



# BUKU AJAR EKONOMETRIKA

**Tim Penulis :**

Dr. Erric Wijaya, S.E., M.E

Dr. Degdo Suprayitno

Urwawuska Ladini, S.Stat., M.Si

Titin Agustin Nengsih, Ph.D

Dr. Sumiyarti, S.E., M.E

Sella Nofriska Sudrimo, S.Si., M.Si

Sri Yani Kusumastuti

Dr. Nurhayati, S.E., M.E

Dalizanolo Hulu, S.E., M.E., CRM., CRP

**SONPEDIA.COM**

PT. Sonpedia Publishing Indonesia

# BUKU AJAR EKONOMETRIKA

## **Tim Penulis :**

Dr. Erric Wijaya, S.E., M.E

Dr. Degdo Suprayitno

Urwawuska Ladini, S.Stat., M.Si

Titin Agustin Nengsih, Ph.D

Dr. Sumiyarti, S.E., M.E

Sella Nofriska Sudrimo, S.Si., M.Si

Sri Yani Kusumastuti

Dr. Nurhayati, S.E., M.E

Dalizanolo Hulu, S.E., M.E., CRM., CRP

**Penerbit**

**SONPEDIA.COM**

PT. Sonpedia Publishing Indonesia

# **BUKU AJAR EKONOMETRIKA**

## **Tim Penulis :**

Dr. Erric Wijaya, S.E., M.E  
Dr. Degdo Suprayitno  
Urwawuska Ladini, S.Stat., M.Si  
Titin Agustin Nengsih, Ph.D  
Dr. Sumiyarti, S.E., M.E  
Sella Nofriska Sudrimo, S.Si., M.Si  
Sri Yani Kusumastuti  
Dr. Nurhayati, S.E., M.E  
Dalizanolo Hulu, S.E., M.E., CRM., CRP

**ISBN : 978-623-8598-48-9**

## **Editor :**

Efitra

## **Penyunting :**

Nur Safitri

## **Desain sampul dan Tata Letak :**

Yayan Agusdi

## **Penerbit :**

PT. Sonpedia Publishing Indonesia

## **Redaksi :**

Jl. Kenali Jaya No 166 Kota Jambi 36129 Telp. +6282177858344

Email : [sonpediapublishing@gmail.com](mailto:sonpediapublishing@gmail.com)

Website : [www.buku.sonpedia.com](http://www.buku.sonpedia.com)

**Anggota IKAPI : 006/JBI/2023**

Cetakan Pertama, April 2024

Hak cipta dilindungi undang-undang  
Dilarang memperbanyak karya tulis ini dalam bentuk dan dengan  
cara Apapun tanpa ijin dari penerbit

## KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa, yang telah memberikan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan buku ini dengan baik. Buku ini berjudul "**BUKU AJAR EKONOMETRIKA**". Tidak lupa kami ucapkan terima kasih bagi semua pihak yang telah membantu dalam penulisan dan penerbitan buku ini.

Buku Ajar ini disusun sebagai buku panduan komprehensif yang menjelajahi kompleksitas dan mendalamnya tentang ilmu ekonometrika. Buku ini dapat digunakan oleh pendidik dalam melaksanakan kegiatan pembelajaran di bidang ilmu ekonometrika dan diberbagai bidang Ilmu terkait lainnya. Buku ini dapat digunakan sebagai panduan dan referensi mengajar mata kuliah ekonometrika dan menyesuaikan dengan Rencana Pembelajaran Semester tingkat Perguruan Tinggi masing-masing.

Secara garis besar, buku ajar ini pembahasannya mulai dari pengantar, ruang lingkup, dan pemodelan ekonometrika hingga teknik analisis seperti korelasi, regresi, multikolinieritas, dan heteroskedastisitas. Materi juga mencakup model-model regresi termasuk variabel dummy dan data panel, serta model persamaan simultan. Buku ajar ini disusun secara sistematis, ditulis dengan bahasa yang jelas dan mudah dipahami, dan dapat digunakan dalam kegiatan pembelajaran.

Buku ini mungkin masih terdapat kekurangan dan kelemahan. Oleh karena itu, saran dan kritik para pemerhati sungguh penulis harapkan. Semoga buku ajar ini memberikan manfaat dan menambah khasanah ilmu pengetahuan dalam pembelajaran.

Jakarta, April 2024

**Tim Penulis**

# DAFTAR ISI

<b>KATA PENGANTAR .....</b>	<b>ii</b>
<b>DAFTAR ISI.....</b>	<b>iii</b>
<b>KEGIATAN BELAJAR 1 PENGANTAR DAN KONSEP DASAR</b>	
<b>EKONOMETRIKA .....</b>	<b>1</b>
DESKRIPSI, KOMPETENSI DAN PETA KONSEP PEMBELAJARAN	
A. PENGERTIAN EKONOMETRIKA .....	2
B. JENIS-JENIS EKONOMETRIKA.....	5
C. METODOLOGI EKONOMETRIKA .....	7
D. STRUKTUR DATA EKONOMETRIKA .....	12
E. RANGKUMAN .....	13
F. TES FORMATIF .....	14
G. LATIHAN.....	14
<b>KEGIATAN BELAJAR 2 RUANG LINGKUP DAN PEMODELAN</b>	
<b>EKONOMETRIKA .....</b>	<b>16</b>
DESKRIPSI, KOMPETENSI DAN PETA KONSEP PEMBELAJARAN	
A. PENDAHULUAN EKONOMETRIKA.....	17
B. RUANG LINGKUP EKONOMETRIKA.....	18
C. DASAR DASAR MODEL EKONOMETRIK.....	19
D. PEMILIHAN MODEL DAN UJI HIPOTESIS.....	22
E. TANTANGAN DAN PELUANG DALAM EKONOMETRIKA .....	24
F. RANGKUMAN .....	25
G. TES FORMATIF .....	25
H. LATIHAN.....	26
<b>KEGIATAN BELAJAR 3 KORELASI.....</b>	<b>27</b>
DESKRIPSI, KOMPETENSI DAN PETA KONSEP PEMBELAJARAN	

A.	KONSEP DASAR KORELASI .....	28
B.	SIFAT-SIFAT KORELASI .....	29
C.	JENIS-JENIS KORELASI.....	29
D.	KORELASI PEARSON.....	30
E.	KORELASI RANK SPEARMAN.....	34
F.	UJI HIPOTESIS KORELASI.....	36
G.	RANGKUMAN .....	40
H.	TES FORMATIF .....	40
I.	LATIHAN.....	41
<b>KEGIATAN BELAJAR 4 REGRESI.....</b>		<b>42</b>
DESKRIPSI, KOMPETENSI DAN PETA KONSEP PEMBELAJARAN		
A.	PENDAHULUAN .....	43
B.	REGRESI LINEAR SEDERHANA.....	45
C.	REGRESI LINEAR BERGANDA .....	48
D.	RANGKUMAN .....	52
E.	TES FORMATIF .....	52
F.	LATIHAN.....	53
<b>KEGIATAN BELAJAR 5 MULTIKOLINEARITAS .....</b>		<b>54</b>
DESKRIPSI, KOMPETENSI DAN PETA KONSEP PEMBELAJARAN		
A.	PENGERTIAN MULTIKOLINEARITAS.....	55
B.	KONSEKUENSI MULTIKOLINEARITAS.....	58
C.	CARA MENDETEKSI.....	60
D.	CARA MENGATASI .....	63
E.	RANGKUMAN .....	65
F.	TES FORMATIF .....	65
G.	LATIHAN.....	67

<b>KEGIATAN BELAJAR 6 HETEROSKEDASTISITAS .....</b>	<b>68</b>
DESKRIPSI, KOMPETENSI DAN PETA KONSEP PEMBELAJARAN	
A. PENGERTIAN .....	69
B. KONSEP DASAR .....	69
C. DAMPAK HETEROSKEDASTISITAS .....	71
D. TEKNIK MENDETEKSI HETEROSKEDASTISITAS .....	72
E. TEKNIK MENGATASI HETEROSKEDASTISITAS .....	79
F. RANGKUMAN .....	81
G. TES FORMATIF .....	81
H. LATIHAN.....	82
<b>KEGIATAN BELAJAR 7 MODEL REGRESI DENGAN VARIABEL DUMMY.....</b>	<b>83</b>
DESKRIPSI, KOMPETENSI DAN PETA KONSEP PEMBELAJARAN	
A. PENGERTIAN VARIABEL DUMMY.....	84
B. SIFAT DARI VARIABEL DUMMY .....	85
C. MODEL ANOVA.....	86
D. PERHATIAN DALAM PENGGUNAAN VARIABEL DUMMY .....	95
E. RANGKUMAN .....	97
F. TES FORMATIF .....	98
G. LATIHAN.....	99
<b>KEGIATAN BELAJAR 8 MODEL REGRESI DATA PANEL .....</b>	<b>103</b>
DESKRIPSI, KOMPETENSI DAN PETA KONSEP PEMBELAJARAN	
A. PENGERTIAN REGRESI DATA PANEL .....	104
B. ESTIMASI MODEL DALAM REGRESI LINIER DATA PANEL.....	107
C. PEMILIHAN ESTIMASI MODEL REGRESI DATA PANEL .....	110
D. LANGKAH-LANGKAH INTERPRETASI OUTPUT MODEL REGRESI DATA PANEL .....	114

E.	RANGKUMAN .....	119
F.	TES FORMATIF .....	120
G.	LATIHAN.....	121
<b>KEGIATAN BELAJAR 9 MODEL PERSAMAAN SIMULTAN .....</b>		<b>123</b>
DESKRIPSI, KOMPETENSI DAN PETA KONSEP PEMBELAJARAN		
A.	PENGERTIAN MODEL SIMULTAN.....	124
B.	PERSAMAAN STRUKTURAL DAN REDUCE FORM.....	125
C.	IDENTIFIKASI PERSAMAAN SIMULTAN .....	128
D.	METODE KUADRAT TERKECIL TIDAK LANGSUNG .....	133
E.	METODE KUADRAT TERKECIL DUA TAHAP.....	134
F.	RANGKUMAN .....	136
G.	TES FORMATIF .....	137
H.	LATIHAN.....	138
<b>DAFTAR PUSTAKA .....</b>		<b>140</b>
<b>TENTANG PENULIS .....</b>		<b>145</b>

# **KEGIATAN BELAJAR 1**

## **PENGANTAR DAN KONSEP DASAR EKONOMETRIKA**

### **DESKRIPSI PEMBELAJARAN**

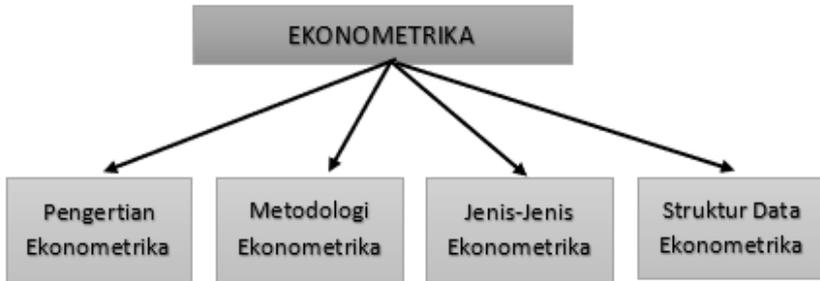
Pada bab ini mahasiswa mempelajari pengenalan dan konsep dasar dalam ekonometrika. Diharapkan mahasiswa memiliki wawasan dan pemahaman mengenai metodologi ekonometrika dan jenis-jenis data yang digunakan dalam menganalisis

### **KOMPETENSI PEMBELAJARAN**

Setelah mengikuti perkuliahan ini diharapkan mahasiswa dan mahasiswi memiliki pengetahuan dan kemampuan :

1. Mampu memahami dan menjelaskan definisi ekonometrika beserta jenis-jenisnya
2. Mampu memahami dan menjelaskan metodologi yang ada pada ekonometrika
3. Mampu memahami dan menjelaskan jenis-jenis ekonometrika yang ada
4. Mampu memahami dan menjelaskan struktur-struktur data ekonometrika

## PETA KONSEP PEMBELAJARAN



### A. PENGERTIAN EKONOMETRIKA

Dalam kehidupan sehari-hari aktivitas ekonomi dan sosial dilakukan oleh masyarakat. Dalam aktivitas ekonomi, masyarakat melakukan transaksi pembelian dan penjualan barang dan jasa. Aktivitas pembelian barang dan jasa ini dapat dipengaruhi oleh beberapa faktor, yaitu: tingkat harga, jumlah produksi, kepentingan akan mengkonsumsi suatu barang, tingkat pendapatan, selera, dan lain-lain. Keterkaitan antar faktor ini dapat dijelaskan dengan menggunakan pendekatan ekonometrika. Berdasarkan hal tersebut, maka ekonometrika secara sederhana dapat diartikan sebagai alat pengukuran ekonomi.

Terdapat beberapa definisi ekonometrika, yaitu:

- Arthur S Goldberger, *Econometrics may be defined as the social science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena*
- P.A. Samuelson, et, al, *Econometrics, the result of a certain outlook on the role of economics, consists of the application of mathematical statistics to economic data to lend empirical*

*support to the models constructed by mathematical economics and to obtain numerical results*

- Maddala, *Econometrics is the application of statistical and mathematical methods to the analysis of economic data, with the purpose of giving empirical content to economic theories and verifying them or refuting them.*

Secara umum, ekonometrika merupakan gabungan dari tiga bidang ilmu, yaitu ilmu ekonomi, ilmu matematika, dan ilmu statistik. Ilmu ekonomi digunakan untuk mendeskripsikan teori atau dalil yang terjadi dalam kehidupan sehari-hari. Biasanya dalam menjelaskan fenomena sehari-hari menggunakan suatu pernyataan atau hipotesa secara kualitatif. Pendekatan teori ekonomi mengambil peran yang penting dalam pendekatan ekonometrika. Misalkan teori yang umum dalam ilmu ekonomi adalah teori permintaan. Teori permintaan menyatakan bahwa jika tingkat harga mengalami kenaikan maka permintaan konsumsi barang dan jasa akan menurun, begitu pula sebaliknya dengan asumsi faktor lain tetap (*ceteris paribus*).

Pendekatan matematika digunakan untuk menyatakan pendekatan teori ekonomi dibuat dalam bentuk pendekatan matematika (persamaan matematika) meskipun teori ekonomi tersebut belum tentu dapat diukur atau belum diuji secara empiris. Pendekatan matematika ini digunakan dalam pendekatan ekonometrika untuk menguji secara empiris. Hal ini membutuhkan kemampuan khusus untuk mengkonversi pendekatan matematika menjadi pendekatan ekonometrika.

Pendekatan statistika bertujuan untuk mengumpulkan, mengolah, dan menyajikan data ekonomi dalam bentuk table, grafik dan lain-lain. Sehingga pengumpulan data merupakan proses pendekatan statistika yang utama. Data tersebut akan digunakan oleh ahli ekonometrika untuk dilakukan pengujian dan analisis dikaitkan dengan teori ekonomi yang telah ditentukan.

Pendekatan matematika dan statistika menyediakan banyak perangkat Analisa yang digunakan tetapi ahli ekonometrika seringkali harus menggunakan metode-metode tertentu untuk mengetahui hal-hal yang lebih khusus dari Sebagian besar data ekonomi. Ahli ekonometrika sama dengan ahli metodologi yaitu sangat bergantung kepada data yang kadangkala tidak dapat dikendalikan atau dikontrol.

Pendapat Spanos menyatakan bahwa “dalam ekonometrika, para pembuat model ekonometrika seringkali berhadapan dengan data yang bertentangan antara observasi dan hasil eksperimen. Keadaan ini memiliki dua implikasi atau dampak penting pada permodelan empiris dari ekonometrika. Pertama, para pembuat model harus menguasai banyak ketrampilan yang dibutuhkan untuk menganalisa data eksperimen. Kedua, oleh karena pengumpul data dan analisis data adalah orang yang berbeda, maka pembuat model harus memahami secara menyeluruh mengenai sifat-sifat dan struktur data yang digunakan.

Tujuan ekonometrika adalah sebagai berikut.

### **1. Analisis struktural**

Model ekonometrika yang dihasilkan harus dapat menunjukkan struktur hubungan antar variabel ekonomi yang menggambarkan kejadian ekonomi, menganalisis, dan menguji teori ekonomi berdasarkan data empiris.

### **2. Peramalan**

Model ekonometrika yang telah distrukturkan harus dapat meramalkan berbagai variabel ekonomi yang akan datang. Untuk dapat melakukannya maka model tersebut harus valid.

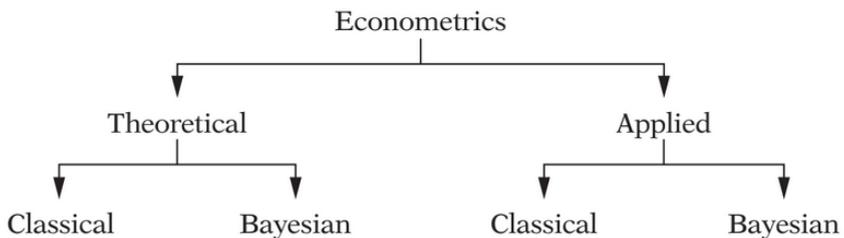
### **3. Evaluasi kebijakan**

Model ekonometrika yang telah valid dapat digunakan untuk mengevaluasi dan merencanakan kebijakan. Contohnya jika Perusahaan menaikkan harga produknya, maka akan dapat diketahui dampak jangka pendek dan jangka panjang terhadap

jumlah produksi, permintaan, keuntungan perusahaan, dan lain-lain.

## B. JENIS-JENIS EKONOMETRIKA

Jenis-jenis ekonometrika dapat dilihat pada gambar dibawah ini.



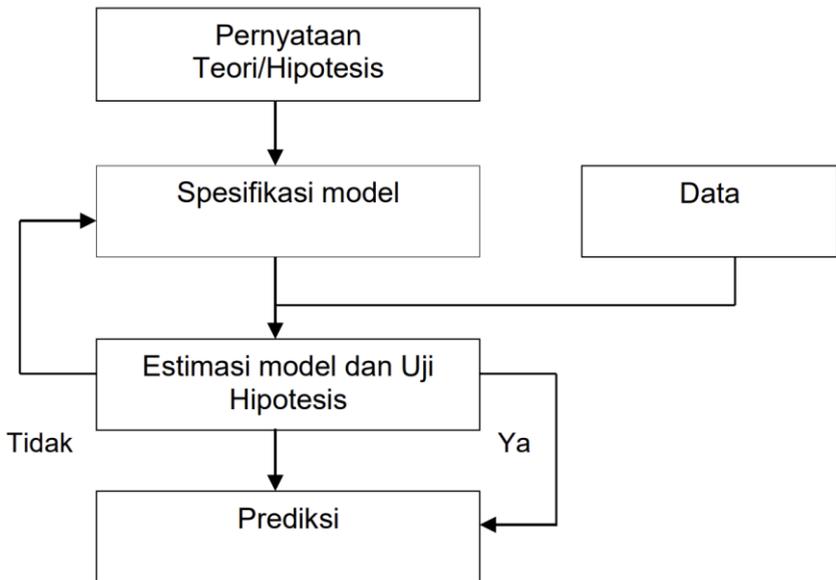
Sumber: Gujarati

Gambar 1.1 Jenis-Jenis Ekonometrika

Berdasarkan Gambar 1.1 Ilmu ekonometrika dibagi menjadi dua, yaitu ekonometrika teori (*theoretical econometrics*) dan ekonometrika aplikasi (*applied econometrics*). Pendekatan ekonometrika teori menjelaskan perkembangan metode-metode yang layak dipakai untuk mengukur perilaku ekonomi menggunakan model-model ekonometrika. Sehingga dapat disimpulkan bahwa ekonometrika teori sangat banyak menggunakan pendekatan matematika statistika.

Pendekatan ekonometrika aplikasi menggunakan teori ekonomi untuk mempelajari beberapa bidang khusus dalam ekonomi dan bisnis seperti fungsi produksi, fungsi permintaan, fungsi penawaran, fungsi konsumsi, dan lain-lain.

Pendekatan klasik memiliki alur yang dapat dilihat pada Gambar 1.2.



Sumber: Agus Wiharjono  
 Gambar 1.2. Alur Pendekatan Klasik

Pendekatan dimulai dari pernyataan teori atau hipotesis. Untuk membuktikan kebenaran teori atau hipotesis yang telah ditetapkan maka selanjutnya membuat suatu model ekonometrika. Setelah spesifikasi model dibangun maka langkah selanjutnya adalah melakukan estimasi parameter model tersebut berdasarkan data yang dikumpulkan. Kemudian setelah itu melakukan verifikasi bagi estimasi parameter melalui uji statistik. Uji statistik ini diperlukan karena estimasi model sebagian besar berasal dari data sampel. Uji statistik ini dengan demikian dilakukan untuk membuat sebuah generalisasi. Jika verifikasi ini sesuai dengan teori atau hipotesis yang ditetapkan di awal maka dapat langsung bisa menggunakan parameter estimasi tersebut untuk melakukan prediksi atau peramalan. Namun jika verifikasi ternyata tidak sesuai dengan teori atau hipotesis maka harus meninjau kembali spesifikasi model yang dibuat. Pembentukan model harus diulang kembali pada langkah kedua.

Pendekatan bayesian berdasarkan Teorema Bayes dari Pendeta Thomas Bayes (1701–1761), yang pertama kali memberikan persamaan yang memungkinkan bukti baru untuk memperbarui keyakinan. Bayes menulis essay menuju penyelesaian masalah dalam Ilmu Kemungkinan (1763). Teorema ini menjelaskan kemungkinan suatu peristiwa, berdasarkan pengetahuan sebelumnya tentang kondisi yang mungkin terkait dengan peristiwa tersebut. Misalnya, jika penyakit kanker terkait dengan usia, maka, dengan menggunakan teorema Bayes, usia seseorang dapat digunakan untuk menilai probabilitas kanker secara akurat, dibandingkan dengan penilaian kemungkinan kanker yang dibuat tanpa sepengetahuan usia seseorang. Penerapan metode Bayesian umumnya juga dilakukan pada sampel kecil dengan informasi yang terbatas. Penggunaan metode Bayesian melalui penefisian distribusi prior dapat menjadi solusi ketidaklengkapan informasi yang diperoleh dari data. Metode Bayesian memberikan hasil yang baik untuk ukuran sampel kecil dibandingkan metode ekonometrika lainnya, misalnya maksimum likelihood.

### **C. METODOLOGI EKONOMETRIKA**

Terdapat beberapa tahapan yang dilakukan dalam menggunakan pendekatan ekonometrika. Secara umum tahapan-tahap tersebut adalah sebagai berikut.

#### **1. Membuat pernyataan teori atau hipotesis**

Pertama-tama menentukan teori yang digunakan terkait hal yang akan diteliti. Teori yang ada dapat dibuktikan apakah sesuai dengan keadaan saat ini atau tidak. Dalam pembuktian teori menggunakan pendekatan hipotesa, misalkan pada teori permintaan, jika harga mengalami kenaikan maka permintaan akan menurun, begitupula sebaliknya jika harga mengalami penurunan maka permintaan akan meningkat dengan asumsi faktor lain tetap.

## 2. Mengumpulkan data

Pengumpulan data merupakan hal yang sangat penting dalam pendekatan ekonometrika. Terdapat istilah GIGO (*Garbage In Garbage Out*) yang berarti jika data yang digunakan tidak valid maka hasil yang akan didapatkan dari pendekatan ekonometrika juga akan bias. Sehingga data yang digunakan haruslah valid dari sumber atau referensi yang resmi dan dapat dipercaya. Data yang digunakan tidak hanya data kuantitatif (data berbentuk angka) tetapi juga dapat menggunakan data kualitatif (data tidak berbentuk angka) .

## 3. Spesifikasi teori dengan model matematika

Teori yang digunakan adalah teori permintaan, dimana diasumsikan permintaan hanya dipengaruhi oleh harga walaupun tidak hanya tingkat harga yang mempengaruhi permintaan. Teori ini diterjemahkan dalam bentuk model matematika sebagai berikut.

$$Q_d = \beta_1 + \beta_2 P \quad \beta_2 < 0$$

Dimana:

$Q_d$  = permintaan barang dan jasa

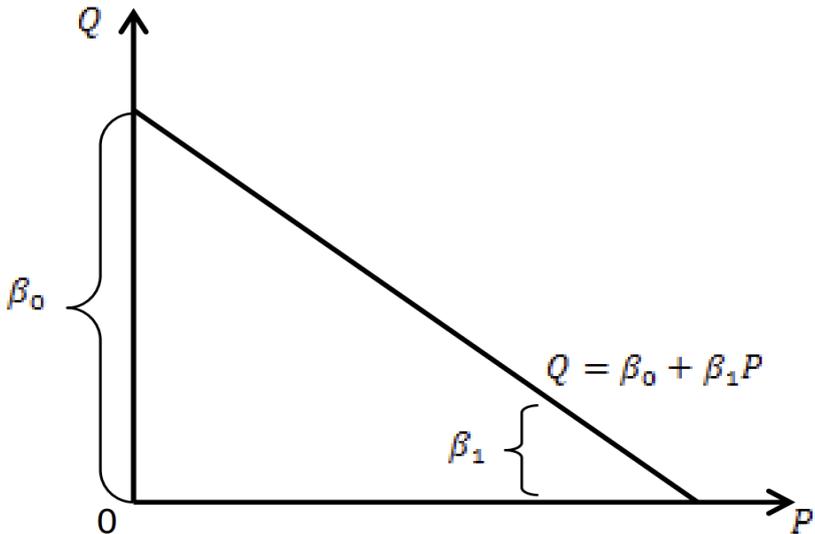
$P$  = tingkat harga

$\beta_1$  dan  $\beta_2$  = parameter

$\beta_1$  = intercept

$\beta_2$  = slope

Terdapat hubungan negatif antara permintaan dan tingkat harga yang ditunjukkan dengan adanya kemiringan garis atau slope yang negatif. Persamaan matematika tersebut dapat digambarkan dalam bentuk kurva sebagai berikut.



Gambar 1.3. Kurva Permintaan

#### 4. Spesifikasi teori dengan model ekonometrika

Pendekatan model matematika sebelumnya menganut prinsip hubungan yang deterministik (pasti) antara permintaan dan harga. Tetapi pada kenyataannya hubungan-hubungan ekonomi tersebut tidak pernah pasti. Hal ini disebabkan terdapat variabel-variabel lain selain harga yang mempengaruhi permintaan, misalkan pendapatan, selera, dan lain-lain. Untuk dapat menggambarkan hubungan yang tidak pasti diantara variabel-variabel ekonomi tersebut, maka model persamaan ekonometrika menjadi.

$$Q_d = \beta_1 + \beta_2 P + \varepsilon \quad \beta_2 < 0$$

Dimana:

$Q_d$  = permintaan barang dan jasa

$P$  = tingkat harga

$\beta_1$  dan  $\beta_2$  = parameter

$\beta_1$  = intercept

$\beta_2$  = slope

$\varepsilon$  = gangguan, kesalahan, atau galat

Pendekatan ekonometrika memasukkan variabel pengganggu atau kesalahan ini karena faktor yang mempengaruhi jumlah permintaan suatu barang dan jasa tidak hanya harga barang tersebut tetapi juga dipengaruhi variabel lain seperti pendapatan konsumen, selera, dan lain-lain. Adanya variabel  $\epsilon$  inilah menunjukkan bahwa model tersebut adalah model ekonometrika.

#### **5. Estimasi parameter dengan model ekonometrika yang dipilih**

Setelah menentukan model ekonometrika dan mendapatkan data yang valid sesuai dengan variabel yang akan diketahui, maka langkah berikutnya adalah mengestimasi parameter-parameter dari persamaan tersebut. Salah satu metode yang digunakan adalah analisis regresi linear untuk mendapatkan hasil estimasi. Misalkan hasil yang didapatkan menjadi.

$$Q_d = 400 - 0.02P$$

Simbol  $\beta_1$  dan  $\beta_2$  telah berganti dengan angka 400 dan -0.02 dimana angka ini akan diinterpretasikan yang akan dijelaskan pada bab-bab selanjutnya.

#### **6. Mengetahui akurasi model: pengujian spesifikasi model**

Pengujian ini dilakukan untuk mengetahui apakah model ekonometrika yang telah ditentukan apakah sudah tepat atau belum. Maka harus memastikan apakah model yang ditetapkan cukup menggunakan 1 variabel independen ataukah lebih dari satu variabel independen. Harus ada pertimbangan-pertimbangan khusus untuk memasukkan banyak variabel independen, karena jika memasukan semua variabel independen yang diduga mempengaruhi variabel dependen maka model ekonometrika yang dibuat akan sangat berat, sehingga pada akhirnya model yang dipilih adalah model replikasi dari model-model sebelumnya.

## 7. Menguji hipotesa dari model ekonometrika

Setelah yakin bahwa model ekonometrika yang ditetapkan sudah tepat, maka langkah selanjutnya menguji hipotesa yang telah ditentukan. Secara statistik, terdapat dua pengujian hipotesa yaitu uji t dan uji F. Pengujian uji t atau uji partial dilakukan untuk pengujian masing-masing variabel independen terhadap variabel dependen. Sedangkan uji F merupakan uji simultan dimana pengujian secara bersama-sama variabel independen terhadap dependen. Terdapat dua hasil dari pengujian ini, yaitu  $H_0$  ditolak (*rejected H<sub>0</sub>*) dan  $H_0$  tidak dapat ditolak (*don't rejected H<sub>0</sub>*). Hasil ini dapat digunakan sebagai rekomendasi temuan variable-variabel yang diteliti.

## 8. Menggunakan model untuk prediksi dan peramalan

Persamaan regresi yang telah ditentukan dapat digunakan untuk memprediksi atau peramalan. Peramalan digunakan untuk mengetahui seberapa besar nilai variabel dependen atas dasar nilai harapan di masa mendatang (*expected future value*) dari variabel independen. Misalkan hasil persamaan regresi sebagai berikut.

$$Q_d = 400 - 0.02P$$

Jika harga diperkirakan Rp. 10.000, maka.

$$Q_d = 400 - 0.02(10000) = 200$$

Hal ini berarti jika tingkat harga diperkirakan Rp.10.000, maka jumlah permintaan diperkirakan menjadi sebesar 200 unit. Harga-harga yang lain juga dapat dimasukkan kedalam persamaan ini untuk mengetahui berapa perkiraan permintaan. Selain itu, hasil estimasi bisa digunakan untuk membuat sebuah kebijakan bagi para pelaku ekonomi. Dari hasil estimasi tersebut bisa ditentukan berapa tingkat harga agar tingkat penghasilan (revenue) adalah maksimum.

## D. STRUKTUR DATA EKONOMETRIKA

Data ekonomi yang digunakan untuk melakukan analisa dengan metode pendekatan ekonometrika terdapat beberapa jenis. Beberapa metode pendekatan ekonometrika dapat menggunakan data yang ada dan juga ada data yang dimodifikasi dan dieksploitasi lebih lanjut. Berikut ini adalah beberapa struktur data yang utama dalam pendekatan ekonometrika.

### 1. **Data *cross section***

Merupakan data yang dikumpulkan pada satu titik waktu tertentu. Contoh data sensus, data kuesioner dapat dikategorikan data *cross section*, contoh lain data Tingkat inflasi bulan Januari 2024 masing-masing provinsi di Indonesia.

### 2. **Data *time series***

Merupakan data runtut waktu pada sekumpulan observasi. Contoh: data inflasi Indonesia dari tahun 2000 – 2023, data IHSG harian, data kurs dollar harian dan lain-lain. Secara umum data *time series* memiliki empat macam pola data, yaitu horizontal, trend, musiman, dan siklis (Hanke dan Wichren, 2005).

- a) Pola Data Horizontal, merupakan pola yang terjadi apabila data berfluktuasi di sekitar rata-rata yang konstan;
- b) Pola Data Trend terjadi apabila data mengalami kenaikan atau penurunan sekuler dalam jangka panjang.
- c) Pola Data Musiman merupakan pola data yang terjadi apabila terdapat fluktuasi dari data yang terjadi secara periodik dalam kurun waktu satu tahun, seperti triwulan, kuartalan, bulanan, mingguan, ataupun harian.
- d) Pola Data Siklis merupakan pola yang dipengaruhi oleh fluktuasi dari data untuk waktu yang lebih dari satu tahun.

### 3. **Data *pool cross section***

Beberapa data kadangkala memiliki karakteristik *cross section* dan *time series*. Contoh terdapat dua survey rumah

tangga pada tahun 2015 dan tahun 2020. Pada tahun 2015, rumah tangga diambil secara acak dalam survey yang berhubungan dengan pendapatan, jumlah anak, pendapatan, dan lain-lain. Pada tahun 2020, survei baru secara acak rumah tangga dengan pertanyaan yang sama. Pendekatan ini disebut *pooled cross section* dengan menggabungkan dua tahun survey tersebut..

#### 4. Data panel atau *longitudinal data*

Merupakan sekelompok data runtut waktu untuk masing-masing cross section pada sekumpulan data. Contoh, misalkan kita mempunyai riwayat gaji, pendidikan, dan pekerjaan untuk sekelompok individu selama periode sepuluh tahun. Contoh lain kita mengumpulkan informasi, seperti data investasi dan keuangan, tentang sekumpulan perusahaan yang sama dalam jangka waktu lima tahun. Perbedaan utama panel data dan pool cross section adalah unit *cross section* yang sama (individu, Perusahaan, atau wilayah) diikuti dengan runtut waktu selama periode tertentu.

## E. RANGKUMAN

Secara umum, ekonometrika merupakan gabungan dari tiga bidang ilmu, yaitu ilmu ekonomi, ilmu matematika, dan ilmu statistik. Ilmu ekonomi digunakan untuk mendeskripsikan teori atau dalil yang terjadi dalam kehidupan sehari-hari. Pendekatan matematika digunakan untuk menyatakan pendekatan teori ekonomi dibuat dalam bentuk pendekatan matematika (persamaan matematika) meskipun teori ekonomi tersebut belum tentu dapat diukur atau belum diuji secara empiris.. Pendekatan statistika bertujuan untuk mengumpulkan, mengolah, dan menyajikan data ekonomi dalam bentuk table, grafik dan lain-lain. Secara umum, Ilmu ekonometrika dibagi menjadi dua, yaitu ekonometrika teori (*theoretical econometrics*) dan ekonometrika aplikasi (*applied econometrics*). Dalam menggunakan pendekatan

ekonometrika, sangat dibutuhkan peran data, terdapat beberapa jenis data, yaitu data cross section, data time series, data pool cross section, dan data panel. Metodologi ekonometrika terdiri dari: 1. Membuat pernyataan teori atau hipotesis; 2. Mengumpulkan data; 3. Spesifikasi teori dengan model matematika; 4. Spesifikasi teori dengan model ekonometrika; 5. Estimasi parameter dengan model ekonometrika yang dipilih; 6. Mengetahui akurasi model: pengujian spesifikasi model; 7. Menguji hipotesa dari model ekonometrika; 8. Menggunakan model untuk prediksi dan peramalan.

## F. TES FORMATIF

1. Ilmu Ekonometrika merupakan gabungan dari ilmu dibawah ini, salah satunya yaitu ?
  - a) matematika
  - b) kimia
  - c) komunikasi
  - d) hukum
  - e) sosiologi
  
2. Tujuan ekonometrika adalah sebagai berikut, kecuali ?
  - a) analisis struktural
  - b) analisis statis
  - c) peramalan
  - d) evaluasi kebijakan
  - e) salah semua

## G. LATIHAN

Misalkan kita ingin membuktikan teori permintaan dimana tinggi rendahnya permintaan dipengaruhi oleh tinggi rendahnya tingkat harga, sehingga model regresi yaitu  $Demand = \beta_0 + \beta_1 Harga +$

$\mu$ . Terkait dengan model tersebut, menurut saudara, kita membutuhkan data-data apa saja yang harus digunakan ? dan termasuk jenis data apa ?

## KEGIATAN BELAJAR 2

### RUANG LINGKUP DAN PEMODELAN EKONOMETRIKA

#### DESKRIPSI PEMBELAJARAN

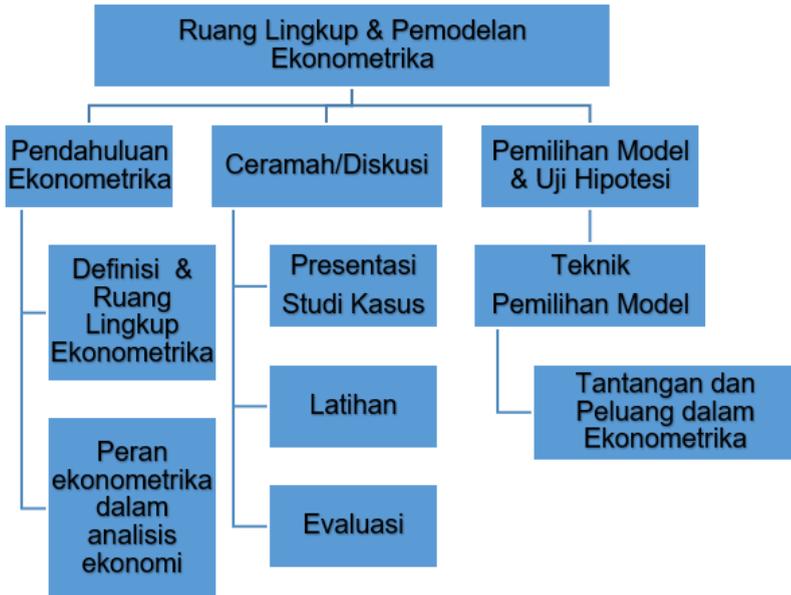
Fokus utama pada bab ini adalah memberikan pemahaman mendalam tentang teknik-teknik pemodelan dan analisis data untuk mengambil keputusan dalam konteks ekonomi. Mahasiswa akan belajar mengembangkan dan menguji model ekonometrik, serta menginterpretasikan hasil analisis statistik untuk mendukung pengambilan keputusan ekonomi yang lebih informasional.

#### KOMPETENSI PEMBELAJARAN

Setelah mengikuti kelas ini, mahasiswa diharapkan memiliki beberapa kompetensi utama:

1. Mengembangkan Pemahaman Dasar Ekonometrika: Mahasiswa akan memahami konsep dasar ekonometrika, termasuk variabel dependen dan independen, model linear, dan non-linear.
2. Pemilihan Model: Memahami proses pemilihan model yang tepat, termasuk identifikasi variabel relevan dan penggunaan kriteria pemilihan model.
3. Aplikasi dalam Konteks Ekonomi: Mengaplikasikan pemodelan ekonometrik untuk menganalisis masalah-masalah ekonomi konkret, seperti dampak kebijakan, prediksi variabel ekonomi, dan perencanaan ekonomi.

## PETA KONSEP PEMBELAJARAN



### A. PENDAHULUAN EKONOMETRIKA

#### ***Definisi Ekonometrika***

Ekonometrika adalah cabang ilmu ekonomi yang menggabungkan prinsip-prinsip ekonomi dengan metode-metode statistika untuk memodelkan dan menganalisis hubungan-hubungan kuantitatif dalam konteks ekonomi. Secara lebih spesifik, ekonometrika berfokus pada pengembangan model matematis yang mencerminkan hubungan antarvariabel ekonomi, serta penerapan metode statistika untuk menguji dan mengestimasi model tersebut. Dengan kata lain, ekonometrika memungkinkan ekonom untuk menyusun model yang dapat digunakan untuk memprediksi perilaku ekonomi, menguji hipotesis ekonomi, dan merumuskan kebijakan berdasarkan analisis data empiris.

## B. RUANG LINGKUP EKONOMETRIKA

Dengan ruang lingkup yang luas, ekonometrika memainkan peran kunci dalam memberikan dasar empiris untuk pengambilan keputusan ekonomi, membantu merancang kebijakan yang efektif, dan memahami dinamika ekonomi dalam masyarakat.

1. Model Ekonometrik: Linear vs. Nonlinear: Ekonometrika melibatkan pengembangan model matematis, baik yang bersifat linear maupun nonlinear, tergantung pada sifat hubungan antarvariabel yang diamati.
2. Variabel Dependan dan Independen: Ruang lingkup ekonometrika mencakup identifikasi variabel dependen (yang ingin dijelaskan) dan variabel independen (yang digunakan untuk menjelaskan).
3. Estimasi Parameter: Metode Kuadrat Terkecil (*OLS, Ordinary Least Square*): Ekonometrika menggunakan teknik estimasi parameter, khususnya metode kuadrat terkecil, untuk menentukan nilai-nilai optimal parameter dalam model.
4. Pemilihan Model: Identifikasi Variabel Relevan: Proses pemilihan model melibatkan identifikasi variabel yang paling relevan dalam menjelaskan variabel dependen.
5. Uji Hipotesis: Validitas Model: Ekonometrika mencakup pengujian hipotesis untuk memastikan bahwa model yang dikembangkan memiliki validitas statistik dan relevansi ekonomi.
6. Aplikasi dalam Keputusan Ekonomi:  
Studi Kasus: Ekonometrika digunakan untuk menganalisis dampak kebijakan ekonomi, memprediksi perilaku konsumen, dan menyusun model-model ekonometrik untuk mendukung pengambilan keputusan ekonomi.

## C. DASAR DASAR MODEL EKONOMETRIK

### 1. Variabel Dependen dan Independen

Variabel Dependen: Variabel dependen, juga dikenal sebagai variabel respons atau variabel terikat, adalah variabel dalam suatu studi atau model statistik yang nilainya bergantung pada atau dipengaruhi oleh variabel lain dalam analisis tersebut. Variabel dependen adalah fokus utama dalam penelitian atau analisis, dan tujuannya adalah untuk memahami atau memprediksi bagaimana perubahan nilai variabel independen akan mempengaruhi atau menyebabkan perubahan pada variabel dependen.

#### **Contoh:**

Dalam studi mengenai pendapatan rumah tangga, pendapatan rumah tangga dapat dianggap sebagai variabel dependen karena nilainya bergantung pada faktor-faktor lain seperti tingkat pendidikan, pekerjaan, dan jumlah anggota keluarga.

Variabel Independen: Variabel independen, juga dikenal sebagai variabel prediktor atau variabel bebas, adalah variabel yang dianggap sebagai penyebab atau variabel yang mempengaruhi variabel dependen. Dalam konteks analisis statistik atau ekonometrika, variabel independen digunakan untuk menjelaskan variasi atau perubahan yang terjadi pada variabel dependen.

#### **Contoh:**

Dalam studi pendapatan rumah tangga, variabel seperti tingkat pendidikan, jenis pekerjaan, dan jumlah anggota keluarga dapat dianggap sebagai variabel independen karena perubahan dalam variabel ini diharapkan mempengaruhi pendapatan rumah tangga. Dalam rumus matematis, variabel dependen (Y) sering dinyatakan sebagai fungsi dari satu atau lebih variabel independen (X), seperti dalam persamaan regresi:

$Y=f(X_1, X_2, \dots, X_n)$  Di mana  $Y$  adalah variabel dependen, dan

$X_1, X_2, \dots, X_n$  adalah variabel independen yang mempengaruhi  $Y$ .

## 2. Model Linear Dan Non-Linear

Model linear adalah jenis model matematis di mana hubungan antara variabel dependen dan satu atau lebih variabel independen dapat diungkapkan sebagai persamaan linear. Persamaan linear dapat ditulis dalam bentuk  $Y=a+bX$ , di mana  $Y$  adalah variabel dependen,  $X$  adalah variabel independen,  $a$  adalah intercept, dan  $b$  adalah koefisien kemiringan. Model linear umumnya lebih sederhana dan lebih mudah diinterpretasikan.

### Contoh:

$Income = \beta_0 + \beta_1 \times Education + \beta_2 \times Experience$

Model Non-Linear:

Model non-linear adalah jenis model matematis di mana hubungan antara variabel dependen dan variabel independen tidak dapat diungkapkan sebagai persamaan linear. Bentuk fungsi matematis yang digunakan dalam model non-linear dapat beragam dan kompleks. Model non-linear dapat memodelkan hubungan yang lebih kompleks dan fleksibel, namun interpretasinya mungkin lebih sulit dibandingkan dengan model linear.

**Contoh:**  $Sales = \alpha_0 + \alpha_1 \times Price + \alpha_2 \times Advertising + \alpha_3$

Dalam contoh ini, hubungan antara penjualan ( $Sales$ ) dengan harga ( $Price$ ) dan biaya iklan ( $Advertising$ ) bersifat non-linear karena melibatkan eksponen ( $\alpha_2$  dan  $\alpha_3$ ). Model non-linear sering digunakan ketika hubungan antarvariabel bersifat kompleks atau tidak dapat dijelaskan dengan baik oleh model linear.

### 3. Estimasi parameter dengan metode OLS

Metode OLS (Ordinary Least Squares) adalah suatu pendekatan statistik yang digunakan untuk mengestimasi parameter dalam model linier. Tujuan utama metode OLS adalah menemukan nilai-nilai parameter yang meminimalkan jumlah kuadrat selisih antara nilai-nilai yang diobservasi dan nilai-nilai yang diprediksi oleh model. Langkah-langkah Estimasi OLS sebagai berikut:

1. Penyusunan Model Linier: Model linier diwakili oleh persamaan  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$ , di mana  $Y$  adalah variabel dependen,  $X_1, X_2, \dots, X_n$  adalah variabel independen,  $\beta_0, \beta_1, \dots, \beta_n$  adalah parameter, dan  $\epsilon$  adalah kesalahan acak.
2. Penyesuaian Parameter: Penyesuaian parameter dilakukan dengan mencari nilai-nilai  $\beta_0, \beta_1, \dots, \beta_n$  yang meminimalkan jumlah kuadrat kesalahan (sum of squared errors) antara nilai observasi dan nilai yang diprediksi.
3. Menggunakan Kriteria Kuadrat Terkecil: Metode OLS memilih parameter-parameter sedemikian rupa sehingga jumlah kuadrat kesalahan ( $\epsilon\epsilon$ ) menjadi minimum. Ini dikenal sebagai kriteria kuadrat terkecil atau Least Squares Criterion.
4. Penghitungan Koefisien: Koefisien  $\beta_0, \beta_1, \dots, \beta_n$  dihitung dengan menggunakan rumus matematika yang mengoptimalkan fungsi kuadrat terkecil.
5. Evaluasi Model: Model yang diestimasi kemudian dievaluasi menggunakan berbagai metrik statistik, termasuk uji hipotesis, R-squared, dan lainnya, untuk memastikan validitas dan signifikansi model.

## D. PEMILIHAN MODEL DAN UJI HIPOTESIS

### 1. Teknik pemilihan model

Teknik pemilihan model ekonometrika adalah suatu proses untuk memilih variabel-variabel independen yang paling relevan dan model yang paling sesuai untuk menganalisis hubungan antarvariabel ekonomi. Pemilihan model merupakan langkah penting dalam pembuatan model ekonometrik, karena variabel yang kurang relevan atau model yang terlalu kompleks dapat menghasilkan hasil yang bias atau tidak dapat diinterpretasikan dengan baik. Berikut adalah beberapa teknik pemilihan model ekonometrika:

*Forward Selection:* Dimulai dengan variabel independen yang paling signifikan secara statistik dan secara bertahap menambahkan variabel satu per satu hingga tidak ada variabel lagi yang dapat meningkatkan signifikansi model.

*Backward Elimination:* Dimulai dengan seluruh variabel independen yang ada dalam model dan secara bertahap menghilangkan variabel yang paling tidak signifikan satu per satu hingga hanya tinggal variabel yang signifikan.

*Stepwise Regression:* Gabungan dari *forward selection* dan *backward elimination*, di mana variabel dapat ditambahkan atau dihilangkan pada setiap langkah pemodelan berdasarkan signifikansinya.

*Criteria Information (AIC, BIC):* Menggunakan kriteria informasi seperti AIC (*Akaike Information Criterion*) atau BIC (*Bayesian Information Criterion*) untuk memilih model terbaik. Model dengan nilai AIC atau BIC yang lebih rendah dianggap lebih baik.

*Validasi Silang (Cross-Validation):* Memisahkan dataset menjadi subset pelatihan dan subset pengujian. Model

dievaluasi berdasarkan performanya pada subset pengujian untuk memastikan generalitas model terhadap data baru.

**Uji Fungsi Kesalahan (LRT):** Menggunakan uji statistik, seperti Uji Likelihood Ratio (LRT), untuk membandingkan model yang lebih sederhana dengan model yang lebih kompleks. Model yang lebih kompleks dapat diterima jika memberikan peningkatan signifikan dalam penjelasan variabilitas.

**Bootstrap:** Melakukan bootstrap untuk menghasilkan sejumlah sampel acak dari dataset dan membangun model pada setiap sampel. Melihat sejauh mana hasilnya konsisten di antara sampel-sampel tersebut.

**Analisis Varians Inflasi Faktor (VIF):** Memeriksa multikolinearitas dengan menganalisis faktor inflasi varians. Variabel dengan VIF yang tinggi dapat dianggap redundan dan dieliminasi. Pemilihan model ekonometrika bukanlah proses yang serba mutlak, dan teknik yang digunakan dapat bervariasi tergantung pada konteks dan tujuan analisis. Penting untuk memilih model yang sesuai dengan teori ekonomi, menghindari overfitting, dan mempertimbangkan asumsi-asumsi yang mendasari pemodelan ekonometrik.

## **2. Uji hipotesis terkait dengan model validitas**

Uji hipotesis terkait dengan model ekonometrika memiliki peran penting dalam mengevaluasi kebermaknaan variabel-variabel independen, menilai validitas model, dan memastikan hasil analisis dapat diandalkan. Beberapa uji hipotesis umum yang sering digunakan terkait dengan model ekonometrika adalah:

- Uji Hipotesis Koefisien (t-Test) Uji t digunakan untuk menentukan apakah variabel independen memiliki dampak yang signifikan pada variabel dependen.
- Uji Keseluruhan Model (Uji F) Uji F digunakan untuk menentukan signifikansinya model secara keseluruhan.

- Uji Hipotesis pada Model Nonlinear (Wald Test): Wald Test digunakan untuk menguji kebermaknaan parameter dalam model non-linear.
- Uji Normalitas Residual (Jarque-Bera Test): Jarque-Bera Test digunakan untuk menguji apakah residual dari model memiliki distribusi normal.
- Uji Homoskedastisitas (Breusch-Pagan Test): Breusch-Pagan Test digunakan untuk menguji apakah varians residual konstan di sepanjang rentang nilai variabel independen.
- Uji Multikolinearitas (VIF atau Uji Kondisi Indeks Varians): Uji Varians Inflasi Faktor (VIF) atau Uji Kondisi Indeks Varians digunakan untuk mengidentifikasi adanya multikolinearitas.
- Uji Kesalahan Residual Independen (Durbin-Watson Test): Durbin-Watson Test digunakan untuk mendeteksi adanya autokorelasi dalam residual.

Penting untuk mencatat bahwa uji hipotesis ini bersifat kontekstual dan pemilihannya bergantung pada karakteristik spesifik dari model ekonometrika yang sedang dianalisis. Sebelum mengambil keputusan berdasarkan hasil uji hipotesis, peneliti harus memahami asumsi-asumsi yang mendasari setiap uji dan mempertimbangkan kebermaknaan hasil tersebut dalam konteks analisis keseluruhan.

## **E. TANTANGAN DAN PELUANG DALAM EKONOMETRIKA**

Tantangan: Aspek-aspek seperti heteroskedastisitas, multikolinearitas, dan asumsi-asumsi lainnya merupakan tantangan yang harus diatasi dalam pemodelan ekonometrik.

Peluang: Penggunaan big data dan inovasi teknologi memberikan peluang baru dalam meningkatkan akurasi dan kompleksitas model ekonometrik.

## F. RANGKUMAN

Ruang lingkup dan pemodelan ekonometrika melibatkan pemahaman mendalam tentang konsep-konsep dasar, aplikasi dalam konteks ekonomi, serta tantangan dan peluang yang dihadapi dalam analisis data ekonometrik.

Dengan demikian, pembelajaran ruang lingkup dan pemodelan ekonometrika memberikan landasan yang kokoh bagi mahasiswa untuk memahami, menerapkan, dan mengatasi tantangan dalam analisis data ekonometrik, sekaligus membuka pintu menuju inovasi dalam bidang ini.

## G. TES FORMATIF

1. Apa tujuan utama dari pemodelan ekonometrik?
  - a. **Menjelaskan sebab-akibat dalam ekonomi.**
  - b. Meramalkan perubahan nilai tukar mata uang.
  - c. Menghasilkan rumus matematis yang kompleks.
  - d. Menyederhanakan kompleksitas ekonomi.
  
2. Apa yang dimaksud dengan variabel independen dalam model ekonometrik?
  - a. Variabel yang nilainya diprediksi oleh model.
  - b. Variabel yang dijelaskan oleh model.
  - c. Variabel yang berubah secara otomatis.
  - d. **Variabel yang mempengaruhi variabel lain dalam model.**

3. Mengapa heteroskedastisitas dan multikolinearitas dianggap sebagai tantangan dalam analisis ekonometrik?
  - a. **Kedua kondisi tersebut dapat menyebabkan hasil estimasi yang tidak efisien.**
  - b. Kedua kondisi tersebut tidak mempengaruhi validitas hasil analisis.
  - c. Hanya heteroskedastisitas yang dapat mempengaruhi hasil analisis.
  - d. Multikolinearitas hanya terjadi jika ada hubungan linier sempurna antar variabel.

## H. LATIHAN

1. Jelaskan apa yang dimaksud dengan Varians Inflasi Faktor (VIF) dan bagaimana VIF digunakan untuk mengidentifikasi adanya multikolinearitas dalam model ekonometrik.
2. Berikan contoh model ekonometrik yang bersifat non-linear dan jelaskan mengapa model tersebut dipilih untuk menganalisis suatu fenomena ekonomi tertentu.
3. Sebuah perusahaan ingin menggunakan model ekonometrik untuk merencanakan strategi pemasaran. Bagaimana Anda akan menjelaskan manfaat penggunaan model ini dalam mendukung pengambilan keputusan pemasaran?

## KEGIATAN BELAJAR 3

### KORELASI

#### DESKRIPSI PEMBELAJARAN

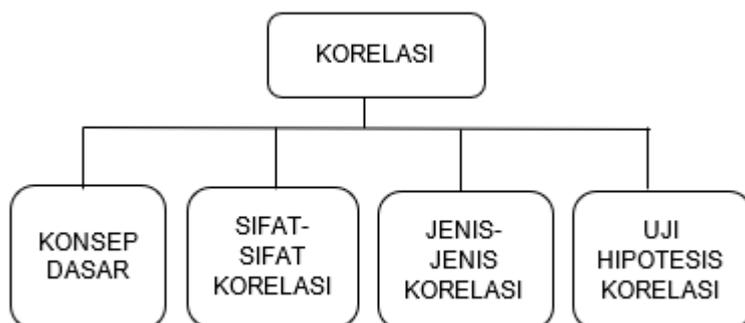
Pada bab ini mahasiswa mempelajari konsep dasar dan analisis korelasi. Diharapkan mahasiswa memiliki wawasan dan pemahaman untuk modal dasar mempelajari metode statistika untuk penelitian dibidang ekonomi lebih lanjut.

#### KOMPETENSI PEMBELAJARAN

Setelah mengikuti perkuliahan ini diharapkan mahasiswa dan mahasiswi memiliki pengetahuan dan kemampuan :

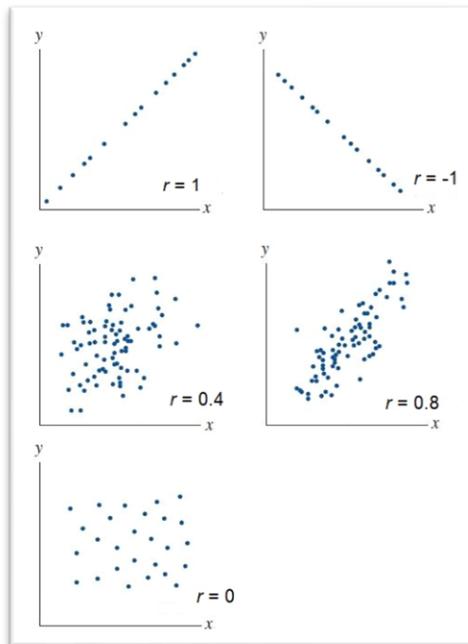
1. Mampu memahami konsep dasar korelasi
2. Mampu melakukan analisis korelasi
3. Mampu menginterpretasikan korelasi

#### PETA KONSEP PEMBELAJARAN



## A. KONSEP DASAR KORELASI

Dalam ekonometrika, metode analisis statistika yang sering digunakan salah satunya adalah analisis korelasi. Korelasi digunakan untuk mengetahui arah dan kekuatan (tren linear/ garis lurus) hubungan linear antar dua variabel. Nilai korelasi diukur oleh nilai koefisien korelasi yang dilambangkan dengan  $r$  untuk data sampel atau  $\rho$  untuk data populasi dengan besar nilainya yaitu  $-1 \leq r \leq +1$  atau  $-1 \leq \rho \leq +1$ . Nilai korelasi yang positif menunjukkan hubungan yang positif, sedangkan nilai korelasi yang negatif menunjukkan hubungan yang negatif. Semakin dekat nilai korelasi dengan nilai 1, maka semakin dekat titik-titik data terhadap garis lurus yang artinya semakin kuat hubungan linearnya, sedangkan semakin dekat *nilai* korelasi dengan nilai 0, maka semakin lemah hubungan linearnya. Berikut ini perbedaan gambar *scatterplot* berdasarkan nilai korelasinya.



Gambar 3.1: Beberapa Scatterplot beserta Nilai Korelasinya

Korelasi  $r$  bernilai  $-1$  dan  $+1$  hanya jika titik-titik data mengikuti pola garis lurus secara sempurna seperti terlihat pada grafik teratas pada **Gambar 3.1**. Ketika  $r = +1$  terjadi, garisnya miring ke atas dan hubungannya positif. Sebaliknya, ketika  $r = -1$ , garisnya miring ke bawah dan hubungannya negatif.

## **B. SIFAT-SIFAT KORELASI**

Korelasi mempunyai sifat-sifat sebagai berikut.

1. Korelasi selalu berada diantara  $-1$  dan  $+1$ . Semakin dekat nilainya dengan  $-1$  atau  $1$ , maka semakin kuat hubungan linear (garis lurus), karena titik-titik data berada lebih dekat ke garis lurus, sedangkan semakin dekat nilainya dengan  $0$ , maka kedua variabel cenderung semakin tidak memiliki hubungan.
2. Korelasi positif menunjukkan hubungan positif, sedangkan korelasi negatif menunjukkan hubungan negatif.
3. Nilai korelasi tidak bergantung pada satuan variabel. Misalnya, satu variabel adalah pendapatan penduduk dalam rupiah. Jika kita mengubah pengamatan dalam kedalam satuan dolar atau ribuan rupiah, kita akan mendapatkan hasil korelasi yang sama.
4. Analisis korelasi antara dua variabel tidak memperhatikan mana antara variabel bebas dan variabel terikat, karena hanya melihat besar dan arah hubungan kedua variabel tersebut.

## **C. JENIS-JENIS KORELASI**

Ada beberapa jenis analisis korelasi, diantaranya:

1. Korelasi Pearson

Ukuran hubungan antar dua variabel yang menggunakan data rasio dapat dihitung menggunakan korelasi Pearson.

## 2. Korelasi Rank Spearman

Ukuran hubungan antar dua variabel yang menggunakan data ordinal dapat dihitung menggunakan korelasi Rank Spearman.

### D. KORELASI PEARSON

Misalkan, dua variabel yaitu variabel  $X$  dan variabel  $Y$ , maka nilai korelasi kedua variabel tersebut dapat dihitung menggunakan rumus pada **Tabel 1** berikut ini.

**Tabel 1.** Rumus Koefisien Korelasi

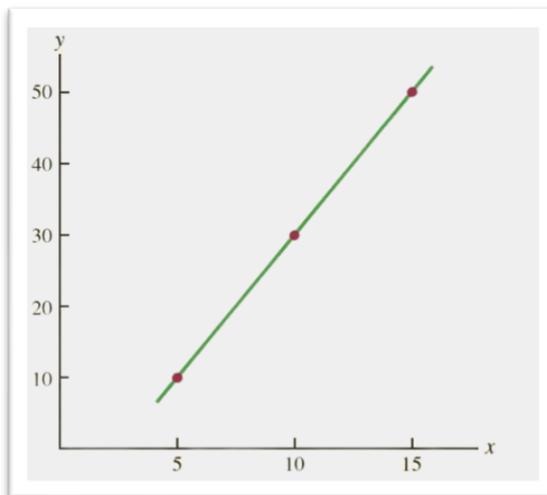
Data	Rumus Korelasi
Sampel	$r_{xy} = \frac{\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n - 1}}{\sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}} \sqrt{\frac{\sum(y_i - \bar{y})^2}{n - 1}}}$ <p>Ket: <math>x_i</math> = nilai observasi ke-<math>i</math> pada variabel <math>X</math> <math>y_i</math> = nilai observasi ke-<math>i</math> pada variabel <math>Y</math> <math>\bar{x}</math> = nilai rata-rata sampel variabel <math>X</math> <math>\bar{y}</math> = nilai rata-rata sampel variabel <math>Y</math> <math>n</math> = banyaknya sampel</p>
Populasi	$\rho_{xy} = \frac{\frac{\sum(x_i - \mu_x)(y_i - \mu_y)}{N}}{\sqrt{\frac{\sum(x_i - \mu_x)^2}{N}} \sqrt{\frac{\sum(y_i - \mu_y)^2}{N}}}$ <p>Ket: <math>x_i</math> = nilai observasi ke-<math>i</math> pada variabel <math>X</math> <math>y_i</math> = nilai observasi ke-<math>i</math> pada variabel <math>Y</math> <math>\mu_x</math> = nilai rata-rata populasi variabel <math>X</math> <math>\mu_y</math> = nilai rata-rata populasi variabel <math>Y</math> <math>N</math> = banyaknya populasi</p>

**Data 1** merupakan contoh sederhana yang mengilustrasikan konsep hubungan linear positif sempurna, dengan data sampel seperti berikut.

Data 1

$x_i$	5	10	15
$y_i$	10	30	50

Gambar *scatterplot* yang menunjukkan hubungan antara variabel  $x$  dan  $y$  adalah sebagai berikut.



*Gambar 3.2: Scatterplot yang menggambarkan Hubungan Linear Positif Sempurna*

Untuk menerapkan rumus pada **Tabel 1** guna menghitung korelasi sampel, berikut ini tahapan perhitungan seperti dibawah ini:

				Total ( $\Sigma$ )	Rata-rata
$x_i$	5	10	15	30	$\bar{x} = 10$
$y_i$	10	30	50	90	$\bar{y} = 30$
$x_i - \bar{x}$	-5	0	5	0	

$(x_i - \bar{x})^2$	25	0	25	50	
$y_i - \bar{y}$	-20	0	20	0	
$(y_i - \bar{y})^2$	400	0	400	800	
$(x_i - \bar{x})(y_i - \bar{y})$	100	0	100	200	

$$\begin{aligned}
 r_{xy} &= \frac{\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}}{\sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}} \sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}}} \\
 &= \frac{\frac{200}{3-1}}{\sqrt{\frac{50}{3-1}} \sqrt{\frac{800}{3-1}}} \\
 &= \frac{100}{\sqrt{25}\sqrt{400}} = \frac{100}{5 \times 20} = 1
 \end{aligned}$$

Secara umum dapat ditunjukkan bahwa jika semua titik dalam suatu kumpulan data berada pada garis lurus dengan kemiringan positif, maka nilai koefisien korelasi sampel adalah +1 dimana koefisien korelasi sampel sebesar +1 menunjukkan hubungan linear positif sempurna antara  $x$  dan  $y$ . Apalagi jika titik-titik dalam kumpulan data terletak pada garis lurus yang kemiringannya negatif, maka nilai koefisien korelasi sampelnya adalah -1 dimana koefisien korelasi sampel -1 menunjukkan hubungan linear negatif sempurna antara  $x$  dan  $y$ .

Misalkan kumpulan data tertentu menunjukkan hubungan linier positif antara  $x$  dan  $y$  tetapi hubungan tersebut tidak sempurna. Nilai  $r_{xy}$  akan lebih kecil dari 1, menunjukkan bahwa titik-titik pada diagram sebar tidak semuanya berada pada satu garis lurus. Ketika titik-titik semakin menyimpang dari hubungan linear positif sempurna, nilai  $r_{xy}$  menjadi semakin kecil. Nilai  $r_{xy}$  yang sama dengan nol menunjukkan tidak adanya hubungan linier antara  $x$  dan  $y$ , dan nilai  $r_{xy}$  yang mendekati nol menunjukkan hubungan linear yang lemah.

Korelasi memberikan ukuran hubungan linear dan belum tentu sebab akibat. Korelasi yang tinggi antara dua variabel tidak berarti perubahan pada satu variabel akan menyebabkan perubahan pada variabel lainnya. Misalnya, kita mungkin menemukan bahwa peringkat kualitas dan harga makanan khas di restoran berkorelasi positif. Namun, menaikkan harga makanan di restoran saja tidak akan menyebabkan peningkatan peringkat kualitas.

Contoh lain misalkan pada Data 2 berikut ini:

Data 2

Minggu ke-	Jumlah Iklan (X)	Volume Penjualan (Y)
1	2	50
2	5	57
3	1	41
4	3	54
5	4	54
6	1	38
7	5	63
8	3	48
9	4	59
10	2	46

akan dihitung nilai korelasi antara jumlah iklan dengan volume penjualan seperti dibawah ini:

											Total ( $\Sigma$ )	Rata- rata
$x_i$	2	5	1	3	4	1	5	3	4	2	30	$\bar{x} = 3$
$y_i$	50	57	41	54	54	38	63	48	59	46	510	$\bar{y} = 51$
$x_i - \bar{x}$	-1	2	-2	0	1	-2	2	0	1	-1	0	
$(x_i - \bar{x})^2$	1	4	4	0	1	4	4	0	1	1	20	
$y_i - \bar{y}$	-1	6	-10	3	3	-13	12	-3	8	-5	0	
$(y_i - \bar{y})^2$	1	36	100	9	9	169	144	9	64	25	566	
$(x_i - \bar{x})(y_i - \bar{y})$	1	12	20	0	3	22	24	0	8	5	99	

$$\begin{aligned}
 r_{xy} &= \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n - 1} \\
 &= \frac{99}{\sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}} \sqrt{\frac{\sum(y_i - \bar{y})^2}{n - 1}}} \\
 &= \frac{10 - 1}{\sqrt{\frac{20}{10 - 1}} \sqrt{\frac{566}{10 - 1}}} \\
 &= \frac{11}{(1.49)(7.93)} = 0.93
 \end{aligned}$$

Jadi, korelasi antara jumlah iklan dengan volume penjualan sebesar 0.93.

## E. KORELASI RANK SPEARMAN

Ukuran hubungan nonparametrik antar dua variabel  $X$  dan  $Y$  ditunjukkan oleh koefisien korelasi rank spearman, dengan rumus berikut ini.

$$r_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2$$

dimana  $d_i$  adalah selisih antara ranking yang diberikan pada  $x_i$  dan  $y_i$ , dan  $n$  adalah jumlah pasangan data.

**Data 3** berikut ini adalah data kandungan tar dan nikotin dari beberapa merk rokok.

Data 3

Merk	Kandungan Tar	Kandungan Nikotin
A	14	0.9
B	17	1.1
C	28	1.6
D	17	1.3

E	16	1.0
F	13	0.8
G	24	1.5
H	25	1.4
I	18	1.2
J	31	2.0

Pertama-tama kita tetapkan ranking pada setiap rangkaian pengukuran, dengan ranking 1 diberikan pada angka terendah pada, ranking 2 pada angka terendah kedua, dan seterusnya, hingga ranking 10 ditetapkan pada pengukuran.

Berikut ini tabel yang menunjukkan ranking masing-masing pengukuran dan perbedaan ranking untuk sepuluh pasang pengamatan.

Merk	Kandungan Tar ( $x_i$ )	Kandungan Nikotin ( $y_i$ )	$d_i$	$d_i^2$
A	2.0	2.0	0.0	0.0
B	4.5	4.0	0.5	0.25
C	9.0	9.0	0.0	0.0
D	4.5	6.0	-1.5	2.25
E	3.0	3.0	0.0	0.0
F	1.0	1.0	0.0	0.0
G	7.0	8.0	-1.0	1.0
H	8.0	7.0	1.0	1.0
I	6.0	5.0	1.0	1.0
J	10.0	10.0	0.0	0.0
Total				5.5

$$r_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2 = 1 - \frac{6(5.5)}{10(10^2 - 1)} = 0.967$$

menunjukkan korelasi positif yang tinggi antara kandungan tar dan nikotin yang ditemukan dalam rokok.

## F. UJI HIPOTESIS KORELASI

Misalkan pada Data 2 sebelumnya ingin dibuktikan apakah terdapat korelasi yang signifikan antara jumlah iklan dengan volume penjualan dengan taraf nyata 5% atau 0.05. Berikut ini tahapan uji hipotesisnya.

Hipotesis:

$$H_0 : \rho = 0$$

$$H_1 : \rho \neq 0$$

Tingkat signifikansi:

$$\alpha = 0.05$$

Statistik uji:

$$t_{hit} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.93\sqrt{10-2}}{\sqrt{1-0.93^2}} = \frac{2.63}{\sqrt{0.135}} = 7.16$$

Daerah kritis:

Tolak  $H_0$  jika  $t_{hit} > t_{tabel}$

Keputusan:

Tolak  $H_0$  karena  $t_{hit} > 2.306$  (lihat Tabel Nilai Kritis Distribusi-T)

Kesimpulan:

Dengan tingkat 95% terbukti bahwa jumlah iklan berkorelasi dengan volume penjualan secara signifikan.

**Tabel 2.** Tabel Nilai Kritis Distribusi-T

$v$	$\alpha$						
	0.40	0.30	0.20	0.15	0.10	0.05	0.025
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.538	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
$\infty$	0.253	0.524	0.842	1.036	1.282	1.645	1.960

Misalkan pada Data 3 sebelumnya ingin dibuktikan apakah terdapat korelasi positif yang signifikan antara kandungan tar dan nikotin pada rokok dengan taraf nyata 1% atau 0.01. Berikut ini tahapan uji hipotesisnya.

Hipotesis:

$$H_0 : \rho = 0$$

$$H_1 : \rho > 0$$

Tingkat signifikansi:

$$\alpha = 0.01$$

Statistik uji:

$$r_s = 0.967$$

Daerah kritis:

$$\text{Tolak } H_0 \text{ jika } r_s > r_{\text{tabel}}$$

Keputusan:

Tolak  $H_0$  karena  $r_s > 0.745$  (lihat Tabel Koefisien Korelasi Rank Spearman)

Kesimpulan:

Dengan tingkat 99% terbukti bahwa kandungan tar berkorelasi positif dengan kandungan nikotin dalam rokok.

**Tabel 3.** Tabel Nilai Kritis Koefisien Korelasi Spearman

$n$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$	$\alpha = 0.005$
5	0.900			
6	0.829	0.886	0.943	
7	0.714	0.786	0.893	
8	0.643	0.738	0.833	0.881
9	0.600	0.683	0.783	0.833
10	0.564	0.648	0.745	0.794
11	0.523	0.623	0.736	0.818
12	0.497	0.591	0.703	0.780
13	0.475	0.566	0.673	0.745
14	0.457	0.545	0.646	0.716
15	0.441	0.525	0.623	0.689
16	0.425	0.507	0.601	0.666
17	0.412	0.490	0.582	0.645
18	0.399	0.476	0.564	0.625
19	0.388	0.462	0.549	0.608
20	0.377	0.450	0.534	0.591
21	0.368	0.438	0.521	0.576
22	0.359	0.428	0.508	0.562
23	0.351	0.418	0.496	0.549
24	0.343	0.409	0.485	0.537
25	0.336	0.400	0.475	0.526
26	0.329	0.392	0.465	0.515
27	0.323	0.385	0.456	0.505
28	0.317	0.377	0.448	0.496
29	0.311	0.370	0.440	0.487
30	0.305	0.364	0.432	0.478

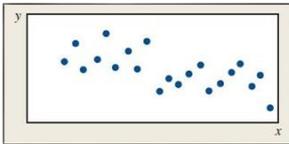
## G. RANGKUMAN

Berdasarkan uraian diatas, analisis korelasi merupakan salah satu metode yang dapat digunakan dalam analisis di bidang ekonomi, untuk melihat arah dan kekuatan hubungan antara dua variabel yang diukur oleh koefisien korelasi, dimana semakin mendekati nilai  $-1$  atau  $+1$ , maka korelasi antar dua variabel semakin kuat, baik pada hubungan positif maupun negatif, sedangkan semakin mendekati nilai  $0$ , maka kedua variabel tersebut cenderung semakin tidak memiliki hubungan. Namun, korelasi antara dua variabel ini tidak memperhatikan hubungan sebab-akibat.

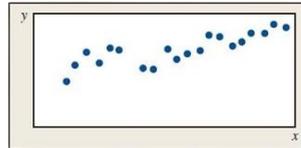
## H. TES FORMATIF

Perhatikan gambar *scatterplot* berikut ini:

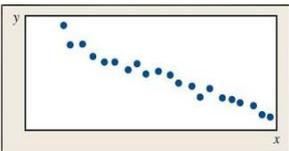
A.



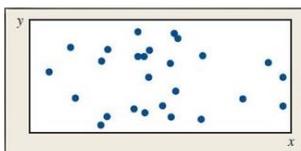
B.



C.



D.



1. Dari gambar diatas manakah yang mempunyai pola hubungan paling kuat?
  - a. A
  - b. B
  - c. C
  - d. D
2. Dari gambar diatas manakah yang mempunyai pola hubungan linear negatif?
  - a. A dan B
  - b. B dan C
  - c. C dan D
  - d. D dan A

- b. B dan C      d. B dan D

## I. LATIHAN

Panel konsumen menguji sembilan merk oven microwave untuk mengetahui kualitas keseluruhan. Peringkat yang ditetapkan oleh panel dan harga eceran yang disarankan adalah sebagai berikut.

Pabrik	Peringkat	Harga (ribu rupiah)
A	6	480
B	9	395
C	2	575
D	8	550
E	5	510
F	1	545
G	7	400
H	4	465
I	3	420

Hitunglah nilai koefisien korelasi antara kualitas oven dan harganya!

## KEGIATAN BELAJAR 4

### REGRESI

#### DESKRIPSI PEMBELAJARAN

Pada bab ini mahasiswa mempelajari pengenalan dan konsep dasar Regresi Linear baik Regresi Linear Sederhana ataupun Regresi Linear Berganda. Diharapkan mahasiswa memiliki wawasan dan pemahaman untuk melakukan pemodelan dalam suatu kasus atau persoalan yang ada.

#### KOMPETENSI PEMBELAJARAN

Setelah mengikuti perkuliahan ini diharapkan mahasiswa dan mahasiswi memiliki pengetahuan dan kemampuan :

1. Mampu menguraikan definisi Regresi Linear baik Regresi Linear Sederhana ataupun Regresi Linear Berganda.
2. Mampu menjelaskan fungsi model dari Regresi Linear Sederhana ataupun Regresi Linear Berganda.
3. Mampu melakukan prediksi pemodelan Regresi Linear baik Regresi Linear Sederhana ataupun Regresi Linear Berganda.

#### PETA KONSEP PEMBELAJARAN



## A. PENDAHULUAN

Regresi mempunyai definisi secara umum adalah peramalan penaksiran atau pendugaan. Regresi pertama kali diperkenalkan pada tahun 1877 oleh Sir Francis Galton (1822-1911). Analisis Regresi digunakan untuk menentukan bentuk dari hubungan antara beberapa variabel. Adapun tujuan utama dalam penggunaan analisis regresi adalah untuk meramalkan atau memperkirakan nilai dari suatu variabel dalam hubungannya dengan variabel yang lain.

Dalam kehidupan sehari-hari sering kita berjumpa atau terlibat dalam suatu sistem baik sederhana maupun kompleks dan berhubungan dengan dua atau lebih peubah baik kualitatif maupun kuantitatif. Hubungan pada suatu sistem ini dapat kita definisikan dengan suatu model matematis. Salah satu model ini berfungsi untuk peramalan/prediksi keadaan/hasil yang akan datang, dan penjelasan hubungan sebab-akibat antara satu peubah terhadap peubah lainnya.

Permasalahan muncul ketika model yang dirancang tidak menghasilkan sesuatu yang sesuai atau diharapkan dengan melihat pada pertimbangan secara teoritis dan pengalaman sebelumnya. Dalam Statistika terdapat analisis yang memanfaatkan hubungan antara dua atau lebih peubah kuantitatif sehingga salah satu peubah dapat diramalkan dari peubah lainnya, ini disebut Analisis Regresi. Dengan menggunakan analisis ini diharapkan model yang dirancang tidak jauh menyimpang/mendekati dari kenyataannya.

Analisis regresi dapat dibedakan berdasarkan:

1. Jumlah variabel dalam model
  - a. Analisis Regresi Sederhana (Simple regression analysis) : mempelajari ketergantungan atau variabel tak bebas hanya pada suatu variabel bebas.

- b. Analisis Regresi Berganda (Multiple regression analysis) : mempelajari ketergantungan atau variabel tak bebas pada lebih dari satu variabel bebas.
2. Jumlah persamaan dalam model
- a. Analisis regresi dengan persamaan tunggal : hanya ada satu persamaan yang digunakan dalam model.
  - b. Analisis regresi dengan persamaan simultan : terdapat lebih dari satu persamaan yang digunakan dalam model.
3. Sifat linearitas
- a. Analisis regresi linier : suatu persamaan regresi di mana semua koefisien parameter dan semua variabel yang digunakan ( baik variabel bebas maupun variabel tak bebas) dalam persamaan tersebut bersifat biner.
  - b. Analisis regresi non linear : jika salah satu atau lebih variabel yang digunakan khusus *Dependent variable* dalam persamaan tersebut bersifat Non linier, tetapi koefisien parameter tetap bersifat linear.

Alasan penggunaan sifat linearitas pada model berdasarkan pada beberapa alasan :

- 1. Banyak fenomena ekonomi yang pada dasarnya dapat diestimasi secara linear.
- 2. Banyak model yang dapat ditransformasi ke dalam model linear.
- 3. Fungsi linear dapat pula dipakai untuk mengaproksimasi persamaan linear dengan menggunakan teknik-teknik tertentu, seperti teknik ekspansi dari Taylor meskipun persamaannya bersifat non linier, tetapi selama parameternya linear maka dapat diestimasi dengan model linear.

## B. REGRESI LINEAR SEDERHANA

Belajar analisis yang paling sederhana adalah analisis regresi dua variabel, yaitu suatu model di mana hanya ada satu variabel bebas dan satu variabel tak bebas yang dinyatakan sebagai fungsi linear. Salah satu contoh ada suatu daerah dengan jumlah total penduduk (populasi) sebanyak 70 keluarga. Misalkan kita berminat untuk mempelajari pengaruh antara belanja konsumsi keluarga bulanan dengan pendapatan keluarga yang dapat dibelanjakan (disposable income) atau pendapatan setelah dipotong pajak bulanan. Lebih spesifik lagi, asumsikan bahwa kita ingin meramalkan rata-rata (populasi) tingkat belanja konsumsi bulanan dengan mengetahui pendapatan bulanan keluarga itu. Oleh karena itu kita bisa membuat suatu bentuk model dari dua hal yang akan kita teliti.

Regresi linier sederhana adalah yang memiliki satu variabel dependen dan satu variabel independen. Modal persamaan regresi linier sederhana dengan rumus sebagai berikut :

$$Y = a + bX + e$$

dimana :

Y = subyek dalam variabel dependen yang diprediksi

a = harga Y ketika harga X = 0 (harga konstan)

b = angka arah atau koefisien regresi, yang menunjukkan angka peningkatan ataupun penurunan variabel dependen yang didasarkan pada perubahan variabel independen.

Bila (+) arah garis naik, dan bila (-) maka arah garis turun.

e = *error terms*

Adapun perhitungan koefisien regresi linear sederhana *a* dan *b* dapat diformulasikan sebagai berikut:

$$b = \frac{n\sum(xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$a = \frac{\sum y - b(\sum x)}{n}$$

$$se = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n - 2}}$$

$$sb = \frac{se}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}}$$

$$thitung = \frac{b}{sb}$$

Asumsi-asumsi dasar yang sering digunakan dalam analisis regresi yang disebut dengan asumsi klasik dalam metode OLS (Ordinary Least Squares) yang sangat sering dilanggar di dalam melakukan estimasi sebuah model regresi. Hal ini menyebabkan parameter yang diperoleh menjadi menyimpang atau bias atau jauh dari harapan, tidak konsisten, dan tidak efisien. Apabila dalam analisis regresi tidak didasarkan pada analisis yang benar maka akan mengakibatkan hasil pendugaan regresi akan menyimpang dari harapan. Misalnya, apabila dalam peubah bebas terjadinya kolinieritas ganda yang sempurna akan menyebabkan matriks  $X'X$  menjadi singular, sehingga tidak mempunyai determinan dan akibatnya koefisien regresi tidak dapat diduga.

Untuk melakukan analisis regresi yang benar berdasarkan metode OLS, maka diperlukan asumsi-asumsi yang harus dipenuhi di antaranya adalah:

1. Nilai tengah (mean value) dari komponen pengganggu  $u_i$ , yang ditimbulkan variabel eksplanatori atau variabel bebas  $X$  harus sama dengan nol.
2. Varians dari komponen pengganggu  $u_i$  harus konstan atau harus memenuhi syarat homoskedastisitas atau setiap variabel bebas  $X$  mempunyai varians komponen pengganggu  $u_i$  harus sama.
3. Tidak terjadi autokorelasi antar komponen pengganggu  $u_i$  atau harus konstan atau tidak terjadi korelasi antar  $X_t$  dengan  $X_{t+1}$  dan seterusnya.
4. Variabel eksplanatori atau variabel bebas  $X$  nilainya harus non stokastik atau apabila bersifat stokastik harus menyebar bebas dari komponen pengganggu.
5. Tidak terjadi multikolinieritas antar variabel eksplanatori atau variabel bebas  $X$
6. Komponen pengganggu  $u_i$  harus menyebar menurut sebaran normal dengan nilai tengah  $\mu = 0$  dengan varians sebesar  $\sigma^2$  hal ini kalau terpenuhi, maka asumsi 1) dan asumsi 2) secara otomatis telah terpenuhi. .

Dengan keenam asumsi tersebut di atas dapat diketahui bahwa estimator OLS dari koefisien regresi  $\beta_i$  ternyata bersifat BLUE (Best Linier Unbias Estimator), dan atas dasar asumsi normalitas maka estimator–estimator tersebut akan menyebar mengikuti sebaran normal. Sehingga, hasilnya memungkinkan untuk mendapatkan suatu kisaran atau range yang dapat diuji kebenarannya terhadap koefisien regresi populasi  $\beta_i$ . Dalam uraian selanjutnya tidak akan dibahas lebih mendalam untuk asumsi-asumsi pertama, keempat, dan keenam, karena pelanggaran asumsi tersebut tidaklah serius akibatnya dalam analisis regresi.

### C. REGRESI LINEAR BERGANDA

Regresi linear berganda adalah regresi dimana variabel terikatnya (Y) dihubungkan atau dijelaskan lebih dari satu variabel, mungkin dua, tiga dan seterusnya variabel bebas ( $x, x_1, x_2, \dots, x_n$ ) namun masih menunjukkan diagram hubungan yang linear.

Penambahan variabel bebas ini diharapkan dapat lebih menjelaskan karakteristik hubungan yang ada walaupun masih saja ada variabel yang terabaikan.

Bentuk umum dari persamaan linear berganda dapat ditulis sebagai berikut:

$$\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 \dots \dots \dots b_kx_k + e$$

Keterangan

$\hat{y}$  : Variabel terikat (nilai duga y)

a,  $b_1, b_2, b_3, \dots, b_k$  : koefisien regresi

$x_1, x_2, x_3, \dots, x_k$  : variabel bebas

e : kesalahan pengganggu

Pada analisis regresi, guna menghindari ketidaksesuaian model maka digunakan metode kuadrat terkecil yang mendasarkan pada asumsi-asumsi berikut:

1.  $\varepsilon_i$  menyebar saling bebas mengikuti sebaran normal ( $\varepsilon_i \sim NID(0, \sigma^2)$ );
2. Ragam dari  $\varepsilon_i$  konstan atau  $V(\varepsilon_i) = \sigma^2$  untuk  $i=1,2,\dots,n$  (homoskedastisitas);
3. Tidak ada autokorelasi sisaan atau  $Cov(\varepsilon_i, \varepsilon_j) = 0$  dimana  $i \neq j$ ;
4. Tidak ada multikolinieritas di antara peubah penjelas (X) atau  $Cov(X_i, X_j) = 0$  dimana  $i \neq j$ .

## Pendugaan dan Pengujian Koefisien Regresi

Kesalahan baku atau selisih taksir standar regresi adalah nilai menyatakan seberapa jauh menyimpangnya nilai regresi tersebut terhadap nilai sebenarnya. Nilai ini digunakan untuk mengukur tingkat ketepatan suatu pendugaan dalam menduga nilai. Jika nilai ini sama dengan nol maka penduga tersebut memiliki tingkat ketepatan 100%.

Kesalahan baku atau selisih taksir standar regresi berganda dirumuskan

$$S_e = \sqrt{\frac{\sum y^2 - b_1(\sum x_1y) + b_2(\sum x_2y)}{n - m}}$$

Keterangan

$S_e$  : Kesalahan baku regresi berganda

$n$  : Jumlah pasangan observasi

$m$  : jumlah konstant dalam persamaan regresi berganda.

Untuk koefisien  $b_1$  dan  $b_2$  kesalahan bakunya dirumuskan

$$S_{b_1} = \frac{S_e}{\sqrt{(\sum x_1^2 - n\bar{x}_1^2)(1 - r^2y_1)}}$$

$$S_{b_2} = \frac{S_e}{\sqrt{(\sum x_2^2 - n\bar{x}_2^2)(1 - r^2y_1)}}$$

Pendugaan interval koefisien regresi berganda (parameter  $B_1$  dan  $B_2$ )

Parameter  $B_1$  dan  $B_2$  sering juga disebut sebagai koefisien regresi parsial. Pendugaan parameter  $B_1$  dan  $B_2$  menggunakan distribusi

t dengan derajat bebas  $db = n - m$  secara umum pendugaan parameter  $B_1$  dan  $B_2$  adalah :

$$b_1 - t_{\alpha/2n-m} S_{b_i} \leq B_i \leq b_1 + t_{\alpha/2n-m} S_{b_i}$$

$$i = 2,3$$

Pengujian hipotesis koefisien regresi berganda (parameter  $B_1$  dan  $B_2$ )

Pengujian hipotesis bagi koefisien regresi berganda atau regresi parsial parameter  $B_1$  dan  $B_2$  dapat dibedakan menjadi 2 bentuk, yaitu pengujian hipotesis serentak dan pengujian hipotesis individual.

Pengujian hipotesis individual yaitu merupakan pengujian hipotesis koefisien regresi berganda dengan hanya satu  $B$  ( $B_1$  dan  $B_2$ ) yang mempunyai pengaruh  $Y$ . pengujian hipotesis serentak merupakan pengujian hipotesis koefisien regresi berganda dengan  $B_1$  dan  $B_2$  serentak atau bersama-sama mempengaruhi  $Y$ .

### **Peramalan dengan Regresi Linear Berganda**

Peramalan terhadap nilai  $Y$  dengan menggunakan regresi linear berganda, dapat dilakukan apabila persamaan garis regresinya sudah diestimasi dan nilai variabel bebas  $x_1, x_2$  sudah diketahui.

Suatu persamaan garis regresi linear berganda dapat dipakai dalam peramalan dengan terlebih dahulu melakukan pengujian hipotesis terhadap koefisien-koefisien regresi parsialnya. Tujuan ialah mengetahui variabel-variabel bebas yang digunakan itu memiliki pengaruh yang nyata atau tidak terhadap  $y$  tersebut. Variabel bebas  $x_1$  dan  $x_2$  disebut memiliki pengaruh yang nyata apabila dalam pengujian hipotesis koefisien parsialnya  $H_0 : B_1 = B_2 = 0$  ditolak atau  $H_1 : B_1 \neq B_2 \neq 0$  diterima, khususnya pada taraf nyata 1%

Kelebihan peramalan  $y$  dengan menggunakan regresi linear berganda adalah dapat diketahui besarnya pengaruh secara kuantitatif setiap variabel bebas ( $x_1$  atau  $x_2$ ) apabila pengaruh variabelnya dianggap konstan. Misalnya sebuah persamaan regresi berganda

$$y = a + b_1x_1 + b_2x_2$$

Keterangan :

$y$  : Nilai statistik mahasiswa

$x_1$  : Nilai inteligensi mahasiswa

$x_2$  : Frekuensi membolos mahasiswa

$b_1$  : Pengaruh  $x_1$  terhadap  $y$  jika  $x_2$  konstan

$b_2$  : Pengaruh  $x_2$  terhadap  $y$  jika  $x_1$  konstan

jika  $a = 17,547$ ;  $b_1 = 0,642$ ;  $b_2 = - 0,284$  maka persamaan regresi linear bergandanya menjadi

$$\hat{y} = 17,547 + 0,624 (x_1) - 0,284 (x_2)$$

Dengan persamaan regresi linear berganda tersebut, nilai  $y$  (nilai statistik maha siswa) dapat diramalkan dengan mengetahui nilai  $x_1$  (nilai inteligensi mahasiswa) dan  $x_2$  (frekuensi membolos mahasiswa) misalkan, nilai  $x_1 = 75$  dan  $x_2 = 24$  maka ramalan nilai  $y$  adalah

$$\begin{aligned}\hat{y} &= 17,547 + 0,624 (75) - 0,284 (24) \\ &= 63.211\end{aligned}$$

Penulisan persamaan garis regresi linear berganda biasanya disertai dengan kesalahan baku masing-masing variabel bebas dan koefisien determinasi berganda  $r^2$ , sebagai ukuran tepat atau tidaknya garis tersebut sehingga pendekatan.

## D. RANGKUMAN

Berdasarkan uraian di atas dapat di simpulkan bahwa analisis regresi membentuk persamaan garis lurus (linear) yang digunakan sebagai bentuk pemodelan variabel bebas dan satu variabel terikat. Analisis Regresi juga dapat digunakan sebagai persamaan untuk membuat perkiraan (*prediction*) nilai suatu variabel terikat jika nilai variabel bebas yang berhubungan dengannya sudah ditentukan. Jika variabel bebas terdiri hanya satu variabel, maka pengaruh dari variabel ini terhadap variabel terikat dapat menggunakan analisis regresi Linear Sederhana. Adapun analisis Regresi Linear Berganda digunakan untuk melihat pengaruh beberapa variabel bebas dan satu variabel terikat.

## E. TES FORMATIF

1. Analisis yang digunakan untuk melihat pengaruh satu variabel terikat dan satu variabel bebas adalah
  - a) Korelasi Sederhana
  - b) Korelasi Berganda
  - c) Regresi Linear sederhana
  - d) Regresi linear berganda
  - e) Salah semua
  
2. Analisis yang digunakan untuk melihat pengaruh satu variabel terikat dan beberapa variabel bebas adalah
  - a) Korelasi Sederhana
  - b) Korelasi Berganda
  - c) Regresi Linear sederhana
  - d) Regresi linear berganda
  - e) Salah semua

3. Salah satu fungsi analisis regresi adalah, kecuali?
  - a) Memodelkan
  - b) Mengidentifikasi pengaruh
  - c) Memprediksi
  - d) Membuat pola
  - e) Salah semua

## **F. LATIHAN**

1. Apakah perbedaan antara korelasi dan regresi? Jelaskan!
2. Tuliskan persamaan model regresi sederhana? Kemudian Jelaskan masing-masing komponen yang anda tulis.
3. Sebutkan tahapan-tahapan yang perlu dilakukan dalam analisis regresi sederhana? Jelaskan!
4. Apa yang anda ketahui mengenai metode OLS? Jelaskan!
5. Sebutkan asumsi-asumsi dalam regresi linier sederhana? Jelaskan!

## KEGIATAN BELAJAR 5

### MULTIKOLINEARITAS

#### DESKRIPSI PEMBELAJARAN

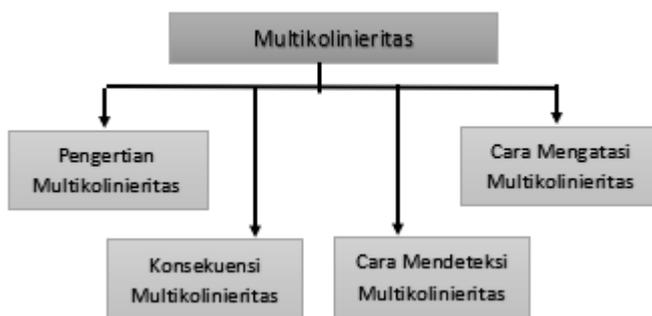
Pada bab ini mahasiswa mempelajari pengenalan dan konsep dasar teoritis multikolinearitas. Melalui pembelajaran ini diharapkan mahasiswa memiliki pengetahuan dan pemahaman terkait dengan multikolinieritas, serta bagaimana mendeteksi dan mengatasi jika terjadi multikolinearitas pada model regresi yang dihasilkan.

#### KOMPETENSI PEMBELAJARAN

Setelah mengikuti perkuliahan ini diharapkan mahasiswa dan mahasiswi memiliki pengetahuan dan kemampuan :

1. Memahami dan menjelaskan penegertian multikolinieritas.
2. Menjelaskan konsekuensi multikolinieritas pada model regresi.
3. Mendeteksi adanya multikolinieritas pada model regresi.
4. Mengatasi jika terdapat multikolinieritas pada model regresi.

#### PETA KONSEP PEMBELAJARAN



## A. PENGERTIAN MULTIKOLINEARITAS

Analisis regresi berganda dan interpretasi hasilnya dapat dipengaruhi secara signifikan oleh beberapa fenomena, salah satunya adalah multikolinieritas. Secara sederhana, multikolinieritas dapat diartikan sebagai adanya hubungan linier antarvariabel bebas (independen) dalam model regresi berganda. Istilah multikolinieritas merujuk kepada definisi yang dinyatakan oleh Ragnar Frisch (1934) sebagai *“the existence of “perfect” or exact linear relationship among some or all explanatory variables in regression model”*, yaitu adanya hubungan linier yang “sempurna”, atau eksak, di antara beberapa atau seluruh variabel penjelas model regresi.

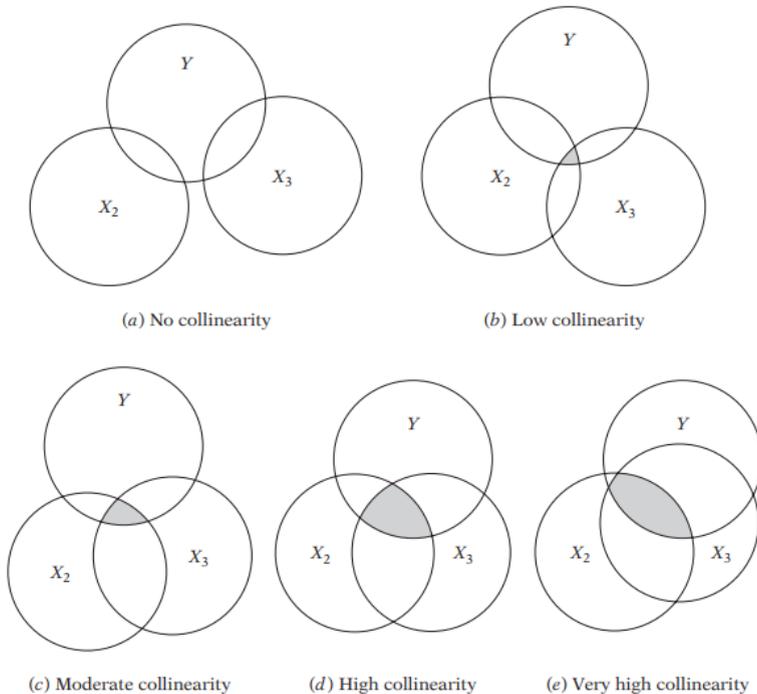
Multikolinieritas antarvariabel dapat dijelaskan dengan menggunakan ilustrasi data hipotesis sebagai berikut.

	$X_2$	$X_3$	$X_3^*$
	10	50	52
	15	75	75
	18	90	97
	24	120	129
	30	150	152

Berdasarkan tabel tersebut terlihat jelas bahwa  $X_3 = 5X_2$ . Oleh karena itu, terdapat kolinieritas sempurna antara  $X_2$  dan  $X_3$ , karena koefisien korelasi antara  $X_2$  dan  $X_3$  adalah satu kesatuan. Sementara variabel  $X_3^*$  dibuat dari variabel  $X_3$  dengan hanya menambahkan angka-angka 2,0,7,9,2 ke dalam angka berikutnya, yang diambil dari tabel secara acak. Antara variabel  $X_3$  dan  $X_3^*$  tidak berkorelasi secara sempurna. Namun jika dilakukan perhitungan, kedua variabel bebas tersebut akan memiliki koefisien korelasi dengan tingkatan atau derajat yang tinggi yakni sebesar 0,9959.

Setiap tingkatan atau derajat multikolinieritas memiliki konsekuensi yang berbeda pada proses analisis regresi beserta interpretasi hasilnya. Maka dari itu, menjadi hal yang penting untuk memahami derajat multikolinieritas yang terjadi dalam model regresi.

Derajat multikolinieritas dapat digambarkan menggunakan diagram sebagai berikut. Diasumsikan gambar lingkaran  $Y$ , mewakili variasi  $Y$  (variabel terikat), sedangkan gambar lingkaran  $X_2$ , dan  $X_3$  masing-masing mewakili variasi  $X_2$  serta  $X_3$  (variabel penjelas atau variabel bebas). Selanjutnya, derajat kolinieritas antarvariabel  $X_2$  dan  $X_3$  diukur dari besarnya wilayah tumpang tindih atau area yang diarsir dari lingkaran  $X_2$ , dan  $X_3$ .



*Gambar 5.1: The Ballentine view of multicollinearity*  
*Sumber: Gujarati & Porter (2009)*

Mengacu kepada gambar 4.1 tersebut, derajat multikolinieritas dapat dikategorikan menjadi 5 (lima) tingkatan. Berikut penjelasan secara singkat:

- a. Tidak ada kolinieritas (*No Collinearity*)  
Fenomena ini terjadi ketika tidak ada korelasi yang signifikan antara variabel-variabel independen dalam model regresi, ditunjukkan dengan tidak ada area terasir antara variabel  $X_2$  dan  $X_3$ . Setiap variabel independen secara unik memberikan informasi yang berbeda kepada model.
- b. Kolinieritas rendah (*Low Collinearity*)  
Fenomena ini terjadi dimana terdapat sedikit korelasi antara variabel-variabel independen ( $X_2$  dan  $X_3$ ), tetapi korelasinya tidak cukup kuat untuk mempengaruhi estimasi koefisien regresi. Pada kondisi ini, model regresi masih bisa diandalkan untuk dilakukan analisis.
- c. Kolinieritas sedang (*Moderate Collinearity*)  
Fenomena yang terjadi pada tingkatan ini terjadi ketika korelasi antarvariabel independen menjadi lebih kuat, menyebabkan koefisien regresi menjadi tidak stabil dan interval kepercayaan menjadi lebih lebar. Namun, dampaknya masih dapat dikelola dengan hati-hati dengan memeriksa asumsi dan hasil analisis.
- d. Kolinieritas tinggi (*High Collinearity*)  
Pada fenomena ini, korelasi antara variabel-variabel independen cukup kuat, sehingga estimasi koefisien regresi menjadi tidak stabil dan sulit untuk diinterpretasikan. Maka dari itu diperlukan analisis atau langkah-langkah yang lebih cermat untuk mengurangi multikolinieritas. Langkah ini perlu dilakukan untuk mendapatkan hasil regresi yang lebih dapat diandalkan.
- e. Kolinieritas sangat tinggi (*Very High Collinearity*)  
Fenomena ini terbilang sangat ekstrim karena setiap variabel independen sangat berkorelasi satu sama lain. Dalam hal ini, estimasi koefisien regresi menjadi sangat tidak stabil, dan interpretasi model menjadi sulit atau bahkan tidak mungkin.

dilakukan tanpa tindakan yang tepat untuk mengatasi multikolinearitas.

Multikolinieritas dapat disebabkan karena beberapa hal seperti penggunaan variabel yang terlalu mirip baik arti maupun pengukurannya dan variabel yang tidak relevan. Penggunaan sampel yang kecil juga dapat menjadi penyebab terjadinya multikolinieritas.

## B. KONSEKUENSI MULTIKOLINEARITAS

Konsekuensi atau dampak dari multikolinearitas dalam analisis regresi secara signifikan dapat mempengaruhi keandalan, validitas, serta interpretasi model regresi. Konsekuensi ini juga dapat bervariasi, tergantung pada tingkat keparahan dan konteks spesifik dari masalah yang sedang dihadapi.

Berikut merupakan beberapa konsekuensi atau dampak dari multikolinearitas:

### a. Varian koefisien regresi menjadi besar

Ketika variabel independen yang terdapat dalam model regresi sangat berkorelasi satu sama lain, maka dihasilkan varian yang tinggi dalam perkiraan koefisien regresi. Korelasi antar varian ini dapat dilihat melalui formula untuk mengukur besarnya varian  $b_1$  dibawah ini:

$$Var(b_1) = \frac{\sigma^2}{\sum x_{1i}(1 - r_{X_1X_2}^2)}$$

Dimana  $r_{X_1X_2}$  dalam formula tersebut menunjukkan korelasi antara variabel bebas  $X_1$  dan  $X_2$ . Formula tersebut menunjukkan bahwa semakin besar korelasi antara variabel bebas  $X_1$  dan  $X_2$  berakibat pada semakin besarnya varian. Nilai varian yang semakin besar artinya variabilitas semakin tinggi. Variabilitas yang tinggi ini yang kemudian membuat timbulnya kesulitan

dalam menarik kesimpulan yang dapat diandalkan dari model regresi.

- b. Interval kepercayaan (*confidence interval*) menjadi lebar  
Variansi yang besar menghasilkan standar error juga besar, berakibat pada interval kepercayaan yang lebar. Hal ini bisa dibuktikan berdasarkan formula pembuatan interval kepercayaan berikut:

$$b_1 \pm Z_{\alpha/2} \text{ s.e} (b_1)$$

Dikarenakan  $\text{s.e} (b_1)$  merupakan akar dari  $\text{Var} (b_1)$  maka dari itu apabila  $\text{Var} (b_1)$  besar maka  $\text{s.e} (b_1)$  juga besar. Dengan begitu interval yang dihasilkan juga akan besar. Besarnya nilai interval ini yang kemudian dapat menyebabkan adanya pengurangan ketepatan estimasi serta interpretasi hasil. Sebab koefisien yang diestimasi dapat bervariasi secara signifikan dalam sampel yang berbeda.

- c. Banyak variabel bebas yang tidak signifikan berdasarkan uji t, meskipun Uji- $F$  signifikan dan Koefisien determinasi ( $R^2$ ) tinggi. Pada dasarnya Uji- $F$  dilakukan guna mengukur apakah terdapat suatu variabel independen yang secara signifikan mempengaruhi variabel dependen. Uji- $F$  yang signifikan dalam multikolinieritas mungkin disebabkan dari total variasi dalam model masih cukup besar. Sedangkan koefisien determinasi  $R^2$  berfungsi untuk mengukur seberapa baik variabilitas dalam variabel dependen yang dijelaskan oleh model regresi. Nilai  $R^2$  yang tinggi dapat disebabkan oleh inklusi variabel yang sebenarnya berkorelasi dengan variabel dependen atau variabel lain dalam model karena multikolinearitas.

$$F = \frac{R^2}{1 - R^2} (n - k - 1)$$

Formula tersebut menunjukkan hubungan antara Uji- $F$  dengan  $R^2$ . Hubungan kedua ukuran tersebut adalah apabila  $R^2$  besar, maka nilai Uji- $F$  juga besar.

- d. Koefisien regresi bisa mempunyai nilai yang tidak sesuai dengan substansi, misalnya hubungannya seharusnya positif tetapi malah negatif, atau sebaliknya. Hal ini bisa menyesatkan interpretasi hasil regresi.

### C. CARA MENDETEKSI

Deteksi multikolinearitas merupakan sebuah langkah penting dalam analisis regresi. Sebab multikolinearitas dapat memiliki beberapa konsekuensi yang merugikan. Diantaranya seperti koefisien regresi yang tidak stabil, estimasi yang tidak akurat, serta ketidakpastian dalam interpretasi. Proses deteksi ini berfungsi untuk memastikan keandalan dan interpretasi yang tepat dari model.

Terdapat sejumlah metode atau pendekatan yang dapat dilakukan untuk mendeteksi multikolinearitas, diantaranya sebagai berikut:

1. Uji korelasi sederhana antara variabel-variabel independen  
Uji korelasi ini dapat dilakukan dengan menghitung koefisien korelasi antara setiap pasangan variabel independen. Jika terdapat korelasi yang kuat antara dua atau lebih variabel independen, ini bisa menjadi indikasi adanya multikolinearitas.

Persamaan untuk menghitung koefisien korelasi antara dua variabel  $x_i$  dan  $x_j$  adalah sebagai berikut:

$$r_{ij} = \frac{\sum(x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sqrt{\sum(x_i - \bar{x}_i)^2 \sum(x_j - \bar{x}_j)^2}}$$

Dimana:

- $r_{ij}$  adalah koefisien relasi antara variabel  $x_i$  dan  $x_j$
- $\bar{x}_i$  adalah rata-rata dari variabel  $x_i$
- $\bar{x}_j$  adalah rata-rata dari variabel  $x_j$

Melalui hasil yang diperoleh dari persamaan tersebut dapat diketahui apakah korelasi antara variabel independen kuat atau tidak. Dikatakan memiliki korelasi yang tinggi apabila  $|r_{ij}| > 0.7$ , sehingga hal ini dapat menunjukkan adanya multikolinearitas yang signifikan. Akan tetapi diperlukan juga perhitungan dengan metode lainnya agar hasilnya lebih akurat.

## 2. Analisis varians (ANOVA) atau Uji-F

Proses mendeteksi multikolinearitas dengan metode ini dapat dilakukan dengan uji hipotesis tentang signifikansi keseluruhan dari model regresi. Jika model secara keseluruhan tidak signifikan, ini dapat menjadi sebuah indikasi adanya multikolinearitas. Pengujian dilakukan menggunakan Uji-F, dengan hipotesis yaitu:

- $H_0: \beta_1 = 0$  , dimana tidak ada variabel independen yang memiliki dampak signifikan terhadap variabel dependen (hipotesis nol / H0)
- $H_A: \beta_1 \neq 0$  , dimana setidaknya satu variabel memiliki dampak yang signifikan terhadap variabel dependen (hipotesis alternatif / H1).

Jika nilai p dari uji F tidak signifikan (misalnya,  $p > \alpha$ , dengan  $\alpha$  sebagai tingkat signifikansi yang dipilih), ini menunjukkan bahwa model secara keseluruhan tidak signifikan. Dalam konteks deteksi multikolinearitas, ini dapat diartikan sebagai adanya masalah multikolinearitas, karena variabel independen secara keseluruhan tidak dapat menjelaskan variasi yang signifikan dalam variabel dependen.

3. Melihat nilai koefisien determinasi ( $R^2$ )

Multikolinearitas juga dapat dideteksi dengan metode perhitungan koefisien determinasi ( $R^2$ ), dimana tingkat penjelasan variasi dalam variabel dependen yang dihasilkan oleh model regresi perlu diperhatikan. Nilai koefisien determinasi ( $R^2$ ) dapat menginterpretasikan seberapa besar variasi dalam variabel dependen dapat dijelaskan oleh model regresi. Koefisien determinasi ( $R^2$ ) dapat dihitung melalui model persamaan sebagai berikut:

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

Dimana:

- SSR merupakan jumlah kuadrat dari residu model (*residual sum of squares*), yaitu jumlah kuadrat perbedaan antara nilai yang diprediksi oleh model dan nilai aktual dari variabel dependen
- SST merupakan jumlah kuadrat total (*total sum of squares*), yaitu jumlah kuadrat perbedaan antara nilai aktual dari variabel dependen dan rata-rata nilai dari variabel dependen.

$R^2$  dapat dikatakan tinggi apabila nilainya lebih dari 0,7 atau 0,8. Nilai yang tinggi ini dapat diartikan bahwa model secara keseluruhan dapat menjelaskan sebagian besar variasi dalam variabel dependen. Satu hal penting yang perlu diketahui yaitu  $R^2$  yang tinggi belum tentu menjadi indikasi adanya multikolinearitas. Tetapi jika nilai koefisien determinasi tinggi, tetapi koefisien regresi individu tidak signifikan, ini menjadi sebuah indikasi adanya multikolinearitas.

4. Uji statistik *VIF* (*Variance Inflation Factor*)

Uji statistik *VIF* digunakan untuk mengukur berapa banyak varians dari koefisien regresi yang diperbesar karena

multikolinearitas. Persamaan untuk menghitung  $VIF$  yakni sebagai berikut:

$$VIF_i = \frac{1}{1 - R_i^2}$$

Dimana:

- $VIF_i$  adalah Variance Inflation Factor untuk variabel independen ke-i
- $R_i^2$  adalah koefisien determinasi yang dihasilkan dari regresi variabel independen ke-i pada semua variabel independen lainnya

Melalui persamaan tersebut, uji statistik  $VIF$  dilakukan pada setiap variabel independen dalam model. Hasil dari uji statistik ini memiliki Indikasi multikolinearitas apabila nilai  $VIF$  yang diperoleh tinggi (biasanya di atas 10). Sebab hal ini menunjukkan bahwa variabel independen ke-i berkorelasi kuat dengan variabel lainnya.

Penting untuk menggunakan beberapa metode untuk mendeteksi multikolinearitas dan memeriksa konsistensi hasil. Jika multikolinearitas terdeteksi, langkah-langkah untuk mengatasi masalah tersebut harus dipertimbangkan. Seperti penghapusan variabel, transformasi data, atau penggunaan teknik regresi yang tahan terhadap multikolinearitas.

#### **D. CARA MENGATASI**

Multikolinearitas adalah kondisi di mana dua atau lebih variabel independen dalam model regresi memiliki korelasi tinggi satu sama lain. Hal Ini dapat menyebabkan masalah dalam interpretasi model regresi dan membuat estimasi koefisien menjadi tidak stabil. Mengatasi masalah multikolinearitas penting untuk memastikan bahwa model regresi atau model statistik lainnya dapat memberikan estimasi parameter yang konsisten, akurat,

dan dapat diandalkan, serta memastikan interpretasi yang tepat terhadap hubungan antara variabel independen dan dependen. Berikut adalah beberapa cara untuk mengatasi multikolinearitas:

1. Memilih variabel dengan bijak  
Perlu adanya pertimbangan yang matang dalam memilih variabel yang paling relevan dan penting untuk model regresi yang digunakan. Diperlukan juga penggunaan domain pengetahuan atau analisis eksploratif data untuk memilih variabel yang memiliki dampak yang signifikan terhadap variabel dependen. Variabel yang tidak relevan atau memiliki korelasi tinggi dengan variabel lain dapat dihapus dari model.
2. Transformasi Variabel  
Variabel yang dapat ditransformasi dalam hal ini adalah yang memiliki distribusi tidak normal atau memiliki hubungan non-linear dengan variabel dependen. Contoh transformasinya seperti transformasi logaritmik, akar kuadrat, atau kuadratik.
3. Penggabungan Variabel  
Beberapa variabel dapat digabungkan menjadi satu bila memungkinkan. Seperti apabila terdapat beberapa variabel yang mengukur konsep yang serupa, indeks atau skor gabungan dapat dibuat dengan merepresentasikan konsep tersebut. Dengan mengurangi jumlah variabel, hal ini memungkinkan untuk mengurangi potensi multikolinearitas.
4. Menambah jumlah data atau observasi juga menjadi salah satu cara untuk mengatasi multikolinieritas yang terjadi antar variabel bebas.

Setiap metode memiliki kelebihan dan kelemahan tersendiri, dan pemilihan metode tergantung pada konteks data, tujuan analisis, dan preferensi peneliti. Dalam prakteknya, seringkali kombinasi dari beberapa metode yang disebutkan di atas digunakan untuk mengatasi multikolinearitas.

## **E. RANGKUMAN**

Berdasarkan uraian yang telah disampaikan mulai dari pengertian, konsekuensi, cara mendeteksi sampai cara mengatasi, dapat dirangkumkan beberapa hal. Pertama, multikolinieritas merupakan fenomena yang tidak dapat dihindari dalam model regresi. Multikolinieritas adalah kondisi dimana dalam sebuah model regresi terjadi korelasi yang signifikan antarvariabel bebas. Derajat kolineritas antarvariabel bebas berbeda-beda pada setiap model regresi. Kedua, keberadaan multikolinieritas pada model regresi membawa dampak tidak dapat dilakukannya interpretasi hasil regresi, atau nilai koefisien regresi menjadi tidak stabil dan mempengaruhi keandalan sebuah model. Ketiga, untuk mendeteksi adanya multikolinieritas dalam model regresi dapat dilakukan dengan beberapa cara. Deteksi awal adanya multikolinieritas pada model regresi dapat diindikasikan dari nilai  $R^2$  yang tinggi, namun uji-t pada masing-masing koefisien regresi tidak signifikan. Keempat, multikolinieritas pada hasil regresi dapat menyesatkan interpretasi hasil regresi atau bahkan tidak dapat diinterpretasikan karena tidak sesuai dengan substansi atau kondisi yang diterima secara akal sehat. Oleh karena itu keberadaan multikolinieritas harus ditanggulangi agar hasil regresi menjadi lebih baik dan dapat digunakan untuk melakukan prediksi. Pada dasarnya tidak ada cara spesifik untuk mengatasi multikolinieritas. Beberapa cara dapat dikombinasikan penggunaannya agar penanggulangan multikolinieritas dapat memberikan hasil regresi yang lebih baik dan andal, dengan nilai koefisien regresi dapat diestimasi.

## **F. TES FORMATIF**

1. Multikolinieritas menunjukkan adanya hubungan yang signifikan di antara:
  - a) Variabel dependen dengan variabel independen

- b) Antarvariabel dependen
  - c) Variabel dependen dengan variabel pengganggu
  - d) Variabel independen dengan variabel pengganggu
  - e) Salah semua
2. Apabila dalam sebuah model regresi terdapat multikolinieritas maka berdampak kepada:
- a) Hasil regresi menyesatkan
  - b) Koefisien regresi tidak sesuai substansi
  - c) Koefisien regresi tidak signifikan
  - d) Hasil regresi sulit atau tidak dapat diinterpretasikan
  - e) Betul semua
3. Salah satu indikasi adanya multikolinieritas dalam sebuah model regresi adalah:
- a) Nilai F statistik tidak signifikan, dengan  $R^2$  yang tinggi, nilai t-statistik setiap variabel tidak signifikan
  - b) Nilai F statistik tidak signifikan, dengan  $R^2$  yang rendah, nilai t-statistik setiap variabel tidak signifikan
  - c) Nilai F statistik signifikan, dengan  $R^2$  yang rendah, nilai t-statistik setiap variabel tidak signifikan
  - d) Nilai F statistik signifikan, dengan  $R^2$  yang tinggi, nilai t-statistik setiap variabel tidak signifikan
  - e) Salah semua
4. Berikut adalah cara mengatasi multikolinieritas, kecuali:
- a) Mengganti semua variabel bebas
  - b) Menghilangkan variabel bebas yang kolinier dari model
  - c) Menambah jumlah data
  - d) Melihat informasi sejenis yang ada
  - e) Tidak ada yang benar

## G. LATIHAN

Lakukan regresi berganda JUB dengan G dan PDB menggunakan data berikut, kemudian ujilah apakah terdapat multikolinieritas pada model regresi yang diperoleh serta cara penganggulangnya!

<b>Obs</b>	<b>JUB</b>	<b>G</b>	<b>PDB</b>
<b>1983</b>	21469	585	75832
<b>1984</b>	18385	412	62665
<b>1985</b>	23417	766	86554
<b>1986</b>	28661	971	93638
<b>1987</b>	35885	1075	113718
<b>1988</b>	42998	1304	134105
<b>1989</b>	54704	1829	156851
<b>1990</b>	86470	2495	198597
<b>1991</b>	97105	2771	228450
<b>1992</b>	118053	3554	269884
<b>1993</b>	145303	3744	287976
<b>1994</b>	186514	4504	372221
<b>1995</b>	224368	4960	456381
<b>1996</b>	366534	5955	557659
<b>1997</b>	178120	2945	283782

## KEGIATAN BELAJAR 6

### HETEROSKEDASTISITAS

#### DESKRIPSI KOMPETENSI

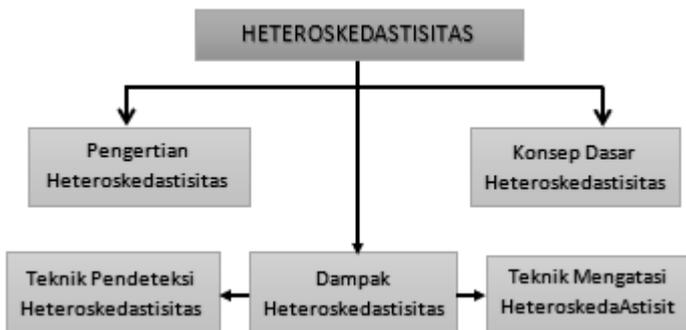
Pada bab ini mahasiswa mempelajari tentang heteroskedastisitas, teknik mendeteksi heteroskedastisitas, dan dampak timbulnya heteroskedastisitas, serta teknik mengatasi terjadinya heteroskedastisitas. Diharapkan mahasiswa memiliki wawasan dan pemahaman tentang konsep heteroskedastisitas sehingga untuk modal dasar mempelajari asumsi klasik model regresi dan analisis data lebih lanjut.

#### KOMPETENSI PEMBELAJARAN

Setelah mengikuti perkuliahan ini, diharapkan mahasiswa memiliki pengetahuan dan kemampuan antara lain:

1. Mampu menjelaskan definisi dan konsep heteroskedastisitas
2. Mampu menjelaskan teknik yang digunakan untuk mendeteksi heteroskedastisitas
3. Mampu menjelaskan teknik yang digunakan untuk mengatasi heteroskedastisitas

#### PETA KONSEP PEMBELAJARAN



## A. PENGERTIAN

Analisis data menggunakan metode analisis regresi perlu memenuhi beberapa asumsi klasik, salah satunya adalah memenuhi asumsi homoskedastisitas atau tidak terjadi heteroskedastisitas. Asumsi klasik heteroskedastisitas penting untuk diuji, sebab melalui uji heteroskedastisitas dapat diketahui apabila terdapat ketidaksamaan varian sisa untuk semua pengamatan pada model regresi. Menurut Ghozali (2021:178) dan Gunawan (2020:128), uji heteroskedastisitas bertujuan untuk mengetahui apakah dalam suatu model regresi terdapat ketidaksamaan varian sisa antar satu pengamatan dengan pengamatan lainnya. Apabila varian sisa dari suatu pengamatan ke pengamatan lainnya tetap maka disebut homoskedastisitas

## B. KONSEP DASAR

Model regresi yang baik apabila memiliki dugaan parameter yang bersifat BLUE (*Best Linier Unbiased Estimate*). Pendugaan parameter agar bersifat BLUE maka varian sisa ( $u_i$ ) harus konstan (tetap), dengan kata lain semua sisa atau *error* mempunyai varian yang sama. Kondisi pada saat varian sisa sama atau tetap sering dikenal dengan homoskedastisitas, sedangkan kondisi pada saat varian sisa tidak konstan atau berubah-ubah sering dikenal dengan heteroskedastisitas. Ketidaksamaan varian sisa dari satu pengamatan ke pengamatan lainnya menimbulkan permasalahan yang cukup serius. Model regresi yang dihasilkan dalam sebuah analisis data dapat dinyatakan tidak valid sebagai alat peramalan apabila memiliki varian sisa yang tidak konstan atau dengan kata lain telah terjadi gangguan homoskedastisitas data.

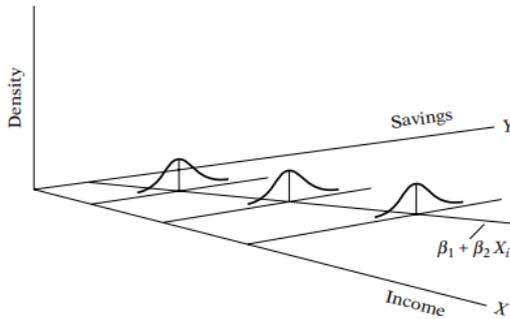
Menurut Gujarati (2003: 387), asumsi homoskedastisitas memiliki varian yang sama atau dapat ditulis:

$$E(u_i^2) = \sigma^2$$

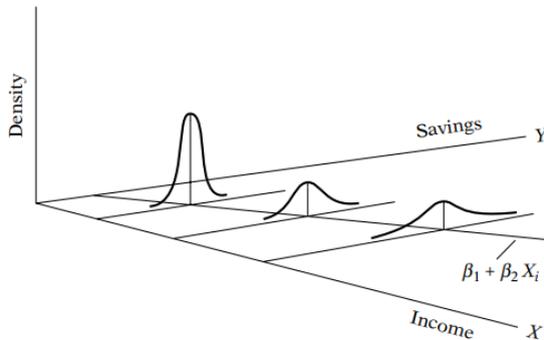
(1)

dimana  $i = 1, 2, \dots, n$

Perhatikan gambar berikut:



Gambar 6.1: Gangguan Homoskedastisitas



Gambar 6.2: Gangguan Heteroskedastisitas

Gambar 6.1 menunjukkan bahwa varian dari  $Y_i$  tergantung pada  $X_i$  yang diberikan dan nilainya tetap sama meskipun dari nilai yang diberikan variabel  $X$ . Sedangkan Gambar 6.2 menunjukkan yang sebaliknya. Gambar 6.2 menunjukkan bahwa varian dari  $Y_i$  memiliki nilai yang terus meningkat seiring dengan meningkatnya  $X_i$ .

Perhatikan rumus persamaan berikut untuk mencari nilai var ( $\beta_1$ ) pada regresi sederhana  $Y = \beta_0 + \beta_1 X_1$ , yaitu:

$$Var(b_1) = \frac{var(u_i)}{\sum x_i} \quad (2)$$

Jika varian sisa  $Var(u_i) = \sigma^2$  tidak tetap, maka persamaan 2 akan berubah dan dapat ditulis dengan persamaan berikut:

$$Var(b_1) = \frac{\sum x_i \sigma_i^2}{(\sum x_i)^2} \quad (3)$$

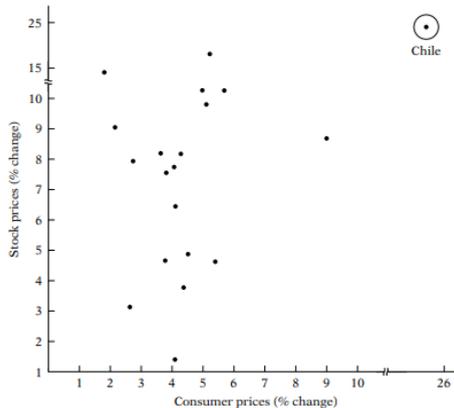
Berdasarkan persamaan 3, dalam pengaplikasiannya  $Var(u_i)$  akan semakin membesar atau dapat dituliskan  $\sigma_1^2 > \sigma_2^2 > \sigma_3^2 > \dots > \sigma_k^2$ . Sehingga dapat dikatakan apabila terjadi heteroskedastisitas akan mengakibatkan  $Var(b_1)$  akan semakin membesar.

### C. DAMPAK HETEROSKEDASTISITAS

Menurut Kurniawan dan Yuniarto (2016: 145), dampak-dampak yang ditimbulkan oleh heteroskedastisitas antara lain: Heteroskedastisitas merusak efisiensi estimator OLS dan Heteroskedastisitas merusak varians. Heteroskedastisitas merusak efisiensi estimator OLS, hal ini dapat menyebabkan hasil pengujian hipotesisnya menjadi tidak dapat dipercaya (meragukan). Selain itu dampak adanya heteroskedastisitas pada data yaitu dapat menghasilkan varian atau kuadrat standar deviasi yang terlalu lebar atau terlalu sempit sehingga sulit untuk mengukur standar deviasi yang sebenarnya. Varian yang membesar mengakibatkan selang kepercayaan menjadi melebar sehingga hal itu tidak baik dalam pemilihan model regresi.

#### D. TEKNIK MENDETEKSI HETEROSKEDASTISITAS

Heteroskedastisitas dapat timbul salah satunya akibat adanya pencilan (*outlier*). Menurut Gujarati (2003: 391), pencilan adalah suatu nilai pengamatan yang jauh berbeda dari nilai rata-rata pengamatan baik memiliki nilai yang sangat kecil maupun nilai yang sangat besar. Perhatikan gambar berikut:



Gambar 6.3: Hubungan Antara Harga Saham dan Harga Konsumen

Pada gambar 6.3 menunjukkan persentase tingkat perubahan harga saham dan harga konsumen. Negara Chile yang ditunjukkan pada gambar 6.3 dianggap sebagai pencilan karena memiliki nilai harga saham dan harga konsumen yang jauh lebih besar dibandingkan negara lain. Kondisi seperti ini yang dapat menyebabkan asumsi homoskedastisitas tidak terpenuhi dengan demikian menyebabkan terjadinya heteroskedastisitas.

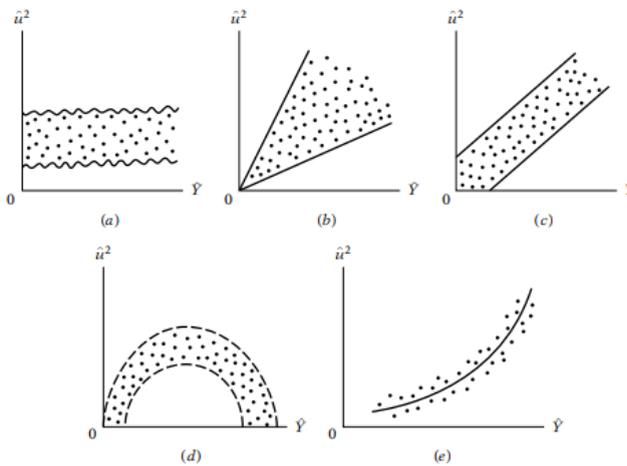
Pencilan data menyebabkan data tidak normal dan terjadinya heteroskedastisitas, namun apakah data yang terlihat normal (tidak memiliki pencilan) tidak akan terjadi heteroskedastisitas? Sebagaimana telah kita pelajari sebelumnya, bahwa heteroskedastisitas adalah kondisi dimana  $Var(u_i)$  atau varian sisa tidak konstan. Oleh karena itu, suatu nilai variabel bebas akan

mempunyai nilai  $Var(u_i)$  yang berbeda dengan variabel bebas lainnya.

Selanjutnya akan dibahas beberapa teknik untuk mendeteksi terjadinya heteroskedastisitas, antara lain:

a. Metode Grafik

Metode grafik adalah metode yang paling mudah untuk mendeteksi adanya masalah heteroskedastisitas. Metode grafik akan membantu membuat pola dari residual data. Apabila data membentuk pola. Perhatikan gambar berikut:

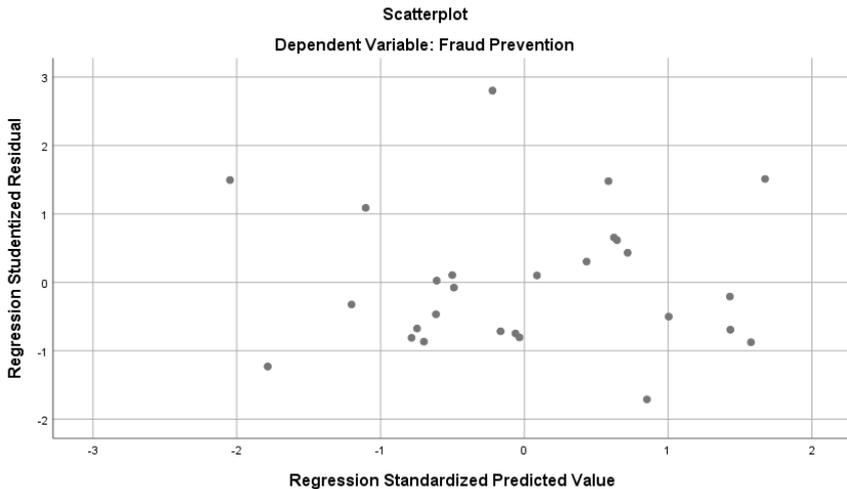


Gambar 6.4: Pola Residual Kuadrat dan Dugaan dari  $Y$

Gambar 6.4 merupakan beberapa plot pola hipotesis residual kuadrat atau  $Var(u_i)$  dan taksiran dari  $Y$  ( $\hat{Y}$ ). Berdasarkan Gambar 6.4, plot (a) tidak membentuk suatu pola yang artinya bahwa tidak ada perbedaan  $Var(u_i)$  pada suatu tingkat ( $\hat{Y}$ ). Dengan kata lain,  $Var(u_i)$  konstan untuk semua nilai ( $\hat{Y}$ ) dan dikenal dengan homoskedastisitas. Sedangkan plot lainnya pada selain Gambar 6.4 (a) membentuk suatu pola. Gambar 6.4 (b) menunjukkan adanya pola sistematis dimana semakin besar nilai ( $\hat{Y}$ ), maka fluktuasi  $u_i^2$  semakin besar. Gambar 6.4 (c) menunjukkan adanya pola trend (kecenderungan). Begitu

pula Gambar 6.4 (d) membentuk pola melengkung mengikuti fungsi kuadrat dan (e) membentuk pola mengikuti fungsi logaritma.

Perhatikan gambar ilustrasi yang diperoleh dari program SPSS berikut:



*Gambar 6.5: Scatterplot Residual Data*

Gambar 6.5 menunjukkan bahwa hubungan antara residual dan taksiran nilai Y tidak membentuk pola tertentu. Hal ini dapat diartikan bahwa data contoh tidak terjadi heteroskedastisitas.

b. Uji Park

Salah satu teknik untuk mendeteksi heteroskedastisitas adalah dengan uji Park dengan memanfaatkan persamaan regresi. Menurut Ghozali (2018: 182), data dinyatakan tidak terjadi heteroskedastisitas apabila hasil uji Park menunjukkan nilai signifikansi di atas 0,05. Menurut Kurniawan dan Yuniarto (2016:147), langkah-langkah uji park sebagai berikut:

- (i) Membuat persamaan regresi antara variabel bebas dan residual yang dapat dituliskan dengan model berikut:

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$$

(4)

Namun dalam beberapa kasus variannya tidak diketahui, maka Park memberikan penyelesaian dengan menggunakan error, dengan demikian perubahan persamaan regresinya dapat ditulis sebagai berikut:

$$\ln e_i^2 = \ln \sigma^2 + \beta \ln X_i + \varepsilon_i = a + \beta \ln X_i + \varepsilon_i$$

(5)

(ii) Lakukan uji-t

Adapun hipotesis untuk uji ini yaitu

$H_0$ : Tidak terjadi heteroskedastisitas (homogenitas)

$H_1$ : Terjadi heteroskedastisitas

Apabila nilai uji-t lebih besar dibandingkan nilai uji-t kritis maka hipotesis nol ditolak yang artinya varian residual tidak konstan atau terjadi heteroskedastisitas data.

c. Uji Goldfeld-Quandt

Uji Goldfeld-Quandt adalah uji yang diperkenalkan pertama kali oleh Arthur Goldfeld dan Richard Quandt pada sekitar tahun 1960 dengan tujuan untuk mendeteksi adanya heteroskedastisitas dalam suatu model regresi. Uji ini dilakukan dengan melibatkan pembagian data lanjut ke dalam subkelompok dan menganalisis varian sisanya. Apabila varian sisa kelompok berbeda nyata maka terjadi heteroskedastisitas. Pembagian kelompok berdasarkan nilai variabel bebasnya, apabila terdapat lebih dari satu variabel bebas dalam model regresi, maka perlu menentukan subkelompoknya.

d. Uji White

Uji White (*White's General Heteroscedasticity Test*) adalah uji pendeteksian heteroskedastisitas yang paling mudah. Uji ini menganalisis residual dari model regresi dengan menggunakan

distribusi Chi-Square untuk pengujiannya. Adapun tahapan prosedur pengujian, yaitu:

1. Langkah pertama membentuk persamaan regresi yang dapat dituliskan sebagai berikut:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i \quad (6)$$

Berdasarkan persamaan regresi yang telah dibentuk, akan menghasilkan nilai error yaitu  $\hat{u}_i^2$

2. Membuat persamaan regresi untuk nilai error atau nilai residual:

$$\hat{u}_i^2 = a_0 + a_1 X_{1i} + a_2 X_{2i} + a_3 X_{1i}^2 + a_4 X_{2i}^2 + a_5 X_{1i} X_{2i} + v_i \quad (7)$$

3. Menentukan nilai  $R^2$  sebagai regresi tambahan. Sampe berukuran  $n$  dan koefisien determinasi  $R^2$  yang mengikuti distribusi Chi-Square dengan derajat bebasnya adalah jumlah variabel bebas atau jumlah koefisien regresi maka formulasi Uji White dapat ditulis sebagai berikut:

$$nR^2 \sim \chi^2$$

Menguji hipotesis

Adapun hipotesis untuk uji ini yaitu

$H_0$ : Tidak terjadi heteroskedastisitas (homogenitas)

$H_1$ : Terjadi heteroskedastisitas

Apabila nilai uji chi-square lebih besar dibandingkan nilai uji chi-square kritis maka hipotesis nol ditolak yang artinya varian residual tidak konstan atau terjadi heteroskedastisitas data.

e. Uji Korelasi Spearman

Uji Korelasi Spearman adalah uji untuk mendeteksi heteroskedastisitas. Koefisien korelasi rank Spearman dapat ditulis sebagai berikut:

$$r_s = 1 - 6 \left[ \frac{\sum d_i^2}{n(n^2-1)} \right] \quad (8)$$

Dimana  $d_i$  adalah selisih dalam rank antar dua pengamatan dan  $n$  adalah jumlah pengamatan. Adapun langkah-langkah menggunakan uji korelasi spearman yaitu:

- i. Menghitung regresi  $y$  terhadap  $x$ , dan hitung  $e_i$
- ii. Menghitung rank dari  $|e_i|$  dan  $x_i$ , selanjutnya hitung korelasi Spearman dimana  $d_i$  adalah selisih rank dari 2 karakteristik yang berbeda.
- iii. Menghitung pengujian hipotesis dengan uji statistik berikut:

$$t_{stat} = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}} \sim r_n - 2 \quad (9)$$

Apabila nilai  $|t_{stat}|$  lebih kecil dari nilai  $|t_{tabel}|$  atau nilai signifikansi lebih besar dari 0,05 maka tidak terjadi heteroskedastisitas.

f. Uji Glejser

Uji glejser yaitu uji dengan cara meregresikan variabel-variabel bebas terhadap nilai absolut residualnya. Uji ini mendeteksi heteroskedastisitas dengan cara meregresikan nilai mutlak residual  $|u_i|$  terhadap variabel bebas yang dianggap berkaitan dengan varian heteroskedastisitas  $u_i^2$ . Adapun langkah-langkah Uji Glejser, yaitu:

1. Meregresi data  $Y$  dan  $X$  dan memperoleh nilai residual  $e_i$ .

2. Meregresi variabel bebas X yang dianggap berkaitan dengan varian heteroskedastisitas dengan nilai mutlak residual  $|u_i|$ .
3. Pilih persamaan dengan nilai  $R^2$  yang paling tinggi untuk menyatakan heteroskedastisitas.
4. Menghitung uji t persamaan regresi pada langkah 3.

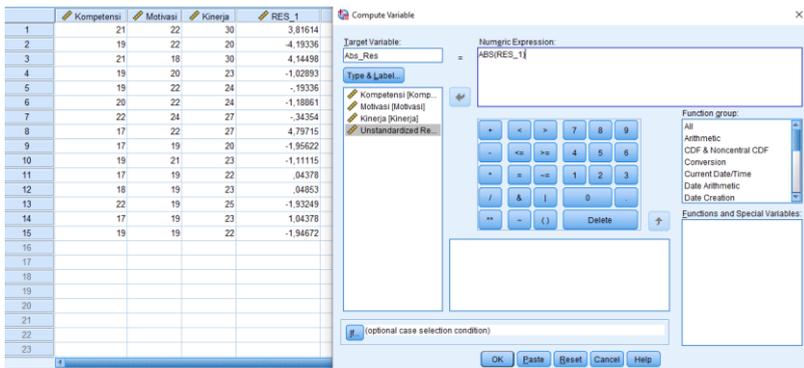
Adapun hipotesis untuk uji ini yaitu

$H_0$ : Tidak terjadi heteroskedastisitas (homogenitas)

$H_1$ : Terjadi heteroskedastisitas

Apabila nilai uji t lebih besar dibandingkan nilai uji t kritis maka hipotesis nol ditolak yang artinya varian residual tidak konstan atau terjadi heteroskedastisitas data.

Contoh pendeteksian heteroskedastisitas menggunakan SPSS:



Gambar 6.6: Pengerjaan Uji Glejser

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

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-,594	6,517		-,091	,929
	Kompetensi	,093	,277	,101	,337	,742
	Motivasi	,032	,277	,035	,117	,909

a. Dependent Variable: Abs\_Res

*Gambar 6.7: Hasil Uji Glejser*

Gambar 6.7 menunjukkan bahwa nilai signifikansi dari variabel kompetensi dan motivasi terhadap residualnya lebih besar dari 0,05 yang artinya bahwa tidak terjadi heteroskedastisitas. Dengan kata lain, varian residual adalah konstan sehingga dapat dilanjutkan untuk membantuk model regresi.

## E. TEKNIK MENGATASI HETEROSKEDASTISITAS

Menurut Nachrowi dan Usman (2020: 120) serta Kurniawan dan Yuniarto (2016:148) ada beberapa teknik untuk mengatasi heteroskedastisitas yaitu:

### 1. Metode Generalized Least Squares (GLS)

Metode ini juga dikenal dengan metode kuadrat terkecil tertimbang. Metode ini digunakan apabila varian  $\sigma_j^2$  diketahui. Model regresi  $Y_j = \beta_0 + \beta_1 X_j + u_j$  dengan  $Var(u_j) = \sigma_j^2$ . Jika setiap ruas dikalikan dengan  $\frac{1}{\sigma_j}$ , maka:

$$Y_j \cdot \frac{1}{\sigma_j} = \beta_0 \left( \frac{1}{\sigma_j} \right) + \beta_1 X_j \left( \frac{1}{\sigma_j} \right) + u_j \left( \frac{1}{\sigma_j} \right)$$

$$\frac{Y_j}{\sigma_j} = \beta_0 \left( \frac{1}{\sigma_j} \right) + \beta_1 \left( \frac{X_j}{\sigma_j} \right) + \left( \frac{u_j}{\sigma_j} \right) \quad (10)$$

Persamaan 10 dapat telah homoskedastisitas. Misalkan  $\frac{1}{\sigma_j}$  kita simbolkan dengan \* maka ini dapat dibuktikan dengan persamaan berikut:

$$E(u_j^{*2}) = E\left(\frac{u_j^2}{\sigma_j^2}\right) = \frac{1}{\sigma_j^2} E(u_j^2) = \frac{1}{\sigma_j^2} (\sigma_j^2) = \text{konstan} \quad (11)$$

Persamaan 11 membuktikan bahwa model sudah memenuhi homoskedastisitas, sehingga persamaan 10 dapat diduga dengan OLS dan pendugaannya akan memperoleh hasil yang bersifat BLUE.

## 2. Menggunakan Robust Standar Error

Heteroskedastisitas mengakibatkan standar error yang dihasilkan dari regresi menjadi bias sehingga uji hipotesis yang dihasilkan menjadi meragukan. Maka dari itu perlu dilakukan koreksi model khususnya pada standar error regresi. Standar error yang kuat (robust) terhadap heteroskedastisitas dapat dihitung dengan rumus berikut:

$$\text{Var}(\hat{\beta}_j) = \frac{\sum_{i=1}^n \hat{r}_{ij}^2 \hat{u}_j^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (12)$$

$$SE(\hat{\beta}_j) = \sqrt{\text{Var}(\hat{\beta}_j)} \quad (13)$$

dimana  $\hat{r}_{ij}^2$  merupakan residual ke-i dari variabel  $x_j$  terhadap variabel bebas lainnya.

## 3. Transformasi dengan Logaritma Natural

Transformasi logaritma yang digunakan untuk mengatasi heteroskedastisitas adalah transformasi logaritma natural. Cara ini yang paling mudah untuk menyelesaikan masalah heteroskedastisitas. Transformasi ini bertujuan untuk memperkecil skala antar variabel bebas dengan cara mentransformasi data kedalam bentuk logaritma natural (Ln)

yang mengubah error menjadi kecil. Hal ini mengakibatkan heteroskedastisitas akan berkurang. Model regresi dalam bentuk logaritma natural dapat ditulis sebagai berikut:

$$\ln Y = \ln \alpha + \beta_1 \ln x_1 + \beta_2 \ln x_2 + e$$

## F. RANGKUMAN

Berdasarkan uraian di atas, heteroskedastisitas adalah suatu kondisi dimana varian residual yang tidak konstan atau selalu berubah-ubah disetiap pengamatan. Terjadinya heteroskedastisitas akan merusak efisiensi estimator OLS sehingga dapat menyebabkan hasil pengujian hipotesisnya meragukan. Dalam model regresi, heteroskedastisitas tidak boleh terjadi karena akan menyebabkan model regresi kurang baik. Perlu dilakukan pendeteksian heteroskedastisitas sebelum memodelkan regresi. Teknik yang pendeteksi heteroskedastisitas yaitu metode grafik, uji park, uji Goldfeld-Quand, uji White, uji korelasi spearman, uji Glejser, dan lainnya. Data yang terjadi heteroskedastisitas perlu diatasi dengan cara yaitu Metode Generalized Least Squares (GLS), metode Robust Standar Error, Transformasi Logaritma Natural, dan lainnya.

## G. TES FORMATIF

1. Berikut adalah kumpulan teknik mengatasi heteroskedastisitas, kecuali?
  - a) Weight least squares dan ordinary least square
  - b) Metode robust standar error dan menghilangkan data
  - c) Weight least square dan transformasi error
  - d) Robust standar error dan transformasi logaritma natural

2. Perhatikan pilihan berikut ini

- i. Uji Glejser
- ii. Uji Normalitas
- iii. Uji Korelasi Pearson
- iv. Uji Goldfeld-Quantd
- v. Metode Grafik

Manakah pilihan teknik untuk mendeteksi heteroskedastisitas yang benar!

- a) i, ii, dan iii
- b) i, ii, dan iv
- c) ii, iii dan iv
- d) i, iii dan v

## H. LATIHAN

Ujilah adanya heteroskedastisitas dari kasus pengaruh data hasil pendapatan domestik regional bruto (PDRB) terhadap pertumbuhan ekonomi! Kemudian jelaskan! Apabila terdapat heteroskedastisitas, atasilah kondisi tersebut dengan metode transformasi logaritma natural!

## **KEGIATAN BELAJAR 7**

### **MODEL REGRESI DENGAN VARIABEL DUMMY**

#### **DESKRIPSI PEMBELAJARAN**

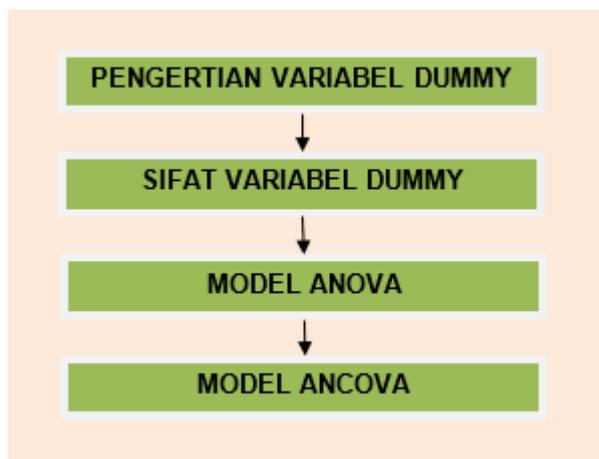
Pada bab ini mahasiswa mempelajari model regresi dengan variable dummy. Diharapkan mahasiswa memiliki wawasan dan pemahaman tentang macam-macam model regresi yang menggunakan variable dummy sebagai cara untuk penggunaan variable kualitatif.

#### **KOMPETENSI PEMBELAJARAN**

Setelah mengikuti perkuliahan ini diharapkan mahasiswa memiliki pengetahuan dan kemampuan:

- (1) Mampu menjelaskan tentang variable dummy
- (2) Mampu menjelaskan tentang sifat variable dummy dan penggunaannya
- (3) Mampu menjelaskan dan mengestimasi Model ANOVA
- (4) Mampu menjelaskan dan mengestimasi Model ANCOVA

#### **PETA KONSEP PEMBELAJARAN**



## A. PENGERTIAN VARIABEL DUMMY

Pada sebelumnya, kita telah membahas secara singkat empat jenis variabel yang umumnya dijumpai dalam analisis empiris, yaitu: **skala rasio, skala interval, skala ordinal, dan skala nominal**. Jenis-jenis variabel yang telah kita temui dalam bab-bab sebelumnya adalah pada dasarnya adalah skala rasio. Namun, hal ini tidak boleh memberikan kesan bahwa model regresi hanya dapat hanya dapat menangani variabel berskala rasio. Model regresi juga dapat menangani jenis-jenis variabel jenis lain dari variabel yang telah disebutkan sebelumnya. Pada bab ini, kita akan membahas model-model yang mungkin melibatkan tidak hanya variabel berskala rasio tetapi juga variabel berskala nominal. Variabel-variabel tersebut juga juga dikenal sebagai **variabel indikator, variabel kategorikal, variabel kualitatif, atau variabel dummy**.

**Variabel Dummy:** Variabel dummy disebut juga variabel indikator, biner, kategori, kualitatif, atau variabel dikotomi. Ini digunakan untuk memodelkan hubungan antara variabel independen kategorikal dan variabel terikat. Misalnya, kita ingin memahami pengaruh jenis kelamin terhadap pendapatan. Kita dapat membuat variabel dummy dengan nilai **1** untuk perempuan dan **0** untuk laki-laki. **Model regresi dengan variabel dummy** digunakan ketika kita ingin memasukkan **variabel kualitatif** (seperti jenis kelamin, kategori, atau status) ke dalam model regresi. Variabel dummy hanya memiliki dua nilai: **1** dan **0**.

### Penggunaan Variabel Dummy

- (1) **Regresi dengan Dummy Variabel:** Ketika kita memiliki variabel bebas kualitatif dan variabel terikat kuantitatif, kita dapat menggunakan regresi dengan variabel dummy. Contohnya adalah mengukur pengaruh jenis kelamin terhadap pendapatan.

- (2) **Regresi Logistik atau Multinomial:** Ketika variabel terikat berbentuk data kategori (misalnya, puas/tidak puas), kita perlu menggunakan variabel dummy. Misalnya, kita ingin memahami pengaruh kualitas produk terhadap kepuasan konsumen.

**Contoh Model Regresi dengan Variabel Dummy:**

Misalkan kita ingin memahami pengaruh lokasi (kota/desa) terhadap harga produk. Model regresi yang digunakan:

$$\text{Harga produk} = \beta_0 + \beta_1 D + e$$

$$D = 1 \text{ (kota)}$$

$$D = 0 \text{ (desa)}$$

Hasil estimasi dengan analisis regresi (OLS) dapat memberikan informasi tentang perbedaan harga antara daerah perkotaan dan pedesaan. Dengan menggunakan variabel dummy, kita dapat memasukkan karakteristik kualitatif ke dalam model regresi dan memahami pengaruhnya terhadap variabel terikat.

**B. SIFAT DARI VARIABEL DUMMY**

Dalam analisis regresi, variabel tidak bebas, sering kali tidak hanya dipengaruhi oleh variabel berskala rasio (misalnya, pendapatan, output, harga, biaya, tinggi, suhu) tetapi juga oleh variabel yang pada dasarnya bersifat kualitatif, atau skala nominal, seperti jenis kelamin, jenis kelamin, warna kulit, agama, kebangsaan, wilayah geografis, pergolakan politik, dan afiliasi partai. Sebagai contoh, dengan menganggap semua faktor lain konstan, pekerja perempuan ternyata mendapat upah lebih rendah daripada pekerja laki-laki. Pola ini dapat diakibatkan oleh diskriminasi jenis kelamin, tetapi apa pun alasannya, variabel kualitatif seperti jenis kelamin tampaknya mempengaruhi regresi dan jelas harus dimasukkan di antara variabel bebas.

Karena variabel-variabel tersebut biasanya menunjukkan ada atau tidaknya "kualitas" atau atribut, seperti pria atau wanita, desa atau kota, sebelum kebijakan atau sesudah kebijakan, pada dasarnya variabel-variabel tersebut adalah variabel berskala nominal. Salah satu cara kita dapat "mengukur" atribut tersebut adalah dengan membuat variabel buatan yang memiliki nilai 1 atau 0, 1 menunjukkan menunjukkan keberadaan (atau kepemilikan) atribut tersebut dan 0 menunjukkan ketiadaan atribut tersebut. Sebagai contoh, 1 dapat mengindikasikan bahwa seseorang adalah perempuan dan 0 dapat menunjukkan laki-laki; atau 1 dapat mengindikasikan bahwa seseorang adalah lulusan perguruan tinggi, dan 0 bahwa orang tersebut bukan lulusan perguruan tinggi, dan seterusnya. Variabel yang mengasumsikan nilai 0 dan 1 seperti itu disebut variabel dummy. Variabel semacam itu dengan demikian pada dasarnya merupakan alat untuk mengklasifikasikan data ke dalam kategori yang saling terpisah seperti laki-laki atau perempuan.

Variabel dummy dapat dimasukkan ke dalam model regresi semudah variabel kuantitatif. Faktanya, sebuah model regresi dapat berisi regressor yang semuanya adalah dummy, yang seluruhnya bersifat dummy, atau kualitatif. Model-model seperti itu disebut model Analisis Varians (ANOVA).

### **C. MODEL ANOVA**

Model ANOVA digunakan untuk menilai signifikansi statistik dari hubungan antara kuantitatif dan kualitatif atau variable bebas boneka. Mereka sering digunakan untuk membandingkan perbedaan nilai rata-rata dari dua atau lebih kelompok atau kategori.

Untuk mengilustrasikan model ANOVA, terlihat pada contoh berikut:

Tabel 1 menyajikan data mengenai Indeks Pembangunan Manusia (IPM) di 38 provinsi di Indonesia pada tahun 2023. Ke-38 provinsi ini diklasifikasikan ke dalam tiga wilayah geografis: (1) Wilayah Indonesia Barat (WIB) ada 16 provinsi, (2) Wilayah Indonesia Tengah (WITA) ada 14 provinsi, dan (3) Wilayah Indonesia Timur (WIT) ada 8 provinsi.

Misalkan kita ingin mengetahui apakah IPM berbeda di antara tiga wilayah geografis di Indonesia. Jika kita mengambil rata-rata aritmatika sederhana dari IPM di tiga wilayah tersebut, kita akan menemukan bahwa rata-rata untuk ketiga wilayah tersebut adalah sebagai berikut: 74.4219 (WIB), 72.0243 (WITA), dan 65.4338 (WIT). Angka-angka ini terlihat berbeda, tetapi apakah mereka berbeda secara statistik satu sama lain? Ada berbagai teknik statistik untuk membandingkan dua atau lebih nilai rata-rata, yang umumnya dikenal dengan nama **analisis varians**. Tetapi tujuan yang sama dapat dicapai dalam kerangka analisis regresi.

Untuk melihat hal ini, dipergunakan model berikut ini:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i \quad (1)$$

di mana:

$Y_i$  = (rata-rata) IPM di provinsi  $i$

$D_{1i} = 1$  jika provinsi tersebut berada di Wilayah Indonesia Barat  
= 0 jika tidak (yaitu, di wilayah lain di Indonesia)

$D_{2i} = 1$  jika provinsi tersebut berada di Wilayah Indonesia Tengah  
= 0 jika tidak (yaitu, di wilayah lain di Indonesia)

Perhatikan bahwa Persamaan (1) seperti model regresi berganda pada umumnya yang dibahas sebelumnya, kecuali bahwa, kita hanya memiliki regresi kualitatif, atau boneka, yang mengambil nilai 1 jika observasi termasuk dalam kategori tertentu dan 0 jika tidak termasuk dalam kategori atau kelompok tersebut. Tabel 1 menunjukkan variabel-variabel dummy yang disusun.

**Tabel 1 Indeks Pembangunan Manusia dan Laju Pertumbuhan Produk Domestik Regional Bruto Atas Dasar Harga Konstan 2010 (Persen) tahun 2023 menurut Provinsi**

No	Provinsi	Laju Pertumbuhan Produk Domestik Regional Bruto Atas Dasar Harga Konstan 2010 (Persen)	Indeks Pembangunan Manusia	d1	d2
1	ACEH	4.23	73.4	1	0
2	SUMATERA UTARA	5.01	73.37	1	0
3	SUMATERA BARAT	4.62	73.75	1	0
4	RIAU	4.21	74.04	1	0
5	JAMBI	4.66	72.77	1	0
6	SUMATERA SELATAN	5.08	71.62	1	0
7	BENGKULU	4.26	72.78	1	0
8	LAMPUNG	4.55	71.15	1	0
9	KEP. BANGKA BELITUNG	4.38	72.85	1	0
10	KEP. RIAU	5.2	77.11	1	0
11	DKI JAKARTA	4.96	82.46	1	0
12	JAWA BARAT	5	73.74	1	0
13	JAWA TENGAH	4.98	73.39	1	0
14	DI YOGYAKARTA	5.07	81.07	1	0
15	JAWA TIMUR	4.95	73.38	1	0
16	BANTEN	4.81	73.87	1	0
17	BALI	5.71	77.1	0	1
18	NUSA TENGGARA BARAT	1.8	70.2	0	1
19	NUSA TENGGARA TIMUR	3.52	66.68	0	1
20	KALIMANTAN BARAT	4.46	69.41	0	1
21	KALIMANTAN TENGAH	4.14	72.2	0	1
22	KALIMANTAN SELATAN	4.84	72.5	0	1
23	KALIMANTAN TIMUR	6.22	78.2	0	1
24	KALIMANTAN UTARA	4.94	72.49	0	1
25	SULAWESI UTARA	5.48	74.36	0	1
26	SULAWESI TENGAH	11.91	70.95	0	1
27	SULAWESI SELATAN	4.51	73.46	0	1
28	SULAWESI TENGGARA	5.35	72.79	0	1
29	GORONTALO	4.5	70.45	0	1
30	SULAWESI BARAT	5.25	67.55	0	1
31	MALUKU	5.21	70.94	0	0
32	MALUKU UTARA	20.49	70.21	0	0
33	PAPUA BARAT	5.18	66.66	0	0
34	PAPUA BARAT DAYA	1.82	66.66	0	0
35	PAPUA	4.2	62.25	0	0
36	PAPUA SELATAN	4.27	62.25	0	0
37	PAPUA TENGAH	5.95	62.25	0	0
38	PAPUA PEGUNUNGAN	4.78	62.25	0	0

Keterangan:

D1 = 1 jika provinsi ada di Wilayah Indonesia bagian Barat (WIB),  
0 jika lainnya

$D_2 = 1$  jika provinsi ada di Wilayah Indonesia bagian Tengah (WITA), 0 jika lainnya

Sumber: Badan Pusat Statistik, Februari 2024

Apa arti model (1)? Dengan mengasumsikan bahwa *error term* memenuhi asumsi OLS yang biasa, dengan mengambil ekspektasi dari Persamaan (1) di kedua sisi, kita memperoleh:

Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia Barat:

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

Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia Tengah:

$$E(Y_i | D_{1i}= 0, D_{2i}=1) = \beta_0 + \beta_2 \quad (3)$$

Kita mungkin bertanya-tanya bagaimana kita mengetahui Indeks Pembangunan Manusia di Wilayah Indonesia Timur. Jika Anda menebak bahwa ini sama dengan  $\beta_0$ , Anda benar sekali, karena:

Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia Timur:

$$E(Y_i | D_{1i}= 0, D_{2i}=0) = \beta_0 \quad (4)$$

Dengan kata lain, Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia Timur ditunjukkan oleh intersep,  $\beta_0$ , dalam regresi berganda (1), dan koefisien "kemiringan"  $\beta_1$  dan  $\beta_2$  menyatakan seberapa besar perbedaan Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia Barat dan Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia tengah berbeda dengan Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia Timur. Tapi bagaimana kita tahu apakah

perbedaan ini signifikan secara statistik? Sebelum kita menjawab pertanyaan ini, mari kita sajikan hasil yang didasarkan pada regresi (1). Dengan menggunakan data yang diberikan pada Tabel 1, kita memperoleh hasil sebagai berikut:

Dependent Variable: IPM				
Method: Least Squares				
Sample: 1 38				
Included observations: 38				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	65.43375	1.164493	56.19074	0.0000
D1	8.988125	1.426207	6.302117	0.0000
D2	6.590536	1.459770	4.514778	0.0001
R-squared	0.533387	Mean dependent var		71.64632
Adjusted R-squared	0.506723	S.D. dependent var		4.689609
S.E. of regression	3.293685	Akaike info criterion		5.297548
Sum squared resid	379.6926	Schwarz criterion		5.426831
Log likelihood	-97.65340	Hannan-Quinn criter.		5.343545
F-statistic	20.00429	Durbin-Watson stat		1.557556
Prob(F-statistic)	0.000002			

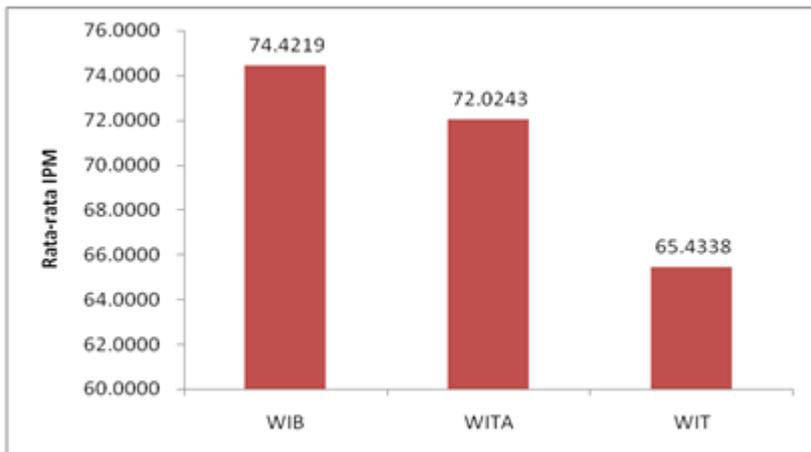
Jika dituliskan dalam bentuk persamaan:

$$\begin{aligned}
 \text{IPM} &= 65.4338 + 8.9881 D1 + 6.5905 D2 & (5) \\
 \text{se} &(1.1645) \quad (1.4262) \quad (1.4598) \\
 \text{t} &(59.1907) \quad (6.6321) \quad (4.5148)
 \end{aligned}$$

Seperti yang ditunjukkan oleh hasil regresi ini, Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia Timur adalah sekitar 65.4338, Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia Barat lebih tinggi sekitar 8.9881 dan Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia Tengah sekitar 6.5905. Rata-rata aktual di dua wilayah terakhir dapat dengan mudah diperoleh dengan menambahkan perbedaan IPM ini ke Rata-rata Indeks Pembangunan Manusia di Wilayah Indonesia Timur seperti yang ditunjukkan pada Persamaan (3) dan (4). Dengan melakukan ini, kita akan menemukan bahwa Rata-rata

Indeks Pembangunan Manusia di Wilayah Indonesia Barat dan Tengah adalah sekitar 74.4219 dan 72.0243.

Namun, bagaimana kita tahu bahwa Rata-rata Indeks Pembangunan Manusia di Barat dan tengah ini secara statistik berbeda dengan rata-rata Indeks Pembangunan Manusia di Timur, sebagai kategori pembandingan? Itu cukup mudah. Yang harus kita lakukan adalah mencari tahu apakah setiap koefisien "kemiringan" dalam Persamaan (5) signifikan secara statistik. Seperti yang dapat dilihat dari regresi ini, estimasi koefisien kemiringan untuk Wilayah Barat signifikan secara statistik, karena nilai p-nya adalah 0.000 persen, dan koefisien kemiringan Wilayah Tengah juga signifikan secara statistik, karena nilai p-nya adalah 0.0001 persen. Oleh karena itu, kesimpulan keseluruhannya adalah bahwa secara statistik rata-rata Indeks Pembangunan Manusia di Timur, Tengah, dan Barat berbeda secara signifikan. Secara grafis, ditunjukkan pada Gambar 1.7



Gambar 7.1 Rata-rata Indeks Pembangunan Manusia di tiga wilayah di Indonesia, 2023

Perlu kehati-hatian dalam menafsirkan perbedaan-perbedaan ini. Variabel dummy hanya akan menunjukkan perbedaan, jika ada, tetapi mereka tidak menunjukkan alasan untuk perbedaan-perbedaan tersebut. Perbedaan tingkat pendidikan, indeks biaya hidup, pertumbuhan ekonomi mungkin memiliki pengaruh terhadap perbedaan yang diamati. Oleh karena itu, kecuali kita memperhitungkan semua variabel lain yang dapat mempengaruhi IPM, kita tidak akan dapat menentukan penyebab perbedaan tersebut.

Dari pembahasan sebelumnya, jelas bahwa yang harus dilakukan adalah melihat apakah koefisien yang melekat pada berbagai variabel boneka secara individual signifikan secara statistik. Contoh ini juga menunjukkan betapa mudahnya memasukkan variabel bebas kualitatif, atau dummy, dalam model regresi.

### **Model ANCOVA: Model Regresi dengan Variabel Bebas Kuantitatif dan Kualitatif**

Model ANCOVA umum digunakan dalam bidang-bidang seperti sosiologi, psikologi, pendidikan, dan penelitian pasar, namun tidak hanya dalam bidang ekonomi. Biasanya, dalam kebanyakan penelitian ekonomi, model regresi berisi beberapa variabel penjelas yang bersifat kuantitatif dan ada juga yang bersifat kualitatif. Model ANCOVA adalah perluasan dari model ANOVA. Model ANCOVA dalam hal ini menyediakan metode untuk mengontrol efek secara statistik dari kuantitatif, yang disebut kovariat atau variabel kontrol, dalam model yang mencakup baik kuantitatif maupun kualitatif, atau dummy. Berikut ini akan diilustrasikan model ANCOVA.

Mari kita lihat kembali Contoh 1 dimana rata-rata IPM berbeda di tiga wilayah tanpa memperhitungkan variabel-variabel yang mungkin mempengaruhi IPM. misalnya, variabel pertumbuhan ekonomi. Untuk melihat apakah hal ini terjadi, kita mengembangkan model berikut ini:

$$IPM_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 GPDRB_i + u_i$$

(5)

di mana:

$IPM_i$  = (rata-rata) IPM di provinsi i

$GPDRB_i$  = Pertumbuhan Produk Domestik Regional Bruto Atas Dasar Harga Konstan 2010 (Persen)

$D_{1i}$  = 1 jika provinsi tersebut berada di Wilayah Indonesia Barat  
= 0 jika tidak (yaitu, di wilayah lain di Indonesia)

$D_{2i}$  = 1 jika provinsi tersebut berada di Wilayah Indonesia Tengah  
= 0 jika tidak (yaitu, di wilayah lain di Indonesia)

Data GPDRB ada pada Tabel 1. Perlu diingat bahwa kita memperlakukan provinsi di wilayah Timur sebagai kategori tolok ukur.

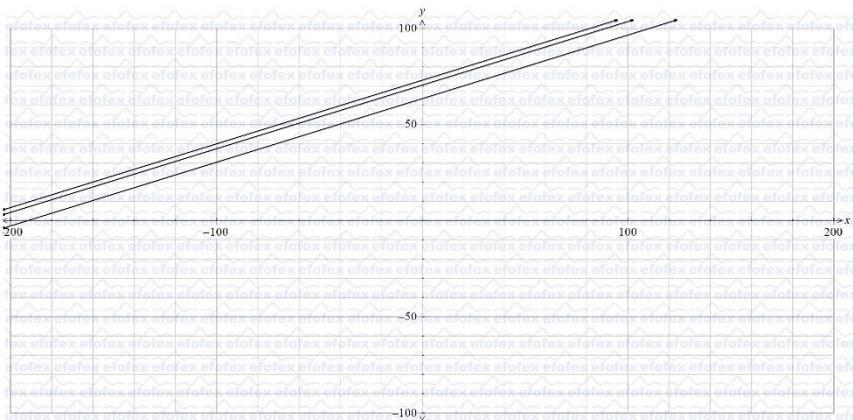
Dari data pada Tabel 1, hasil estimasi persamaan model (5) adalah sebagai berikut:

Dependent Variable: IPM				
Method: Least Squares				
Sample: 1 38				
Included observations: 38				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	63.28829	1.645382	38.46418	0.0000*
D1	9.563348	1.419845	6.735489	0.0000*
D2	7.020336	1.435823	4.889415	0.0000*
GPDRB	0.330707	0.184452	1.792912	0.0819*
R-squared	0.573692	Mean dependent var		71.64632
Adjusted R-squared	0.536077	S.D. dependent var		4.689609
S.E. of regression	3.194183	Akaike info criterion		5.259840
Sum squared resid	346.8953	Schwarz criterion		5.432218
Log likelihood	-95.93697	Hannan-Quinn criter.		5.321171
F-statistic	15.25152	Durbin-Watson stat		1.860713
Prob(F-statistic)	0.000002			

\*P-value lebih kecil dari alpha 5 persen untuk uji satu sisi

Seperti yang ditunjukkan oleh hasil estimasi ini, ceteris paribus: ketika pertumbuhan PDRB naik 1 persen, rata-rata IPM naik sekitar 0.3307. Koefisien intersep diferensial signifikan, baik untuk wilayah Barat maupun untuk wilayah Tengah. Hasil ini sama dengan hasil sebelumnya bahkan perbedaannya menjadi lebih besar. Jika kita tuliskan persamaan regresi untuk setiap wilayah, akan diperoleh persamaan sebagai berikut:

Wilayah Timur:	$IPM_i = 63.28829 + 0.330707$
GPDRB <sub>i</sub>	
Wilayah Barat:	$IPM_i = 63.28829 + 9.563348 +$
0.330707 GPDRB <sub>i</sub>	$= 72.85164 + 0.330707$
GPDRB <sub>i</sub>	
Wilayah Tengah:	$IPM_i = 63.28829 + 7.020336 +$
0.330707 GPDRB <sub>i</sub>	$= 70.30863 + 0.330707$
GPDRB <sub>i</sub>	



Gambar 7.2. Garis Regresi Pertumbuhan PDRB terhadap IPM di tiga wilayah di Indonesia tahun 2023

Gambar 2 menunjukkan tiga garis regresi untuk tiga wilayah. Garis regresi tersebut tidak sama untuk ketiga wilayah tersebut. Ketiganya menunjukkan garis regresi yang sejajar, artinya ketiga garis tersebut mempunyai intersep yang berbeda dengan kemiringan garis atau slope yang sama. (Mengapa?)

#### D. PERHATIAN DALAM PENGGUNAAN VARIABEL DUMMY

Meskipun mudah untuk dimasukkan ke dalam model regresi, kita harus menggunakan variabel dummy dengan hati-hati. Secara khusus, pertimbangkan aspek-aspek berikut:

1. Jika variabel kualitatif memiliki  $m$  kategori, perkenalkan hanya  $(m - 1)$  variabel dummy. Dalam contoh kita, karena variabel kualitatif "wilayah" memiliki tiga kategori, kita hanya memasukkan dua variabel dummy. Jika tidak mengikuti aturan ini, kita akan jatuh ke masuk dalam jebakan variabel dummy, yaitu situasi kolinieritas sempurna atau perfect multicollinearity, jika ada lebih dari satu hubungan yang sempurna di antara variabel-variabel. Ini ini juga berlaku jika kita memiliki lebih dari satu variabel kualitatif dalam model.
2. Kategori yang tidak diberi variabel dummy dikenal sebagai basis, tolok ukur, kontrol, perbandingan, referensi, atau kategori. Dan semua perbandingan dibuat dalam kaitannya dengan kategori tolok ukur.
3. Nilai intersep ( $\beta_0$ ) mewakili nilai rata-rata dari kategori tolok ukur. Dalam Dalam Contoh 1, kategori tolok ukurnya adalah wilayah Timur. Oleh karena itu, dalam regresi (5) nilai intersep sekitar 63.28829 mewakili rata-rata IPM di provinsi di Wilayah Timur.
4. Koefisien yang melekat pada variabel dummy dalam Persamaan (1) dikenal sebagai **koefisien intersep diferensial** karena koefisien tersebut menunjukkan seberapa besar nilai kategori yang menerima nilai 1 berbeda dengan koefisien

intersep dari kategori pembanding. Sebagai contoh, dalam Persamaan (5), nilai sekitar 9.563348 memberi tahu kita bahwa rata-rata IPM di provinsi di wilayah Barat lebih besar sekitar 9.563348 daripada rata-rata IPM provinsi di wilayah Timur.

5. Jika variabel kualitatif memiliki lebih dari satu kategori, seperti dalam contoh ilustrasi kami, maka pilihan kategori tolok ukur sepenuhnya tergantung pada peneliti. Terkadang pilihan tolok ukur ditentukan oleh masalah tertentu yang dihadapi. Dalam contoh ilustrasi, kita bisa saja memilih wilayah Barat sebagai kategori tolok ukur. Dalam hal ini hasil regresi yang diberikan dalam Persamaan (5) akan berubah, karena sekarang semua perbandingan dibuat dalam kaitannya dengan rata-rata IPM provinsi di wilayah Barat. Tentu saja, hal ini tidak akan mengubah kesimpulan keseluruhan dari contoh kita (mengapa?).
6. Kita sudah membahas tentang jebakan variabel boneka. Ada cara untuk menghindari jebakan ini yaitu dengan memasukkan variabel dummy sebanyak jumlah kategori dari variabel tersebut, asalkan kita tidak memasukkan intersep dalam model tersebut. Dengan demikian, jika kita menghilangkan intersep, sehingga diperoleh model sebagai berikut berikut,

$$IPM_i = \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i$$

(6)

Dengan persamaan ini kita tidak masuk ke dalam jebakan variabel dummy, karena tidak ada lagi kolinieritas yang sempurna. Tetapi pastikan bahwa ketika kita menjalankan regresi ini, kita menggunakan opsi no-intercept dalam model regresi kita.

Bagaimana kita menginterpretasikan regresi (6)? Jika kita mengestimasi Persamaan (6), kita akan menemukan hasil bahwa:

- $\beta_1$  = rata-rata IPM provinsi bagian Barat
- $\beta_2$  = rata-rata IPM provinsi bagian Tengah

$\beta_3$  = rata-rata IPM provinsi bagian Timur

Dependent Variable: IPM				
Method: Least Squares				
Sample: 1 38				
Included observations: 38				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D1	74.42188	0.823421	90.38130	0.0000
D2	72.02429	0.880274	81.82028	0.0000
D3	65.43375	1.164493	56.19074	0.0000
R-squared	0.533387	Mean dependent var		71.64632
Adjusted R-squared	0.506723	S.D. dependent var		4.689609
S.E. of regression	3.293685	Akaike info criterion		5.297548
Sum squared resid	379.6926	Schwarz criterion		5.426831
Log likelihood	-97.65340	Hannan-Quinn criter.		5.343545
Durbin-Watson stat	1.557556			

Dengan kata lain, dengan menampilkan hanya intersep, dan variabel dummy untuk setiap kategori, kita akan memperoleh nilai rata-rata langsung dari berbagai kategori. Seperti yang kita lihat, koefisien dummy memberikan secara langsung nilai rata-rata IPM di tiga wilayah Barat, Tengah, dan Timur.

## E. RANGKUMAN

- (1) Variabel dummy, dengan nilai 1 dan nol, merupakan cara untuk menggunakan variable kualitatif dalam model regresi.
- (2) Variabel dummy adalah alat pengklasifikasian data yang membagi sampel menjadi beberapa subkelompok berdasarkan kualitas atau atribut (jenis kelamin, status perkawinan, agama, dll.) dan secara implisit memungkinkan seseorang untuk menjalankan regresi individual untuk setiap subkelompok. Jika ada perbedaan dalam respon regresi terhadap variasi dalam variabel kualitatif dalam berbagai subkelompok, mereka akan tercermin dalam perbedaan dalam

intersep atau atau koefisien kemiringan, atau keduanya, dari berbagai regresi subkelompok.

- (3) Meskipun merupakan alat yang serbaguna, teknik variabel dummy perlu ditangani dengan hati-hati. Pertama, jika regresi mengandung istilah konstan, jumlah variabel dummy harus lebih sedikit dari jumlah klasifikasi setiap variabel kualitatif. Kedua, koefisien yang melekat pada variabel dummy harus selalu ditafsirkan dalam kaitannya dengan kelompok dasar, atau referensi, yaitu kelompok yang menerima nilai nol. Basis yang dipilih akan tergantung pada tujuan penelitian yang sedang dilakukan. Ketiga, jika sebuah model memiliki beberapa variabel kualitatif dengan beberapa kelas, penggunaan variabel dummy dapat mengurangi sejumlah besar derajat kebebasan. Oleh karena itu, seseorang harus selalu mempertimbangkan jumlah variabel boneka yang akan digunakan terhadap jumlah total observasi yang tersedia untuk analisis.

## **F. TES FORMATIF**

### **Soal 1**

Sebuah penelitian ingin mengetahui pengaruh jenis kelamin terhadap gaji karyawan. Data yang tersedia adalah gaji karyawan dan jenis kelamin (laki-laki = 1, perempuan = 0).

- (a) Jelaskan bagaimana regresi variabel dummy dapat digunakan untuk menjawab pertanyaan penelitian ini!
- (b) Tuliskan model regresi yang sesuai untuk penelitian ini!
- (c) Jelaskan interpretasi koefisien regresi untuk variabel dummy jenis kelamin!

### **Soal 2**

Sebuah perusahaan ingin mengetahui pengaruh pendidikan karyawan terhadap produktivitas kerja. Data yang tersedia adalah produktivitas kerja karyawan dan tingkat pendidikan (SD = 1, SMP = 2, SMA = 3, Diploma = 4, Sarjana = 5).

- (a)Jelaskan bagaimana regresi variabel dummy dapat digunakan untuk menjawab pertanyaan penelitian ini!
- (b)Tuliskan model regresi yang sesuai untuk penelitian ini!
- (c)Jelaskan bagaimana cara menguji hipotesis bahwa produktivitas kerja karyawan dengan pendidikan SMA berbeda dengan produktivitas kerja karyawan dengan pendidikan SD!

### Soal 3

Sebuah penelitian ingin mengetahui pengaruh jenis kelamin dan pendidikan terhadap gaji karyawan. Data yang tersedia adalah gaji karyawan, jenis kelamin (laki-laki = 1, perempuan = 0), dan tingkat pendidikan (SD = 1, SMP = 2, SMA = 3, Diploma = 4, Sarjana = 5).

- (a)Jelaskan bagaimana model regresi variabel dummy dapat digunakan untuk menjawab pertanyaan penelitian ini!
- (b)Tuliskan model regresi yang sesuai untuk penelitian ini!
- (c)Jelaskan interpretasi koefisien regresi untuk variabel dummy jenis kelamin dan pendidikan!
- (d)Bagaimana cara menguji hipotesis bahwa gaji karyawan laki-laki berbeda dengan gaji karyawan perempuan?
- (e)Bagaimana cara menguji hipotesis bahwa gaji karyawan dengan pendidikan SMA berbeda dengan gaji karyawan dengan pendidikan SD?

## G. LATIHAN

Gunakan data yang diberikan pada Lampiran 1 dan pertimbangkan model berikut ini:

$$\text{Tabungans}_i = \beta_0 + \beta_1 \text{Pendapatan}_i + \beta_3 D_i + u_i$$

di mana  $D_i = 1$  untuk tahun 2010-2023 dan  $D_i=0$  untuk 1998-2009  
 Pertanyaan:

- (a)Estimasi model diatas dan interpretasikan hasil yang Anda peroleh.
- (b)Berapa nilai intersep dari fungsi tabungan dalam dua subperiode dan dan bagaimana anda menginterpretasikannya?

Year	Konsumsi (KONS)	Pendapatan (DPI)	D
1947	976.4	1035.2	0
1948	998.1	1090.0	0
1949	1025.3	1095.6	0
1950	1090.9	1192.7	0
1951	1107.1	1227.0	0
1952	1142.4	1266.8	0
1953	1197.2	1327.5	0
1954	1221.9	1344.0	0
1955	1310.4	1433.8	0
1956	1348.8	1502.3	0
1957	1381.8	1539.5	0
1958	1393.0	1553.7	0
1959	1470.7	1623.8	0
1960	1510.8	1664.8	0
1961	1541.2	1720.0	0
1962	1617.3	1803.5	0
1963	1684.0	1871.5	0
1964	1784.8	2006.9	0
1965	1897.6	2131.0	0
1966	2006.1	2244.6	0
1967	2066.2	2340.5	0
1968	2184.2	2448.2	0
1969	2264.8	2524.3	0
1970	2317.5	2630.0	0
1971	2405.2	2745.3	0

Year	Konsumsi (KONS)	Pendapatan (DPI)	D
1972	2550.5	2874.3	0
1973	2675.9	3072.3	0
1974	2653.7	3051.9	1
1975	2710.9	3108.5	1
1976	2868.9	3243.5	1
1977	2992.1	3360.7	1
1978	3124.7	3527.5	1
1979	3203.2	3628.6	1
1980	3193.0	3658.0	1
1981	3236.0	3741.1	1
1982	3275.5	3791.7	1
1983	3454.3	3906.9	1
1984	3640.6	4207.6	1
1985	3820.9	4347.8	1
1986	3981.2	4486.6	1
1987	4113.4	4582.5	1
1988	4279.5	4784.1	1
1989	4393.7	4906.5	1
1990	4474.5	5014.2	1
1991	4466.6	5033.0	1
1992	4594.5	5189.3	1
1993	4748.9	5261.3	1
1994	4928.1	5397.2	1
1995	5075.6	5539.1	1
1996	5237.5	5677.7	1
1997	5423.9	5854.5	1
1998	5683.7	6168.6	1
2022	5968.4	6320.0	1
2023	6257.8	6539.2	1

Catatan

:

Year	Konsumsi (KONS)	Pendapatan (DPI)	D
------	--------------------	---------------------	---

Tahun = tahun kalender.

Kons = pengeluaran konsumsi riil dalam miliaran dolar

DPI= pendapatan riil pribadi yang dapat dibelanjakan dalam miliaran dolar

## **KEGIATAN BELAJAR 8**

### **MODEL REGRESI DATA PANEL**

#### **DESKRIPSI PEMBELAJARAN**

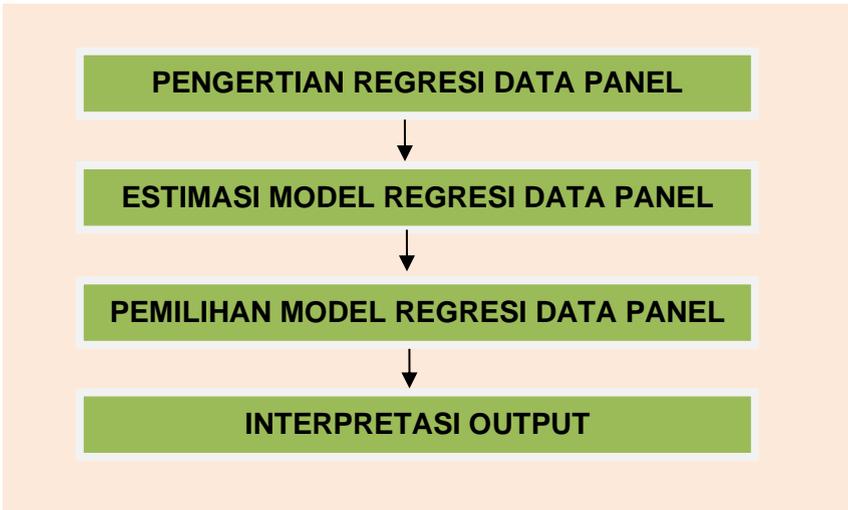
Pada bab ini mahasiswa mempelajari model regresi data panel. Diharapkan mahasiswa memiliki wawasan dan pemahaman tentang model estimasi didalam regresi data panel serta teori dalam pemilihan model estimasi terbaik yang akan digunakan. Selain itu, pada bab ini akan diberikan langka-langkah interpretasi output regresi linier data panel.

#### **KOMPETENSI PEMBELAJARAN**

Setelah mengikuti perkuliahan ini diharapkan mahasiswa memiliki pengetahuan dan kemampuan:

1. Mampu menjelaskan tentang konsep regresi data panel
2. Mampu menjelaskan tiga model estimasi dalam regresi data panel
3. Mampu menjelaskan pemilihan model estimasi terbaik
4. Mampu menjelaskan output olahan data regresi data panel

## PETA KONSEP PEMBELAJARAN



### A. PENGETERIAN REGRESI DATA PANEL

Pada sebelumnya, kita membahas model regresi logistik yang merupakan suatu teknik analisis data yang menggunakan matematika untuk menemukan hubungan antara dua faktor data, dimana hubungan ini bertujuan untuk memprediksi nilai dari salah satu faktor tersebut berdasarkan faktor yang lain. Prediksi biasanya memiliki jumlah hasil yang terbatas, seperti ya atau tidak, yang dikoding dengan angka nol dan satu. Pada bab ini kita fokus mempelajari model regresi data panel.

Model regresi data panel adalah model regresi dengan struktur data panel yaitu gabungan antara data *cross section* dengan data *time series*. Data panel diperkenalkan oleh Howles pada tahun 1950. Sebelum kita lebih jauh, mari kita ulang kembali pengertian dari data *cross section*, data *time series* dan data panel (*pooling data*).

(1) Data *cross section* merupakan data yang didapat dengan mengamati banyak subyek dalam satu waktu yang sama atau data yang pengamatannya dilakukan pada satu waktu dengan banyak objek.

Contoh: Data harga saham bank yang listed di BEI tahun 2023; Data inflasi 38 provinsi di Indonesia tahun 2023; Data gender karyawan PT.X tahun 2023

(2) Data *time series* merupakan data yang diperoleh dari amatan satu objek dari beberapa periode waktu atau data yang pengamatannya dilakukan dari waktu ke waktu (satu objek, banyak waktu).

Contoh: Data inflasi Indonesia selama periode 10 tahun; Data Tingkat suku Bunga Bank Indonesia selama periode 5 tahun; Data Laba Perusahaan PT.X selama periode 10 tahun

(3) Data panel (*pooling data*) merupakan gabungan antara data *time series* dan data *cross-section*, artinya periode waktu pengamatan dan objek yang digunakan lebih satu. Data panel umumnya digunakan ketika peneliti dihadapkan pada masalah keterbatasan jumlah sample yang terlalu sedikit sehingga untuk menambah jumlah sample dapat dilakukan data panel.

Contoh: Data Tingkat suku bunga bank Persero selama kurun waktu 5 tahun; Data inflasi 38 provinsi di Indonesia periode 2020-2023; Data Laba Perusahaan otomotif kurun waktu 5 tahun

Terdapat beberapa kelebihan penggunaan data panel dibandingkan data cross section dan data time series (Gujarati, 2012), diantaranya:

(1) Data panel menyediakan data yang lebih informatif, variabilitasnya menjadi lebih besar dan kolinearitas yang rendah. Data panel menghasilkan derajat bebas (*degrees of freedom*) yang lebih besar dan membuat lebih efisien sehingga mampu meningkatkan presisi dari estimasi yang dilakukan.

Secara matematis *degrees of freedom* adalah jumlah sample dikurangi banyaknya variabel didalam penelitian  $(n-k)$ .

- (2) Data panel mampu mengakomodasi tingkat heterogenitas individu yang tidak diobservasi namun dapat mempengaruhi hasil dari *individual heterogeneity*. Hal ini tidak dapat dilakukan oleh *time series* maupun *cross section*.
- (3) Data panel dapat mengidentifikasi dan juga mengukur efek yang tidak dapat diperoleh dari data *cross section* murni maupun data *time series* murni.
- (4) Data panel dapat memberikan informasi bagaimana kondisi individu pada waktu tertentu dibandingkan pada kondisi pada waktu yang lain.
- (5) Data panel yang memungkinkan untuk membentuk dan menguji model yang bersifat lebih rumit dibandingkan data *cross section* murni maupun data *time series* murni.
- (6) Data panel dapat meminimalkan bias yang dihasilkan oleh agregasi individu yang disebabkan oleh unit observasi yang terlalu banyak.

Model regresi data panel memiliki tujuan yang sama dengan model regresi lainnya, yaitu pengujian besarnya pengaruh satu atau lebih variabel independen terhadap sebuah variabel dependen. Namun model regresi data panel memiliki kelebihan dibandingkan dengan teknik analisis regresi lainnya, yaitu memperoleh hasil estimasi yang lebih baik karena seiring dengan peningkatan jumlah observasi yang otomatis berimplikasi pada peningkatan *degree of freedom* dan menghindari kesalahan penghilangan variabel (*omitted variable problem*). Menurut Gujarati (2012), dalam menentukan model regresi data panel terdapat beberapa kemungkinan yang terjadi antar intersep, koefisien dari slope dan variabel pengganggu (*error term*), diantaranya:

- (1) Intersep dan koefisien dari *slope* konstan sepanjang periode waktu dan individu, *error term* berbeda sepanjang periode waktu dan individu.

- (2) Koefisien *slope* konstan, tetapi intersep bervariasi sepanjang individu.
- (3) Koefisien *slope* konstan, tetapi intersep bervariasi sepanjang waktu dan individu.
- (4) Intersep dan koefisien *slope* bervariasi sepanjang individu.
- (5) Intersep dan koefisien *slope* bervariasi sepanjang individu.

Beberapa kemungkinan yang akan terjadi di atas menunjukkan bahwa semakin banyak variabel independen maka semakin kompleks estimasi parameter. Oleh karena itu model regresi data panel membutuhkan beberapa metode untuk melakukan estimasi dari parameter yang akan menunjukkan besarnya pengaruh variabel independen terhadap variabel dependen. Metode yang akan digunakan dalam model regresi data panel antara lain menggunakan pendekatan model estimasi *common effect*, *fixed effect* dan *random effect*.

## **B. ESTIMASI MODEL DALAM REGRESI LINIER DATA PANEL**

Dalam mengestimasi model regresi panel, metode yang akan digunakan sangat bergantung pada asumsi yang dibuat yaitu mengenai intersep, koefisien *slope* dan *error term*. Berdasarkan dari berbagai asumsi dan faktor pembentukannya, model regresi data panel dibagi menjadi tiga pendekatan yaitu:

### **(1) *Common Effect Model (CEM)***

Merupakan suatu teknik untuk mengestimasi parameter model data panel, dengan mengkombinasikan data *cross section* dan *time series* sebagai satu kesatuan tanpa melihat adanya perbedaan waktu dan entitas (individu). Dimana pendekatan yang sering dipakai adalah metode *Ordinary Least Square (OLS)*. Model *Common Effect* mengabaikan adanya perbedaan dimensi individu maupun waktu atau dengan kata lain perilaku data antar individu sama dalam berbagai kurun

waktu. Persamaan *common effect model* dapat dinyatakan sebagai berikut:

$$Y_{it} = \beta_0 + \sum_{k=1}^n \beta_k X_{it} + e_{it}$$

Keterangan:

$Y_{it}$  = Variabel dependen pada unit observasi ke-i dan waktu ke-t.

$\beta_0$  = Intersep model regresi pada unit observasi ke-i dan waktu ke-t.

$\beta_1$  = Koefisien slope.

$X_{it}$  = Variabel independen untuk unit observasi ke-i periode waktu ke-t.

$e_{it}$  = *error term* atau residual pada unit observasi ke-i dan waktu ke-t.

$i$  = Unit *cross section* (1, 2, 3, ..., N).

$t$  = Unit *time series* (1, 2, 3, ..., T).

$k$  = Jumlah variabel independen (1, 2, 3, ..., n)

## (2) **Fixed Effect Model (FEM)**

*Fixed Effect Model* mengasumsikan bahwa intersep dari setiap individu adalah berbeda sedangkan slope antar individu adalah tetap (sama). Adanya variabel-variabel yang tidak semuanya masuk dalam model memungkinkan adanya intersep yang tidak konstan. Intersep ini bisa berubah untuk setiap individu dan waktu.

Teknik ini menggunakan variabel dummy untuk menangkap adanya perbedaan intersep antar individu, perbedaan intersep bisa terjadi karena perbedaan budaya kerja, manajerial, dan insentif. Namun demikian sloponya sama antar Perusahaan. Model estimasi ini sering juga disebut dengan teknik *Least Squares Dummy Variable* (LSDV). Persamaan *fixed effect model* dapat dinyatakan sebagai berikut:

$$Y_{it} = \beta_0 + \sum_{k=1}^n \beta_k X_{it} + e_{it}$$

Keterangan:

$Y_{it}$  = Variabel dependen pada unit observasi ke- $i$  dan waktu ke- $t$ .

$\beta_0$  = Intersep model regresi pada unit observasi ke- $i$  dan waktu ke- $t$ .

$\beta_1$  = Koefisien slope.

$X_{it}$  = Variabel independen untuk unit observasi ke- $i$  periode waktu ke- $t$ .

$e_{it}$  = *error term* atau residual pada unit observasi ke- $i$  dan waktu ke- $t$ .

$i$  = Unit *cross section* (1, 2, 3, ..., N).

$t$  = Unit *time series* (1, 2, 3, ..., T).

$k$  = Jumlah variabel independen (1, 2, 3, ..., n)

### (3) **Random Effect Model (REM)**

Model *Random Effect* mengasumsikan setiap perusahaan mempunyai perbedaan intersep, yang mana intersep tersebut adalah variabel random atau stokastik. Model ini sangat berguna jika individu (entitas) yang diambil sebagai sampel adalah dipilih secara random dan merupakan wakil populasi. Teknik yang memungkinkan bahwa error mungkin berkorelasi sepanjang *time series* dan *cross section*.

Pada model *Fixed Effect*, perbedaan karakteristik individu dan waktu diakomodasi pada intersep sehingga intersepanya dapat berubah antar individu dan antar waktu, sedangkan pada model *Random Effect* perbedaan intersep diakomodasi oleh error masing-masing individu, error komponen waktu dan error gabungan. Model ini juga disebut dengan *Error Component Model* (ECM) atau teknik *Generalized Least Square* (GLS). Persamaan *random effect model* dapat dinyatakan sebagai berikut:

$$Y_{it} = \beta_0 + \sum_{k=1}^n \beta_k X_{it} + \mu_i + e_{it}$$

Keterangan:

$Y_{it}$  = Variabel dependen pada unit observasi ke-i dan waktu ke-t.

$\beta_0$  = Intersep model regresi pada unit observasi ke-i dan waktu ke-t.

$\beta_1$  = Koefisien slope.

$X_{it}$  = Variabel independen untuk unit observasi ke-i periode waktu ke-t.

$\mu_{it}$  = *error term* atau residual pada unit observasi ke-i.

$e_{it}$  = *error term* atau residual pada unit observasi ke-i dan waktu ke-t.

$i$  = Unit *cross section* (1, 2, 3, ..., N).

$t$  = Unit *time series* (1, 2, 3, ..., T).

$k$  = Jumlah variabel independen (1, 2, 3, ..., n)

### C. PEMILIHAN ESTIMASI MODEL REGRESI DATA PANEL

Pada model *Fixed Effect*, perbedaan karakteristik individu dan Pemilihan estimasi model regresi data panel ini bertujuan untuk memilih model terbaik yang tepat dan sesuai dari ketiga model estimasi regresi data panel yang terdiri dari tiga estimasi yaitu *Common effect Model*, *Fixed Effect Model*, dan *Random effect Model*. Dalam pemilihan estimasi model regresi data panel terbaik dapat menggunakan dua pendekatan yaitu secara kepakaran ahli ekonometrika dan secara statistic.

Menurut Judge et al dalam Gujarati (2012), pada saat kita dihadapkan dengan pilihan untuk memilih antara *Fixed Effect Model* atau *Random effect Model* menurut beberapa ahli ekonometrika menyatakan:

- (1) Pada saat periode waktu lebih banyak daripada jumlah individu yang digunakan ( $T > N$ ) maka pilihlah estimasi menggunakan *Fixed Effect Model*, karena ada kemungkinan perbedaan nilai parameter yang diestimasi dengan *Fixed effect model* dan *Random effect model* cukup kecil.
- (2) Pada saat periode waktu lebih sedikit daripada jumlah individu yang digunakan ( $T < N$ ) maka pilihlah estimasi menggunakan *Random Effect Model*. Pada kondisi ini estimasi antara *Fixed effect model* dan *random effect model* akan menghasilkan estimasi yang berbeda secara signifikan, pilihan ditentukan berdasarkan keyakinan apakah individu yang diobservasi merupakan sample acak yang diambil dari populasi tertentu atau tidak. Jika diambil dari sample acak lebih baik menggunakan *Random effect model*, dan sebaliknya jika diambil dari sample bukan acak lebih baik menggunakan *Fixed effect model*.
- (3) *Random Effect Model* mempunyai parameter lebih sedikit sehingga berakibat pada degrees of freedom nya lebih besar, namun pada *Fixed Effect Model* pada membedakan efek individual dan efek waktu. *Fixed Effect Model* juga tidak perlu mengasumsikan bahwa komponen error tidak berkorelasi dengan variabel bebas yang mungkin sulit dipenuhi.

Pemilihan estimasi model yang paling tepat secara statistik dapat menggunakan tiga pengujian yaitu *Chow Test*, *Hausman Test* dan *Langrange Multiplier Test*.

#### **(1) Chow Test**

Chow test adalah pengujian untuk menentukan model *Fixed Effect* atau *Random Effect* yang paling tepat digunakan dalam mengestimasi data panel.

#### **Hipotesa**

$H_0$ : *Common Effect Model*

$H_a$ : *Individual Effect (Fixed Effect Model)*

## Pengujian Statistik

$$F_{hitung} = \frac{(RSS_1 - RSS_2)/(K - 1)}{RSS_2/(KT - K - P)} \sim F_{(\alpha, (K-1), (KT-K-P))}$$

Keterangan:

K = jumlah individu observasi.

T = jumlah periode waktu observasi.

P = jumlah parameter dalam model *fixed effect*.

RSS<sub>1</sub> = residual sum of squares dari model *common effect*.

RSS<sub>2</sub> = residual sum of squares dari *fixed effect*.

### Pengambilan Keputusan

Jika nilai Prob <  $\alpha$  (Alpha: 1%, 5%, 10%) maka H<sub>0</sub> ditolak, dan sebaliknya. Atau

Jika nilai F<sub>hitung</sub> > F<sub>tabel</sub> ( $F_{(\alpha, (K-1), (KT-K-P))}$ ) maka H<sub>0</sub> ditolak, dan sebaliknya

### (2) Hausman Test

Hausman test adalah pengujian statistik untuk memilih apakah model Fixed Effect atau Random Effect yang paling tepat digunakan

#### Hipotesa

H<sub>0</sub>: *Random Effect Model*

H<sub>a</sub>: *Fixed Effect Model*

### Pengujian Statistik

$$w = X_{(p)}^2 = [b - \beta]' \varphi^{-1} [b - \beta]$$

dimana

$$\varphi = Var [b] - Var [\beta]$$

Keterangan:

$b$  = parameter (tanpa intersep) model *random effect*.

$\beta$  = parameter model *fixed effect*.

$\text{Var}[b]$  = matriks kovarian parameter (tanpa intersep) model *random effect*.

$\text{Var}[\beta]$  = matriks kovarian parameter model *fixed model*.

### **Pengambilan Keputusan**

Jika nilai  $\text{Prob} < \alpha$  (Alpha: 1%, 5%, 10%) maka  $H_0$  ditolak dan sebaliknya. Atau

Jika nilai  $W_{\text{hitung}} > X_{\alpha,p}^2$  maka  $H_0$  ditolak dan sebaliknya

### **(3) Langrange Multiplier Test**

Uji Lagrange Multiplier (LM) adalah uji untuk mengetahui apakah model *Random Effect* lebih baik daripada metode *Common Effect* (OLS) yang digunakan.

#### **Hipotesa**

$H_0$ : *Common Effect Model*

$H_a$ : *Random Effect Model*

#### **Pengujian Statistik**

$$LM = \frac{KT}{2(T-1)} \left[ \frac{\sum_{i=1}^K [\sum_{t=1}^T e_{it}]^2}{\sum_{i=1}^K \sum_{t=1}^T e_{it}^2} - 1 \right]^2 \sim X_{\alpha,1}^2$$

Dimana:

$K$  = jumlah individu

$T$  = jumlah periode waktu

$e_{it}$  = residual atau error dari model *common effect*.

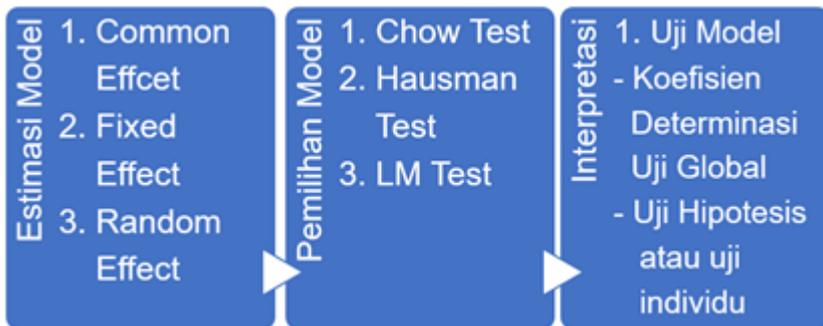
#### **Pengambilan Keputusan**

Jika nilai  $\text{Prob} < \alpha$  (Alpha: 1%, 5%, 10%) maka  $H_0$  ditolak dan sebaliknya. Atau

Jika nilai  $LM_{\text{hitung}} > X_{\alpha-1}^2$  maka  $H_0$  ditolak dan sebaliknya

#### D. LANGKAH-LANGKAH INTERPRETASI OUTPUT MODEL REGRESI DATA PANEL

Pada tahap ini kita akan membuat simulasi proses atau Langkah-langkah interpretasi dari output model regresi data panel, dengan tahapan yang dapat dilihat pada gambar 1. Tahap pertama kita melakukan estimasi model regresi data panel untuk mengeluarkan tiga estimasi yaitu *Common Effect Model*, *Fixed Effect Model* dan *Random Effect Model*. Tahap kedua memilih dari ketiga estimasi yang dihasilkan ditahap pertama mana yang terbaik mengestimasi model kita, dilakukan dengan tiga kali pengujian yaitu *Chow Test*, *Hausman Test* dan *LM Test*. Tahap ketiga, setelah kita mengetahui secara statistik model mana yang paling baik mengestimasi persamaan yang diajukan maka tahap selanjutnya melakukan interpretasi untuk Uji Model yang terdiri dari Uji Koefisien Determinasi dan Uji Global (Uji F), kemudian dilanjutkan dengan Uji hipotesa atau Uji individu (Uji t).



Gambar 8.1. Tahapan Analisis Model Regresi Data Panel

#### Contoh Kasus

Suatu penelitian bertujuan untuk menguji pengaruh dari aspek kesejahteraan masyarakat terhadap pertumbuhan inklusif di Indonesia. Aspek kesejahteraan dalam penelitian ini merupakan variabel independen yang terdiri dari Gini Ratio (%), Rata-Rata

Lama Sekolah (Tahun), Rata-Rata Pengeluaran Perkapita (Ribu Rp) dan Angka Harapan Hidup (Tahun), sedangkan variabel dependen adalah pertumbuhan inklusif (%). Semua data yang digunakan merupakan data sekunder yang bersumber dari Badan Pusat Statistik (BPS) dan BAPPENAS, yang menggunakan data tingkat provinsi sebanyak 34 provinsi di Indonesia dengan periode penelitian 2018-2022. Hasil pengujian menggunakan software *Eview's 10*, ditunjukkan tabel 1.

**Tabel 1 Hasil Pengolahan Data Model Regresi Data Panel**

Variabel	CEM		FEM		REM	
	Beta	Prob	Beta	Prob	Beta	Prob
Constanta	-0,013245	0,9886	-45,00977	0,0000	-1,300570	0,3913
GR	-3,007358	0,0001	-3,150005	0,1391	-4,117538	0,0003
RPP	0,000176	0,0000	0,000558	0,0000	0,000180	0,0000
RLS	0,126254	0,0012	-0,697690	0,0192	0,175985	0,0041
AHH	0,056597	0,0001	0,742081	0,0000	0,073907	0,0017
<b>Goodness of Fit Test</b>						
R-square	0,710048		0,902492		0,487742	
Adj <sup>2</sup> square	0,703019		0,875160		0,475323	
F-stat	101,0149		33,01971		39,27576	
Prob F-stat	0,000000		0,000000		0,000000	
<b>Uji Pemilihan Model</b>						
<i>Chow Test</i>	Cross-section X <sup>2</sup>		185,2620		Prob	0,0000
<i>Hausman Test</i>	Cross-section Random		50,1025		Prob	0,0000
<i>LM Test</i>	Breusch Pagan (Both)		274,2181		Prob	0,0000

Sumber: Data Diolah (2024)

### Hipotesa

- H<sub>1</sub>: Terdapat pengaruh negatif Gini Ratio terhadap Pertumbuhan Inklusif di Indonesia
- H<sub>2</sub>: Terdapat pengaruh positif Rata-Rata Pengeluaran Perkapita terhadap Pertumbuhan Inklusif di Indonesia
- H<sub>3</sub>: Terdapat pengaruh positif Rata-Rata Lama Sekolah terhadap Pertumbuhan Inklusif di Indonesia
- H<sub>4</sub>: Terdapat pengaruh positif Angka Harapan Hidup terhadap Pertumbuhan Inklusif di Indonesia

H<sub>5</sub>: Paling tidak terdapat satu variabel independen yang berpengaruh terhadap Pertumbuhan Inklusif di Indonesia

### **Pemilihan Model Estimasi**

Berdasarkan hasil pengujian diketahui bahwa *Chow Test* menghasilkan nilai prob dari *Cross-section Chi Square* lebih kecil dari 0,05 (alpha 5%) maka diputuskan menggunakan *individual Effect* yang diwakili oleh *Fixed Effect Model*. Tahap selanjutnya pengujian melalui *LM Test* untuk melihat apakah benar tidak ada perbedaan antar individu. Hasil pengujian *LM Test* menunjukkan nilai prob kurang dari 0,05 artinya ada perbedaan antar individu, maka tahap selanjutnya memilih dari *individual effect (Fixed effect vs Random effect)* mana yang terbaik untuk mengestimasi model melalui *Hausman Test*. Hasil pengujian dari *Hausman Test* menunjukkan nilai prob lebih kecil dari 0,05 (alpha 5%) maka disimpulkan model menggunakan hasil estimasi *Fixed Effect Model*. Keputusan uji pemilihan model regresi data panel menyimpulkan model yang terbaik untuk mengestimasi adalah dengan *Fixed Effect Model*, maka persamaan ekonometrika yang akan dibentuk adalah sebagai berikut:

$$Y_{it} = -45,00977 - 3,150005 GR_{it} + 0,000558 RPP_{it} - 0,697690 RLS_{it} + 0,742081 AHH_{it} + e_{it}$$

### **Goodness of Fit Test**

#### **(1) Uji Koefisien Determinasi**

Berdasarkan hasil pengolahan data model yang terpilih yaitu *Fixed Effect Model* diketahui nilai dari *Adj R2* sebesar 0,875160 artinya kemampuan variabel independen (*Gini Ratio*, *Rata-Rata Lama Sekolah*, *Rata-Rata Pengeluaran Perkapita* dan *Angka Harapan Hidup*) dalam menjelaskan variasi dari variabel dependen (*Pertumbuhan Inklusif*) sebesar 87,51% sedangkan sisa nya dijelaskan oleh variabel independen lain yang tidak dimasukkan kedalam model.

## **(2) Uji Global (Uji F)**

Hasil pengujian global (Uji F) menunjukkan bahwa nilai probabilitas dari  $F_{\text{stat}}$  sebesar  $0.0000 < 0,05$  (alpha 5%) maka  $H_5$  diterima, disimpulkan secara statistik paling tidak terdapat satu variabel independen dalam penelitian ini yang berpengaruh terhadap pertumbuhan inklusif di Indonesia.

## **Uji Individu (Uji t)**

### **(1) Pengaruh Gini Ratio terhadap Pertumbuhan Inklusif di Indonesia**

Berdasarkan hasil pengujian statistik diketahui besarnya koefisien dari **Gini Ratio (GR)** adalah sebesar  $-3,150005$  artinya jika **Gini Ratio** naik sebesar satu satuan maka secara rata-rata **Pertumbuhan Inklusif** akan turun sebesar  $3,150005$  satuan. Hasil pengujian menunjukkan tanda koefisien yang sesuai dengan teori yang diajukan dalam penelitian ini, dimana secara teori Gini Ratio memiliki pengaruh negatif terhadap Pertumbuhan Inklusif. Oleh karena itu pengujian signifikansi dapat dilanjutkan. Hasil pengujian menunjukkan nilai Prob sebesar  $0,0695$  ( $0.1391/2$ )  $< 0.10$  (alpha 10%) maka diputuskan  $H_1$  diterima. Disimpulkan secara statistik pada tingkat kepercayaan 90% terdapat pengaruh negatif Gini Ratio terhadap Pertumbuhan Inklusif, yang artinya Gini Ratio mampu menurunkan Pertumbuhan Inklusif di Indonesia.

### **(2) Pengaruh Rata-Rata Pengeluaran Perkapita terhadap Pertumbuhan Inklusif di Indonesia**

Berdasarkan hasil pengujian statistik diketahui besarnya koefisien dari Rata-Rata Pengeluaran Perkapita (RPP) adalah sebesar  $0,000558$  artinya jika Rata-Rata Pengeluaran Perkapita naik sebesar satu satuan maka secara rata-rata Pertumbuhan Inklusif akan naik sebesar  $0,000558$  satuan. Hasil pengujian menunjukkan tanda koefisien yang sesuai dengan teori yang diajukan dalam

penelitian ini, dimana secara teori Rata-Rata Pengeluaran Perkapita memiliki pengaruh positif terhadap Pertumbuhan Inklusif. Oleh karena itu pengujian signifikansi dapat dilanjutkan. Hasil pengujian menunjukkan nilai Prob sebesar  $0,0000 (0.0000/2) < 0.01$  (alpha 1%) maka diputuskan  $H_2$  diterima. Disimpulkan secara statistik pada tingkat kepercayaan 99% terdapat pengaruh positif Rata-Rata Pengeluaran Perkapita terhadap Pertumbuhan Inklusif, yang artinya Rata-Rata Pengeluaran Perkapita mampu meningkatkan Pertumbuhan Inklusif di Indonesia.

### **(3) Pengaruh Rata-Rata Lama Sekolah terhadap Pertumbuhan Inklusif di Indonesia**

Berdasarkan hasil pengujian statistik diketahui besarnya koefisien dari Rata-Rata Lama Sekolah (RLS) adalah sebesar  $-0,697690$  artinya jika Rata-Rata Lama Sekolah naik sebesar satu satuan maka secara rata-rata Pertumbuhan Inklusif akan turun sebesar  $0,697690$  satuan. Hasil pengujian menunjukkan tanda koefisien yang tidak sesuai dengan teori yang diajukan dalam penelitian ini, dimana secara teori Rata-Rata Lama Sekolah memiliki pengaruh positif terhadap Pertumbuhan Inklusif. Oleh karena itu pengujian signifikansi tidak dapat dilanjutkan dan diputuskan  $H_3$  ditolak. Disimpulkan secara statistik tidak terdapat pengaruh Rata-Rata Lama Sekolah terhadap Pertumbuhan Inklusif, yang artinya Rata-Rata Lama Sekolah tidak mampu meningkatkan Pertumbuhan Inklusif di Indonesia.

### **(4) Pengaruh Angka Harapan Hidup terhadap Pertumbuhan Inklusif di Indonesia**

Berdasarkan hasil pengujian statistik diketahui besarnya koefisien dari Angka Harapan Hidup (AHH) adalah sebesar  $0,742081$  artinya jika Angka Harapan Hidup naik sebesar satu satuan maka secara rata-rata Pertumbuhan Inklusif

akan naik sebesar 0,742081 satuan. Hasil pengujian menunjukkan tanda koefisien yang sesuai dengan teori yang diajukan dalam penelitian ini, dimana secara teori Angka Harapan Hidup memiliki pengaruh positif terhadap Pertumbuhan Inklusif. Oleh karena itu pengujian signifikansi dapat dilanjutkan. Hasil pengujian menunjukkan nilai Prob sebesar  $0,0000$  ( $0.0000/2$ )  $< 0.01$  (alpha 1%) maka diputuskan  $H_4$  diterima. Disimpulkan secara statistik pada tingkat kepercayaan 99% terdapat pengaruh positif Angka Harapan Hidup terhadap Pertumbuhan Inklusif, yang artinya Angka Harapan Hidup mampu meningkatkan Pertumbuhan Inklusif di Indonesia.

## E. RANGKUMAN

1. Model regresi data panel memiliki tujuan yang sama dengan regresi lain pada umumnya, yaitu penguji pengaruh variabel independen terhadap variabel dependen. Hanya saja struktur data yang digunakan ada data panel (gabungan *data cross section* dan *time series*).
2. Model regresi data panel dapat mengatasi beberapa kelemahan seperti jumlah sample yang sedikit, karena penggunaan data panel menyebabkan peningkatan jumlah observasi yang otomatis berimplikasi pada peningkatan derajat kebebasan dan menghindari kesalahan penghilangan variabel kompleksitas.
3. Model regresi data panel akan menghasilkan tiga estimasi yang terdiri dari *Common Effect Model*, *Fixed Effect Model* dan *Random Effect Model*. Masing-masing metode yang digunakan adalah *Ordinary Least Square*, *Least Squares Dummy Variable* dan *Generalized Least Square*.
4. Terdapat tiga pengujian untuk memilih estimasi terbaik dalam model yaitu *Chow Test* (*Common Effect Model* vs *Fixed Effect Model*), *LM Test* (*Common Effect Model* vs *Random Effect*

Model) dan Hausman Test (Fixed Effect Model vs Random Effect Model).

## F. TES FORMATIF

Pengangguran merupakan permasalahan yang tidak bias dipisahkan dari perekonomian suatu negara. Masalah pengangguran menjadi perhatian pemerintah, oleh karena itu suatu penelitian dilakukan untuk menguji factor yang mempengaruhi Tingkat Pengangguran Terbuka (%) Kabupaten/Kota di propinsi Banten selama kurun waktu 2017-2022. Dalam penelitian ini menggunakan variable independent antara lain Jumlah penduduk (Ribu JIwa), Upah Minimum Propinsi (Ribu Rp), dan Pendapatan Dosmestik Regional Bruto (Juta Rp). Hasil pengolahan data menunjukkan rangkuman sebagai berikut:

Variabel	CEM		FEM		REM	
	Beta	Prob	Beta	Prob	Beta	Prob
Constanta	-9,12082	0,0000	9,835139	0,1641	-5,671898	0,3828
POP	1,135057	0,0000	0,341849	0,0000	0,373045	0,0000
UMP	0,059432	0,0005	0,618207	0,0001	0,000298	0,8939
PDRB	-0,167650	0,3066	-1,040667	0,0657	-1,203838	0,1784
<b>Goodness of Fit Test</b>						
R-square	0,303851		0,998125		0,132695	
Adj <sup>2</sup> square	0,280333		0,997741		0,103394	
F-stat	12,91966		2600,655		4,528692	
Prob F-stat	0,000000		0,000000		0,000717	
<b>Uji Pemilihan Model</b>						
<i>Chow Test</i>	Cross-section X <sup>2</sup>		830,762349		Prob	0,0000
<i>Hausman Test</i>	Cross-section Random		24,652963		Prob	0,0002
<i>LM Test</i>	Breusch Pagan (Both)		370,2871		Prob	0,0000

Pertanyaan:

- a. Buatlah hipotesa yang akan diajukan, jelaskan secara teori ekonomi.

- b. Lakukanlah estimasi *Common Effect Model*, *Fixed Effect Model* dan *Random Effect Model*, menurut anda lebih baik menggunakan hasil estimasi model apa.
- c. Lakukanlah interpretasi sampai ke pengujian secara individu.

## G. LATIHAN

Suatu Penelitian bertujuan untuk melihat factor-faktor yang mempengaruhi *Stock Underpricing* (UNPRICE) Di Bursa Efek Indonesia (BEI), dimana variabel independen dalam Penelitian ini antara lain: Umur perusahaan (UMUR), ukuran perusahaan (SIZE), prosentase saham yang ditawarkan (PST), dan *earning per share* (TPD). Hasil pengolahan data menunjukkan rangkuman sebagai berikut:

Variabel	CEM		FEM		REM	
	Beta	Prob	Beta	Prob	Beta	Prob
Constanta	23,98234	0,0060	144,4804	0,0936	24,86963	0,1508
UMUR	0,012950	0,0960	0,008343	0,0042	0,000169	0,0775
SIZE	0,023728	0,0098	0,009495	0,1490	0,000515	0,0340
PST	0,024498	0,0219	0,032604	0,0993	0,012793	0,0065
TDP	0,003817	0,1408	0,003298	0,0865	0,024256	0,0296
<b>Goodness of Fit Test</b>						
R-square	0,874845		0,953498		0,849756	
Adj <sup>2</sup> square	0,699627		0,813993		0,639414	
F-stat	4,992912		6,834860		4,039882	
Prob F-stat	0,007370		0,012521		0,015996	
<b>Uji Pemilihan Model</b>						
<i>Chow Test</i>	Cross-section X <sup>2</sup>	19,270657		Prob	0,0007	
<i>Hausman Test</i>	Cross-section Random	2,14679		Prob	0,0562	
<i>LM Test</i>	Breusch Pagan (Both)	70,30001		Prob	0,0052	

Pertanyaannya, berdasarkan lampiran data:

- a) Buatlah hipotesa yang akan diajukan, jelaskan secara teori ekonomi.

- b) Lakukanlah estimasi *Common Effect Model*, *Fixed Effect Model* dan *Random Effect Model*, menurut anda lebih baik menggunakan hasil estimasi model apa.
- c) Lakukanlah interpretasi sampai ke pengujian secara individu.

## KEGIATAN BELAJAR 9

### MODEL PERSAMAAN SIMULTAN

#### DESKRIPSI PEMBELAJARAN

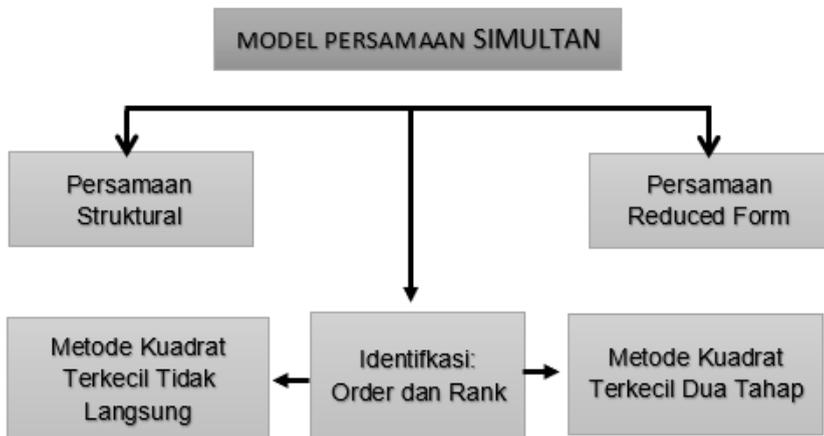
Pada bab ini mahasiswa dibekali kemampuan memahami, menjelaskan dan menghitung, serta interpretasi tentang model persamaan simultan. Hasil metode identifikasi dapat digunakan

#### KOMPETENSI PEMBELAJARAN

Setelah mengikuti perkuliahan ini diharapkan mahasiswa dan mahasiswi memiliki pengetahuan dan kemampuan :

1. Memahami masalah simultanitas dan konsekuensinya.
2. Memahami identifikasi persamaan simultan.
3. Memahami dan menggunakan metode kuadrat terkecil tidak langsung
4. Memahami dan menggunakan metode kuadrat terkecil dua tahap.

#### PETA KONSEP PEMBELAJARAN



## A. PENGERTIAN MODEL SIMULTAN

Pada kenyataannya banyak situasi dimana hubungan sebab akibat tidak hanya terjadi satu arah, tetapi terjadi dua arah. Sistem atau model persamaan simultan mengacu pada kasus di mana variabel terikat dalam satu atau lebih persamaan juga merupakan variabel penjelas dalam beberapa persamaan sistem lainnya. Secara khusus, variabel Y tidak hanya ditentukan oleh X saja, tetapi juga X, pada gilirannya, ditentukan oleh Y, sehingga Y dan X ditentukan secara bersama-sama atau bersamaan.

Untuk kasus seperti ini, pendugaan koefisien fungsi dengan menggunakan model persamaan tunggal akan menyebabkan bias dan tidak konsisten. Sehingga untuk menduga model seperti itu digunakan model persamaan simultan. Bias persamaan simultan mengakibatkan hasil estimasi yang terlalu tinggi atau terlalu rendah terhadap parameter struktural yang diperoleh dari penerapan OLS pada persamaan struktural model persamaan simultan. Bias ini terjadi karena variabel endogen sistem yang juga merupakan variabel penjelas berkorelasi dengan error term, sehingga melanggar asumsi OLS

Jika variabel terikat dalam suatu persamaan juga merupakan variabel penjelas dalam persamaan lain, maka kita mempunyai sistem atau model persamaan simultan. Variabel terikat dalam sistem persamaan simultan disebut variabel endogen. Variabel yang ditentukan oleh faktor di luar model disebut variabel eksogen.

Karena itu, persamaan simultan mengenal beberapa istilah variabel berikut. Variabel endogen yaitu merupakan variabel terikat dalam sistem persamaan simultan. Ini adalah variabel-variabel yang ditentukan oleh sistem, meskipun variabel-variabel tersebut juga muncul sebagai variabel penjelas dalam beberapa persamaan sistem lainnya. Variabel eksogen adalah variabel yang ditentukan di luar model. Ini juga termasuk variabel endogen lag,

karena nilainya sudah diketahui pada periode sebelumnya. Variabel eksogen dan variabel endogen yang lag juga disebut variabel yang telah ditentukan sebelumnya (*predetermined variable*).

## **B. PERSAMAAN STRUKTURAL DAN REDUCE FORM**

Model persamaan *reduced form* adalah persamaan yang dibentuk dari persamaan struktural sedemikian rupa sehingga masing-masing variabel endogen dalam model merupakan fungsi dari semua variabel *predetermined* dan *error*. Karena itu, persamaan *reduced form* menjadi penaksir koefisien regresi di fungsi persamaan struktural.

### **1) Reduced Form Model Uang Beredar**

Berikut diberikan dua persamaan struktural di model makroekonomi. Persamaan struktural pertama adalah jumlah uang beredar ( $M_t$ ) yang ditentukan pendapatan nasional ( $Y_t$ ) dan faktor gangguan ( $u_{1t}$ ). Persamaan structural kedua adalah pendapatan nasional ( $Y_t$ ) yang juga ditentukan oleh uang beredar ( $M_t$ ) dan investasi ( $I_t$ ).

Karena  $M$  bergantung pada  $Y$  pada persamaan pertama dan  $Y$  bergantung pada  $M$  dan  $I$  pada persamaan kedua, maka  $M$  dan  $Y$  ditentukan secara bersama-sama, sehingga merupakan model persamaan simultan.  $M$  dan  $Y$  merupakan variabel endogen, sedangkan  $I$  merupakan variabel eksogen atau ditentukan di luar model. Perubahan  $u_{1t}$  mempengaruhi  $M_t$  pada persamaan pertama. Hal ini, pada gilirannya, mempengaruhi  $Y$ , pada persamaan kedua. Akibatnya,  $Y_t$ , dan  $u_{1t}$ , berkorelasi, sehingga menyebabkan estimasi OLS menjadi bias dan tidak konsisten pada persamaan  $M$  dan  $Y$

$$(11.1) \quad M_t = a_0 + a_1 Y_t + u_{1t}$$

$$(11.2) \quad Y_t = b_0 + b_1 M_t + b_2 I_t + u_{2t}$$

Persamaan *reduced-form* pertama diperoleh dengan mensubstitusi persamaan (11.2) ke persamaan (11.1) menghasilkan:

$$M_t = a_0 + a_1(b_0 + b_1 M_t + b_2 I_t + u_{2t}) + u_{1t}$$

$$M_t = \frac{a_0 + a_1 b_0}{1 - a_1 b_1} + \frac{a_1 b_2}{1 - a_1 b_1} I_t + \frac{u_{1t} + a_1 u_{2t}}{1 - a_1 b_1}$$

Menyederhakan penulisan persamaan, yaitu

$$\pi_0 = \frac{a_0 + a_1 b_0}{1 - a_1 b_1} \quad \pi_1 = \frac{a_1 b_2}{1 - a_1 b_1} \quad v_{1t} = \frac{u_{1t} + a_1 u_{2t}}{1 - a_1 b_1}$$

Maka diperoleh

$$(11.3) \quad M_t = \pi_0 + \pi_1 I_t + v_{1t}$$

Persamaan *reduced-form* kedua diperoleh dengan mensubstitusi persamaan (11.1) ke persamaan (11.2) menghasilkan:

$$Y_t = b_0 + b_1(a_0 + a_1 Y_t + u_{1t}) + b_2 I_t + u_{2t}$$

$$Y_t = \frac{a_0 b_1 + b_0}{1 - a_1 b_1} + \frac{b_2}{1 - a_1 b_1} I_t + \frac{b_1 u_{1t} + u_{2t}}{1 - a_1 b_1}$$

Menyederhakan penulisan persamaan, yaitu

$$\pi_2 = \frac{a_0 b_1 + b_0}{1 - a_1 b_1} \quad \pi_3 = \frac{b_2}{1 - a_1 b_1} \quad v_{2t} = \frac{b_1 u_{1t} + u_{2t}}{1 - a_1 b_1}$$

$$(11.4) \quad Y_t = \pi_2 + \pi_3 I_t + v_{2t}$$

## 2) Reduced Form Model Keseimbangan

Namun, analisis ekonomi menunjukkan bahwa harga dan kuantitas biasanya ditentukan secara bersamaan oleh proses pasar, dan oleh karena itu model pasar yang lengkap tidak ditangkap oleh satu persamaan tunggal namun terdiri dari tiga persamaan yang berbeda: fungsi permintaan, fungsi penawaran, dan kondisi keseimbangan di pasar produk. Jadi kita punya:

$$(11.4) \quad Q_t^d = \beta_1 + \beta_2 P_t + \beta_3 Y_t + u_{1t}$$

$$(11.5) \quad Q_t^s = \gamma_1 + \gamma_2 P_t + u_{2t}$$

$$(11.6) \quad Q_t^d = Q_t^s$$

Di mana,  $Q_t^s$  menunjukkan kuantitas yang ditawarkan,  $Q_t^d$  adalah kuantitas yang diminta,  $P_t$  adalah harga relatif barang tersebut, dan  $Y_t$  adalah pendapatan. Persamaan (11.4), (11.5) dan (11.6) disebut persamaan struktural model persamaan simultan, dan koefisien  $\beta$  dan  $\gamma$  disebut parameter struktural.

Karena harga dan kuantitas ditentukan secara bersama-sama maka merupakan variabel endogen, dan karena pendapatan tidak ditentukan dalam model maka dikategorikan sebagai variabel eksogen. Perhatikan di sini bahwa dalam model persamaan tunggal kami menggunakan istilah variabel eksogen dan variabel penjelas secara bergantian, namun hal ini tidak lagi mungkin dilakukan dalam model persamaan simultan. Jadi mempunyai harga sebagai variabel penjelas tetapi bukan sebagai variabel eksogen juga.

Menyamakan (11.4) dengan (11.5) dan menyelesaikan  $P_t$  diperoleh:

$$(11.7) \quad P_t = \frac{\beta_1 - \gamma_1}{\beta_2 - \gamma_2} + \frac{\beta_3}{\beta_2 - \gamma_2} Y_t + \beta_3 Y_t + \frac{u_{1t} - u_{2t}}{\beta_2 - \gamma_2}$$

yang dapat ditulis ulang:

$$(11.8) \quad P_t = \pi_1 + \pi_2 Y_t + \pi_{1t}$$

Mengganti Persamaan (11.8) ke (11.5) diperoleh

$$(11.7) \quad \begin{aligned} Q &= \gamma_1 + \gamma_2(\pi_1 + \pi_2 Y_t + v_{1t}) + u_{1t} \\ &= \gamma_1 + \gamma_2 \pi_1 + \gamma_2 \pi_2 Y_t + \gamma_2 v_{1t} + u_{1t} \\ &= \pi_3 + \pi_4 Y_t + v_{2t} \end{aligned}$$

Sekarang Persamaan (11.5) dan (11.7) menentukan masing-masing variabel endogen hanya dalam variabel eksogen, parameter model, dan kesalahan stokastik. Kedua persamaan ini dikenal sebagai persamaan *reduced form* dan  $\pi$  dikenal sebagai parameter *reduced form*. Secara umum, persamaan *reduced form* diperoleh dengan menyelesaikan setiap variabel endogen dalam bentuk variabel eksogen, parameter yang tidak diketahui, dan kesalahan stokastik.

### C. IDENTIFIKASI PERSAMAAN SIMULTAN

Identifikasi mengacu pada kemungkinan menghitung parameter struktural model persamaan simultan dari parameter bentuk tereduksi. Ada dua syarat yang diperlukan agar suatu persamaan dapat diidentifikasi yaitu syarat urutan (*order condition*) dan syarat peringkat (*rank condition*).

## Order Condition

- Just atau exact identification: kondisi di mana koefisien suatu persamaan dapat ditentukan secara tepat dari koefisien persamaan *reduced form*. Metode yang digunakan untuk menduga model adalah *Indirect Least Square* (ILS).
- Under identification (tidak dapat diidentifikasi). Kondisi di mana dari persamaan *reduced form* tidak dapat digunakan untuk menduga koefisien model struktural (Contoh di atas fungsi permintaannya). Model tidak dapat diduga.
- Over identification: kondisi di mana dari koefisien persamaan *reduced form* dapat menghasilkan lebih dari satu nilai salah satu koefisien persamaan struktural. Metode yang digunakan untuk menduga model adalah Two Stage Least Square (2SLS atau TSLS).

Untuk melakukan identifikasi dengan order condition dapat digunakan rumus sebagai berikut:

$$(K-M) \geq (G-1)$$

di mana:

G	=	Jumlah persamaan dalam model
K	=	Jumlah variabel dalam model (variabel endogen dan variable <i>predetermined</i> )
M	=	Jumlah variabel dipersamaan yang diidentifikasi

Jika  $(K - M) < (G - 1)$  : Persamaan tidak teridentifikasi (*not identified*)

Jika  $(K - M) = (G - 1)$  : Persamaan teridentifikasi secara tepat (*exactly identified*); dan

Jika  $(K - M) > (G - 1)$  : persamaan teridentifikasi berlebihan (*over identified*)

Melakukan identifikasi di masing-masing persamaan berikut ini.

Persamaan 1: Permintaan  $D_t = \beta_1 + \beta_2 P_t + \beta_3 Y_t + \beta_4 P_t^S + u_{1t}$

Persamaan 2: Penawaran  $S_t = \alpha_1 + \alpha_2 P_t + \alpha_3 W_t + u_{2t}$

Persamaan 3:  $D_t = S_t$

Keseimbangan

Diketahui:

G = 3 persamaan dalam model = Persamaan permintaan, penawaran dan keseimbangan

K = 6 variabel dalam model = Variabel  $D_t$ ,  $P_t$ ,  $Y_t$ ,  $P_t^S$  dan  $W_t$

Identifikasi Persamaan (1)  $D_t = \beta_1 + \beta_2 P_t + \beta_3 Y_t + \beta_4 P_t^S + u_{1t}$

M	=	2 variabel yang diidentifikasi	=	Variabel $M_t$ dan $Y_t$
---	---	--------------------------------	---	--------------------------

$$(K-M) \geq (G-1)$$

(K - M)	vs	(G - 1)	
(6 - 4)	vs	(3 - 1)	
2	=	2	<i>Exactly identified</i>

Identifikasi Persamaan (2)  $S_t = \alpha_1 + \alpha_2 P_t + \alpha_3 W_t + u_{2t}$

M	=	3 variabel yang diidentifikasi	=	Variabel $S_t$ , $P_t$ dan $W_t$
---	---	--------------------------------	---	----------------------------------

$$(K-M) \geq (G-1)$$

(K - M)	vs	(G - 1)	
(6 - 3)	vs	(3 - 1)	
3	>	2	<i>over identified</i>

*Order condition* merupakan syarat perlu, tetapi masih belum menjadi syarat cukup. Artinya, meskipun *order condition* terpenuhi, tidak dapat memastikan bahwa persamaan tersebut teridentifikasi. Karena itu, masih diperlukan *rank condition* (syarat peringkat) untuk menyimpulkannya.

### Rank Condition

Tahapan identifikasi dengan *rank condition* (syarat peringkat) bertujuan menentukan apakah persamaan ini teridentifikasi atau tidak teridentifikasi. Karena itu, diperlukan membuat masing-masing persamaan dan koefisien ke tabel untuk menghitung determinan matriks. *Pertama*, masing-masing persamaan jadikan sebagai fungsi implisit yaitu variabel  $D_t$  dan  $S_t$  seperti berikut ini.

$$\text{Persamaan 1} \quad -D_t + \beta_1 + \beta_2 P_t + \beta_3 Y_t + \beta_4 P_t^S + u_{1t}$$

$$\text{Persamaan 2} \quad -S_t + \alpha_1 + \alpha_2 P_t + \alpha_3 W_t + u_{2t}$$

$$\text{Persamaan 3} \quad -D_t + S_t$$

*Kedua*, membuat tabel yang terdiri dengan kolom setiap variabel dan baris setiap persamaan. Setiap persamaan, masukkan nilai koefisien masing-masing variabel ke dalam kolom, jika tidak, masukkan 0. Bebas Letak variabel di kolom. *Ketiga*, menghapus baris persamaan dan kolom sesuai dengan posisi persamaan dan variabel yang diidentifikasi. Keempat, jika hasil determinan matriks tidak sama dengan nol maka persamaannya teridentifikasi, sebaliknya jika nol, maka tidak akan teridentifikasi.

Identifikasi Persamaan 1:  $-D_t + \beta_1 + \beta_2 P_t + \beta_3 Y_t + \beta_4 P_t^S + u_{1t}$

Coret baris dan kolom variabel persamaan 1. (diberikan warna)

	$D_t$	$S_t$	$P_t$	$P_t^S$	$Y_t$	$W_t$
Persamaan 1	-1	0	$\beta_2$	$\beta_4$	$\beta_3$	0
Persamaan 2	0	-1	$\alpha_2$	0	0	$\alpha_3$
Persamaan 3	-1	1	0	0	0	0

Setelah mencoret baris pertama dan kolom variabel di persamaan 1, tersisa matriks persegi yaitu 2x2 berikut ini:

$$\Delta = \begin{vmatrix} -1 & \alpha_3 \\ 1 & 0 \end{vmatrix} = (-1 \times 0) - (1 \times \alpha_3) = -\alpha_3 \neq 0:$$

Teridentifikasi

Identifikasi Persamaan 2:  $-S_t + \alpha_1 + \alpha_2 P_t + \alpha_3 W_t + u_{2t}$

Coret baris dan kolom variabel persamaan 2. (diberikan warna)

	$D_t$	$S_t$	$P_t$	$P_t^S$	$Y_t$	$W_t$
Persamaan 1	-1	0	$\beta_2$	$\beta_4$	$\beta_3$	0
Persamaan 2	0	-1	$\alpha_2$	0	0	$\alpha_3$
Persamaan 3	-1	1	0	0	0	0

Setelah mencoret baris pertama dan kolom variabel di persamaan 2, tersisa matriks tidak bersegi yaitu 3x2 berikut ini:

$$\begin{vmatrix} -1 & \beta_4 & \beta_3 \\ -1 & 0 & 0 \end{vmatrix}$$

Jadikan menjadi matriks persegi yaitu 2x2 agar diperoleh determinan.

$$\Delta = \begin{vmatrix} -1 & \beta_4 \\ 1 & 0 \end{vmatrix} = (-1 \times 0) - (1 \times \beta_4) = -\beta_4 \neq 0:$$

Teridentifikasi

#### D. METODE KUADRAT TERKECIL TIDAK LANGSUNG

Metode kuadrat terkecil tidak langsung (ILS) adalah metode penghitungan nilai parameter struktural untuk persamaan yang diidentifikasi secara tepat. ILS melibatkan penggunaan OLS untuk memperkirakan persamaan *reduced form* dari sistem dan kemudian menggunakan koefisien estimasi untuk menghitung parameter struktural.

Metode ini hanya dapat digunakan bila persamaan-persamaan dalam model persamaan simultan ditemukan teridentifikasi secara tepat. Prosedur ILS melibatkan tiga langkah berikut:

- Langkah 1: Temukan persamaan *reduced form*.
- Langkah 2: Perkirakan parameter *reduced form* dengan menerapkan OLS pada persamaan *reduced form*.
- Langkah 3: Dapatkan estimasi parameter struktural dari estimasi parameter persamaan bentuk *reduced form* pada langkah 2.

Sebagai latihan penerapan metode ILS diperlukan beberapa data. Menggunakan Tabel 11.1 yang terdiri dari 4 variabel yaitu M = Jumlah uang beredar, Y = Produk domestik bruto, I = Investasi domestik swasta bruto I, dan G = Pengeluaran pemerintah G yang terdiri dari 18 periode waktu.

Latihan penerapan metode ILS maupun TSLS ini dapat menggunakan *Microsoft Excel* dengan mengaktifkan “Data Analysis” agar persamaan regresi dapat digunakan. Setelah dipahami dengan baik, menggunakan Eviews atau lainnya menjadi mudah pemahaman konsep.

Estimasi persamaan bentuk tereduksi dari Contoh 2 adalah

$$\hat{M}_t = 312.0608 + 0.5693\hat{I}_t \quad R^2 = 0.67$$

(2.98)                      (5.65)

$$\hat{Y}_t = 852.3203 + 5.3522I_t \quad R^2 = 0.93$$

(2.17)                      (14.18)

$$\hat{a}_1 = \frac{\hat{\pi}_1}{\hat{\pi}_3} = \frac{0.5693}{5.3522} = 0.1064$$

$$\hat{a}_0 = \hat{\pi}_0 - a_1\hat{\pi}_3 = 312.0608 - 0.1064(852.3203) = 221.3739$$

Jadi persamaan M diestimasi dengan ILS adalah

$$\hat{M}_t = 221.3739 + 0.1064Y_t$$

Persamaan yang sama jika diestimasi dengan OLS akan menghasilkan hasil yang tidak tepat yaitu:

$$\hat{M}_t = 162.7044 + 0.1159\hat{Y}_t \quad R^2 = 0.85$$

(2.13)                      (9.70)

## E. METODE KUADRAT TERKECIL DUA TAHAP

Kuadrat terkecil dua tahap (TSLS) adalah metode memperkirakan parameter struktural yang konsisten. Persamaan yang terverifikasi, misalnya persamaan yang teridentifikasi secara tepat, penerapan metode 2SLS memberikan hasil yang sama dengan metode ILS.

Metode 2SLS melibatkan regresi setiap variabel endogen pada semua variabel eksogen dan kemudian menggunakan nilai prediksi variabel endogen untuk memperkirakan persamaan struktural model. Ide dasar di balik metode TSLS adalah mengganti regressor endogen stokastik (yang berkorelasi dengan faktor gangguan (*error term*) yang menyebabkan bias) dengan



$\hat{a}_1 = 0.1153$  adalah perkiraan yang konsisten  $a_1$

Periode	1	2	3	4	5	6	7	8	9
M	474.30	520.79	551.20	619.28	724.20	749.61	786.25	792.49	824.41
Y	3,315.60	3,688.80	4,033.50	4,319.30	4,537.50	4,891.60	5,258.30	5,588.00	5,847.30
I	483.50	639.50	743.60	762.30	737.10	831.60	842.00	866.70	812.80
G	710.10	742.70	829.00	905.10	963.20	1,019.30	1,060.70	1,123.90	1,213.10
Periode	10	11	12	13	14	15	16	17	18
M	896.34	1,024.31	1,129.69	1,150.08	1,126.80	1,081.06	1,073.94	1,097.37	1,122.96
Y	6,080.70	6,469.80	6,795.50	7,217.70	7,529.30	7,981.40	8,478.60	8,974.90	9,559.70
I	832.10	909.80	995.80	1,146.10	1,155.60	1,284.30	1,434.50	1,590.80	1,723.70
G	1,239.50	1,281.80	1,307.10	1,344.00	1,374.50	1,438.90	1,508.20	1,567.20	1,688.80

## F. RANGKUMAN

1. Kuadrat terkecil tidak langsung (ILS) adalah metode penghitungan nilai parameter struktural yang konsisten untuk persamaan yang teridentifikasi secara tepat dalam suatu sistem persamaan simultan.
2. Metode ILS melibatkan penggunaan OLS untuk memperkirakan persamaan bentuk tereduksi dari sistem dan kemudian menggunakan estimasi parameter bentuk tereduksi untuk menghitung estimasi parameter struktural keuangan konsolidasi menunjukkan besarnya dan ruang lingkup operasi perusahaan sependali.
3. Metode kuadrat terkecil dua tahap (TSLs) adalah metode memperkirakan nilai parameter struktural yang konsisten untuk persamaan *exactly identification* atau *over identification* dalam sistem persamaan simultan. Untuk persamaan yang teridentifikasi secara tepat, TSLs memberikan hasil yang sama seperti ILS.

## G. TES FORMATIF

1. Pada sistem persamaan simultan, metode estimasi OLS tidak dapat digunakan karena akan menghasilkan estimator yang tidak konsisten. Hal ini dikarenakan:
  - a. satu atau lebih variabel penjelas berkorelasi dengan *error term*
  - b. satu atau lebih variabel penjelas tidak berkorelasi *error term*
  - c. Satu atau lebih variabel penjelas bersifat stokastik
  - d. *Error term* hanya dipengaruhi oleh variabel penjelas untuk setiap persamaan dalam sistem
  
2. Pernyataan yang benar mengenai tahap endogenitas adalah:
  - a. Pendugaan persamaan *reduced form* terhadap variabel-variabel yang diduga bersifat endogen dilakukan untuk memperoleh nilai *error term* yang digunakan dalam pengujian endogenitas.
  - b. Pendugaan persamaan *reduced form* terhadap variabel-variabel yang diduga bersifat endogen dilakukan untuk memperoleh nilai *error term* yang digunakan dalam pengujian eksogenitas.
  - c. *Error term* yang diperoleh dari hasil estimasi *reduced form* diestimasi bersama persamaan struktural dengan estimasi model 2SLS
  - d. *Error term* yang diperoleh dari hasil estimasi *reduced form* diestimasi persamaan strukturalnya dengan estimasi model OLS
  
3. Persamaan berikut merupakan model persamaan simultan:

$$K_1 = \alpha_1 K_2 + \beta_1 z_1 + u_1$$

$$K_2 = \alpha_2 K_1 + \beta_2 z_2 + u_2$$

Estimasi OLS akan mengalami bias simultanitas jika:

- a.  $u_1$  berkorelasi dengan  $z_1$ .

- b.  $z_1$  berkorelasi dengan  $z_2$ .
  - c.  $K_2$  berkorelasi dengan  $u_1$ .
  - d.  $K_1$  berkorelasi dengan  $u_1$
4. Manakah dari berikut ini yang dengan *exactly identification* karakteristik persamaan struktural?
- a. Persamaan struktural harus mengandung jumlah variabel terikat dan bebas yang sama.
  - b. Persamaan struktural harus mengandung jumlah variabel endogen dan eksogen yang sama.
  - c. Persamaan struktural harus mempunyai interpretasi perilaku, ceteris paribus tersendiri.
  - d. Persamaan struktural tidak boleh mengandung kesalahan struktural.

## H. LATIHAN

1. Pertimbangan model berikut ini.

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + u_{1t}$$

$$I_t = \beta_0 + \beta_1 R_t + \beta_2 I_{t-1} + u_{2t}$$

$$R_t = \gamma_0 + \gamma_1 Y_t + \gamma_2 M_t + u_{3t}$$

$$Y_t = C_t + I_t + G_t$$

Di mana  $C_t$  = menunjukkan konsumsi riil,  $Y_t$  = Pendapatan nasional riil,  $I_t$  = Investasi riil,  $M_t$  = Jumlah uang beredar riil,  $R_t$  = Tingkat bunga, dan  $G_t$  = Pengeluaran pemerintah riil.

### Pertanyaan:

- a. Lakukan identifikasi model dengan menerapkan *order condition* dan *rank condition*.
- b. Jelaskan saran Anda mengestimasi model, apakah ILS atau TSLS.

c. Tentukan persamaan *reduced form* dari model persamaan struktural.

2. Perhatikan model dua persamaan berikut:

$$Y_{1t} = A_1 + A_2 Y_{2t} + A_3 X_{1t} + u_{1t}$$

$$Y_{2t} = B_1 + B_2 Y_{1t} + B_3 X_{2t} + u_{2t}$$

di mana Y adalah variabel endogen, X adalah variabel eksogen, dan u adalah suku kesalahan stokastik.

Pertanyaan:

- a. Dapatkan regresi *reduced form*.
- b. Tentukan persamaan mana yang teridentifikasi.
- c. Untuk persamaan yang teridentifikasi, metode estimasi manakah yang akan Anda gunakan dan mengapa?
- d. Misalkan, secara apriori, diketahui  $A_3 = 0$ . Bagaimana jawaban Anda terhadap pertanyaan sebelumnya berubah? Mengapa?

## DAFTAR PUSTAKA

- Agresti, Alan dan Christine Franklin. 2013. *Statistics The Art and Science of Learning from Data Third Edition*. Pearson
- Ananta, Aris. 1987. *Landasan Ekonometrika*. Jakarta: Penerbit PT Gramedia
- Anderson, David R, dkk. 2014. *Statistics for Business and Economics 12e*. South-Western Cengage Learning
- Asteriou, D., & Hall, S. G. (2021). *Applied econometrics*. Bloomsbury Publishing.
- Damodar N. Gujarati, Dawn C. Porter (2009). *Basic econometrics*, 5th ed. New York: McGraw-Hill.
- Damodar N. Gujarati, Dawn C. Porter (2012). *Basic econometrics*, 5th ed. New York: McGraw-Hill.
- Dyanasari, Asnah, Siswadi. 2018. *Pengantar Penerapan Ekonometrika*. Yogyakarta, Deepublish.
- Ghozali, Imam. 2021. *Aplikasi Analisis Multivariate dengan Program IBM SPSS 26*. 10th ed. Semarang: Badan Penerbit Universitas Dipenogoro.
- Gio, Ugiana, Prana. Irawan, Erwin, Dasapta. 2016. *Belajar Statistika dengan R*. Medan : USU Press.
- Gujarati, D. N. (2021). *Essentials of econometrics*. Sage Publications.
- Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics*. McGraw-hill.

- Gujarati, D. N., & Porter, D. C. (2019). Basic Econometrics. McGraw-Hill Education.
- Gujarati, Damodar N. & Porter, Dawn C. 2009. Basic Econometrics. Fifth Edition. McGraw-Hill Companies.
- Gujarati, Damodar N. 2003. Basic Econometrics. 4th ed. US: McGraw Hill.
- Gujarati. Damodar N, Porter. Dawn C. 2010. Essentials of Econometrics 4th Edition. United States; McGraw-Hill Irwin
- Gunawan, Ce. 2020. Mahir Menguasai SPSS Panduan Praktis Mengolah Data Penelitian. Yogyakarta: Deepublish.
- Hanke, J.E., and Wichern, D.W. 2005. Business Forecasting. Prentice Hall, New York
- Hasan, Arifin, Johan. 2008. Statistik Bisnis Terapan. Jakarta : PT. Alex Media Komputindo.
- Hill, R Carter, William E G dan Guay C L. 2011. Principles of Econometrics Fourth Edition. Wiley
- Illowsky, Barbara. Dean, Susan. 2018. Introductory Statistics. Openstax. Rice University
- Irawan, Nata. 2016. Statistika Ekonomi dan Bisnis. Denpasar. Universitas Udayana.
- Kleiber, Christian. Zeileis, Achim. 2008. Applied Econometrics With R. USA : Springer
- Kurniawan, Robert & Yuniarto, Budi. 2016. Analisis Regresi: Dasar dan Penerapannya dengan R. 1st ed. Jakarta: Kencana.
- Lane, M. David. Introduction to Statistic. Rice University.

- Lind, Douglas, William Marchal, Samuel Wathen (2021). *Statistical Techniques in Business and Economics* (18th Edition). New York: McGraw-Hill
- Maddala, G.S (1992). *Introduction to Econometric*, 2nd Edition, MacMillan Publishing Company, New York. Sumodiningrat, Gunawan. 2001. *Ekonometrika Pengantar*. Yogyakarta: PFEYogyakarta
- Maddala, GS. Lahiri, Kjal. 2009. *Introduction to Econometrics* 4th Edition. New Delhi : John Wiley & Sons Inc.
- Mendenhall, William, Robert J B, dan Barbara M B. 2009. *Introduction to Probability and Statistics*. Brooks/Cole
- Nachrowi, Djalal & Usman, Hardius. 2020. *Pendekatan Populer dan Praktis Ekonometrika untuk Analisis Ekonomi dan Keuangan*. Jakarta: Lembaga Penerbit Fakultas EKonomi Universitas Indonesia.
- Nahrowi, Nahrowi D. & Usman, Hardius. 2006. *Pendekatan Populer dan Praktis Ekonometrika Untuk Analisis Ekonomi dan Keuangan*. Jakarta: Lembaga Penerbit Fakultas Ekonomi Universitas Indonesia.
- Salvatore, D., & Reagle, D. P. (2002). *Schaum's Outline off Theory and Problems off Statistics and Econometrics*. McGraw-Hill.
- Stock, J. H., & Watson, M. W. (2019). *Introduction to Econometrics*. Pearson.
- Stock, J. H., Watson, M. W., Wooldridge, J. M., & Wooldridge, J. M. *Introductory Econometrics: A Modern Approach* (4 th International).

- Subagyo, Pangestu. (2003). Statistik Deskriptif. Yogyakarta : BFFE Yogyakarta
- Sumodiningrat, Gunawan. 2001. Ekonometrika Pengantar. Yogyakarta: PFEYogyakarta
- Suparmi, Christina. 2014. Statistika Ekonomi. Tangerang: Universitas Terbuka.
- Titin Agustin Nengsih, Bella Arisha, Yuliana Safitri. (2022). Statistika Deskriptif dengan Program R. SONPEDIA, Jambi.
- Titin Agustin Nengsih, Fadhlul Mubarak, Vinny Yuliani Sundara. (2020). Pemrograman R Dasar. Forum Pemuda Aswaja, NTB.
- Tri Basuki, Agus & Prawoto, Nano. 2016. Analisis Regresi dalam Penelitian Ekonomi & Bisnis: Dilengkapi Aplikasi SPSS & Eviews. Jakarta: Rajawali Pers.
- Walpole, Ronald E, dkk. 2012. Probability and Statistics for Engineers and Scientist Ninth Edition. Prentice Hall
- Wibisono, Yusuf. 2009. Metode Statistik. Yogyakarta : Gajah Mada University Press.
- Widarjono, Agus. 2018. Ekonometrika.Tangerang Selatan: Universitas Terbuka
- Wooldridge, J. M. (2015). Introductory Econometrics: A Modern Approach. Cengage Learning.
- Wooldridge, J. M. (2016). Introductory econometrics: A modern approach (6th ed.). Boston: Cengage Learning
- Wooldridge, J. M. (2016). Introductory econometrics: A modern approach (6th ed.). Boston: Cengage Learning

- Wooldridge, J. M., Wadud, M., & Lye, J. (2016). *Introductory Econometrics: Asia Pacific Edition With Online Study Tools 12 Months*. Cengage AU.
- Wooldridge, J.M. 2013. *Introductory Econometrics: A Modern Approach 5th Edition*. Mason: South-Western, Cengage Learning
- Wooldridge, Jeffrey M. 2020. *Introductory Econometrics: A Modern Approach*. Amerika: Penerbit. Cengage.
- Yudiatmaja, Fridayana. 2013. *Analisis Regresi dengan Menggunakan Aplikasi Komputer Statistik SPSS*. Jakarta: Kompas Gramedia.

## TENTANG PENULIS



**Dr. Erric Wijaya, SE.,ME**, merupakan dosen tetap Program Studi S1 Manajemen STIE Indonesia Banking School Jakarta. Lahir di Palembang, 8 Desember 1975. Penulis merupakan anak pertama dari empat bersaudara dari pasangan Bapak Husaini dan Ibu Alm Hikmah. Menyelesaikan studi S1 di Universitas Sriwijaya tahun 1998 pada Program Studi Ilmu Ekonomi dan Studi Pembangunan. Selanjutnya gelar Magister Ekonomi diraih dari Universitas Indonesia pada tahun 2003 pada Program Ilmu Ekonomi. Gelar Doktor pada Program Studi Ilmu Ekonomi diraih pada tahun 2015 di Universitas Padjadjaran. Penulis pernah memiliki pengalaman sebagai tim peneliti di Lembaga Penyelidikan Ekonomi Masyarakat (LPEM) Universitas Indonesia, Tim Peneliti Insititute for Development of Economics and Finance (INDEF), dan Tim Peneliti PRANATA Universitas Indonesia. Berbagai tulisan jurnal ilmiah serta prosiding tingkat nasional dan internasional telah dihasilkan oleh penulis.



### **Dr. Degdo Suprayitno**

Saat ini menjabat sebagai Ketua Program Studi Manajemen Logistik, Institut Ilmu Sosial dan Manajemen STIAMI, Jakarta. Berpengalaman lebih dari 20 tahun dalam bidang: *Ship Management, Sea Transport, Logistics & Shipping*. Aktif di berbagai organisasi seperti INSA (*Indonesia Shipowner Association*), PRAMARIN (*Indonesian Maritime Practitioner*), IARSI (*Indonesian*

*Supply Chain Expert Association*), ALI (*Indonesian Logistics Association*) dan ISLI (*Indonesian Institute of Supply Logistics and Supply Chain*) Meraih berbagai sertifikasi dan pelatihan dari dalam dan luar negeri. Assessor BNSP No. MET.000.000067 Warehousing Systems, Internal Auditor *Quality Management System ISO 9001:2008; ISO 9001:2015; Management System & Occupational Safety ISO 14001:2015 and OHSAS 18001:2007*.



**Urwawuska Ladini, S.Stat., M.Si.**

Seorang penulis dan dosen tetap Prodi Statistika Fakultas Sains dan Teknologi Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi. Lahir di Kab.Semarang, 20 Mei 1994 Jawa Tengah. Penulis merupakan anak kedua dari tiga bersaudara dari pasangan Bapak Mudzakkir Ahmad dan Ibu Joesinta Ramprihatini. Pendidikan program Sarjana (S1) Universitas Islam Indonesia pada Prodi Statistika dan menyelesaikan program Pascasarjana (S2) di Institut Pertanian Bogor pada Prodi Statistika Terapan. Karya Ilmiah yang telah ditulis dan terbit diantaranya berjudul: “*Generalized Structured Component Analysis* pada Hubungan antara Pengelolaan Lingkungan, Faktor Sosial, dan Kualitas Lingkungan”; “*The Impact of Islamic Financial Development, GDP, and Population on Environmental Quality in Indonesia*”; “*The Influences of Good Corporate Governance and Profitability on Earnings Management at Commercial Banks by Panel Data Regression Analysis*”; “*Modeling of Quality of Education in Junior High School using Multivariate Adaptive Regression Splines (MARS) Method*”.



**Titin Agustin Nengsih, S.Si., M.Si., Ph.D.**

Penulis dengan latar belakang pendidikan Statistika yang lulus Sarjana (S1) dan Magister (S2) di bidang Statistika pada tahun 2000 dan 2010 dari Institut Pertanian Bogor (IPB). Pendidikan doktoral (S3) di Perancis pada tahun 2020 dalam bidang penelitian adalah modelling dan missing value. Penulis saat ini adalah dosen Prodi Statistika di Fakultas Sains dan Teknologi Universitas Islam Negeri Sulthan Thaha Saifuddin Jambi. Saat ini peneliti menjadi bagian dari Unit Pelaksana Teknis di UIN Sulthan Thaha Saifuddin Jambi yaitu Kepala Unit Teknologi Informasi dan Pangkalan Data (Ka. UTIPD). Peneliti juga sedang melakukan penelitian di bidang Arkeologi dalam bidang penelitian modelling, visualisasi, dan penanganan data hilang dengan pendanaan riset dari BRIN yang di mulai tahun 2023 hingga 2026 (multi years) di pulau Kalimantan dan pulau Sumatera.



**Dr. Sumiyarti, ME.**

Seorang dosen tetap Prodi Ekonomi Pembangunan Fakultas Ekonomi Universitas Trisakti Jakarta. Lahir di Bantul, 22 Januari 1969. Pendidikan program Sarjana (S1) diselesaikan dari Prodi Ilmu Ekonomi Dan Studi Pembangunan Fakultas Ekonomi Universitas Gadjah Mada, kemudian melanjutkan studi ke Program Pasca Sarjana Ilmu Ekonomi Fakultas Ekonomi Universitas Indonesia. Terakhir menamatkan Program Doktor Ilmu Ekonomi Konsentrasi Kebijakan Publik di Fakultas Ekonomi dan Bisnis Universitas Trisakti, Jakarta. Beberapa artikel telah diterbitkan dengan bidang kajian ekonomi makro, pembangunan ekonomi regional, perdagangan internasional.

Buku yang diterbitkan diantaranya adalah Kinerja Perdagangan Luar Negeri Indonesia: *Suatu Kajian Empiris, Metode Kuantitatif untuk Ekonomi dan Bisnis*.



**Sella Nofriska Sudrimo, M.Si.**

Seorang penulis dan dosen tetap Program Studi Pendidikan Agama Islam di IAIN Sorong. Lahir di Bandar Lampung, 26 November 1994. Penulis merupakan anak ke-dua dari dua bersaudara dari pasangan bapak Sudrimo dan Ibu Kamila. Pendidikan program Sarjana (S1) Universitas Lampung Fakultas MIPA Program Studi Matematika dan menyelesaikan program Pascasarjana (S2) di Institut Pertanian Bogor Fakultas MIPA Departemen Statistika

Terapan. Buku yang telah ditulis dan terbit berjudul di antaranya: *Statistika Bisnis* dan *Matematika Ekonomi*.



**Sri Yani Kusumastuti**

Dosen tetap di Program Studi Ekonomi Pembangunan Fakultas Ekonomi dan Bisnis Universitas Trisakti Jakarta. Lulus dari Program Sarjana (S1) Program Studi Ilmu Ekonomi dan Sudi Pembangunan Universitas Gadjah Mada dan menyelesaikan program Pasca Sarjana (S2) di Program Magister Sains Ilmu Ekonomi Universitas Gadjah Mada. Pengalaman mengajar Mikroekonomika, Statistika, dan Ekonometrika. Banyak menulis

artikel di bidang Ekonomi, Keuangan, dan Perbankan, Menjadi Editor di beberapa buku, antara lain: Kinerja perdagangan luar negeri Indonesia pada masa krisis: suatu kajian empiris (2002), Desain eksperimental: aplikasi dalam riset ilmu ekonomi,

manajemen dan akuntansi (2011). Penulis buku: Lembaga Jasa Keuangan di Indonesia (2018).



### **Nurhayati**

Merupakan dosen tetap di Program Studi Ekonomi Pembangunan Fakultas Ekonomi dan Bisnis Universitas Trisakti Jakarta. Lulus dari Program Sarjana (S1) Program Studi Ilmu Ekonomi dan Sudi Pembangunan Universitas Trisakti dan menyelesaikan program Pasca Sarjana (S2) di Program Magister Perencanaan dan Kebijakan Publik Universitas Indonesia dan melanjutkan Pasca Sarjana (S3) di Program Kebijakan Publik Universitas Trisakti. Pengalaman mengajar Statistika, Ekonometrika dan Praktikum Alat Analisa Kuantitatif. Banyak menulis artikel di bidang Ekonomi, Regional, dan Pembangunan Berkelanjutan. Penulis aktif sebagai pengurus Jurnal sebagai Managing Editor pada Jurnal Media Ekonomi. Penulis juga aktif sebagai Ketua Lembaga Pengolahan Data dan Statistik di Fakultas Ekonomi dan Bisnis Universitas Trisakti.



### **Dalizanolu Hulu, S.E, M.E, CRM, CRP**

Memperoleh S1 Ilmu Ekonomi dan Studi pembangunan dari Universitas HKBP Nommensen, Medan. Selanjutnya S2 Ilmu Ekonomi dari Universitas Indonesia, Jakarta. Kompeten di bidang Manajemen Risiko di peroleh CRM (*Certified Risk Management*) dari *Quint Wellington Redwood*, Australia. Selanjutnya CRP (*Certified Risk Professional*) dari BNSP, LSP-PM (Lembaga Sertifikasi Profesi Pasar Modal), Jakarta. Sekarang

Dosen Tetap di Prodi Manajemen, Universitas Pembangunan Jaya (UPJ), Kota Tangerang Selatan. Mengampu mata kuliah Ekonomi Mikro, Ekonomi Makro, Ekonometrika, Matematika Bisnis, Statistik Bisnis, Manajemen Investasi, Lembaga Keuangan dan Pasar Modal, Manajemen Risiko, Manajemen Keuangan, dan Manajemen Operasi.

**Penerbit :**

PT. Sonpedia Publishing Indonesia

Buku Gudang Ilmu, Membaca Solusi  
Kebodohan, Menulis Cara Terbaik  
Mengikat Ilmu. Everyday New Books

**SONPEDIA.COM**  
**PT. Sonpedia Publishing Indonesia**

**Redaksi :**

Jl. Kenali Jaya No 166

Kota Jambi 36129

Tel +6282177858344

Email: [sonpediapublishing@gmail.com](mailto:sonpediapublishing@gmail.com)

Website: [www.buku.sonpedia.com](http://www.buku.sonpedia.com)