Rizky Kusumawardhani, S.Si., M.Si
Zulfanita Dien Rizqiena, S.Si., M.Si
Septin Puji Astuti, S.Si., M.T., Ph.D

EKONOMERTIKA SUATU PENGANTAR

EKONOMERTIKA SUATU PENGANTAR

Rizky Kusumawardhani, S.Si., M.Si

Zulfanita Dien Rizqiena, S.Si., M.Si

Septin Puji Astuti, S.Si., M.T., Ph.D

EKONOMERTIKA SUATU PENGANTAR/Rizky Kusumawardhani,
Zulfanita Dien Rizqiena, Septin Puji Astuti
—cet.1.—Yogyakarta: Gerbang Media, 2021 viii + 126 hal.
Uk 15,5 x 23 cm

ISBN: 978-623-6666-48-7

Cetakan 1 Oktober 2021

Penerbit:
CV Gerbang Media Aksara (Anggota IKAPI)

Alamat. Jl sampangan No 58A, Rt 01 Banguntapan, Bantul,
Yogyakarta Telp. (0274) 4353651

PENGANTAR

BUKU ekonometrika ini disusun untuk memberikan penjelasan mengenai Mata Kuliah Ekonometrika yang sebagian besar diambil oleh mahasiswa dari Program Studi-Program Studi Ekonomi: Manajemen, Manajemen Syariah, Akuntansi, Akuntansi Syariah, Ekonomi Pembangunan, Ekonomi Islam, dan Perbankan. Buku ini adalah buku pengantar yang hanya menjabarkan dasar-dasar ekonometrika.

Buku ini terdiri dari tujuh bab. Ketujuh Bab tersebut dimulai dengan Bab 1 yang berisi Pendahuluan. Di bab ini menjelaskan apa ekonometrika, asal mula Ekonometrika, penerapannya di dunia nyata. Di Bab 2 analisis regresi yang diawali dengan memberi pendahuluan apa itu analisis regresi. Kemudian dilanjutkan dengan pembahasan analisis korelasi dan terminology-terminologi yang digunakan dalam analisis regresi, hubungan linear antar variabel dan bagaimana pengujian hipotesisnya. Rumus-rumus dan contoh soal disampaikan dalam buku ini beserta bagaimana menyelesaikan permasalahan-permasalahan ekonomi dengan ekonometrika.

Di Bab 3 dijelaskan analisis regersi berganda. Asumsi yang melandasi metode estimasi data, dan bagaimana melakukan estimasinya dijelaskan pada bab ini. Selanjutnya di Bab 4 dijelaskan mengenai regresi dummy. Asumsi-asumsi yang mendasari analisis regresi semuanya dijelaskan pada Bab 5. Dilanjutkan di Bab 6 yang membahas diagnostik model. Terakhir, di Bab 7 membahasa implementasi analisis regresi linear, baik yang sederhana maupun berganda.

Buku yang kami susun ini masih jauh dari sempurna. Oleh karena itu semua saran mohon disampaikan kepada kami para penulis. Akhirul kata, semoga buku ini bisa bermanfaat dan memberi berkah untuk semua pembacanya.

Tim Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

BAB 1.
PENDAHULUAN

1.1. Pengantar

Penggunaan istilah *econometrics* atau **ekonometrika** diawali oleh Powel Ciompa yang pertama kali menggunakan istilah ekonometrika dalam Bahasa Jerman yaitu "Oekonometrie" pada tahun 1910. Powel Ciompa adalah ahli ekonomi Polandia yang juga seorang professor di *Higher School of Economics di Cracow*. Dia juga direktur akuntansi di Bank Negara Federal Galicia di Lemberg. Powel Ciompa mendeskripsikan istilah ekonometrika sebagai berikut:

> *Seperti mekanika, akustik, dinamika, dan fenomena lain dalam fisika, dan fenomena masa pada geometri, juga fenomena ekonomi harus direpresentasikan dan ditampilkan mengikuti doktrin, yang saya bayangkan sebagai economografis. Ekonomografis akan membentuk suatu ekonomi deskriptif, yang akan berdasarkan pada ilmu ekonomi, matematika, dan geometri. Tugas utama seperti doktrin secara geometris akan mewakili nilai. Bagian dari ekonomografis ini dinamakan ekonometrik. Penerapan parktis dari ekonometrika ke bentuk matematika dari suatu nilai dan perubahannya adalah akuntansi. Ekonometrik kemudian hanya menjadi teori dalam akuntansi* (Israel, 2018).

Menurutnya, ekonometrika adalah bagian dari ekonomografis dan hubungan antara akuntasi dengan ekonometrika itu sangat erat sebagaimana hubungan antara matematika dan aljabar.

Pada tahun 1926 Ragnar Frisch menggunakan istilah ekonometrika dalam Bahasa Perancis yaitu "économétrie". Ragnar Frisch adalah ahli ekonomi dari Norwegia yang sekaligus penerima nobel pertama bidang ekonomi bersama Dutchman Jan Tinbargen pada tahun 1969. Frisch mendefinisikan économétrie sebagai;

> *Penghubung antara matematika, statistic, dan ekonomi politik, kami menemukan disiplin baru yang kami sebut dengan ekonometrika (econometrics). Tujuan dari ekonometrika adalah untuk menundukkan hukum abstrak ekonomi politik teoritis atau ekonomi murni untuk verifikasi eksperimental atau numerik, dan untuk mengubah ekonomi murni menjadi suatu ilmu dalam arti kata yang ketat* (Israel, 2018).

Di bidang ekonometrika, Ragnar Frisch lebih dikenal dibandingkan Pawel Ciompa.

Pada tahun 1970an, ekonometrika mengalami reformasi, seperti yang dibahas dalam buku *A History of Econometrics: The Reformation from the 1970s* (Ruggins, 2016). Pada saat itu mulai muncul pendekatan *Cowle's Commission*.

Para ahli ekonomi yang lebih banyak menggunakan analisis-analisis ekonomi secara kuantitatif harus menguasai ekonometrika. Ekonometrika dikenal sebagai suatu ilmu yang merupakan gabungan dari ilmu ekonomi, statistika, dan matematika. Dalam banyak masalah ekonomi, ekonometrika telah berperan dalam penyelesaiannya. Analisisnya digunakan sebagai dasar kebijakan ekonomi dan untuk meramalkan kondisi ekonomi suatu negara.

Pada awalnya, ekonometrika ini hanya digunakan untuk melakukan analisis makro ekonomi yang sering pula dikenal dengan nama makroekonometrika. Contoh-contoh dari makro ekonomi adalah seperti mengetahui apa yang menjadi penyebab pengangguran; bagaimana hubungan antara pengangguran, ketenaga kerjaan dengan pertumbuhan ekonomi; bagaimana

dampak inflasi terhadap pertumbuhan ekonomi; dan masih banyak lagi lainnya. Namun, pada akhirnya, ekonometrika diterapkan juga untuk analisis-analisis ekonomi mikro atau dikenal juga dengan istulah mikroekonometrika. Contoh dari mikroekometrika diterapkan pada penelitian perilaku konsumsi masyarakat, masalah produksi dan distribusi.

Verbeek (2017) berpendapat bahwa model ekonometrik itu dapat dikelompokkan ke dalam beberapa kategori. Dia menamakannya dengan model tipe pertama, kedua, ketiga dan keempat. **Model tipe pertama** adalah membandingkan masa lalu dan sekarang. Model ini dikenal dengan model *time series* yang seringnya digunakan untuk meramalkan adanya perubahan. **Model tipe kedua** membuat hubungan antara kuantitas ekonomi pada periode tertentu. Pada model ini digunakan untuk memberi informasi bagaimana hubungan suatu variabel ekonomi dengan variabel ekonomi lainnya. **Model tipe ketiga** adalah model yang menggambarkan hubungan antar variabel yang diukur pada waktu tertentu untuk suatu unit yang berbeda. **Model tipe keempat** adalah model yang membuat hubungan antar variabel yang berbeda yang diukur pada unit yang berbeda pada jangka waktu yang lebih panjang (paling sedikit dua periode). Model ini menggunakan data panel.

Pada bab ini akan dibahas sejarah ekonometrika. Selanjutnya dibahas berbagai definisi ekonometrika dari berbagai pakar ekonomi dan ekonometrika. Berbagai penerapan ekonometrika juga akan dijabarkan pada bab ini.

1.2. Apa Ekonometrika

Istilah ekonometrika diadopsi dari istilah bahasa Inggris yaitu *econometrics*. Kata ini gabungan dari dua kata yaitu *economics* yang artinya adalah ilmu ekonomi dan *metrics* yang berarti pengukuran. Jadi, ekonometrika dalam arti sempit adalah pengukuran ekonomi. Namun, sampai sekarang, banyak ahli yang

telah mendefinisikan ekonometrika. Berikut adalah beberapa definisi ekonometrika dari beberapa ahli

Menurut Pawel Ciompa (1910)
Ekonometrika adalah sekumpulan alat yang digunakan untuk menyampaikan informasi dalam bidang akuntansi (Israel, 2018).

Menurut Ragnas Frisch (1926)
Ekonometrika adalah pendekatan kuantitatif untuk membuat teori ekonomi sebagai suatu unifikasi dari teori ekonomi, matematik, dan statistika (Israel, 2018).

Menurut T. Haavelmo (1944)
Tujuan dari metode penelitian ekonometrika adalah menghubungkan teori ekonomi dan pengukuran aktual menggunakan teori-teori dan teknik-teknik statistika inferensia sebagai jembatan (Gujarati, 2003).

Menurut P. A. Samuelson, T.C. Koopmans, J.R.N. Stone (1954)
Ekonometrika didefinisikan sebagai analisis dari fenomena ekonomi actual yang didasarkan pada pengembangan teori dan pengamatan secara bersamaan, terkait dengan metode inferensia yang tepat (Gujarati, 2003).

Menurut Arthur S. Goldberger (1964)
Ekonometrika didefinisikan sebagai ilmu social yang menggunakan alat-alat teori ekonomi, matematika, dan statistika inferensia untuk menganalisis fenomena ekonomi (Gujarati, 2003).

Menurut Gerhard Tintner (1968)
Ekonometrika adalah hasil dari pandangan tertentu pada ekonomi, terdiri dari penerapan matematika statistika pada data ekonomi untuk memberi memberi bukti empiris untuk mendukung model terstruktur dari matematika ekonomi dan untuk mendapatkan hasil numerik (Gujarati, 2003).

Menurut H. Theil (1971)
Ekonometrika berhubungan dengan penentuan empiris dari hukum-hukum ekonomi (Gujarati, 2003)

Menurut E. Malinvaud (1966)
Seni dari para pakar ekonometrika terdiri dari pencarian asumsi bahwa kecukupan spesifik dan kecukupan realistik memboleh-kan dia mendapatkan keuntungan yang mungkin dari data yang ada padanya (Gujarati, 2003).

Menurut Adrian C. Darnell, J. Lynne Evans (1990)
Ahli ekonomi membantu dengan positif dalam menghilangkan kurangnya citra umum dari ekonometrika (kualitatif) sebagai kotak kosong yang dibuka dengan mengasumsikan bahwa ke-beradaan pembuka kaleng untuk mengetahui isinya dimana 10 ahli ekonomi akan menginterpretasikan dalam 11 cara (Gujarati, 2003).

Menurut Verbeek (2017)
Ekonometrika bertugas untuk mengkuantifikasi hubungan antar berbagai kuantitas berdasarkan data dan menggunakan tekni-k-teknik statistika dan untuk menginterpretasikan menggunakan atau memanfaatkan hasil outcome secara tepat.

Dari berbagai definisi tersebut, ekonomi, matematika, dan statistika adalah tiga bidang keilmuan yang dimanfaatkan dalam analisis ekonometrika. Hukum-hukum ekonomi dalam analisis ekonometrika digunakan sebagai hipotesis. Sementara, mate-matika dan statistika digunakan dalam mengolah data-data eko-nomi berdasarkan dari pernyataan ekonomi.

1.3. Metode dalam Ekonometrika

Gujarati (2003) mengusulkan metodologi ekonometrika ada delapan tahapan. Tahapan tersebut adalah membuat **per-nyataan atas teori**, menentukan **model matematika** dari teori,

menentukan **model statistika, mendapatkan data-data**, melakukan **estimasi parameter** dari model statistika atau model ekonometrik, **menguji hipotesis**, **melakukan peramalan** atau **prediksi** berdasarkan analisis model ekonometrik, dan terakhir adalah menggunakan model untuk **tujuan pengendalian** atau **kebijakan**.

Sementara itu, Walters dan Johnston (1965) memberi panduan dalam melakukan penelitian dengan menggunakan ekonometrika. Ada 12 tahap yang dia usulkan yaitu membuat hipotesis, menggambarkan teori dengan menggunakan diagram hubungan, menyiapkan data, melakukan estimasi dan interpretasi hasil, menganalisis varians, uji multikolinearitas, variabel batasan dan dummy, heteroskedastisitas, autokorelasi, stasioneritas, kausalitas dan kointegrasi, dan essay berdasarkan penelitian tersebut. Tahapan ini cenderung untuk melakukan tahapan penelitian yang memanfaatkan analisis regresi untuk analisis datanya.

Dari beberapa pendapat tersebut, tahapan ekonometrika dapat diringkas menjadi tahapan berikut:

Tahap pertama adalah membuat **pernyataan atas teori** atau dikenal dengan **hipotesis**. Teori ekonomi itu bisa teori makroekonomi maupun teori mikroekonomi. Teori makroekonomi seperti teori pertumbuhan, produktifitas, model fluktuasi, penentu konsumsi, investasi, ketenaga kerjaan, inflansi, tingkat bunga, keseimbangan perdagangan dan lain sebagianya. Sementara itu teori mikroekonomi seperti permintaan komoditas, biaya produksi, harga pasar, laba, pendapatan industry, tingkat upah, investasi langsung luar negeri, struktur pasar, dan lain sebagainya. Selain itu, juga ada teori perdagangan dan keuangan public.

Tahap kedua adalah menentukan **model matematika** dari teori tersebut. Dari teori-teori ekonomi tersebut dibuat model matematika. Dalam membuat model matematika dibuat diagram yang menghubungkan antar variabel.

Tahap ketiga adalah menentukan **model statistika**. Setelah dibuat model matematika maka ditentukan analisis statistika

yang tepat untuk mendapatkan kesimpulan yang tepat. Pada umummya, analisis ekonometrika menggunakan analisis regresi dan analisis *time series*. Analisis regresi juga bermacam-macam. Ada analisis regresi linear, analisis regresi non-linear, analisis regresi menggunakan variabel dummy, analisis regresi logistik dan lain sebagainya. Sementara itu, untuk data *time series* menggunakan data

Tahap keempat adalah **mendapatkan data-data**. Dalam mendapatkan data bisa diperoleh dengan data primer maupun data sekunder. Data primer diperoleh melalui survey atau pengambilan data langsung. Di era sekarang, survey bisa dilakukan secara online. Sehingga cakupan survey bisa lebih luas. Sementara, data sekunder diperoleh dari pihak kedua yang bisa jadi juga data survey maupun data dokumentasi selama bertahun-tahun. Data-data banyak yang tersedia secara online. Di Indonesia, ada Badan Pusat Statistik yang menyediakan data-data ekonomi, data dari Bank Indonesia, dan dari sumber lainnya.

Tahap kelima adalah melakukan **estimasi parameter** dari model statistika atau model ekonometrik dan interpretasi hasil. Estimasi atau menaksir parameter dari model yang telah ditentukan pada tahap ketiga. Dalam estimasi parameter ini sangat bergantung pada analisis statistik yang akan digunakan dan juga karakteristik data dan model yang akan dibangun.

Tahap keenam adalah **pemeriksaan model** untuk **memilih model terbaik**. Setelah dilakukan estimasi parameter, dalam analisis statistika perlu dilakukan pemeriksaan model apakah sudah memenuhi semua asumsi yang dipersyaratkan. Untuk itu perlu dilakukan pemeriksaan model seperti uji kecukupan model dengan melihat R^2, F atau X^2 . Juga uji multikolinearitas, uji heteroskedastisitas, autokolerasi, stasioneritas, dan lain-lain. Dari model tersebut akan diperoleh model terbaik yang mampu menggambarkan data.

Tahap ketujuh adalah **menguji hipotesis**. Pada tahap ini adalah pembuktian apakah dugaan kita di tahap awal adalah

benar. Hipotesis ini diambil berdasarkan hasil analisis data statistika. Namun, hasil analisis statistika hanyalah petunjuk kuantitatif, yang tetap harus dilakukan penyelidikan lebih dalam apakah hasil dari analisis ini sudah memenuhi teori atau memang memungkinkan akan memunculkan teori baru.

Tahap kedelapan adalah **melakukan peramalan** atau **prediksi** berdasarkan analisis model ekonometrik. Prediksi dilakukan dari model yang terbaik yang dapat mewakili data sesuai dari hasil tahap keenam.

Tahap kesembilan, menggunakan model untuk **tujuan pengendalian** atau **kebijakan**. Hasil analisis ini kemudian dijadikan pedoman untuk melakukan kebijakan apa.

1.4. Data yang digunakan dalam Ekonometrika

Dalam analisis ekonometrika, tentu saja, data-data ekonomi yang digunakan. Data tersebut bisa bersumber dari data primer atau data sekunder. Data primer di sini adalah data yang diperoleh dari sumbernya langsung. Ini bisa dilakukan dengan melakukan pengukuran atau survey yang dilakukan sendiri oleh peneliti. Data sekunder diambil melalui pihak ketiga. Data ini juga bisa dilakukan melalui pengukuran atau survey. Namun, data survey tersebut diperoleh dari pihak ketiga seperti lembaga negara, lembaga survey, atau perusahaan swasta yang melayani data.

Data yang akan diolah dalam ekonometrika ini bisa dalam berbagai bentuk. Namun, secara umum, dalam pengolahan data ada empat skala. Secara umum ada empat skala data yaitu skala data nominal, ordinal, interval, dan rasio. Astuti (2020) meringkas penjelasan jenis skala data tersebut dalam bentuk tabel seperti dalam Tabel 1.1.

Tabel 1.1 Ciri setiap skala data

	Dapat dibedakan	**Mempunyai urutan**	**Interval sama**	**Mempunyai nilai nol mutlak**
Nominal	√			
Ordinal	√	√		
Interval	√	√	√	
Rasio	√	√	√	√

Data nominal adalah data yang hanya dapat dibedakan. Contohnya adalah data nama kota, nama produk bank, jenis lembaga keuangan, kelompok, dan lain sebagainya. Nama kota seperti Jakarta, Bandung, Semarang, dan seterusnya itu sifatnya hanya membedakan saja. Sama halnya dengan jenis lembaga keuangan seperti lembaga keuangan seperti bank sentral dan bank umum dan lembaga keuangan non-bank seperti pegadaian, koperasi, dan perusahaan asuransi. Sementara itu, untuk **data ordinal** selain dapat dibedakan juga memiliki urutan yang bisa dari besar ke kecil atau kecil ke besar. Contoh dari skala data ini adalah kelompok usia tenaga kerja dan jenis inflasi. Usia tenaga kerja ada jenjang urutannya, sama juga dengan jenis inflasi seperti inflasi *creeping inflation*, *galloping inflation*, dan *hyper inflation*. Dua skala data ini dikategorikan sebagai data kategori atau data diskrit.

Jika data nominal dan ordinal adalah data diskrit, data interval dan rasio adalah data yang sifatnya kontinyu. Jika dilihat dengan menggunakan Tabel 1.1, **data interval** adalah data yang dapat dibedakan, memiliki urutan, memiliki interval yang sama, namun tidak memiliki nilai nol mutlak. Jika memiliki nilai nol, maka data tersebut tidak berarti benar-benar nol, atau kosong. Sementara itu, **data rasio** adalah data yang cirinya lengkap, data ini selain bisa dibedakan, memiliki urutan, dan memiliki interval yang sama, data rasio memiliki nol mutlak atau nilai nol yang sebenarnya. Contohnya jumlah uang. Jika jumlah uang bernilai 0 maka berarti tidak memiliki uang. Contoh lainnya adalah per-

tumbuhan ekonomi yang bernilai nol. Dalam hal ini berarti memang tidak ada pertumbuhan ekonomi sama sekali.

Selain harus mengenal skala data, tipe data juga harus dipahami oleh ahli ekonometrika. Tipe data yang dikenal dalam ekonometrika ada tiga yaitu data *time series*, data *cross section*, dan *pooled*. Data ***time series*** adalah data yang tergantung pada waktu. Misalnya, data inflasi tiap bulan dari tahun 2000 hingga 2021, data pengangguran tiap bulan dari tahun 1980 hingga 2021, jumlah penduduk miskin per tahun dari tahun 1980 hingga 2021, dan seterusnya. Data-data tersebut bergantung pada waktu dan pada umumnya, ada korelasi antar periode waktu. Data ***cross section*** adalah sekumpulan variabel data yang diambil dalam satu waktu. Contohnya, data kemiskinan, data pengangguran, data pendapatan penduduk pada tahun 2021. Terakhir, data ***pooled***, adalah kombinasi dari data *time series* dan data *cross section*. Contohnya, data kemiskinan, pengangguran, pendapatan penduduk padaa tahun 2000 hingga 2021.

Pengenalan skala data dan tipe data ini sangat diperlukan sebelum melakukan pengolahan ekonometrika. Hal ini karena skala dan jenis data mempengaruhi jenis analisis statistik yang akan digunakan. Kesalahan dalam mengidentifikasi data akan mengarahkan pada kesalahan dalam melakukan analisis data. Tentu, hal ini akan menghasilkan output yang tidak dapat dipercaya.

1.5. Aplikasi Ekonometrika

Analisis ekonometrika diterapkan pada berbagai kasus ekonomi. Sudah banyak ahli yang menggunakan ekonometrika untuk dijadikan dasar dalam pengambilan kebijakan ekonomi. Beberapa di antaranya akan dibahas di bab ini.

Ekonometrika untuk analisis krisis moneter

Pada tahun 2000, Fukuchi (2000) melakukan analisis ekonometrika pada krisis moneter di Indonesia pada tahun 1996 hingga

1998. Dia menggunakan data *time series* bulanan untuk melihat dampak krisis keuangan pada indeks produksi dan tenaga kerja industri pada masa itu. Dengan menggunakan analisis tersebut diperoleh beberapa pelajaran berharga selama krisis moneter bisa dijadikan acuan untuk kebijakan selanjutnya. Sementara itu, Green (2004) menggunakan ekonometrika untuk melakukan analisis perilaku investasi selama krisis moneter. Haan et al. (2001) menggunakan data makroekonomi 1993-1999 dari Republik Makedonia untuk melakukan analisis atas keuangan atau guncangan fiscal, eksternal dan pasar tenaga kerja.

Ekonometrika untuk analisis pertumbuhan

Márquez et al. (2010) menggunakan ekonometrika membuat analisis ekonomi regional di Spanyol. Data yang digunakan adalah data *panel* pertumbuhan untuk melihat pengaruh spasial.

Glytsos (2002) melakukan analisis ekonometrika untuk melihat pengaruh remiten terhadap konsumsi, investasi, impor dan output di negara-negara Mediterania.

Ekonometrika untuk analisis kemiskinan

Kemiskinan sangat erat hubungannya dengan kemiskinan. Ekonometrika yang digunakan untuk analisis kemiskinan sudah banyak dilakukan oleh peneliti. Misalnya, Zheng (1994) membuat analisis ekonometrika untuk melihat apakah kemiskinan itu relatif atau absolut. Selanjutnya, Duque et al. (2015) menerapkan spatial ekonometrik untuk mengukur kemiskinan Medellin, Kolombia.

Ekonometrika untuk analisis konsumsi

Konsumsi masyarakat perlu dianalisis untuk menentukan kebutuhan. Kebutuhan pokok, seperti beras, garam, gula, dan lain sebagainya sangat diperlukan prediksi kebutuhan agar dapat dipersipkan jauh-jauh hari supaya persediaan di dalam negeri tidak mengalami kekurangan. Analisis ekonometrika sangat berguna di dalam hal ini. Selain kebutuhan pokok, kebutuhan

lain yang tidak pokok juga perlu diprediksi. Misalnya analisis kepemilikan kendaran seperti yang dilakukan oleh Dargay (2001). Kepemilikan kendaraan berkaitan dengan pendapatan masyarakat dan itu mempengaruhi penerimaan pajak kendaraan.

Ekonometrika untuk analisis kebutuhan energi

Konsumsi energi memang sangat erat hubungannya dengan Produk Domesti Bruto (PDB) suatu negara. Semakin tinggi PDB akan semakin tinggi konsumsi energinya. Akibatnya, kebutuhan energi di negara tersebut akan ikut meningkat pula. Oleh karenanya, diperlukan kebijakan untuk penyediaan energi agar masyarakat tidak kekurangan pasokan energi. Penerapan ekonometrika untuk analisis energi banyak dilakukan oleh peneliti. Larsen dan Nesbakken (2004) misalnya, melakukan analisis konsumsi listrik warga Norwegia. Data diperoleh melalui survey kepada pelanggan listrik. Arabatzis dan Malesios (2011) melakukan analisis penggunakan kayu bakar di pegunungan di Yunani Utara. Faktor-faktor apa yang membuat masyarakat membeli kayu bakar dia analisis untuk membuat kebijakan terkait analisis energi. Sementara Parajuli et al. (2014) juga melakukan analisis ekonomi energi di Nepal.

Analisis ekonometrik untuk ekonomi lingkungan

Ketersediaan sumber daya erat kaitannya dengan masalah ekonomi. Keterbatasan sumberdaya akan mempengaruhi ekonomi. Sementara, ekploitasi sumber daya yang berlebihan akan berdampak pada lingkungan. Oleh karenanya, kemudian muncul ekonomi lingkungan. Analisis ekonometrika sangat banyak dimanfaatkan untuk analisis-analisis ekonomi lingkungan seperti ini. Seperti misalnya Zwane (2007) yang melakukan analisis akan kemiskinan dan dampaknya terhadap penggundulan hutan di Peru. Li et al. (2017) membuat analisis kebijakan polusi udara dan keuntungannya secara ekonomi. Sementara (Wang et al., 2016) melakukan analisis spatio temporal untuk mengevaluasi emisi CO_2 di Cina pada tahun 1995-2012.

BAB 2.
ANALISIS REGRESI

2.1 Pendahuluan

Analisis regresi adalah salah satu analisis statistika yang cukup populer. Analisis ini banyak diterapkan di berbagai bidang, selain di dalam bidang ekonomi. Analisis regresi yang awalnya diterapkan dalam dunia kedokteran. Sampai sekarang banyak digunakan di dunia industri, ekonomi, dan bisnis.

Analisis regresi dikenalkan oleg Francis Galton yang pada saat itu melihat bahwa ada kecenderungan orangtua yang tinggi akan memiliki anak yang tinggi. Sebaliknya, jika orangtua pendek akan cenderung memiliki anak yang pendek. Dia mengungkapkan ini pada tulisannya berjudul *Family Likeness in Stature* yang dipublikasikan pada *Proceeing of the Royal Society* pada Januari 1886. Kemudian, Karl Pearson mengonfirmasi hasil penelitian Galton dan menuliskannya di sebuah artikel berjudul *On the Law of Inheritance* yang dipublikasikan oleh Karl Pearson da Alice Lee di Journal *Biometrica* Volume 2 Nomor 4 tahun 1903.

Analisis regresi berkaitan dengan ilmu yang mempelajari ketergantungan suatu variabel pada satu atau beberapa variabel lain untuk memperkirakan atau memprediksi rata-rata dari sampel. Analisis regresi menurut Iriawan dan Astuti (2006) dapat digunakan untuk tiga hal. Pertama, analisis regresi digunakan untuk mengukur kekuatan hubungan antar variabel respons dan prediktor. Kedua, analisis regresi juga digunakan untuk mengetahui pengaruh variabel prediktor terhadap variabel respons. Ketiga, analisis regresi digunakan untuk mempredikasi pengaruh

suatu variabel prediktor terhadap variabel respons. Namun, lagi-lagi perlu dicatat adalah bahwa pengaruh variabel independen terhadap variabel dependen itu harus didukung dengan teori yang kuat, tidak cukup hanya sekedar dibuat analisis statistika saja.

Dalam analisis regresi lebih fokus pada ketergantungan secara statistik antar variabel yang itu bukan hubungan fungsional atau deterministrik, namun variabel tersebut berhubungan dengan variabel acak (*random*) atau *stochastic* yang probabilistik.

2.2. Analisis Korelasi

Seringkali, analisis regresi dihubungkan dengan analisis korelasi. **Analisis korelasi** merupakan analisis untuk mencari hubungan antar dua variabel. Namun analisis regresi mampu mendeteksi pola hubungan suatu variabel dengan beberapa variabel lainnya. Baik analisis korelasi maupun analisis regresi tidak bisa menentukan kausalitas jika tanpa ada teori pendukungnya. Oleh karenanya, dalam ekonometrika, penentuan teori diletakkan di tahap awal sebelum melakukan analisis regresi.

Analisis korelasi digunakan untuk mendeteksi kekuatan hubungan linear antar dua variabel. Sebut saja kedua variabel tersebut adalah variabel x dan y. Korelasi dinotasikan dengan ρ. Persamaan untuk menghitung korelasi adalah sebagai berikut:

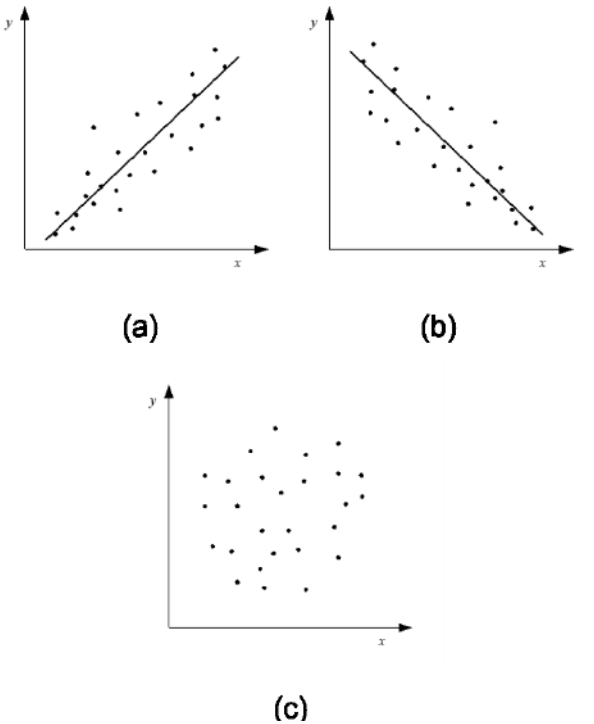
$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad (2.1)$$

dimana ρ_{xy} = korelasi *product moment Pearson*
x_i = data ke-i dari variabel x
y_i = data ke-i dari variabel y
$\bar{x}$ = rata-rata dari variabel x
$\bar{y}$ = rata-rata dari variabel y
n = banyaknya data

Nilai korelasi berada pada rentang $-1 < \rho < 1$. Jika korelasinya kuat, maka ρ akan bernilai mendekati -1 atau 1. Sebaliknya, jika nilai korelasi mendekati 0, maka dikatakan memiliki korelasi rendah. Tanda minus (-) atau plus menunjukkan arah hubungan. Apabila hubungannya positif, maka ρ akan bernilai di atas 0. Sementara, jika hubungannya negative, maka ρ bernilai di bawah 0 atau bernilai negatif.

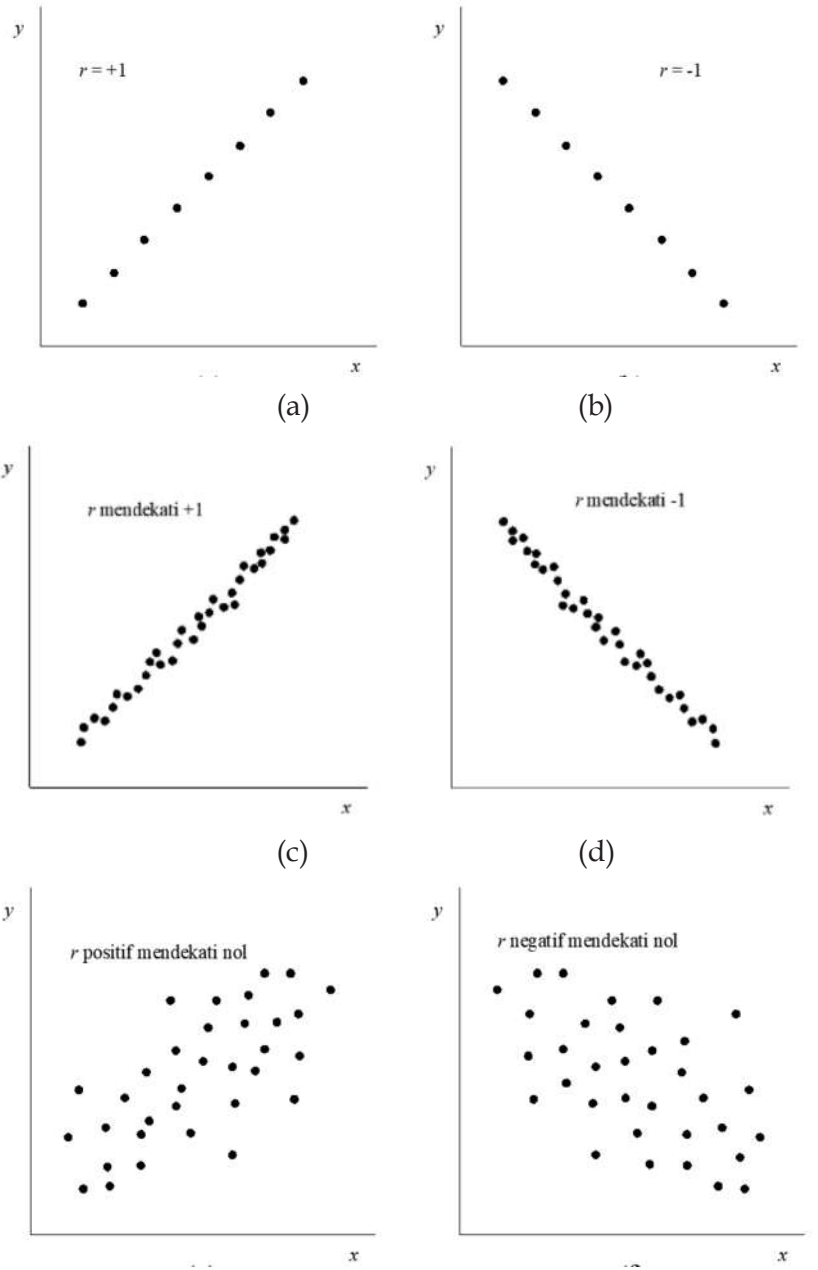
Dua variabel memiliki pola hubungan positif artinya, jika salah satu variabel dinaikkan maka variabel lainnya akan ikut naik. Begitu juga sebaliknya, jika variabel satu turun, maka akan cenderung membuat variabel lainnya turun. Sebagai contoh, jika pendapatan dinaikkan, maka akan menyebabkan kenaikan pada konsumsi. Berbeda dengan pola hubungan negatif. Jika suatu variabel dinaikkan, maka akan menyebabkan variabel lainnya naik. Contohnya adalah, jika harga (penawaran) dinaikkan, maka akan menurunkan permintaan.

Pola hubungan ini lebih detilnya dapat dilihat pada Gambar 2.1. Pada Gambar 2.1(a) adalah *scatterplot* dari dua variabel yang korelasinya positif. Sedangkan Gambar 2.2(b) adalah bentuk hubungan dua variabel yang korelasinya negatif. Jika dalam *scatterplot* tidak jelas arahnya, bisa dikatakan tidak ada hubungan antar variabel tersebut. Seperti yang terlihat pada Gambar 2.1(c).



Gambar 2.1. Pola hubungan antar dua variabel

Dalam praktiknya, data tidak selalu berada dalam garis persis. Gambar 2.2 memberikan ilustrasi bagaimana besaran nilai korelasi dan bagaimana bentuk datanya.



Gambar 2.2. Berbagai macam pola hubungan antar variabel dan sebaran datanya.

Pada gambar tersebut terlihat, semakin data mendekati garis, maka nilai korelasinya akan semakin mendekati nilai korelasi sempurna, yaitu 1 atau -1. Seperti yang dalam Gambar 2.2 (a) dan (b), data tepat berada dalam garis dan nilai korelasinya adalah sama dengan 1 atau -1. Sedangkan pada Gambar 2.2(c) dan (d), data menyebar di sekitar garis. Data semakin mengumpul dalam garis, maka korelasinya akan semakin mendekati korelasi sempurna. Sebaliknya, jika semakin menyebar, korelasinya akan semakin mendekati nilai nol atau tidak ada korelasi. Seperti yang terlihat pada Gambar 2.2(e) dan (f).

Sebagai ilustrasi, akan ditunjukkan contoh analisis korelasi pada ekonometrika. Data pada Tabel 2.1 adalah tingkat pengangguran dan jumlah penduduk miskin di tiap provinsi di Indonesia pada tahun 2019.

Tabel 2.1. Data ilustrasi 2.1

Provinsi	Tingkat pengangguran terbuka	Persentase penduduk miskin
Aceh	6,20	15,01
Sumatera Utara	5,41	8,63
Sumatera Barat	5,33	6,29
Riau	5,97	6,90
Jambi	4,19	7,51
Sumatera Selatan	4,48	12,56
Bengkulu	3,39	14,91
Lampung	4,03	12,30
Kep. Bangka Belitung	3,62	4,50
Kep. Riau	6,91	5,80
DKI Jakarta	6,22	3,42
Jawa Barat	7,99	6,82
Jawa Tengah	4,49	10,58
DI Yogyakarta	3,14	11,44
Jawa Timur	3,92	10,20
Banten	8,11	4,94
Bali	1,52	3,61
Nusa Tenggara Barat	3,42	13,88
Nusa Tenggara Timur	3,35	20,62
Kalimantan Barat	4,45	7,28
Kalimantan Tengah	4,10	4,81
Kalimantan Selatan	4,31	4,47
Kalimantan Timur	6,09	5,91
Kalimantan Utara	4,40	6,49
Sulawesi Utara	6,25	7,51
Sulawesi Tenggara	3,15	13,18
Sulawesi Selatan	4,97	8,56
Sulawesi Tenggara	3,59	11,04
Gorontalo	4,06	15,31
Sulawesi Barat	3,18	10,95
Maluku	7,08	17,65
Maluku Utara	4,97	6,91
Papua Barat	6,24	21,59
Papua	3,65	26,55

Dengan menggunakan rumus pada persamaan 2.1, korelasi dari antara tingkat pengangguran terpadu dan kemiskinan adalah sebesar -0,28. Hasil korelasi sebesar -0,28 menunjukkan bahwa tingkat penganguran terpadu dan kemiskinan memiliki hubungan negatif, artinya bahwa peningkatan pengangguran terpadu akan menurunkan kemiskinan. Nilai korelasi sebesar -0,28 ini juga mengindikasikan nilai korelasi yang lemah karena kurang dari 0,5. Korelasi yang lemah bukan berarti dua variabel tersebut hubungannya diabaikan. Bisa jadi, hubungan yang kuat terjadi antara variabel kemiskinan dengan variabel lainnya, akan tetapi kecil korelasinya dengan peningkatan pengangguran terpadu.

2.3 Terminology dan Persamaan dalam Analisis Regresi

Persamaan umum pada analisis regresi adalah seperti yang ditunjukkan pada Persamaan 2.2.

$$\boldsymbol{Y=XB} = \boldsymbol{\varepsilon} \qquad (2.2)$$

Dimana $\boldsymbol{Y}$ adalah matriks $\begin{bmatrix} Y_1 \\ Y_1 \\ \vdots \\ Y_1 \end{bmatrix}$, sedangkan $\boldsymbol{X}$ adalah matriks $\begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{k1} \\ 1 & X_{12} & X_{22} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{kn} \end{bmatrix}$, $\boldsymbol{B}$ adalah matriks $\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$, dan $\boldsymbol{\varepsilon}$ adalah matriks $\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$. Variabel $\boldsymbol{Y}$ adalah **variabel dependen** atau **variabel tidak bebas**. Dikatakan sebagai variabel tidak bebas karena variabel ini dipengaruhi oleh variabel bebas atau variabel independen. Variabel ini disebut juga sebagai **variabel respons** karena merespons dari variabel independen. Dalam Persamaan 2.2 tersebut, **variabel bebas** atau **variabel independen** adalah $\boldsymbol{X}$.

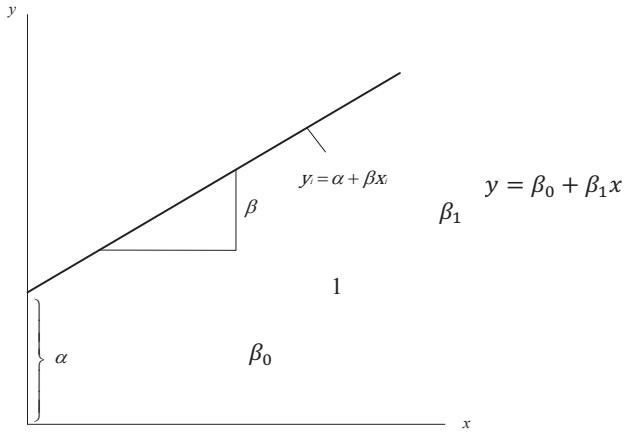
Variabel ini mempengaruhi variabel dependen, sehingga disebut juga sebagai **variabel prediktor**. Sementara ***B*** adalah parameter model yang nilainya diperoleh dengan menaksir model. Variabel ***ε*** adalah **variabel residual** atau **error**. Variabel ini terjadi karena kesalahan dalam melakukan pemodelan. Semakin besar kesalahannya berarti semakin besar kesalahan model tersebut mampu memprediksi kenyataan yang ada.

Contoh-contoh variabel dependen dan independen dalam analisis regresi yang digunakan pada ekonometrika adalah sebagai berikut: pendapatan personal sebagai variabel independen dan konsumsi personal sebagai variabel dependen. Pola hubungan antara permintaan dan harga, dimana permintaan sebagai variabel independen dan harga sebagai variabel dependen. Rata-rata upah dan tingkat pengangguran adalah dua variabel independent dan variabel dependen. Contoh lainnya, perubahan temperatur, hujan, sinar matahari, dan pupuk sebagai variabel independen dan hasil panen sebagai variabel independent.

Model regresi tersebut dapat berbentuk bivariat atau multivariat. Model regresi bivariat hanya terdiri dari dua variabel, yaitu satu variabel independen dan satu variabel dependen. Model regresi multivariat memiliki satu variabel dependen dan banyak variabel independent. Model regresi bivariat dikenal juga sebagai model regresi sederhana. Bentuknya ditunjukkan pada Persamaan 2.3.

$$y = \beta_0 + \beta_1 x + \varepsilon \qquad (2.3)$$

Dalam persamaan tersebut hanya ada satu variabel x. Sementara parameternya ada dua yaitu β_0 dan β_1. Dalam hal ini β_0 adalah konstanta atau *intercept*. Konstanta ini menentukan nilai awal dari variabel dependen pada saat nilai variabel independent bernilai nol. Sementara itu, β_1 adalah parameter dari variabel prediktor (x) atau *slope*. Gambaran dari model regresi linear sederhana ditunjukkan pada Gambar 2.3. Terlihat jelas bagaimana β_0 dan β_1 dalam diagram kartesius tersebut.

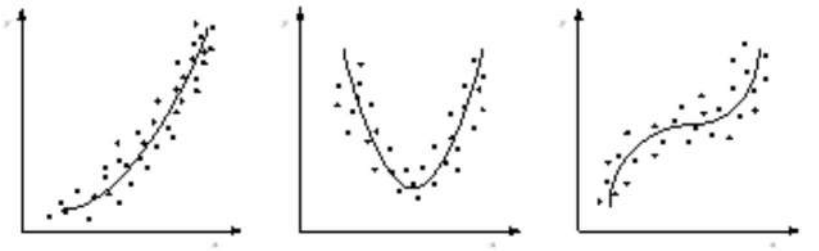


Gambar 2.3: Grafik regresi

2.4 Linearitas Hubungan antar Variabel

Hubungan antar variabel bisa bersifat linear atau tidak linear. Jika hubungan antar variabel tidak linear maka digunakan model linear. Sebaliknya, jika hubungannya tidak linear, maka perlu mencari model non-linear yang tepat untuk mencari hubungan antar variabel tersebut. Sebelum melakukan analisis regresi, perlu mengetahui pola hubungan antar data. Dalam buku ini, hanya akan dibahas mengenai hubungan linear. Detail mengenai hubungan linearitas dalam analisis regresi akan dibahas lebih mendalam di Bab 6.

Gambar 2.4 menunjukkan pola-pola hubungan variabel yang tidak linear. Gambar 2.4(b) sama dengan Gambar 2.2(h) yang polanya membentuk model kuadratik. Sementara itu model 2.4(a) dan 2.2(c) adalah pola eksponensial dan kubik. Persamaan untuk Gambar 2.4(a), (b), dan (c) adalah sebagai berikut:



(a) (b) (c)

Gambar 2.4. Contoh hubungan non-linear pada dua variabel

Model eksponensial,

$$y = \beta_0+\beta_1 x_1+\beta_2 x_1^2+\varepsilon \quad (2.4)$$

Model kuadratik,

$$y = e^{\beta_0+\beta_1 x_1} + \varepsilon \quad (2.5)$$

Model kubik

$$y=\beta_0+\beta_1 x_1+\beta_2 x_1^2+\beta_3 x_1^3+\varepsilon \quad (2.6)$$

2.5 Estimasi Analisis Regresi Sederhana

Untuk mendapatkan koefisien regresi sederhana dapat digunakan metode *Ordinary Least Square* yang bertujuan untuk meminimumkan jumlah kuadrat error. Penduga OLS yang memenuhi kriteria dapat ditemukan dalam du acara sebagai berikut

Pendekatan Pertama, digunakan suatu prosedur pencarian numerik. Prosedur ini digunakan untuk berbagai nilai dugaan β_0 dan β_1 yang berbeda sampai diperoleh nilai dugaan yang minimum.

Pendekatan kedua yaitu dengan meminimumkan jumlah kuadrat error.

$$\sum\varepsilon^2=\sum(y_i-\beta_0-\beta_1 x_i)^2 \quad (2.7)$$

Untuk memperoleh β_0 dan β_1, maka dapat mendeferensialkan persamaan diatas terhadap β_0 dan β_1, sehingga diperoleh

$$\frac{\partial Q}{\partial \beta_0} = -2\sum(y_i - \beta_0 - \beta_1 x_i) \quad (2.8)$$

$$\frac{\partial Q}{\partial \beta_1} = -2\sum x_i(y_i - \beta_0 - \beta_1 x_i) \quad (2.9)$$

Selanjutnya kedua turunan parsial ini disama dengankan nol, maka diperoleh

$$\sum (y_i - \beta_0 - \beta_1 x_i) = 0 \quad (2.10)$$

$$\sum x_i (y_i - \beta_0 - \beta_1 x_i) = 0 \quad (2.11)$$

Dengan menyelesaikan persamaan-persamaan diatas, maka dapat diperoleh persamaan sebagai berikut

$$\widehat{\beta_1} = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i{}^2 - (\sum x_i)^2} \quad (2.12)$$

$$\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{\bar{x}} \quad (2.13)$$

2.6 Standar Error dari OLS

Analisis regresi pada dasarnya disebut sebagai regresi sampel karena mengestimasi regresi populasi. Oleh karena itu, penduga $\widehat{\beta_0}$ dan $\widehat{\beta_1}$ yang diperoleh dari metode OLS adalah variabel yang sifatnya acak dimana nilainya dapat berubah dari satu sampel ke sampel yang lain. Adanya variabilitas penduga ini maka dibutuhkan ketepatan penduga $\widehat{\beta_0}$ dan $\widehat{\beta_1}$. Untuk mengetahui ketepatan penduga OLS ini diukur dengan menggunakan standar error. Dengan kata lain standard error ini mengukur ketepatan estimasi penduga $\widehat{\beta_0}$ dan $\widehat{\beta_1}$. Adapun rumus standar error untuk $\widehat{\beta_0}$ dan $\widehat{\beta_1}$ adalah sebagai berikut

$$var(\widehat{\beta_0}) = \frac{\sum x_i^2}{n\sum x_i^2}\sigma^2 \quad (2.14)$$

$$Se(\widehat{\beta_0}) = \sqrt{var(\widehat{\beta_0})} = \sqrt{\frac{\sum x_i^2}{n\sum x_i^2}\sigma^2} \quad (2.15)$$

$$var(\widehat{\beta_1}) = \frac{\sigma^2}{\sum x_i^2} \tag{2.16}$$

$$Se(\widehat{\beta_1}) = \sqrt{var(\widehat{\beta_1})} = \sqrt{\frac{\sigma^2}{\sum x_i^2}} \tag{2.17}$$

Dimana var adalah varian atau ragam, *Se* adalah standar error dan σ^2 adalah varian konstan. Nilai penduga σ^2 dapat dihitung menggunkan rumus sebagai berikut

$$\hat{\sigma}^2 = \frac{\sum \hat{e}_i^2}{n-k} \tag{2.18}$$

Dimana, $x_i = X_i - \bar{x}$, n adalah jumlah observasi, k adalah jumlah parameter estimasi, dan $\sum \hat{e}_i^2$ adalah jumlah kuadrat error, $n - k$ dikenal sebagai derajat bebas. Semakin kecil standar error dari penduga maka semakin kecil variabilitas dari angka penduga dan berarti semakin dipercaya nilai penduga yang didapat.

2.5 Pengujian Hipotesis

Hipotesis merupakan pernyataan tentang sifat populasi sedangkan uji hipotesis adalah suatu prosedur untuk pembuktian kebenaran sifat populasi berdasarkan data sampel. Hipotesis dapat dibagi menjadi dua yaitu H_0 dan H_1 yang biasanya disebut hipotesis tandingan. Pada regresi sederhana pengujian hipotesis biasanya digunakan untuk mengetahui apakah variabel independen berpengaruh terhadap variabel dependennya. Adapun Langkah uji hipotesis dalam regresi sederhana yaitu sebagai berikut.

1. Membuat hipotesis

   $H_0 : \beta_1 = 0$

   $H_0 : \beta_1 \neq 0$

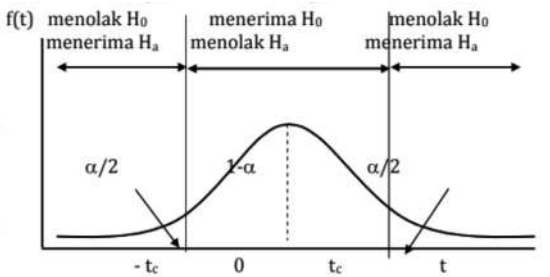
   Apabila ingin menguji penduga β_0, maka hipotesisnya adalah

   $H_0 : \beta_0 = 0$

   $H_0 : \beta_0 \neq 0$

2. Menetapkan nilai kritis t tabel

   Pada regresi sederhana pengujian hipotesis menggunakan uji t. Sehingga titik kritis pada pengujian hipotesis regresi sederhana, Tolak H_0 apabila *t hitung* > *t tabel* (t_{α},*db)*.



Gambar 2.6.

Daerah penolakan (penerimaan) H_0: $\beta_1=0$ dan H_a: $\beta_1 \neq 0$

Gambar 2.5. Daerah Kritis Uji Hipotesis

3. Menghitung nilai *t* statistik

$$t = \frac{\hat{\beta}_1 - \beta_1}{Se(\beta_1)}$$

4. Membandingkan nilai *t* statistik dengan titik kritisnya
   - Jika nilai *t* hitung > nilai *t* tabel maka H_0 ditolak artinya variabel independent secara signifikan berpengaruh terhadap variabel dependennya
   - Jika nilai *t* hitung < nilai *t* tabel maka H_0 diterima artinya variabel independen tidak berpengaruh terhadap variabel dependennya

Kunci dari analisis regresi linear adalah pada dependensi suatu variabel ke satu variabel lainnya. Tujuan dari analisis ini adalah untuk menaksir atau memprediksi nilai rata-rata dari variabel dependen atas variabel dependen (*exploratory variable*) yang sudah diketahui atau ditetapkan. Keberhasilan dari analisis regresi pada ketersediaan dan ketepatan data. Oleh karena itu, sumber data yang jelas, definisi, metode pengumpulannya dan keterbatasan-keterbatasannya harus dijelaskan di awal.

BAB 3.
ANALISIS REGRESI LINEAR BERGANDA

3.1 Pengantar Analisis Regresi Linear Berganda

Dalam prakteknya banyak model regresi sederhana tidak mencerminkan kondisi variabel dependen yang sebenarnya, karena variabel dependen dapat dipengaruhi tidak hanya dipengaruhi oleh satu faktor saja namun dapat dipengaruhi oleh banyak faktor. Misalkan permintaan barang X oleh konsumen tidak hanya dipengaruhi oleh faktor harga, tetapi juga bisa dipengaruhi oleh faktor harga barang lain, pendapatan konsumen dan sebagainya. Oleh karena itu, analisis yang dapat digunakan untuk menganalisis satu variabel dependen dengan satu/lebih variabel independen, maka digunakan analisis regresi berganda.

Sama seperti halnya pada regresi sederhana, untuk mendapatkan koefisien pada regresi berganda digunakan metode *Ordinary Least Square*. Ada beberapa asumsi OLS yang digunakan dalam regresi berganda, yang akan dijelaskan pada subbab berikutnya. Salah satu asumsi OLS yang harus dipenuhi pada regresi berganda yaitu multikolinieritas. Multikolinieritas merupakan hubungan linieritas antar variabel. Oleh karena pada regresi berganda terdapat lebih dari 1 variabel independen, maka asumsi multikolinieritas ini harus terpenuhi. Penjelasan lebih lanjut mengenai multikolinieritas akan dijelaskan pada bab berikutnya.

Dalam hal pengartian koefisien regresi berganda, β, terdapat sedikit perbedaan dari regresi sederhana. Misalkan terdapat dua koefisien regresi berganda yaitu β_1 dan β_2. β_1 dapat diartikan

yaitu mengukur perubahan rata – rata y atau variabel dependen terhadap perubahan per unit x_1 dengan asumsi variabel x_2 tetap. Begitu pula dengan β_2 adalah dapat diartikan yaitu mengukur perubahan rata–rata y atau variabel dependen terhadap perubahan per unit x_2 dengan asumsi variabel x_1 tetap.

3.2 Ordinary Least Square

Ordinary Least Square (OLS) atau yang biasa disebut metode kuadrat terkecil digunakan untuk mengestimasi penduga regresi dengan cara meminimumkan residual. Residual atau *error* dalam model diharapkan seminimal mungkin karena semakin minimal residualnya maka kesalahan ramalan model bisa menggambarkan data sebenarnya. Kuadrat di sini dilakukan untuk menghasilan nilai yang sekecil mungkin. Biasanya, residul itu bernilai kurang dari 1. Jika nilai kurang dari satu dikuadratkan maka akan menghasilkan nilai yang sangat kecil. Misalnya diperoleh ε = 0,3. Maka nilai ε^2= 0,09. Misalkan untuk model regresi dengan k peubah sebagai berikut ini

$$y_i=\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+...+\beta_k x_{ki}+\varepsilon_i, \qquad (3.1)$$

Dimana i = 1, 2, …, n. Dengan β_0 adalah intersep, β_1,.... , β_k adalah penduga koefisien regresi, ε adalah error, i adalah amatan ke- i, dan n adalah ukuran populasi atau jumlah observasi. Notasi pada persamaan (3.1) dapat dinyatakan dalam persamaan simultan sebagai berikut:

$$y_1=\beta_0+\beta_1 x_{11}+\beta_2 x_{21}+....+\beta_k x_{k1}+\varepsilon_i \qquad (3.2)$$

$$y_2=\beta_0+\beta_1 x_{12}+\beta_2 x_{22}+....+\beta_k x_{k2}+\varepsilon_i$$

$$\vdots$$

$$y_n=\beta_0+\beta_1 x_{1n}+\beta_2 x_{2n}+....+\beta_k x_{kn}+\varepsilon_i$$

Persamaan simultan (3.2) dapat dituliskan kedalam bentuk matriks sebagai berikut:

$$\begin{bmatrix} y_1 \\ y_1 \\ \vdots \\ y_1 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

atau dapat diubah dalam bentuk lain yaitu,

$$\widehat{\boldsymbol{Y}} = \boldsymbol{X}\widehat{\boldsymbol{B}} + \hat{\boldsymbol{\varepsilon}} \tag{3.3}$$

Pada regresi linier berganda dengan *k* peubah, pendugaan metode OLS yaitu dengan meminimumkan errornya.

$$\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} \tag{3.4}$$

Dengan $\hat{\boldsymbol{\varepsilon}} = \widehat{\boldsymbol{Y}} - \boldsymbol{X}\widehat{\boldsymbol{B}}$. Sehingga didapatkan,

$$\boldsymbol{J} = \hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}} = \left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right)'\left(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right) \tag{3.5}$$

$$= \boldsymbol{Y}'\boldsymbol{Y} - 2\widehat{\boldsymbol{\beta}}\boldsymbol{X}'\boldsymbol{Y} + \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} \tag{3.6}$$

Pada metode OLS untuk mendapatkan nilai penduga $\hat{\beta}$, maka persamaan (3.6) dapat diturunkan terhadap β, sebagai berikut,

$$\frac{\partial \boldsymbol{J}}{\partial \beta_0} = -2\sum(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) = 0 \tag{3.7}$$

$$\frac{\partial \boldsymbol{J}}{\partial \beta_1} = -2\sum(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) x_{i1} = 0$$

$$\frac{\partial \boldsymbol{J}}{\partial \beta_2} = -2\sum(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) x_{i2} = 0$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\frac{\partial \boldsymbol{J}}{\partial \beta_k} = -2\sum(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \cdots - \beta_k x_{ik}) x_{ik} = 0$$

Kemudian apabila persamaan (3.7) diubah dengan cara mengganti parameter – parameter yang ada penduganya, dan meletakkan Y di kanan persamaan, maka persamaan diatas dapat dituliskan sebagai berikut

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum x_{i1} + \hat{\beta}_2 \sum x_{i2} + \ldots + \hat{\beta}_k \sum x_{ik} = \sum Y_i \quad (3.8)$$

$$\hat{\beta}_0 \sum x_{i1} + \hat{\beta}_1 \sum x_{i1}^2 + \hat{\beta}_2 \sum x_{i2}x_{i1} + \ldots + \hat{\beta}_k \sum x_{ik}x_{i1} = \sum y_i x_{i1}$$

$$\hat{\beta}_0 \sum x_{i2} + \hat{\beta}_1 \sum x_{i1}x_{i2} + \hat{\beta}_2 \sum x_{i2}^2 + \ldots + \hat{\beta}_k \sum x_{ik}x_{i1} = \sum y_i x_{i2}$$

$$\hat{\beta}_0 \sum x_{ik} + \hat{\beta}_1 \sum x_{i1}x_{ik} + \hat{\beta}_2 \sum x_{i2}x_{ik} + \cdots + \hat{\beta}_k \sum x_{ik}^2 = \sum y_i x_{ik}$$

Persamaan (3.8) dapat diubah menjadi bentuk matriks sebagai berikut,

$$\begin{bmatrix} n & \sum x_{i1} & \ldots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \ldots & \sum x_{ik} \sum x_{i1} \\ \ldots & \ldots & \ldots & \\ \sum x_{1k} & \sum x_{i2}x_{i1} & \ldots & \sum x_{ik}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \ldots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_{11} & x_{12} & \ldots & x_{1k} \\ \ldots & \ldots & & \ldots \\ x_{11} & x_{11} & \ldots & x_{11} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \ldots \\ y_k \end{bmatrix}$$

Secara sederhana dapat dituliskan sebagai berikut,

$$(\boldsymbol{X'X})\widehat{\boldsymbol{\beta}} = \boldsymbol{X'Y}$$

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

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y} \quad (3.9)$$

Sehingga persamaan yang digunakan untuk menduga estimator β, yaitu,

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y} \quad (3.10)$$

3.3 Asumsi Model Regresi Linier dalam Notasi Matriks

Adapun asumsi regresi linier berganda dengan beberapa penyesuaian yaitu

- Diasumsikan E(ε) = **0** dengan ε dan **0** adalah vector kolom berdimensi n x 1 atau dapat didefinisikan sebagai berikut

$$E\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix} = \begin{bmatrix}E(\varepsilon_1)\\ E(\varepsilon_2)\\ \vdots\\ E(\varepsilon_n)\end{bmatrix} = \begin{bmatrix}0\\ 0\\ \vdots\\ 0\end{bmatrix} = \mathbf{0} \quad (3.11)$$

- Nilai E ($\boldsymbol{\varepsilon\dot{\varepsilon}}$) = $\sigma^2 \boldsymbol{I}$ dengan *I* adalah matriks identtas berukuran *n x n*.

$$\mathrm{E}(\boldsymbol{\varepsilon\varepsilon}') = \begin{bmatrix}\sigma^2 & 0 & \cdots & 0\\ 0 & \sigma^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma^2\end{bmatrix} = \sigma^2 \boldsymbol{I} \quad (3.12)$$

Persamaan (3.12) disebut matriks varians kovarians dari *error* (ε),

- Matriks $\boldsymbol{X}$ yang berdimensi $n \; x \; (k + 1)$ adalah nonstokastik atau dianggap tidak berubah.
- Rank Matriks *X* adalah p(*X*) = *k*, dengan *k* adalah banyak kolom pada matriks *X* dan *k* kurang dari jumlah amatan n. Artinya kolom–kolom pada matriks *X* saling bebas linier atau dengan kata lain tidak terdapat multikolinieritas.
- Vektor ε berditribusi normal, $\varepsilon \sim N(\mathbf{0}, \sigma 2\boldsymbol{I})$

3.4 Matriks Varians-Kovarians

Matriks varians dan kovarians digunakan tidak hanya untuk menghitung varians $\widehat{\beta_1}$, namun dapat digunakan juga untuk menghitung kovarians dari dua elemen pada β. Matriks varians kovarians didefinsikan sebagai berikut

$$var - covar(\hat{\boldsymbol{\beta}}) = E\left[\left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)\left(\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right)'\right] \quad (3.13)$$

Bentuk umum matriks varians-kovarians $\hat{\beta}$ adalah sebagai berikut,

$$var - cov(\hat{\boldsymbol{\beta}}) = \begin{bmatrix} var(\hat{\beta}_1) & cov(\hat{\beta}_1, \hat{\beta}_2) & \cdots & cov(\hat{\beta}_1, \hat{\beta}_k) \\ cov(\hat{\beta}_2, \hat{\beta}_1) & var(\hat{\beta}_2) & \cdots & cov(\hat{\beta}_2, \hat{\beta}_k) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\hat{\beta}_k, \hat{\beta}_1) & cov(\hat{\beta}_k, \hat{\beta}_1) & \cdots & var(\hat{\beta}_k) \end{bmatrix} \quad (3.14)$$

Pada bagian sebelumnya sudah didapatkan bahwa penduga estimator β, yaitu

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{3.15}$$

Apabila disubstitusikan pada $\boldsymbol{Y} = \boldsymbol{X\beta} + \varepsilon$ke persamaan (3.15), maka diperoleh

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X\beta} + \boldsymbol{\varepsilon})$$

$$= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$$

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \tag{3.16}$$

Selanjutnya masukkan persamaan (3.16) kedalam definsi varians – kovarians,

$$var - covar(\widehat{\boldsymbol{\beta}}) = E\left[\left(\widehat{\boldsymbol{\beta}} - E(\widehat{\boldsymbol{\beta}})\right)\left(\widehat{\boldsymbol{\beta}} - E(\widehat{\boldsymbol{\beta}})\right)'\right]$$

$$= E\left[\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\right)\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\right)'\right]$$

$$= E\left[\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\right)\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\right] \tag{3.17}$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E((\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1})$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

$$= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

Penduga tak bias untuk $\hat{\sigma}^2$ pada regresi linier berganda dengan 2 peubah yaitu $\hat{\sigma}^2 = \sum \hat{\varepsilon}_i^2/(n-2)$. Penduga tak bias untuk $\hat{\sigma}^2$ pada regresi linier berganda dengan 3 peubah yaitu $\sigma\hat{}^2 = \sum \hat{\varepsilon}_i^2/(n-3)$. Penduga tak bias untuk $\hat{\sigma}^2$ pada regresi linier berganda dengan k peubah

$$\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_i^2}{n-k} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n-k} = \frac{Y'Y - \hat{\beta}X'Y}{n-k} \tag{3.18}$$

3.5 Uji Hipotesis Parsial Koefisien Regresi Linier Berganda

Uji hipotesis parsial koefisien pada regresi linier berganda bertujuan untuk mengetahui adanya pengaruh antara masing – masing penduga β_i terhadap variabel dependen Y. Pada re-

gresi linier berganda yang mempunyai lebih dari satu variabel independen, salah satu asumsi yang harus dipenuhi yaitu asumsi normalitas. Apabila variabel error ε_i berdistribusi normal, maka variabel dependen Y juga akan berdistribusi normal.

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_k X_{ki} + \varepsilon_i \quad (3.19)$$

i = 1,2,... , n

Maka $\varepsilon_i \sim N(0,\sigma^2)$ dan $Y_i \sim N(0,\sigma^2)$. Penduga β_i adalah fungsi linier terhadap variabel dependen Y, sehingga penduga β_i akan mempunyai distribusi normal dengan rata-rata β_i dan varian sebesar var (β_i).

$$\hat{\beta}_k \sim N[\beta_k, var(\beta_k)] \quad (3.20)$$

Apabila $\hat{\beta}_k$ dikurangi dengan rata – ratanya kemudian dibagi dengan akar variannya atau standar *error*nya, maka melakukan transformasi variabel random $\hat{\beta}_k$ yang berdistribusi normal variabel Z yang mempunyai standar normal sebagai berikut:

$$Z = \frac{\hat{\beta}_k - \beta_k}{\sqrt{var(\beta_k)}} \sim N(0,1) \quad untuk\ k = 1,2,\dots,k \quad (3.21)$$

Apabila $var(\beta_k)$ diganti dengan $var(\hat{\beta}_k)$, maka adakan didapatkan variabel random t sebagai berikut :

$$t = \frac{\hat{\beta}_k - \beta_k}{\sqrt{var(\hat{\beta}_k)}} \sim t_{(n-k)} \text{ atau } t = \frac{\hat{\beta}_k - \beta_k}{se(\hat{\beta}_k)} \sim t_{(n-k)} \quad (3.22)$$

Dengan derajat bebas (db) sebesar *n-k*, bergantung pada jumlah variabel independen yang terdapat pada model regresi linier berganda.

Prosedur untuk pengujian hipotesis uji parsial pada regresi berganda yaitu sebagai berikut.

1. Membuat hipotesis untuk penduga koefisien regresi β_k
   a. Uji hipotesis satu arah
      $H_0: \beta_k \leq 0$
      $H_1: \beta_k > 0$

Atau

H_0:$\beta_k \geq 0$

H_1:$\beta_k < 0$

b. Uji hipotesis dua arah

H_0:$\beta_k = 0$

H_1:$\beta_k \neq 0$

2. Menentukan wilayah kritis untuk distribusi t dengan db = *n-k*
3. Menghitung nilai t hitung untuk masing – masing penduga koefisien regresi

$$t = \frac{\hat{\beta}_k - \beta_k}{se(\hat{\beta}_k)}$$

4. Menarik keputusan dengan cara
   a. Apabila nilai $t_{hitung} > t_{tabel}$, maka H_0 ditolak atau penduga koefisien β_k berpengaruh secara signifikan
   b. Apabila nilai $t_{hitung} < t_{tabel}$, maka H_0 ditolak atau penduga koefisien β_k tidak berpengaruh secara signifikan

3.5 Uji Hipotesis Simultan Koefisien Regresi Linier Berganda

Uji hipotesis simultan koefisien regresi linier berganda bertujuan untuk menguji secara keseluruhan koefisien penduga ($\hat{\beta}_k$) regresi linier berganda. Pengujian hipotesis simultan koefisien regresi linier berganda menggunakan analisis varians. Dengan mengasumsikan *error* ε_i berdistribusi normal dengan H_0:β_1=β_2=...=β_k dan statistik uji.

$$F = \frac{\left(\widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} - n\bar{y}^2\right)/(k-1)}{\dfrac{\boldsymbol{Y}'\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y}}{n-k}} \qquad (3.23)$$

Dengan derajat bebas (db) sebesar *k* - 1 dan *n* - *k*, dimana *n* adalah jumlah observasi dan *k* adalah parameter penduga

termasuk konstanta. Keputusan uji hipotesis simultan koefisien regresi berganda yaitu Apabila nilai $t_{hitung} > t_{tabel}$, maka H_0 ditolak atau penduga koefisien β_k berpengaruh secara signifikan. Apabila nilai $t_{hitung} < t_{tabel}$, maka H_0 diterima atau penduga koefisien β_k tidak berpengaruh secara signifikan.

Sumber variasi	JK	Db	KT
Regresi	$\widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} - n\bar{y}^2$	k-1	$\frac{\widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} - n\bar{y}^2}{k-1}$
Sisaan	$\boldsymbol{Y}'\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y}$	n-k	$\frac{\boldsymbol{Y}'\boldsymbol{Y} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y}}{n-k}$
Total	$\boldsymbol{Y}'\boldsymbol{Y} - n\bar{y}^2$	n-1	

3.6 Koefisien Determinasi

Koefisien determinasi R^2 digunakan untuk menjelaskan seberapa besar keragaman variabel dependen yang dijelaskan oleh variabel independen. Koefisien determinasi R^2 didefinsikan sebagai rasio dari jumlah kuadrat regresi dibagi dengan jumlah kuadrat total. Koefisien detreminasi R^2 didefiniskan sebagai berikut,

$$R^2 = \frac{JKR}{JKT} \tag{3.24}$$

Koefisien determinasi R^2 regresi linier berganda pada kasus 2 variabel independen

$$R^2 = \frac{JKR}{JKT} = \frac{\hat{\beta}_1^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.25}$$

Koefisien determinasi R^2 regresi linier berganda pada 3 variabel independent,

$$R^2 = \frac{JKR}{JKT} = \frac{\hat{\beta}_1 \sum_{i=1}^{n} y_i(x_{2i} - \bar{x}_2) + \hat{\beta}_2 \sum_{i=1}^{n} y_i(x_{3i} - \bar{x}_3)}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3.26}$$

Koefisien determinasi R^2 regresi linier berganda pada k peubah,

$$R^2 = \frac{JKR}{JKT} = \frac{\hat{\beta}_1 \sum_{i=1}^{n} y_i(x_{2i} - \bar{x}_2) + \cdots + \hat{\beta}_k \sum_{i=1}^{n} y_i(x_{ki} - \bar{x}_k)}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Atau dapat pula dituliskan dalam bentuk notasi matriks

$$R^2 = \frac{JKR}{JKT} = \frac{\widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} - n\bar{y}^2}{\boldsymbol{Y}'\boldsymbol{Y} - n\bar{y}^2} \tag{3.27}$$

Koefisien determinasi R^2 akan semakin besar jika variabel independen ditambah ke dalam suatu model. Hal ini terjadi karena $\sum_{i=1}^{n}(y_i - \bar{y})^2$ atau *JKT* bukan fungsi yang berasal dari variabel independen *X*, sedangkan *JKR* bergantung pada jumlah variabel independen *X* dalam model. Sehingga apabila jumlah variabel independen *X* bertambah, maka nilai *JKR* akan menurun. Nilai koefisien determinasi yang tidak pernah menurun ini mengakibatkan suatu kehatia – hatian dalam membandingkan dua buah model regresi yang mempunyai variabel dependen *Y* yang sama tetapi memiliki jumlah variabel independen *Y* yang berbeda. Hal ini dikarenakan bahwa tujuan metode OLS adalah utuk mendapatkan nilai koefisien determinasi yang tinggi. Sebagai alternative untuk mengatasi persoalan diatas adalah dengan menggunakan koefisien determinasi R^2 yang disesuaikan (adjusted R^2). Koefisien determinasi R^2 yang disesuaikan didefinsikan sebagai berikut

$$R^2 = \frac{JKR}{JKT} = \frac{\widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y} - n\bar{y}^2/(n-k)}{\boldsymbol{Y}'\boldsymbol{Y} - n\bar{y}^2/(n-1)} \tag{3.28}$$

3.7 Asumsi yang Mendasari OLS

Metode *Ordinary Least Square* (OLS) merupakan metode pendugaan yang sering digunakan karena selain mudah juga karena memiliki beberapa sifat teoritis yang kokoh, yang diringkas dalam Teorema Gauss-Markov. Pada Teorema Gauss Markov metode OLS akan menghasilkan estimator yang mempunyai sifat tidak bias, linier, dan memiliki varian minimum (*best linier unbiased*

estimator/BLUE). Namun, metode kuadrat terkecil tidak membuat asumsi probabilistic terhadap ε_i. Sehingga, untuk melakukan inferensi terhadap fungsi regresi populasi dan fungsi regresi sampel, Teorema Gauss Markov tidak dapat digunakan. Oleh sebab itu, *error* ε_i diasumsikan menyebar normal. Penambahan asumsi ini pada model regresi klasik dikenal sebagai regresi linier normal klasik.

Asumsi model regresi normal mengasumsikan masing – masing ε_i menyebar normal dengan nilai tengah $E(\varepsilon_i)=0$, varians $E(\varepsilon_i\text{-}E(\varepsilon_i))^2=E(\varepsilon_i^2)=\sigma^2$ dan kovarian cov $(\varepsilon_i,\varepsilon_j)=E(\varepsilon_i\ \varepsilon_j)=0$ untuk $i\neq j$ atau biasa disingkat dengan $\varepsilon_i \sim N(0,\sigma^2)$. Atau diasumsikan saling bebas dan berdistribusi identik, atau IID normal. Ada beberapa alasan kenapa *error* ε_i menurut Gujarati (2004) yaitu :

- Teorema limit pusat menyatakan jika terdapat sejumlah besar variabel acak yang berdistribusi bebas dan identic, dengan sedikit pengecualian, distribusi dari jumlahnya menuju distribusi normal. *Error* ε_i menyatakan pengaruh kombinasi dari variable independen dan variabel dependen yang tidak secara eksplisit dimasukkan pada model regresi. Sehingga diharapka variabel yang dihilangkan atau diabaikan ini kecil dan acak.
- Dalam Teorema Limit Pusat juga dijelaskan bahwa bahkan jika jumlah variabel tidak sangat besar atau jika variabel – variabel ni tidak saling bebas tegas, jumlah mungkin tetap normal
- Dengan asumsi kenormalan, distribusi peluang dari penduga kuadrat terkecil dapat diperoleh dengan mudah karena sifat kenormalan yaitu sebarang fungsi linier dari variabel yang menyebar normal juga akan menyebar normal. Penduga koefisien regresi $\widehat{\beta_0}$ dan $\widehat{\beta_1}$. adalah fungsi linier dari ε_i. Dengan demikian apabila ε_i menyebar normal, maka $\widehat{\beta_0}$ dan $\widehat{\beta_1}$. akan menyebar normal
- Distribusi normal merupakan distribusi yang cukup sederhana yang hanya melibatkan dua parameter, yaitu nilai

tengah dan varians. Sifat – sifat distribusi ini telah banyak dipelajari secara ekstensif

- Data dengan ukuran kecil atau berhingga, missal 100, asumsi kenormalan memegang peranan penting. Asumsi ini membantu untuk mendapatkan distribusi peluang eksak dari penduga kuadrat terkecil.

Berdasarkan asumsi bahwa ε_i berdistribusi normal, penduga kuadrat terkecil memiliki sifat – sifat berikut :

1. Bersifat tak bias
2. Memiliki varian/ ragam minimum (penduga efisien)
3. Konsisten. Artinya, seiring meningkatkatnya ukuran sampel menuju tak berhingga, penduga tersebut konvergen menuju nilai parameter populasi
4. Konstanta $\hat{\beta}_0 \sim N(\beta_0, \sigma^2_{\hat{\beta}_0})$, dengan penduga varian/ ragam

$$\sigma^2_{\hat{\beta}_0} = \frac{\sigma^2 \sum X_i^2}{n \sum (X_i - \bar{X})^2}$$

5. Penduga $\beta_1 \sim N(\beta_0, \sigma^2_{\hat{\beta}_0})$, dengan penduga varian/ ragam

$$\sigma^2_{\hat{\beta}_1} = \frac{\sigma^2}{n \sum (X_i - \bar{X})^2}$$

6. Kuantitas $(n-2)(\hat{\sigma}^2/\sigma^2)$ berdistribusi X^2 dengan derajat bebas (n-2)
7. Pendugaann $\hat{\beta}$ menyebar dan saling bebas dari $\hat{\sigma}^2$

3.8 Estimasi Interval

Estimasi untuk mendapatkan suatu penduga tidak hanya didapatkan dari pendugaan titik saja. OLS atau pendugaan kuadrat terkecil adalah suatu pendugaan titik. Selain pendugaan titik, terdapat pendugaan selang dimana parameter populasi berada diantara nilai interval atau selang dengan tingkat signifikansi tertentu. Pada estimasi interval, dua estimator $\hat{\theta}_1$ dan $\hat{\theta}_2$ pada estimasi selang dapat didefinsiskan sebagai berikut.

$$P\left(\hat{\theta}_1 < \theta < \hat{\theta}_2\right) = 1 - \alpha, 0 < \alpha < 1 \quad (3.29)$$

Interval ini disebut dengan selang kepercayaan atau selang kepercayaan *1 - a* untuk θ. Nilai *1 - a* disebut koefisien kepercayaan. Jika $a = 0.05$, maka 1 - a = 0.95 artinya kita percaya bahwa 95% estimator ada pada selang kepercayaan dengan tingkat siginifikansi 5%.

Pada regresi, selang kepercayaan untuk penduga β didefinisikan sebagai berikut

$$P\left(-t_c \leq \frac{\hat{\beta}_k - \beta_k}{se(\hat{\beta}_k)} \leq t_c\right) = 1 - \alpha \quad (3.30)$$

dengan t_c adalah nilai kritis table distribusi t dengan derajat bebas (n - k) sehingga $P(t \geq t_c) = a/2$. Persamaan (4.30) dapat ditulis kembali sebagai berikut:

$$P\left[\hat{\beta}_k - t_c se(\hat{\beta}_k) \leq \beta_k \leq \hat{\beta}_k + t_c se(\hat{\beta}_k)\right] = 1 - \alpha \quad (3.31)$$

Atau dapat disederhanakan menjadi sebagai berikut:

$$\left[\hat{\beta}_k - t_c se(\hat{\beta}_k), \hat{\beta}_k + t_c se(\hat{\beta}_k)\right] \quad (3.32)$$

dimana (1 - *a)* merupakan selang kepercayaan untuk koefisien β. Apabila $a = 5\%$, artinya bahwa kita percaya bahwa 95% penduga β ada pada selang kepercayaan dengan tingkat signifikansi sebesar $a = 5\%$.

BAB 4.
VARIABEL DUMMY

SKALA data pada statistika dapat dikategorikan menjadi empat yaitu skala nominal, skala ordinal, skala interval, dan skala rasio. Bahasan tentang skala data ini sudah dijelaskan pada Bab 1

Pada umumnya, analisis regresi menggunakan data–data yang berasal dari skala rasio, baik untuk variabel dependen maupun variabel independen. Namun, analisis regresi juga dapat digunakan apabila variabel dependen maupun variabel independennya berasal dari variabel yang diklasifikasikan sebagai jenis data kategorik atau data berskala nominal. Pada Bab ini, akan berfokus pada variabel independen yang berskala nominal yang sering disebut variabel dummy.

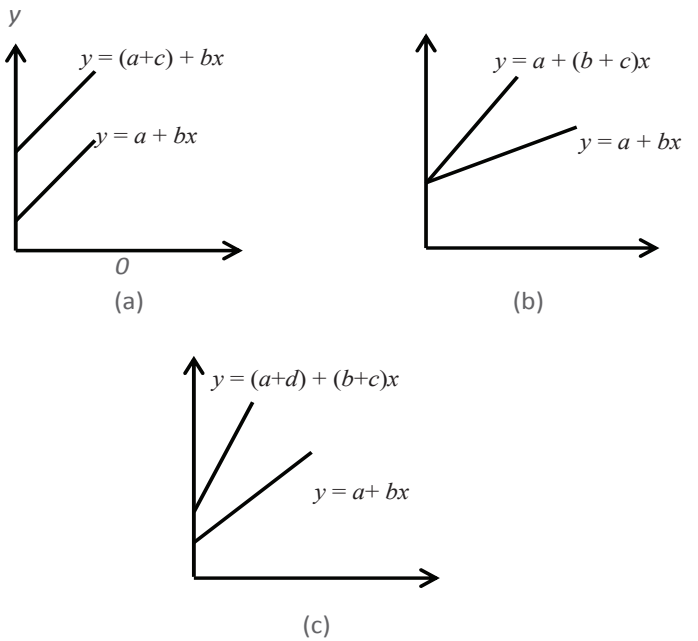
Variabel dummy adalah variabel yang digunakan untuk mengkuantifikasikan variabel – variabel yang bersifat kualitatif dan diduga berpengaruh terhadap variabel dependen. Cara yang digunakan untuk mengkuantifikasikan variabel kualitatif yaitu dengan membangun variabel buatan/boneka dengan memberikan nilai 1 dan 0. Gujarati (2004) mengatakan bahwa variabel dummy adalah variabel yang bernilai 1 dan 0. Nilai 1 diartikan sebagai adanya kepemilikan atribut, sedangkan nilai 0 menunjukkan tidak adanya kepemilikan atribut. Misalkan pada variabel jenis kelamin, nilai 1 untuk jenis kelamin laki – laki dan nilai 0 untuk jenis kelamin perempuan. Variabel jenis wilayah , nilai 1 untuk wilayah maju dan nilai 0 untuk wilayah tertinggal.

4.2 Model Regresi Variabel Dummy

Variabel dummy biasanya dinotasikan sebagai D dan memiliki 2 nilai kategori yaitu 1 dan 0. Penulisan untuk variabel dummy biasanya dituliskan sebagai D=1 untuk menuliskan adanya kepemilikian atribut (wanita, Jawa, Islam, dan lain-lain) dan D=0 untuk menuliskan tidak adanya kepemilikan atribut. Variabel dummy digunakan sebagai solusi untuk mengetahui keeratan hubungan antara variabel independen yang bersifat kualitatif dengan variabel dependen. Adapun model yang bisa dibangun menggunakan variabel dummy dalam analisis regresi adalah sebagai berikut:

1. $y=a+bx+cD_1$ (Model Dummy Instersep)
2. $y=a+bx+cD_1 x$ (Model Dummy Slope)
3. $y=a+bx+cD_1 x+dD_1$ (Model Dummy Kombinasi)

Dalam bentuk grafis, bentuk-bentuk model dummy tersebut digambarkan pada Gambar 3.1.



Gambar 3.1. Berbagai model dummy (Basuki, 2017)

4.2 Aturan Pembuatan Variabel Dummy

Gujarati (2004) mengemukakan beberapa aturan pembuatan variabel dummy. Aturannya adalah sebagai berikut:

- Jika variabel bebas kualitatif memiliki m kategori, maka banyaknya variabel dummy yang terbentuk adalah sebanyak m – 1 variabel.
- Kategori yang tidak memiliki variabel dummy merupakan kategori kontrol, kategori acuan, kategori dasar, kategori pembanding, kategori referensi, atau kategori yang dihilangkan.
- Nilai intersepsi β_1 merupakan nilai rata – rata dari kategori acuan.
- Koefisien regresi yang melekat pada variabel dummy disebut sebagai koefisien intersepsi diferensial karena menunjukkan perbedaan nilai dari kategori yang variabel dummynya bernilai 1 dengan kategori yang dijadikan sebagai variabel acuan.
- Jika variabel kualitatif memiliki lebih dari 1 kategori, maka pemilihan kategori yang akan dijadikan acuan diserahkan kepada peneliti sepenuhnya
- Berhati – hati terhadap jebakan variabel dummy. Yaitu kesalahan dala menentukan jumlah variabel dummy dalam model regresi. Cara yang dapat digunakan agar terhindar dari kesalahan ini adalah dengan menggunakan variabel dummy sejumlah kategori yang ada dalam pembuatan model regresinya dan menghilangkan konstanta dalam model. Jika ingin menggunakan cara ini, pastikan untuk menggunakan pilihan tanpa konstanta dalam .

4.3 Model ANOVA

Pada umumnya model regresi dapat mengandung variabel independen dengan data bersifat kualitatif maupun kuantitatif. Namun, ada keadaan dimana model regresi dapat mengandung variable independent yang semua datanya bersifat kualitatif.

Model regresi ini disebut sebagai model analisis varians (ANOVA). Perhatikan ilustrasi dibawah ini .

Tabel 4.1 Data PDRB Indonesia Tahun 2020

Provinsi	PDRB	Wilayah 1 (D1)	Wilayah 2 (D2)	Wilayah 3 (D3)
Aceh	166377	1	0	0
Sumatera Utara	811283	1	0	0
Sumatera Barat	242119	1	0	0
Riau	729167	1	0	0
Jambi	206846	1	0	0
Sumatera Selatan	458430	1	0	0
Bengkulu	73337	1	0	0
Lampung	354632	1	0	0
Bangka Belitung	75534	1	0	0
Kep. Riau	254253	1	0	0
DKI Jakarta	2772381	0	1	0
Jawa Barat	2088039	0	1	0
Jawa Tengah	1348600	0	1	0
DI Yogyakarta	138389	0	1	0
Jawa Timur	2299465	0	1	0
Banten	626437	0	1	0
Bali	224214	0	1	0
NTB	133522	0	0	1
NTT	106506	0	0	1
Kalbar	214002	0	0	1
Kalteng	152191	0	0	1
Kalsel	179151	0	0	1
Kaltim	607321	0	0	1
Kalut	100544	0	0	1
Sulut	132299	0	0	1
Sulteng	197441	0	0	1
Sulsel	504479	0	0	1
Sultengg	130184	0	0	1
Gorontalo	41726	0	0	1
Sulbar	45909	0	0	1
Maluku	46264	0	0	1
Maluku Utara	42142	0	0	1
Papua Barat	83566	0	0	1
Papua	198929	0	0	1

Tabel 4.1 adalah data PDRB tahun 2020 atas dasar harga berlaku (miliar rupiah) untuk 34 provinsi di Indonesia yang terbagi menjadi 3 wilayah. Wilayah tersebut dibedakan atas dasar jarak masing – masing provinsi dengan DKI Jakarta. Wilayah 1 adalah wilayah Indonesia bagian barat kecuali pulau Jawa dan Bali, wilayah 2 adalah provinsi yang berada di Pulau Jawa dan Bali, sedangkan wilayah 3 adalah provinsi – provinsi Indonesia tengah dan timur kecuali Bali. Untuk mengetahui rata – rata perbedaan PDRB pada ketiga wilayah tersebut, model analisis varians yang terbentuk adalah sebagai berikut :

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \varepsilon_i$$

Dengan

Y_i = rata – rata PDRB pada wilayah ke-i

D_{1i} = 1, untuk wilayah 2

= 0, untuk wilayah lainnya

D_{2i} = 1, untuk wilayah 3

= 0, wilayah lainnya

Maka rata – rata penghasilan responden adalah sebagai berikut :

a) Rata – rata PDRB untuk wilayah 1

$$E(Y_i \mid D_{1i}=0, D_{2i}=0) = \beta_0$$

b) Rata – rata PDRB untuk wilayah 2

$$E(Y_i \mid D_{1i}=1, D_{2i}=0) = \beta_0 + \beta_1$$

c) Rata – rata PDRB untuk wilayah 3

$$E(Y_i \mid D_{1i}=0, D_{2i}=1) = \beta_0 + \beta_2$$

Data diatas diolah dengan menggunakan software SPSS. Adapun langkah – langkah untuk menganalisa data dengan SPSS yaitu sebagai berikut :

1. Masukkan Data pada Data View dan berikan nama variabel pada Variable View
2. Klik Analyze → Regression → Linier Regression

3. Setelah kotak dialog Linier Regression muncul, masukkan variabel dependen pada kolom Dependent dan variabel independen pada kolom Independet (s).
4. Klik OK,

Tabel 4.2 Model Summary Regresi Dummy ANOVA

Model	R	R Square	Adj. R Square	Std. Error of Estimate
1	0.695	0.483	0.449	4.98797E5

Tabel 4.3 ANOVA Model Regresi Dummy ANOVA

Model	SS	df	MS	F	Sig
Regression	7.194E12	2	3.597E12	14.458	0.000
Residual	7.713E12	31	2.488E11		
Total	1.491E13	33			

Tabel 4.4 Coefficients Model Regresi Dummy ANOVA

Model	Unstandardized Coeff		Standardized Coeff	t	Sig
	B	Std. Error	Beta		
Constant	337197.800	157733.319		2.138	0.041
Wilayah3	-165658.035	198783.742	-0.125	-0.833	0.411
Wilayah2	1.020E6	245908.623	0.623	4.418	0.000

Berdasarkan ouput SPSS diatas, model regresi dummy ANOVA untuk PDRB setiap Provinsi di Indonesia adalah sebagai berikut:

$\hat{Y}$=337197,8 + 1020000 D_1 - 165658.035D_2

Nilai rata – rata PDRB untuk wilayah 1 yaitu,

$\hat{Y}$=337197,8 + 102000000 (0) - 165658.035(0)

=337197.2

Artinya bahwa nilai rata – rata PDRB untuk wilayah 1 yaitu 337197.2 miliar.

Nilai rata – rata PDRB untuk wilayah 2 yaitu,

$\hat{Y}$=337197,8+1020000 (1)-165658.035(0)

=1357197.8

Artinya bahwa nilai rata – rata PDRB untuk wilayah 2 adalah 1357197.8 miliar atau 1.3 triliun.

Dan nilai rata – rata PDRB untuk wilayah 3 yaitu,

$\hat{Y}$=337197,8+1020000 (1)-165658.035(0)

=171539.765

Artinya nilai rata – rata PDRB untuk wilayah 3 yaitu 171539.765 miliar. Setelah masing – masing nilai rata – rata PDRB untuk tiap wilayah didapatkan, dapat disimpulkan bahwa diantara ketiga wilayah tersebut, wilayah 2 memiliki nilai PDRB yang tertinggi dan wilayah 3 memiliki nilai rata – rata PDRB paling kecil.

Berdasarkan hasil output SPSS diatas, dapat diketahui nilai R^2 sebesar 0.483 atau 48.3%. Hal ini menunjukkan bahwa keragaman yang dapat dijelaskan oleh model sebesar 48.3%, sebanyak 51.7% lainnya dijelaskan oleh variabel independen lain yang belum dimasukkan kedalam model. Pada uji secara simultan koefisien regresi didapatkan nilai p-value sebesar 0.000. Hal ini dapat diartikan bahwa setidaknya ada 1 variabel independen dalam model yang berpengaruh dan dibuktikan pada uji parsial, wilayah 2 memiliki nilai p – value sebesar 0.000, yang artinya berpengaruh secara signifikan terhadap variabel dependennya atau dalam hal ini adalah PDRB.

4.4 Model ANCOVA

Pada umumnya model ANOVA seperti pada subbab 4.3 di atas tidak umum atau jarang terjadi dalam bidang ekonomi. Pada kebanyakan riset atau penelitian bidang ekonomi, model regresi yang dibangun mengandung sedikitnya 1 variabel independen kuantitatif dan variabel independen kualitatif. Model regresi yang

mengandung campuran data demikian disebut model analisis kovarian (ANCOVA). Model ANCOVA merupakan kelanjutan dari model ANOVA dimana adanya variabel independen kuantitatif ini dikontrol secara statistic dan dinamakan covariate. Perhatikan ilustrasi dibawah ini.

Diketahui data hasil pengamatan selama 15 tahun untuk menduga permintaan terhadap komoditas A selama periode 2003-2017 yang dipengaruhi oleh faktor X_1 (harga) dan kondisi pertumbuhan ekonomi (D) disajikan pada tabel berikut:

Y = jumlah permintaan komoditas A (ribu ton)
X_1 = harga komoditas A (Ribu rupiah per ton)
D = kondisi ekonomi (variabel dummy):
D = 1, jika kondisi ekonomi baik; dan
D = 0, jika kondisi ekonomi tidak baik *(buruk).*

Tabel 4.5 Data Permintaan Komoditas A Periode 2003 - 2017

Tahun	Y	X1	Kondisi Ekonomi (D)
2003	6.55	5.00	0
2004	8.70	5.10	1
2005	5.92	4.90	0
2006	4.30	5.90	0
2007	3.30	5.60	0
2008	6.23	4.90	0
2009	10.97	5.60	1
2010	9.14	8.50	1
2011	5.77	7.70	0
2012	6.45	7.10	0
2013	7.60	6.10	1
2014	11.47	5.80	1
2015	13.46	7.10	1
2016	10.24	7.60	1
2017	5.99	9.70	0

Data diatas diolah dengan menggunakan software SPSS. Adapun langkah – langkah untuk menganalisa data dengan SPSS yaitu sebagai berikut :

1. Masukkan Data pada Data View dan berikan nama variabel pada Variable View
2. Klik Analyze → Regression → Linier Regression
3. Setelah kotak dialog Linier Regression muncul, masukkan variabel dependen pada kolom Dependent dan variabel independen pada kolom Independet (s).
4. Klik OK,

Tabel 4.6 Model Summary Regresi Dummy ANCOVA

Model	R	R Square	Adj. R Square	Std. Error of Estimate
1	0.848	0.719	0.672	1.62989

Tabel 4.7 ANOVA Model Regresi Dummy ANCOVA

Model	SS	df	MS	F	Sig
Regression	81.500	2	40.750	15.340	0.000
Residual	31.878	12	2.657		
Total	113.379	4			

Tabel 4.8 Coefficients Model Regresi Dummy ANCOVA

Model	**Unstandardized Coeff**		**Standardized Coeff**	**t**	**Sig**
	B	**Std. Error**	**Beta**		
Constant	4.860	1.995		2.437	0.031
X1	0.111	0.301	0.057	0.368	0.719
D	4.641	0.846	0.842	5.488	0.000

Berdasarkan output diatas dapat diperoleh model ANCOVA untuk permintaan terhadap komoditas A sebagai berikut.

$\hat{Y}=4.860+0.111X_1+4.641D$

Model untuk permintaan terhadap komoditas A pada saat kondisi ekonomi baik (D=1) yaitu

$\hat{Y}=4.860+0.111X_1+4.641(1)$

$\hat{Y}=(4.860+4.641)+0.111X_1$

$\hat{Y}=9.501+0.111X_1$

Model untuk permintaan terhadap komoditas A pada saat kondisi ekonomi tidak baik (D=0) yaitu

$\hat{Y}=4.860+0.111X_1+4.641(0)$

$\hat{Y}=4.860+0.111X_1$

Berdasarkan output SPSS diatas dapat diketahui R^2 sebesar 0.719. Hal ini dapat diartikan bahwa keragaman yang dapat dijelaskan oleh model sebesar 71.9%. Sisanya yaitu sebesar 28.1% dijelaskan oleh variabel lain yang belum dimasukkan kedalam model. Apabila dilihat pada output diatas, untuk uji secara simultan, nilai p-value sebesar 0.000, yang artinya sedikitnya ada 1 variabel independen yang berpengaruh secara signifikan. Hal ini dibuktikan dengan uji parsial koefisien regresi bahwa variabel kondisi ekonomi memiliki nilai p-value sebesar 0.000, yang artinya variabel kondisi ekonomi berpengaruh secara signifikan terhadap permintaan komoditas A.

4.5 Model Regresi dengan Dummy Lebih dari Satu

Suatu model regresi dapat mengandung variabel independen kualitatif lebih dari satu, oleh karena itu kategori yang akan menjadi kontrol harus diperhatikan. Misalkan suatu penelitian ingin diketahui prediksi ketahanan kerja terhadap kepuasan kerja pegawai pada suatu Bank Syariah seperti pada tabel 5.3. Kepuasan kerja juga dipengaruhi oleh jabatan dan bidang kerja. Variabel jabatan terdiri dari 2 kategori yaitu manajerial dan pelaksana, sedangkan variabel bidang kerja terdiri dari dua kategori yaitu operasional, marketing, dan back office. Model regresi yang terbentuk yaitu sebagai berikut :

$$Y_i=\beta_0+\beta_1 X_i+\beta_2 D_{1i}+\beta_3 D_{2i}+\beta_4 D_{3i}+\varepsilon_i$$

Dengan :

Y_i = kepuasan kerja

X = ketahanan kerja

D_{1i} = 1 , untuk jabatan manajerial

= 0, untuk jabatan pelaksana

D_{2i} = 1, untuk bidang operasional

= 0, untuk bidang lainnya

D_{3i} = 1, bidang marketing

= 0, bidang lainnya

Tabel 4.9 Data Kepuasan Kerja Pegawai Bank Syariah

No	Y	X	D1	D2	D3
1	3	4	0	1	0
2	4	5	0	1	0
3	3	6	0	1	0
4	4	6	0	0	1
5	3	4	0	1	0
6	4	5	0	1	0
7	3	6	0	1	0
8	5	6	0	0	1
9	4	5	0	0	1
10	5	6	0	0	1
11	6	6	1	0	0
12	8	9	1	1	0
13	9	8	1	1	0
14	7	7	1	0	1
15	8	6	1	0	1
16	9	5	1	0	1
17	9	9	1	0	0
18	9	8	1	0	0
19	8	8	1	0	0
20	7	9	1	0	0

Data diatas diolah dengan menggunakan software SPSS. Adapun langkah – langkah untuk menganalisa data dengan SPSS yaitu sebagai berikut :

1. Masukkan Data pada Data View dan berikan nama variabel pada Variable View
2. Klik Analyze → Regression → Linier Regression
3. Setelah kotak dialog Linier Regression muncul, masukkan variabel dependen pada kolom Dependent dan variabel independen pada kolom Independet (s).
4. Klik OK,

Tabel 4.10 Model Summary Regresi Dummy

Model	R	R Square	Adj. R Square	Std. Error of Estimate
1	0.940	0.884	0.853	0.89469

Tabel 4.11 ANOVA Regresi Dummy

Model	SS	df	MS	F	Sig
Regression	91.793	4	22.948	28.668	0.000
Residual	12.007	15	0.800		
Total	103.800	19			

Tabel 4.12 Coefficients Regresi Dummy

Model	Unstandardized Coeff		Standardized Coeff	t	Sig
	B	Std. Error	Beta		
Constant	1.703	1.357		1.255	0.229
D1	3.762	0.603	0.826	6.234	0.000
D2	0.267	0.651	0.057	0.410	0.688
D3	0.975	0.635	0.204	1.536	0.145
X.1	0.2921	0.192	0.200	1.5200.229	0.149

Berdasarkan output SPSS dapat diperoleh model regresi dummy untuk kepuasaan kerja sebagai berikut :

$$\hat{Y}=1.703+0.292X_i+3.762D_{1i}$$

$$+ 0.267D_{2i}+0.975D_{3i}+\varepsilon_i$$

Berdasarkan output pada diatas dapat diketahui bahwa nilai R^2 sebesar 0.884. Hal ini dapat diartikan bahwa keragaman yang dapat dijelaskan oleh model yaitu sebesar 88.4%. Sebesar 11.6% sisanya dijelaskan oleh variabel lain yang belum dimasukkan ke dalam model. Pada uji simultan model regresi dummy kepuasaan kerja, nilai p-value sebesar 0.000, sehingga dapat disimpulkan bahwa setidaknya ada 1 variabel independen yang berpengaruh secara signifikan. Hal ini diperkuat dengan uji parsial pada variabel independen D1 atau jabatan kerja memiliki nilai p – value sebesar 0.000. Hal ini dapat diartikan bahwa jabatan kerja memiliki pengaruh secara signifikan terhadap kepuasan kerja pegawai Bank Syariah.

BAB 5.
ASUMSI-ASUMSI DALAM ANALISIS REGRESI

SALAH satu tujuan analisis regresi linear adalah mendapatkan estimasi parameter model yang menunjukkan besarnya pengaruh variabel independen terhadap variabel dependen berdasarkan data sampel. Estimasi parameter yang dihasilkan diharapkan dapat menggambarkan parameter populasi yang sebenarnya. Tujuan ini dapat dicapai dengan akurat apabila memenuhi beberapa asumsi, yaitu

- variabel independen dan variabel dependen memiliki hubungan linear
- *residual* atau *error* yang dihasilkan dari model regresi identik (*homoscedastic*), independen (*independent*), dan berdistribusi normal.

Menurut Best & Wolf (2015), jika asumsi 1 dan 2 terpenuhi maka *teorema* Gauss-Markov menjamin estimasi parameter OLS yang dihasilkan *Best Linear Unbiased Estimator* (BLUE). Selain kedua asumsi tersebut terdapat asumsi lain dalam analisis regresi linear dengan lebih dari satu variabel independen yang harus dipenuhi yaitu antara variabel independen tidak boleh saling berkorelasi atau tidak *multicollinearity* (Best & Wolf, 2015).

Ilustrasi 5.1

Pemodelan Indeks Harga Konsumen (IHK) berdasarkan BI *rate*, jumlah uang beredar, nilai bersih ekspor, dan nilai tukar rupiah terhadap dolar digunakan sebagai ilustrasi pada bab ini untuk menunjukkan asumsi-asumsi yang ada pada analisis

regresi linear. IHK digunakan sebagai variabel dependen (y), sedangkan BI *rate* (x_1), jumlah uang beredar (x_2), nilai bersih ekspor (x_3), dan nilai tukar rupiah terhadap dolar (x_4) digunakan sebagai variabel independen. Pemilihan variabel independen dan variabel dependen ini mengacu pada penelitian yang telah dilakukan oleh Sumantri, F. & Latifah, U. (2019). Data yang digunakan untuk ilustrasi diambil dari *website* resmi Badan Pusat Statistika (BPS) Indonesia. Data dapat dilihat pada Lampiran 2.

Langkah awal yang dilakukan sebelum masuk ke dalam pembahasan asumsi pada analisis regresi linear adalah mengecek apakah terdapat hubungan antara variabel dependen dan variabel independen yang digunakan. Hubungan antara kedua variabel ini ditunjukkan dengan pengujian koefisien korelasi. Tabel 5.1 berikut ini menyajikan hasil uji korelasi antara variabel dependen dan variabel independen yang digunakan pada Ilustrasi 5.1.

Tabel 5.1. Uji Korelasi Antara Variabel Dependen dan Variabel Independen Data Ilustrasi 5.1

Variabel dependen-variabel independen	**Nilai Korelasi**	**p_{value}**	**Kesimpulan**
y-x_1	0,397	0,002	Berkorelasi
y-x_2	-0,435	0,001	Berkorelasi
y-x_3	-0,598	0,000	Berkorelasi
y-x_4	-0,263	0,042	Berkorelasi

Kesimpulan yang dapat diambil dari Tabel 5.1 adalah variabel x_1, x_2, x_3, x_4 berkorelasi dengan variabel y (α yang digunakan 0,05). Korelasi yang terbentuk cukup kuat yang ditunjukkan dengan koefisien korelasi antara 0,263-0,598. Variabel x_2, x_3, dan x_4 memiliki hubungan negatif terhadap variabel y, sedangkan variabel x_1 memiliki hubungan positif. Langkah selanjutnya yang harus dilakukan setelah hubungan antara variabel dependen dan variabel dependen diketahui adalah mengecek hubungan antar variabel independennya yang akan dibahas pada subbab berikut ini.

5.1 Multikolinearitas

Multikolinearitas merupakan sebuah kondisi dimana antara satu pasang variabel independen atau lebih memiliki korelasi yang kuat. Jika antara variabel independen memiliki korelasi yang kuat maka akan muncul masalah dalam analisis regresi linear yang dibentuk (Best & Wolf, 2015), salah satunya masalah serius pada estimasi parameter model regresi linear menggunakan OLS (Montgomery, Peck, & Vining, 2012). Menurut Adeboye, Fagoyinbo, & Olatayo (2014) multikolinearitas berkontribusi dalam peningkatan *standard residual* koefisien regresi, sehingga menyebabkan parameter yang diestimasi kurang efisien dan kurang siginifikan dalam kelas estimator OLS.

Salah satu cara yang dapat digunakan untuk mendeteksi adanya multikolinearitas adalah dengan menguji koefisien korelasi antara dua variabel independen. Akan tetapi cara ini tidak menjamin bahwa salah satu koefisien korelasi antara variabel independen nilainya besar. Pada umumnya deteksi multikolinearitas tidak cukup hanya dengan melihat koefisien korelasi antara dua variabel independen (Montgomery, Peck, & Vining, 2012). Pada subbab ini akan ditunjukkan cara mendeteksi adanya multikolinearitas pada data Ilustrasi 5.1 menggunakan pendekatan uji korelasi dan VIF.

Deteksi multikolinearitas menggunakan uji korelasi

Pembahasan lebih mendalam tentang korelasi sudah dibahas pada Bab 2, sehingga pada bab ini hanya ditunjukkan langkah analisis untuk melakukan uji koefisien korelasi untuk mendeteksi adanya multikolinearitas.

Langkah 1. Membentuk hipotesis uji koefisien korelasi.

H_0: $\rho = 0$.

H_1: $\rho \neq 0$.

Langkah 2. Menentukan statistik uji yang digunakan

$$t_{hitung} = r\sqrt{\frac{n-2}{1-r^2}} \tag{5.1}$$

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{5.2}$$

(Wilcox, 2012).

dimana

t_{hitung} statistik uji
r koefisien korelasi
n jumlah pengamatan
x , y_i variabel independen ke-i, i=1,2,…,n
$\bar{x}$, y rata-rata variabel independen.

Langkah 3. Menentukan nilai α dan daerah kritis

tolak H_0 jika $\left| t_{hitung} \right| > t_{(n-2);(1-\alpha/2)}$ atau

tolak H_0 jika $p_{value} < \alpha$

dimana

$p_{value}=P(T<t_{hitung})$.

Langkah 4. Mengambil kesimpulan.

Berikut ini adalah hasil analisis data Ilustrasi 5.1 menggunakan langkah-langkah analisis yang telah dijelaskan sebelumnya.

Tabel 5.2. Uji Korelasi Antara Variabel Independen Data Ilustrasi 5.1

Variabel independen	x_1	x_2	x_3	x_4
x_1	1	-0,498 (0,000)*	-0,371 (0,003)*	-0,122 (0,353)*
x_2		1	0,183 (0,163)*	0,719 (0,000)*
x_3			1	-0,093 (0,479)*
x_4				1

(….)* menunjukkan nilai $p_{value.}$

Table 5.2 memberikan informasi bahwa variabel x_1 dengan x_2, x_1 dengan x_3, dan x_2 dengan x_4 berkorelasi, kesimpulan ini didapatkan dengan cara membandingkan nilai p_{value} yang dihasilkan dengan α=0,05. Korelasi yang dihasilkan dari ketiga pasang variabel independen sebenarnya tidak terlalu kuat, namun korelasi ini berpengaruh pada estimasi parameter model regresi yang dihasilkan.

Deteksi multikolinearitas menggunakan nilai VIF

Nilai VIF digunakan untuk mengukur sejauh mana variabel independen ke-j bergantung pada kumpulan variabel independen yang lain (Best & Wolf, 2015). Cara ini sama-sama menggunakan koefisien korelasi untuk mendeteksi adanya multikolinearitas, bedanya disini nilai korelasi didapatkan dari hubungan antara satu variabel independen dengan sekelompok variabel independen yang lain. Persamaan VIF dapat dilihat pada persamaan berikut ini

$$VIF_j = \frac{1}{1 - R_j^2} \tag{5.3}$$

dimana
R_j^2 menunjukkan koefisien korelasi kuadrat antara variabel independen x_j dengan variabel independen yang lain.

Nilai VIF berkisar antara 1 sampai +∞, nilai yang tinggi menunjukkan adanya hubungan ketergantungan yang kuat antara satu variabel independen dengan sekelompok variabel independen yang lain. Nilai VIF ≥10 menunjukkan adanya potensi bahaya multikolinearitas dalam analisis regresi yang dibuat. Nilai VIF 10 menunjukkan R_j^2 sebesar 0,9. Hasil analisis nilai VIF data Ilustrasi 5.1 dapat dilihat pada Tabel 5.3.

Tabel 5.3. Nilai VIF

Variabel independen	Nilai VIF
x_1	1,658
x_2	3,305
x_3	1,225
x_4	2,583

Tabel 5.3 menunjukkan nilai VIF dari masing-masing variabel independen. Kesimpulan yang dapat diambil adalah tidak adanya multikolinearitas dalam analisis regresi yang dibentuk, karena nilai VIF yang dihasilkan kurang dari 10.

Hasil pendeteksian multikolinearitas menggunakan uji koefisien korelasi dan VIF menunjukkan hasil yang berbeda. Hasil yang sebaiknya digunakan adalah hasil dari uji koefisien korelasi, karena nilai VIF hanya mampu mendeteksi kondisi multikolinieritas jika koefisien korelasi menunjukkan nilai 0,9. Menurut Yoo, et al., (2014) adanya korelasi antara dua variabel independen atau lebih dapat menyebabkan estimasi parameter model regresi bias. Pada penelitian yang dihasilkannya estimasi parameter model regresi linear yang dihasilkan berubah jika terdapat korelasi antara variabel independennya (berapapun besar nilai koefisien korelasinya).

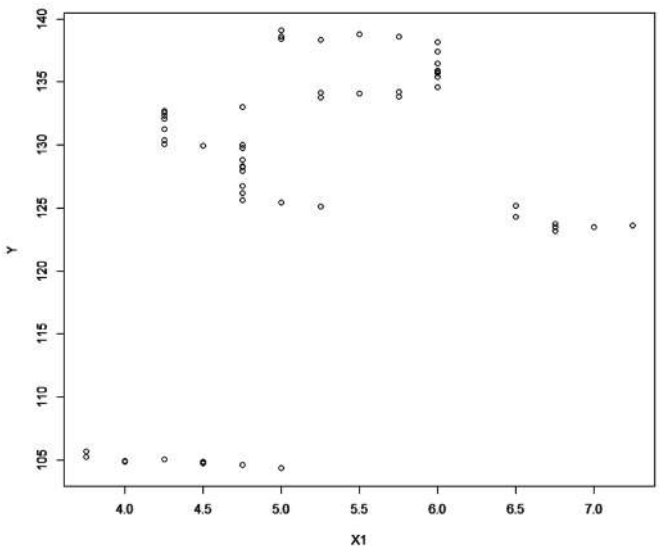
Berdasarkan hasil analisis data Ilustrasi 5.1 dapat disimpulkan terdapat kondisi multikolinearitas, sehingga pemodelan menggunakan regresi linear kurang cocok untuk digunakan. Jika tetap ingin menggunakan model regresi linear untuk analisisnya, maka peneliti harus menambah data yang digunakan. Selain itu terdapat beberapa pemodelan yang cocok untuk data yang memuat kondisi multikolinearitas yaitu dengan menggunakan pemodelan *ridge regression* atau *principal-component regression*

5.2 Linearitas

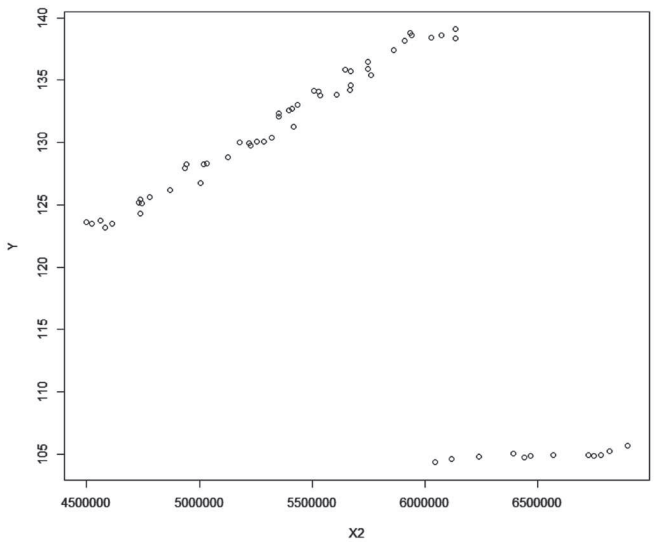
Linearitas yang dimaksudkan disini berkaitan dengan posisi parameter model regresi dan variabel independen yang sejajar

dengan variabel dependen seperti yang digambarkan pada persamaan (2.3). Asumsi linearitas penting untuk dibahas, karena berkaitan dengan estimasi parameter yang dihasilkan dalam model regresi linear. Nilai ekspektasi parameter model regresi akan sama dengan nilai dalam populasinya apabila asumsi ini terpenuhi. Jika asumsi ini dilanggar maka hasil estimasi parameter model regresi yang dihasilkan dapat bias (Best & Wolf, 2015).

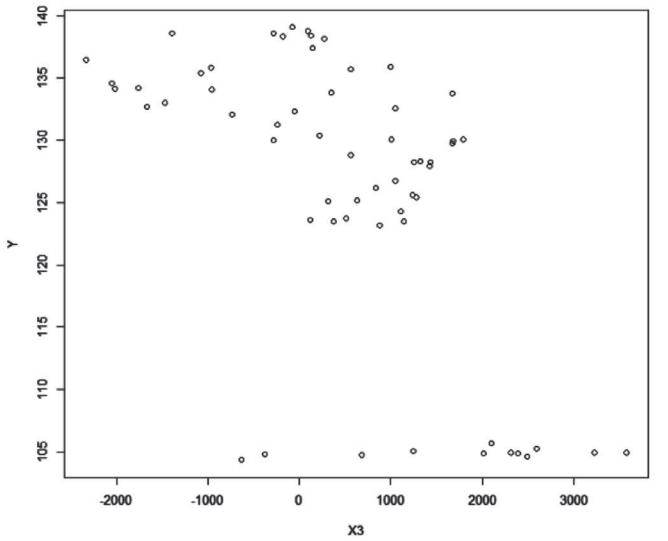
Linearitas dapat dideteksi dari *residual* yang dihasilkan model regresi. Jika *residual* yang dihasilkan 0, maka dapat disimpukan terdapat hubungan linear yang sempurna antara variabel dependen dengan variabel independennya. Namun dikondisi ini jarang ditemui pada penelitian di dunia nyata. Linearitas juga dapat dilihat dari koefisien korelasi, menurut Johnson & Bhattacharyya (2019) koefisien korelasi menunjukkan kekuatan hubungan linear antara variabel x dan y. Selain itu linearitas juga dapat dideteksi menggunakan grafik dan diuji menggunakan *lack-of-fit test* (Best & Wolf, 2015). Pada subbab ini akan ditunjukkan pendeteksian asumsi linearitas pada model regresi menggunakan grafik *Scatterplot* berdasarkan penjelasan yang disampaikan oleh Johnson & Bhattacharyya (2019).



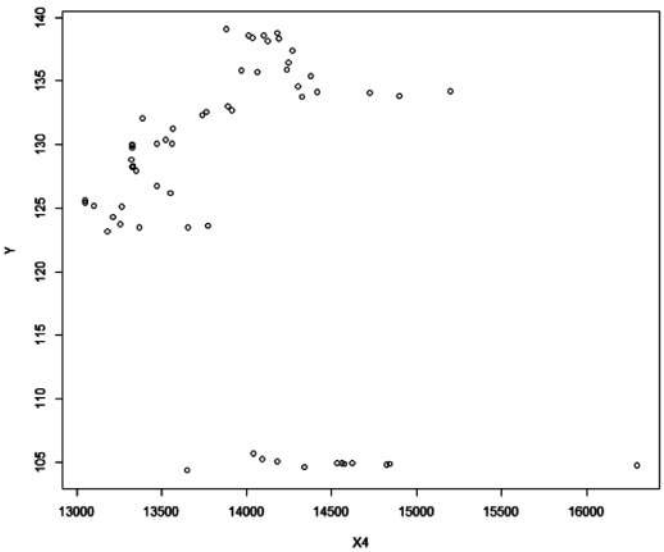
(a) Scatterplot x_1 dengan y



(b) scatterplot x_2 dengan y



(c) Scatterplot x_3 dengan y



(d) *Scatterplot* x_4 dengan y

Gambar 5.1. *Scatterplot* variabel independen dengan variabel dependen

Scatterplot merupakan grafik dua dimensi yang menunjukkan hubungan antara dua variabel. Grafik ini dapat digunakan untuk mendeteksi apakah terdapat hubungan linear antara 2 variabel. Pendeteksian atau pengujian asumsi linearitas dapat dilakukan diawal, yaitu sebelum, sesudah, atau bersamaan dengan pengujian korelasi. *Scatterplot* data Ilustrasi 5.1 dapat dilihat pada Gambar 5.1, terdapat 4 gambar yang disajikan, setiap gambar menunjukkan grafik *scatterplot* antara variabel independen dengan variabel dependen. Sumbu x pada *scatterplot* menunjukkan variabel independen dan sumbu y menunjukkan variabel dependen. Berdasarkan Gambar 5.1 didapatkan informasi bahwa tidak terdapat pola linear antara variabel x_1 dengan y, x_3 dengan y, dan x_4 dengan y, sedangan antara variabel x_2 dengan y pola yang terbentuk hampir linear.

Jika antara variabel pada data yang digunakan tidak memiliki pola hubungan linear, maka sebaiknya analisis regresi yang digunakan menyesuaikan pola hubungan yang terbentuk. Misal-

nya menggunakan model regresi polinomial. Dalam beberapa situasi cara lain yang dapat dilakukan untuk memodelkan data yang tidak berpola linear adalah dengan melakukan transformasi. Variabel dependen atau variabel independen ditransformasi sehingga pola hubungan yang baru terbentuk hampir linear. Pada beberapa situasi hubungan nonlinear yang spesifik sangat disarankan berdasarkan pertimbangan teoritis. Tabel 5.4 berikut ini.

Tabel 5.4. Model nonlinear dan Transformasi Linear

Model non linear	Transformasi		Model Transformasi	
			$y' = \beta_0 + \beta_1 x'$	
a. $y=ae^{bx}$	$y'=\log_e y$	$x' = x$	$\beta_0 = \log_e a$	$\beta_1 = b$
b. $y=ax^b$	$y'=\log y$	$x' = \log x$	$\beta_0 = \log a$	$\beta_1 = b$
c. $y = \frac{1}{a+bx}$	$y' = \frac{1}{y}$	$x' = x$	$\beta_0 = a$	$\beta_1 = b$
d. $y = a + b\sqrt{x}$	$y'= y$	$x' = \sqrt{x}$	$\beta_0 = a$	$\beta_1 = b$

(Johnson & Bhattacharyya, 2019).

5.3 Residual identik, independen, dan berdistribusi normal

Dalam pembahasan pendeteksian dan pengujian asumsi *residual* identik, independen, dan berdistribusi normal digunakan data Ilustrasi 5.1 dengan asumsi model yang digunakan adalah model regresi linear dengan x_1, x_2, x_3, x_4 dimasukkan dalam model.

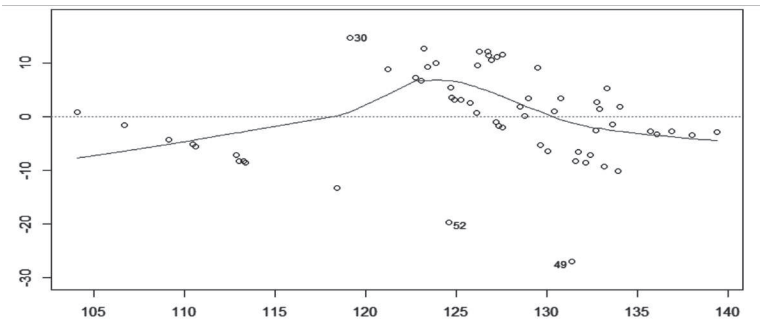
5.3.1 Identik (*homoscedasticity*)

Asumsi ini menggambarkan varians *residual* konstan untuk setiap kovarians X, $Var(\varepsilon_i \mid x_{i1},\ldots,x_{ik})=\sigma^2$. Artinya, jika *residual* tidak berkorelasi, maka matrik varians kovarians *residual* dapat ditulis seperti ini $Var(\varepsilon_i)=\sigma^2 I$ (Best & Wolf, 2015). Asumsi ini digunakan untuk menentukan varian distribusi parameter model regresi menggunakan estimasi OLS. Jika asumsi ini dilanggar

maka estimasi parameter yang dihasilkan tidak bias tetapi tidak efisien. Terdapat dua cara untuk menentukan apakah *residual* model regresi yang dihasilkan konstan atau tidak, yaitu dengan cara pendeteksian atau pengujian. Pendeteksian asumsi lebih mudah diaplikasikan karena hanya menggunakan grafik, sedangkan pengujian lebih rumit diaplikasikan karena harus menggunakan urutan pengujian hipotesis statistik.

Pendeteksian *residual* identik

Pendeteksian *residual* identik dapat dilakukan dengan cara membuat grafik dua dimensi antara *residual* dengan prediksi variabel dependen yang dihasilkan. Jika *residual* memiliki pola acak disekitar rata-rata 0 maka dapat dikatakan varians *residual* identik.



Gambar 5.2. Grafik *Residual* vs Prediksi Variabel Dependen Data Ilustrasi 5.1

Sumbu x pada Gambar 5.2 menunjukkan prediksi variabel dependen dan sumbu y menunjukkan *residual* hasil pemodelan regresi data Ilustrasi 5.1. Pola residual yang dihasilkan pada Gambar 5.2 belum acak disekitar nilai 0 dan bergerombol di beberapa titik, sehingga dapat dikatakan bahwa varians *residual* tidak identik.

Pengujian *residual* identik

Pendeteksian menggunakan grafik akan memberikan hasil yang kurang objektif, sehingga dalam menguji asumsi *residual* model regresi sebaiknya menggunakan sebuah pengujian.

Pengujian *residual* identik dapat dilakukan menggunakan beberapa metode yaitu *goldfeld-Quandt test, breusch-pagan-godfrey test, white test*, dsb. Pada subbab ini akan ditunjukkan cara pengujian asumsi *residual* identik menggunakan *white test*, karena lebih mudah diaplikasikan dan tidak terlalu sensitif terhadap asumsi normal (Gujarati, 2004).

Langkah 1. Membentuk hipotesis uji *residual* identik.

H_0: $E(\varepsilon_i \mid x_{i1,}\ldots,x_{ik})=\sigma^2$.

H_1: $E(\varepsilon_i \mid x_{i1},\ldots,x_{ik})\neq\sigma^2$.

Langkah 2. Menentukan statistik uji yang digunakan

$$\mathcal{X}^2_{hitung} = nR^2_{\hat{\varepsilon}^2} \tag{5.4}$$

dimana

$\mathcal{X}^2_{hitung}$ statistik uji

n jumlah pengamatan

$R^2_{\hat{\varepsilon}^2}$ pemodelan *residual*.

Berikut ini adalah langkah-langkah *white test* untuk mendapatkan $R^2_{\hat{\varepsilon}^2}$.

1. Mendapatkan nilai *residual* $\widehat{\varepsilon_i}$ berdasarkan persamaan berikut ini.

   $$y_i=\beta_0+\beta_1 x_{1i}+\beta_2 x_{2i}+....+\beta_k x_{ki}+\varepsilon_i \tag{5.5}$$

2. Membuat pemodelan regresi *(auxiliary)*, jika diaplikasikan pada persoalan data Ilustrasi 5.1 maka didapatkan persamaan.

   $$\begin{aligned}\hat{u}_i{}^2 = {} & \alpha_0 + \alpha_1 x_{1i} + \cdots + \alpha_4 x_{4i} + \alpha_5 x_{1i}^2 + \cdots + \alpha_8 x_{4i}^2 \\ & + \alpha_9 x_{1i} x_{2i} + \cdots + \alpha_{14} x_3 x_4 \\ & + v_i\end{aligned} \tag{5.6}$$

   Nilai $R^2_{\hat{\varepsilon}^2}$ didapatkan pada tahapan ini.

Langkah 3. Menentukan nilai α dan daerah kritis

tolak H_0 jika $\mathcal{X}^2_{hitung} > \mathcal{X}^2_{k,\alpha}$ dimana k menunjukkan jumlah variabel independen yang digunakan dalam pemodelan regresi (*auxiliary)*.

Langkah 4. Mengambil kesimpulan.

Berdasarkan data Ilustrasi 5.1 yang diuji menggunakan *white test* didapatkan nilai $R^2_{\hat{\varepsilon}^2}$ sebesar 0,462. Nilai $\mathcal{X}^2_{hitung} = 60\text{x}0{,}462 = 27{,}72$. Jika digunakan α sebesar 0,05 maka didapatkan nilai $\mathcal{X}^2_{(14;0,05)}$ sebesar 23,685. Berdasarkan nilai $\mathcal{X}^2_{hitung}$ dan $\mathcal{X}^2_{(14;0,05)}$ dapat disimpulkan bahwa varians *residual* tidak identik atau dalam pemodelan regresi data Ilustrasi 5.1 terdapat *heteroscedasticity*. Peneliti harus berhati-hati dalam menggunakan *white test* karena disini yang digunakan nilai R^2 yang sangat dipengaruhi oleh jumlah variabel independen yang digunakan. Jika terdapat kasus seperti ini yaitu adanya *heteroscedasticity*, maka model regresi yang dibentuk dapat diperbaiki dengan cara melakukan transformasi pada variabel dependen yang digunakan. Cara yang lain adalah menggunakan estimasi *weighted least squares* (WLS) dalam OLS. Estimasi WLS dapat digunakan ketika σ_i^2 diketahui.

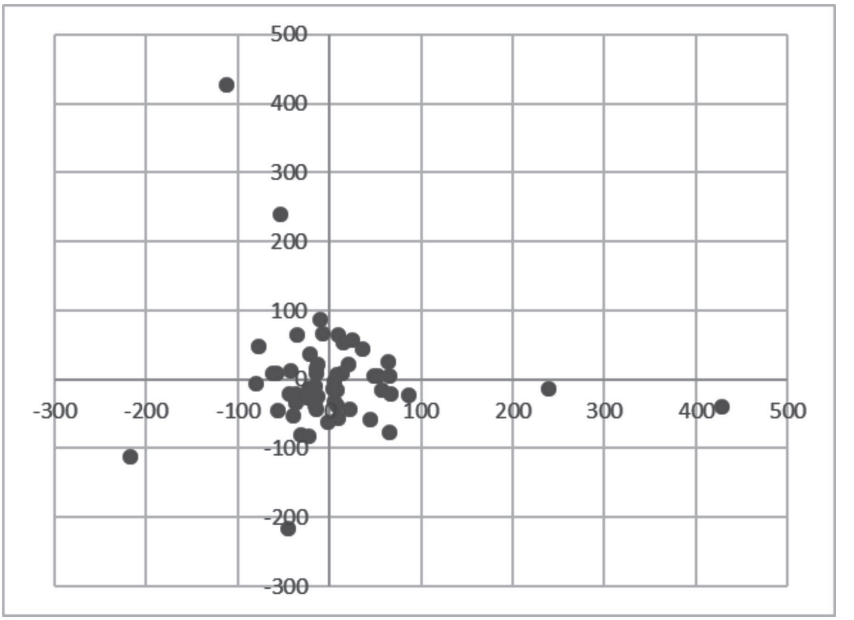
5.3.2 Independen

Residual yang dihasilkan dari model regresi diharapkan tidak berpola, yang artinya residual antar pengamatan tidak saling berkorelasi. Dalam dunia nyata, sering kali dijumpai pemodelan regresi yang tidak memenuhi asumsi ini. Salah satu penyebabnya adalah pemilihan data yang kurang tepat untuk pemodelan regresi. Penggunakan data *time series* dan data pengelompokan pengamatan akan mempengaruhi korelasi antar *residual* dalam model regresi yang dibentuk. Pelanggaran asumsi residual tidak independen merupakan masalah yang sangat serius karena dapat berpengaruh dalam pengambilan keputusan estimasi parameter model regresi. Terdapat beberapa cara yang dapat dilakukan untuk mengecek apakah *residual* yang dihasilkan

sudah independen atau belum, yaitu dengan pendeteksian dan pengujian sama seperti pengecekan asumsi *residual* identik.

Pendeteksian *residual* independen.

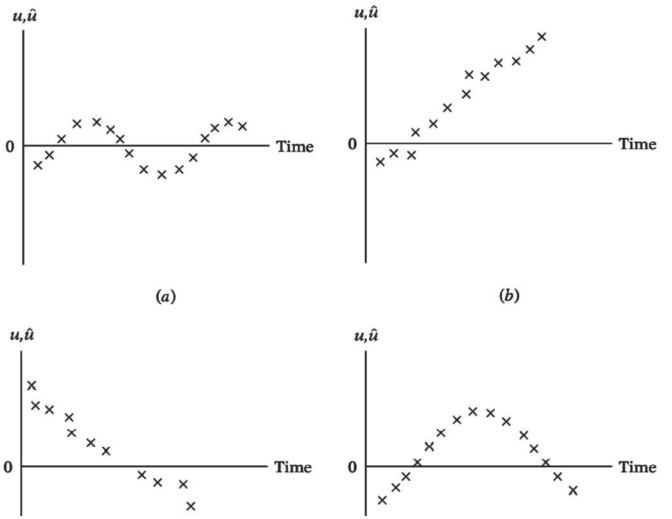
Pada subab ini akan ditunjukkan salah satu grafik *residual* yang dapat digunakan untuk mendeteksi apakah terdapat korelasi antar residual pada model regresi yang dibentuk. Grafik yang digunakan merupakan grafik dua dimensi dengan sumbu x menunjukkan *residual* pada pengamatan ke (t-1) dan sumbu y menunjukkan *residual* pada pengamatan ke t. Gambar 5.3 berikut ini menunjukkan grafik *residual* ke- t-1 dengan *residual* ke t data Ilustrasi 5.1.



Gambar 5.3. Grafik *Residual* ke t-1 dengan *Residual* ke t

Residual tidak indepenen jika grafik yang dihasilkan berpola tidak acak atau membentuk pola tertentu, misalkan pola garis lurus atau yang lain. Berdasarkan hasil pada Gambar 5.3 dapat diperoleh informasi bahwa *residual* yang dihasilkan dari pemodelan regresi data Ilustrasi 5.1 tidak independen, karena titik-titik *residual* yang digambarkan mengelompok di tengah

perpotongan sumbu koordinat. Gambar 5.4 berikut ini menunjukkan contoh pola korelasi antar *residual* yang lain. Grafik yang terdapat pada Gambar 5.4 merupakan grafik antara taksiran *residual* dengan pengamatannya. Jika dilihat dengan seksama, *residual* yang digambarkan memiliki pola tertentu seperti ada yang bergelombang dan berupa garis lurus. Korelasi positif antar *residual* dapat dilihat pada Gambar 5.4.b, sedangkan korelasi negatif dapat dilihat pada Gambar 5.4.c.



(Gujarati, 2004).

Gambar 5.4. Pola *Residual* yang Berkorelasi

Pengujian *residual* independen.

Terdapat beberapa jenis alat uji *residual* independen yang dapat digunakan, yaitu *run test*, *durbin Watson test, Breusch-Godfrey (BG) test*, dsb. Pada subbab ini akan ditunjukkan cara penggunaan BG *test* untuk menguji apakah *residual* sudah independen atau belum. Menurut Gujarati (2004) penggunaan BG *test* dapat menghindari beberapa jebakan dalam *durbin-watson test* dan memungkinkan digunakan untuk variabel independen yang tidak stokastik. BG *test* juga dikenal dengan LM *test*. Berikut ini adalah langkah analisis untuk BG *test*.

Langkah 1. Membentuk hipotesis uji *residual* independen.
Bentuk umum: H_0: $\rho_1=\rho_2=....=\rho_p=0$.

p menunjukkan jumlah lag waktu/pengamatan yang akan diuji. Sebagai ilustrasi untuk pengujian data Ilustrasi 5.1 digunakan nilai p sebesar 2, sehingga dapat dirumuskan hipotesis awalnya sebagai berikut ini.

$H_0: \rho_1 = \rho_2 = 0$.

$H_1: \rho_1 = \rho_2 \neq 0$.

Langkah 2. Menentukan statistik uji yang digunakan

$$\mathcal{X}^2_{hitung} = (n-p)R^2 \qquad (5.7)$$

dimana

$\mathcal{X}^2_{hitung}$ statistik uji

n jumlah pengamatan.

Berikut ini adalah langkah-langkah *white test* untuk mendapatkan R^2.

1. Mendapatkan nilai *residual* $\hat{\varepsilon_i}$ berdasarkan persamaan berikut ini.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i \qquad (5.8)$$

2. Membuat pemodelan regresi *(auxiliary,* jika diaplikasikan pada persoalan data Ilustrasi 5.1 maka didapatkan persamaan.

$$\hat{u}_i = \alpha_0 + \alpha_1 X_{1i} + \cdots + \alpha_4 X_{4i} + \hat{\rho}_1 \hat{\varepsilon}_{i-1} + \cdots + \hat{\rho}_p \hat{\varepsilon}_{i-p} + u_i \qquad (5.9)$$

Nilai R^2 didapatkan pada tahapan ini, p menunjukkan jumlah lag yang akan diuji.

Langkah 3. Menentukan nilai α dan daerah kritis

tolak H_0 jika $\mathcal{X}^2_{hitung} > \mathcal{X}^2_{p;\alpha}$ dimana p menunjukkan jumlah lag yang diinginkan

atau

tolak H_0 jika $p_{value}<\alpha$ dimana $p_{value}=1-P(X^2<\mathcal{X}^2_{hitung})$.

Langkah 4. Mengambil kesimpulan.

Tabel 5.5 berikut ini menyajikan hasil pengujian *residual* independen menggunakan BG *test*.

Tabel 5.5. Hasil BG *test*

$\mathcal{X}^2_{hitung}$	P	p_{value}
30,186	2	$2{,}787\text{x}10^{-7}$

Jika digunakan α=0,05 maka didapatkan $\mathcal{X}^2_{2;0,05}$ sebesar 5,991. Kesimpulan yang dapat diambil dari Tabel 5.5 adalah *residual* hasil pemodelan regresi data Ilustrasi 5.1 tidak independen, karena nilai $\mathcal{X}^2_{hitung}>\mathcal{X}^2_{2;0,05}$ dan p_{value} yang dihasilkan kurang dari 0,05. Jenis data yang digunakan merupakan salah satu penyebab asumsi ini tidak terpenuhi. Data Ilustrasi 5.1 merupakan data *time series*, sehingga resiko terjadinya *residual* tidak independen tinggi. Cara mengatasi pelanggaran adalah dengan membuat pemodelan yang melibatkan data *time series*, misalkan dengan menggunakan model lag terdistribusi atau sejenisnya.

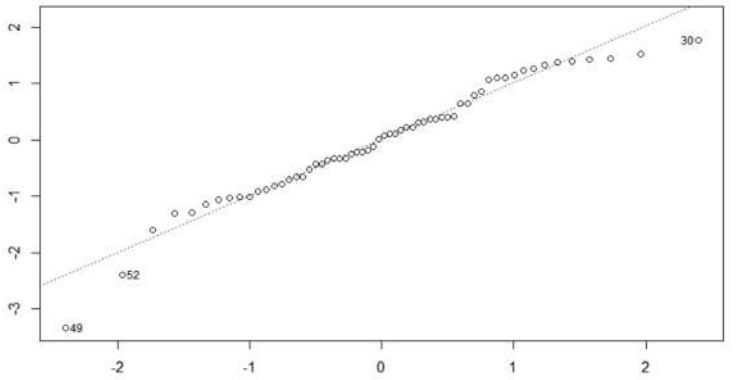
5.3.3 Berdistribusi Normal

Asumsi *residual* berdistribusi normal digunakan untuk menentukan bentuk fungsi distribusi sampling dari estimasi parameter regresi. Dalam penelitian ilmu sosial asumsi ini sering dilanggar. Menurut Gujarati (2004) pelanggaran asumsi ini tidak menyebabkan estimasi parameter tidak BLUE. Selain itu uji statistik estimasi parameter model regresi cukup kuat terhadap pelanggaran asumsi *residual* berdistribusi normal. Jika ukuran sampel cukup besar, estimasi parameter regresi akan mendekati distribusi t (Best & Wolf, 2015). Dalam pemodelan regresi linear pengujian asumsi *residual* berdistribusi normal tidak bermakna sama dengan menguji distribusi normal pada variabel dependennya. Pada subbab ini akan ditunjukkan cara

mendeteksi dan menguji asumsi *residual* berdistribusi normal dari pemodelan regresi data Ilustrasi 5.1.

Pendeteksian asumsi *residual* berdistribusi normal.

Pendeteksian *residual* berdistribusi normal dapat menggunakan *normal quantiles-quantiles plot* (Q-Q plot) atau plot probabilitas normal. Grafik yang terbentuk merupakan grafik 2 dimensi dengan sumbu x dan sumbu y merupakan nilai *quantiles residual*. Jika titik-titik pengamatan dalam Q-Q *plot* menunjukkan garis lusur maka dapat dikatakan *residual* berdistribusi normal.



Gambar 5.5. Q-Q *Plot Residual* Data Ilustrasi 5.1

Titik-titik *quantiles* pada grafik di Gambar 5.5 menunjukkan pola garis lurus. Terdapat beberapa titik (3) yang jauh dari sekumpulan titik yang membentuk garis lurus. Berdasarkan Gambar 5.5 dapat diperoleh informasi bahwa *residual* pemodelan regresi data Ilustrasi 5.1 berdistribusi normal.

Pengujian asumsi *residual* berdistribusi normal.

Terdapat beberapa statistik uji yang dapat digunakan untuk menguji asumsi *residual* berdistribusi normal, yaitu *Shapiro-Wilks* (SW), *Kolmogorof-Smirnov* (KS), *Liliefors* (LF), *Anderson-Darling* (AD), dsb. Pada penelitian ini digunakan *Shapiro-Wilks* untuk menguji asumsi *residual* berdistribusi normal. Menurut Razali & Wah (2011) statistik uji SW merupakan statistik uji paling kuat di-

bandingkan KS, LF, dan AD. Akan tetapi kekuatan ujinya masih rendah untuk ukuran sampel kecil. Royston (1982) mengatakan bahwa SW menyediakan uji normalitas paling baik tetapi masih terbatas pada sampel 3 sampai 50 dan diperpanjang sampai 2000. Jadi jika ingin menggunakan uji statistik SW ukuran sampel minimumnya adalah 3. Berikut ini adalah langkah analisis untuk melakukan uji *residual* berdistribusi normal menggunakan statistik uji SW.

Langkah 1. Membentuk hipotesis uji *residual* berdistribusi normal.

H_0: *Residual* berdistribusi normal.

H_1: *Residual* tidak berdistribusi normal.

Langkah 2. Menentukan statistik uji yang digunakan

$$W = \frac{b^2}{((n-1)s^2)} \tag{5.10}$$

$$b = \sum_{i=1}^{n/2} a_{n-i+1}(x_{n-i+1} - x_i) \tag{5.11}$$

dimana

W statistik uji SW

s^2 varians *residual*

a koefisien SW

n jumlah pengamatan (Thode, 2002).

Langkah 3. Menentukan nilai α dan daerah kritis

tolak H_0 jika p_{value}<α.

Langkah 4. Mengambil kesimpulan.

Berdasarkan data Ilustrasi 5.1 dengan menggunakan statistik uji SW didapatkan W=0,963 dengan p_{value} 0,066. Jika digunakan α=0,05 maka dapat disimpulkan bahwa *residual* pemodelan regresi data Ilustrasi 5.1 berdistribusi normal. Jika *residual* pemodelan regresi yang dihasilkan tidak berdistribusi normal,

maka dapat digunakan model yang dapat menyesuaikan pola data seperti menggunakan *generalized linear model* (GLM). Selain itu pelanggaran asumsi ini dapat diperbaiki dengan melakukan transformasi pada variabel dependennya.

Analisis pengecekan asumsi *residual* sangat penting dilakukan supaya model regresi yang dihasilkan dapat menjadi alat untuk pengambilan keputusan yang tepat. Jika salah satu atau lebih asumsi tidak terpenuhi, baiknya pelanggaran-pelanggaran tersebut diperbaiki atau dicari pemodelan lain yang sesuai dengan kondisi data yang digunakan. Berdasarkan data Ilustrasi 6.1 dapat kita rangkum bahwa semua variabel independen berpengaruh terhadap variabel dependennya; terdapat 3 pasang variabel independen yang saling berkorelasi; hanya ada 1 variabel independen yang memiliki hubungan linear dengan variabel dependennya; *residual* yang dihasilkan tidak identik dan independen tetapi berdistribusi normal. Jika pemodelan regresi tetap dipaksakan pada data Ilustrasi 5.1, maka pemodelan yang dihasilkan kurang bagus berdasarkan kriteria kebaikan model regresi dan pastinya estimasi parameter regresi yang dihasilkan tidak BLUE.

BAB 6.
SPESIFIKASI MODEL DAN UJI DIAGNOSTIK

PADA bab ini akan dibahas mengenai penentuan model regresi linear yang paling baik dengan cara melihat signifikansi estimasi parameter regresi secara serentak dan parsial. Selain itu juga akan dibahas mengenai metode pemilihan variabel independen (seleksi variabel) yang berpengaruh dalam penentuan model regresi linear terbaik.

6.1 Seleksi Variabel

Berikut ini adalah strategi atau langkah-langkah yang dapat dilakukan untuk seleksi variabel dan pembentukan model regresi linear.

1. Bentuk model regresi linear menggunakan seluruh variabel independen.
2. Lakukan analisis menyeluruh terhadap model yang dibentuk pada nomer 1, mulai dari uji serentak, uji multikolinearitas, uji linearitas, uji asumsi *residual* identik, independen, dan berdistribusi normal.
3. Tentukan apakah perlu dilakukan transformasi terhadap variabel dependen atau independen yang digunakan.
4. Tentukan apakah semua variabel independen yang digunakan *feasible*. Gunakan uji t untuk mengevaluasi model secara parsial. Jika tidak *feasible* maka lakukan seleksi variabel.
5. Lakukan analisis menyeluruh pada model yang telah dievaluasi atau model yang baru, terutama pada analisis residualnya untuk menentukan kecukupan model.

Secara umum dalam pemodelan regresi linear diperlukan sebuah seleksi variabel untuk menentukan variabel independen mana saja yang memiliki pengaruh signifikan terhadap variabel dependennya. Seleksi variabel memiliki banyak keunggulan diantaranya adalah

1. seleksi variabel dapat meningkatkan presisi estimasi parameter dari variabel independen yang dipertahankan (Montgomery, Peck, & Vining, 2012)
2. investigasi kecukupan model regresi memiliki keterkaitan dengan seleksi variabel.

Terdapat 2 aspek dalam seleksi variabel yang harus dilakukan, yaitu menghasilkan model subset dan memutuskan model mana yang terbaik. Terdapat beberapa kriteria untuk mengevaluasi dan membandingkan model regresi subset, yaitu R^2, *adjusted* R^2, *Residual Mean Square* (MS_{Res}), *Mallows's* C_p *Statistic, Akaike Information Criterion (AIC), Bayesian Analogues (BICs)*. Pemilihan kritesia ini berkaitan dengan tujuan penggunaan model regresi linear yang dibentuk. Terdapat beberapa kemungkinan tujuan penggunaan model regresi linear, yaitu

1. deskripsi data
2. prediksi dan estimasi
3. estimasi parameter
4. kontrol, dsb (Montgomery, Peck, & Vining, 2012).

Selain menggunakan kriteria di atas, pengalaman (penelitian sebelumnya), penilaian profesional dalam subjek bidang materi, dan pertimbangan subjektif juga harus digunakan dalam seleksi variabel. Terdapat dua acara yang dapat digunakan untuk mendapatkan model regresi subset dalam seleksi variabel, yaitu:

1. *All possible regression*
2. *Stepwise-type procedures*.

Pada subbab ini akan dibahas mengenai *stepwise-type procedures* karena jika menggunakan *all possible regression* akan memberatkan proses komputasi karena terdapat banyak model

regresi subset yang dihasilkan berdasarkan jumlah variabel independen yang dihasilkan.

Dalam *stepwise-type procedures* terdapat 3 metode seleksi variabel yang dapat digunakan, yaitu *backward elimination, forward selection,* dan *stepwise method*. *Forward selection* dimulai dengan asumsi bahwa dalam model tidak ada variabel independen yang dimasukkan kecuali *intercept*, kemudian variabel independen yang digunakan dimasukkan satu persatu dengan memperhatikan nilai korelasi yang dibentuk antara variabel independen dengan dependennya. *Backward elimination* merupakan kebalikan dari metode sebelumnya, asumsi awalnya adalah semua variabel independen dimasukkan semua dalam model, kemudian dikeluarkan satu persatu variabel independen yang tidak memiliki pengaruh signifikan terhadap model. Sedangkan s*tepwise method* merupakan modifikasi dari *forward selection* dan *backward elimination.* Ilustrasi terkait seleksi variabel dapat dilihat pada subbab 6.3.

6.1.1 Kriteria Pemilihan Model

Uji signifikansi pada model regresi linear merupakan uji yang digunakan untuk menentukan adanya hubungan linear antara variabel independen dengan variabel dependen (Montgomery, Peck, & Vining, 2012). Prosedur ini biasanya disebut dengan uji keseleruhan dari kecukupan model atau biasanya disebut dengan uji serentak. Secara umum dalam pemodelan regresi terdapat dua pengujian untuk menentukan model terbaiknya, yaitu pengujian serentak dan pengujian parsial. Pengujian serentak menggunakan uji F yang biasanya dirangkum dalam tabel A*nalysis Of Variance* (ANOVA) dan pengujian parsial menggunakan uji t. Dalam pemilihan model terbaik selain menggunakan pengujian estimasi parameter koefisien regresi juga digunakan nilai determinasi.

Pengujian Serentak.

Hipotesis yang digunakan dalam pengujian serentak adalah

H_0 : $\beta_0=\beta_1= .... =\beta_k=0$

H_0 : $\beta_0=\beta_1$, paling tidak ada satu j.

Jika hasil yang diperoleh adalah tolak H_0, maka terdapat minimal satu variabel independen yang bermakna atau berkontribusi secara signifikan terhadap model regresi linear yang dibentuk. Nilai-nilai yang digunakan dalam pengujian serentak dirangkum dalam tabel ANOVA seperti yang disajikan pada Tabel 6.1 berikut ini.

Tabel 6.1 ANOVA

Source of Variation (Sumber Variasi)	*Sum of Squares* (Jumlah Kuadrat)	*Degrees of Freedom* (Derajat Bebas)	*Mean Square*	F_{hitung}
Regression (Regresi)	SS_R	*K*	MS_R	MS_R/MS_{Res}
Residual	SS_{Res}	*n-k-1*	MS_{Res}	
Total	SS_T	n-1		

$$SS_R = \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$SS_{Res} = \boldsymbol{y}'\boldsymbol{y} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}$$

$$SS_T = \boldsymbol{y}'\boldsymbol{y} - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$$MS_R = \frac{SS_R}{k}$$

$$MS_{Res} = \frac{SS_{Res}}{n-k-1}$$

$$F_{hitung} = \frac{MS_R}{MS_{Res}} \qquad (6.1)$$

dimana

n jumlah pengamatan

k jumlah variabel independen (Montgomery, Peck, & Vining, 2012).

Pengujian Parsial.

Jika hasil uji serentak menyatakan tolak H_0 maka tahapan selanjutnya yang harus dilakukan adalah melakukan uji parsial. Uji ini digunakan untuk mengecek variabel independen mana yang berpengaruh signifikan terhadap variabel dependennya. Penambahan variabel independen akan meningkatkan varian nilai taksiran dari variabel dependen, sehingga peneliti harus hati-hati dalam memasukkan variabel independen dalam model. Hanya variabel independen yang memiliki nilai nyata dalam menjelaskan variabel dependen yang harus dimasukkan. Jika variabel independen yang tidak penting masuk dalam model maka akan meningkatkan MS_{Res} nya.

Hipotesis yang digunakan dalam pengujian parsial adalah

H_0: $\beta_j=0$

H_0: $\beta_j\neq0$, paling tidak ada satu j.

Jika H_0 gagal ditolak maka variabel independen x_j harus dihapus atau dihilangkan dari model regresi linear yang dibentuk. Statistik uji yang digunakan dalam uji parsial adalah uji t. Persamaannya yang digunakan adalah

$$t_{hitung} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \qquad (6.2)$$

tolak H_0 jika $|t_{hitung}|>t_{(\alpha/2;n-k-1)}$ atau tolak H_0 jika $p_{value}<\alpha$.

Koefisien Determinasi.

Dua cara lain yang dapat digunakan untuk menilai kecukupan model adalah dengan melihat nilai R^2 dan *adjusted* R^2 (R^2_{Adj}) (Montgomery, Peck, & Vining, 2012). Pada subbab ini akan dibahas (R^2_{Adj}) untuk menilai kecukupan model karena interpretasinya yang lebih mudah dibandingkan R^2. Rumus (R^2_{Adj}) adalah.

$$R^2_{Adj} = 1 - \frac{SS_{Res}/(n-p)}{SS_T/(n-1)} \quad (6.3)$$

Nilai R^2_{Adj} meningkat jika jumlah variabel independen yang digunakan juga meningkat. Peningkatan jumlah variabel independen yang digunakan akan mengurangi total variabilitas. Jika sebuah model subset memiliki nilai R^2_{Adj} paling tinggi, maka model ini yang dipilih sebagai model terbaik dengan catatan semua asumsi regresinya terpenuhi

6.3 Pengaplikasian Seleksi Variabel dan Kriteria Pemilihan Model Berdasarkan Data Ilustrasi 6.1.

Subbab seleksi variabel dan kriteria pemilihan model saling mendukung untuk menentukan model terbaik dalam analisis regresi linear. Pada subbab ini akan diilustrasikan cara mendapatkan model regresi terbaik berdasarkan data Ilustrasi 5.1. R^2_{Adj} digunakan sebagai salah satu kriteria pemilihan model. Model yang terbaik adalah model yang uji serentak dan parsialnya signifikan, asumsi *residual* terpenuhi, dan R^2_{Adj} tinggi. Terdapat 4 ilustrasi model regresi subset yang akan ditunjukkan disini, yaitu model dengan semua variabel independen masuk dalam model, model hasil *forward selection*, model hasil *backward elimination*, dan model hasil *stepwise method*. Langkah-langkah dalam menentukan model terbaik adalah.

1. Menentukan model regresi *subset* (dapat menggunakan seleksi variabel)
2. Melihat kecukupan model berdasarkan uji serentak, parsial, dan nilai R^2_{Adj}.
3. Melakukan uji asumsi *residual*.
4. Membuat rangkuman hasil 1-3 berdasarkan hasil setiap model regresi subset.
5. Menentukan model terbaik.

Model regresi subset dengan x_1, x_2, x_3, x_4 dalam model regresi (Model regresi subset 1).

Jika data Ilustrasi 5.1 dimodelkan menggunakan regresi linear dengan semua variabel independen dimasukkan dalam model maka didapatkan ANOVA seperti pada Tabel 6.2 berikut ini.

Tabel 6.2 ANOVA Model Regresi Subset 1

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	F_{hitung}
Regression	3810,373	4	952,593	12,945
Residual	4047,200	55	73,585	
Total	7857,573	59		

Statistik uji yang digunakan adalah F, untuk mengambil kesimpulan F yang didapatkan dari perhitungan dibandingkan dengan $F_{(\alpha;k;(n-k-1)}$. Daerah kritis pengujian hipotesisnya adalah tolah H_0 jika $F_{hitung} > F_{(\alpha;k;(n-k-1).}$ Jika digunakan α sebesar 0,05 maka didapatkan $F_{(0,05;4;55)}$ sebesar 2,539. Kesimpulan yang dapat diambil dari ANOVA adalah tolak H_0, artinya terdapat minimal satu estimasi parameter koefisien regresi yang signifikan berpengaruh dalam model regresi linear yang dihasilkan. Nilai R^2_{Adj} yang dihasilkan dari model ini adalah 0,448. Hasil uji parsial dapat dilihat pada Tabel 6.3 berikut ini.

Tabel 6.3 Uji Parsial Model Regresi Subset 1

Simbol	Estimasi Parameter	$se(\hat{\beta} j)$	t_{hitung}	p_{value}
Intercept	188,6	29,440	6,406	$3{,}530 \times 10^{-8}$
x_1	1,248	1,621	0,770	0,445
x_2	$-2{,}473 \times 10^{-6}$	$3{,}106 \times 10^{-6}$	-0,796	0,429
x_3	$-4{,}886 \times 10^{-3}$	$9{,}418 \times 10^{-4}$	-5,186	$3{,}15 \times 10^{-6}$
x_4	$-3{,}781 \times 10^{-3}$	$2{,}899 \times 10^{-3}$	1,304	0,198

Berdasarkan hasil pada Tabel 6.3 didapatkan informasi bahwa estimasi parameter yang signifikan berpengaruh pada model regresi linear yang dibentuk adalah *intercept* dan $x_{3.}$ Adanya variabel independen yang tidak bermakna pada model akan menyebabkan nilai SS_{Res} tinggi.

Setelah dilakukan uji serentak dan parsial, selanjutnya dapat dilakukan pendeteksian atau pengujian asumsi *residual*. Tabel 6.4 berikut ini menyajikan rangkuman pengujian asumsi *residual* model regresi subset 1 yang telah dilakukan pada bab sebelumnya.

Tabel 6.4. Pengujian Asumsi *Residual* Model Regresi Subset 1

Asumsi Residual	Statistik Uji	Nilai Tabel	p_{value}
Identik	χ^2_{hitung} =27,72	$\chi^2_{(14;0,05)}$ =23,685	-
Independen	χ^2_{hitung} =30,186	$\chi^2_{2;0,05}$ = 5,991	2,787x10^{-7}
Berdistribusi Normal	W=0,963	-	0,066

Jika digunakan α=0,05 maka dapat disimpulkan bahwa *residual* model regresi subset dengan x_1, x_2, x_3, x_4 dalam model tidak identik, tidak independen, dan berdistribusi normal.

Model regresi subset dengan seleksi variabel *forward selection* (Model regresi subset 2).

Jika data Ilustrasi 5.1 dimodelkan menggunakan regresi linear dengan menggunakan *forward selection* untuk menentukan variabel independen dalam model maka didapatkan ANOVA seperti pada Tabel 6.4 berikut ini.

Tabel 6.5 ANOVA Model Regresi Subset 2

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	F_{hitung}
Regression	3766,755	3	1255,585	17,188
Residual	4090,818	56	73,050	
Total	7857,573	59		

Statistik uji yang digunakan adalah F, untuk mengambil kesimpulan F yang didapatkan dari perhitungan dibandingkan dengan $F_{(\alpha;k;(n-k-1)}$. Daerah kritis pengujian hipotesisnya adalah tolah H_0 jika $F_{hitung} > F_{(\alpha;k;(n-k-1).}$ Jika digunakan α sebesar 0,05 maka didapatkan $F_{(0,05;3;56)}$ sebesar 2,769. Kesimpulan yang dapat diambil dari ANOVA adalah tolak H_0, artinya terdapat minimal satu estimasi parameter koefisien regresi yang signifikan berpengaruh dalam model regresi linear yang dihasilkan. Nilai R^2_{Adj} yang dihasilkan dari model ini adalah 0,451. Hasil uji parsial dapat dilihat pada Tabel 6.5 berikut ini.

Tabel 6.6 Uji Parsial Model Regresi Subset 2

Simbol	Estimasi Parameter	$se(\hat{\beta}_j)$	t_{hitung}	p_{value}
Intercept	192,212	28,954	6,638	0,000
x_3	-0,005	0,001	-5,536	0,000
x_2	0,000	0,000	-1,423	0,160
x_4	-0,003	0,003	-1,122	0,267

Berdasarkan hasil pada Tabel 6.5 didapatkan informasi bahwa estimasi parameter yang signifikan berpengaruh pada model regresi linear yang dibentuk adalah *intercept* dan X_3. Hasilnya sama dengan model regresi subset sebelumnya. Adanya variabel independen yang tidak bermakna pada model akan menyebabkan nilai SS_{Res} tinggi.

Setelah dilakukan uji serentak dan parsial, selanjutnya dapat dilakukan pendeteksian atau pengujian asumsi *residual*. Tabel 6.6 berikut ini menyajikan rangkuman pengujian asumsi *residual* model regresi subset 2.

Tabel 6.7 Pengujian Asumsi *Residual* Model Regresi Subset 2

Asumsi Residual	Statistik Uji	Nilai Tabel	p_{value}
Identik	χ^2_{hitung}=26,177	$\chi^2_{(9;0,05)}$=16,919	-
Independen	χ^2_{hitung}=29,811	$\chi^2_{(2;0,05)}$= 5,991	$3{,}363 \times 10^{-7}$
Berdistribusi Normal	W=0,959	-	0,044

Jika digunakan α=0,05 maka dapat disimpulkan bahwa *residual* model regresi subset dengan *forward selection* dalam model tidak identik, tidak independen, dan tidak berdistribusi normal.

Model regresi subset dengan seleksi variabel *backward elimination* (Model regresi subset 3).

Pembentukan model regresi subset menggunakan seleksi variabel *backward elimination* menghasilkan model yang sama dengan model regresi subset menggunakan seleksi variabel *forward elimination*. Sehingga tidak diulas lagi dalam subbab ini.

Model regresi subset dengan seleksi variabel *stepwise method* (Model regresi subset 4).

Jika data Ilustrasi 5.1 dimodelkan menggunakan regresi linear dengan menggunakan *stepwise method* untuk seleksi variabel ANOVA yang dihasilkan dapat dilihat pada Tabel 6.7 berikut ini.

Tabel 6.8 ANOVA Model Regresi Subset 4

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	F_{hitung}
Regression	3674,875	2	1837,438	25,04
Residual	4182,698	57	73,381	
Total	7857,573	59		

Statistik uji yang digunakan adalah F, untuk mengambil kesimpulan F yang didapatkan dari perhitungan dibandingkan dengan $F_{(\alpha;k;(n-k-1)}$. Daerah kritis pengujian hipotesisnya adalah tolah H_0 jika $F_{hitung} > F_{(\alpha;k;(n-k-1).}$ Jika digunakan α sebesar 0,05 maka didapatkan $F_{(0,05;2;57)}$ sebesar 3,159. Kesimpulan yang dapat diambil dari ANOVA adalah tolak H_0, artinya terdapat minimal satu estimasi parameter regresi yang signifikan berpengaruh dalam model regresi linear yang dihasilkan. Nilai R^2_{Adj} yang dihasilkan

dari model ini adalah 0,449. Hasil uji parsial dapat dilihat pada Tabel 6.9 berikut ini.

Tabel 6.9 Uji Parsial Model Regresi Subset 4

Simbol	Estimasi Parameter	$se(\hat{\beta}_j)$	t_{hitung}	p_{value}
Intercept	161,584	9,640	16,762	0,000
x_3	-0,005	0,001	-5,459	0,000
x_2	0,000	0,000	-3,431	0,001

Berdasarkan hasil pada Tabel 6.9 didapatkan informasi bahwa semua estimasi parameter dalam model signifikan. Jadi model regresi yang dibentuk memuat variabel independen yang benar-benar berpengaruh terhadap variabel dependennya.

Setelah dilakukan uji serentak dan parsial, selanjutnya dapat dilakukan pendeteksian atau pengujian asumsi *residual*. Tabel 6.10 berikut ini menyajikan rangkuman pengujian asumsi *residual* model regresi subset 4.

Tabel 6.10 Pengujian Asumsi *Residual* Model Regresi Subset 4

Asumsi Residual	Statistik Uji	Nilai Tabel	p_{value}
Identik	χ^2_{hitung}=25,076	$\chi^2_{(9;0,05)}$=11,071	-
Independen	χ^2_{hitung}=28,04	$\chi^2_{(2;0,05)}$= 5,991	$8{,}15\text{x}10^{-7}$
Berdistribusi Normal	W=0,962	-	0,062

Jika digunakan α=0,05 maka dapat disimpulkan bahwa *residual* model regresi subset dengan *stepwise method* dalam model tidak identik, tidak independen, dan berdistribusi normal.

Rangkuman.

Hasil model regresi subset yang telah dianalisis sebelumnya jika dirangkum hasilnya dapat dilihat pada Tabel 6.11 berikut ini.

Tabel 6.11 Rangkuman Model Regresi Subset

Model Regresi Subset	Variabel Independen dalam Model	Hasil Uji Serentak	Hasil Uji Parsial (Variabel Independen dengan Estimasi Parameter Yang Signifikan)	R^2_{Adj}	Asumsi *Residual*		
					Identik	Independen	Berdistribusi Normal
1	X_1, X_2, X_3, X_4	Tolak H_0	X_3	0,448	Tidak Identik	Tidak Independen	Berdistribusi normal
2	X_2, X_3, X_4	Tolak H_0	X_3	0,451	Tidak Identik	Tidak Independen	Tidak berdistribusi normal
3	X_2, X_3, X_4	Tolak H_0	X_3	0,451	Tidak Identik	Tidak Independen	Tidak berdistribusi normal
4	X_2, X_3	Tolak H_0	X_2, X_3	0,449	Tidak Identik	Tidak Independen	Berdistribusi normal

Model regresi subset yang telah dibentuk memberikan hasil yang berbeda-beda. Sebenarnya dari keempat model regresi subset yang telah disajikan tidak ada yang terbaik, tapi yang lebih baik dibandingkan yang lain adalah model regresi subset 4. Model regresi subset 4 lebih baik dibandingkan model yang lain karena hasil uji parsial menunjukkan bahwa semua variabel independen yang dimasukkan dalam model berpengaruh signifikan, namun asumsi residual identik dan independen belum terpenuhi. Dalam penentuan model regresi subset 4 ini belum memasukkan unsur teoritis dari model, sehingga baiknya dalam penentuan model dimasukkan unsur teoritis dari data yang digunakan. Berdasarkan hasil keseluruhan dari data Ilustrasi 5.1, sebaiknya model regresi linear yang dibentuk tidak digunakan dan dilanjutkan dalam pemodelan lain untuk mengatasi permasalahan pada asumsi *residual* yang tidak identik dan independen. Sebaiknya dilakukan pengecekan apakah terdapat unsur *spatiotemporal* dalam data.

Software yang digunakan dalam analisis regresi linear ini adalah *software R* dan Ms. Excel. Terdapat beberapa syntax, langkah-langkah, dan hasil yang tidak dicantumkan disini, termasuk setiap tahapan dalam *forward selection, backward elimination, stepwise method.* Pembaca dapat melihat hal-hal yang tidak dicantumkan dalam buku ini di Lampiran 3.

BAB 7.
IMPLEMENTASI ANALISIS REGRESI LINEAR

PADA bab ini akan dibahas tentang implementasi analisis regresi linear pada permasalahan ekonomi. Analisis regresi linear yang digunakan adalah analisis regresi linear sederhana dan analisis regresi linear berganda.

7.1 Regresi Sederhana

Pada subbab ini akan diilustrasikan implementasi analisis regresi linear sederhana pada permasalahan pemodelan Indeks Harga Konsumen (IHK) terhadap Indeks Harga Saham Gabungan (IHSG). Variabel dependennya adalah IHSG dan variabel independennya adalah IHK. Data diperoleh dari *yahoo finance* dan BPS. Struktur data dapat dilihat pada Lampiran 4.

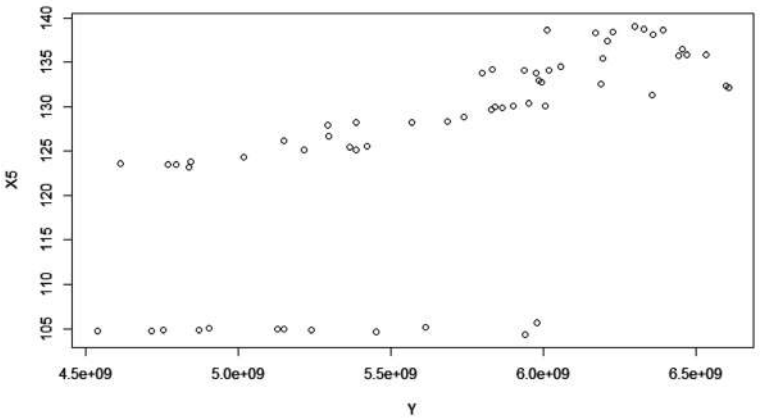
Langkah analisis yang dilakukan untuk menyelesaikan permasalahan pemodelan IHSG menggunakan analisis regresi linear adalah

1. Melakukan identifikasi hubungan antara variabel independen dan dependen menggunakan *Scatterplot.*
2. Melakukan uji korelasi antara variabel dependen dan independen.
3. Membuat pemodelan regresi subset linear sederhana.

   Berdasarkan model regresi linear yang dihasilkan di Langkah nomer 3, untuk setiap model dilakukan analisis
4. Uji serentak

5. Uji parsial
6. Menghitung nilai R^2_{adj}
7. Uji asumsi residual
8. Membandingkan model regresi subset linear sederhana yang didapatkan.
9. Memilih model regresi linear sederhana yang terbaik.

Berikut ini adalah hasil analisis pemodelan regresi linear IHSG (y) dan IHK (x). Gambar 7.1 menunjukkan *scatterplot* antara y dengan x.



Gambar 7.1 *Scatterplot* x dengan y

Pola hubungan antara variabel x dengan y acak dan membentuk garis lurus sehingga antara variabel x dengan y tampak memiliki hubungan linear. Berdasarkan pengujian korelasi antara variabel x dan y diketahui bahwa antara variabel x dan y memiliki nilai korelasi sebesar 0,681, hubungan yang terbentuk positif dan kuat. Ketika dilakukan pengujian korelasi didapatkan hasil

Langkah 1. Membentuk hipotesis uji koefisien korelasi.

H_0: $\rho = 0$.

H_1: $\rho \neq 0$.

Langkah 2. Menghitung statistik uji

t_{hitung}= 7,080

dengan p_{value}=2,15x10^{-9}.

Langkah 3. Menentukan nilai α dan daerah kritis

α=0,05

tolak H_0 jika p_{value}< α

Langkah 4. Mengambil kesimpulan.

Berdasarkan hasil perhitungan nilai t, p_{value} dibandingkan dengan α=0,05 dapat diambil kesimpulan bahwa antara variabel x dengan y memiliki hubungan atau saling berkorelasi.

Setelah dianalisis hubungan antara variabel x dengan y, Langkah selanjutnya adalah melakukan pemodelan regresi, karena variabel independen yang digunakan 1 maka hanya ada 1 model regresi yang terbentuk. Hasil ANOVA pemodelan regresi linear antara variabel x dengan y ditunjukkan pada Tabel 7.1 berikut ini.

Tabel 7.1 ANOVA Regresi Linear Sederhana Variabel x dengan y

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	F_{hitung}
Regression	545,5	1	545,49	4,327
Residual	7312,1	58	126,07	
Total	7857,6			

Statistik uji yang digunakan adalah F, untuk mengambil kesimpulan F yang didapatkan dari perhitungan dibandingkan dengan $F_{(\alpha;k;(n-k-1)}$. Daerah kritis pengujian hipotesisnya adalah tolah H_0 jika F_{hitung}> $F_{(\alpha;k;(n-k-1).}$ Jika digunakan α sebesar 0,05 maka didapatkan $F_{(0,05;1;58)}$ sebesar 4,007. Kesimpulan yang dapat

diambil dari ANOVA adalah tolak H_0, artinya terdapat minimal satu estimasi parameter koefisien regresi yang signifikan berpengaruh dalam model regresi linear yang dihasilkan. Nilai R^2_{Adj} yang dihasilkan dari model ini adalah 0,053. Hasil uji parsial dapat dilihat pada Tabel 7.2 berikut ini.

Tabel 7.2 Uji Parsial Regresi Linear Sederhana Variabel x dengan y

Simbol	Estimasi Parameter	$se(\hat{\beta}_j)$	t_{hitung}	p_{value}
Intercept	194,420	32,910	5,908	$1{,}93x10^{-7}$
X	-0,005	0,002	-2,080	0,042

Informasi yang diperoleh berdasarkan Tabel 8.2 adalah *intercept* dan variabel x berpengaruh signifikan terhadap variabel y yang ditunjukkan oleh nilai p_{value}<0,05. Selanjutnya dilakukan uji asumsi residual yang hasilnya ditunjukkan pada Tabel 7.3 berikut ini.

Tabel 7.3 Pengujian Asumsi *Residual* Regresi Linear Sederhana Variabel x dengan y

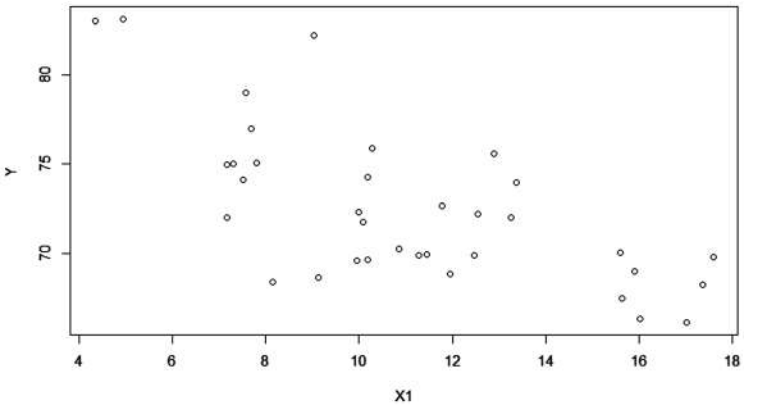
Asumsi Residual	Statistik Uji	Nilai Tabel	p_{value}
Identik	$\mathcal{X}^2_{hitung}$ =26,712	$\mathcal{X}^2_{(3;0,05)}$ =7,815	-
Independen	$\mathcal{X}^2_{hitung}$ =48,111	$\mathcal{X}^2_{(2;0,05)}$= 5,991	3,571x10-11
Berdistribusi Normal	W=0,9108	-	0,000

Jika digunakan α=0,05 maka dapat disimpulkan bahwa *residual* model regresi linear sederhana antara x dengan y tidak identik, tidak independen, dan tidak berdistribusi normal. Berdasarkan data ilustrasi regresi linear berganda didapatkan model rgresi terbaik yaitu,

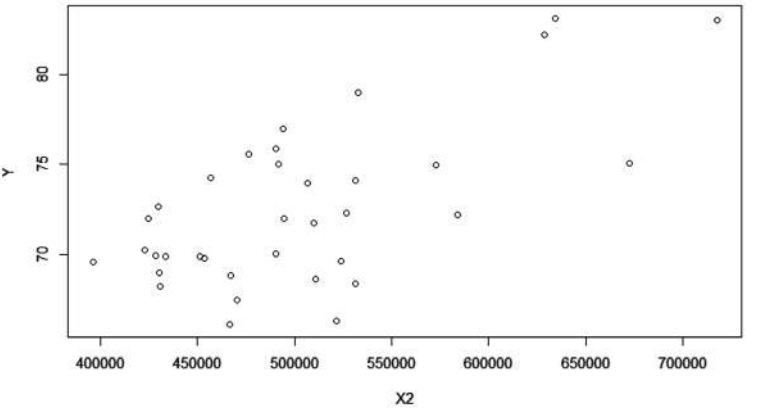
$\hat{y} = 194{,}420 - 0{,}005\ x.$

7.2 Regresi Berganda

Pada subbab ini akan diilustrasikan analisis regresi linear berganda menggunakan data sosial ekonomi tahun 2020 yang didapatkan dari BPS. Unit penelitiannya adalah kabupaten/kota di provinsi Jawa Tengah. Variabel dependen yang digunakan adalah Indeks Pembangunan Manusia (y), variabel independen yang digunakan adalah %kemiskinan (x_1) dan rata-rata pengeluaran untuk konsumsi makanan (x_2). Data dapat dilihat pada Lampiran 5.



(a) Scatterplot x_1 dengan y



(b) *Scatterplot* x_2 dengan y

Gambar 7.2 *Scatterplot* Variabel X_1, X_2 dengan Y

Gambar 7.2 menunjukkan visualisasi pola hubungan yang terbentuk antara variabel x dan y dalam *scatterplot*. Gambar 7.2 (a) menunjukkan pola hubungan antara variabel x_1 dengan y, berdasarkan *scatterplot* variabel x_1 dengan y memiliki pola hubungan lurus menurun. Gambar 7.2 (b) menunjukkan pola hubungan antara variabel x_2 dengan y, berdasarkan *scatterplot* variabel x_2 dengan y membentuk pola hubungan lurus meningkat. Jika dilakukan pengujian korelasi, seperti yang ditunjukkan pada Tabel 7.4 tampak bahwa antara variabel x_1 dengan y memiliki hubungan negatif yang sesuai dengan hasil *scatterplot*nya.

Tabel 7.4 Rangkuman Hasil Uji Korelasi

Variabel dependen-variabel independen	Nilai Korelasi	p_{value}	Kesimpulan
$y\text{-}x_1$	-0,694	$3{,}774 \times 10^{-6}$	berkorelasi
$y\text{-}x_2$	0,666	$1{,}28 \times 10^{-5}$	berkorelasi

Kesimpulan yang dapat diambil berdasarkan Tabel 7.4 adalah variabel x_1 dan x_2 berkorelasi dengan variabel y. Oleh karena itu, antara variabel x_1, x_2, dan y dapat dimodelkan menggunakan regresi linear berganda.

Sebelum analisis regresi linear berganda dibentuk, diperlukan Langkah analisis untuk mengecek asumsi multikolinieritas karena variabel independen yang digunakan lebih dari 1.

Tabel 7.5 Uji Korelasi Antara Variabel Independen

Variabel independen	x_1	x_2
x_1	1	-0,532 $_{(0,000)*}$
x_2		1

Tabel 7.5 memberikan informasi bahwa antara variabel x_1 dengan x_2 saling berkorelasi yang mengindikasikan adanya multikolineritas pada data yang digunakan. Kondisi ini sebenarnya sangat berpengaruh pada hasil analisis regresi linear

berganda. Kondisi multikolinieritas sebaiknya diatasi, salah satunya dengan menggunakan metode seleksi variabel yang lebih modern seperti menggunakan algoritma genetika yang tidak dibahas dalam bab ini.

Pada analisis regresi linear berganda dibutuhkan pembentukan model regresi subset untuk menentukan model regresi yang terbaik karena jumlah variabel independen yang digunakan lebih dari 1. Pembentukan model regresi subset menggunakan bantuan metode seleksi variabel. Berikut ini adalah analisis dari masing-masing regresi subset yang terbentuk.

Model regresi subset dengan x_1, x_2 dalam model regresi (Model regresi subset 1).

Hasil uji serentak parameter koefisien regresi model subset 1 menunjukkan bahwa terdapat minimal satu variabel independen yang berpengaruh signifikan terhadap variabel dependennya. Hasilnya dapat dilihat pada Tabel 7.6 berikut ini.

Tabel 7.6 ANOVA Model Regresi Subset 1

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	F_{hitung}
Regression	402,04	2	201,02	24,41
Residual	263,57	32	8,24	
Total	665,61			

Variabel independen yang berpengaruh signifikan adalah variabel x_1 dan x_2, serta *intercept*nya. Hasil pengujian parsial model regresi subset 1 dapat dilihat pada Tabel 7.7 berikut ini.

Tabel 7.7 Uji Parsial Model Regresi Subset 1

Simbol	Estimasi Parameter	$se(\hat{\beta}_j)$	t_{hitung}	p_{value}
Intercept	6,661 x 10	5,185	12,847	$3{,}56 \times 10^{-14}$
x_1	$-5{,}950 \times 10^{-1}$	$1{,}650 \times 10^{-1}$	-3,606	0,001
x_2	2,475 x 10-5	$7{,}872 \times 10^{-6}$	3,145	0,004

Setelah diperoleh informasi tentang pengujian signifikansi estimasi parameter koefisien regresi selanjutnya adalah melakukan pengujian asumsi residual. Hasilnya dapat dilihat pada Tabel 7.8 berikut ini.

Tabel 7.8 Pengujian Asumsi *Residual* Model Regresi Subset 1

Asumsi Residual	Statistik Uji	Nilai Tabel	p_{value}
Identik	χ^2_{hitung}=7,784	$\chi^2_{(3;0,05)}$=12,592	-
Independen	χ^2_{hitung}=15,025	$\chi^2_{2;0,05}$= 5,991	0,001
Berdistribusi Normal	W=0,982	-	0,824

Berdasarkan hasil uji asumsi *residual* pada Tabel 7.8, jika digunakan α=0,05 maka dapat disimpulkan bahwa *residual* model regresi subset 1 tidak identik, tidak independen, dan berdistribusi normal.

Seleksi variabel *forward, backward,* dan *stepwise* memberikan hasil yang sama yaitu x_1, x_2 masuk dalam model regresi linear berganda. Berdasarkan data ilustrasi regresi linear berganda didapatkan model rgresi terbaik yaitu,

$\hat{y}$=6,661x10-5,950 x10^{-1} x_1+2,475x10^{-5} x_2 .

DAFTAR PUSTAKA

Adeboye, N. O., Fagoyinbo, I. S., & Olatayo, T. O. (2014). Estimation of the Effect of Multicollinearity on the Standard Error for Regression Coefficients Vol 10. Journal of Mathematics (IOSR-JM), 16-20.

Arabatzis, G., & Malesios, C. (2011). An econometric analysis of residential consumption of fuelwood in a mountainous prefecture of Northern Greece. *Energy Policy*, *39*(12), 8088–8097. https://doi.org/10.1016/j.enpol.2011.10.003

Astuti, S. P. (2020). *Statistika* (Cetakan 1). Gerbang Media Aksara.

Basuki AT. (2017). Pengantar Ekonometrika. Sleman : Danisa Media

Bawono A dan Shina AFI. (2018). Ekonometrika Terapan Untuk Ekonomi dan Bisnis Islam Aplikasi Eviews. Salatiga : Lembaga Penelitian dan Pengabdian kepada Masyarakat IAIN Salatiga

Best, H., & Wolf, C. (2015). *Regression Analysisi and Causal Inference.* Croydon: SAGE Publications.

BPS. (2021). *Beranda.* Retrieved from Badan Pusat Statistika: https://www.bps.go.id/subject/3/inflasi.html#subjekViewTab3

Dargay, J. M. (2001). The effect of income on car ownership: Evidence of asymmetry. *Transportation Research Part A: Policy and Practice*, *35*(9), 807–821. https://doi.org/10.1016/S0965-8564(00)00018-5

Duque, J. C., Patino, J. E., Ruiz, L. A., & Pardo-Pascual, J. E. (2015). Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. *Landscape and Urban Planning*, *135*, 11–21. https://doi.org/10.1016/j.landurbplan.2014.11.009

Fukuchi, T. (2000). Econometric analysis of the effect of Krismon shocks on Indonesia's industrial subsectors. *Developing Economies*, *38*(4), 490–517.

Glytsos, N. P. (2002). *Dynamic Effects of Migrant Remittances on Growth : An Econometric Model with an Application to Mediterranean Countries* (Issue 74).

Green, D. J. (2004). Investment behavior and the Economic Crisis in Indonesia. *Journal of Asian Economics*, *15*(2), 287–303. https://doi.org/10.1016/j.asieco.2004.02.003

Gujarati, D. N. (2003). *Basic Econometrics* (Internatio). McGraw-Hill.

Gujarati, D. (2004). *Basic Econometrics.* Hill: McGraw.

Haan, L. de, Naumovska, A., & Peeters, H. M. M. (2001). Makmodel: A macroeconometric model for the Republic of Macedonia. In *Research Memorandum WO&E no. 665/0120*.

Iriawan, N., & Astuti, S. P. (2006). *Mengolah Data Statistika dengan Mudah Menggunakan Minitab 14*. Andi Offset.

Israel, K.-F. (2018). Pawel Ciompa and the Meaning of Econometrics: A Comparison of Two Concepts. *SSRN Electronic Journal*, *33* (0). https://doi.org/10.2139/ssrn.3288520

Johnson, R. A., & Bhattacharyya, G. K. (2019). *Statistics: Principles and Methods Eighth Edition.* Hoboken: Wiley.

Larsen, B. M., & Nesbakken, R. (2004). Household electricity end-use consumption: Results from econometric and engineering models. *Energy Economics*, *26*(2), 179–200. https://doi.org/10.1016/j.eneco.2004.02.001

Li, X., Qiao, Y., & Shi, L. (2017). The aggregate effect of air pollution regulation on CO2 mitigation in China's manufacturing

industry: an econometric analysis. *Journal of Cleaner Production*, *2*, 976–984. https://doi.org/10.1016/j.jclepro.2016.03.015

Márquez, M. A., Ramajo, J., & Hewings, G. J. . (2010). A spatio-temporal econometric model of regional growth in Spain. *Journal of Geographical Systems*, *12*(2), 207–226. https://doi.org/10.1007/s10109-010-0119-3

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. Hoboken: John Wiley & Sons.

Parajuli, R., Østergaard, P. A., Dalgaard, T., & Pokharel, G. R. (2014). Energy consumption projection of Nepal: An econometric approach. *Renewable Energy*, *63*, 432–444. https://doi.org/10.1016/j.renene.2013.09.048

Razali, N. M., & Wah, Y. B. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Liliefors and Anderson-Darling Test. *Journal of Statistical Modeling and Analytics Vol.2 No.1*, 21-33.

Royston, J. P. (1982). An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. *Journal of the Royal Statistical Society: Series C Volume 31*, 115-124.

Ruggins, S. M. (2016). A History of Econometrics: The Reformation from the 1970s. *Journal of Cultural Economy*, *9*(2), 226–228. https://doi.org/10.1080/17530350.2015.1009148

Sumarjaya IW. (2017). Modul Ekonometrika. Denpasar : Universitas Udayana

Sumantri, F., & Latifah, U. (2019). Faktor-Faktor yang Mempengaruhi Indeks Harga Konsumen. *Jurnal Sekretari dan Manajemen*, 25-34.

Thode, H. C. (2002). *Testing for Normality*. New York: Marcel Dekker.

Verbeek, M. (2017). *A Guide to Modern Econom*terics (Fifth edit). John Wiley & Sons, lnc.

Walters, A. A., & Johnston, J. (1965). Econometric Methods. *Economica*, *32*(126), 231. https://doi.org/10.2307/2552555

Wang, S., Zhou, C., Li, G., & Feng, K. (2016). CO2, economic growth, and energy consumption in China's provinces: Investigating the spatiotemporal and econometric characteristics of China's CO2 emissions. *Ecological Indicators*, *69*, 184–195. https://doi.org/10.1016/j.ecolind.2016.04.022

Wilcox, R. (2012). *Introduction to Robust Estimation and Hypothesis Testing 3rd Edition*. San Diego: Elsevier.

Yoo, W., Mayberry, R., Bae, S., Singh, K., Qinghua, & Jr, J. W. (2014). A Study of Effects of MultiCollinearity in the Multivariable Analysis. *International Journal of Applied Science and Technology*.

Zheng, B. (1994). Can a Poverty Index be Both Relative and Absolute? *Econometrica*, *62*(6), 1453–1458.

Zwane, A. P. (2007). Does poverty constrain deforestation? Econometric evidence from Peru. *Journal of Development Economics*, *84*(1), 330–349. https://doi.org/10.1016/j.jdeveco.2005.11.007

LAMPIRAN

Lampiran 1 *Output* Model Regresi Dummy

Model Regresi Dummy ANOVA

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.695[a]	.483	.449	4.98797E5

a. Predictors: (Constant), Wilayah2, Wilayah3

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7.194E12	2	3.597E12	14.458	.000[a]
	Residual	7.713E12	31	2.488E11		
	Total	1.491E13	33			

a. Predictors: (Constant), Wilayah2, Wilayah3

b. Dependent Variable: Proovinsi

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	337197.800	157733.319		2.138	.041
	Wilayah3	-165658.035	198783.742	-.125	-.833	.411
	Wilayah2	1.020E6	245809.623	.623	4.148	.000

a. Dependent Variable: Proovinsi

Model Regresi Dummy ANCOVA

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.848[a]	.719	.672	1.62989

a. Predictors: (Constant), D, X1

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	81.500	2	40.750	15.340	.000[a]
	Residual	31.878	12	2.657		
	Total	113.379	14			

a. Predictors: (Constant), D, X1

b. Dependent Variable: Y

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.860	1.995		2.437	.031
	X1	.111	.301	.057	.368	.719
	D	4.641	.846	.842	5.488	.000

a. Dependent Variable: Y

Regresi Dummy Lebih Dari Satu

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.940[a]	.884	.853	.89469

a. Predictors: (Constant), X.1, D3, D1, D2

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	91.793	4	22.948	28.668	.000[a]
	Residual	12.007	15	.800		
	Total	103.800	19			

a. Predictors: (Constant), X.1, D3, D1, D2

b. Dependent Variable: Y.1

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	1.703	1.357		1.255	.229
	D1	3.762	.603	.826	6.234	.000
	D2	.267	.651	.057	.410	.688
	D3	.975	.635	.204	1.536	.145
	X.1	.292	.192	.200	1.520	.149

a. Dependent Variable: Y.1

Lampiran 2 Data Ilustrasi 4.1

Observasi	IHK (Y)	BIRate (X1)	Jumlah Uang Beredar (X2)	Ekspor Bersih (X3)	Nilai Tukar Rupiah Terhadap Dolar (X4)
1	123,62	7,25	4498361,28	114,9	13770
2	123,51	7	4521951,2	1141,1	13367
3	123,75	6,75	4561872,52	510,4	13255
4	123,19	6,75	4581877,87	876,1	13180
5	123,48	6,75	4614061,82	376,7	13655
6	124,29	6,5	4737451,23	1110,9	13210
7	125,15	6,5	4730379,68	632,3	13097
8	125,13	5,25	4746026,68	316,5	13265
9	125,41	5	4737630,76	1282,3	13047
10	125,59	4,75	4778478,89	1236,5	13047
11	126,18	4,75	4868651,16	833,5	13550
12	126,71	4,75	5004976,79	1049,9	13470
13	127,94	4,75	4936881,99	1423,9	13347
14	128,24	4,75	4942919,76	1256,6	13331
15	128,22	4,75	5017643,55	1435,3	13323
16	128,33	4,75	5033780,29	1319,1	13327
17	128,83	4,75	5125383,79	561,4	13321
18	129,72	4,75	5225165,76	1669,8	13325
19	130	4,75	5178078,75	-278,7	13324
20	129,91	4,5	5219647,63	1678,8	13324
21	130,08	4,25	5254138,51	1792	13470
22	130,09	4,25	5284320,16	1003,4	13560
23	130,35	4,25	5321431,77	221,2	13524
24	131,28	4,25	5419165,05	-240	13565
25	132,1	4,25	5351684,67	-733,1	13387
26	132,32	4,25	5351650,33	-53,1	13740
27	132,58	4,25	5395826,04	1047	13760

Observasi	IHK (Y)	BIRate (X1)	Jumlah Uang Beredar (X2)	Ekspor Bersih (X3)	Nilai Tukar Rupiah Terhadap Dolar (X4)
28	132,71	4,25	5409088,81	-1666,1	13910
29	132,99	4,75	5435082,93	-1464,6	13890
30	133,77	5,25	5534149,83	1673,8	14325
31	134,14	5,25	5507791,75	-2012,4	14415
32	134,07	5,5	5529451,81	-953	14725
33	133,83	5,75	5606779,89	346,3	14900
34	134,2	5,75	5667512,1	-1758,5	15200
35	134,56	6	5670975,24	-2050,1	14300
36	135,39	6	5760046,2	-1074,8	14375
37	135,83	6	5644985	-963,3	13970
38	135,72	6	5670778	562,6	14060
39	135,87	6	5747247	996,7	14235
40	136,47	6	5746732	-2331,1	14245
41	137,4	6	5860509	145,1	14270
42	138,16	6	5908509	267,9	14125
43	138,59	5,75	5941133	-280,1	14012
44	138,75	5,5	5934562	92,7	14180
45	138,37	5,25	6134178	-183,3	14190
46	138,4	5	6026908	122,4	14032
47	138,6	5	6074377	-1396	14100
48	139,07	5	6136552	-78	13880
49	104,33	5	6046651	-632,3	13650
50	104,62	4,75	6116495	2494	14340
51	104,72	4,5	6440457,39	679,1	16300
52	104,8	4,5	6238267	-375,4	14825
53	104,87	4,5	6468193,5	2014	14575
54	105,06	4,25	6393743,8	1246,5	14180
55	104,95	4	6567725,02	3225,6	14530

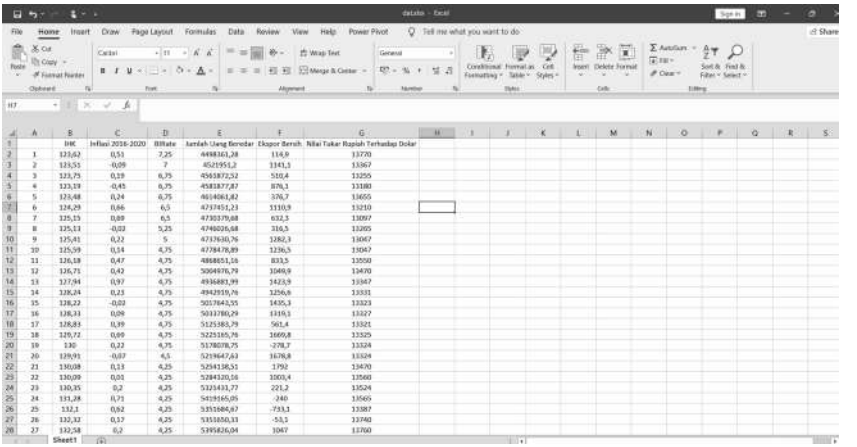
Observasi	IHK (Y)	BIRate (X1)	Jumlah Uang Beredar (X2)	Ekspor Bersih (X3)	Nilai Tukar Rupiah Terhadap Dolar (X4)
56	104,9	4	6726135,25	2312,9	14560
57	104,85	4	6748574,03	2386,1	14840
58	104,92	4	6780844,54	3577,4	14620
59	105,21	3,75	6817456,68	2594	14090
60	105,68	3,75	6900049,49	2101,2	14040

Lampiran 3 Analisis regresi Menggunakan RStudio

Ilustrasi imlementasi regresi linear sederhana dan berganda diolah menggunakan bantuan RStudio yang merupakan Bahasa pemograman pengolahan data statistika. RStudi bersifat *open source* dan komersial, sehingga siapapun yang akan menggunakan tidak diwajibkan untuk membeli lisensinya. RStudio dapat di-download melalui RStudio|Open source & professional software for data science teams - RStudio. Pengoperasian RStudio sangat mudah jika pembaca mengetahui *syntax* yang digunakan. Video tutorial penggunaan RStudio dapat dilihat disini: https://youtube.com/playlist?list=PLI12lTYMsYm4mnzGpwjQg_pUCUokpNlBo.

Berikut ini akan diberikan Langkah-langkah analisis regresi linear berganda berdasarkan data pada Lampiran 2 menggunakan RStudio.

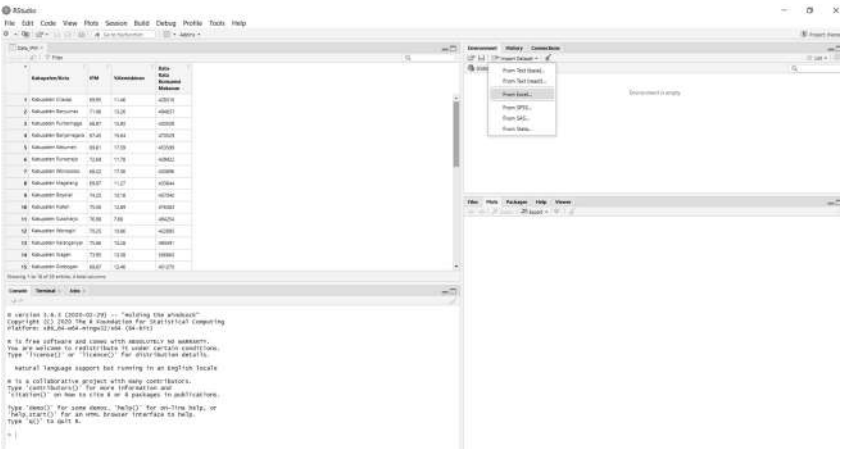
Langkah 1. Menyiapkan data pada Ms. Excel

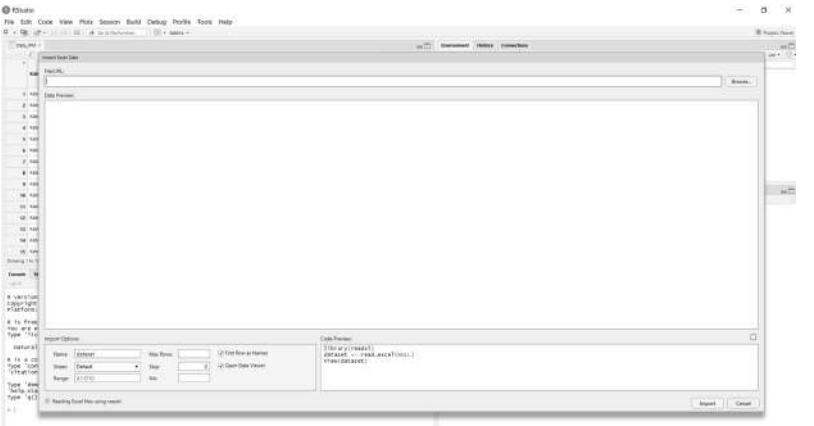


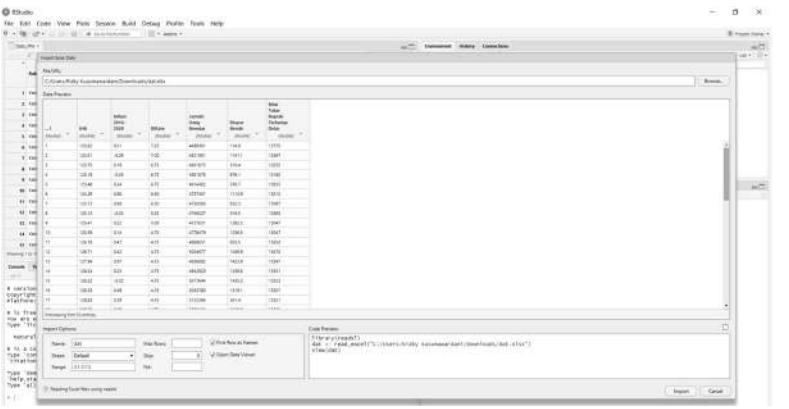
	A	B	C	D	E	F	G
1		IHK	Inflasi 2016-2020	BIRate	Jumlah Uang Beredar	Ekspor Bersih	Nilai Tukar Rupiah Terhadap Dolar
2	1	123,62	0,51	7,25	4498361,28	114,9	13770
3	2	123,51	0,09	7	4521951,2	1141,1	13367
4	3	123,75	0,19	6,75	4563872,52	510,4	13255
5	4	123,19	-0,45	6,75	4583877,87	876,1	13180
6	5	123,48	0,24	6,75	4614061,82	376,7	13655
7	6	124,29	0,66	6,5	4737451,23	1110,9	13210
8	7	125,15	0,69	6,5	4730379,68	632,3	13097
9	8	125,13	-0,02	5,25	4746026,68	316,5	13265
10	9	125,41	0,22	5	4737630,76	1282,3	13047
11	10	125,59	0,14	4,75	4778478,89	1236,5	13047
12	11	126,18	0,47	4,75	4868651,16	833,5	13550
13	12	126,71	0,42	4,75	5004976,79	1049,9	13470
14	13	127,94	0,97	4,75	4936881,99	1423,9	13347
15	14	128,24	0,23	4,75	4942559,76	1256,6	13331
16	15	128,22	-0,02	4,75	5017643,55	1435,3	13323
17	16	128,33	0,09	4,75	5033786,29	1319,1	13327
18	17	128,83	0,39	4,75	5125383,79	561,4	13321
19	18	129,72	0,69	4,75	5229165,79	1699,8	13325
20	19	130	0,22	4,75	5178078,75	-278,7	13324
21	20	129,91	-0,07	4,5	5219647,63	1678,8	13324
22	21	130,08	0,13	4,25	5254138,51	1792	13470
23	22	130,09	0,01	4,25	5284320,16	1003,4	13560
24	23	130,35	0,2	4,25	5321433,77	221,2	13524
25	24	131,28	0,71	4,25	5419165,05	-240	13565
26	25	132,1	0,62	4,25	5353684,67	-733,1	13387
27	26	132,32	0,17	4,25	5353850,33	-53,1	13740
28	27	132,58	0,2	4,25	5395826,04	1047	13760

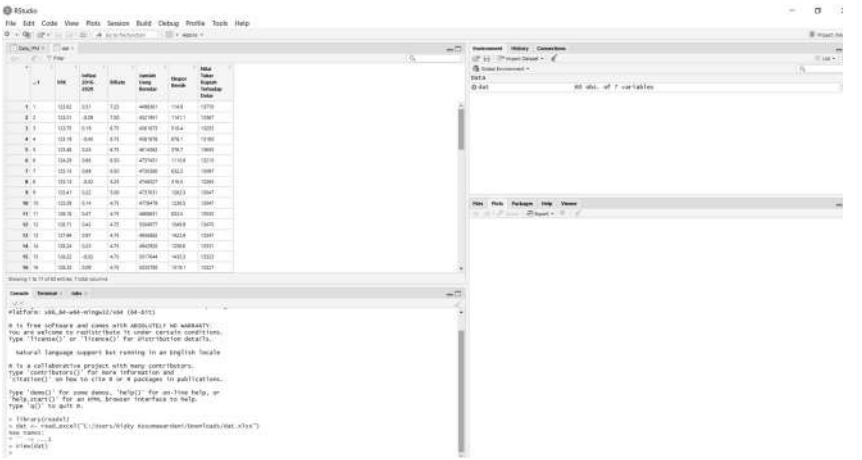
Langkah 2. Memasukkan data dalam RStudio

- Pilih *import dataset*
- Klik *frim excel*
- Klik *browse* untuk mencari data dalam penyimpanan komputer
- Cari data yang akan digunakan, lalu *double* klik
- Klik *import*









Tampilan data yang sudah di*import*.

Langkah 3. Mendefinisikan variabel dependen (y) dan independen (x) yang digunakan

Syntax

Variabel=Nama data yang masuk dalam RStudio$Nama Kolom

```
> y=dat$IHK
> x1=dat$BIRate
> x2=dat$`Jumlah Uang Beredar`
> x3=dat$`Ekspor Bersih`
> x4=dat$`Nilai Tukar Rupiah Terhadap Dolar`
```

Langkah 4. Membuat *scatterplot* x dengan y

```
> plot(x1,y)
> plot(x2,y)
> plot(x3,y)
> plot(x4,y)
```

Langkah 5. Menguji korelasi antara x dengan y

```
> cor.test(x1,y)
> cor.test(x2,y)
> cor.test(x3,y)
> cor.test(x4,y)
```

Langkah 6. Membentuk model regresi linear berganda

> reg=lm(formula=y~x1+x2+x3+x4)

reg merupakan istilah bantuan yang digunakan untuk menyimpan hasil pemodelan regresi linear berganda.

Langkah 7. Menampilkan ringkasannya

> summary(reg)

Summary digunakan untuk menampilkan uji parsial, R^2, nilai F dalam ANOVA, dsb.

Langkah 8. Menampilkan ANOVA

> anova(reg)

Hasil ANOVA dapat digunakan untuk melakukan uji serentak, namun untuk SSE dan MSE *regression* harus dijumlahkan secara manual.

Langkah 9. Menampilkan *residual*

> res=residuals(reg)

res merupakan istilah bantuan yang digunakan untuk menyimpan hasil *residual* yang dapat digunakan untuk pengujian asumsi *residual* identik.

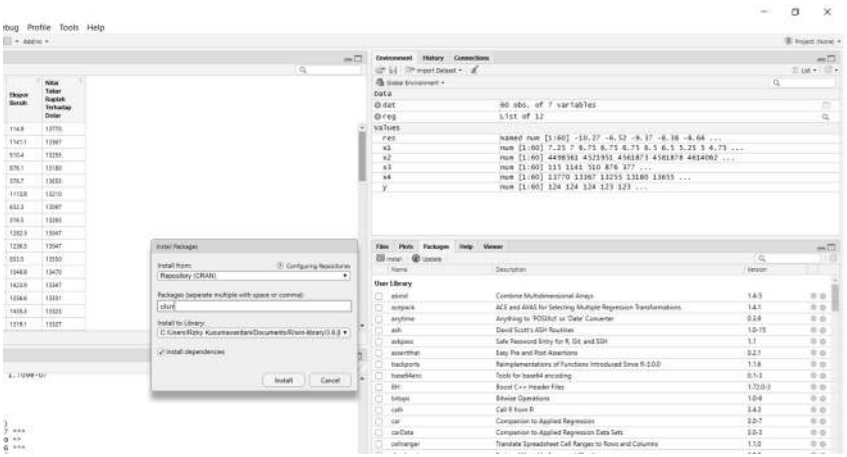
Langkah 10. Seleksi variabel

Seleksi variabel menggunakan *library olssr*, sebelum menggunakan *syntax* seleksi variabel maka *library* ini harus diaktifkan terlebih dahulu.

> library(olsrr)

Jika dalam komputer belum terinstal, maka *library* ini harus diinstal terlebih dahulu. Caranya:

a. Klik *package*
b. Klik *install*
c. *Install from*: *Respository* (CRAN)
d. *Package*:olsrr
e. Klik *install.*



Pada saat menginstal komputer harus tersambung dengan internet.

Backward

> ols_step_backward_p(reg)

```
                        Elimination Summary
-------------------------------------------------------------------------------
        Variable                    Adj.
Step    Removed     R-Square      R-Square      C(p)        AIC         RMSE
-------------------------------------------------------------------------------
   1    x1            0.4794       0.4515     3.5928    433.6020     8.5469
-------------------------------------------------------------------------------
```

Forward

> ols_step_forward_p(reg)

```
                          Selection Summary
-------------------------------------------------------------------------------
        Variable                    Adj.
Step    Entered     R-Square      R-Square      C(p)        AIC         RMSE
-------------------------------------------------------------------------------
   1    x3            0.3578       0.3467    12.5782    442.1972     9.3277
   2    x2            0.4677       0.4490     2.8414    432.9347     8.5663
   3    x4            0.4794       0.4515     3.5928    433.6020     8.5469
-------------------------------------------------------------------------------
```

Stepwise

> ols_step_both_p(reg)

```
                        Stepwise Selection Summary
-------------------------------------------------------------------------------
                    Added/                    Adj.
Step    Variable    Removed     R-Square      R-Square      C(p)        AIC         RMSE
-------------------------------------------------------------------------------
   1        x3      addition       0.358        0.347    12.5780    442.1972     9.3277
   2        x2      addition       0.468        0.449     2.8410    432.9347     8.5663
-------------------------------------------------------------------------------
```

Langkah 11. Menguji asumsi *residual* identik

Terdapat beberapa tahapan dalam pengujian asumsi ini, yaitu

a. Menghitung *residual* dipangkatkan dua
b. Menghitung x_1 dipangkatkan dua
c. Menghitung x_2 dipangkatkan dua
d. Menghitung x_3 dipangkatkan dua
e. Menghitung x_4 dipangkatkan dua
f. Menghitung x_1 dikalikan x_2
g. Menghitung x_1 dikalikan x_3
h. Menghitung x_1 dikalikan x_4
i. Menghitung x_2 dikalikan x_3
j. Menghitung x_2 dikalikan x_4
k. Menghitung x_3 dikalikan x_4
l. Membuat model regresi dengan variabel dependen=point a dan variabel independen=poinb-k
m. Mendapatkan nilai R^2
n. Menghitung F_{hitung}= nxR^2

```
> reskuadrat=res^2
> x1kuadrat=x1^2
> x2kuadrat=x2^2
> x3kuadrat=x3^2
> x4kuadrat=x4^2
> x1x2=x1*x2
> x1x3=x1*x3
> x1x4=x1*x4
> x2x3=x2*x3
> x2x4=x2*x4
> x3x4=x3*x4
> reg2=lm (formula = reskuadrat ~ x1 + x2 + x3 + x4 + x1kuadrat
  + x2kuadrat + x3kuadrat + x4kuadrat + x1x2 + x1x3 + x1x4 +
  x2x3 + x2x4 + x3x4)
```

> summary(reg2)

```
Call:
lm(formula = reskuadrat ~ x1 + x2 + x3 + x4 + x1kuadrat + x2kuadrat +
    x3kuadrat + x4kuadrat + x1x2 + x1x3 + x1x4 + x2x3 + x2x4 +
    x3x4)

Residuals:
    Min      1Q  Median      3Q     Max
-216.89  -35.24  -13.53   17.02  428.09

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.439e+04  1.092e+04  -1.317  0.19442
x1           1.581e+03  1.069e+03   1.479  0.14614
x2           3.840e-03  1.658e-03   2.315  0.02521 *
x3          -2.825e-01  3.880e-01  -0.728  0.47030
x4          -5.759e-02  1.185e+00  -0.049  0.96147
x1kuadrat   -5.139e+01  4.016e+01  -1.280  0.20728
x2kuadrat    3.068e-11  1.361e-10   0.225  0.82274
x3kuadrat   -7.103e-07  1.049e-05  -0.068  0.94630
x4kuadrat    5.533e-05  5.284e-05   1.047  0.30064
x1x2        -1.161e-04  1.191e-04  -0.975  0.33476
x1x3         3.278e-03  2.453e-02   0.134  0.89431
x1x4        -2.888e-02  9.765e-02  -0.296  0.76878
x2x3        -1.291e-07  4.369e-08  -2.955  0.00496 **
x2x4        -2.466e-07  1.726e-07  -1.429  0.15989
x3x4         7.185e-05  3.615e-05   1.987  0.05299 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 93.02 on 45 degrees of freedom
Multiple R-squared:  0.4623,    Adjusted R-squared:  0.295
F-statistic: 2.764 on 14 and 45 DF,  p-value: 0.004942
```

> Rkuadrat=0.4623

> n=60

> Fhit=n*Rkuadrat

> Fhit

[1] 27.738

Langkah 12. Menguji asumsi *residual* independen

Pengujian sumsi *residual* independen menggunakan *library lmtest*, sebelum menggunakan *syntax* bgtest maka *library* ini harus diaktifkan terlebih dahulu.

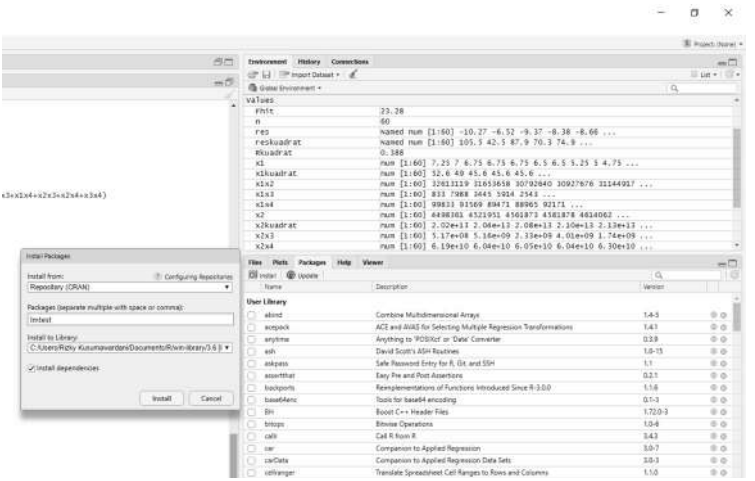
> library(lmtest)

Jika dalam komputer belum terinstal, maka *library* ini harus diinstal terlebih dahulu. Caranya:

a. Klik *package*
b. Klik *install*
c. *Install from*: *Respository* (CRAN)

d. *Package*:lmtest

e. Klik *install.*



> bgtest(reg,order=2)

reg menunjukkan model regresi yang akan diuji *residualnya,* order menunjukkan jumlah *lag* yang diuji. Hasilnya adalah

```
        Breusch-Godfrey test for serial correlation of order up to 2

data:  reg
LM test = 30.186, df = 2, p-value = 2.787e-07

>
```

Langkah 13. Menguji asumsi *residual* berdistribusi normal

> shapiro.test(res)

```
        Shapiro-Wilk normality test

data:  res
W = 0.96304, p-value = 0.0663

>
```

Lampiran 4 Data Ilustrasi Regresi Linear Sederhana Bab 7

Observasi	IHSG (Y)	IHK (X)
1	4.615.163.086	123,62
2	4.770.956.055	123,51
3	4.845.371.094	123,75
4	4.838.583.008	123,19
5	4.796.869.141	123,48
6	5.016.646.973	124,29
7	5.215.994.141	125,15
8	5.386.082.031	125,13
9	5.364.804.199	125,41
10	5.422.541.992	125,59
11	5.148.910.156	126,18
12	5.296.710.938	126,71
13	5.294.103.027	127,94
14	5.386.691.895	128,24
15	5.568.105.957	128,22
16	5.685.297.852	128,33
17	5.738.154.785	128,83
18	5.829.708.008	129,72
19	5.840.938.965	130
20	5.864.059.082	129,91
21	5.900.854.004	130,08
22	6.005.784.180	130,09
23	5.952.138.184	130,35
24	6.355.653.809	131,28
25	6.605.630.859	132,1
26	6.597.217.773	132,32
27	6.188.986.816	132,58
28	5.994.595.215	132,71
29	5.983.586.914	132,99
30	5.799.236.816	133,77

Observasi	IHSG (Y)	IHK (X)
31	5.936.442.871	134,14
32	6.018.459.961	134,07
33	5.976.553.223	133,83
34	5.831.649.902	134,2
35	6.056.124.023	134,56
36	6.194.498.047	135,39
37	6.532.969.238	135,83
38	6.443.348.145	135,72
39	6.468.754.883	135,87
40	6.455.352.051	136,47
41	6.209.117.188	137,4
42	6.358.628.906	138,16
43	6.390.504.883	138,59
44	6.328.470.215	138,75
45	6.169.102.051	138,37
46	6.228.316.895	138,4
47	6.011.830.078	138,6
48	6.299.539.063	139,07
49	5.940.047.852	104,33
50	5.452.704.102	104,62
51	4.538.930.176	104,72
52	4.716.402.832	104,8
53	4.753.611.816	104,87
54	4.905.392.090	105,06
55	5.149.626.953	104,95
56	5.238.486.816	104,9
57	4.870.039.063	104,85
58	5.128.225.098	104,92
59	5.612.415.039	105,21
60	5.979.073.242	105,68

Lampiran 5 Data Ilustrasi Regresi Linear Berganda Bab 7

Kabupaten/Kota	IPM	%Kemiskinan	Rata-Rata Konsumsi Makanan
Kabupaten Cilacap	69,95	11,46	428316
Kabupaten Banyumas	71,98	13,26	494857
Kabupaten Purbalingga	68,97	15,9	430509
Kabupaten Banjarnegara	67,45	15,64	470529
Kabupaten Kebumen	69,81	17,59	453599
Kabupaten Purworejo	72,68	11,78	429922
Kabupaten Wonosobo	68,22	17,36	430 896
Kabupaten Magelang	69,87	11,27	433 844
Kabupaten Boyolali	74,25	10,18	457 040
Kabupaten Klaten	75,56	12,89	476 383
Kabupaten Sukoharjo	76,98	7,68	494 254
Kabupaten Wonogiri	70,25	10,86	422 895
Kabupaten Karanganyar	75,86	10,28	490 491
Kabupaten Sragen	73,95	13,38	506 860
Kabupaten Grobogan	69,87	12,46	451 270
Kabupaten Blora	68,84	11,96	467 340
Kabupaten Rembang	70,02	15,6	490 540
Kabupaten Pati	71,77	10,08	509 800
Kabupaten Kudus	75	7,31	491 693
Kabupaten Jepara	71,99	7,17	424 772
Kabupaten Demak	72,22	12,54	584 033
Kabupaten Semarang	74,1	7,51	531 461
Kabupaten Temanggung	69,57	9,96	396 295
Kabupaten Kendal	72,29	9,99	526 553
Kabupaten Batang	68,65	9,13	510 704
Kabupaten Pekalongan	69,63	10,19	523 869
Kabupaten Pemalang	66,32	16,02	521 620

Kabupaten/Kota	IPM	%Kemiskinan	Rata-Rata Konsumsi Makanan
Kabupaten Tegal	68,39	8,14	531 480
Kabupaten Brebes	66,11	17,03	466 924
Kota Magelang	78,99	7,58	532 707
Kota Surakarta	82,21	9,03	628 884
Kota Salatiga	83,14	4,94	634 461
Kota Semarang	83,05	4,34	717 494
Kota Pekalongan	74,98	7,17	572 653
Kota Tegal	75,07	7,8	672 593

BUKU ekonometrika ini disusun untuk memberikan penjelasan mengenai Mata Kuliah Ekonometrika yang sebagian besar diambil oleh mahasiswa dari Program Studi-Program Studi Ekonomi: Manajemen, Manajemen Syariah, Akuntansi, Akuntansi Syariah, Ekonomi Pembangunan, Ekonomi Islam, dan Perbankan. Buku ini adalah buku pengantar yang hanya menjabarkan dasar-dasar ekonometrika.

UNIVERSITAS ISLAM NEGERI
RADEN MAS SAID
SURAKARTA

ISBN:978-623-6666-48-7