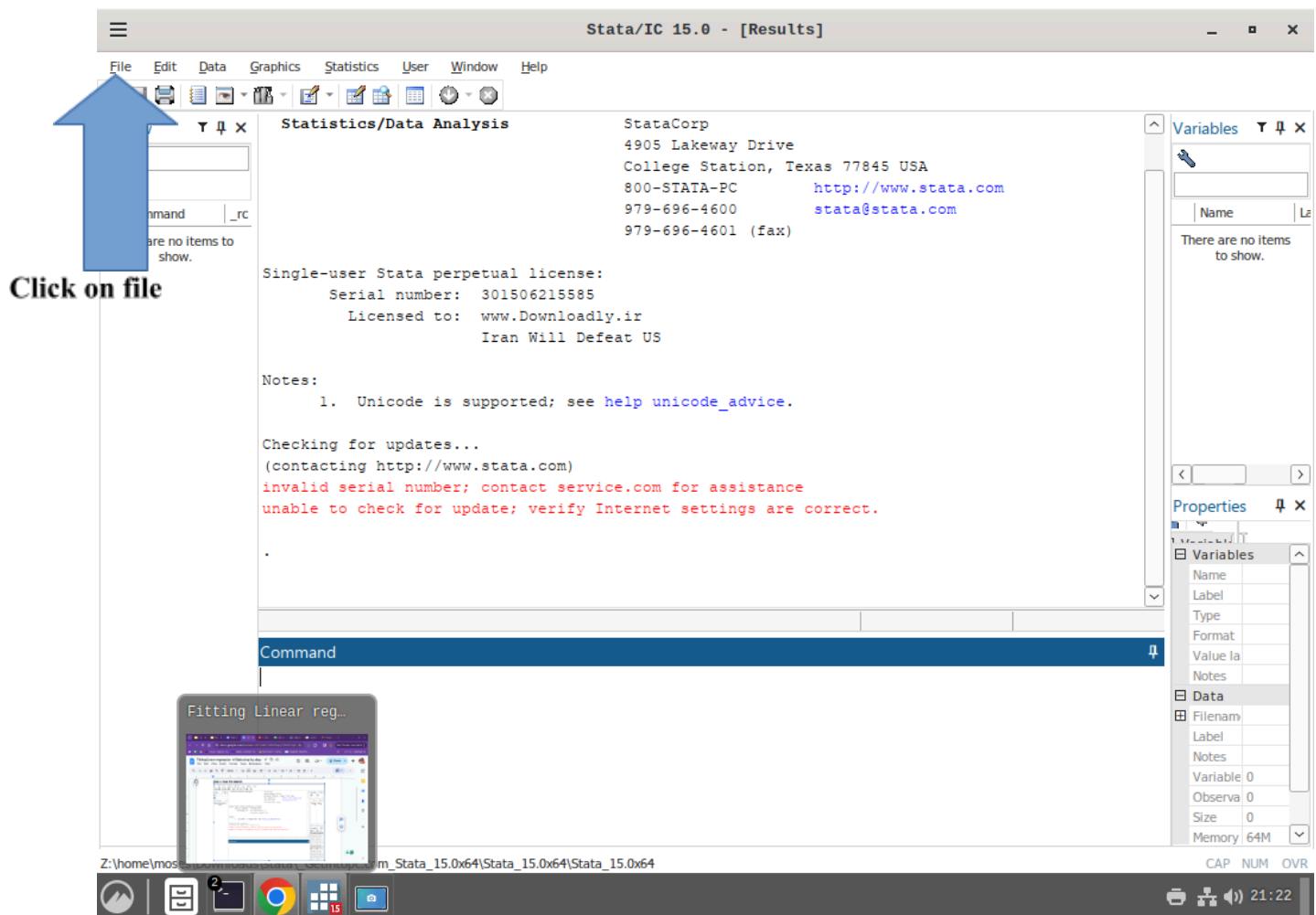
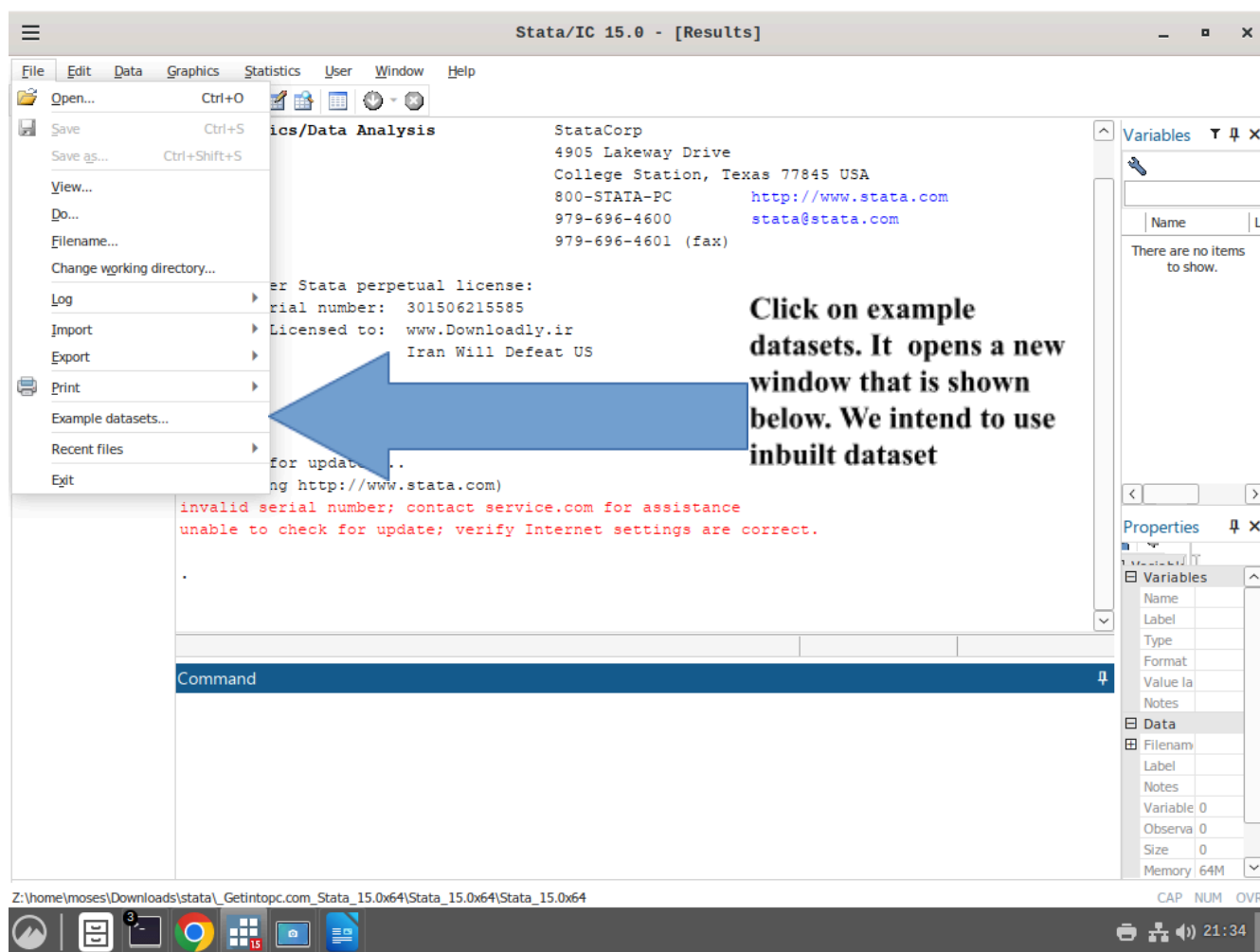


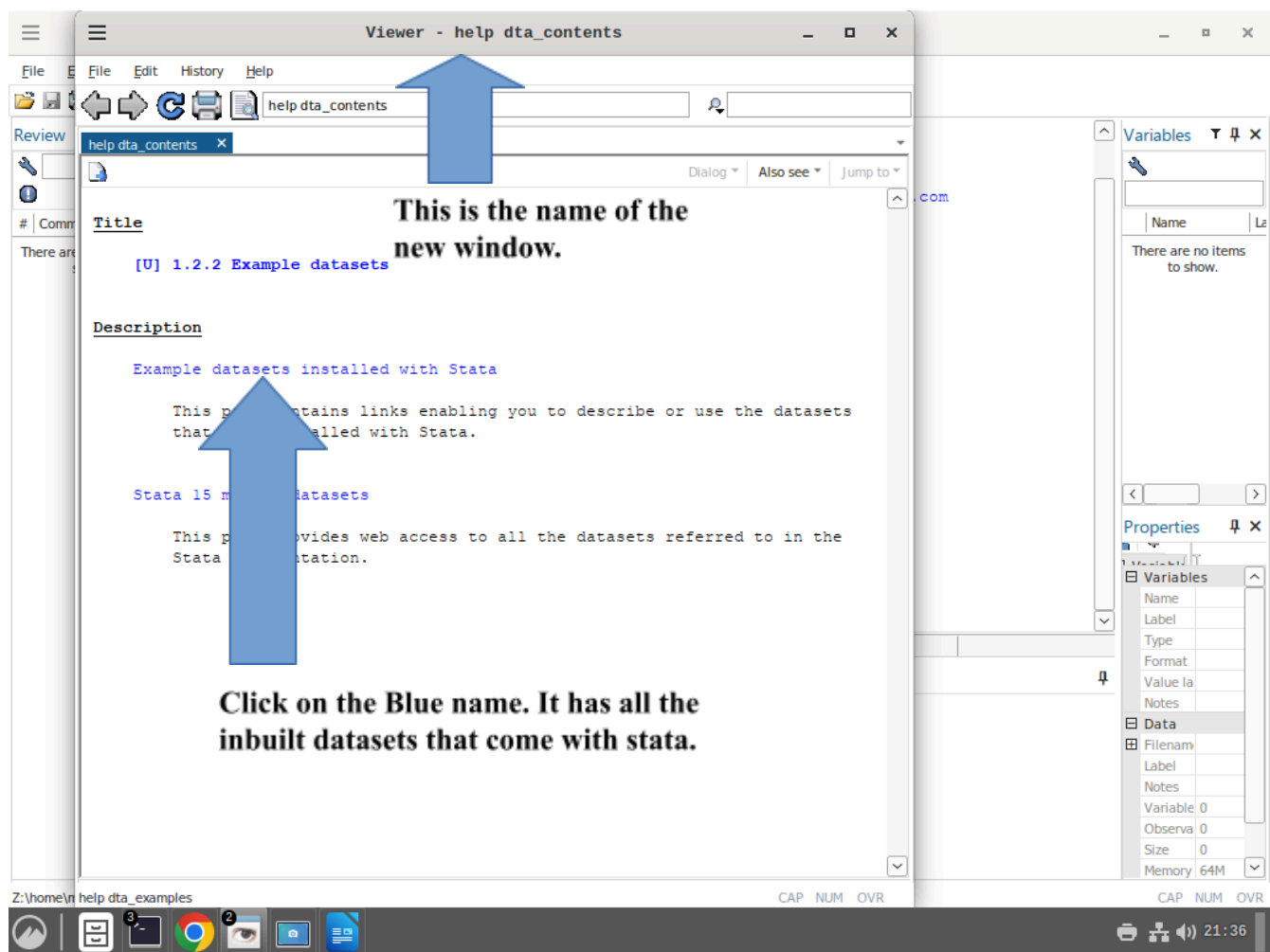
Step 1: load the data



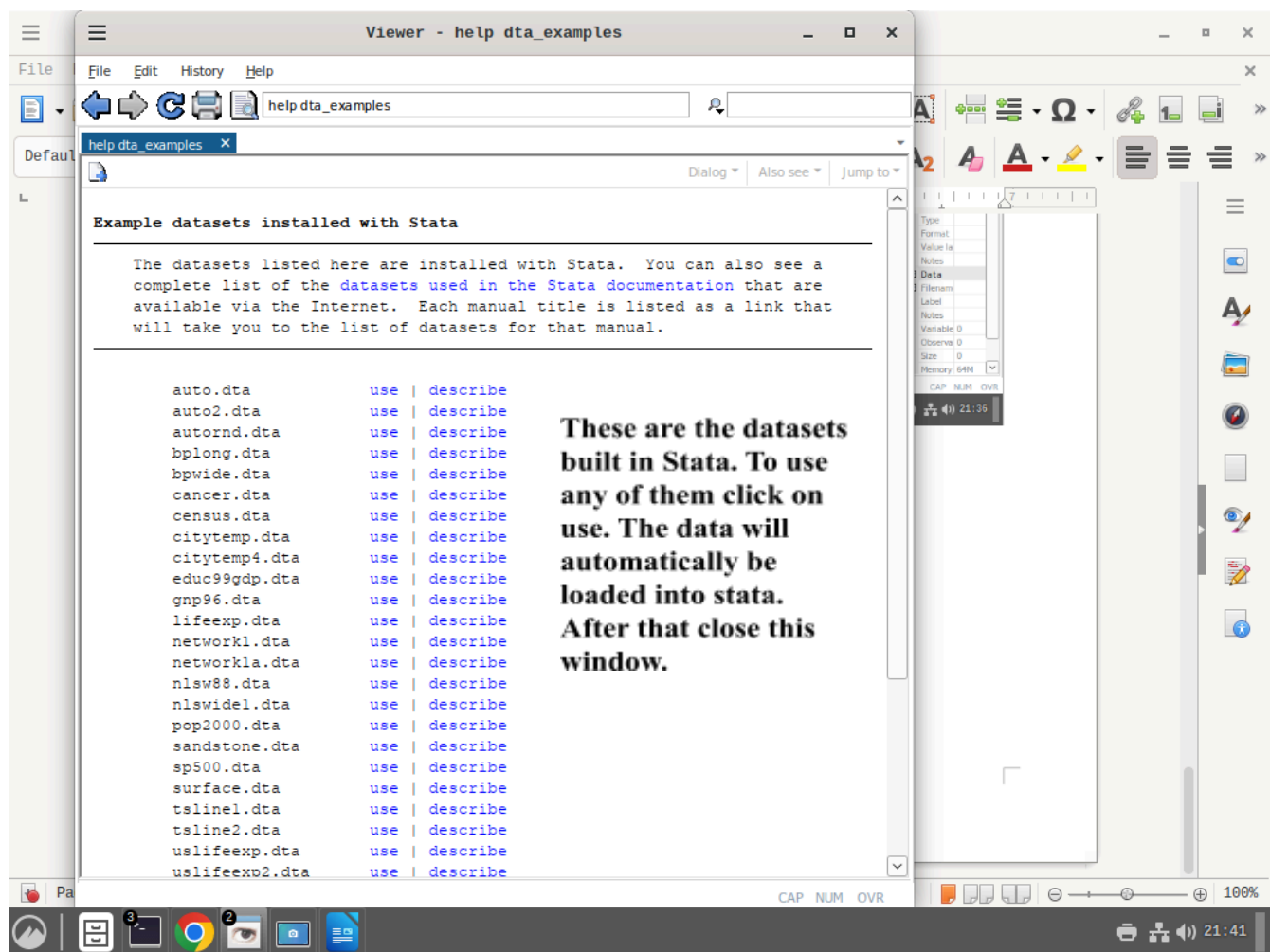
Now we have to get the inbuilt dataset



Lets go to the new windows that opens up after this



Let's view the datasets



How do we know that the dataset has been loaded? We selected auto.dta

Stata/IC 15.0 - Z:\home\moses\Downloads\stata_Getintopc.com_Stata_15.0x64\Stata_15.0x64\Stata_15.0x64\ado\bas...

File Edit Data Graphics Statistics User Window Help

Review

4905 Lakeway Drive
College Station, Texas 77845 USA
800-STATA-PC <http://www.stata.com>
979-696-4600 stata@stata.com
979-696-4601 (fax)

Single-user Stata perpetual license:
Serial number: 301506215585
Licensed to: www.Downloadly.ir
Iran Will Defeat US

Notes:
1. Unicode is supported; see [help unicode](#).

Checking for updates...
(contacting <http://www.stata.com>)
invalid serial number; contact [service.com](#) for assistance
unable to check for update; verify Internet settings are correct.

. sysuse auto.dta
(1978 Automobile Data)

Com

You will see variable names on this part. Like price, The column names are the Names, and labels describe the data.

This is another hint that shows the data has been loaded. You will see the code here and also some basic information about the dataset.

Variables

Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 1978
headroom	Headroom (in.)
trunk	Trunk space (cu. ft.)
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu. ...
gear_ratio	Gear Ratio
foreign	Car type

Properties

Variables

Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 1978
headroom	Headroom (in.)
trunk	Trunk space (cu. ft.)
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu. ...
gear_ratio	Gear Ratio
foreign	Car type

Data

Filename	auto.dta
Label	1978 Automobile Data
Variables	12
Observations	74
Size	3.11K
Memory	64M

Z:\home\moses\Downloads\stata_Getintopc.com_Stata_15.0x64\Stata_15.0x64\Stata_15.0x64

CAP NUM OVR

21:44

We have the dataset. Lets get the basic descriptive statistics

The screenshot shows the Stata software interface. The 'Statistics' menu is open, displaying a list of statistical options. A blue arrow points to the 'Summaries, tables, and tests' option. Another blue arrow points to the 'Regression' option at the bottom of the menu. The 'Variables' window is open on the right, showing a list of variables with their labels and types. The 'Properties' window is also open, showing details for the selected variable 'auto.dta'.

Statistics Menu Options:

- Summaries, tables, and tests
- Linear models and related
- Binary outcomes
- Ordinal outcomes
- Categorical outcomes
- Count outcomes
- Fractional outcomes
- Generalized linear models
- Time series
- Multivariate time series
- Spatial autoregressive models
- Longitudinal/panel data
- Multilevel mixed-effects models
- Survival analysis
- Epidemiology and related
- Endogenous covariates
- Sample-selection models
- Treatment effects
- SEM (structural equation modeling)
- LCA (latent class analysis)
- FMM (finite mixture models)
- IRT (item response theory)
- Survey data analysis
- Multiple imputation
- Nonparametric analysis
- Multivariate analysis
- Exact statistics
- Resampling

Variables Window:

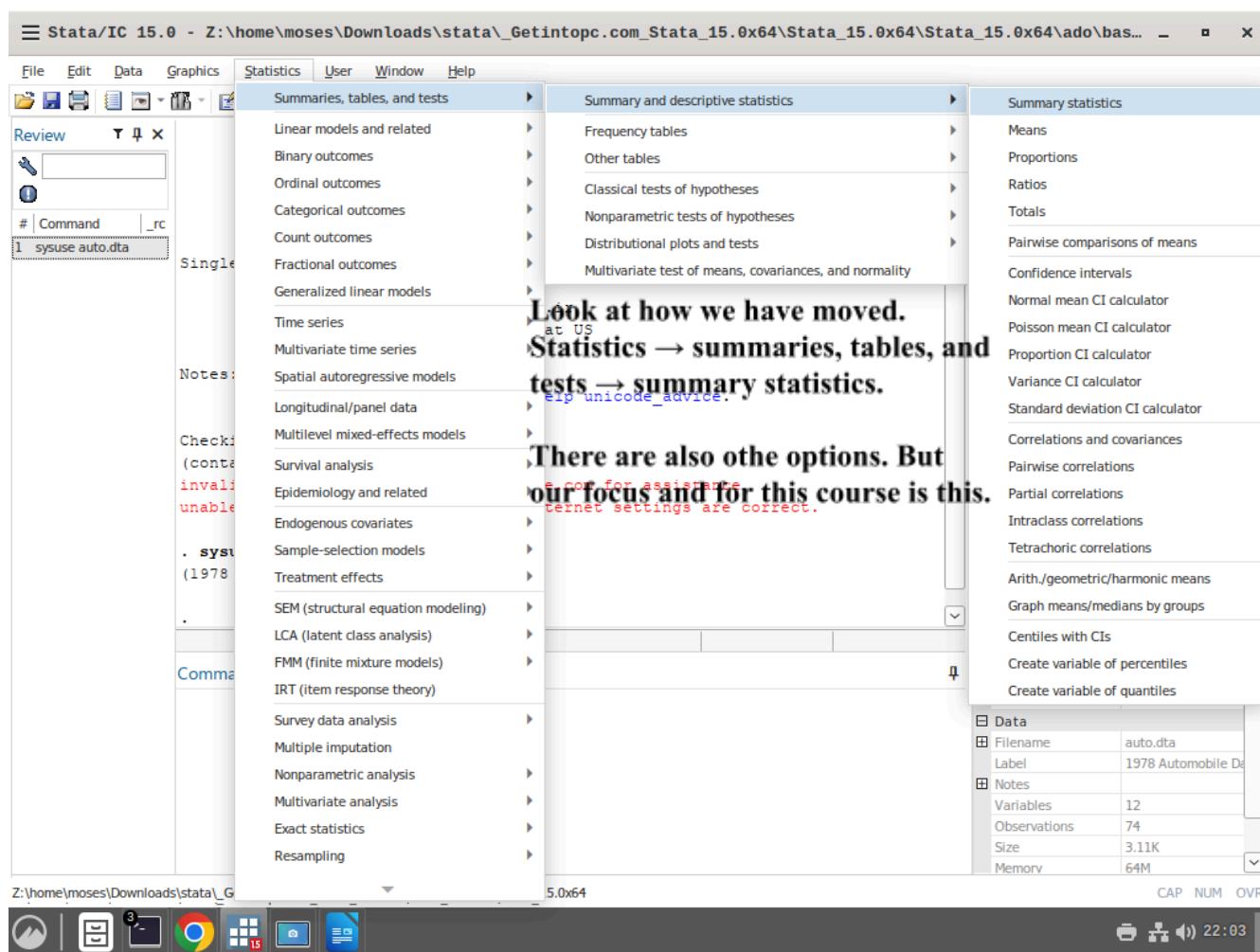
Name	Label	Type
price	Price	Price
mpg	Mileage (mpg)	Price
rep78	Repair Record 1978	Price
headroom	Headroom (in.)	Price
trunk	Trunk space (cu. ft.)	Price
weight	Weight (lbs.)	Price
length	Length (in.)	Price
turn	Turn Circle (ft.)	Price
displacement	Displacement (cu. ...)	Price
gear_ratio	Gear Ratio	Price
foreign	Car type	Price

Properties Window:

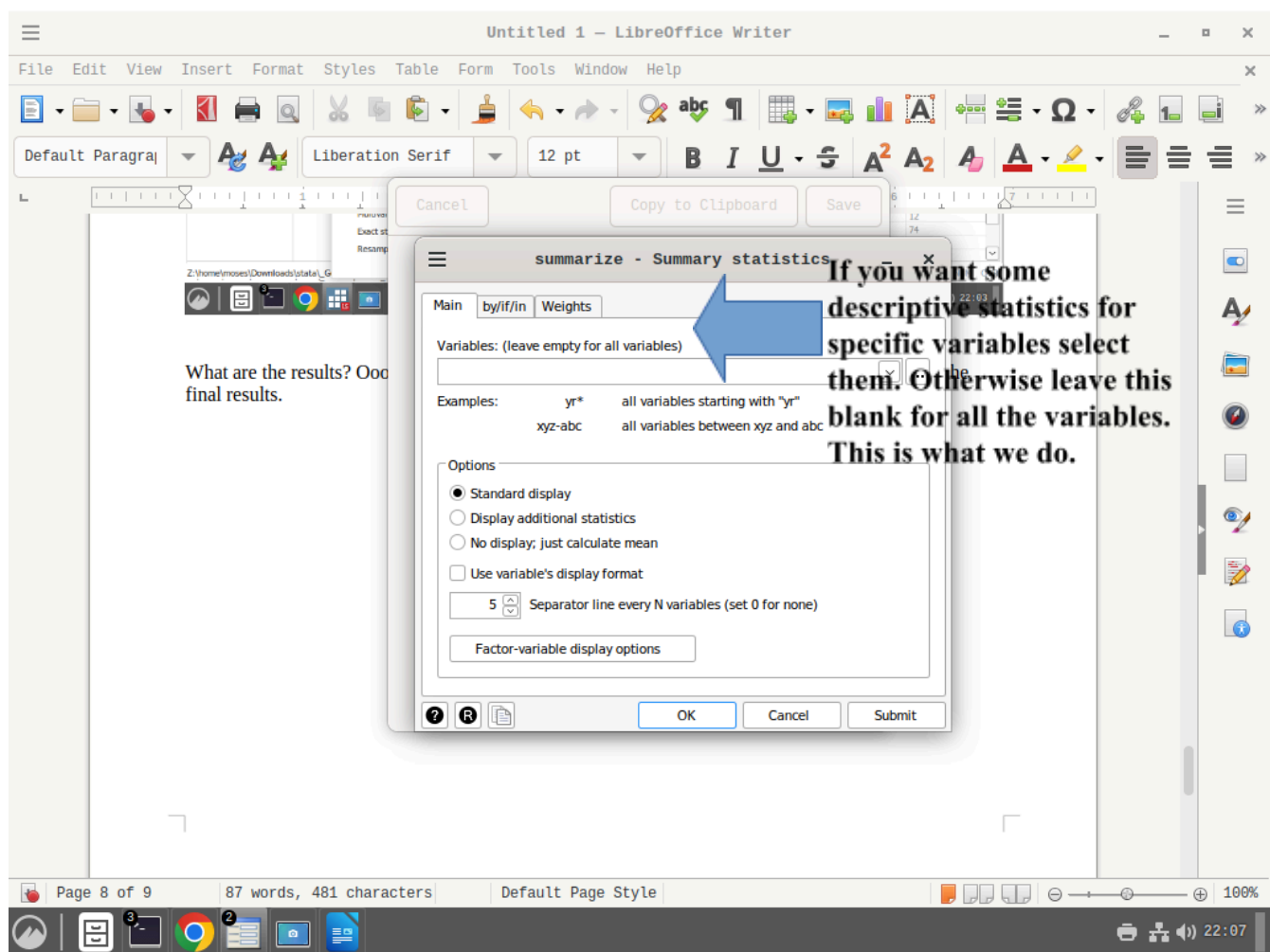
Variable	Value
Filename	auto.dta
Label	1978 Automobile Data
Notes	
Variables	12
Observations	74
Size	3.11K
Memory	64M

This box shows that there are other options. We use it to select the options post estimation analysis. We will use it for regression model diagnostic.

lets do descriptive statistics. We select the first options summaries, tables and tests.



What are the results? Ooops! Not the output but a new small window that you customize before the final results.



The descriptive statistics table.

Stata/IC 15.0 - Z:\home\moses\Downloads\stata_Getintopc.com_Stata_15.0x64\Stata_15.0x64\Stata_15.0x64\ado\bas...

File Edit Data Graphics Statistics User Window Help

Review (1978 Automobile Data)

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	74	6165.257	2949.496	3291	15906
mpg	74	21.2973	5.785503	12	41
rep78	69	3.405797	.9899323	1	5
headroom	74	2.993243	.8459948	1.5	5
trunk	74	13.75676	4.277404	5	23
weight	74	3019.459	777.1936	1760	4840
length	74	187.9324	22.26634	142	233
turn	74	39.64865	4.399354	31	51
displacement	74	197.2973	91.83722	79	425
gear_ratio	74	3.014865	.4562871	2.19	3.89
foreign	74	.2972973	.4601885	0	1

Command

This is the descriptive statistics table. It contains the measures of central tendency and dispersion.

Variables

Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 19
headroom	Headroom (in.)
trunk	Trunk space (cu.
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu
gear_ratio	Gear Ratio
foreign	Car type

Properties

Variables

Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 19
headroom	Headroom (in.)
trunk	Trunk space (cu.
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu
gear_ratio	Gear Ratio
foreign	Car type

Data

Filename	Label
auto.dta	1978 Automobile

Notes

Variables	Observations	Size	Memory
12	74	3.11K	64M

22:10

Fitting a regression model

Stata/IC 15.0 - Z:\home\moses\Downloads\stata_Getintopc.com_Stata_15.0x64\Stata_15.0x64\Stata_15.0x64\ado\bas...

File Edit Data Graphics Statistics User Window Help

Review (1978 Automobile Data)

```
. summarize
```

Statistics

- Summaries, tables, and tests
- Linear models and related
 - Linear regression
 - Regression diagnostics
 - ANOVA/MANOVA
 - Constrained linear regression
 - Nonlinear least-squares estimation
 - Nonparametric regression
 - Censored regression
 - Truncated regression
 - Hurdle regression
 - Heteroskedastic linear regression
 - Endogenous covariates
 - Sample-selection models
 - Box-Cox regression
 - Fractional polynomials
 - Quantile regression
 - Errors-in-variables regression
 - Frontier models
 - Panel data
 - Mixed-effects linear regression
 - Mixed-effects nonlinear regression
 - Spatial autoregressive models
 - Multiple-equation models
 - Treatment effects
- Binary outcomes
- Ordinal outcomes
- Categorical outcomes
- Count outcomes
- Fractional outcomes
- Generalized linear models
- Time series
- Multivariate time series
- Spatial autoregressive models
- Longitudinal/panel data
- Multilevel mixed-effects models
- Survival analysis
- Epidemiology and related
- Endogenous covariates
- Sample-selection models
- Treatment effects
- SEM (structural equation modeling)
- LCA (latent class analysis)
- FMM (finite mixture models)
- IRT (item response theory)
- Survey data analysis
- Multiple imputation
- Nonparametric analysis

Again, we start from statistics → linear models and related → linear regression. What does this do? It opens a small window where we can customize the variables.

Variables

Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 19
headroom	Headroom (in.)
trunk	Trunk space (cu.
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu
gear_ratio	Gear Ratio
foreign	Car type

Properties

Variables

Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 19
headroom	Headroom (in.)
trunk	Trunk space (cu.
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu
gear_ratio	Gear Ratio
foreign	Car type

Data

Filename	Label
auto.dta	1978 Automobile

Notes

Variables	Observations	Size	Memory
12	74	3.11K	64M

Defining the regression model

Stata/IC 15.0 - Z:\home\moses\Downloads\stata_Getintopc.com_Stata_15.0x64\Stata_15.0x64\Stata_15.0x64\ado\bas...

File Edit Data Graphics Statistics User Window Help

Review (1978 Automobile Data)

. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	74				
mpg	74				
headroom	74				
trunk	74				
weight	74				
length	74				
turn	74				
displacement	74				
gear_ratio	74				
foreign	74				

Command

regress - Linear regression

Model by/lf/in Weights SE/Robust Reporting

Dependent variable: price Independent variables: mpg rep78

Treatment of constant

☐ Suppress constant term

☐ Has user-supplied constant

☐ Total SS with constant (adjusted R²)

OK Cancel Submit

Variables

Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
repair	Repair Record 15
headroom	Headroom (in.)
trunk	Trunk space (cu.
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu
ratio	Gear Ratio
car_type	Car type

Notes

Variables	12
Observations	74
Size	3.11K
Memory	64M




what is the

Define the independent variables. Here we have selected two. You can add more. Remember that this can be empty as model will assume a constant.

output?

Stata/IC 15.0 - Z:\home\moses\Downloads\stata_Getintopc.com_Stata_15.0x64\Stata_15.0x64\Stata_15.0x64\ado\bas...

File Edit Data Graphics Statistics User Window Help

Review   

1 sysuse auto.dta
2 summarize
3 regress price
4 regress price 10 198
5 regress price ...

```
. regress price 10
10 invalid name
r(198);

. regress price mpg rep78
```

Source	SS	df	MS	Number of obs	=	69
Model	144754063	2	72377031.7	F(2, 66)	=	11.06
Residual	432042896	66	6546104.48	Prob > F	=	0.0001
				R-squared	=	0.2510
				Adj R-squared	=	0.2283
Total	576796959	68	8482308.22	Root MSE	=	2558.5

	price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
mpg		-271.6425	57.77115	-4.70	0.000	-386.9864 -156.2987
rep78		666.9568	342.3559	1.95	0.056	-16.5789 1350.492
_cons		9657.754	1346.54	7.17	0.000	6969.3 12346.21

Command

This is the Regression model output. This is the table to interpret. For more information look at the regression model interpretation powerpoint.

Variables

Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 1
headroom	Headroom (in.)
trunk	Trunk space (cu.
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu
gear_ratio	Gear Ratio
foreign	Car type

Properties

Variables

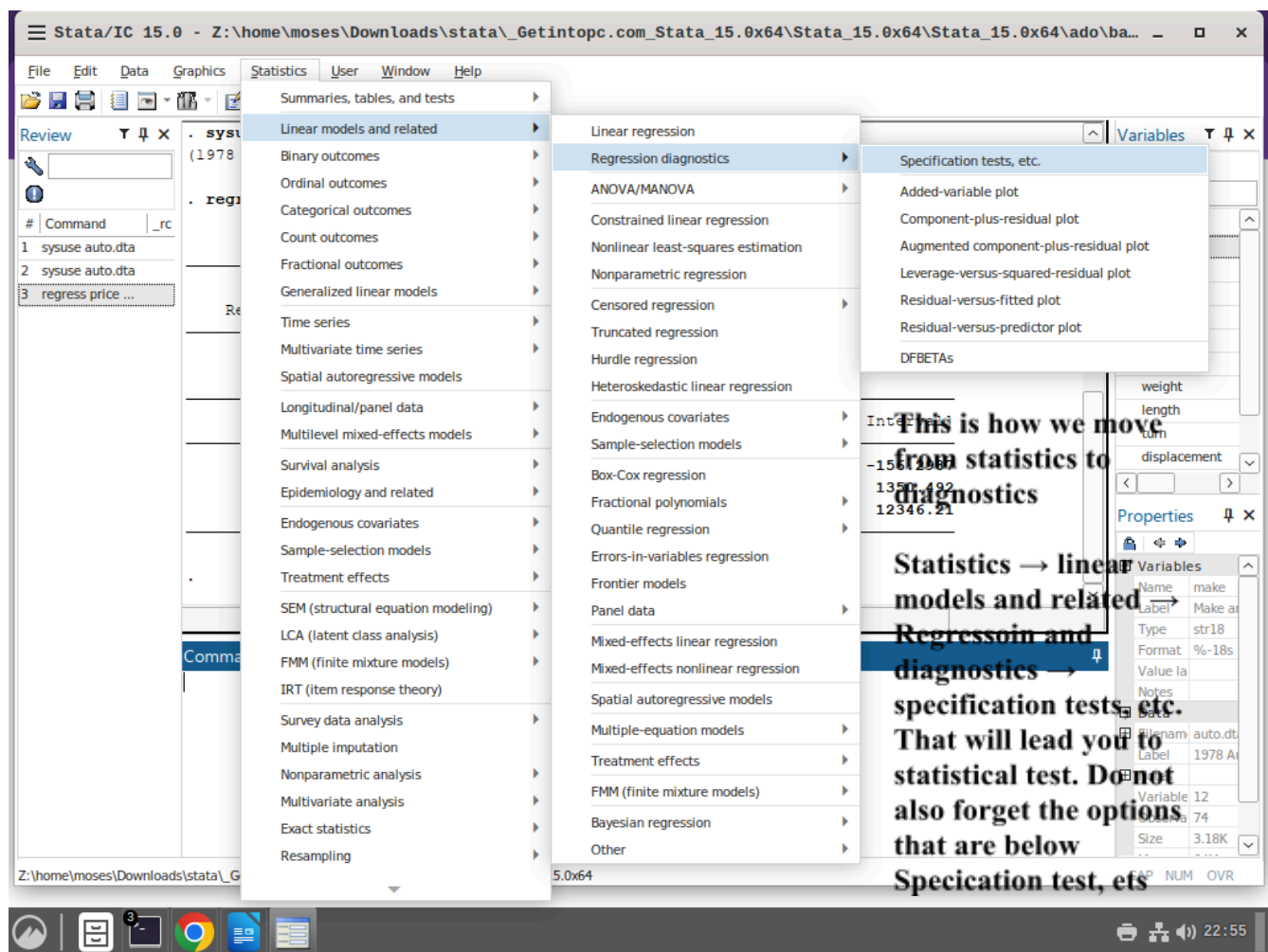
Name	Label	Type	Format	Value label	Notes
Filename	auto.dta				
Label	1978 Automobile				
Variables	12				
Observations	74				
Size	3.18K				
Memory	64M				

Z:\home\moses\Downloads\stata_Getintopc.com_Stata_15.0x64\Stata_15.0x64\Stata_15.0x64

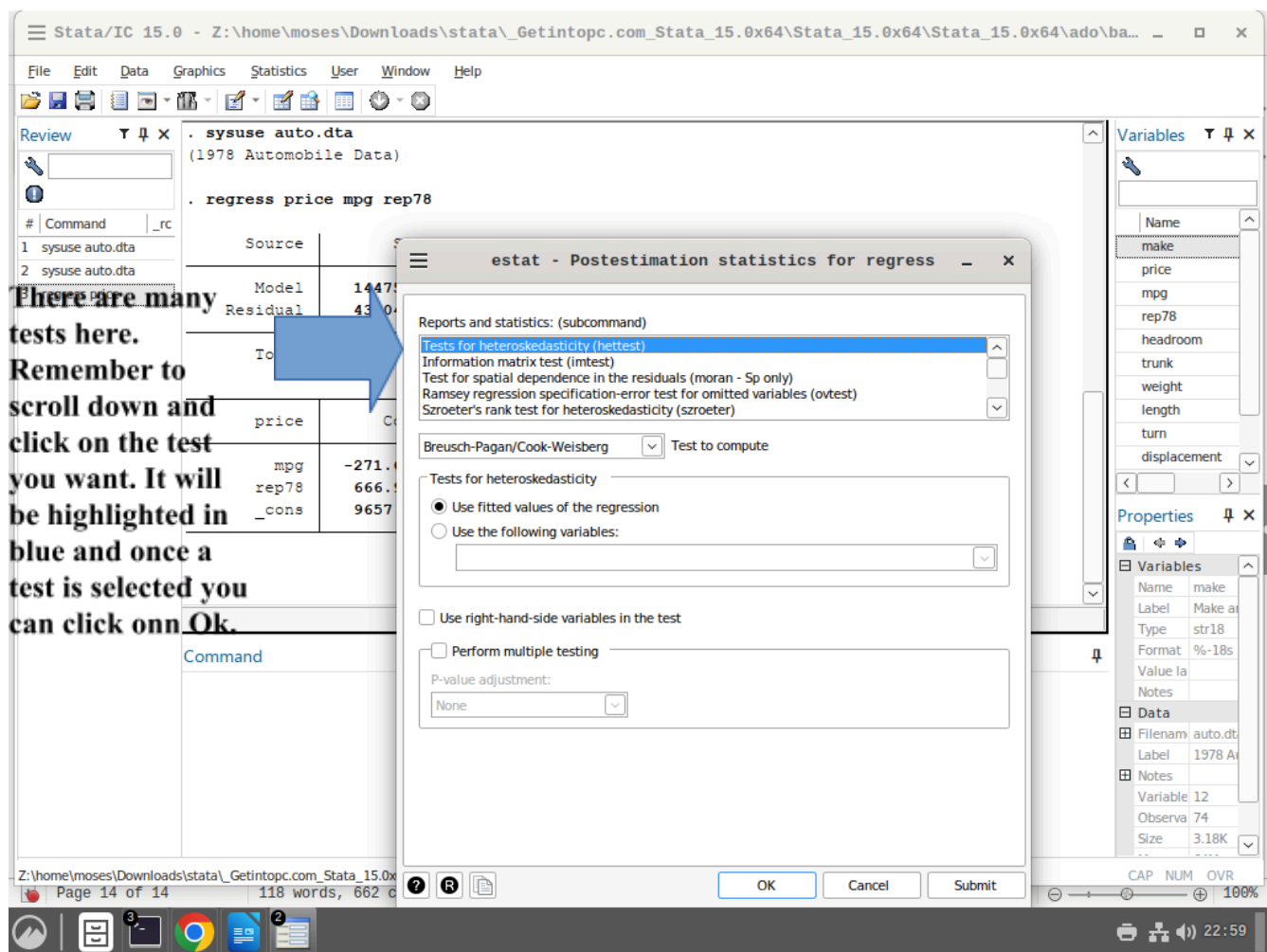
CAP NUM OVR

22:22

Lets go to model diagnostics



lets click on the last part and see the results



Lets end with a summary of the diagnostics tests

In this chapter, we have used a number of tools in Stata for determining whether our data meets the regression assumptions. Below, we list the major commands we demonstrated organized according to the assumption the command was shown to test.

- **Detecting Unusual and Influential Data**

- **predict** — used to create predicted values, residuals, and measures of influence.
- **rvpplot** — graphs a residual-versus-predictor plot.

- **rvfplot** — graphs residual-versus-fitted plot.
- **lvr2plot** — graphs a leverage-versus-squared-residual plot.
- **dfbeta** — calculates DFBETAs for all the independent variables in the linear model.
- **avplot** — graphs an added-variable plot, a.k.a. partial regression plot.

- **Tests for Normality of Residuals**

- **kdensity** — produces kernel density plot with normal distribution overlayed.
- **pnorm** — graphs a standardized normal probability (P-P) plot.
- **qnorm** — plots the quantiles of varname against the quantiles of a normal distribution.
- **iqr** — resistant normality check and outlier identification.
- **swilk** — performs the Shapiro-Wilk W test for normality.

- **Tests for Heteroscedasticity**

- **rvfplot** — graphs residual-versus-fitted plot.
- **hettest** — performs Cook and Weisberg test for heteroscedasticity.
- **whitetst** — computes the White general test for Heteroscedasticity.

- **Tests for Multicollinearity**

- **vif** — calculates the variance inflation factor for the independent variables in the linear model.

- **collin** — calculates the variance inflation factor and other multicollinearity diagnostics
- **Tests for Non-Linearity**
 - **acprplot** — graphs an augmented component-plus-residual plot.
 - **cprplot** — graphs component-plus-residual plot, a.k.a. residual plot.
- **Tests for Model Specification**
 - **linktest** — performs a link test for model specification.
 - **ovtest** — performs regression specification error test (RESET) for omitted variables.

Questions

1. In the same manner that this analysis has used inbuilt dataset, you are going to use the inbuilt dataset `nls88.dta` . Create a linear regression model with your dependent variable and independent variable. Interpret the model and do model diagnostics.
2. Using the `laptop.csv` dataset, use variables inches, RAM (GB), and Weight (kg) as the independent variables and Price as the dependent variables to model a linear regression model. Ensure that you plot graphs, and do diagnostic analysis.