, regresi non-linear mengakui bahwa perubahan dalam variabel independen dapat memiliki efek yang bervariasi terhadap variabel dependen tergantung pada nilai variabel tersebut.

Pentingnya regresi non-linear tidak dapat diremehkan dalam berbagai bidang aplikasi. Dalam biologi, pertumbuhan populasi seringkali mengikuti kurva logistik yang tidak dapat dimodelkan secara akurat dengan regresi linear (Verhulst, 1838). Dalam ekonomi, hubungan antara produksi dan input seringkali mengikuti hukum diminishing returns yang bersifat non-linear (Cobb & Douglas, 1928). Dalam farmakologi, hubungan dosis-respons seringkali mengikuti model sigmoidal yang memerlukan pendekatan non-linear untuk pemodelan yang akurat (Hill, 1910).

8.2. Konsep Dasar Regresi Non-Linear

8.2.1 Definisi dan Karakteristik

VRegresi non-linear dapat didefinisikan sebagai metode statistik yang digunakan untuk memodelkan hubungan antara variabel respons dan variabel prediktor melalui fungsi yang tidak linear dalam parameter (Ratkowsky, 1990). Secara matematis, model regresi non-linear dapat dinyatakan sebagai:

$$Y = f(X, \theta) + \varepsilon$$

dimana:
Y adalah variabel respons
X adalah vektor variabel prediktor
θ adalah vektor parameter yang akan diestimasi
f(X, θ) adalah fungsi non-linear
ε adalah error term yang diasumsikan berdistribusi normal

Karakteristik utama yang membedakan regresi non-linear dari regresi linear adalah sifat non-linearitas dalam parameter. Hal ini berarti bahwa turunan parsial dari fungsi *f* terhadap parameter θ mengandung parameter itu sendiri, sehingga tidak dapat diselesaikan dengan metode least squares biasa yang digunakan dalam regresi linear (Bates & Watts, 1988).

8.2.2 Perbedaan dengan Regresi Linear

Perbedaan fundamental antara regresi linear dan non-linear terletak pada beberapa aspek kunci.

1. Pertama, dalam hal interpretasi parameter, regresi linear memiliki interpretasi yang langsung dan intuitif dimana setiap koefisien merepresentasikan perubahan rata-rata dalam variabel respons untuk setiap unit perubahan dalam variabel prediktor. Sebaliknya, dalam regresi non-linear, interpretasi parameter seringkali lebih kompleks dan tergantung pada konteks spesifik model yang digunakan (Draper & Smith, 1998).
2. Kedua, dalam hal estimasi parameter, regresi linear menggunakan metode ordinary least squares (OLS) yang memiliki solusi analitik yang closed-form. Regresi non-linear, di sisi lain, memerlukan metode iteratif seperti Gauss-

Newton, Levenberg-Marquardt, atau algoritma optimasi lainnya untuk mencari estimasi parameter yang optimal (Nocedal & Wright, 2006).

3. Ketiga, sifat statistik dari estimator dalam regresi non-linear berbeda secara signifikan. Sementara estimator OLS dalam regresi linear bersifat unbiased dan memiliki varians minimum (BLUE - *Best Linear Unbiased Estimator*), estimator dalam regresi non-linear umumnya bersifat biased dalam sampel kecil meskipun konsisten secara asimptotik (Seber & Wild,2003).

8.3 Jenis-Jenis Model Regresi Non-Linear

8.3.1 Model Eksponensial

Model eksponensial merupakan salah satu bentuk regresi non-linear yang paling umum digunakan, terutama dalam modeling pertumbuhan atau peluruhan. Model ini dapat dinyatakan dalam berbagai bentuk, dengan yang paling sederhana adalah:

$$Y = \alpha * e^{(\beta X)} + \varepsilon$$

dimana α adalah parameter skala dan β adalah parameter laju pertumbuhan atau peluruhan. Model eksponensial sangat berguna dalam berbagai aplikasi, termasuk modeling pertumbuhan populasi, peluruhan radioaktif, dan pertumbuhan ekonomi (Malthus, 1798).
Variasi lain dari model eksponensial adalah model pertumbuhan eksponensial terbatas atau model asimptotik:

$$Y = \alpha * (1 - e^{(-\beta X)}) + \varepsilon$$

Model ini sangat berguna ketika variabel respons memiliki batas atas teoretis yang tidak dapat dilampaui, seperti dalam kasus pembelajaran atau adopsi teknologi (Rogers, 2003).

8.3.2 Model Logaritmik

Model logaritmik menunjukkan hubungan dimana variabel respons berubah secara logaritmik terhadap variabel prediktor. Bentuk umum model logaritmik adalah:

$$Y = \alpha + \beta * \ln(X) + \varepsilon$$

Model ini sangat berguna dalam ekonomi untuk modeling fungsi utilitas, dalam psikologi untuk modeling persepsi

sensori (hukum Weber-Fechner), dan dalam berbagai bidang lain dimana efek marginal menurun seiring dengan peningkatan input (Weber, 1834).

8.3.3 Model Polinomial

Meskipun secara teknis masih dapat dikategorikan sebagai regresi linear (karena linear dalam parameter), model polinomial seringkali dikelompokkan dalam regresi non-linear karena hubungan non-linear antara variabel prediktor dan respons:

$$Y = \alpha + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \ldots + \beta_n X^n + \varepsilon$$

Model polinomial berguna untuk menangkap hubungan yang kompleks dengan titik balik, seperti hubungan kuadratik yang memiliki maksimum atau minimum (Box & Draper, 2007).

8.3.4 Model Logistik

Model logistik atau kurva S merupakan salah satu model non-linear yang paling penting dan banyak digunakan. Model ini dinyatakan sebagai:

$$Y = \alpha / (1 + \beta * e^{(-\gamma X)}) + \varepsilon$$

dimana α adalah kapasitas carrying atau asimptot atas, β adalah parameter yang menentukan posisi kurva, dan γ adalah parameter laju pertumbuhan. Model logistik sangat berguna dalam modeling pertumbuhan populasi dengan batasan sumber daya, difusi inovasi, dan berbagai fenomena yang menunjukkan pola pertumbuhan yang dimulai lambat, kemudian cepat, dan akhirnya melambat ketika mendekati batas atas (Pearl & Reed, 1920).

8.4 Metode Estimasi Parameter

8.4.1 Metode Gauss-Newton

Metode Gauss-Newton merupakan salah satu algoritma iteratif yang paling populer untuk estimasi parameter dalam regresi non-linear. Algoritma ini merupakan modifikasi dari metode Newton untuk masalah least squares non-linear (Gauss, 1809).

Prinsip dasar metode Gauss-Newton adalah linearisasi fungsi non-linear di sekitar nilai parameter saat ini menggunakan ekspansi

Taylor orde pertama. Jika $\theta^{(k)}$ adalah estimasi parameter pada iterasi ke-k, maka estimasi pada iterasi selanjutnya dihitung sebagai:

$$\theta^{(k+1)} = \theta^{(k)} - (J^T J)^{-1} J^T r$$

dimana J adalah matriks Jacobian dari residual dan r adalah vektor residual.

Keunggulan metode Gauss-Newton adalah konvergensi yang relatif cepat ketika nilai awal parameter sudah dekat dengan solusi optimal. Namun, metode ini dapat mengalami masalah konvergensi jika matriks J^T J singular atau nearsingular (Björck,1996).

8.4.2 Metode Levenberg-Marquardt

Metode Levenberg-Marquardt merupakan kombinasi dari metode Gauss-Newton dan metode steepest descent, dirancang untuk mengatasi masalah konvergensi yang seringkali dihadapi oleh metode Gauss-Newton (Levenberg,1944; Marquardt, 1963).

Formula update parameter dalam metode Levenberg-Marquardt adalah:

$$\theta^{(k+1)} = \theta^{(k)} - (J^T J + \lambda I)^{-1} J^T r$$

dimana λ adalah parameter damping yang disesuaikan secara adaptif seVlama iterasi. Ketika λ kecil, metode ini berperilaku seperti Gauss-Newton, memberikan konvergensi yang cepat. Ketika λ besar, metode ini berperilaku seperti steepest descent, memberikan stabilitas yang lebih baik meskipun konvergensi lebih lambat.

8.4.3 Metode Maximum Likelihood

Dalam konteks tertentu, terutama ketika distribusi error diketahui, metode maximum likelihood dapat digunakan untuk estimasi parameter dalam regresi non-linear. Metode ini melibatkan maksimisasi fungsi likelihood:

$$L(\theta) = \prod_i f(y_i \mid x_i, \theta)$$

dimana f adalah fungsi densitas probabilitas dari distribusi error (McCullagh & Nelder, 1989).

Keunggulan metode maximum likelihood adalah sifat asimptotik yang baik dari estimator yang dihasilkan, termasuk konsistensi, efisiensi asimptotik, dan distribusi asimptotik yang normal. Namun, metode ini memerlukan spesifikasi yang benar dari distribusi error.

8.5 Evaluasi dan Validasi Model

8.5.1 Kriteria Goodness of Fit

Evaluasi kualitas model regresi non-linear melibatkan berbagai kriteria statistik. Coefficient of determination (R^2) yang disesuaikan untuk model non-linear dihitung sebagai:

$$R^2 = 1 - (SSE/SST)$$

dimana SSE adalah sum of squared errors dan SST adalah total sum of squares. Namun, interpretasi R^2 dalam konteks non-linear harus dilakukan dengan hati-hati karena sifat non-linear dari model (Kvålseth, 1985).

Kriteria lain yang penting adalah *Root Mean Square Error* (RMSE):

$$RMSE = \sqrt{(SSE/n)}$$

dimana n adalah jumlah observasi. RMSE memberikan ukuran rata-rata kesalahan prediksi dalam unit yang sama dengan variabel respons.

8.5.2 Kriteria Informasi

Akaike Information Criterion (AIC) dan Bayesian Information Criterion (BIC) seringkali digunakan untuk pemilihan model dalam konteks non-linear:

$$AIC = n * \ln(SSE/n) + 2p \quad BIC = n * \ln(SSE/n) + p * \ln(n)$$

dimana p adalah jumlah parameter dalam model (Akaike, 1974; Schwarz, 1978).

Kriteria ini mempertimbangkan *trade-off* antara *goodness of fit* dan kompleksitas model, dengan model yang memiliki nilai AIC atau BIC lebih rendah dianggap lebih baik.

8.5.3 Analisis Residual

Analisis residual dalam regresi non-linear mengikuti prinsip yang sama dengan regresi linear, tetapi dengan pertimbangan tambahan terkait sifat non-linear dari model. Residual standardized dihitung sebagai:

$$r_i = (y_i - \hat{y}_i) / \sqrt{(MSE * (1 - h_{ii}))}$$

dimana h_{ii} adalah elemen diagonal dari hat matrix yang dihitung berdasarkan matriks Jacobian.

8.6 Aplikasi dan Studi Kasus

8.6.1 Aplikasi dalam Biologi

Dalam biologi, model pertumbuhan logistik seringkali digunakan untuk modeling pertumbuhan populasi. Contoh klasik adalah pertumbuhan populasi bakteri dalam medium terbatas. Model logistik dapat dinyatakan sebagai:

$N(t) = K / (1 + ((K-N_0)/N_0) * e^{(-rt)})$

dimana

N(t) adalah ukuran populasi pada waktu t,

K adalah carrying capacity,

N_0 adalah ukuran populasi awal,

dan r adalah laju pertumbuhan intrinsik (Pearl & Reed, 1920).

8.6.2 Aplikasi dalam Ekonomi

Dalam ekonomi, fungsi produksi Cobb-Douglas merupakan contoh klasik model non-linear:

$$Y = A * L^{\alpha} * K^{\beta}$$

dimana Y adalah output,

L adalah tenaga kerja,

K adalah modal,

A adalah faktor produktivitas total, dan α serta β adalah elastisitas output terhadap tenaga kerja dan modal masing-masing (Cobb & Douglas, 1928).

8.6.3 Aplikasi dalam Farmakologi

Model Hill dalam farmakologi digunakan untuk menggambarkan hubungan dosis-respons:

$E = (E\_max * C^n) / (EC_{50}^n + C^n)$

dimana E adalah efek, E_max adalah efek maksimum, C adalah konsentrasi, EC_{50} adalah konsentrasi yang menghasilkan 50% efek maksimum, dan n adalah koefisien Hill (Hill, 1910).

8.7 Tantangan dan Limitasi

8.7.1 Masalah Konvergensi

Salah satu tantangan utama dalam regresi non-linear adalah masalah konvergensi algoritma iteratif. Konvergensi dapat gagal karena berbagai alasan, termasuk pemilihan nilai awal parameter

yang buruk, ill-conditioning dari matriks Hessian, atau spesifikasi model yang tidak tepat (Bates & Watts, 1988).

8.7.2 Sensitivitas terhadap Nilai Awal

Berbeda dengan regresi linear yang memiliki solusi unik, algoritma untuk regresi non-linear sangat sensitif terhadap pemilihan nilai awal parameter. Pemilihan nilai awal yang buruk dapat menyebabkan konvergensi ke local minimum daripada global minimum (Seber & Wild, 2003).

8.8 Perkembangan Terkini dan Masa Depan

8.8.1 Machine Learning dan Deep Learning

Perkembangan dalam machine learning dan deep learning telah membuka perspektif baru dalam modeling non-linear. Neural networks, khususnya deep neural networks, dapat dianggap sebagai bentuk ekstrem dari regresi non-linear dengan kemampuan untuk menangkap pola yang sangat kompleks (Goodfellow et al., 2016).

8.8.2 Regularisasi dalam Model Non-Linear

Teknik regularisasi seperti Ridge dan Lasso yang populer dalam regresi linear telah diadaptasi untuk konteks non-linear. Regularisasi membantu mengatasi masalah *overfitting* dan meningkatkan generalisasi model (Hastie et al., 2009).

8.8.3 Bayesian Non-Linear Regression

Pendekatan Bayesian untuk regresi non-linear memberikan *framework* yang principled untuk menangani ketidakpastian dalam estimasi parameter dan prediksi. *Markov Chain Monte Carlo* (MCMC) dan Variational Inference menjadi tools utama dalam implementasi Bayesian non-linear regression (Gelman et al., 2013).

DAFTAR PUSTAKA

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/tac.1974.1100705

Ake Bjorck. (1996). *Numerical Methods for Least Squares Problems*. SIAM.

Bates, D. M., & Watts, D. G. (2007). *Nonlinear Regression Analysis and Its Applications*. Wiley-Interscience.

Cobb, C. W., & Douglas, P. H. (1928). A Theory of Production. *The American Economic Review*, *18*(1), 139–165. https://www.jstor.org/stable/1811556

Cox, D. R. (2007). Response Surfaces, Mixtures, and Ridge Analyses, 2nd Edition by George E.P. Box, Norman R. Draper. *International Statistical Review*, *75*(2), 265–266. https://doi.org/10.1111/j.1751-5823.2007.00015_17.x

Draper, N. R., & Smith, H. (1998). Applied Regression Analysis. *Wiley Series in Probability and Statistics*, *3*. https://doi.org/10.1002/9781118625590

Gauss, C. F. (2025). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium : C. F. Gauss : Free Download, Borrow, and Streaming : Internet Archive*. Internet Archive. https://archive.org/details/bub_gb_ORUOAAAAQAAJ

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. Routledge & CRC Press. https://www.routledge.com/Bayesian-Data-Analysis/Gelman-Carlin-Stern-Dunson-Vehtari-Rubin/p/book/9781439840955

George, & Wild, C. J. (1989). *Nonlinear Regression*. John Wiley & Sons.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. Deeplearningbook.org; MIT Press. https://www.deeplearningbook.org/

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. In *Springer Series in Statistics*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Hill, A. V. (1910). *A. V. Hill, "The Possible Effects of The Aggregation of The Molecules of Haemoglobin on its Dissociation Curves," The*

Journal of Physiology, Vol. 40, 1910, pp. iv-vii. - References - Scientific Research Publishing. Scirp.org. https://www.scirp.org/reference/referencespapers?referenceid=415941

Kvalseth, T. O. (1985). Cautionary Note about R 2. *The American Statistician*, *39*(4), 279. https://doi.org/10.2307/2683704

LEVENBERG, K. (1944). A METHOD FOR THE SOLUTION OF CERTAIN NONLINEAR PROBLEMS IN LEAST SQUARES. *Quarterly of Applied Mathematics*, *2*(2), 164–168. JSTOR. https://doi.org/10.2307/43633451

Malthus, T. (1798). *An Essay on the Principle of Population* . Oll.libertyfund.org. https://oll.libertyfund.org/titles/malthus-an-essay-on-the-principle-of-population-1798-1st-ed

Marquardt, D. W. (1963). An Algorithm for LeastSquares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, *11*(2), 431–441. JSTOR. https://doi.org/10.2307/2098941

McCullagh, P., & Nelder, J. A. (1989). *McCullagh, P. and Nelder, J.A. (1989) Generalized Linear Models. 2nd Edition, Chapman and Hall, London. - References - Scientific Research Publishing*. Www.scirp.org. https://www.scirp.org/reference/referencespapers?referenceid=1851775

Montgomery, D. C., Elizabeth , A., & Vining , G. G. (2021). *Introduction to Linear Regression Analysis, 6th Edition | Wiley*. Wiley.com. https://www.wiley.com/en-us/Introduction+to+Linear+Regression+Analysis%2C+6th+Edition-p-9781119578727

Nocedal , J., & Wright, S. J. (2006). Numerical Optimization. In *Springer Series in Operations Research and Financial Engineering*. Springer New York. https://doi.org/10.1007/978-0-387-40065-5

Pearl, R., & Reed, L. (1920). Since 1790 and its Mathematical Representation On the Rate of Growth of the Population of the United States. *Proceedings of the National Academy of Sciences*, *6(6)*(275-288).

https://doi.org/10.1073/pnas.6.6.275
Ratkowsky, D. A. (1990). *Ratkowsky, D.A. (1990) Handbook of Nonlinear Regression Models. Marcel Dekker Inc., New York. - References - Scientific Research Publishing*. Scirp.org. https://www.scirp.org/reference/referencespapers?referenceid=2428575
Rothenberg, T. J. (1971). Identification in Parametric Models. *Econometrica*, *39*(3), 577–577. https://doi.org/10.2307/1913267
Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.
Turner, R. J. (2007). Diffusion of Innovations. *Journal of Minimally Invasive Gynecology*, *14*(6), 776. https://doi.org/10.1016/j.jmig.2007.07.001
Verhulst, P. F. (1838). *Verhulst, P.F. (1838) Notice sur la loi que la population suit dans son accroissement. Correspondence Mathematique et Physique (Ghent), 10, 113-121. - References - Scientific Research Publishing*. Scirp.org. https://www.scirp.org/reference/referencespapers?referenceid=2399035
Weber, E. H. (1834). *De pulsu, resorptione, auditu et tactu : annotationes anatomicae et physiologicae | WorldCat.org*. Worldcat.org. https://search.worldcat.org/title/de-pulsu-resorptione-auditu-et-tactu-annotationes-anatomicae-et-physiologicae/oclc/20047057

BIODATA PENULIS

Joni Wilson Sitopu
Dosen Universitas Simalungun

Joni Wilson Sitopu. Telah Menyelesaikan Studi Program Doktor Ilmu Matematika, FMIPA USU Medan. Sebelumnya mengikuti Pendidikan Program S1 Matematika, FMIPA USU dan S2 di Program Studi Pendidikan Matematika UNIMED. Ia adalah dosen tetap Program Studi Teknik Sipil, Fakultas Teknik, Universitas Simalungun. Mengampu mata kuliah Matematika, Statitika, Analisis Numerik, Metode Penelitian, Statistik Ekonomi dan Matematika Ekonomi. Selama ini terlibat aktif sebagai dosen pembimbing dan Penguji mahasiswa. Telah menulis 42 Buku referensi dan ratusan artikel penelitian dan pkm baik terindeks google schoolar, Garuda, Sinta dan Scopus. email: jwsitopu@gmail.com

BIODATA PENULIS

Dodi, S.Pd., M.Pd.
Dosen Program Studi Ilmu Kelautan
Fakultas Ilmu Pengetahuan Alam dan Kelautan
Universitas OSO

Penulis buku ini adalah Dodi, yang lahir pada tanggal 20 Agustus 1990 di sebuah daerah pesisir yang kaya akan budaya dan kearifan lokal, yaitu Padang Tikar, Kalimantan Barat. Lingkungan tempat tinggal yang menyatu dengan alam serta kehidupan masyarakat yang bergantung pada laut telah memberi warna tersendiri dalam cara pandangnya terhadap ilmu pengetahuan, terutama pada bidang kelautan dan pendidikan.

Sejak kecil, Dodi dikenal sebagai anak yang gemar membaca dan memiliki rasa ingin tahu yang tinggi terhadap fenomena-fenomena alam, angka, serta pola-pola logika. Hal inilah yang kemudian menumbuhkan kecintaan mendalam terhadap ilmu Matematika dan Sains. Ketertarikannya pada dunia pendidikan membawanya untuk menempuh studi di Program Studi Pendidikan Matematika, Universitas Tanjungpura (UNTAN), di mana ia meraih gelar sarjana (S1) dengan prestasi yang memuaskan. Tidak berhenti di situ, ia melanjutkan ke jenjang magister (S2) pada program studi yang sama di UNTAN, dengan fokus pada pengembangan pembelajaran matematika yang kontekstual dan aplikatif.

Saat ini, Dodi Andri berkarier sebagai dosen tetap di Program Studi Ilmu Kelautan, Fakultas Ilmu Pengetahuan Alam dan Kelautan,

Universitas OSO. Di lingkungan kampus, ia dikenal sebagai pendidik yang tidak hanya menguasai teori, tetapi juga senantiasa berusaha membumikan konsep-konsep abstrak dalam matematika dan sains agar dapat lebih mudah dipahami oleh mahasiswa. Ia kerap mengaitkan teori-teori ilmiah dengan fenomena kelautan dan kehidupan sehari-hari, menjadikannya sebagai pendekatan pembelajaran yang lebih hidup dan relevan.

Di luar aktivitas mengajar, Dodi juga aktif menulis karya ilmiah, artikel populer, serta buku-buku yang mengangkat keterkaitan antara matematika, sains, dan pendidikan. Fokus penulisannya berada pada upaya untuk menjembatani kesenjangan antara dunia akademik dan masyarakat umum, agar ilmu pengetahuan tidak hanya berhenti di ruang kelas atau jurnal, tetapi dapat menjadi bagian dari budaya belajar masyarakat luas. Dengan pendekatan bahasa yang komunikatif dan kontekstual, Dodi berharap setiap karyanya bisa menjadi sarana edukasi yang menyenangkan, inspiratif, dan membangkitkan semangat belajar.

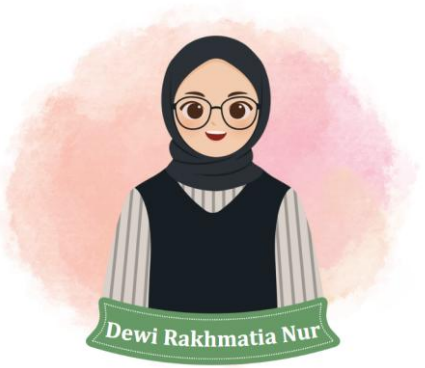
Dodi percaya bahwa matematika dan sains bukanlah sekadar mata pelajaran atau kumpulan rumus semata, tetapi merupakan cara berpikir dan kerangka logika yang dapat diterapkan dalam berbagai aspek kehidupan. Oleh karena itu, melalui buku ini dan karya-karya lainnya, ia ingin menularkan semangat belajar serta menginspirasi generasi muda, guru, dan siapa pun yang memiliki ketertarikan terhadap ilmu pengetahuan.

Ia juga terbuka untuk berdiskusi dan berbagi pengalaman dengan sesama akademisi, pelajar, maupun masyarakat umum. Bagi pembaca yang ingin berkomunikasi lebih lanjut, penulis dapat dihubungi melalui:

a. Email pribadi: dodi_a1@yahoo.com
b. Email institusi: dodi@oso.ac.id
c. Nomor HP/WhatsApp: 0896-9415-6039

Semoga karya ini dapat menjadi salah satu kontribusi kecil dalam membangun budaya literasi dan sains di Indonesia, khususnya di bidang matematika.

BIODATA PENULIS



Dosen Program Studi Matematika
Institut Teknologi dan Sains Nahdlatul Ulama Lampung

Penulis dilahirkan di Bandar Lampung pada 4 Agustus 1994. Penulis adalah dosen tetap pada Program Studi Matematika. Menyalesaikan studi S1 Jurusan Matematika di Universitas Sriwijaya dan S2 Jurusan Matematika di Universitas Lampung. Penulis dapat dihubungi melalui e-mail: dewirakhmatian@gmail.com

BIODATA PENULIS

Tedi Hartoyo, Ir. MSc.
Dosen tetap di Program Studi Agribisnis Fakultas Pertanian
Universitas Siliwangi

Selain sebagai Dosen tetap di Program Studi Agribisnis Fakultas Pertanian Universitas Siliwangi. Penulis mengajar Statistika di Jurusan Kimia Analis BTH Tasikmalaya , Keperawatan Respati Tasikmalayadan di keperawatan Gigi Tasikmalaya. Penulis lahir di Tasikmalaya tanggal 26 Juli 1962. Penulis adalah Dosen Program Studi Agribisnis Fakultas Pertanian Universitas Siliwangi. Menyelesaikan pendidikan S1 pada Jurusan Statistika IPB Bogor dan melanjutkan S2 pada Agriculture Development, Gent Unversity Belgium. Saat ini penulis mengampu beberapa mata kuliah, diantaranya Matematika, Statistika, Ekonometrika, Metode Penelitian Sosial Ekonomi dan Perancangan Percobaan.

BIODATA PENULIS

Ade Astuti Widi Rahayu

Ade Astuti Widi Rahayu, lahir di Jakarta, 26 Juni 1984. Menyelesaikan pendidikan Program Sarjana Teknik Industri di Fakultas Teknik Universitas Widya Mandala Madiun, dan melanjutkan Program Pascasarjana Magister Teknik Industri (S-2) di Fakultas Teknologi Industri Universitas Mercu Buana, Jakarta. Saat ini, penulis sebagai pengajar di Program Studi Teknik Industri, Fakultas Teknik, Universitas Buana Perjuangan Karawang (2017 – Sekarang). Penulis saat ini aktif sebagai pengelola Jurnal Ilmiah "***Industry Xplore***" Teknik Industri Universitas Buana Perjuangan Karawang dan juga aktif menulis di berbagai jurnal ilmiah. Penulis dapat dihubungi melalui email: ade.widiastuti@ubpkarawang.ac.id

BIODATA PENULIS

Ir. Hamdan Gani, S.Kom., MT., Ph.D
Dosen Program Studi Otomasi Sistem Permesinan
Politeknik ATI Makassar

Penulis lahir di Ujung Pandang tanggal 23 November 1988. Penulis adalah dosen tetap pada Program Studi Program Studi Otomasi Sistem Permesinan Politeknik ATI Makassar. Menyelesaikan pendidikan S1 pada Jurusan Teknik Informatika UIN Alauddin Makassar, S2 pada Jurusan Teknik Elektro Universitas Hasanuddin dan Melanjutkan S3 di Kyushu University Jepang dan terakhir melanjutkan pendidikan profesi Insinyur pada Universitas Muslim Indonesia (UMI).

BIODATA PENULIS

Dr.Ir. Patrich Phill Edrich Papilaya. M.Sc.Forest.Trop.
Dosen Program Studi Pascasarjana Manajemen Hutan
Universitas Pattimura

Penulis lahir di Ambon Tanggal 14 Pebruari 1967. Penulis adalah dosen tetap pada Program Studi Manajemen Hutan Pascasarjana. Universitas Pattimura. Ambon. Menyelesaikan pendidikan S1 pada Jurusan Kehutanan Fakultas Pertanian Universitas Pattimura. menyelesaikan Pendidikan S2 pada Georg-August Universitas. Faculty of Forestry. Göttingen. dalam bidang Remote sensing dan GIS. Pendidikan S3 diselesaikan pada IPB University. Bogor