

STATA FOR BEGINNERS

Modul Pelatihan Universitas Trilogi

4 Februari 2019



BAB I.

PENGENALAN STATA DAN PERTAMA KALI MENGGUNAKAN STATA

1.1 Informasi Singkat STATA

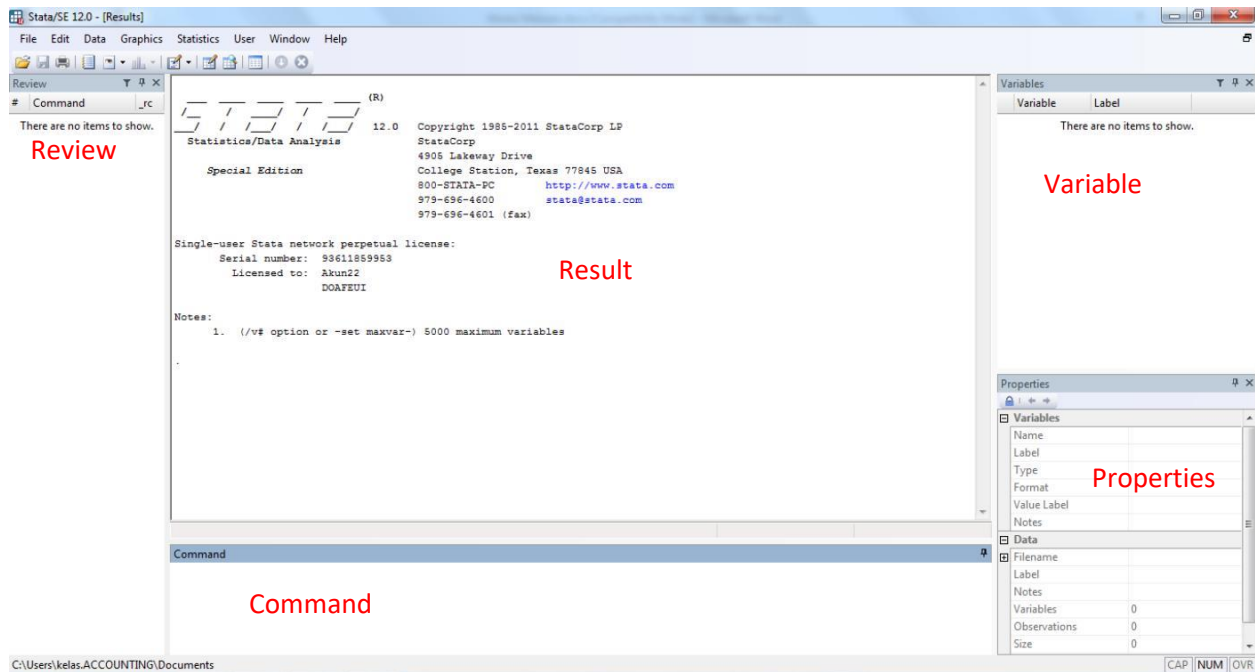
STATA adalah suatu perangkat lunak statistika terintegrasi dan lengkap yang menyediakan berbagai fitur untuk analisis data, manajemen data, dan grafik. STATA menawarkan fitur analisis yang sangat komprehensif untuk keperluan analisis, seperti:

- a. Statistika Dasar
- b. Model Linear
- c. Data Panel/Longitudinal
- d. Model generalized linear
- e. ANOVA/MANOVA
- f. Analisis Time Series
- g. Analisis Survival
- h. Analisis Baynesian
- i. Structural Equation Modeling (SEM)
- j. Metode Survey
- k. Forecasting
- l. Serta berbagai fitur lainnya yang dapat dilihat secara lengkap di alamat berikut ini <http://www.stata.com/features/>

Perangkat lunak yang dikeluarkan Perusahaan di Amerika Serikat ini saat ini memiliki versi 14 sebagai versi terbarunya. STATA banyak digunakan oleh para peneliti dari berbagai belahan dunia karena memiliki *interface* yang sangat ramah pengguna.

1.2 Tampilan STATA

Pada bagian ini akan dijelaskan pengenalan singkat mengenai tampilan utama STATA. 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*.



Gambar 1.1. Tampilan Utama STATA

Menu *data* digunakan untuk melakukan berbagai tindakan/manipulasi terkait data yang digunakan termasuk didalamnya untuk pengaturan jenis data, pembuatan variabel-variabel, dan pembuatan matrix. 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.

Pada tampilan utama STATA, terdapat lima panel yakni:


- Results* (panel bagian tengah), yang menunjukkan hasil pengolahan perintah yang telah dijalankan.
- Command* (panel bagian bawah), yang digunakan untuk memasukkan perintah-perintah pengolahan/analisis.
- 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.
- Variables* (panel kanan atas), menunjukkan daftar variabel yang digunakan dalam set data yang sedang diolah.

- e. *Properties* (panel kanan bawah), berisikan informasi/karakteristik dari variable 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.

1.3 Perintah Dasar Terkait 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*.
 - 3. *Paste* data yang sebelumnya telah di-*copy*.

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, 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, 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.

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

Setelah data dimasukan 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*.

1.4 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, unitoptions**.

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***.

- Data *string* dapat memiliki nilai berbentuk huruf, tetapi tidak dapat dilakukan operasi matematis/statistik.
- 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:

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

```
. describe
```

Contains data

```
obs:      154
vars:      11
size:     20,482
```

	storage	display	value
variable name	type	format	label
Individu	int	%10.0g	Individu
Tahun	int	%10.0g	Tahun
Identifier	str7	%9s	Identifier
CompanyName	str49	%49s	Company Name
GICSIndustryName	str32	%32s	GICS Industry Name
DFIN	double	%10.0g	DFIN
PTROA	double	%10.0g	PTROA
LEV	double	%10.0g	LEV
SIZE	double	%10.0g	SIZE
DKomAk_KA	byte	%10.0g	DKomAk_KA
ETR	double	%10.0g	ETR

Gambar 1.2. 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 “lnadv” yang merupakan hasil operasi logaritma natural dari variabel “adv”, maka perintah yang dimasukan:

Generate lnadv = ln(adv)

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:

- Ganti nama variable DKomAk_KA menjadi DKA
- Buatlah variabel baru DFINKA yang merupakan perkalian dari variable DFIN dan DKA

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 “DKA”, untuk mempermudah dapat ditambahkan informasi “latar belakang komite audit” pada label variabel sehingga pengguna mengetahui dengan jelas bahwa “DKA” merupakan variabel yang berisikan latar belakang pendidikan anggota komite audit.

BAB II. ANALISIS DESKRIPTIF

Sebelum melakukan analisis utama terhadap data yang dimiliki, seorang yang sedang melakukan penelitian perlu memahami karakteristik dari data yang digunakan sehingga dapat lebih tajam dalam penarikan kesimpulan yakni dengan menggunakan analisis deskriptif. Dalam statistika deskriptif hanya akan dilakukan upaya 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, median, modus.
- 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

Variable	Obs	Mean	Std. Dev.	Min	Max
Tahun	154	2013.5	.5016313	2013	2014
Identifier	0				
Individu	154	39	22.29863	1	77
CompanyName	0				
GICSIndust~e	0				
GICSSector~e	0				
DFIN	154	.0035063	.1157588	-.6138974	.4492697
PTROA	154	.0870754	.1155064	-.1972726	.5828636
LEV	154	.2759935	.2124637	0	1.066142
SIZE	154	27.9612	1.669777	24.14989	33.09497
DKA	154	.7012987	.4591819	0	1
ETR	154	.2462326	.2659673	-.0006402	2.280867

Gambar 2.1. Hasil Perintah *Summarize*

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

DFIN				
	Percentiles	Smallest		
1%	-.3753061	-.6138974		
5%	-.1496383	-.3753061		
10%	-.1023661	-.2821642	Obs	154
25%	-.0370799	-.2734996	Sum of Wgt.	154
50%	.0011132		Mean	.0035063
		Largest	Std. Dev.	.1157588
75%	.0518107	.2705645		
90%	.1129188	.272181	Variance	.0134001
95%	.1675777	.3714365	Skewness	-.5990937
99%	.3714365	.4492697	Kurtosis	9.838495
ETR				
	Percentiles	Smallest		
1%	-.0005401	-.0006402		
5%	0	-.0005401		
10%	0	0	Obs	154
25%	.0980674	0	Sum of Wgt.	154
50%	.228834		Mean	.2462326
		Largest	Std. Dev.	.2659673
75%	.2797667	.961134		
90%	.4010261	.9636932	Variance	.0707386
95%	.8865912	1.032363	Skewness	3.738959
99%	1.032363	2.280867	Kurtosis	25.18173

Gambar 2.2. Contoh Hasil Perintah: Summarize DFIN, ETR, detail

- c. Jika hendak menayangkan data dalam bentuk daftar, dapat menggunakan perintah ***list***.
- d. Untuk menunjukan tabulasi suatu atau dua variable (*dummy*), dapat menggunakan perintah

tabulate nama_variabel untuk tabulasi satu variable

tabulate DKA

. tabulate DKA

DKomAk_KA	Freq.	Percent	Cum.
0	46	29.87	29.87
1	108	70.13	100.00
Total	154	100.00	

Gambar 2.4. Contoh tabulasi satu variabel

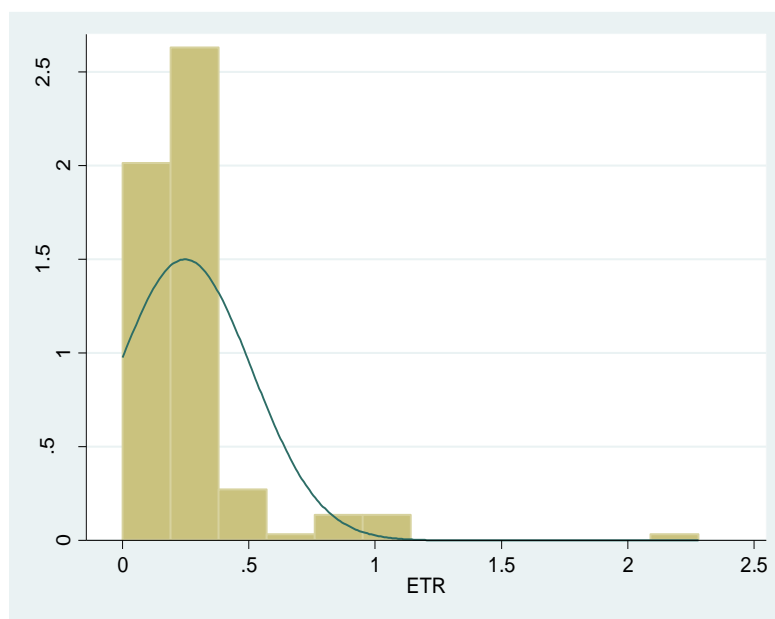
atau perintah tabulate ***nama_variabel1 nama_variabel2***, untuk tabulasi antar dua variabel

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

2.2 Menggunakan Tabel dan Grafik Statistika Deskriptif pada STATA

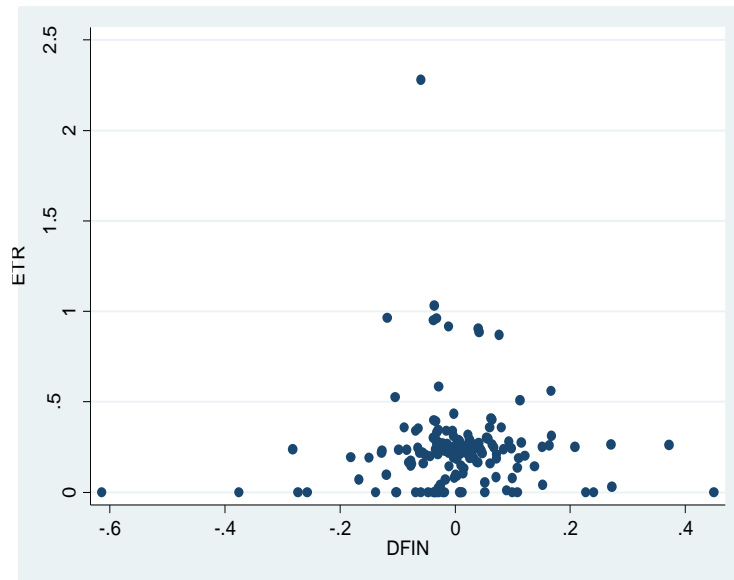
Selain menggunakan nilai-nilai statistic seperti yang ditunjukkan pada bagian sebelumnya, analisis deskriptif juga dapat dilakukan menggunakan bantuan tabel/grafik/histogram. Penggunaan tabel/grafik/histogram memberikan kemudahan bagi pengguna untuk mendapatkan visualisasi pemusatan dan penyebaran data 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 menunjukan frekuensi setiap kelompok histogram. Perintah ***discrete*** dapat digunakan untuk mengetahui diagram batang untuk setiap nilai.

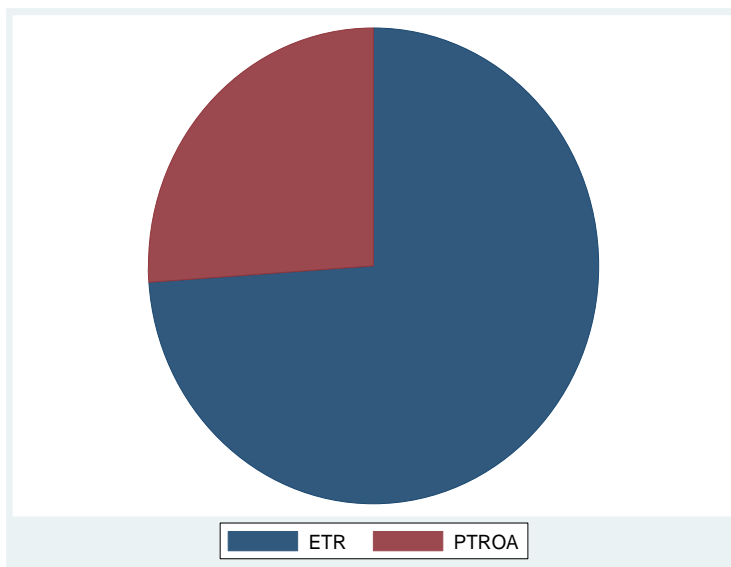


- b. Jika hendak melihat grafik antarvariabel, dapat menggunakan perintah ***twoway jenis_plot nama_variabel1 nama_variabel2***. Terdapat dua jenis plot yang dapat digunakan yakni plot titik (***scatter***) dan garis (***line***). Untuk menambahkan garis tren dapat ditambahkan pilihan opsi ***lfit*** (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 detil terkait penggunaan opsi-opsi ini.



- c. Perintah **graph** dapat digunakan untuk memperlihatkan suatu grafik. Contoh untuk menunjukkan grafik pie untuk membandingkan variable ETR dan PTROA.



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:

dengan:

Y = Variabel dependen yang hendak diprediksi

X = Variabel independen

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

β = 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.

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

- a. Parameter menghasilkan varians error yang paling minimal (**B**est)
- b. Persamaan yang diestimasi adalah persamaan linear (garis lurus) (**L**inear)
- c. Parameter estimasi tidak bias (**U**nbiased)
- d. Parameter mampu untuk mengestimasi secara baik parameter populasi (**E**stimator)

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 salah
- 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 Z-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 H_0 ketika H_0 Benar
- b. Kesalahan tipe II (Type II Error), menerima H_0 ketika H_0 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

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

Maka sintaks STATA yang digunakan adalah

reg (atau regress) yield or nonor model alliance adv size lev gnp reg1 reg2 reg3

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

. reg ETR DFIN PTROA LEV SIZE

Source	SS	df	MS	Number of obs	=	154
Model	.257033926	4	.064258481	F(4, 149)	=	0.91
Residual	10.5659688	149	.070912542	Prob > F	=	0.4621
				R-squared	=	0.0237
				Adj R-squared	=	-0.0025
Total	10.8230027	153	.07073858	Root MSE	=	.26629

ETR	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
DFIN	.0252263	.1975479	0.13	0.899	-.3651309	.4155834
PTR0A	-.098037	.2089651	-0.47	0.640	-.5109548	.3148808
LEV	.0101293	.1082672	0.09	0.926	-.2038081	.2240667
SIZE	.0240623	.0131441	1.83	0.069	-.0019107	.0500353
_cons	-.4209266	.3631306	-1.16	0.248	-1.138477	.2966243

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

- 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* (Prob > F) untuk dibandingkan dengan α .
- Menunjukkan *goodness of fit* dari suatu model yakni seberapa banyak variasi yang terjadi pada variabel dependen dapat dijelaskan menggunakan model yang diujikan.
- Nilai konstanta (α), yakni nilai variabel dependen ketika seluruh variabel independen bernilai = 0.
- Menunjukkan koefisien estimasi (untuk setiap variabel independen yang
- digunakan pada model. Tanda + atau – menunjukkan arah pengaruh variabel independen terhadap variabel dependen.
- Menunjukkan standar deviasi dari koefisien estimasi

- 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.
- Alternatif untuk melakukan pengujian untuk mengetahui signifikansi suatu variabel. Nilai ini akan dibandingkan dengan nilai α .
- 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 seperti contoh di atas!

3.4. Pendeteksian dan Pengoreksian Pelanggaran Asumsi

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 indikasi model yang digunakan memiliki masalah multikolinearitas.
2. Menguji korelasi antarvariabel yang digunakan dalam model. Suatu variabel dikatakan memiliki korelasi yang kuat jika memiliki korelasi (lebih besar dari $\pm 0,8$) terhadap

variabel lainnya. Jika menggunakan STATA pengujian korelasi ini dapat menggunakan perintah:

```
. corr ETR DFIN PTROA LEV SIZE
(obs=154)
```

	ETR	DFIN	PTROA	LEV	SIZE
ETR	1.0000				
DFIN	0.0093	1.0000			
PTROA	-0.0238	0.3085	1.0000		
LEV	0.0380	0.0391	-0.2919	1.0000	
SIZE	0.1478	0.0739	0.1175	0.1129	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-variabel yang digunakan dalam model seperti contoh di atas! Silakan interpretasikan hasil pengujian tersebut!

- 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 nama_variabelx, sig

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

	— ETR	DFIN	PTROA	LEV	SIZE
ETR	1.0000				
DFIN	0.0093 0.9086	1.0000			
PTROA	-0.0238 0.7696	0.3085 0.0001	1.0000		
LEV	0.0380 0.6398	0.0391 0.6303	-0.2919 0.0002	1.0000	
SIZE	0.1478 0.0674	0.0739 0.3624	0.1175 0.1466	0.1129 0.1631	1.0000

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**.

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

Variable	VIF	1/VIF
PTROA	1.26	0.795558
LEV	1.14	0.875927
DFIN	1.13	0.886296
SIZE	1.04	0.962166
Mean VIF	1.14	

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

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

Source	chi2	df	p
Heteroskedasticity	19.06	14	0.1626
Skewness	7.82	4	0.0986
Kurtosis	1.21	1	0.2715
Total	28.09	19	0.0818

Gambar 3.4. Contoh hasil pengujian white

3. Menggunakan pengujian Breush-Pagan / Godfrey, dapat menggunakan sintaks:

estat hettest

berikut contoh hasil pengujian white untuk model yang diujikan sebelumnya

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

chi2(1)      =    50.32
Prob > chi2   =    0.0000
```

Gambar 3.5. Contoh hasil pengujian Breush-Pagan

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

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, kemudian lakukan langkah koreksi atas pelanggaran asumsi heterokedastisitas yang ditemukan!

c. Autokorelasi

Autokorelasi adalah kondisi terdapat hubungan error antarwaktu pada data yang digunakan. Kasus autokorelasi banyak ditemukan ketika menggunakan data *time-series* juga data *panel*. 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)

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:

H0: error terdistribusi normal

H1: error tidak terdistribusi normal

Pengujian dapat dilakukan dengan sintaks:

swilk nama_variabel (shapiro-wilk)

sfrancia nama_variabel (shapiro-francia)

ksmirnov nama_variabel (Kolmogorov-Smirnov)

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 seperti ***ln-skewness*** atau ***box-cox***, dengan sintaks:

lnskew variabel_baru = ekspresi

bcskew variabel_baru = ekspresi

3. 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

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.