

APLIKASI PENELITIAN KEUANGAN
DAN EKONOMI SYARIAH
DENGAN STATA



Oleh:
Dr. Faisol, S.Pd., MM
Dr. Agus Eko Sujianto, SE., MM

PENERBIT CAHAYA ABADI
TULUNGAGUNG
2020

APLIKASI PENELITIAN KEUANGAN DAN EKONOMI SYARIAH DENGAN STATA

Oleh:

Dr. Faisol, S.Pd., MM dan Dr. Agus Eko Sujianto, SE., MM

Hak Cipta Dilindungi dengan Undang-Undang, dilarang keras mengutip, menjiplak, memfotokopi sbagian atau seluruh isi buku ini serta memperjualbelikannya tanpa izin tertulis dari Penerbit CAHAYA ABADI.

Diterbitkan oleh:

Penerbit CAHAYA ABADI

Email: agusekosujianto@gmail.com

Telp. 085234635471

Editor: Khusnul Mufidati

Dr. Faisol, S.Pd., MM dan Dr. Agus Eko Sujianto, SE., MM

Aplikasi Penelitian Keuangan dan Ekonomi Syariah dengan STATA; Dr. Faisol, S.Pd., MM dan Dr. Agus Eko Sujianto, SE., MM : editor, Khusnul Mufidati – Tulungagung : Cahaya Abadi, 2020.
111 hlm. ; 24 cm

ISBN: 978-602-8569-77-4

Anggota Ikatan Penerbit Indonesia No. 121/JTI/2010

KATA PENGANTAR

Pertama-tama penulis mengucapkan puji syukur kepada Allah SWT yang telah memberikan kemampuan pikir dan pencerahan untuk berbuat lebih baik daripada hari kemarin, dan dengan penuh harapan juga akan lebih baik pada hari yang akan datang, dimana penulis telah membuat dan menyusun buku ini, sehingga buku ini dapat tersaji dengan segala kekurangan dan keterbatasan kami sebagai penulis.

Dengan semakin kompleks dan berkembangnya data-data penelitian ekonomi, dan keuangan, penguasaan perangkat lunak pengolahan dan analisis data adalah sangat relevan dan dibutuhkan untuk membentuk mahasiswa memiliki kemampuan penelitian yang memadai. Dalam hal ini STATA memberikan banyak varian dan menyediakan berbagai fitur untuk analisis data, manajemen data, dan grafik. STATA memberikan kita berbagai fasilitas alat analisis yang dibutuhkan untuk mengatur dan mengelola data kita dan juga memperoleh serta menghasilkan hasil analisis statistik. STATA memiliki beberapa dokumentasi dasar (seperti User's Guide) yang dapat digunakan oleh pembaca untuk memahami informasi yang lebih detail terkait setiap bahasa perintah yang digunakan di dalam program

Buku ini ditujukan untuk memperkenalkan mahasiswa kepada pengolahan dan penganalisisan data penelitian yang ada. Diharapkan melalui buku ini, mahasiswa dapat menggunakan perangkat lunak untuk membantu melakukan analisis penelitian kuantitatif di bidang ekonomi dan keuangan.

Topik yang dibahas dan dijelaskan dalam buku ini hanya difokuskan pada pembelajaran STATA tingkat dasar, yakni pengantar singkat STATA kemudian dilanjutkan dengan bagaimana mengoperasikan STATA. Kemudian, dijelaskan beberapa perintah dalam analisis deskriptif dengan tabel dan grafik. Di Bab 4 dijelaskan analisis regresi data panel di lengkapi dengan tahapan pengujian pemilihan model yakni common effect, fixed effect dan random effect, selanjutnya konsep asumsi klasik adalah bagian penting dalam model regresi, diuraikan secara sederhana dan spesifik. Di Bab 4 memberikan contoh kasus tahapan analisis time series yakni VECM, Uji Stasioner, uji lag optimal, uji kointegrasi, uji model VEC, dan dilengkapi dengan IRF dan FEVD. Di bagian akhir, penulis menjelaskan secara singkat analisis data primer dengan SEM, yakni uji validitas dan uji reliabilitas

Kami menyadari bahwa masih terdapat banyak kekurangan dalam modul ini, oleh karena itu kami sangat mengharapkan saran dan kritik konstruktif untuk perbaikan modul ini di kemudian hari.

Penulis mengucapkan banyak terima kasih kepada banyak pihak yang berkontribusi memberikan masukan buku ini, baik dari segi motivasi maupun

dari segi materi. Juga kepada penerbit buku ini, penulis mengucapkan terima kasih sehingga buku ini dapat sampai ke tangan para pembaca. Buku ini dilengkapi dengan CD yang berisi data dalam file excel yang dapat digunakan dalam analisis, workfile program STATA. Harapan penulis buku ini akan melengkapi buku-buku statistik yang telah terbit sebelumnya. Tentunya buku ini masih banyak kekurangannya, untuk itu penulis menunggu masukan dalam rangka perbaikan lebih lanjut.

Kediri, 21 Desember 2020
Penulis

DAFTAR ISI

BAB	I	Pengenalan Tentang STATA.....	1
1.1		Pengantar singkat STATA	1
1.2		Menginstall STATA	3
1.3		Tampilan STATA	4
1.4		Perintah Dasar untuk Penggunaan Awal	7
1.5		Data dan Variabel 11	
BAB	II	ANALISIS DESKRIPTIF	17
2.1.		Analisi Deskriptif dengan STATA.....	17
2.2.		Menggunakan Tabel dan Grafik Statistik Deskriptif pada STATA	21
BAB	III	ANALISIS REGRESI.....	25
3.1.		Analisis Regresi Linier	25
3.2.		Hipotesis	26
3.3.		Analisis Regresi Sederhana	27
3.4.		Pendeteksian dan Pengoreksian Pelanggaran Asumsi	31
BAB	IV	ANALISIS REGRESI DENGAN DATA PANEL	45
4.1.		Analisis Menggunakan Data Panel.....	45
4.2.		Analisis Regresi Data Panel.....	51
		4.2.1. Menentukan Model Estimasi Data Panel	55
		4.2.2. Uji Asumsi Klasik	60
BAB	V	Vector Error Correction Model (VECM).....	63
5.1.		Tahapan Model Vector Error Correctin	63
5.2.		Uji Stasionaritas	65
5.3.		Uji Lag Optimal	68
5.4.		Uji Kointegrasi	71
5.5.		Uji Model Vector Error Correction	72
5.6.		Analisis IRF dan FEVD	76
		5.6.1 Impulse Response Function	76
		5.6.2 Forcasting Error variance Decomposition	78
BAB VI		REGRESI LOGIT dan PROBIT	81
6.1.		Regrei Logit	81
6.2.		Regresi Probit	82
6.3.		Marginal Effect	82

6.4.	Menjalankan Regresi Logit dengan STATA	83
6.5.	Menjalan Regresi Probit dengan STATA	84
6.6.	Tahapan melakukan regresi Logit dan Probit	85
BAB VII ANALISIS KUESIONER MENGGUNAKAN STATA		99
7.1.	Analisis Data Primer dengan STATA	99
7.2.	Uji Validitas	102
7.3.	Uji Reliabilitas	106
Daftar Referensi		109
BIODATA PENULIS		111

BAB I

PENGENALAN TENTANG STATA

1.1 Pengantar Singkat STATA



(sumber: www.stata.com)

STATA adalah salah satu program statistik yang powerful yang dikembangkan oleh StataCorp dan banyak digunakan di dalam analisis bidang ilmu ekonomi dan keuangan. Dengan STATA peneliti dapat dengan mudah menganalisis data kuantitatif baik itu menggunakan data cross-section, panel atau time series. STATA menyediakan berbagai fitur untuk analisis data, manajemen data, dan grafik. STATA memberikan kita berbagai fasilitas alat analisis yang dibutuhkan untuk mengatur dan mengelola data kita dan juga memperoleh serta menghasilkan hasil analisis statistik seperti

1. ANOVA/MANOVA
2. Model linear
3. Metode Survey
4. Model generalised linear
5. Logistic regresi
6. Probit regresi
7. Regresi linear
8. Analisis cluster
9. Data Panel/Longitudinal
10. Analisis Time Series (seperti ARIMA, ARMAX, ARCH, VAR, VECM),
11. Analisis Model Persamaan Struktural (Structural Equation Modeling).

Bagi banyak pengguna, STATA merupakan program statistik yang menyediakan fitur-fitur yang dapat digunakan untuk membaca data, membuat variabel baru, menghitung analisis statistik dan menggambar grafik. Bagi pengguna lain, STATA merupakan program dengan bahasa perintah yang umumnya digunakan untuk mengeksekusi data untuk menjalankan langkah-langkah analisis tanpa intervensi. Beberapa menganggap STATA sebagai bahasa pemrograman untuk mengembangkan data atau bahasa perintah STATA dari penambahan data manajemen, statistisk atau kemampuan grafik.

Dalam bukunya Henky Latan (2014) menguraikan terdapat beberapa jenis program STATA yang dapat digunakan oleh pembaca antara lain Small STATA (SmallStata) yang memiliki beberapa keterbatasan dalam jumlah variabel dan data observasi sehingga cocok untuk pemula, STATA special edition (STATASE) yang digunakan untuk menghandle data set yang besar dan memiliki kemampuan yang sama dengan StataMP, STATA IC (StataIC) adalah standar STATA, STATA Multiple Processor (StataMP) yang dapat digunakan di dalam parallel untuk sistem operasi Windows, Mac, Linux dan Solaris yang merupakan jenis STATA yang paling cepat.

Program STATA memiliki beberapa dokumentasi dasar (seperti User's Guide) yang dapat digunakan oleh pembaca untuk memahami informasi yang lebih detail terkait setiap bahasa perintah yang digunakan di dalam program. Selain itu, berbagai fitur lainnya yang dapat dilihat secara lengkap di alamat website <http://www.stata.com/features/>, <https://stats.idre.ucla.edu/>, yang menyediakan informasi mengenai buku-buku tentang STATA yang sudah terbit diseluruh dunia.

1.2 Menginstall STATA

Sebelum kita menginstall program STATA, dalam Henky Latan (2014) menjelaskan hal hal-hal yang harus kita perhatikan adalah

1. Pilih jenis program STATA yang akan di install, apakah SmallStata, StataSE, StataIC ataukah StataMP.
2. Gunakan license key yang sesuai dengan jenis program STATA yang akan diinstall (Catatan: jika kita ingin mengganti jenis STATA yang digunakan pada laptop, kita memilih modify pada saat instalasi.
3. Tentukan dimana kita akan menginstall program STATA pada hard disk drive, dalam hal ini direkomendasikan pada
C:\Program Files\Stata14 on 32-bit Windows dan
C://Program Files (x86)\Stata14 on 64-bit Windows
4. Jika kita sudah memiliki program STATA versi sebelumnya (*old version*) pada sistem komputer/laptop, kita dapat langsung melakukan update ke STATA versi terbaru (catatan: tidak disarankan untuk menggunakan dua versi STATA pada satu sistem komputer).

Langkah langkah instalasi program STATA dengan menggunakan sistem operasi Windows, sebagai berikut: (Sumber: Henky Latan, 2014)

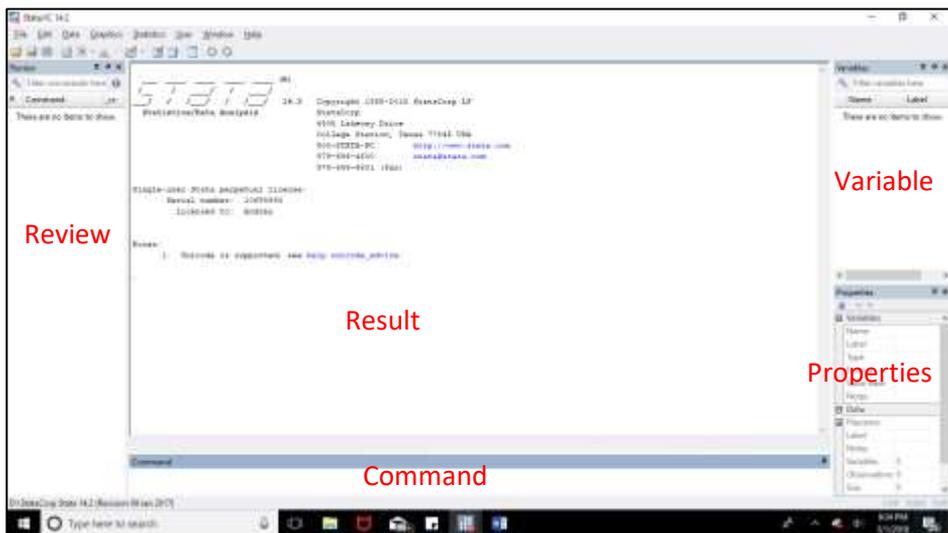
1. Double klik installer program STATA
2. Klik accept ***software license agreement*** untuk memulainya proses instalasi
3. Pilih jenis STATA yang ingin diinstall sesuai dengan license key
4. Jika kita menggunakan komputer 64 bit dengan 64 bit Microsoft Windows, kita harus menginstall STATA versi 64 bit. Sebaliknya, jika kita memiliki komputer 32 bit, kita bisa menginstall STATA versi 32 bit.
5. Proses instalasi akan menanyakan dimana program STATA akan disimpan, dalam hal ini kita memilih default yaitu seperti yang sudah disarankan di atas.

6. Proses instalasi juga akan menanyakan dimana direktori pekerjaan akan disimpan. Direktori pekerjaan yang dimaksud adalah lokasi untuk data set, grafik dan file-file STATA lainnya. Dalam hal ini disarankan untuk memilih ***Use Each User's Documents Folder***
7. Klik ***next*** kemudian klik ***install*** dan tunggu sampai proses instalasi selesai kemudian klik ***finish***
8. Setelah selesai instalasi, untuk menjalankan program STATA terlebih dahulu harus memasukkan license key program (catatan: license key yang dimasukkan harus sesuai dengan jenis STATA yang diinstall) dengan memilih menu Start > All Program? Stata kemudian double klik program STATA yang diinstall dan isikan serial number license key lalu klik OK
9. Buat shortcut program dengan mengcopy paste STATA ikon ke layar desktop untuk memudahkan kita mengakses program STATA
10. Jika terjadi masalah dalam instalasi program STATA, kita dapat mencari informasi faktor penyebabnya di website <http://www.stata.com/support/installation-qualification/>.

4.1 Tampilan STATA

Pada bagian ini akan dijelaskan pengenalan singkat mengenai tampilan utama STATA. Dalam kesempatan ini, versi STATA yang digunakan adalah STATA versi 14. Pada bagian atas panel utama terdapat menu bar (mulai dari *File* hingga *Help*) yang berisikan seluruh perintah detil yang dapat dilakukan oleh STATA. Menu *File* dan *Edit* memiliki fungsi yang secara umum sama dengan perangkat lunak untuk OS Windows yakni terkait manipulasi file-file secara umum (termasuk *save*, *print*, *copy*, *paste*, dan sebagainya). Menu spesifik yang ada pada STATA yang perlu mendapat perhatian adalah *data*, *graphic*, dan *statistics*.

Menu **data** digunakan untuk melakukan berbagai tindakan/manipulasi terkait data yang digunakan termasuk didalamnya untuk pengaturan jenis data, pembuatan variabel-variabel, dan pembuatan matriks. Selanjutnya, menu **graphics** digunakan untuk perintah-perintah terakhir produksi grafik, diagram, data plot, dan sejenisnya. Sementara menu **statistics** merupakan menu utama yang dimiliki STATA yang berisikan perintah-perintah pengolahan dan analisis data termasuk analisis deskriptif, regresi, ANOVA, SEM, dan berbagai perangkat analisis lain yang dimiliki STATA sesuai versinya.



Gambar 1.1. Tampilan Utama STATA

STATA memiliki setidaknya 12 tombol toolbar yang terdapat pada bagian atas program seperti berikut:



Setiap tombol toolbar berisi beberapa pilihan untuk memberikan akses yang cepat ke program sama seperti menggunakan fitur command.

Penjelasan untuk masing-masing tombol toolbar tersebut adalah:

- Open  = Berfungsi untuk membuka dataset STATA
- Save  = Berfungsi untuk menyimpan dataset STATA yang sedang berjalan di dalam memori ke disk
- Print  = Berfungsi untuk mencetak hasil analisis STATA
- Log  = Berfungsi untuk membuat sebuah log-File baru
- Viewer  = Berfungsi untuk membuka help advice
- Graph  = Berfungsi untuk menampilkan hasil analisis grafik
- Do-File Editor  = Berfungsi untuk membuat sebuah Do-File baru
- Data Editor (Edit)  = Berfungsi untuk membuka dan menginput data
- Data Editor (Browser)  = Berfungsi untuk mencari dataset
- Variables Manager  = Berfungsi untuk melakukan pengaturan variabel
- Clear-more-condition  = Berfungsi melanjutkan proses STATA ketika di pause
- Break  = Berfungsi menghentikan semua proses STATA yang sedang berlangsung

Dari tampilan utama STATA yaitu gambar 1,1 terdapat lima panel yakni:

- a. **Results** (panel bagian tengah), yang menunjukkan hasil (output) pengolahan perintah-perintah yang telah dijalankan
- b. **Command** (panel bagian bawah), yang digunakan untuk memasukkan syntax perintah-perintah pengolahan/analisis.

- c. **Review** (panel bagian kiri), berisikan daftar perintah-perintah yang telah dijalankan sebelumnya secara historis. Jadi, pengguna tidak perlu mengetikkan ulang atas perintah yang sama, cukup mengklik pada panel ini perintah yang hendak digunakan kembali. Perintah yang tercatat pada panel ini pun dapat disimpan untuk digunakan kemudian hari.
- d. **Variables** (panel kanan atas), menunjukkan daftar variabel yang digunakan dalam set data yang sedang diolah.
- e. **Properties** (panel kanan bawah), berisikan informasi/karakteristik dari variabel-variabel yang dipilih.

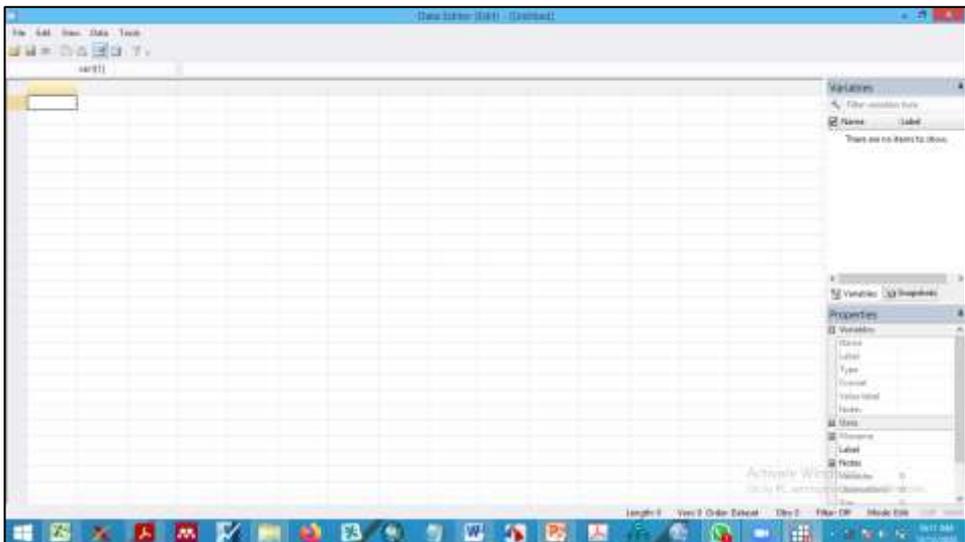
Selain menu bar dan panel-panel di atas, pada STATA terdapat informasi terkait *working directory* yang dapat dilihat pada bagian bawah program. Bagian ini menunjukkan lokasi penyimpanan file STATA yang sedang dibuka. Seluruh manipulasi dan perintah yang dimasukkan akan tersimpan secara default pada direktori ini kecuali dilakukan perubahan. Sangat disarankan untuk melakukan penyimpanan hasil analisis menggunakan STATA pada direktori yang sama dengan lokasi file-file lain yang terkait untuk mempermudah pekerjaan anda.

Setiap tampilan kolom menu diatas dapat diatur dan dipindahkan posisinya sesuai dengan yang dkehendaki oleh pengguna (catatan: untuk tampilan kolom menu utama program STATA tidak dapat dipindahkam posisinya, hanya berada diatas). Tampilan commad dan result adalah dua komponen yang penting dari program STATA. Ini berarti kita harus mempelajari bagaimana cara untuk memasukkan perintah yang terdiri dari satu baris yang diikuti dengan menekan Enter.

1.4. Perintah Dasar Untuk Penggunaan Awal

Penggunaan STATA pertama kali perlu dimulai dengan penggunaan data. Data yang akan dianalisis dapat bersumber dari:

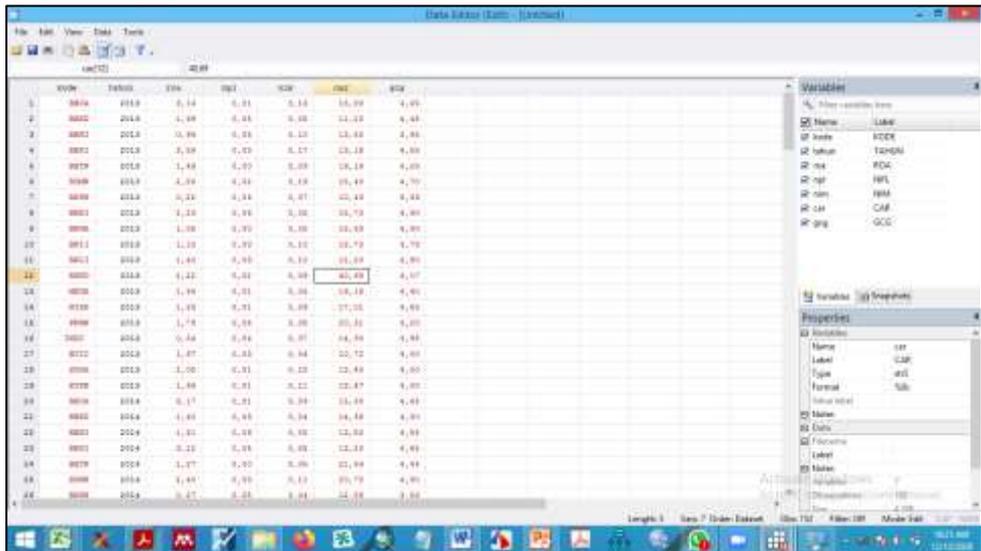
- a. Jika sebelumnya telah melakukan pembuatan *file* dan disimpan dalam pada *file* STATA menggunakan ekstensi *file* **.dat**, maka pengguna dapat mencari file tersebut melalui menu **File – Open** untuk kemudian mencarinya pada direktori penyimpanan. Cara lain menggunakan perintah pada panel *command*: *use nama_file.dta*
- b. Membuka *file* yang berasal dari basis data lainnya melalui perintah *import*, yakni **File – Import**. STATA dapat melakukan import atas *file-file* dengan ekstensi seperti .csv, .txt, .xls, juga .xlsx
- c. Melakukan *copy-paste* secara langsung ke STATA dari data yang sudah dimasukan pada *spreadsheet* (misalkan Microsoft Excel). Untuk melakukannya:
 1. *Copy* data yang hendak digunakan di STATA.
 2. Kembali ke STATA lalu buka *data editor* melalui Data → Data Editor. Akan muncul tabel-tabel seperti yang ditampilkan pada *spreadsheet*, seperti nampak pada gambar 1.2.
 3. *Paste* data yang sebelumnya telah di-*copy*.



Gambar 1.2. Contoh tampilan *data editor* pada STATA

Bila menggunakan cara ini harap diperhatikan:

1. Bila pada data asli yang di-copy-kan telah memiliki nama variabel, maka pilihlah ***treat first row as variable names***. Sementara itu, jika pada data asli belum memiliki variabel, maka pilihlah ***treat first row as data***. Jika baris pertama dianggap sebagai data, maka STATA secara otomatis akan memberi nama setiap variabel dengan var1, var2, var3, dan seterusnya.
2. Dalam pemberian nama variabel, STATA akan membatasi sebanyak 32 karakter dan harus didahului dengan huruf. Selain itu, nama variabel tidak boleh mengandung tanda baca atau spasi.
3. STATA akan mengabaikan baris kosong pada *spreadsheet*, namun tidak untuk kolom yang kosong sepenuhnya. Jika, dilakukan penyalinan terhadap kolom yang kosong sama sekali, STATA akan membacanya sebagai variabel dengan nilai yang kosong untuk setiap observasi.
4. Bila setelah dilakukan *copy-paste* terdapat nilai berwarna merah seperti contoh gambar 1.3., maka data tersebut diidentifikasi sebagai data nonnumerik (bukan angka). Terkadang pengguna memasukkan angka menggunakan separator “,” yang tidak dapat dibaca oleh STATA karena STATA menggunakan “.” Sebagai separator. Jika hal ini terjadi, pengguna diharapkan melakukan modifikasi separator pada data asli.



Gambar 1.3. Contoh Kesalahan Impor Data dari Excel

Latihan: Silahkan pindahkan data yang ada pada *file: data_keuangan.xls* ke STATA!

Setelah data dimasukkan ke *data editor* pada STATA, sekarang data siap digunakan sebagai bahan analisis. Selanjutnya, **Dataset** yang digunakan ini dapat disimpan untuk penggunaan berikutnya melalui **File → Save (Ctrl + S)** atau **Save As (Ctrl + Shift + S)**. Dapat pula menggunakan perintah *save nama_file* pada panel *command*.

Selain fungsi pembuatan dan penyimpanan *dataset* yang sebelumnya telah dijelaskan, fungsi dasar yang akan banyak digunakan dalam STATA ini adalah *doedit* dan *log*. Telah dijelaskan bahwa panel *review* digunakan untuk melihat perintah-perintah yang pernah digunakan untuk menganalisis suatu *dataset* untuk digunakan kembali kemudian. Jika pengguna menghendaki untuk menyimpan perintah-perintah tertentu melalui file dengan ekstensi **.do**. File **.do** dapat dibuka kemudian hari menggunakan perintah **do nama_file** pada panel

command, kemudian jika ingin melakukan pengeditan daftar perintah dapat dilakukan dengan perintah ***doedit nama_file***. Penyimpanan daftar perintah ini sangat membantu pengguna ketika menyusun laporan hasil penelitian yang dibuat sehingga mengetahui secara runut proses perolehan data yang digunakan sebagai basis analisis.

Di sisi lain, perintah *log* digunakan untuk penyimpanan hasil operasi *dataset* yang ditampilkan pada panel *results*. Untuk melakukan penyimpanan, pengguna secara mudah dapat melakukan dengan memasukkan perintah ***log nama_file*** pada panel *commands*. Jika pengguna ingin langsung menggunakan hasil yang ditampilkan pada aplikasi lain, dapat juga dilakukan *copy-paste* hasil pengolahan. Hasil pengolahan yang ditampilkan pada panel *results* memungkinkan untuk di-*copy-paste*-kan dalam bentuk ***table*** yang dapat diatur layaknya table pada spreadsheet atau ***picture*** jika menghendaki tampilan yang sama persis dengan tampilan di panel *results*.

1.5. Data dan Variabel

STATA selayaknya teori statistika secara umum, mengenal tiga jenis data, yakni:

- a. Data *Cross Section*, yakni data beberapa sampel untuk satu periode observasi. Jenis data ini merupakan data dasar (*default*) yang digunakan STATA. Jika pengguna sebelumnya telah mengatur *dataset*-nya kepada bentuk data set lain (dalam hal ini menjadi *time series* atau *panel*), maka untuk mengembalikan kepada dataset cross section dapat dilakukan dengan perintah ***clear tsset*** atau ***clear xtreg*** pada panel *command*.
- b. Data *Time Series*, yakni data satu sampel untuk beberapa periode observasi. Pengguna perlu mengidentifikasi jika dataset yang digunakan adalah data *time series* menggunakan perintah ***tsset variabel_waktu, unitoptions***. *Unitoptions* dapat dapat ditentukan berdasarkan *clocktime*, *daily*, *weekly*,

monthly, quarterly, halfyearly, yearly, generic tergantung data yang dimiliki/dikehendaki pengguna. STATA memiliki data dasar tahunan (*yearly*) atas dataset *time series*.

- c. Data *Panel*, yakni data atas beberapa sampel untuk beberapa periode observasi. Untuk mengatur *dataset* yang digunakan sebagai data panel, dapat dilakukan dengan menggunakan perintah ***xtreg variabel_panel variabel_waktu, unioptions.***

Khusus penggunaan data *time series*, perlu diperhatikan bahwa STATA memiliki format waktu untuk setiap jenis waktunya. STATA memiliki default waktu pada 1 Januari 1960, 00:00:00.000. Untuk memastikan bahwa data yang dimiliki dapat terbaca STATA, maka pastikan variabel penanda waktu mengikut format sebagai berikut:

- Clock time = %tc; 0 = 1Jan1960 00:00:00.000, 1 = 1Jan1960 00:00:00.001, ...
- Daily (harian) = %td; 0 = 1Jan1960, 1 = 2Jan1960, ...
- Weekly (mingguan) = %tw; 0 = 1960w1, 1 = 1960w2, ...
- Monthly (bulanan) = %tm; 0 = 1960m1, 1 = 1960m2, ...
- Quarterly (kuartalan) = %tq; 0 = 1960q1, 1 = 1960q2, ...
- Halfyearly (semesteran) = %th; 0 = 1960h1, 1 = 1960h2, ...
- Yearly (tahunan) = %ty; 0 = 1960, 1 = 1961, ...
- Generic (waktu umum) = %tg; 0 = ?, 1 = ?, ...

Untuk informasi lebih jelas terkait penggunaan data *time series* dapat dilihat pada *help tsset*.

Atas dataset yang dimiliki, pengguna STATA dapat menentukan suatu variabel di dalam dataset tersebut setidaknya ke dalam dua kelompok yakni data ***string*** dan data ***numerik***.

- a. Data *string* dapat memiliki nilai berbentuk huruf, tetapi tidak dapat dilakukan operasi matematis/statistik.
- b. Data numerik hanya menerima nilai berbentuk angka sehingga dapat dilakukan operasi matematis/statistika. Terdapat banyak jenis data numerik seperti integer (int), byte, atau float. Jenis data numerik ini biasanya akan langsung ditentukan STATA saat memindahkan data dari *spreadsheet* ke *data editor* STATA. Pilihan atas data numerik yang digunakan menentukan tingkat keakurasian nilai (misalnya seberapa banyak angka di belakang koma) dan kebutuhan memori yang dibutuhkan (yakni besarnya ukuran *file*). Jika ingin melakukan perubahan *type* variabel, dapat dilakukan secara langsung pada panel *properties* atau menggunakan perintah *recast type nama_variabel*

Untuk mengetahui jenis variabel yang digunakan dalam dataset dapat menggunakan perintah ***describe*** pada panel *command*. Sementara jika pengguna ingin merubah urutan variabel pada dataset yang dimiliki, dapat menggunakan perintah:

- a. ***order nama_variabel***, untuk menjadikan variabel tertentu berada pada urutan atas dalam dataset.
- b. ***order nama_variabel, last***, untuk menjadikan variabel tertentu berada pada urutan akhir dalam dataset.
- c. ***order nama_variabel, before nama_variabel***, untuk menjadikan variabel tertentu berada sebelum variabel tertentu.
- d. ***order nama_variabel, after nama_variabel***, untuk menjadikan variabel tertentu berada setelah variabel tertentu.

```

describe

obs:          152
vars:          8                               3 Dec 2020 20:24
size:         7,752
-----
variable name  storage  display  value  variable label
                type    format   label
-----
KODE           str5    %9s
TAHUN          int     %10.0g
ROA            double  %14.2f
NPL            double  %14.2f
NIM            double  %14.2f
CAR            double  %14.2f
GCG            double  %14.2f
BANK           long    %8.0g      BANK      KODE
-----
Sorted by:  BANK  TAHUN
    
```

Gambar 1.4. Contoh Perintah *Describe*

Latihan: Masukan perintah *describe* seperti contoh di atas, kemudian *copy* hasilnya ke word sebagai gambar!

STATA juga memberikan kemudahan jika pengguna hendak membuat variabel baru yang merupakan operasional dari variabel-variabel yang sudah ada pada dataset. Perintah umum yang dapat digunakan untuk menghasilkan variabel baru adalah ***generate type nama_variabel_baru=ekspresi***. Jenis variabel (*type*) dapat dikosongkan dan STATA akan menentukan secara otomatis sesuai hasil ekspresi yang dimasukkan. Ekspresi merupakan hasil operasi hitung atas variabel yang dipilih dapat penambahan (+), pengurangan (-), pembagian (/), perkalian (*), eksponensial (exp ()), logaritma natural (ln()), dan sebagainya (untuk lebih detil fungsi yang dapat digunakan dapat dilihat menggunakan perintah ***help functions***).

Berikut contoh pembuatan variabel baru yakni “lnROA” yang merupakan hasil operasi logaritma natural dari variabel “adv”, maka perintah yang dimasukan:

Generate lnROA = ln(ROA) atau

Gen lnROA=ln(ROA)

Perintah-perintah lainnya terkait variabel yang dapat digunakan:

- Jika hendak secara langsung mengganti suatu variabel lama dengan nilai baru hasil operasi, dapat digunakan perintah ***replace nama_variabel_lama = ekspresi.***
- Jika ingin menghapus suatu variabel dari dataset dapat menggunakan perintah ***drop nama_variabel.***
- Jika ingin mempertahankan suatu variabel dan menghapus selain variabel tersebut dapat menggunakan perintah ***keep nama_variabel.***
- Jika hendak melakukan konversi terhadap suatu variabel string menjadi variabel numerik untuk dapat dilakukan operasi matematis, maka dapat menggunakan perintah ***encode nama_variabel, generate(variabel_baru).*** Selain dapat pula dilakukan modifikasi terhadap label hasil perintah ***encode*** dengan menggunakan perintah ***recode.***

Latihan: Buatlah variabel baru NIM, CAR da GCG sesuai deskripsi yang diberikan pada penjelasan variabel!

Hal lainnya yang dapat membantu mempermudah pengguna dalam menggunakan STATA adalah keberadaan *Label*. Label digunakan sebagai penanda keseluruhan dataset (*label data*) dan penanda suatu variabel (*label variabel*). Label berguna untuk memudahkan pengguna mengidentifikasi suatu variabel atau dataset dikarenakan nama variabel terkadang terbatas serta tidak dapat menggunakan operator atau spasi. Untuk memberikan keterangan label pada suatu dataset atau variabel, dapat dilakukan pada panel *properties*. Dalam contoh misalnya terdapat variabel "*passrev*", untuk mempermudah dapat

ditambahkan informasi “pendapatan per penumpang” pada label variabel sehingga pengguna mengetahui dengan jelas bahwa “passrev” merupakan variabel yang berisikan nilai pendapatan maskapai atas setiap penumpang.

Latihan: Berikan label untuk setiap variabel sesuai dengan penjelasan variabel yang ada pada file: data_keuangan.xls!

BAB II.

ANALISIS DESKRIPTIF

Sebelum melakukan analisis utama terhadap data yang dimiliki, seorang yang sedang melakukan penelitian perlu memahami karakteristik data yang digunakan sehingga dapat lebih tajam dalam menarik kesimpulan. Salah satu metode yang dapat digunakan yakni dengan menggunakan analisis deskriptif. Statistika deskriptif hanya akan melakukan upaya untuk menguraikan, mempelajari, dan memberikan keterangan mengenai data yang digunakan. Secara umum terdapat dua kelompok besar analisis deskriptif yakni:

- a. Ukuran pemusatan, yakni untuk mengetahui bagaimana distribusi suatu data berpusat. Contohnya antara lain: mean (rerata), median (nilai tengah), modus (nilai terbanyak).
- b. Ukuran penyebaran, yakni untuk mengetahui bagaimana penyimpangan atau menyebarnya suatu data dibandingkan titik pusatnya. Contohnya adalah varians dan standar deviasi.

2.1 Analisis Deskriptif Menggunakan STATA

Sebagai sebuah perangkat lunak statistika, STATA tentunya memiliki kemampuan untuk melakukan analisis deskriptif untuk membantu penggunaanya dalam mendapatkan pemahaman data secara lebih baik. Berikut ini perintah-perintah yang dapat digunakan untuk memperoleh statistika deskriptif menggunakan STATA:

- a. **Summarize**, dapat digunakan untuk menunjukkan nilai-nilai statistika deskriptif yang paling jamak digunakan sebagai informasi awal pemusatan

dan penyebaran data antara lain mean, standar deviasi, nilai observasi minimal, dan nilai observasi maksimal.

```
summarize ROA NPL NIM CAR GCG
```

Variable	Obs	Mean	Std. Dev.	Min	Max
ROA	152	1.239812	1.219919	-.69	4.45
NPL	152	.0190807	.0173788	0	.1052433
NIM	152	.0696934	.0504637	.0024	.546
CAR	152	14.52688	6.013934	0	46.49
GCG	152	4.233257	.491731	2.33	5


```
summarize BANK TAHUN ROA NPL NIM CAR GCG
```

Variable	Obs	Mean	Std. Dev.	Min	Max
BANK	152	10	5.495332	1	19
TAHUN	152	2016.5	2.298862	2013	2020
ROA	152	1.239812	1.219919	-.69	4.45
NPL	152	.0190807	.0173788	0	.1052433
NIM	152	.0696934	.0504637	.0024	.546
CAR	152	14.52688	6.013934	0	46.49
GCG	152	4.233257	.491731	2.33	5

Gambar 2.1. Hasil Perintah *Summarize nama_variabel*

- b. Jika hendak mengetahui secara mendetil karakteristik deskriptif suatu variabel maka dapat menggunakan perintah ***summarize nama_variable, detail.***

```
summarize ROA , detail
```

ROA				

	Percentiles	Smallest		
1%	-.0781953	-.69		
5%	.0028779	-.0781953		
10%	.0113771	-.0763563	Obs	152
25%	.0263232	-.0528255	Sum of Wgt.	152
50%	1.35		Mean	1.239812
		Largest	Std. Dev.	1.219919
75%	2.02	3.69		
90%	3.11	3.99	Variance	1.488203
95%	3.39	4.33	Skewness	.5215226
99%	4.33	4.45	Kurtosis	2.23055

Gambar 2.2. Contoh Hasil Perintah

- c. Perintah **codebook** juga dapat digunakan untuk mengetahui karakteristik masing-masing variabel yang digunakan.

```
codebook ROA
```

ROA					

	type: numeric (double)				
	range: [-.69,4.45]		units: 1.000e-11		
	unique values: 138		missing .: 0/152		
	mean: 1.23981				
	std. dev: 1.21992				
percentiles:	10%	25%	50%	75%	90%
	.011377	.026323	1.35	2.02	3.11

Gambar 2.3. Contoh Hasil Perintah *Codebook*

- d. Jika hendak menayangkan data dalam bentuk daftar, dapat menggunakan perintah **list**.
- e. Untuk menunjukkan tabulasi suatu atau dua variabel, dapat menggunakan perintah **tabulate nama_variabel** untuk tabulasi satu variable atau perintah

tabulate nama_variabel1 nama_variabel2, untuk tabulasi antar dua variable.

```
tabulate ROA
```

ROA	Freq.	Percent	Cum.
-0.69	1	0.66	0.66
-0.08	1	0.66	1.32
-0.08	1	0.66	1.97
0.00	1	0.66	3.95
3.52	1	0.66	96.05
3.57	1	0.66	96.71
3.59	1	0.66	97.37
3.69	1	0.66	98.03
3.99	1	0.66	98.68
4.33	1	0.66	99.34
4.45	1	0.66	100.00
Total	152	100.00	

Gambar 2.4. Contoh tabulasi satu variabel

Pada contoh di atas dapat diketahui berapa banyak observasi yang merupakan perusahaan – perusahaan bank dan beberapa variable yang memiliki kecenderungan mempengaruhi kinerja nya dalam suatu periode tertentu.

Latihan: Lakukan analisis deskriptif atas data yang ada pada file: data_keuangan.xls (setidaknya tunjukkan nilai mean, median, modus, standar deviasi, nilai min, nilai max)

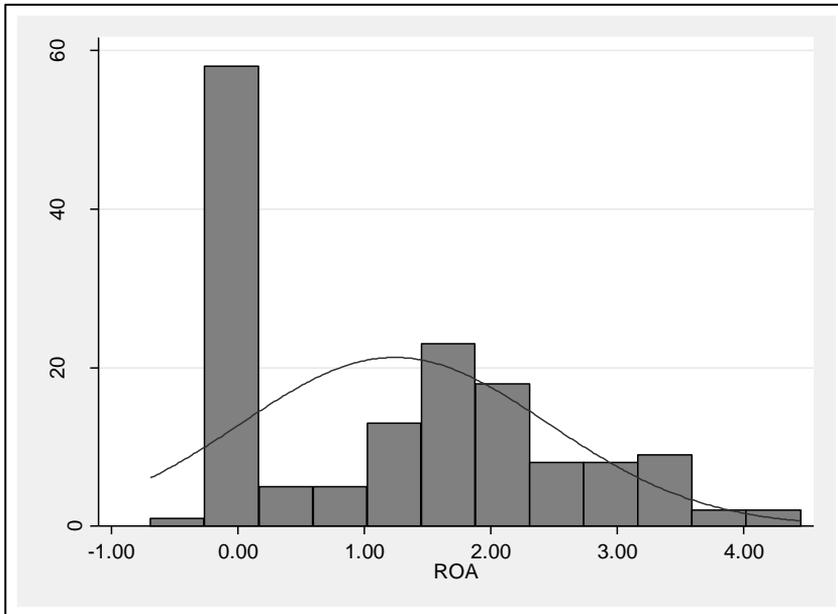
2.2 Menggunakan Tabel dan Grafik Statistika Deskriptif pada STATA

Selain menggunakan nilai-nilai statistika seperti yang ditunjukkan pada bagian sebelumnya, analisis deskriptif juga dapat dilakukan menggunakan bantuan tabel/grafik/histogram. Penggunaan tabel/grafik/histogram memberikan kemudahan bagi pengguna untuk mendapatkan visualisasi pemusatan dan penyebaran data secara lebih cepat. Untuk memperoleh tabel/grafik/histogram, pengguna dapat menggunakan menu utama **Graphic**. Pada menu tersebut, terdapat banyak sekali pilihan jenis visualisasi statistika deskriptif sesuai kebutuhan pengguna, Selain menggunakan menu *Graphics*, tentunya STATA memiliki beberapa perintah yang dapat digunakan antara lain:

- a. Perintah ***histogram nama_variabel*** dapat digunakan untuk menampilkan histogram suatu variabel. Opsi ***normal*** dapat dimasukkan bila menghendaki diperlihatkannya garis distribusi normal. Sedangkan opsi ***frequency*** dapat digunakan untuk menunjukkan frekuensi setiap kelompok histogram. Perintah ***discrete*** dapat digunakan untuk mengetahui diagram batang untuk setiap nilai.

Command nya adalah

histogram ROA, normal frequency

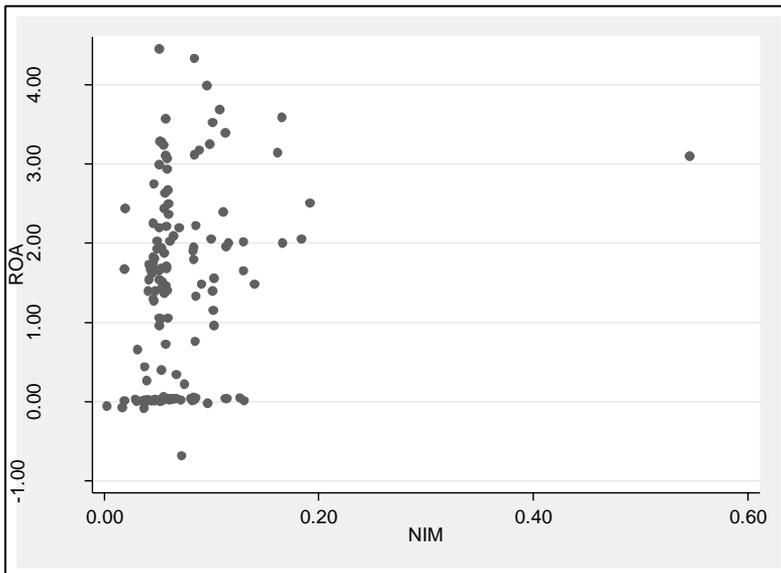


Gambar 2.6. Contoh Histogram untuk variabel ROA dengan tambahan opsi normal dan frekuensi

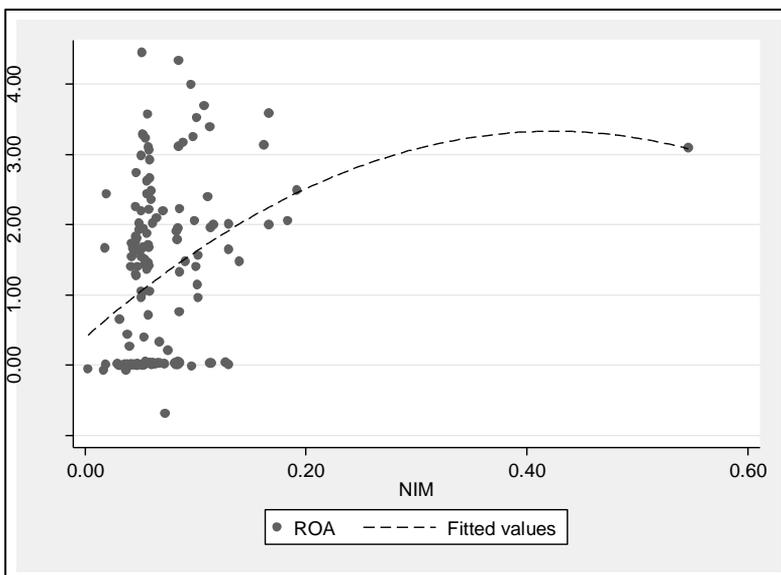
- b. Jika hendak melihat grafik antarvariabel, dapat menggunakan perintah ***twoway jenis_plot nama_variabel1 nama_variabel2***. Terdapat dua jenis plot yang dapat digunakan yakni plot titik (***scatter***) dan garis (***line***). Untuk menambahkan garis tren dapat ditambahkan pilihan opsi ***lfit*** (untuk *linear fit*) atau ***qfit*** (untuk *quadratic fit*). Untuk menyepesifikasikan plot yang dihasilkan dengan beberapa skala axis (x) atau ordinat (y), pada opsi dapat ditambahkan ***yaxis(#)*** dan ***xaxis(#)***. Menu ***Help*** menginformasikan lebih detail terkait penggunaan opsi-opsi ini.

Command nya adaah

twoway scatter ROA NIM

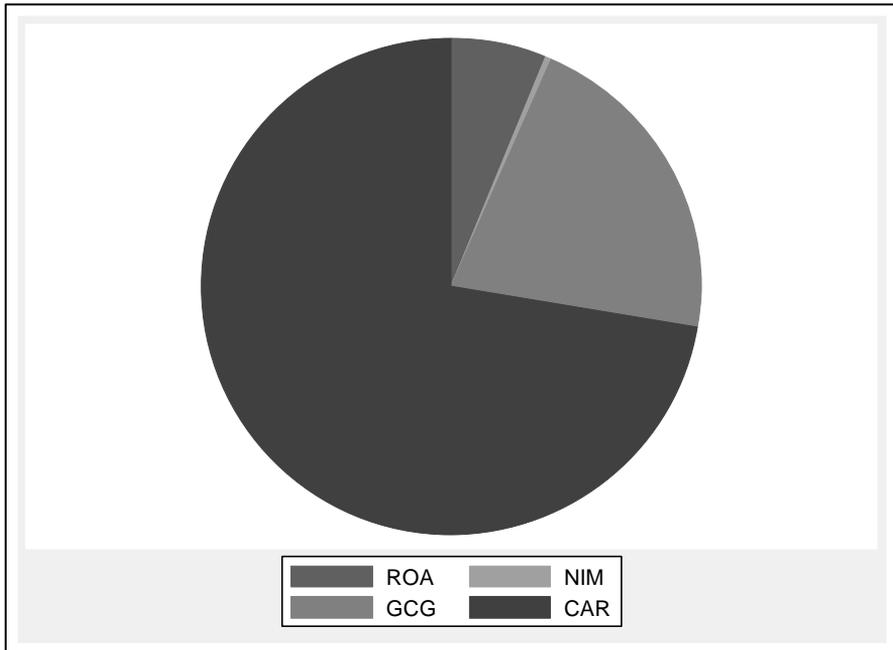


Gambar 2.7. Contoh: *twoway scatter ROA NIM*



Gambar 2.8. Contoh: *twoway (scatter ROA NIM) (qfit ROA NIM)*

- c. Perintah **graph** dapat digunakan untuk memperlihatkan suatu grafik. Contoh untuk menunjukkan grafik pie untuk jenis bank dengan ROA, NIM, CAR dan GCG maka dapat digunakan perintah *graph pie NIM dan variabel lainnya, misalnya graph pie ROA NIM GCG CAR*



Gambar 2.9. Contoh: *graph pie ROA NIM GCG CAR*

BAB III.

ANALISIS REGRESI

3.1. Analisis Regresi Linear

Bila pada bagian sebelumnya telah diketahui gambaran karakteristik data yang dimiliki menggunakan analisis deskriptif, maka pada bagian ini akan dilakukan analisis inferensial. Analisis inferensial dilakukan untuk mengetahui apakah suatu variabel dependen (terikat) dapat diprediksi oleh suatu variabel independen (bebas). Dengan kata lain, melalui analisis ini pengguna diharapkan dapat menentukan apakah kenaikan/penurunan suatu variabel independen dapat mempengaruhi kenaikan/penurunan variabel dependen. Namun perlu dicatat bahwa dalam menentukan variabel-variabel yang akan diujikan, perlu didasarkan pada teori atau konsep yang terkait variabel-variabel tersebut.

Analisis regresi sederhana hanya melibatkan hubungan kausal antara satu variabel dependen dan satu variabel independen. Sementara analisis regresi berganda melibatkan beberapa variabel independen. Adapun persamaan umum regresi yakni:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

dengan:

Y = Variabel dependen yang hendak diprediksi

X = Variabel independen

β_0 = *intercept*/konstanta, yakni nilai variabel dependen pada saat variabel independen = 0

β_1 = koefisien regresi, yakni besarnya kenaikan (penurunan) variabel dependen ketika terjadi kenaikan (penurunan) variabel independen sebesar 1 unit.

ε = term error, perbedaan antara data sampel hasil regresi dengan data populasi.

n / i = jumlah koefisien / variabel

Menurut Gujarati dan Porter (2008) Penggunaan regresi linear sederhana harus memenuhi kriteria yakni:

- a. Parameter menghasilkan varians error yang paling minimal (**Best**)
- b. Persamaan yang diestimasi adalah persamaan linear (garis lurus) (**Linear**)
- c. Parameter estimasi tidak bias (**Unbiased**)
- d. Parameter mampu untuk mengestimasi secara baik parameter populasi (**Estimator**)

Namun sayangnya, sangat sulit ditemukan suatu model yang benar-benar memenuhi asumsi-asumsi tersebut. Pada bagian berikutnya, akan dipelajari teknik-teknik yang dapat digunakan untuk mendeteksi adanya pelanggaran atas asumsi-asumsi tersebut serta bagaimana melakukan koreksi atas pelanggaran yang ditemukan.

3.2. Hipotesis

Analisis regresi merupakan salah satu alat yang dapat digunakan untuk membantu seorang peneliti dalam menarik kesimpulan atas hipotesis. Hipotesis yakni pernyataan statistika mengenai parameter suatu populasi. Pengujian suatu hipotesis akan menentukan apakah suatu hipotesis:

- a. ditolak, yakni hipotesis tersebut tidak tepat

- b. diterima (tidak dapat menolak), yakni tidak terdapat bukti yang memadai untuk menolak hipotesis.

Pengembangan hipotesis selalu menggunakan pasangan hipotesis, yaitu:

Hipotesis nol (**H₀**): tidak adanya perbedaan antara ukuran sampel dengan ukuran populasi

Hipotesis alternatif (**H₁**): adanya perbedaan antara ukuran sampel dengan ukuran populasi

Pengujian hipotesis dapat dilakukan dengan melihat nilai-nilai suatu ukuran statistika, seperti nilai F-stat (untuk pengujian global) atau nilai t-stat (untuk pengujian individual) yang kemudian akan dibandingkan dengan nilai tabel. Selain itu dapat juga menggunakan nilai *p-value* untuk menentukan apakah suatu variabel independen memiliki signifikansi terhadap suatu variabel dependen. **Nilai *p-value* akan dibandingkan dengan nilai α yakni rentang keyakinan (*confidence interval*).** Nilai α yang umumnya digunakan adalah pada level 1% dan 5%, sedangkan pada level α 10% dianggap sebagai marjinal signifikan. Perlu diperhatikan bahwa ada kemungkinan hasil pengujian statistika yang dilakukan menyebabkan dua kesalahan berikut:

- a. Kesalahan tipe I (Type I Error), menolak H₀ ketika H₀ Benar
- b. Kesalahan tipe II (Type II Error), menerima H₀ ketika H₀ salah

3.3. Analisis Regresi Sederhana Menggunakan STATA

STATA sebagai suatu perangkat lunak statistika, memiliki fitur untuk melakukan analisis regresi yang memadai. Melakukan analisis regresi menggunakan STATA dapat menggunakan menu STATISTICS >> Linear Models and related >> Linear regression. Jika menggunakan cara ini pengguna tinggal

memilih dari daftar variabel yang dimiliki untuk setiap variabel dependen dan variabel independen yang akan digunakan. Sementara jika hendak menggunakan sintaks, menggunakan sintaks umum untuk regresi pada stata yakni:

reg variabel_dependen variabel_independen, opsi

Untuk latihan kali ini, dapat digunakan model penelitian berdasarkan informasi berikut:

Penelitian ini mencoba melihat pengaruh dari NPL, NIM, CAR, dan GCG terhadap ROA di beberapa Bank dalam periode 2013-2020. Penjelasan lengkap untuk setiap variabel adalah sebagai berikut:

1. **YEAR:** tahun observasi meliputi tahun ke-1 hingga tahun ke-8 (2013 sd 2020)
2. **BANK:** Perusahaan yang diobservasi
3. **ROA** (*Return On Asset*) adalah rasio untuk mengukur efektifitas bank dalam memperoleh keuntungan secara keseluruhan melalui pengoperasian total aset yang dimiliki bank. Dalam hal ini pengukurannya adalah rasio antara laba sebelum pajak dengan total aset
4. **NPL** (*Net Performing Loan*) merupakan rasio yang mencerminkan besarnya kredit bermasalah yang dihadapi bank. Semakin tinggi rasio ini maka akan semakin buruk kualitas kredit bank, yang menyebabkan jumlah kredit bermasalah semakin besar. Dalam hal ini pengukurannya adalah pembagian antara kredit bermasalah dengan total kredit.
5. **NIM** (*Net Interest Margin*) adalah rasio untuk mengukur kemampuan manajemen bank dalam mengelola aktiva produktifitasnya untuk menghasilkan pendapatan bunga bersih. Jadi semakin besar rasio ini maka akan semakin besar earning uang diperoleh bank dari pendapatan bunga.

Dalam hal ini pengukurannya adalah pembagian antara pendapatan bunga bersih dengan rata-rata total aset produktif

6. **CAR** (*Capital Adequacy Ratio*) adalah rasio yang memperlihatkan seberapa besar jumlah seluruh aktiva bank yang mengandung resiko ikut dibiayai dari modal sendiri. Dalam hal ini pengukurannya adalah pembagian antara modal dengan aktiva tertimbang menurut resiko.
7. **GCG** (*Good Corporate Governance*) merupakan sistem yang mengendalikan perusahaan untuk menciptakan nilai tambah untuk semua stakeholder. Dalam hal ini pengukurannya adalah dari data survey yang dilakukan lembaga Indonesian Institute for Corporate Governance (IICG)

Khusus untuk bahasan ini, akan digunakan model regresi linear sebagai berikut:

$$ROA = \beta_0 + \beta_1 NPL + \beta_2 NIM + \beta_3 CAR + \beta_4 GCG + e$$

Maka sintaks STATA yang digunakan adalah

reg (atau regress) ROA NPL NIM CAR GCG

Pada gambar 3.1, terlihat hasil pengujian model di atas menggunakan STATA.

reg ROA NPL NIM CAR GCG						
Source	SS	df	MS	Number of obs	=	152
-----				F(4, 147)	=	9.96
Model	47.9072824	4	11.9768206	Prob > F	=	0.0000
Residual	176.811404	147	1.20279867	R-squared	=	0.2132
-----				Adj R-squared	=	0.1918
Total	224.718686	151	1.48820322	Root MSE	=	1.0967

ROA	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
NPL	5.491708	5.223975	1.05	0.295	-4.832086	15.8155
NIM	5.79292	1.784418	3.25	0.001	2.266495	9.319345
CAR	.0544075	.0150772	3.61	0.000	.0246114	.0842037
GCG	.7422607	.1859474	3.99	0.000	.3747854	1.109736
_cons	-3.201254	.8709542	-3.68	0.000	-4.922462	-1.480045

Gambar 3.1 Contoh hasil pengujian regresi

Berdasarkan hasil pengujian yang terlihat pada Gambar 3.1., informasi-informasi yang perlu diperhatikan sebagai bahan analisis adalah

- A. *Number of obs* = 152 artinya jumlah sampel atau observasi sebanyak 152 sampel, kemudian, *F (4, 147)* artinya uji F pada DF 4 dan 147. DF 4 artinya jumlah variabel yang diuji-1, yaitu $5-1=4$ variabel. 147 adalah jumlah observasi – jumlah variabel, yaitu $152-5=147$. Uji F ini menunjukkan hasil pengujian global (*global test / F-test*) yakni untuk mengetahui apakah secara menyeluruh model yang digunakan dapat menjelaskan secara signifikan variabel dependen. STATA menyediakan informasi F-test untuk dibandingkan dengan F-tabel maupun nilai *p-value* ($\text{Prob} > F$) untuk dibandingkan dengan α .
- B. $\text{Prob} > F = 0.0000$ adalah Nilai Uji F, artinya jika nilai $\text{Prob} > F < 0.05$ maka Uji F menerima H_1 pada taraf signifikansi 5% atau yang berarti semua variabel independen secara simultan mempunyai pengaruh yang signifikan pada variabel dependen. Kemudian, *R-Squared* adalah Koefisien Determinasi Berganda, artinya seberapa besar secara simultan semua variabel independen dapat menjelaskan variabel dependen sebesar 21,32%. Maka sisainya adalah $100\% - 21,32\% = 78.68\%$ dipengaruhi oleh variabel lain diluar model regresi. Hal ini juga menunjukkan *goodness of fit* dari suatu model yakni seberapa banyak variasi yang terjadi pada variabel dependen dapat dijelaskan menggunakan model yang diujikan.
- C. Nilai konstanta (β_0), yakni nilai variabel dependen ketika seluruh variabel independen bernilai = 0.
- D. Kolom *Coef* adalah nilai Unstandardized Koefisien Beta. Nilai koefisien beta ini menunjukkan koefisien estimasi (β_1) untuk setiap variabel independen

yang digunakan pada model. Tanda + atau – menunjukkan arah pengaruh variabel independen terhadap variabel dependen.

- E. Menunjukkan standar deviasi dari koefisien estimasi
- F. Menunjukkan hasil pengujian individual (*individual test / t-test*) yakni untuk mengetahui apakah secara individu, masing-masing variabel independen yang digunakan memiliki pengaruh yang signifikan terhadap variabel dependen. Nilai ini akan dibandingkan dengan nilai t-tabel.
- G. Alternatif untuk melakukan pengujian untuk mengetahui signifikansi suatu variabel. Nilai ini akan dibandingkan dengan nilai α .
- H. Menunjukkan jangkauan (*range*) setiap parameter estimasi.

Dalam melakukan analisis regresi, hasil pengujian yang perlu diperhatikan selain signifikansi variabel, juga adalah arah pengaruh setiap variabel. Kedua hal ini yang kemudian akan mendukung atau bertentangan dengan hasil penelitian yang telah ada sebagai suatu temuan dari penelitian yang dilakukan.

Latihan: Lakukan pengujian regresi linear pada data yang ada pada file:
data_keuangan.xls, seperti contoh di atas!

3.4. Pendeteksian dan Pengoreksian Pelanggaran Asumsi

Asumsi klasik merupakan syarat yang harus dipenuhi pada model regresi yang menggunakan metode estimasi OLS, supaya menghasilkan karakteristik tidak bias, konsisten dan efisien (disebut *best linier, unbiased estimator, BLUE*). Pengujian atas asumsi klasik dilakukan untuk mengetahui apakah model pengujian yang digunakan memenuhi ketentuan-ketentuan statistika yang harus dipenuhi pada analisis linear berganda. Ketika suatu asumsi dilanggar,

maka pengguna akan melakukan suatu teknik/perlakuan sehingga pelanggaran tersebut dapat diperbaiki.

a. Multikolinearitas

Multikolinearitas adalah pelanggaran asumsi dikarenakan adanya variabel independen yang memiliki hubungan (korelasi) yang kuat di dalam suatu model regresi. Sebagai contoh, penggunaan variabel pendapatan/penjualan (*sales*) bersamaan dengan variabel total aset, secara umum biasanya semakin tinggi total aset suatu perusahaan maka akan memiliki nilai penjualan yang semakin tinggi. Keterkaitan yang erat antardua variabel ini bisa menyebabkan terjadinya multikolinearitas jika digunakan secara bersamaan pada suatu model penelitian. Adapun indikasi dan cara yang dapat dilakukan untuk mengidentifikasi asumsi multikolinearitas adalah:

1. Jika menemukan bahwa model yang digunakan signifikan secara global (f-test signifikan), tetapi tidak ada variabel independen yang signifikan (t-test), maka ada memiliki masalah multikolinearitas.
2. Menguji korelasi antarvariabel. Variabel dikatakan ada korelasi yang kuat jika memiliki korelasi (lebih besar dari +/- 0,8) terhadap variabel lainnya. Jika menggunakan STATA pengujian korelasi ini dapat menggunakan perintah:

Corr nama_variabel1 nama_variabel2

```
corr NPL NIM
(obs=152)

      |      NPL      NIM
-----+-----
NPL |      1.0000
NIM |     -0.0777      1.0000
```

Gambar 3.2 Contoh hasil pengujian korelasi

Selain besaran korelasi, perlu diperhatikan juga arah korelasi untuk menentukan hubungan antarvariabel apakah searah (+) atau berlawanan (-).

Latihan: Lakukan pengujian korelasi untuk variabel yang digunakan dalam model seperti contoh di atas! Silakan interpretasikan hasil pengujian tersebut!

3. Membandingkan nilai koefisien regresi yang dihasilkan (nilai β) dengan koefisien korelasi antara variabel independen tersebut dengan variabel dependen. Indikasi terjadinya multikolinearitas jika ditemukan: (a) terdapat perubahan tanda koefisien (+/-) pada koefisien regresi dan korelasi, (b) terdapat perubahan signifikansi pada koefisien korelasi dan koefisien regresi. Untuk mengetahui signifikansi (*p-value*) suatu koefisien korelasi dapat menggunakan sintaks:

pwcorr nama_variabel1 nama_variabel2, sig

Latihan: Tunjukkan nilai signifikansi dari koefisien korelasi, lakukan identifikasi apakah ada indikasi multikolinearitas pada model!

4. Nilai *Variance Inflation Factor* (VIF) juga dapat digunakan untuk mengecek terjadinya multikolinearitas. Jika nilai VIF suatu variabel lebih dari 10, maka terindikasi ada masalah multikolinearitas. Pada STATA, nilai VIF dapat dimunculkan menggunakan perintah ***vif***.

Syntac yang digunakan adalah

Diawali dengan *me-regres*

```
reg ROA NPL NIM CAR GCG
```

```
estat ic
```

```
estat vif
```

Variable	VIF	1/VIF
GCG	1.05	0.952754
NPL	1.03	0.966442
CAR	1.03	0.968848
NIM	1.02	0.982348
Mean VIF	1.03	

Gambar 3.3 Contoh hasil pengujian vif

Nilai VIF direkomendasikan harus < 5 dengan nilai Tolerance harus > 0.20 . Berdasarkan hasil output di atas nilai VIF yang dihasilkan untuk semua prediktor < 5 dan nilai Tolerance > 0.20 sehingga dapat disimpulkan tidak terjadi problem multikolonieritas di dalam model yang diujikan sebelumnya.

Latihan: Tunjukkan nilai VIF atas variabel-variabel pada model!

5. Menggunakan grafik untuk melihat hubungan antarvariabel menggunakan sintaks ***graph matrix nama_variabel nama_variabel***

Suatu kondisi, jika terdapat indikasi multikolinearitas, maka beberapa teknik/pelakuan yang dapat dilakukan adalah:

1. Menghapus salah satu variabel independen yang diindikasikan mengalami masalah multikolinearitas. Namun, hal ini dapat menyebabkan adanya kesalahan pengukuran karena ada variabel yang diabaikan.

2. Menambah data.
3. Menggunakan metode analisis lainnya seperti *stepwise regression*, *two-stage least square (2SLS)*, *limited-information maximum likelihood (LIML)*, *generalized method of moments (GMM)*. Pada STATA, dapat menggunakan sintaks sebagai berikut:

ivregress estimator nama_variabel

(pilihan estimator: 2sls, liml, gmm)

4. Tidak melakukan apa-apa, selama tidak terjadi masalah serius akibat multikolinearitas seperti perubahan tanda atau perubahan signifikansi.

b. Heterokedastisitas

Suatu analisis regresi menggunakan asumsi homokedastis, yakni varians error tidak berubah (konstan) seiring dengan perubahan nilai variabel independen. Tidak konstannya varians error menyebabkan hasil estimasi tidak efisien. Heterokedastisitas dapat disebabkan oleh (1) kondisi alamiah dari data yang digunakan, (2) adanya kesalahan input data, atau (3) adanya manipulasi data yang menyebabkan error memiliki varian yang sistematis. Untuk mendeteksi adanya permasalahan heterokedastisitas dapat menggunakan teknis sebagai berikut:

1. Menggunakan grafik dengan melakukan plot dari variabel dependen sebagai sumbu Y dan masing-masing variabel independen sebagai sumbu X. Membuat *scatterplot* dengan STATA dapat dilakukan menggunakan sintaks.

plot sumbu_Y sumbu_X

2. Menggunakan Uji White (*white-test*), dapat menggunakan sintaks:

estat imtest, white

berikut contoh hasil pengujian white untuk model yang diujikan sebelumnya

```
estat imtest, white

White's test for Ho: homoskedasticity
      against Ha: unrestricted heteroskedasticity

      chi2(14)      =      19.82
      Prob > chi2   =      0.1358

Cameron & Trivedi's decomposition of IM-test

-----+-----
           Source |           chi2    df    p
-----+-----
      Heteroskedasticity |      19.82    14    0.1358
           Skewness |      10.98     4    0.0267
           Kurtosis |       0.23     1    0.6313
-----+-----
           Total |      31.04    19    0.0400
-----+-----
```

Gambar 3.4 Contoh hasil pengujian white

Berdasarkan hasil pengujian di atas, dapat diketahui bahwa pengujian regresi dengan model yang sebelumnya telah dilakukan tidak mengalami masalah heterokedastisitas.

3. Menggunakan uji Szroeter. Ketik command di bawah ini untuk menjalankan uji tersebut

estat szroeter ROA NPL NIM CAR GCG

berikut adalah hasil pengujian Szroeter untuk model yang diujikan sebelumnya

```
estat szroeter ROA NPL NIM CAR GCG

Szroeter's test for homoskedasticity

Ho: variance constant
Ha: variance monotonic in variable

-----
Variable |      chi2   df      p
-----+-----
      ROA |     19.50    1  0.0000 #
      NPL |      1.62    1  0.2027 #
      NIM |      3.52    1  0.0605 #
      CAR |      0.23    1  0.6282 #
      GCG |      0.05    1  0.8216 #
-----
# unadjusted p-values
```

Gambar 3.5 Contoh hasil pengujian Szroeter

Dari hasil output disimpulkan, model regresi mempunyai varian yang konstan atau tidak terdapat problem heterokedastisitas karena nilai P yang dihasilkan > 0.05, kecuali variabel ROA dimana nilai P yang dihasilkan < 0.05, masalah ini dapat ditimbulkan dari ketidaknormalan dari data sebelumnya.

- 4. Menggunakan pengujian Breusch-Pagan / Godfrey, dapat menggunakan sintaks:

estat hettest

berikut contoh hasil pengujian white

```
estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of ROA

chi2(1)      =      0.01
Prob > chi2  =      0.9232
```

Gambar 3.6 Contoh hasil pengujian Breush-Pagan

Uji heterokedastisitas hanya dilakukan ketika menggunakan estimasi FE dan PLS. Hipotesis yang diterapkan dalam uji ini adalah

H_0 : Homoskedastis

H_1 : Heteroskedastis

Hipotesis nol akan ditolak bila $(\text{Prob} > \chi^2) < \alpha$ atau nilai $\chi^2 >$ nilai kritis t_{tabel}

Berdasarkan hasil pengujian Breusch-Pagan di atas menunjukkan tidak mengalami masalah heterokedastisitas, hal tersebut dibuktikan dengan nilai p value yang dinyatakan dengan nilai $\text{Prob} > \chi^2$ adalah 0.9232, dimana nilai tersebut > 0.05 , maka dapat dinyatakan H_0 adalah homoskedastis tidak dapat ditolak.

Dalam suatu hal, jika terdapat indikasi heterokedastisitas, maka beberapa teknik/pelakuan yang dapat dilakukan adalah:

1. Menambah data sampel.
2. Melakukan transformasi variabel merujuk pada variabel dependen yang dimiliki, seperti melakukan transformasi data menjadi bentuk logaritma, rasio, dan sebagainya.
3. Menggunakan tambahan opsi **robust** ketika melakukan pengujian regresi.
4. Menggunakan metode estimasi lain, seperti *generalized least square* (GLS)
5. Menggunakan model regresi dengan ARCH orde 1.

Latihan: Lakukan pengujian indikasi terjadinya heterokedastisitas pada file: data_keuangan.xls ke STATA, kemudian lakukan langkah koreksi atas pelanggaran asumsi heterokedastisitas yang ditemukan!

c. Autokorelasi

Autokorelasi adalah keadaan dimana terjadi hubungan error antarwaktu pada data yang digunakan. Korelasi dapat dimaknai (i) korelasi antara variabel

dan (ii) korelasi antar periode waktu. Jika terjadi korelasi yang kuat antar variabel dapat mengakibatkan terjadinya masalah multikolinearitas. Sedangkan jika terjadi korelasi yang kuat antar periode waktu dapat mengakibatkan terjadinya autokorelasi.

Kasus autokorelasi banyak ditemukan ketika menggunakan data *time-series*, artinya kondisi sekarang (periode t) dipengaruhi waktu lalu (t-n). Atau suatu kondisi di mana sifat residual regresi yang saling berkaitan antara satu observasi (ke-i) dengan observasi lainnya (ke-j). Oleh karena itu, dalam analisis data time series, masalah autokorelasi menjadi pusat perhatian dalam permodelan. Selain itu, kasus autokorelasi dapat disebabkan oleh (1) tidak dimasukkannya suatu variabel penting pada model regresi yang digunakan, (2) adanya manipulasi data yang menyebabkan error memiliki varian yang sistematis, atau (3) hubungan yang tidak linear antara variabel dependen dan independen. Untuk mendeteksi adanya indikasi autokorelasi, dapat dilakukan dengan cara:

1. Menggunakan nilai Durbin-Watson untuk menguji adanya autokorelasi lag-1. Pada STATA dapat menggunakan sintaks:

estat dwatson

diindikasikan terdapat autokorelasi jika nilai Durbin-Watson jauh dari 2.

2. Menggunakan pengujian Breuch-Godfrey, dapat menggunakan sintaks:

estat bgodfrey, lags(nomor lags yang hendak diujikan)

3. Menggunakan Uji *serial correlation in the idiosyncratic error of a linear panel-data model*. Sintaks yang digunakan:

xtserial varibale dependen variabel-variabel independen

Misalnya dalam uji ini adalah seperti berikut:

\

```
xtserial ROA NPL NIM CAR GCG

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation
    F( 1,      18) =      49.303
      Prob > F =      0.0000
```

Gambar 3.7 Contoh hasil pengujian Serial correlation

Menurut Wooldridge (2002) pernyataan Hipotesis yang diterapkan dalam uji ini adalah

Hipotesis:

H_0 : No autokorelasi

H_1 : Autokorelasi

Hipotesis nol akan ditolak bila $(\text{Prob} > \chi^2) < \alpha$ atau nilai $\chi^2 >$ nilai kritis t-tabel.

Berdasarkan pada uji *serial correlation in the idiosyncratic error of a linear panel-data model* menunjukkan hasil mengalami masalah autokorelasi, hal tersebut dibuktikan dengan nilai $\text{Prob} > F$ lebih kecil dari α .

Dalam kondisi lain, jika terdapat indikasi autokorelasi, maka beberapa teknik/pelakuan yang dapat dilakukan adalah:

1. Menambahkan variabel *lag* pada model pengujian. (autoregressive orde 1 (AR(1)))
2. Menggunakan transformasi Cochrane-Orkutt dengan sintaks: *prais variabel_dependen variabel_independen, corc*

d. Normalitas

Analisis regresi linear mengasumsikan bahwa error terdistribusi secara normal untuk menghasilkan estimasi yang baik. Namun, seringkali asumsi ini

tidak terpenuhi karena (1) terdapat data ekstrem (*outlier*), (2) kondisi alamiah data yang tidak terdistribusi secara normal, melainkan mengikuti pola distribusi lainnya. Adapun untuk melakukan pendeteksian terhadap pelanggaran asumsi normalitas dapat menggunakan:

1. Mengecek grafik plot untuk residual
2. Menggunakan pengujian Shapiro-Wilk, Shapiro-Francia, Kolmogorov-Smirnov, Anderson-Darling, dengan hipotesis:

H_0 : error terdistribusi normal

H_1 : error tidak terdistribusi normal

Hipotesis nol akan ditolak bila $(\text{Prob}>z) < \alpha$

Pengujian dapat dilakukan dengan sintaks:

swilk nama_variabel (*shapiro-wilk*)

sfrancia nama_variabel (*shapiro-francia*)

ksmirnov nama_variabel (*Kolmogorov-Smirnov*)

Misalnya dalam uji ini seperti berikut:

`sfrancia ROA NPL CAR GCG`

```
sfrancia ROA NPL NIM CAR GCG
```

Shapiro-Francia W' test for normal data						
Variable	Obs	W'	V'	z	Prob>z	
ROA	152	0.88962	14.240	5.395	0.00001	
NPL	152	0.82902	22.056	6.283	0.00001	
NIM	152	0.54927	58.144	8.252	0.00001	
CAR	152	0.87010	16.757	5.725	0.00001	
GCG	152	0.93534	8.341	4.308	0.00001	

Gambar 3.8 Contoh hasil pengujian Shapiro Francia

Berdasarkan hasil uji normalitas Shapiro-Francia diatas dapat disimpulkan bahwa data tidak normal untuk ke lima variabel yaitu ROA, NPL, NIM, CAR, dan

GCG dengan nilai probabilitas yang dihasilkan < 0.05 (Catatan: data yang normal seharusnya menghasilkan nilai probabilitas > 0.05).

Solusi, jika terdapat indikasi pelanggaran normalitas, maka beberapa teknik/pelakuan yang dapat dilakukan adalah:

1. Menghapus atau *trimming* data yang dianggap ekstrem pada errornya.
2. Menggunakan transformasi data de dalam beberapa bentuk lain seperti berikut (Baca juga Johnson dan Dinardo, 1997, p.45-47)

Tabel 3.1 Bentuk Transformasi Data

Bentuk Transformasi	Data Transformasi
Log Transformation	$\ln(X)$
Square Root Transformation	$\text{Sqrt}(X)$
Reciprocal Transformation	$1/(X)$
Reverse Score Transformation	-

Misalkan kita mengubah variabel ROA, NPL, NIM, CAR dan GCG ke dalam bentuk akar kuadrat, maka command yang harus dilakukan adalah

$$g \text{ sqROA} = \text{sqrt}(\text{ROA})$$

$$g \text{ sqNPL} = \text{sqrt}(\text{NPL})$$

$$g \text{ sqNIM} = \text{sqrt}(\text{NIM}) \quad \text{dan seterusnya}$$

Menggunakan metode estimasi lainnya seperti regresi nonparametrik atau *bootstrapping*.

e. Hal Lain Terkait Pengujian

Terkadang, suatu regresi dilakukan untuk memperoleh suatu nilai error untuk digunakan dalam model penelitian lainnya. Untuk memperoleh nilai error

dari suatu estimasi dapat menggunakan perintah **predict**. Sintaks lengkap untuk perintah ini:

predict nama_variabel_baru, opsi

predict res, r

maka, akan ada variabel baru di layar data editor yaitu “res” seperti hasil sebagai berikut:

Tabel 3.2 data hasil penambahan variabel residual

ROA	NPL	NIM	CAR	GCG	BANK	res
3.14	0.01	0.16	15.80	4.65	BBCA	1.060362
3.17	0.01	0.09	15.30	4.65	BBCA	1.529249
3.28	0.01	0.05	13.50	5.00	BBCA	1.693147
3.57	0.01	0.06	12.83	4.92	BBCA	2.063627
2.44	0.00	0.06	14.20	5.00	BBCA	.8109925
0.04	0.00	0.06	9.24	5.00	BBCA	-1.357779
0.04	0.01	0.07	9.57	5.00	BBCA	-1.402656
0.04	0.00	0.07	9.66	5.00	BBCA	-1.396706
1.69	0.05	0.05	11.20	4.47	BBKP	.4246621
1.40	0.03	0.04	14.36	4.30	BBKP	.2292366
1.40	0.03	0.05	13.82	4.50	BBKP	.0571714
1.64	0.11	0.05	14.33	3.50	BBKP	.6221375
1.29	0.03	0.05	18.45	4.00	BBKP	.108155
0.02	0.02	0.04	8.81	4.00	BBKP	-.5750588
-0.02	0.05	0.10	7.82	4.00	BBKP	-1.034823
0.01	0.02	0.04	6.98	4.00	BBKP	-.4592064
0.96	0.05	0.10	13.50	3.95	BBNI	-.3904142
1.51	0.06	0.05	13.80	4.55	BBNI	-.0627137
2.21	0.01	0.06	18.60	4.75	BBNI	.4765876
2.49	0.01	0.06	17.60	4.70	BBNI	.8700218
2.67	0.01	0.06	16.70	4.00	BBNI	1.607889
0.03	0.01	0.06	9.44	4.00	BBNI	-.6377522
0.03	0.00	0.06	10.40	4.00	BBNI	-.6817046

Opsi yang dapat dipilih pada perintah *predict* untuk memunculkan nilai prediksi pada suatu variabel baru termasuk di dalamnya:

- a. Xb
- b. Stdp
- c. Stddp

- d. Score
- e. Scores
- f. equation

Untuk keterangan lebih lengkap dapat dilihat pada menu *help* pada STATA.

BAB IV

ANALISIS REGRESI DENGAN DATA PANEL

4.1. Analisis Menggunakan Data Panel

Sebelumnya telah dijelaskan secara singkat mengenai data panel yang merupakan gabungan data dengan beberapa data *cross section* dan beberapa data *time series*. Dapat dinyatakan pula bahwa analisis data panel merupakan observasi yang mengamati beberapa individu untuk kurun waktu tertentu. Sehingga model data panel secara matematis adalah

$$Y_{i,t} = \beta_0 + \beta_1 X_{i,t} + \dots + \beta_n X_{i,t} + \varepsilon_{i,t}$$

Dengan:

Y : variabel terikat

X : variabel bebas

β : parameter

i : data *cross section*

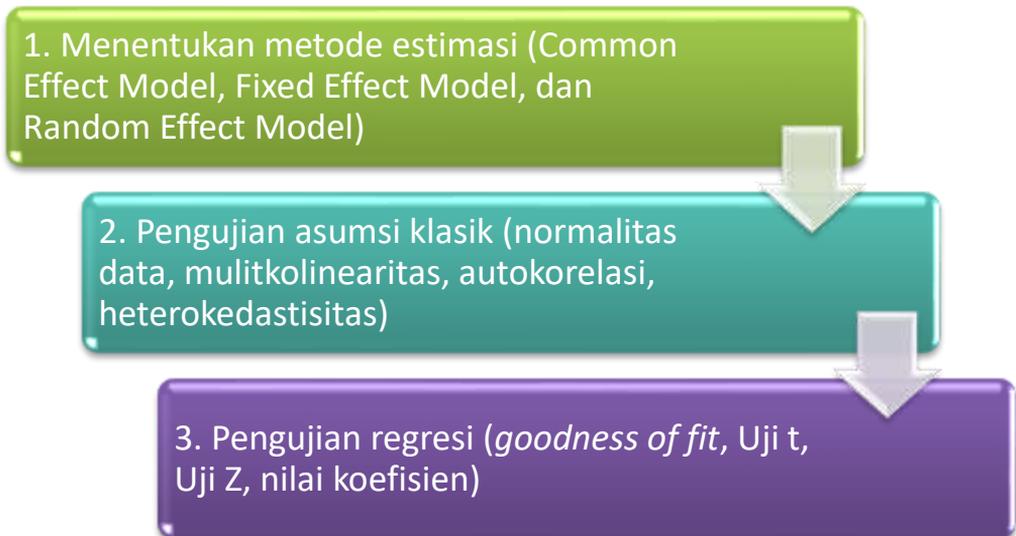
t : data timer series

ε : error term

Penelitian ekonomi, akuntansi dan keuangan banyak menggunakan data panel sehingga memahami analisis menggunakan data panel sangatlah berguna untuk memperkaya alat analisis penelitian. Adapun keunggulan data panel dibandingkan penelitian menggunakan jenis data lainnya antara lain:

1. Menyediakan jumlah observasi yang lebih besar sehingga meningkatkan *degree of freedom*.
2. Memungkinkan memperoleh variasi data yang lebih banyak sehingga diharapkan mengurangi kasus multikolinearitas.
3. Menyediakan informasi yang lebih kaya untuk tujuan analisis fenomena yang terjadi pada populasi.
4. Mengontrol variabel yang tidak dapat diobservasi/diukur.

Pada analisis regresi bab sebelumnya, tahapan analisis cukup dilakukan melalui tahapan uji asumsi klasik dan analisis utama regresi. Namun, untuk analisis menggunakan data panel, perlu dilakukan satu tahapan tambahan untuk menentukan model estimasi panel yang sebaiknya digunakan. Gambar 4.1. memberikan gambaran singkat tahapan estimasi menggunakan data panel



Gambar 4.1. Alur analisis data panel

Tahap pertama yang perlu dilakukan adalah menentukan metode estimasi data panel. Terdapat tiga model estimasi yang dapat digunakan yakni:

1. *Common Effect Model (CEM)*, merupakan metode data panel paling sederhana yang hanya mengombinasikan data *cross section* dan data *time series*. Model ini tidak memperhatikan adanya perbedaan karakteristik dalam *cross section* maupun *time series*, sehingga diasumsikan bahwa perilaku data *cross section* sama dalam berbagai kurun waktu. Metode ini bisa menggunakan pendekatan *panel ordinary least square (POLS)* atau teknik kuadrat terkecil untuk mengestimasi model data panel. Bentuk umum persamaan CEM dalam Gujarati (2012) dapat ditulis seperti berikut:

$$Y_{i,t} = \alpha_0 + \beta_1 X_{i,t} + \varepsilon_{i,t}$$

Dimana:

Y_{it} : variabel terikat pada unit observasi ke-i dan waktu ke-t

X_{it} : variabel bebas pada unit observasi ke-i dan waktu ke-t

β : koefisien slope atau koefisien arah

α : intercept model regresi

ε_{it} : komponen error pada unit observasi ke-i dan waktu ke-t

2. *Fixed Effect Model (FEM)*, model ini mengasumsikan bahwa perbedaan antar-*cross section* diakomodasi oleh nilai konstanta (*intercept*). Bila menggunakan metode ini, teknik estimasinya bisa dengan *least square dummy variable (LSDV)*, dimana variabel dummy disini yang akan menangkap perbedaan konstanta antar-*cross section*. Dummy variabel yang dimaksud adalah unit *cross section* menjadi variabel dalam model. Sehingga persamaan dari model ini dalam Gujarati (2012) adalah seperti berikut

$$Y_{i,t} = (\beta_0 + \alpha_i) + \beta_1 X_{1i,t} + \beta_2 X_{2i,t} + \varepsilon_{i,t}$$

Model ini masih banyak memiliki kekurangan yaitu kekurangan derajat kebebasan (*degree of freedom*) akibat jumlah sample yang terbatas dan

adanya multikolinieritas yang diakibatkan oleh banyaknya variabel dummy yang diestimasi, sedangkan kemampuan etimasinya masih terbatas, terutama jika terdapat variabel dammy yang diestimasi, dan lagi kemungkinan korelasi diantara komponen residual spesifik (*cross section* dan *time series*).

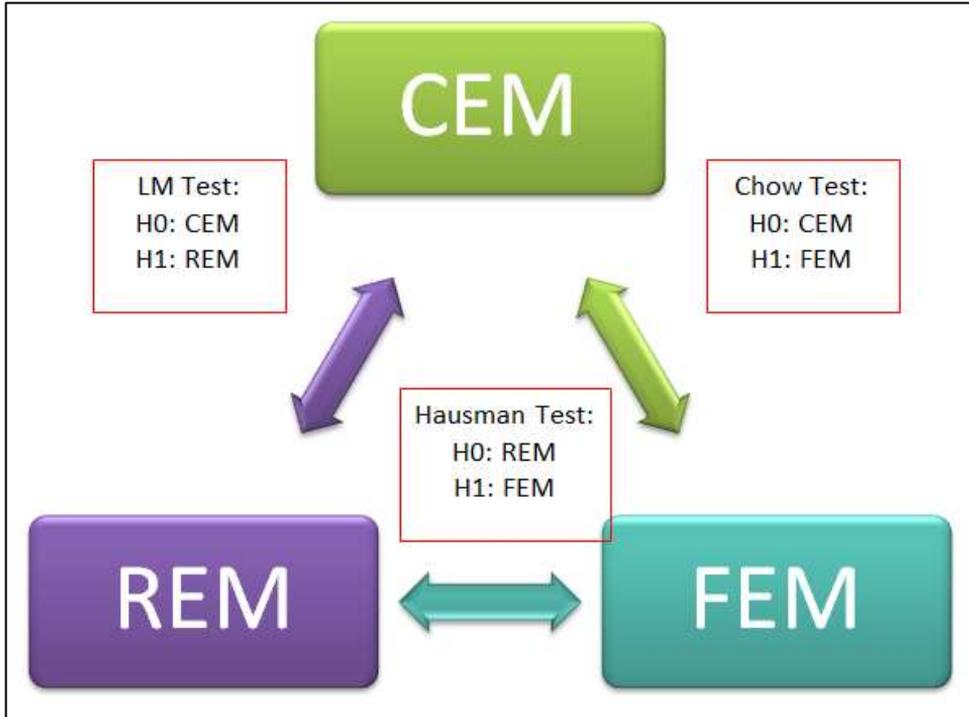
Namun, dalam praktiknya, metode estimasi untuk model FEM lebih banyak menggunakan *fixed effect estimator* dengan panel OLS. Maksudnya, peneliti tidak memilih rekomendasi estimasi LSDV tersebut, namun menggunakan *fixed effect estimator*. Hal tersebut dapat dilakukan karena dalam software yang biasa dipakai, misalnya STATA, EVIEWS, dan lainnya

3. *Random Effect Model (REM)*, model ini mengasumsikan bahwa error memiliki hubungan antarwaktu dan antar-*cross section*. Oleh karena itu, hasil estimasi menggunakan REM akan menyesuaikan nilai konstanta (*intercept*) dengan error setiap *cross section*. Persamaan model ini yang ditulis dalam Gujarati (2012) adalah sebagai berikut:

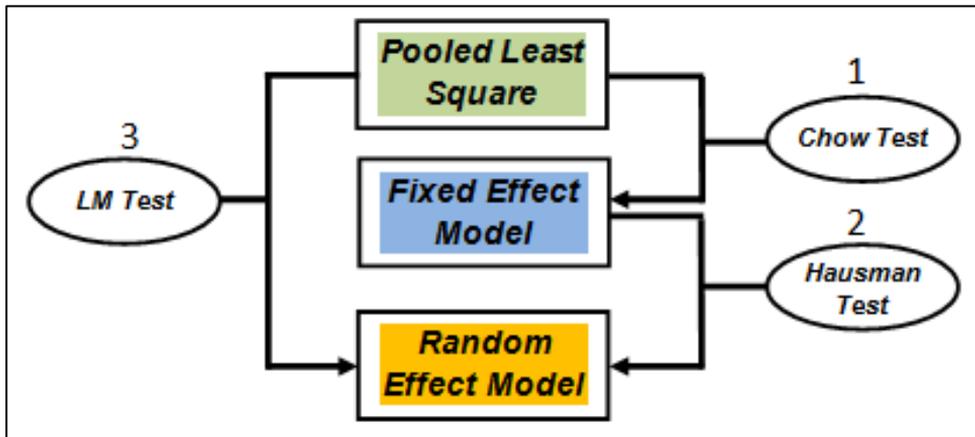
$$Y_{i,t} = \beta_0 + \beta_1 X_{1i,t} + \beta_2 X_{2i,t} + \beta_3 X_{3i,t} \dots \dots \dots + \beta_n X_{ni,t} + \varepsilon_{i,t}$$

Dimana ε_{it} adalah gangguan (*error term*) Yang merupakan gabungan dari *time series* dan *cross section*. Untuk melihat apakah model yang digunakan adalah *Fixed Effect* atau *Random Effects* maka harus dilakukan uji *Correlated Random Effects-Hausman Test*.

Berikut adalah tahapan menentukan model estimasi yang tepat dilakukan dengan membandingkan hasil tiga pengujian seperti yang digambarkan pada Gambar 4.2. di bawah ini:



Gambar 4.2. Skema uji pemilihan model estimasi data panel



Sumber: Gujarati (2012)

Gambar 4.3. Konsep Pemilihan Model Estimasi Data Panel

a. Chow Test

Pengujian Chow digunakan untuk menentukan apakah suatu estimasi sebaiknya menggunakan model *Fixed Effect Model* dibandingkan model *Common Effect Model*, dengan menerapkan hipotesis sebagai berikut:

H_0 : Pilih *Common Effect Model*

H_1 : Pilih *Fixed Effect Model*

Untuk menerima atau menolak hipotesis tersebut, maka dilakukan perbandingan antara perhitungan F-tabel dan F-statistik, jika F-statistik lebih besar dari F-tabel maka H_0 ditolak yang berarti model yang paling cocok diterapkan adalah model *fixed effects*, dan jika F-statistik lebih kecil maka model yang paling cocok yang diterapkan adalah model CEM
 $F\text{-statistik} > F\text{-tabel} = H_0$ ditolak

$F\text{-statistik} < F\text{-tabel} = H_0 =$ diterima

b. Hausman Test

Pengujian Hausman digunakan untuk menentukan model estimasi mana yang sebaiknya digunakan antara model *Random Effect Model* atau *Fixed Effect Model*. Uji ini dilakukan ketika hasil yang ditunjukkan oleh Uji Chow Model *Fixed Effects* lebih baik. Uji hausman menerapkan hipotesis sebagai berikut:

H_0 : Pilih *Random Effect Model*

H_1 : Pilih *Fixed Effect Model*

Untuk menerima atau menolak hipotesis tersebut didasarkan pada perbandingan nilai *Prob>chi2* dengan nilai *alpha* (0.05) dari hasil uji hausman. Jika nilai *Prob>chi2* lebih besar (>) dari nilai *alpha* (0.05) maka model *random effect* (H_0) diterima, dan jika *Prob>chi2* lebih kecil (<) dari

alpha (0.05) maka H_0 di tolak, yang berarti persamaan penelitian cenderung ke kategori model *fixed effect* (Gujarati, Damodar N 2012)

c. Lagrange Multiplier (LM) Test

Uji Lagrange Multiplier dilakukan untuk menentukan model persamaan penelitian masuk pada kategori model *random effect* ataukah model *common effect*. Ketentuan hasil uji ini berdasarkan pada hipotesis sebagai berikut

H_0 : Pilih *common effect* atau hipotesis dapat dinyatakan

H_0 : $\text{Prob} > \text{chibar}2 > \alpha$ (0.05) adalah *common effect*

H_1 : Pilih *random effect* atau dapat dinyatakan

H_1 : $\text{Prob} > \text{chibar}2 < \alpha$ (0.05) adalah *random effect*

Dengan hipotesis tersebut dapat dijelaskan bahwa jika nilai p-value atau nilai $\text{Prob} > \text{chibar}2$ lebih besar ($>$) dari nilai α (0.05) maka H_0 diterima, sebaliknya jika $\text{Prob} > \text{chibar}2$ lebih kecil dari nilai α , maka persamaan penelitian cenderung ke model random effects.

4.2. Analisis Regresi Data Panel Menggunakan STATA

Di sub bab ini, STATA memandu pengguna untuk melakukan analisis regresi atas data panel. Hal pertama yang harus dilakukan adalah menentukan variabel *cross section* dan variabel waktu dengan menggunakan sintaks sebagai berikut:

xtset cross_section waktu, opsi

keterangan:

cross_section: pilihan variabel yang menjadi indikator cross section

waktu: pilihan variabel yang menjadi indikator waktu, secara default STATA akan menentukan jarak antarwaktu pada interval 1.

Perlu diingat jika variabel *cross section* yang dimiliki masih berbentuk data non-numerik, maka harus dilakukan transformasi data menjadi data numerik menggunakan sintaks *encode* seperti yang sudah dijelaskan pada Bab 1.

Sebagai contoh, akan digunakan pengembangan skenario yang sebelumnya digunakan pada Bab 3.

Penelitian ini hendak melakukan pengujian atas pengaruh Net Performing Loan (NPL), Net Interest Margin (NIM), Capital Adequacy (CAR) dan Good Corporate Governance (GCG) terhadap Return On Asset (ROA)

Penjelasan lengkap untuk setiap variabel adalah sebagai berikut:

1. **YEAR:** tahun observasi meliputi tahun ke-1 hingga tahun ke-8 (2013 sd 2020) sebagai data time series
2. **BANK:** Perusahaan yang diobservasi sebagai data cross section
3. **ROA** (*Return On Asset*) adalah sebagai variabel dependen yaitu rasio untuk mengukur efektifitas bank dalam memperoleh keuntungan secara keseluruhan melalui pengoperasian total aset yang dimiliki bank. Dalam hal ini pengukurannya adalah rasio antara laba sebelum pajak dengan total aset
4. **NPL** (*Net Performing Loan*) merupakan rasio yang mencerminkan besarnya kredit bermasalah yang dihadapi bank. Semakin tinggi rasio ini maka akan semakin buruk kualitas kredit bank, yang menyebabkan jumlah kredit bermasalah semakin besar. Dalam hal ini pengukurannya adalah pembagian antara kredit bermasalah dengan total kredit.
5. **NIM** (*Net Interest Margin*) adalah rasio untuk mengukur kemampuan manajemen bank dalam mengelola aktiva produktifitasnya untuk menghasilkan pendapatan bunga bersih. Jadi semakin besar rasio ini maka

akan semakin besar earning uang diperoleh bank dari pendapatan bunga. Dalam hal ini pengukurannya adalah pembagian antara pendapatan bunga bersih dengan rata-rata total aset produktif

6. **CAR** (*Capital Adequacy Ratio*) adalah rasio yang memperlihatkan seberapa besar jumlah seluruh aktiva bank yang mengandung resiko ikut dibiayai dari modal sendiri. Dalam hal ini pengukurannya adalah pembagian antara modal dengan aktiva tertimbang menurut resiko.
7. **GCG** (*Good Corporate Governance*) merupakan sistem yang mengendalikan perusahaan untuk menciptakan nilai tambah untuk semua stakeholder. Dalam hal ini pengukurannya adalah dari data survey yang dilakukan lembaga Indonesian Institute for Corporate Governance (IICG)

Sehingga model lengkap yang digunakan adalah

$$ROA = \beta_0 + \beta_1 NPL + \beta_2 NIM + \beta_3 CAR + \beta_4 GCG$$

Gambar 4.3. menunjukkan hasil sintaks pengesetan data panel. Informasi yang perlu diperhatikan selain indikator panel dan waktu, adalah informasi mengenai balance atau tidaknya data set yang dimiliki. STATA dapat secara langsung menentukan apakah data set yang digunakan:

- a. *strongly balanced*, yakni seluruh cross section memiliki jumlah observasi yang sama antarwaktu
- b. *unbalanced*, yakni cross section memiliki jumlah observasi yang berbeda antarwaktu, misalkan data cross section A memiliki observasi selama 5 tahun, sedangkan cross section B hanya memiliki observasi selama 3 tahun.

```
xtset BANK TAHUN
      panel variable:  BANK (strongly balanced)
      time variable:  TAHUN, 2013 to 2020
                   delta:  1 unit
```

Gambar 4.4. Hasil pengaturan data panel

Latihan: Lakukan pengaturan data panel seperti contoh di atas!

Kemudian untuk mengetahui karakteristik data, dapat dilakukan dahulu analisis deskriptif seperti yang telah dicontohkan pada Bab 3. Sebagai contoh dapat digunakan sintaks

xtsum nama_variabel

```
xtsum ROA NPL NIM CAR GCG
```

Variable		Mean	Std. Dev.	Min	Max	Observations
ROA	overall	1.239812	1.219919	-.69	4.45	N = 152
	between	.5147514	.3615589	2.355005		n = 19
	within	1.111538	-1.078202	5.103178		T = 8
NPL	overall	.0190807	.0173788	0	.1052433	N = 152
	between	.0094277	.0051305	.0419328		n = 19
	within	.0147397	-.0094405	.0826362		T = 8
NIM	overall	.0696934	.0504637	.0024	.546	N = 152
	between	.0236979	.0494409	.122994		n = 19
	within	.0448445	-.0104395	.4985605		T = 8
CAR	overall	14.52688	6.013934	0	46.49	N = 152
	between	2.849022	10.35198	21.38811		n = 19
	within	5.331675	1.797206	40.28559		T = 8
GCG	overall	4.233257	.491731	2.33	5	N = 152
	between	.350123	3.35	4.9025		n = 19
	within	.3534064	3.088257	5.020757		T = 8

Gambar 4.5. Hasil Pengujian xtsum

Latihan: Tunjukkan hasil analisis deskriptif untuk data set yang digunakan!

4.2.1. Menentukan Model Estimasi Data Panel pada STATA

Setelah melakukan pengaturan dan penentuan variabel yang ditentukan sebagai indikator data panel dan indikator waktu, selanjutnya dilakukan pengujian untuk menentukan model estimasi yang sebaiknya digunakan.

1. Lakukan pengujian regresi *common effect, model*, seperti contoh berikut:

```
reg ROA NPL NIM CAR GCG
```

Source	SS	df	MS	Number of obs	=	152
Model	47.9072824	4	11.9768206	F(4, 147)	=	9.96
Residual	176.811404	147	1.20279867	Prob > F	=	0.0000
				R-squared	=	0.2132
				Adj R-squared	=	0.1918
Total	224.718686	151	1.48820322	Root MSE	=	1.0967

ROA	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
NPL	5.491708	5.223975	1.05	0.295	-4.832086 15.8155
NIM	5.79292	1.784418	3.25	0.001	2.266495 9.319345
CAR	.0544075	.0150772	3.61	0.000	.0246114 .0842037
GCG	.7422607	.1859474	3.99	0.000	.3747854 1.109736
_cons	-3.201254	.8709542	-3.68	0.000	-4.922462 -1.480045

```
estimates store CEM
```

Gambar 4.6. Hasil Pengujian CEM

Pengujian *pooled least square* ini menggunakan sintaks yang sama seperti yang telah dipelajari pada Bab 3. Untuk menyimpan hasil pengujian ini dapat menggunakan sintaks ***estimates store CEM***.

2. Lakukan pengujian menggunakan *fixed effect model*. Berikut contoh hasil pengujiannya:

```

xtreg ROA NPL NIM CAR GCG, fe
Fixed-effects (within) regression      Number of obs   =      152
Group variable: BANK                  Number of groups =      19

R-sq:                                  Obs per group:
  within = 0.1446                       min =           8
  between = 0.3927                       avg =          8.0
  overall = 0.1852                       max =           8

corr(u_i, Xb) = 0.0737                  F(4,129)        =      5.45
                                          Prob > F         =      0.0004
-----+-----
      ROA |      Coef.   Std. Err.    t    P>|t|    [95% Conf. Interval]
-----+-----
      NPL |   8.108763   6.170146    1.31  0.191   -4.099023   20.31655
      NIM |   5.705563   2.0302     2.81  0.006    1.688764    9.722363
      CAR |   .0591847   .0171065    3.46  0.001    .025339    .0930304
      GCG |   .3574793   .2574855    1.39  0.167   -.1519621    .8669207
      _cons | -1.685619   1.14068    -1.48  0.142   -3.942482    .5712434
-----+-----
      sigma_u |   .40520004
      sigma_e |   1.1122503
      rho    |   .11716871   (fraction of variance due to u_i)
-----+-----
F test that all u_i=0: F(18, 129) = 0.77          Prob > F = 0.7274
estimates store FEM

```

Gambar 4.7. Hasil pengujian model *fixed effect*

Sintaks yang digunakan untuk estimasi regresi panel secara *default* akan mengikuti model random. Jika hendak melakukan regresi *fixed effect*, cukup menambahkan opsi, *fe* setelah memasukan sintaks variabel. Untuk menyimpan hasil pengujian ini dapat menggunakan sintaks ***estimates store fe***.

3. Lakukan pengujian menggunakan *random effect model*. Berikut contoh hasil pengujiannya

```

xtreg ROA NPL NIM CAR GCG, re

Random-effects GLS regression           Number of obs   =       152
Group variable: BANK                   Number of groups =        19

R-sq:                                  Obs per group:
  within = 0.1310                       min =           8
  between = 0.6645                      avg =          8.0
  overall = 0.2132                      max =           8

Wald chi2(4) =       39.83
Prob > chi2   =       0.0000

corr(u_i, X) = 0 (assumed)

-----+-----
      ROA |      Coef.   Std. Err.   z   P>|z|   [95% Conf. Interval]
-----+-----
      NPL |   5.491708   5.223975   1.05  0.293   -4.747096   15.73051
      NIM |   5.79292    1.784418   3.25  0.001    2.295526    9.290314
      CAR |   .0544075   .0150772   3.61  0.000    .0248567    .0839584
      GCG |   .7422607   .1859474   3.99  0.000    .3778106    1.106711
      _cons | -3.201254   .8709542  -3.68  0.000   -4.908292   -1.494215
-----+-----

      sigma_u |           0
      sigma_e |   1.1122503
      rho     |           0   (fraction of variance due to u_i)
-----+-----

estimates store REM

```

Gambar 4.8. Hasil pengujian model *random effect*

Seperti yang disebutkan sebelumnya bahwa sintaks yang digunakan untuk estimasi regresi panel secara *default* akan mengikuti model random. Untuk menyimpan hasil pengujian ini dapat menggunakan sintaks ***estimates store re.***

4. Untuk membandingkan hasil tiga pengujian di atas dapat menggunakan sintaks berikut

estimates table ols fe re, star stats(N r2 r2_a)

Menggunakan sintaks ini akan dihasilkan tabel perbandingan atas statistika yang diminta yakni N (jumlah sampel), r2 dan r2_a (nilai *goodness of fit*),

serta koefisien setiap variabel beserta identifikasi signifikansinya. Berikut contohnya

```
estimates table CEM FEM REM, star stats(N r2)
```

Variable	CEM	FEM	REM
NPL	5.4917082	8.1087634	5.4917082
NIM	5.79292**	5.7055633**	5.79292**
CAR	.05440753***	.05918472***	.05440753***
GCG	.74226074***	.35747927	.74226074***
_cons	-3.2012535***	-1.6856194	-3.2012535***
N	152	152	152
r2	.2131878	.14460069	

legend: * p<0.05; ** p<0.01; *** p<0.001

Gambar 4.9. Perbandingan hasil tiga model estimasi

5. Lakukan pengujian sesuai Gambar 4.2 atau Gambar 4.3

- Uji Chow dilakukan dengan melihat Prob > F pada hasil pengujian *fixed effect model*
- Uji LM dilakukan dengan menggunakan sintaks

xttest0

contoh hasil pengujian LM:

```
xttest0
```

Breusch and Pagan Lagrangian multiplier test for random effects

$$ROA[BANK, t] = Xb + u[BANK] + e[BANK, t]$$

Estimated results:

	Var	sd = sqrt(Var)
ROA	1.488203	1.219919
e	1.237101	1.11225
u	0	0

Test: Var(u) = 0

chibar2(01) = 0.00
 Prob > chibar2 = 1.0000

Gambar 4.10. Hasil pengujian LM

- Uji hausman dilakukan dengan membandingkan hasil pengujian menggunakan model *fixed effect* dan *random effect*. Pengujian dapat dilakukan sebagai berikut

```
quietly xtreg nama_variabel, fe
estimates store FEM
quietly xtreg nama_variabel, re
estimates store REM
```

Sintaks **quietly** digunakan untuk melakukan suatu perintah estimasi tanpa menunjukkan hasil estimasi pada *Panel Results*. Sementara pengujian hausman menggunakan perintah **hausman FEM REM** Berikut contoh hasil pengujian hausman:

```
quietly xtreg ROA NPL NIM CAR GCG, fe
estimates store FEM
quietly xtreg ROA NPL NIM CAR GCG, re
estimates store REM
hausman FEM REM
```

	---- Coefficients ----			
	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	FEM	REM	Difference	S.E.
NPL	8.108763	5.491708	2.617055	3.283411
NIM	5.705563	5.79292	-.0873567	.9682789
CAR	.0591847	.0544075	.0047772	.0080815
GCG	.3574793	.7422607	-.3847815	.1781078

```
-----
b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

      chi2(4) = (b-B)' [(V_b-V_B)^(-1)] (b-B)
            =      5.04
Prob>chi2 =      0.2830
```

Gambar 4.11. Hasil pengujian hausman

- Berdasarkan tiga pengujian di atas, dapat ditentukan model estimasi yang sebaiknya

- digunakan. Dalam kasus di atas, model **Random Effect** yang sebaiknya digunakan.

Latihan: Lakukan pengujian untuk menentukan model estimasi seperti contoh di atas!

4.2.2. Uji Asumsi Klasik Data Panel Menggunakan STATA

Setelah ditentukan model estimasi yang sebaiknya digunakan, selanjutnya dilakukan uji asumsi klasik yakni:

a. Multikolinearitas

- *Common Effect Model*, pengujian menggunakan mekanisme yang sama seperti yang dijelaskan pada Bab 3.
- *Fixed Effect Model*, dapat menggunakan **vif, uncentered**, berikut contoh hasil pengujiannya

```
vif, uncentered
```

Variable	VIF	1/VIF
GCG	8.54	0.117055
CAR	6.05	0.165269
NIM	2.96	0.337278
NPL	2.11	0.474097
Mean VIF	4.92	

Gambar 4.12. Hasil pengujian vif untuk Fixed Effect

- *Random Effect*, dapat menggunakan **vif, uncentered**

b. Heterokedastisitas

- *Comon Effect Model*, pengujian menggunakan mekanisme yang sama seperti yang dijelaskan pada Bab 3.
- *Fixed Effect Model*, sintaks **xttest3**, berikut contoh hasil pengujiannya

```
xttest3

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (19) =      2280.47
Prob>chi2 =      0.0000
```

Gambar 4.13. Hasil pengujian xttest3 untuk Fixed Effect

- *Random Effect*, jika menggunakan model ini asumsi ini tidak dilanggar karena pengujian akan menggunakan *generalized least square (GLS)*.

c. Autokorelasi

- *Common Effect Model*, pengujian menggunakan mekanisme yang sama seperti yang dijelaskan pada Bab 3.
- *Fixed Effect Model*, sintaks **xtserial nama_variabel**, berikut contoh hasil pengujiannya

```
xtserial ROA NPL NIM CAR GCG

Wooldridge test for autocorrelation in panel data
H0: no first-order autocorrelation

F( 1,      18) =      49.303
Prob > F =      0.0000
```

Gambar 4.14 Hasil pengujian xtserial untuk Fixed Effect

- *Random Effect Model*, jika menggunakan model ini asumsi ini tidak dilanggar karena pengujian akan menggunakan *generalized least square (GLS)*.

Latihan: Lakukan pengujian asumsi klasik untuk menentukan model estimasi seperti contoh di atas dengan menggunakan file: *data_keuangan.xls!*

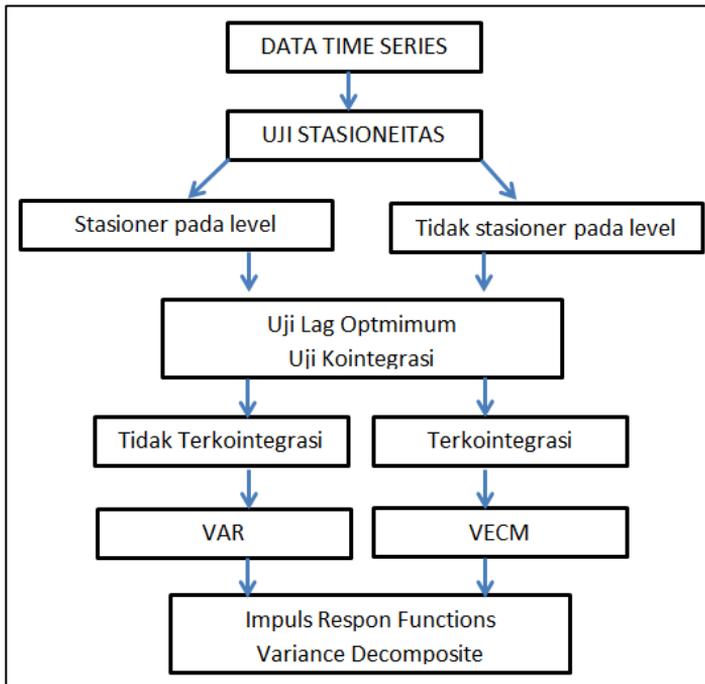
BAB V

VECTOR ERROR CORRECTION MODEL (VECM)

5.1. Tahapan Model *Vector Error Correction* (VECM).

Menurut Green (2000), Model *Error Correction* merupakan model regresi non-linier, namun dalam kenyataannya model ini merupakan linier secara intrinsik dan dapat memberikan kesimpulan singkat dari persamaan yang tertutup secara langsung. VECM digunakan untuk mengidentifikasi keseimbangan pada jangka pendek dan jangka panjang. Menurut Aija et, al (2011), model VECM dapat digunakan jika suatu data time series dalam model VAR telah terbukti memiliki hubungan kointegrasi.

Langkah-langkah pengujian estimasi VECM yaitu pertama; uji stasioneritas dengan menggunakan *Augmented Dicky Fuller*, kedua pencarian *Lag Optimal* yang didahului oleh persamaan uji VAR, dan ke tiga uji kointegrasi dengan menggunakan *Johansen Test*, selanjutnya uji model VECM. Berikut adalah tahapan model VAR dan VECM yang dijelaskan pada gambar 5.1.



Gambar 5.1. Tahapan model VAR – VECM. Sumber: www.diasatria.com

Dalam kasus ini kita menggunakan data dalam CD dengan nama (data_time series.exls) yaitu kita mencoba melihat pengaruh dari Pendapatan Bersih Pribadi (Yd), Kekayaan Riil (Wealth), dan Tingkat Suku Bunga Tahunan di AS (interest Rate) terhadap Pengeluaran Konsumsi (C) di Amerika Serikat (www.diasatria.com). Persamaan dalam estimasi ini sebagai berikut:

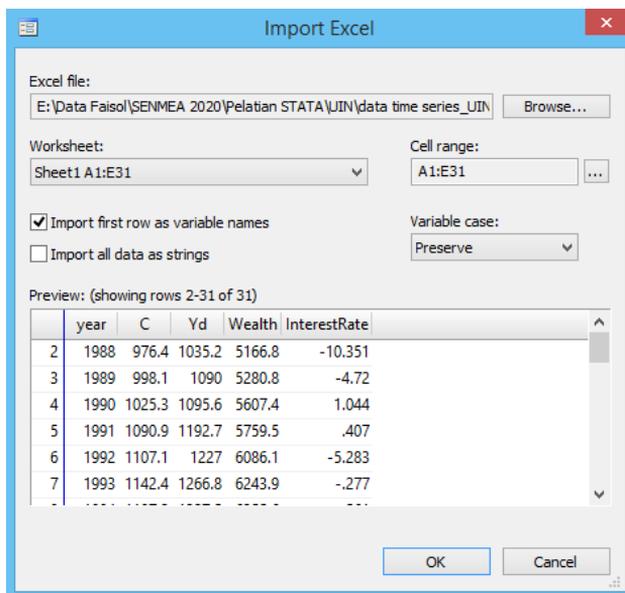
$$C_t = \alpha_1 + \sum \beta_1 t . \sum Y_d t - 1 + \sum \beta_1 t . \text{Wealth} t - 1 + \sum \beta_1 t . \text{Interest Rate } t - 1 + \epsilon_t$$

Dimana

C adalah pengeluaran konsumsi riil, dalam miliar dollar; Yd adalah pendapatan bersih riil, dalam miliar dollar; Wealth adalah kekayaan riil, dalam miliar dollar; Interest rate adalah tingkat suku bunga tahunan AS, dalam persen; ϵ adalah error term dan t adalah time series.

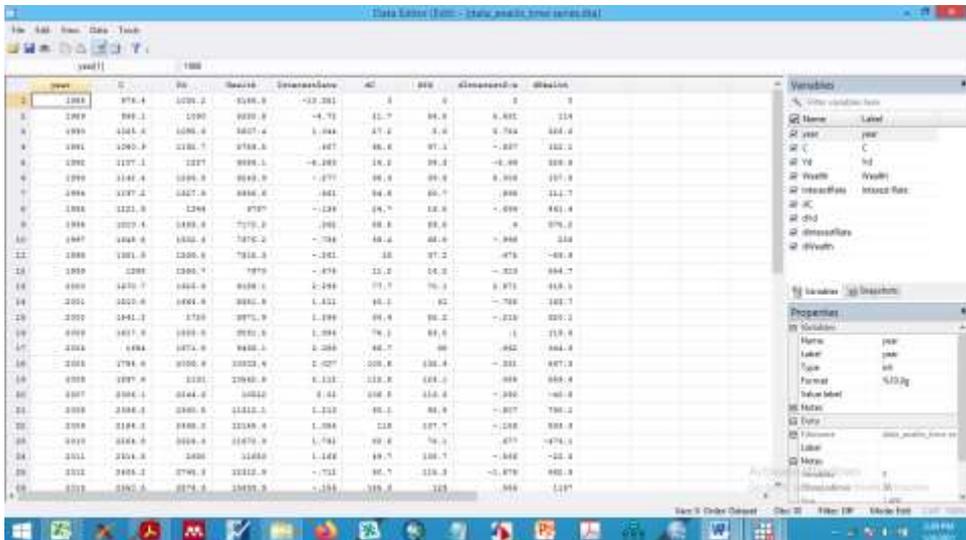
Selanjutnya, tahapan-tahapannya adalah

1. Pertama, copy semua data dari Ms. Excel, lalu masukan ke data editor di STATA atau dengan cara mengimport dari menu STATA dari klik menu file pilih import kemudian pilih *excel spreadsheet (*.xls)*, lalu muncul menu *import excel*, (gambar 5.2) pilih browse untuk melacak dimana file data time series disimpan, lalu klik centang pada indikatornya box pada *import first row as variable names*, selanjutnya klik OK. Seperti gambar berikut:



Gambar 5.2. Import data dari excel

Untuk mengamati dan mengecek data yang digunakan estimasi, sebagai sudah dijelaskan di bab sebelumnya, yaitu dengan mengklik ikon Data Editor (Edit) yaitu seperti gambar berikut:



Gambar 5.3. Data Editor

2. Sebelum melakukan uji apapun pada STATA, dalam estimasi ini kita harus melakukan setting waktu, dengan perintah sintak seperti *tsset year*
3. Jika data analisis sudah masuk di Data Editor dalam warna hitam, menunjukkan data tersebut sudah siap digunakan analisis, jika berwarna mewarna merah menunjukkan STATA tidak bisa membaca data analisis, maka kita perlu mengedit dengan cara memperbaiki data di excel, sebagai sudah dibahas di bab sebelumnya. Contoh; tanda coma bisa di edit dengan tanda titik seperti angka 97,1, 34,3 dan 39, 8, angka tersebut harus diedit dengan tanda titik bukan tanda koma, menjadi 97.1, 34.3 dan 39.8, selanjut dapat melanjutkan tahapan analisis uji stasioneritas.

5.2. UJI STATIONER

Satu satu tahapan penting yang harus dilakukan dalam analisa dengan data *times series* adalah kondisi data yang stasioner atau tidak stasioner. Jika

estimasi dilakukan dengan menggunakan data yang tidak stasioner maka akan menghasilkan regresi yang palsu atau disebut sebagai *spurious regression* (Gujarati, 2004). *Spurious regression* memiliki pengertian bahwa hasil regresi dari satu variabel time series atau beberapa variabel *time series* lainnya cenderung menghasilkan kesimpulan estimasi yang bias yang ditunjukkan dengan karakteristik seperti hasil R^2 yang sangat tinggi (lebih besar dari 0,9) namun pada kenyataannya hubungan antara variable dalam model tidak mempunyai arti. Gujarati menjelaskan bahwa jika $R^2 > d$ (durbin watson statistik), maka kondisi ini merupakan *rule of thumb* yang baik untuk menduga bahwa hasil estimasi kemungkinan besar merupakan *nonsense/spurious regression*.

Ada beberapa metode yang dapat digunakan untuk uji stasioneritas. Pada umumnya, metode yang dapat digunakan adalah metode pengujian *Augmented Dicky Fuller* dan metode pengujian *Philip-Peron* (Gujarati, 2004). Dalam sub bab ini, Uji stasioneritas yang digunakan adalah dengan *Augmented Dicky Fuller*. Uji stasioneritas ini digunakan untuk melihat apakah data analisis yang digunakan bersifat stasioner atau non-stasioner, dimana apabila data yang digunakan bersifat non-stasioner, maka akan dilakukan proses diferensiasi hingga data tersebut pada akhirnya bersifat stasioner (catatan: maksimal diferensi pada derajat kedua (I_2)). Dan jika data yang digunakan bersifat stasioner pada derajat level, maka model estimasi yang akan digunakan adalah model VAR. Maka, untuk mendapatkan kejelasan stasioneritas data adalah dengan perintah STATA sebagai berikut:

Ketikkan sintak *dfuller* diikuti nama variabel, misalnya

dfuller C

```
dfuller C
Dickey-Fuller test for unit root           Number of obs =      29
----- Interpolated Dickey-Fuller -----
          Test          1% Critical      5% Critical      10% Critical
          Statistic      Value          Value          Value
-----
Z(t)          2.779          -3.723          -2.989          -2.625
-----
MacKinnon approximate p-value for Z(t) = 1.0000
```

Gambar 5.4. Hasil uji *dfuller C*

Hasil uji stasioner dengan dfuller diatas untuk variabel Consumption (C). Berdasarkan uji dfuller menunjukkan bahwa data diatas tidak stasioner, yaitu dapat dilihat dari perbandingan *Test Statistic* (2.779) lebih kecil dari nilai *Critical Value* (-2.989) dengan menggunakan tingkat taraf signifikansi 5% dimana hasilnya harus dibawah 5%, dari hasil uji diatas nilai *p-value* atau probabilitasnya diatas 5% yaitu 1.000. Selanjutnya, untuk mengetahui kestasioneritas variabel lainnya dapat dilakukan dengan cara yang sama.

Berdasarkan hasil uji stasioneritas diatas, menunjukkan data tidak stasioner pada tingkat level, maka harus dilakukan stasioneritas pada tingkat different pertama. Sintak yang digunakan untuk mendiferensikan variabel yang tidak stasioner adalah sebagai berikut:

Untuk varibale C (pengeluaran konsumsi riil)

Ketikkan sintak : gen dC=d.C

Kemudian ketik lagi : replace dC=0 if dC==.

Kemudian dfuller dC

dC adalah variabel yang sudah di differnsiasi

berikut adalah hasil uji dfuller untuk variabel dC

dfuller dC

Dickey-Fuller test for unit root		Number of obs = 29		
Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value	
Z(t)	-3.738	-3.723	-2.989	-2.625
MacKinnon approximate p-value for Z(t) = 0.0036				

Gambar 5.5 Uji dfuller variabel dC

Untuk variabel Yd (pendapatan bersih riil)

Ketikkan sintak : gen dYd=d.Yd

Kemudian ketik lagi : replace dYd=0 if dYd==.

Kemudian dfuller dYd

```

dfuller dYd
Dickey-Fuller test for unit root                Number of obs =      29
----- Interpolated Dickey-Fuller -----
      Test          1% Critical      5% Critical      10% Critical
      Statistic      Value           Value           Value
-----
Z(t)          -4.438           -3.723           -2.989           -2.625
MacKinnon approximate p-value for Z(t) = 0.0003
    
```

Gambar 5.6 Uji dfuller variabel Yd

Dari hasil uji dfuller untuk variabel dC dan Yd, kita dapat melihat bahwa nilai Test Statistic dari kedua variabel tersebut lebih besar dari pada Critical Value 5%, maka dapat dikatakan variabel dC dan dYd bersifat stasioner dengan nilai probabilitas (Prob*) atau *p-value* berada dibawah 0.05. Selanjutnya, tahapan ini harus dilakukan pada setiap variabel yang digunakan dalam penelitian. Dalam hal ini semua variabel dalam model bersifat stasioner pada tingkat diferensi pertama.

Tahapan selanjutnya adalah melakukan uji pencarian Lag Optimal, yaitu yang didahului oleh Uji VAR (diawali dengan meregresikan persamaan penelitian dengan uji VAR).

5.3. UJI LAG OPTIMAL (Penentuan Panjang Lag)

Salah satu kondisi yang harus dilakukan dalam estimasi dengan VAR/VECM adalah kondisi penentuan panjang lag akan digunakan. Permasalahan yang muncul apabila panjang lagnya terlalu kecil akan menghasilkan model tersebut tidak dapat digunakan karena kurang mampu menjelaskan hubungannya. Dan sebaliknya, jika panjang lag yang digunakan terlalu besar maka derajat kebebasannya (*degree of freedom*) akan menjadi lebih besar sehingga tidak efisien lagi dalam menjelaskan. Salah satu metode yang paling umum digunakan adalah dengan melihat Akaike Information Criterion (AIC). Dimana rumusnya adalah (Gujarati, 2004):

$$AIC = T \log |\Sigma| + 2N$$

Dimana $|\Sigma|$ adalah determinan dari matrik residual varians atau kovarian sedangkan N adalah jumlah total dari parameter yang diestimasi dalam

persamaan. Gujarati memberika pedoman bahwa untuk melihat nilai AIC, dimana nilai AIC terendah yang diperoleh dari hasil estimasi var dengan berbagai lag menunjukkan bahwa panjang lag tersebut yang paling baik digunakan, dalam STATA ditunjukkan dengan tanda bintang (*) pada *selection order creterio yaitu* LL, LR, FPE, AIC, HQIC dan SBIC

Langkah awal uji lag optimal adalah dengan perintah berikut:

Ketik var diikuti variabel contoh: *var C Yd Wealth InterestRate*

Setelah hasil uji Var muncul, ketik perintah varsoc, untuk pencarian Lag Optimal Berikut hasil uji var dan perintah varsoc

var C Yd Wealth InterestRate

Vector autoregression

Sample: 1990 - 2017	Number of obs	=	28
Log likelihood = -468.9655	AIC	=	36.06896
FPE = 5.97e+10	HQIC	=	36.59259
Det(Sigma_ml) = 4.15e+09	SBIC	=	37.7818

Equation	Parms	RMSE	R-sq	chi2	P>chi2
C	9	26.0149	0.9986	19465.23	0.0000
Yd	9	31.2455	0.9985	18476.62	0.0000
Wealth	9	465.931	0.9770	1189.205	0.0000
InterestRate	9	1.01144	0.7521	84.97055	0.0000

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
C					
C					
L1.	.7636493	.4068694	1.88	0.061	-.0338 1.561099
L2.	.1989704	.3810457	0.52	0.602	-.5478653 .9458062
Yd					
L1.	-.202235	.3230135	-0.63	0.531	-.8353299 .4308599
L2.	.2907767	.2791119	1.04	0.298	-.2562725 .837826
Wealth					
L1.	.073592	.0120914	6.09	0.000	.0498932 .0972908
L2.	-.0761775	.0156822	-4.86	0.000	-.1069141 -.0454408
InterestRate					
L1.	3.812169	3.467922	1.10	0.272	-2.984833 10.60917
L2.	-.0558081	2.066241	-0.03	0.978	-4.105566 3.993949
_cons	-4.046651	39.2463	-0.10	0.918	-80.96798 72.87468

```

-----+-----
Yd
      C
      L1. | 1.144351  .488676  2.34  0.019  .1865632  2.102138
      L2. | .3211977  .4576601  0.70  0.483  -.5757996  1.218195
      Yd
      L1. | -.2732959  .3879598  -0.70  0.481  -1.033683  .4870913
      L2. | .132831   .3352311  0.40  0.692  -.52421   .7898719
      Wealth
      L1. | .0580706  .0145226  4.00  0.000  .0296068  .0865344
      L2. | -.0744244  .0188354  -3.95  0.000  -.111341  -.0375077
      InterestRate
      L1. | 4.022643  4.165195  0.97  0.334  -4.14099  12.18628
      L2. | 1.166511  2.481687  0.47  0.638  -3.697505  6.030528
      _cons | -88.13236  47.1373  -1.87  0.062  -180.5198  4.255045
-----+-----
Wealth
      C
      L1. | -2.260268  7.287099  -0.31  0.756  -16.54272  12.02218
      L2. | 7.80559   6.824592  1.14  0.253  -5.570365  21.18155
      Yd
      L1. | -2.491683  5.785227  -0.43  0.667  -13.83052  8.847155
      L2. | -.5773265  4.998941  -0.12  0.908  -10.37507  9.220418
      Wealth
      L1. | 1.09008   .2165599  5.03  0.000  .6656307  1.51453
      L2. | -.5019349  .2808716  -1.79  0.074  -1.052433  .0485635
      InterestRate
      L1. | 42.69139  62.11108  0.69  0.492  -79.04408  164.4269
      L2. | 59.64735  37.00672  1.61  0.107  -12.88449  132.1792
      _cons | 723.8755  702.9078  1.03  0.303  -653.7985  2101.55
-----+-----
InterestRate
      C
      L1. | -.0225081  .0158188  -1.42  0.155  -.0535123  .0084962
      L2. | -.0167578  .0148148  -1.13  0.258  -.0457942  .0122786
      Yd
      L1. | -.0059672  .0125585  -0.48  0.635  -.0305814  .0186471
      L2. | .0245347  .0108517  2.26  0.024  .0032659  .0458036
      Wealth
      L1. | .0031229  .0004701  6.64  0.000  .0022015  .0040443
      L2. | .0009201  .0006097  1.51  0.131  -.000275  .0021151
      InterestRate
      L1. | -.0429132  .1348302  -0.32  0.750  -.3071754  .2213491
      L2. | -.2899819  .0803339  -3.61  0.000  -.4474334  -.1325304
      _cons | -3.747375  1.525866  -2.46  0.014  -6.738018  -.7567332
-----+-----

```

Gambar 5.7 Hasil Uji Var

varsoc

```

Selection-order criteria
Sample: 1990 - 2017                                Number of obs   =   28
+-----+-----+-----+-----+-----+-----+-----+-----+
|lag|  LL   LR   df   p   FPE   AIC   HQIC   SBIC |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 0 | -598.199                5.6e+13  43.0142  43.0724  43.2046 |
| 1 | -495.504  205.39  16  0.000  1.2e+11  36.8217  37.1126  37.7733* |
| 2 | -468.965  53.077*  16  0.000  6.0e+10*  36.069*  36.5926*  37.7818 |
+-----+-----+-----+-----+-----+-----+-----+
Endogenous:  C Yd Wealth InterestRate
Exogenous:   _cons
    
```

Gambar 5.8 Hasil Uji Lag Optimal

Berdasarkan pada pengujian *Lag Optimal* dengan *Selection Order Criteria* yaitu LL, LR, FPE, AIC, HQIC dan SBIC diperoleh hasil bahwasannya *Lag Optimal* nya adalah di Lag ke dua, yaitu yang menunjukkan paling banyak tanda bintang. Menurut beberapa selection order criteria, yaitu LR, FPE, AIC dan HQIC memberikan rekomendasi lag kedua yang paling sesuai, dimana nilai yang ditandai dengan bintang (*). Dengan diketahuinya *Lag Optimal*, maka tahapan selanjutnya adalah melakukan uji kointegrasi, dalam hal ini menggunakan *Johansen Test*.

5.4. UJI KOINTEGRASI

Salah satu kegunaan uji kointegrasi adalah untuk menentukan model antara model *VAR in level* ataukah *Vector Error Correction Model (VECM)*. Jika pengujian menunjukkan hasil estimasi data yang tidak stasioner dan memiliki kointegrasi dengan variabel data yang lain maka model yang akan digunakan adalah *VECM*. Uji kointegrasi ini didasarkan pada hubungan antara *rank* dari sebuah matrik dengan akar karakteristiknya yang akan dihasilkan nilai *trace statistic* yang dibandingkan dengan *critical value*. Jika nilai *trace statistic* lebih besar dari *critical value* maka dapat diambil kesimpulan bahwa terdapat paling tidak dua hubungan kointegrasi antara variabel.

```
vecrank C Yd Wealth InterestRate, lag(2)
```

Johansen tests for cointegration						
Trend: constant					Number of obs =	28
Sample: 1990 - 2017					Lags =	2
maximum				5%		
rank	parms	LL	eigenvalue	trace statistic	critical value	
0	20	-515.85144	.	93.7719	47.21	
1	27	-492.11483	0.81649	46.2987	29.68	
2	32	-476.41494	0.67418	14.8989*	15.41	
3	35	-469.10721	0.40666	0.2835	3.76	
4	36	-468.96549	0.01007			

Gambar 5.9 Hasil Uji Kointegrasi

Dari hasil uji kointegrasi menggunakan Johansen Test diatas dapat kita lihat bahwa variabel-variabel penelitian terkointegrasi dalam jangka panjang, dimana yang ditunjukkan oleh nilai *trace statistic* > *critical value* 5%. Jadi dalam jangka panjang variabel-variabel tersebut akan mempengaruhi, yang mana ditunjukkan oleh tanda bintang.

Maka, setelah berbagai tahapan pengujian dan bahwa data penelitian stasioner pada tingkat diferensi pertama (*first different*) dan terjadi kointegrasi, maka model estimasi akhir yang sesuai digunakan adalah *Vector Error Correction Model* (VECM)

5.5. UJI MODEL Vector Error Correction Model (VECM)

Analisis VECM adalah analisis yang dikembangkan oleh Engle dan Granger (1987) untuk melihat rekonsiliasi perilaku ekonomi jangka pendek dengan variabel ekonomi jangka panjang (Gujarati, 2004). Konsep penting dalam VECM adalah keseimbangan jangka panjang dari data deret waktu (*time series*) yang sering disebut kointegrasi. Kointegrasi merupakan kombinasi linier dari variabel yang tidak stasioner dan terintegrasi pada ordo yang sama. Kointegrasi bertujuan untuk melihat hubungan keseimbangan dalam jangka panjang antara berubah-peubah yang diamati (Enders, 2004). Analisis VECM disebut juga sebagai model *vector auto-regression* restriksi dengan data yang tidak stasioner namun terdapat hubungan kointegrasi.

Dalam bab ini, sintak yang digunakan untuk pengujian VECM dengan menggunakan lag optimum yaitu lag kedua, adalah seperti berikut

Ketik **vec dC dYd dWealth dInterestRate, lag (2)**

Maka output yang keluar adalah seperti berikut

Aplikasi Penelitian Keuangan dan Ekonomi Syariah dengan STATA

vec dC dYd dWealth dInterestRate, lag (2)

Vector error-correction model

Sample: 1990 - 2017

Number of obs = 28

Log likelihood = -512.7049

AIC = 38.55035

Det(Sigma_ml) = 9.43e+10

HQIC = 38.94307

SBIC = 39.83497

Equation	Parms	RMSE	R-sq	chi2	P>chi2
D_dC	6	20.9502	0.8686	145.3929	0.0000
D_dYd	6	32.6733	0.7689	73.19997	0.0000
D_dWealth	6	514.021	0.5233	24.15487	0.0005
D_dInterestRate	6	2.94689	0.3373	11.19552	0.0825

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
D_dC						
_cel						
L1.	-.6136609	.3092356	-1.98	0.047	-1.219752	-.0075702
dC						
LD.	.1389183	.2617722	0.53	0.596	-.3741459	.6519824
dYd						
LD.	-.5952482	.1689104	-3.52	0.000	-.9263066	-.2641899
dWealth						
LD.	.0707451	.0066771	10.60	0.000	.0576581	.083832
dInterestRate						
LD.	2.194168	1.337972	1.64	0.101	-.4282084	4.816545
_cons	3.580816	3.993266	0.90	0.370	-4.245841	11.40747
D_dYd						
_cel						
L1.	.99144	.4822732	2.06	0.040	.0462018	1.936678
dC						
LD.	-.2309847	.408251	-0.57	0.572	-1.031142	.5691725
dYd						
LD.	-.5320237	.2634269	-2.02	0.043	-1.048331	-.0157165
dWealth						
LD.	.0514752	.0104134	4.94	0.000	.0310652	.0718851
dInterestRate						
LD.	1.874022	2.086655	0.90	0.369	-2.215747	5.96379
_cons	3.925877	6.22776	0.63	0.528	-8.280308	16.13206

D_dwealth						
_cel	L1.	-25.25619	7.587205	-3.33	0.001	-40.12684 -10.38555
dC	LD.	12.71793	6.422674	1.98	0.048	.1297193 25.30614
dYd	LD.	-11.88007	4.144276	-2.87	0.004	-20.0027 -3.757435
dwealth	LD.	.064892	.1638261	0.40	0.692	-.2562013 .3859852
dInterestRate	LD.	-25.38644	32.82761	-0.77	0.439	-89.72738 38.95449
_cons		.0671389	97.97618	0.00	0.999	-191.9626 192.0969

D_dInterestRate						
_cel	L1.	-0.062147	.0434974	-0.14	0.886	-.0914681 .0790387
dC	LD.	-.0013947	.0368212	-0.04	0.970	-.0735629 .0707736
dYd	LD.	-.0124066	.0237591	-0.52	0.602	-.0589736 .0341605
dwealth	LD.	.0024731	.0009392	2.63	0.008	.0006323 .0043139
dInterestRate	LD.	-.2059825	.1882007	-1.09	0.274	-.5748491 .1628841
_cons		-.1611085	.5616975	-0.29	0.774	-1.262015 .9397983

Cointegrating equations			
Equation	Parms	chi2	P>chi2
_cel	3	265.2141	0.0000

Identification: beta is exactly identified

Johansen normalization restriction imposed

beta	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_cel	dC	1	.	.	.	
	dYd	-.8256325	.0571344	-14.45	0.000	-.9376139 -.7136511
	dwealth	.0230653	.0067912	3.40	0.001	.0097549 .0363758
	dInterestRate	.564065	1.085869	0.52	0.603	-1.5642 2.69233
	_cons	-7.861692

Gambar 5.10 hasil uji VECM lag(2)

Dalam pengujian model VECM, terdapat syarat mutlak, yaitu ECT-1 yang ditunjukkan oleh simbol ce_1 dalam persamaan pertama adalah nilai koefisien negatif dan signifikan, dimana hasil tersebut menunjukkan bahwa pengujian VECM yang dilakukan, syarat tersebut telah terpenuhi.

Hasil pengujian VECM diatas dapat diinterpretasikan bahwa model menunjukkan long run causality bersamaan yang ditunjukkan pada nilai p-value < taraf toleransi 5%, yaitu nilai $P > |z| 0.047 < 0.05$ dan dengan nilai koefisien negatif (-.6136609). Selanjutnya, untuk mengetahui causality jangka pendek adalah dengan melihat pada baris berikutnya, berdasarkan pada hasil pengujian tersebut kita dapat menjelaskan secara sederhana dengan 4 blok persamaan yaitu

Blok Persamaan 1

$$dC = \beta_0 + dC_{t-1} + dYd_{t-1} + dwealth_{t-1} + dInteresRate_{t-1}$$

Dalam jangka pendek variabel dC_{t-1} tidak ada causality kepada dC (pengeluaran riil saat ini) yang ditandai dengan nilai $P > |z| 0.596 < 0.05$. Kemudian, variabel dYd_{t-1} menunjukkan ada hubungan causality dari dYd (pendapatan bersih riil) ke pengeluaran konsumsi riil yaitu yang ditandai dengan nilai $P > |z| 0.000 < 0.05$. selanjutnya variabel $dwealth_{t-1}$ memberikan adanya hubungan causality kepada dC (pengeluaran konsumsi riil) yang ditandai dengan nilai $P > |z| 0.000 < 0.05$, namun untuk variabel $dInteresRate_{t-1}$ tidak menunjukkan causality jangka pendek terhadap dC , dimana ditandai nilai $P > |z| 0.101 > 0.05$

Blok Persamaan 2

$$dYd = \beta_0 + dC_{t-1} + dYd_{t-1} + dWealth_{t-1} + dInteresRate_{t-1}$$

Blok Persamaan 3

$$dWealth = \beta_0 + dC_{t-1} + dYd_{t-1} + dWealth_{t-1} + dInteresRate_{t-1}$$

Blok Persamaan 4

$$dInteresRate = \beta_0 + dC_{t-1} + dYd_{t-1} + dWealth_{t-1} + dInteresRate_{t-1}$$

Latihan: Lakukan interpretasi causality jangka panjang dan jangka pendek untuk blok persamaan 2 sampai persamaan 4 dan bisa mencoba menjalankan VECM dengan data `times series.exls`

5.6. Analisis Impulse Response Function (IRF) dan Variance Decomposition (FEVD)

Dalam sub bab ini, ada 2 tambahan dalam analisis VAR/VECM yang akan juga diuji dalam model VECM/VAR yaitu **Impulse Response dan Variance Decomposition**. Impulse Response adalah suatu prosedur yang dapat diterapkan untuk mengestimasi dan melacak respon dari variabel endogen dalam sistem analisis VAR/VECM, karena adanya guncangan (shock) atau perubahan dalam variabel pengganggunya (e) atau dalam bab ini dengan kasus yang sama pada bab sebelumnya, shock apa yang diberikan variabel independen (Y_d) terhadap dependen (C). Impulse respon ini akan melacak respon saat ini dan masa setiap variabel akibat shock suatu variabel.

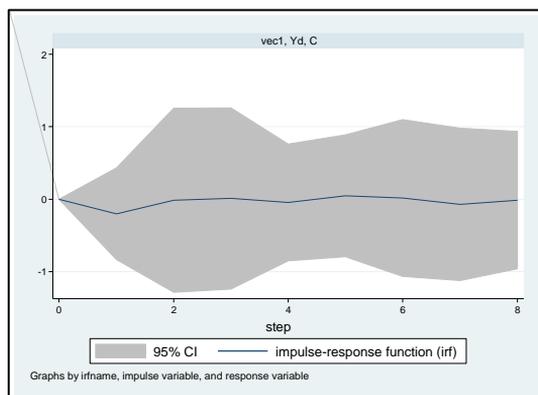
5.6.1. Impulse Response Function

Langkah-langkahnya adalah sebagai berikut

1. Dalam kasus ini kita melihat shock dari variabel pendapatan bersih pribadi riil atau Y_d terhadap variabel konsumsi riil atau C . Command yang digunakan adalah

```
irf create vec1, set (my vec1)
irf graph irf, irf(vec1) impulse(Yd) response(C)
```

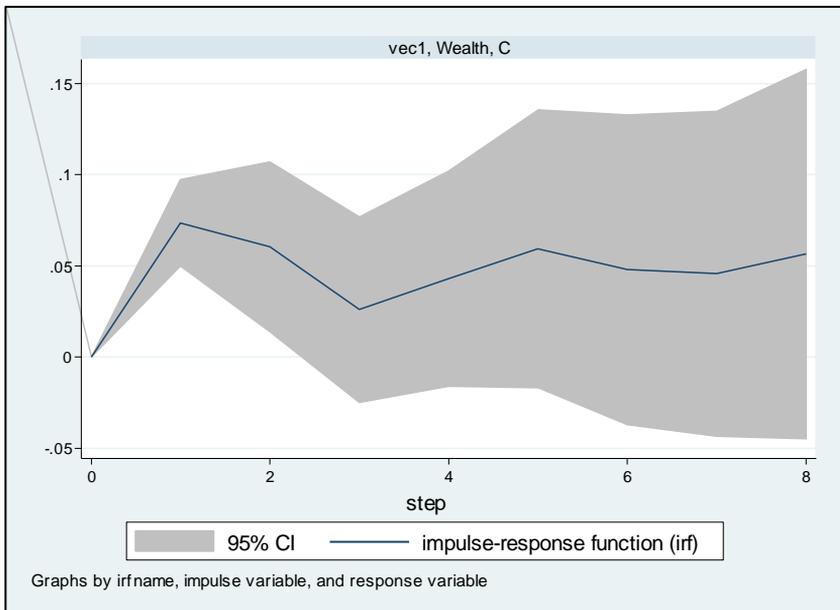
hasil output seperti berikut:



Output impulse respond menunjukkan shock dari variabel pendapatan bersih pribadi riil (Yd) terhadap variabel konsumsi riil (C) yaitu efeknya pada jangka pendek sekitar 1 periode atau dalam jangka pendek pendapatan bersih pribadi riil menurunkan minat konsumsi, dan dalam jangka panjang lamban menaikkan konsumsi.

2. Misalkan, kita mencoba untuk melihat shock varibale kekayaan (*Wealth*) terhadap pengeluaran konsumsi riil (C), maka perintah yang digunakan adalah

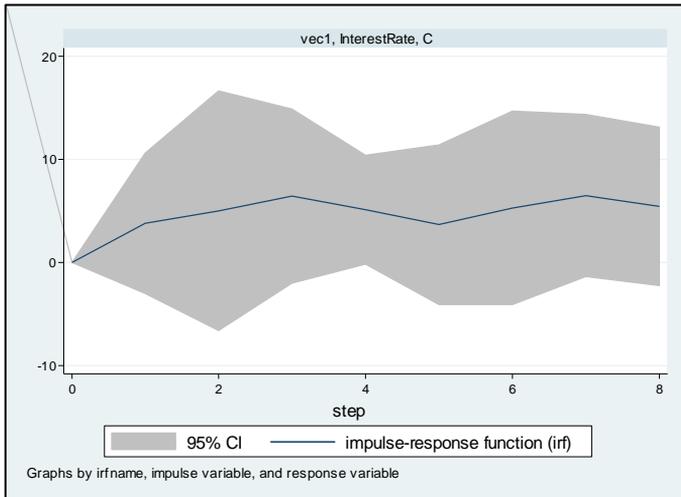
`irf graph irf, irf(vec1) impulse(Wealth) response(C)`
hasil output seperti berikut



3. Selanjutnya, kita melihat shock variabel Interestratae terhadap pengeluaran konsumsi riil, maka perintah yang diketik adalah

`irf graph irf, irf(vec1) impulse(InterestRate) response(C)`

hasil output seperti berikut:

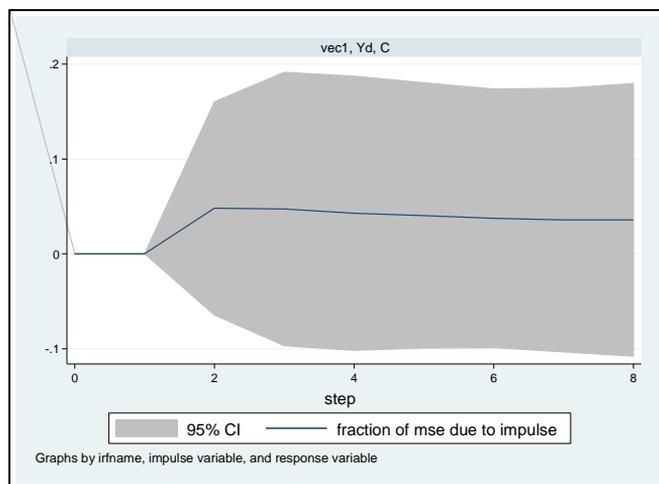


5.6.2. Forecasting Error Variance Decomposition (FEVD)

Variance decomposition digunakan untuk memprediksi kontribusi (persentase) varians setiap variabel karena adanya perubahan variabel tertentu dalam sistem VAR/VECM. Dalam hal ini kita bisa memprediksi variabel-variabel dalam sistem seperti berikut:

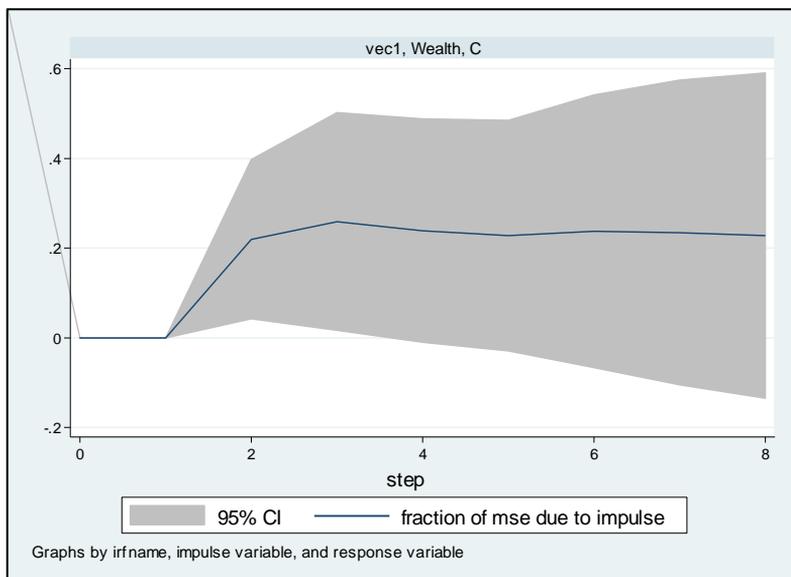
1. Untuk memprediksi seberapa besar variabel Yd, yaitu pendapatan bersih menjelaskan variable pengeluaran konsumsi riil. Command yang digunakan adalah

irf graph fevd, irf(vec1) impulse(Yd) response(C)
hasil output seperti berikut



2. Kita mencoba memprediksi seberapa besar kontribusi variabel 'Wealth' dalam hal ini kekayaan bersih terhadap pengeluaran konsumsi riil. Maka command yang digunakan adalah

`irf graph fevd, irf(vec1) impulse(Wealth) response(C)`
maka muncul hasil fevd - impulse variabel Wealth



BAB VI REGRESI LOGIT dan PROBIT

6.1. Regresi Logit

Regresi logit/logistic adalah salah satu model yang digunakan untuk mencari hubungan antara variabel dependen berskala dikotomi dengan variabel independent baik yang berbentuk kontiniu maupun kategorikal. Skala dikotomi yang dimaksud adalah skala data nominal dengan dua kategori, misalnya: Ya dan Tidak, Baik dan Buruk, atau Tinggi dan Rendah. Apabila pada regresi linier mewajibkan syarat atau asumsi bahwa *error varians (residual)* terdistribusi secara normal. Sebaliknya, pada regresi logistik tidak dibutuhkan asumsi tersebut sebab pada regresi logistik mengikuti distribusi logistik. Dalam praktiknya, regresi logistik menggunakan variabel independen untuk memprediksi peluang kemunculan hasil spesifik (*outcome*) pada variabel dependen. Dengan kata lain, regresi logistik dirancang untuk menggambarkan peluang-peluang yang terkait dengan nilai variabel dependen.

Berdasarkan regresi linier, estimasi dilakukan dengan persamaan $Y_i = a + bX_i + e_i$; sedangkan regresi logistik mengestimasi probabilitas bahwa $y = 1$ sebagai fungsi variabel independen $p = pr[y = 1|x] = F(x'\beta)$. Dalam model logistik, $F(x'\beta)$ mengikuti distribusi logistik berikut:

$$F(x'\beta) = \Lambda(x'\beta) = \frac{e^{x'\beta}}{1 + e^{x'\beta}} = \frac{1 + \exp(x'\beta)}{\exp(x'\beta)}$$

Interpretasi dari koefisien X pada model regresi logit adalah kenaikan (penurunan) nilai pada variabel independent X meningkatkan (menurunkan) peluang bahwa $Y=1$. **Regresi ini hanyalah mencari tahu arah hubungan variabel independen kepada variabel dependen, tetapi tidak bisa mengetahui seberapa besar dampaknya.** Berbeda dengan regresi linier, β dari

model regresi logit tidak menunjukkan besaran tambahan peluang terjadinya $Y=1$ saat terjadi peningkatan nilai variabel independen.

6.2. Regresi Probit

Regresi probit memiliki fungsi yang sama persis dengan regresi logistik, yakni sebagai model yang menjelaskan hubungan variabel dependen berskala dikotomi dengan variabel independen. Hanya saja fungsi probabilitas kumulatif dari regresi probit berbeda. $F(x'\beta)$ dari regresi probit menggunakan distribusi standar normal berikut

$$F(x'\beta) = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \phi(z) dz$$

Interpretasi dari koefisien X pada model regresi probit adalah kenaikan (penurunan) nilai pada variabel X meningkatkan (menurunkan) peluang bahwa $Y=1$. Dari regresi ini yang bisa dilakukan hanyalah mencari tahu arah hubungan variabel independen kepada variabel dependen, namun tidak bisa mengetahui seberapa besar dampak hubungannya. Berbeda dengan regresi linier, β dari model regresi probit tidak menunjukkan besaran tambahan peluang terjadinya $Y=1$ saat terjadi peningkatan nilai variabel independen.

6.3 Marginal Effect

Dalam melakukan estimasi menggunakan model logit ataupun probit, pada umumnya ditunjukkan *marginal effect* setelah presentasi hasil koefisien probit/logit. Besaran *marginal effect* merfleksikan perubahan probabilitas pada $Y=1$ setiap adanya perubahan satu unit pada variabel independen (X). Pada estimasi linier, *marginal effect* adalah koefisien β dan itu tidak terpengaruhi nilai variabel independen. Sedangkan pada model regresi logit dan probit dihitung dengan:

$$\frac{\partial p}{\partial x_j} = F'(x'\beta)\beta_j$$

Marginal effect bergantung pada nilai x , sehingga kita perlu mengestimasi *marginal effect* pada suatu titik tertentu dari x , misalnya pada titik reratanya, minimum, dan maksimumnya.

Pada model logit, *marginal effect* mengikuti formula:

$$\frac{\partial p}{\partial x_j} = \Lambda(x'\beta)[1 - \Lambda(x'\beta)]\beta_j = \frac{e^{x'\beta}}{(1 + e^{x'\beta})^2}\beta_j$$

Adapun pada model probit, *marginal effect* mengikuti formula:

$$\frac{\partial p}{\partial x_j} = \phi(x'\beta)\beta_j$$

6.4. Menjalankan Regresi Logit dengan STATA

Regresi Logit (*logistic regression*) sebetulnya mirip dengan analisis diskriminan yaitu kita ingin menguji apakah probabilitas terjadinya variabel terikat dapat diprediksi dengan variabel bebasnya. Logistic regression juga disebut model regresi binary response, karena variabel terikatnya diukur dengan skala nominal dua kategori. Berikut ini beberapa kasus dan salah satu contoh kasus (yang diambil dari Imam Ghozali (2013) yang umumnya dianalisis dengan logistic regression:

- a. Seorang dokter ingin mengetahui apakah probabilitas seorang pasien terserang penyakit jantung atau tidak terserang dapat diprediksi dari tekanan darah, kadar kolesterol, kalori yang dimakan, jenis kelamin dan gaya hidup
- b. Seorang dosen ingin menentukan probabilitas mahasiswa lulus dengan nilai cumlaude (A) dengan melihat beberapa sikap, indeks prestasi pada saat masuk kuliah, score mata kuliah, dan metode pembelajaran yang digunakan (metode baru atau metode lama)

Kasus-kasus seperti ini sebetulnya dapat dianalisis dengan analisis diskriminan. Namun, asumsi multivariate normal distribution tidak dapat dipenuhi karena variabel bebas adalah campuran antara variabel kontinyu (matrik) dan kategorial (nonmetrik). Dalam hal ini dapat dianalisis dengan logistic regression karena tidak perlu asumsi normalitas data pada variabel bebasnya. Jadi logistic regression umumnya dipakai jika asumsi multivariate normal distribution tidak terpenuhi.

6.5. Menjalankan Regresi Probit dengan STATA

Seperti yang telah dijelaskan pada bagian sebelumnya bahwa, *logistic regression* dalam mengestimasi model menggunakan fungsi kumulatif logistik (*cumulative logistic function*). Fungsi kumulatif logistic bukan satu-satunya *cumulative distribution function* (CDF) yang dapat digunakan. Pada beberapa aplikasi, normal cumulative distribution function lebih sesuai untuk digunakan. Model estimasi yang menggunakan normal CDF sering disebut dengan probit model atau normil model. Secara prinsip kita dapat mengganti logistic CDF dengan normal CDF dan melakukan analisis seperti halnya logistic regression. Untuk membandingkan kedua konsep tersebut, kita bisa merujuk pada Gujarati (2003) bab 15.

Dalam sub bab ini akan memberikan contoh model probit, dan logit dengan menggunakan data: probit.xls di CD, kita akan melihat pengaruh metode pembelajaran, Nilai Mata Kuliah (Score), Indeks Prestasi ketika masuk (IP) terhadap Nilai Akhir Matakuliah (*Cumlaude*), secara sederhana dapat dibuat model persamaan seperti berikut:

$$P_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_{k-1} X_{k-1}, i$$

$$L_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_{k-1} X_{k-1}, i$$

- Pi = Cumlaude (Nilai Akhir) = kode 1 menunjukkan nilai akhir yang diperoleh adalah A
kode 0 menunjukkan nilai akhir yang diperoleh adalah B/C
- X1 = Metode = kode 1 menunjukkan metode pembelajaran baru
Kode 0 menunjukkan metode pembelaran lama
- X2 = Score = Nilai ujian yang diberikan pada awal semester untuk menguji pengetahuan ekonomi makro
- X2 = IP = Indeks prestasi ketika masuk studi

Dalam sub ini, kita akan membahas

1. Melakukan regresi logit dan probit di STATA
2. Uji Diagnostik regresi logit dan probit

6.6. Tahapan melakukan analisis regresi probit dan logit menggunakan

STATA:

1. Mengimport data dari excel ke STATA seperti langkah-langkah sebelumnya, selanjutnya persiapan data dengan menentukan variable dependen dan independent, yaitu dengan menggunakan perintah sebagai berikut:

```
global ylist Cumlaude
global xlist Metode Score IP
describe $ylist $xlist
summarize $ylist $xlist
tabulate $ylist
```

describe \$ylist \$xlist

```
describe $ylist $xlist
```

variable name	storage type	display format	value label	variable label
Cumlaude	byte	%10.0g		Cumlaude
Metode	byte	%10.0g		Metode
Score	byte	%10.0g		Score
IP	double	%10.0g		IP

summarize \$ylist \$xlist

Variable	Obs	Mean	Std. Dev.	Min	Max
Cumlaude	32	.34375	.4825587	0	1
Metode	32	.4375	.5040161	0	1
Score	32	21.9375	3.901509	12	29
IP	32	3.179688	.6293544	2.06	5.53

tabulate \$ylist

Cumlaude	Freq.	Percent	Cum.
0	21	65.63	65.63
1	11	34.38	100.00
Total	32	100.00	

- Selanjutnya, melakukan regresi untuk masing-masing pengujian yakni, regresi logit dan regresi probit, dengan perintah sebagai berikut:

Regresi logit

logit Cumlaude Metode Score IP

estimate store logit

```
Iteration 0: log likelihood = -20.59173
Iteration 1: log likelihood = -15.448627
Iteration 2: log likelihood = -15.236285
Iteration 3: log likelihood = -15.235346
Iteration 4: log likelihood = -15.235346
```

```

Logistic regression                               Number of obs   =       32
LR chi2(3)                                       =       10.71
Prob > chi2                                      =       0.0134
Pseudo R2                                       =       0.2601
Log likelihood = -15.235346
    
```

Cumlaude	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Metode	2.206321	.9555803	2.31	0.021	.3334177	4.079223
Score	.1309219	.1337411	0.98	0.328	-.1312058	.3930497
IP	1.018485	.7421655	1.37	0.170	-.4361328	2.473102
_cons	-8.020292	3.451206	-2.32	0.020	-14.78453	-1.256053

Hasil estimasi logit di atas menunjukkan bahwa variabel “Metode” berpengaruh signifikan positif terhadap probabilitas mendapatkan Cumlaude (Nilai A) yang ditunjukkan dengan koefisien sebesar 2.206321 dan nilai $P>|z|$ sebesar 0.021. secara bersama-sama ketiga variabel berpengaruh signifikan terhadap probabilitas mendapat nilai Cumlaude yang ditunjukkan dengan LR statistic (sebesar 10.71 dan nilai P sebesar 0.0134 yang lebih kecil dari 0.05. Sebagai pembanding hasil logit, kita dapat menggunakan model probit. Caranya sama dengan diatas, perintah yang digunakan adalah sebagai berikut:

Regresi probit

```

probit Cumlaude Metode Score IP
estimate store probit
    
```

```

Iteration 0:  log likelihood = -20.59173
Iteration 1:  log likelihood = -15.200703
Iteration 2:  log likelihood = -15.08453
Iteration 3:  log likelihood = -15.084241
Iteration 4:  log likelihood = -15.084241
    
```

```

Probit regression                               Number of obs   =       32
LR chi2(3)                                       =       11.01
Prob > chi2                                      =       0.0116
Pseudo R2                                       =       0.2675
Log likelihood = -15.084241
    
```

Cumlaude	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Metode	1.356962	.5543741	2.45	0.014	.2704088	2.443515
Score	.0800558	.0801041	1.00	0.318	-.0769454	.2370571
IP	.6213893	.415483	1.50	0.135	-.1929424	1.435721
_cons	-4.903271	1.978069	-2.48	0.013	-8.780216	-1.026326

Secara kualitatif, hasil model logit sama dengan model probit yaitu variabel ‘Metode’ berpengaruh signifikan terhadap variabel ‘Cumlaude’.

Namun, kita tidak dapat memperbandingkan hasil koefisien regresi logit dan probit secara langsung (Gujarati, 2003, hal 613). Untuk model logit, kita dapat menginterpretasikan dalam terminologi odds yang dapat diperoleh dengan antilog setiap koefisien regresi. Misalnya, untuk koefisien variable 'Metode' di atas sebesar 1.356962 jika dihitung antilog-nya adalah sebesar 22,7489. Hal ini menunjukkan bahwa mahasiswa yang mendapatkan metode pembelajaran model baru memiliki probabilitas mendapatkan nilai Cumlaude (nilai A) 22 kali lebih besar dibandingkan dengan mahasiswa yang mendapatkan metode pembelajaran dengan model lama, dengan mengasumsikan faktor-faktor lain bersifat tetap (konstan).

Secara sederhana, kita dapat menyatakan bahwa hasil logit dan probit menunjukkan bahwa probabilitas mendapatkan nilai Cumlaude lebih besar untuk mahasiswa yang mendapatkan metode pembelajaran baru dibandingkan dengan metode pembelajaran lama.

3. Melakukan display hasil regresi logit dan probit, menggunakan perintah sebagai berikut:

Display hasil regresi logit dan probit

```
estimate table probit logit, star stats(N)
```

Variable	probit	logit
Metode	1.356962*	2.2063206*
Score	.08005583	.13092191
IP	.62138929	1.0184848
_cons	-4.9032711*	-8.0202918*
N	32	32

legend: * p<0.05; ** p<0.01; *** p<0.00

4. Menghitung marginal effect setiap regresi, perintah yang digunakan sebagai berikut:

Marginal effect regresi logit

```
quietly logit Cumlaude Metode Score IP  
margins, dydx(*) atmeans post
```

```
Conditional marginal effects      Number of obs   =       32  
Model VCE      : OIM  
  
Expression   : Pr(Cumlaude), predict()  
dy/dx w.r.t. : Metode Score IP  
at           : Metode           =       .4375 (mean)  
              Score            =      21.9375 (mean)  
              IP               =      3.179688 (mean)
```

	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
Metode	.4447938	.1820323	2.44	0.015	.0880171	.8015705
Score	.0263938	.0264961	1.00	0.319	-.0255376	.0783253
IP	.2053263	.1499468	1.37	0.171	-.088564	.4992167

Hasil perhitungan marginal effect regresi logit diatas melihat pengaruh variabel 'Metode' (dengan variabel independent lainnya yang ditetapkan pada nilai rata-rata terhadap probabilitas pada nilai Cumlaude (A) dalam studi. Kemudian Tabel marginal effect memperlihatkan ' dy/dx yang dapat diterjemahkan sebagai pengukuran perubahan pada nilai probabilitas. Misalnya nilai dy/dx untuk variabel 'Metode' sebesar 0.4447938. Nilai tersebut dapat diinterpretasikan bahwa metode pembelajaran meningkat 1 metode baru, maka probabilitas untuk menjadi Cumlaude (Nilai A) akan meningkat 4,4 %.

Marginal effect regresi probit

quietly probit Cumlaude Metode Score IP
 margins, dydx(*) atmeans post

```

Conditional marginal effects      Number of obs      =      32
Model VCE      :      OIM

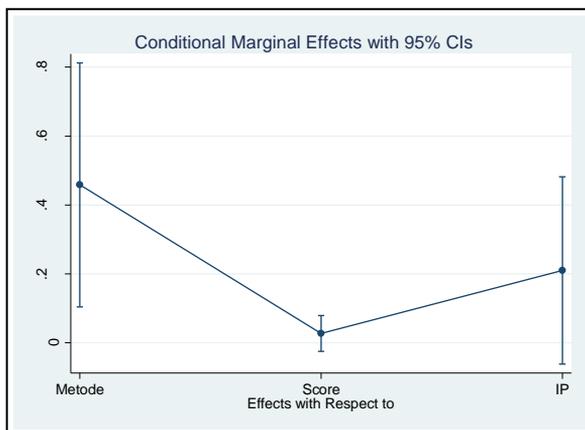
Expression      :      Pr(Cumlaude), predict()
dy/dx w.r.t.   :      Metode Score IP
at              :      Metode          =      .4375 (mean)
                Score                =      21.9375 (mean)
                IP                    =      3.179688 (mean)
    
```

	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
Metode	.4581892	.1806129	2.54	0.011	.1041945	.812184
Score	.0270315	.0266406	1.01	0.310	-.0251831	.0792461
IP	.2098171	.1387182	1.51	0.130	-.0620655	.4816997

5. Melihat probabilitas regresi logit dan probit dari variable independent yang lain dalam bentuk grafik. Perintah yang digunakan adalah ‘marginsplot’ yaitu untuk mengetahui nilai perubahan probabilitas dari variabel independent, berikut hasil output marginsplot.

```

quietly probit Cumlaude Metode Score IP
margins, dydx(*) atmeans post
marginsplot
quietly logit Cumlaude Metode Score IP
margins, dydx(*) atmeans post
marginsplot
    
```



- Melihat pengaruh responden dengan masing-masing variable independent ditetapkan pada nilai minimum atau maksimum terhadap probabilitas menjadi Cumlaude?

Pada titik Minimum

```
quietly probit Cumlaude Metode Score IP
margins, dydx(*) at(Metode=(0) Score=(12) IP=(2.06))
marginsplot
```

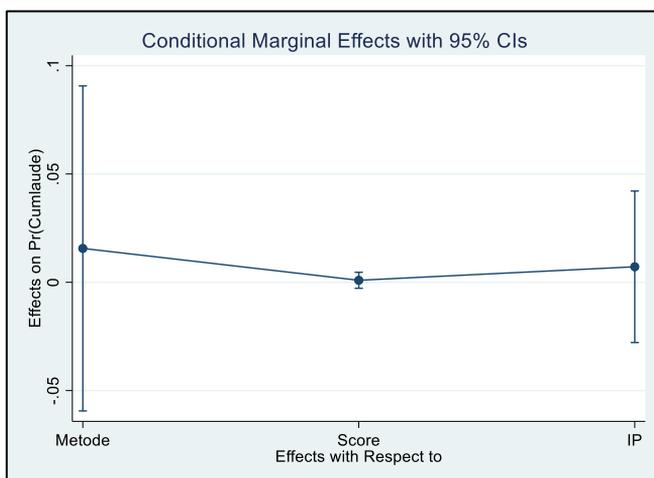
selanjutnya akan muncul output seperti berikut:

```
Conditional marginal effects      Number of obs   =       32
Model VCE      : OIM

Expression     : Pr(Cumlaude), predict()
dy/dx w.r.t.  : Metode Score IP
at             : Metode      =         0
               : Score       =        12
               : IP         =        2.06
```

		Delta-method				[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z			
Metode	.0156349	.0382882	0.41	0.683	-.0594086	.0906784	
Score	.0009224	.0018855	0.49	0.625	-.0027731	.0046179	
IP	.0071597	.0178472	0.40	0.688	-.0278202	.0421395	

```
marginsplot
```



Pada titik maksimum

```
quietly probit Cumlaude Metode Score IP
margins, dydx(*) at (Metode=(1) Score=(29) IP=(5.53))
marginsplot
```

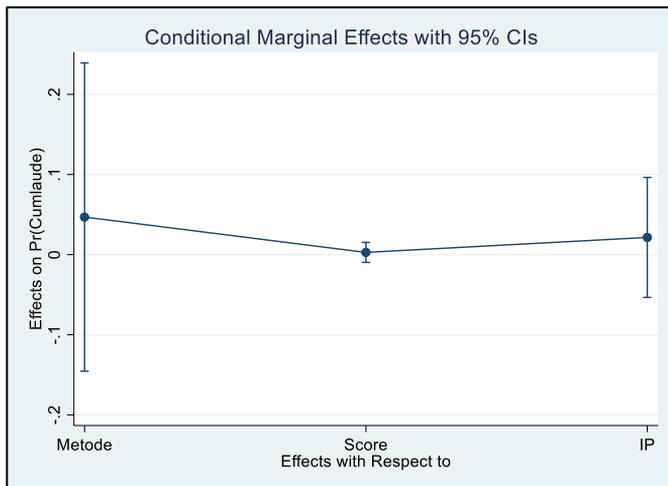
selanjutnya akan muncul output seperti berikut:

```
Conditional marginal effects      Number of obs      =      32
Model VCE      _:      OIM

Expression      _:      Pr(Cumlaude), predict()
dy/dx w.r.t.   _:      Metode Score IP
at              _:      Metode      =      1
                  Score      =      29
                  IP      =      5.53
```

		Delta-method			[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z		
Metode	.0469223	.0981591	0.48	0.633	-.145466	.2393106
Score	.0027682	.0063792	0.43	0.664	-.0097348	.0152713
IP	.021487	.038207	0.56	0.574	-.0533973	.0963712

```
marginsplot
```



7. Membuat prediksi probabilitas logit dan probit, perintah yang digunakan adalah seperti berikut:

```
quietly logit Cumlaude Metode Score IP
predict plogit, pr
```

```
quietly probit Cumlaude Metode Score IP
predict pprobit, pr
```

```
summarize $ylist plogit pprobit
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Cumlaude	32	.34375	.4825587	0	1
plogit	32	.34375	.2692374	.0300246	.823284
pprobit	32	.3395097	.2740866	.0166626	.8275898

8. Menghitung *goodness of fit* (ketepatan hasil prediksi)

Perhitungan ketepatan hasil prediksi

```
quietly logit Cumlaude Metode Score IP
estat classification
```

Logistic model for Cumlaude

Classified	True -----		Total
	D	~D	
+	6	5	11
-	5	16	21
Total	11	21	32

```
Classified + if predicted Pr(D) >= .5
True D defined as Cumlaude != 0
```

Sensitivity	Pr(+ D)	54.55%
Specificity	Pr(- ~D)	76.19%
Positive predictive value	Pr(D +)	54.55%
Negative predictive value	Pr(~D -)	76.19%
False + rate for true ~D	Pr(+ ~D)	23.81%
False - rate for true D	Pr(- D)	45.45%
False + rate for classified +	Pr(~D +)	45.45%
False - rate for classified -	Pr(D -)	23.81%
Correctly classified		68.75%

quietly probit Cumlaude Metode Score IP
estat classification

Probit model for Cumlaude

Classified	True		Total
	D	~D	
+	6	5	11
-	5	16	21
Total	11	21	32

Classified + if predicted Pr(D) >= .5
True D defined as Cumlaude != 0

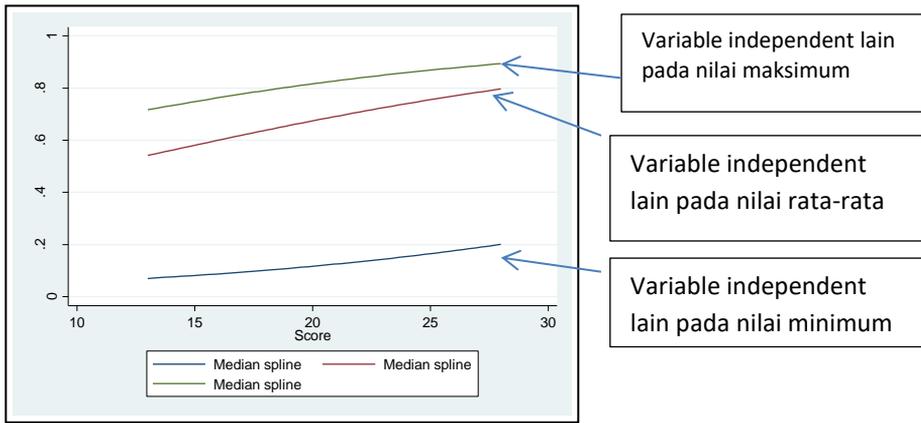
Sensitivity	Pr(+ D)	54.55%
Specificity	Pr(- ~D)	76.19%
Positive predictive value	Pr(D +)	54.55%
Negative predictive value	Pr(~D -)	76.19%
False + rate for true ~D	Pr(+ ~D)	23.81%
False - rate for true D	Pr(- D)	45.45%
False + rate for classified +	Pr(~D +)	45.45%
False - rate for classified -	Pr(D -)	23.81%
Correctly classified		68.75%

Berdasarkan hasil uji goodness of fit pada regresi logit dan regresi probit diatas menunjukkan bahwa sebanyak 11 respondengan yang menjadi Cumlaude, ada 6 responden yang diprediksi secara tepat dan 5 responden yang diprediksi secara tidak tepat. Selanjutnya terdapat 21 total responden yang tidak menjadi Cumlaude, terdapat 5 responden yang diprediksi secara tepat dan ada 16 responden yang diprediksi secara tidak tepat.

- Melihat dan menampilkan efek dari variabel independent lain (selain Score) dalam bentuk plot/kurva. Perintah yang digunakan sebagai berikut:

```
quietly probit Coumlaude Metode Score IP
generate L1=_b[_cons]+_b[Metode]*0+_b[Score]*Score+_b[IP]*2.06
generate Phat1=1/(1+exp(-L1))
```

```
generate  
L2=_b[_cons]+_b[Metode]*0.4375+_b[Score]*Score+_b[IP]*3.179688  
generate Phat2=1/(1+exp(-L2))  
  
generate L3=_b[_cons]+_b[Metode]*1+_b[Score]*Score+_b[IP]*5.53  
generate Phat3=1/(1+exp(-L3))  
  
graph twoway mspline Phat1 Score||mspline Phat2 Score||mspline  
Phat3 Score
```



Gambar 6.3 effect pada titik maksimum, rata-rata, dan minimum

Latihan: Lakukan analisis regresi probit dan logit menggunakan data yang tersedia di CD dengan nama file **probit.exls**

10. Menghitung besarnya nilai probabilitas dari responden untuk menjadi Cumlaude dalam belajar? (untuk semua variabel independent dalam rata-rata). Perintah yang digunakan sebagai berikut:

```
quietly probit Coumlaude Metode Score IP  
mfx, predict (p)
```

```

Marginal effects after probit
y = Pr(Cumlaude) (predict, p)
= .28178346
-----+-----
variable |      dy/dx   Std. Err.   z   P>|z|   [   95% C.I.   ]   X
-----+-----
Metode* |   .4529208   .16395   2.76   0.006   .13159   .774252   .4375
Score |   .0270315   .02664   1.01   0.310   -.025183   .079246   21.9375
IP |   .2098171   .13872   1.51   0.130   -.062065   .4817   3.17969
-----+-----
(*) dy/dx is for discrete change of dummy variable from 0 to 1

```

11. Menghitung besarnya nilai probabilitas dari responden untuk menjadi Cumlaude dalam belajar? (untuk responden yang menggunakan metode pembelajaran baru, Scorenya yang paling tinggi (29)) dan IP=5.53 (indek prestasi yang tinggi saat masuk kuliah). Command yang digunakan seperti berikut:

```
mfx, predict (p) at (Metode=1, Score=29, IP=5.53)
```

```

Marginal effects after probit
y = Pr(Coumlaude) (predict, p)
= .9865026
-----+-----
variable |      dy/dx   Std. Err.   z   P>|z|   [   95% C.I.   ]   X
-----+-----
Metode* |   .1828804   .21626   0.85   0.398   -.240985   .606745   1
Score |   .0027682   .00638   0.43   0.664   -.009735   .015271   29
IP |   .021487   .03821   0.56   0.574   -.053397   .096371   5.53
-----+-----
(*) dy/dx is for discrete change of dummy variable from 0 to 1

```

12. **UJI DIAGNOSTIK (LINK TEST)** merupakan test correctly specified model.

Command yang digunakan sebagai berikut

```
quietly probit Coumlaude Metode Score IP
linktest, nolog
```

hasil output seperti berikut

```

Probit regression                               Number of obs   =       32
                                                LR chi2(2)      =       12.38
                                                Prob > chi2     =       0.0021
Log likelihood = -14.403481                    Pseudo R2      =       0.3005
    
```

Coumlaude	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_hat	.6212533	.4733702	1.31	0.189	-.3065352 1.549042
_hatsq	-.6607898	.6481279	-1.02	0.308	-1.931097 .6095177
_cons	.2879519	.3874014	0.74	0.457	-.4713408 1.047245

Hat : signifikan, good model
Hatsq : tidak signifikan, correctly specified

13. UJI DIAGNOSTIK (GOODNESS OF FIT TEST) : untuk melihat apakah nilai probabilitas sudah tepat dengan data observasinya

quietly probit Coumlaude Metode Score IP
estat gof
Hasil output seperti berikut:

Probit model for Coumlaude, goodness-of-fit test

```

number of observations =       32
number of covariate patterns =       32
Pearson chi2(28) =       26.62
Prob > chi2 =       0.5389
    
```

Prob. tidak signifikan, maka model probit sudah baik

14. UJI DIAGNOSTIK

VIF Test: Uji Multikolinearitas

qui regress Coumlaude Metode Score IP
vif

Variable	VIF	1/VIF
Score	1.21	0.826360
IP	1.20	0.834037
Metode	1.03	0.974490
Mean VIF	1.15	

Nilai tolerance ($1/VIF$) kurang dari 0,2 maka koefisien regresi probit kurang stabil
Sebaliknya, lebih dari 0,2 maka tidak ada masalah multikolinieritas

BAB VII

ANALISIS KUESIONER MENGGUNAKAN STATA

7.1. Analisis Data Primer Menggunakan STATA

Pada bab-bab sebelumnya, telah dijelaskan bagaimana melakukan analisis untuk rancangan penelitian yang menggunakan metode inferensial (regresi) yang banyak menggunakan data sekunder. Pada bagian ini akan dijelaskan penggunaan STATA untuk metode penelitian lain seperti penelitian yang menggunakan data primer.

Penggunaan data primer seperti survei banyak dilakukan untuk tujuan penyusunan tugas akhir di lingkungan mahasiswa S1, S2 maupun S3. Data primer yakni data yang diperoleh langsung dari sumber/objek penelitian. Berbagai cara dapat dilakukan untuk memperoleh data primer seperti wawancara, kuesioner, observasi, atau *focus group discussion*. Salah satu metode pengumpulan data tersebut yang cukup sering digunakan adalah kuesioner.

Memanfaatkan kuesioner untuk memperoleh data primer memiliki beberapa keunggulan seperti memiliki jangkauan yang luas, memperoleh respon yang lebih banyak dibandingkan metode pengumpulan data lain, cukup mudah dilakukan, serta tentunya cenderung lebih murah terutama saat ini penggunaan kuesioner dalam jaringan sudah jamak dilakukan. Namun, kuesioner pun memiliki kelemahan antara lain peneliti akan sulit mengontrol kualitas jawaban serta kedalaman informasi yang menjadi terbatas. Alangkah baiknya untuk menggabungkan berbagai metode pengumpulan data untuk saling melengkapi kelemahan setiap metode.

Setelah data dikumpulkan menggunakan kuesioner, tahap yang selanjutnya perlu dilakukan adalah menyiapkan data yang terkumpul untuk dianalisis. Namun, sebelum digunakan untuk analisis, data yang terkumpul perlu dilakukan pengujian apakah memenuhi syarat validitas dan realibilitas data sebagai indikator kualitas kuesioner yang buat. Validitas data mengukur ketepatan suatu data sebagai alat ukur atas sesuatu yang hendak diuji. Validitas terdiri dari validitas individu (*item*) dan validitas total. Validitas individu melihat ketepatan setiap unit pengukur (dalam hal ini dapat setiap pertanyaan dalam kuesioner), sedangkan validitas total melihat ketepatan keseluruhan alat ukur.

Sementara itu, reliabilitas data digunakan untuk menguji apakah alat ukur yang digunakan dapat diandalkan untuk perolehan informasi serta mencerminkan informasi yang sesuai dengan fakta di lapangan. Salah satu cara yang dapat digunakan untuk menguji keandalan suatu alat ukur adalah **Cronbach Alpha**. Suatu alat ukur dikatakan dapat diandalkan jika memiliki nilai Cronbach Alpha $> 0,70$.

STATA memberikan kemudahan untuk melakukan pengujian validitas dan reliabilitas data hasil pengumpulan kuesioner. Berikut langkah-langkah yang perlu dilakukan:

1. Jika sebelumnya menggunakan kuesioner yang disebarkan secara fisik, maka terlebih dahulu pindahkan data dari kuesioner ke spreadsheet. Jika menggunakan kuesioner dalam jaringan, maka cukup lakukan pengunduhan data dari situs kuesioner dalam jaringan yang digunakan. Beberapa contoh situs yang dapat digunakan untuk melakukan pengumpulan data dengan kuesioner antara lain: esurvey creator (<http://www.esurveycreator.com>); survey monkey (<http://www.surveymonkey.com>); atau yang paling sederhana melalui google form (<http://forms.google.com>).

2. Lakukan pengimporan data yang sudah dimasukkan dalam spreadsheet seperti langkah yang dilakukan pada bab sebelumnya. Baris menyatakan responden, sedangkan kolom menyatakan pertanyaan dalam kuesioner (diperlakukan sebagai variabel).
3. Jika menggunakan survei dalam jaringan dan terdapat respon yang sifatnya non-numerik, maka dilakukan encoding data menggunakan perintah **encode** seperti yang dicontohkan pada Bab I.
4. Lakukan pengujian validitas dan reliabilitas menggunakan sintaks sebagai berikut:

alpha nama_variabel(nama pertanyaan), opsi

Opsi yang digunakan: **asis, casewise, item, lab**. Untuk opsi lebih detil dapat dilihat pada menu *help* pada STATA. Selain menggunakan sintaks di atas, dapat menggunakan menu utama **Statistics > Multivariate Analysis → Cronbach's Alpha**

Hasil yang perlu diperhatikan adalah nilai **item-rest correlation** yang menunjukkan validitas pertanyaan. Nilai ini akan dibandingkan dengan nilai *r*-tabel yang ditentukan berdasarkan nilai *degree of freedom* yakni $n-2$ (dengan n adalah jumlah responden). Pada tahap ini jika ditemukan pertanyaan yang dinyatakan tidak valid, lakukan kembali pengujian alpha setelah menghapuskan pertanyaan yang dinilai tidak valid tersebut. Lakukan proses ini hingga seluruh pertanyaan memiliki nilai valid.

Indikator kedua yang perlu dilihat adalah nilai **Cronbach Alpha** yang ditunjukkan untuk setiap pertanyaan maupun secara keseluruhan. Seperti yang sebelumnya dijelaskan, suatu kuesioner dapat diandalkan jika memiliki nilai Cronbach Alpha $> 0,7$.

Berikut contoh pengujian validitas dan reliabilitas menggunakan data yang tersedia di CD pada **file kuesioner.xls**.

7.2. Uji Validitas

Sebelumnya kita bisa melihat karakter data variabel yang digunakan dalam model secara deskriptif, yaitu dengan perintah *summarize*, yaitu seperti berikut

```
summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
FP01	99	3.689211	.9328645	1	4.941778
FP02	99	3.766687	.8510755	2.371907	5.217636
FP03	99	3.761849	.8982638	1	5.143621
FP04	99	3.739409	.8934767	1	5.217636
ICT01	99	3.128981	.8553419	1	4.467823
ICT02	99	3.84995	.9514296	1	5.157511
ICT03	99	3.692163	.9522964	1	5.107138
ICT04	99	3.450887	.8436987	1	5.272188
HC01	99	3.691533	.9297419	1	4.857758
HC02	99	3.94232	.9064072	2.528165	5.505428
HC03	99	3.742287	.885651	2.212196	4.996763
HC04	99	4.009455	.882817	2	5.200653
OC01	99	4.413582	.9558403	2	5.636177
OC02	99	3.965221	.8713746	2	5.201491
OC03	99	4.110977	.896757	2	5.2664
OC04	99	4.547528	.8512819	2	5.412891
CC01	99	3.731898	.8734477	1	5.016005
CC02	99	3.878905	.8187891	2.331886	5.268997
CC03	99	3.562289	.9214461	1	5.187011
CC04	99	3.636003	.8245546	2.097794	5.192023

Perintah yang digunakan untuk melihat tingkat validitas setiap pertanyaan adalah: ketik alpha diikuti indikator1, indikator2, indikator3, indikator4, dan seterusnya, seperti berikut:

alpha FP01 FP02 FP03 FP04 ICT01 ICT02 ICT03 ICT04 HC01 HC02 HC03
 HC04 OC01 OC02 OC03 OC04 CC01 CC01 CC02 CC03 CC04, item

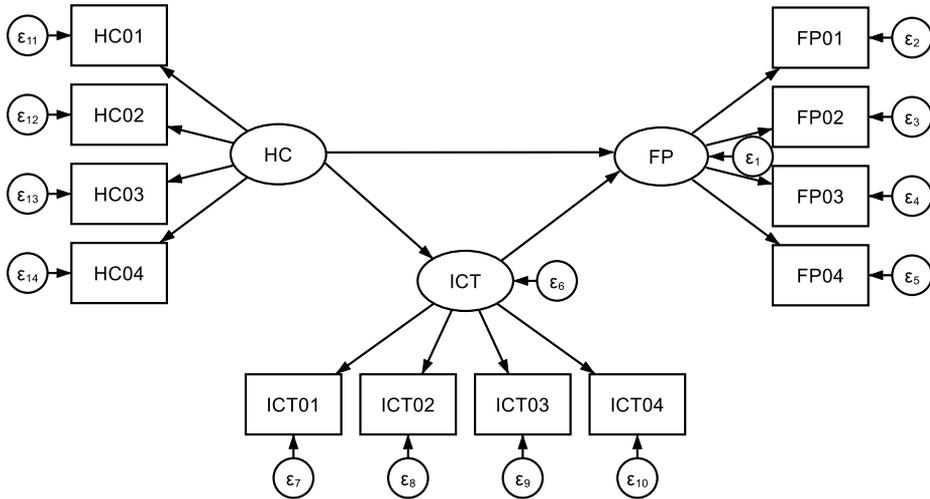
Test scale = mean(unstandardized items)

Item	Obs	Sign	item-test correlation	item-rest correlation	interitem covariance	alpha
FP01	99	+	0.6441	0.5752	.1534191	0.8209
FP02	99	+	0.5749	0.5045	.1582962	0.8248
FP03	99	+	0.5603	0.4839	.1580604	0.8255
FP04	99	+	0.3833	0.2916	.1662726	0.8344
ICT01	99	+	0.3468	0.2570	.1682353	0.8357
ICT02	99	+	0.6322	0.5601	.1535918	0.8215
ICT03	99	+	0.5424	0.4590	.1579645	0.8266
ICT04	99	+	0.3695	0.2823	.1673432	0.8345
HC01	99	+	0.5308	0.4483	.1588981	0.8272
HC02	99	+	0.4498	0.3618	.1630555	0.8313
HC03	99	+	0.5420	0.4648	.1591253	0.8265
HC04	99	+	0.4724	0.3887	.1623307	0.8300
OC01	99	+	0.4797	0.3893	.1609863	0.8301
OC02	99	+	0.5148	0.4362	.1605932	0.8278
OC03	99	+	0.5843	0.5108	.156981	0.8243
OC04	99	+	0.3255	0.2350	.1691986	0.8366
CC01	99	+	0.4775	0.3953	.1622314	0.8297
CC02	99	+	0.4952	0.4200	.1622852	0.8286
CC03	99	+	0.4877	0.4016	.161076	0.8294
CC04	99	+	0.4160	0.3339	.1655502	0.8323

Berdasarkan hasil uji validitas diatas, kita bisa melihat pada nilai *item rest correlatin*. Untuk menentukan validitas setiap pertanyaan yaitu dengan membandingkan nilai *item-rest correlation* dengan nilai *r-tabel* , yang ditentukan berdasarkan nilai *degree of freedom* yaitu $n-2$ (dimana n adalah jumlah responden/observasi). Dalam hal ini jumlah responden 99 ($n=99$). Maka $N (=99) - 2 = 97$. Kita dapatkan nilai r tabel nya 0.1663 dengan taraf signifikasni 0.05. Selanjutnya membandingkan nilai *ir-cor* masing-masing item di tabel diatas dengan r tabel (0.1663). Maka, hasil menunjukkan semua item kuesioner

memiliki nilai ir-cornya lebih besar dari pada r-tabel, artinya semua item menunjukkan valid.

Hasil ini juga bisa kita lakukan analisis model SEM dengan aplikasi STATA seperti gambar 7.1 berikut:



Gambar 7.1 Hubungan 3 variabel laten HC, ICT dan FP

Perintah yang digunakan untuk membuat model seperti di Gambar 7.1 adalah dengan alur seperti alur yang ditunjukkan di Gambar 7.2 yaitu klik menu **statistic** dengan sorot dan pilih **SEM (struktural equation modeling)**, sorot dan pilih dengan meng klik **model building and estimation.**, maka akan muncul seperti pada gambar 7.2.

Faisol\SENMEA 2020\Pelatian STATA\UIN\kuesioner.dta

Statistics User Window Help

	st	average interitem covariance	alpha
Summaries, tables, and tests			
Linear models and related			
Binary outcomes			
Ordinal outcomes			
Categorical outcomes			
Count outcomes			
Fractional outcomes	2	.1534191	0.8209
Generalized linear models	5	.1582962	0.8248
Choice models	9	.1580604	0.8255
Time series	6	.1662726	0.8344
Multivariate time series	0	.1682353	0.8357
Spatial autoregressive models	1	.1535918	0.8215
Longitudinal/panel data	0	.1579645	0.8266
Multilevel mixed-effects models	3	.1673432	0.8345
Survival analysis	3	.1588981	0.8272
Epidemiology and related	8	.1630555	0.8313
Endogenous covariates	8	.1591253	0.8265
Sample-selection models	7	.1623307	0.8300
Treatment effects	3	.1609863	0.8301
SEM (structural equation modeling)	2	.1605932	0.8278
LCA (latent class analysis)	8	.156981	0.8243
FMM (finite mixture models)	0	.1691986	0.8366
IRT (item response theory)	3	.1622314	0.8297
Multivariate analysis	0	.1622852	0.8286
Survey data analysis			
Lasso			
Meta-analysis			
Multiple imputation			

Model building and estimation

Testing and CIs

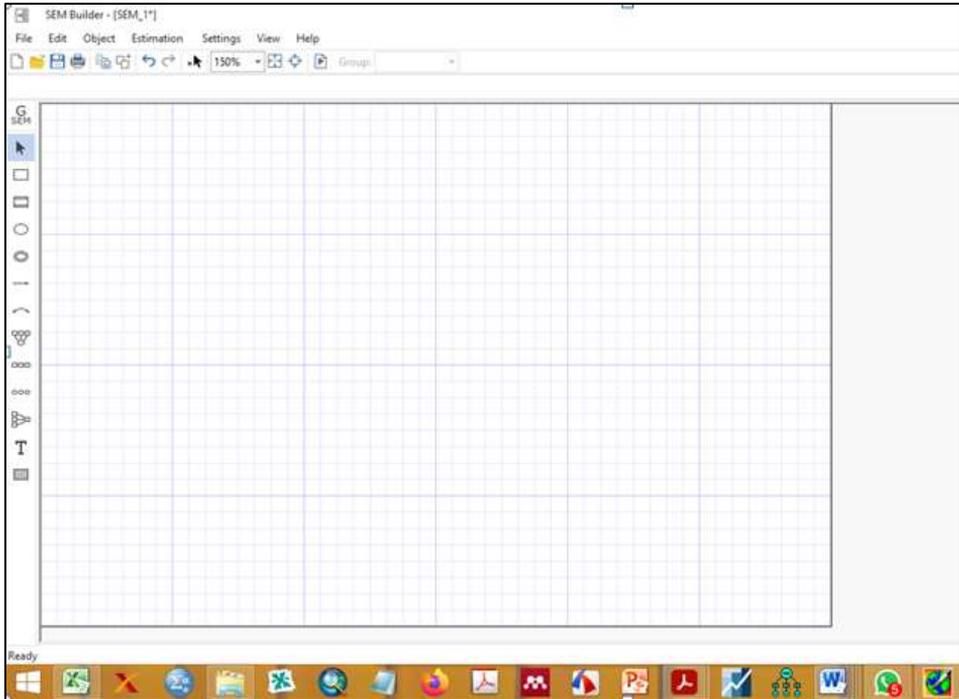
Goodness of fit

Group statistics

Predictions

Other

Gambar 7.2 Alur Analisis SEM



Gambar 7.3 Layar kerja “SEM Builder”

7.3. Uji Reliabilitas

Uji reabilitas , perintah yang digunakan adalah ketik *alpha* dan di ikuti item-item yang diuji, seperti berikut

```
alpha FP01 FP02 FP03 FP04 ICT01 ICT02 ICT03 ICT04 HC01 HC02  
HC03 HC04 OC01 OC02 OC03 OC04 CC01 CC02 CC03 CC04 CC01 CC02  
CC03 CC04
```

hasil output uji reliabilitas,
Test scale = mean(unstandardized items)

```
Average interitem covariance:      .1612747  
Number of items in the scale:      20  
Scale reliability coefficient:      0.8361
```

Latihan: lakukan sebuah survey kecil dengan 10 pertanyaan, uji validitas dan realibilitas atas survey tersebut

DAFTAR REFERENSI

- Gujarati, Damodar N., dan Porter, Dawn C. 2012. *Basic Econometrics: 5th Edition*. New York: McGraw Hil.
- Henky Latan. 2014. *Aplikasi Analisis Data Statistik untuk Ilmu Sosial Sains dengan STATA*. Bandung, Alfabeta
- Imam Ghozali dan Dwi Ratmono. 2013. *Analisis Multivariat dan Ekonometrika Teori, Konsep, dan Aplikasi dengan Eviews*
- Suwardi, A. 2011. *Modul STATA: Tahapan dan Perintah (Syntax) Data Panel*. Depok: Lab Komputasi Departemen Ilmu Ekonomi FEUI.
- Yappy, Benedict J., dan Ihsanuddin, J. 2014. *Pelatihan Komputasi dengan STATA: Modul A Manajemen Data dan Analisis Deskriptif*. Diunduh dari <https://benconomy.files.wordpress.com/2014/02/modul-a2.pdf> (06/1/2020)
- Yappy, Benedict J., dan Ihsanuddin, J. 2014. *Pelatihan Komputasi dengan STATA: Modul B Manajemen Data dan Analisis Deskriptif*. Diunduh dari <https://benconomy.files.wordpress.com/2014/02/modul-a2.pdf> (06/11/2020)
- <http://www.stata.com>
- Teguh I. Maulana, dan Pyan P. S. Amin Muchtar. 2018 Modul Metode Penelitian Akuntansi Politeknik Keuangan Negara STAN <https://staff.blog.ui.ac.id/martani/files/2018/09/Modul-STAN-1.docx> (31/12/2020)

BIODATA PENULIS



Dr. Faisol, S.Pd., M.M. lahir di Kabupaten Kediri, pada tanggal 12 April 1969. Anak ketujuh dari 11 bersaudara dengan ayah bernama Bapak H. Moh. Syafi'i Hambali (almarhum) dan Ibu bernama Hj. Siti Salamah.

Pendidikan Madrasah Ibtidaiyah Negeri Kediri (MIN) sampai Madrasah Aliyah Negeri I diselesaikan di Kabupaten Kediri, Lulus dari MIN pada tahun 1982, lulus dari MTsN I Kota Kediri pada tahun 1985, dan lulus dari MAN I Kabupaten Kediri pada tahun 1988. Pendidikan Program Sarjana (S1) di Fakultas Pendidikan dan Keguruan IKIP PGRI Kota Kediri diselesaikan pada tahun 1993. Pendidikan Program Pascasarjana (S2) di Fakultas Ekonomi dan Bisnis Universitas Islam Kediri diselesaikan pada tahun 2011. Program Doktor Ilmu Ekonomi di Fakultas Ekonomi dan Bisnis Universitas Brawijaya Malang diselesaikan pada tahun 2020. Pengalaman kerja sebagai Dosen Tetap Fakultas Ekonomi dan Bisnis di Universitas Nusantara PGRI Kediri sejak tahun 2003 sampai sekarang. Disamping sebagai dosen, beliau aktif menulis berbagai artikel yang dipublikasikan secara nasional maupun internasional serta artikel yang disampaikan melalui seminar ilmiah. Bidang penelitian yang pernah dilakukan adalah bidang ekonomi regional, makro ekonomi, ekonomi publik, keuangan syariah, SMEs, dan MSDM. Menikah dengan Lisa Lutfiatin, S.Si dan dikaruniai 3 orang anak, yaitu Fadhlana Ni'ma Najwah, Yasmin Marsha Syahida, dan Ma'dan Falakh.

Dr. Agus Eko Sujianto, SE.,MM, Dosen Pascasarjana IAIN Tulungagung. Sebagai lektor kepala penulis aktif memberi perkuliahan pada Fakultas Ekonomi dan Bisnis Islam (S1) dan Pascasarjana (S2 dan S3). Disamping itu, penulis juga aktif menulis artikel baik penelitian maupun pengabdian kepada masyarakat yang dipublikasikan pada jurnal nasional dan internasional.