Tutorial 1

keywords: EViews, variables, data set, sample, population, descriptive analytics, summary statistics, histogram, scatter plots, correlation, simple linear regression model, predictive analytics

estimated reading time: 19 minutes

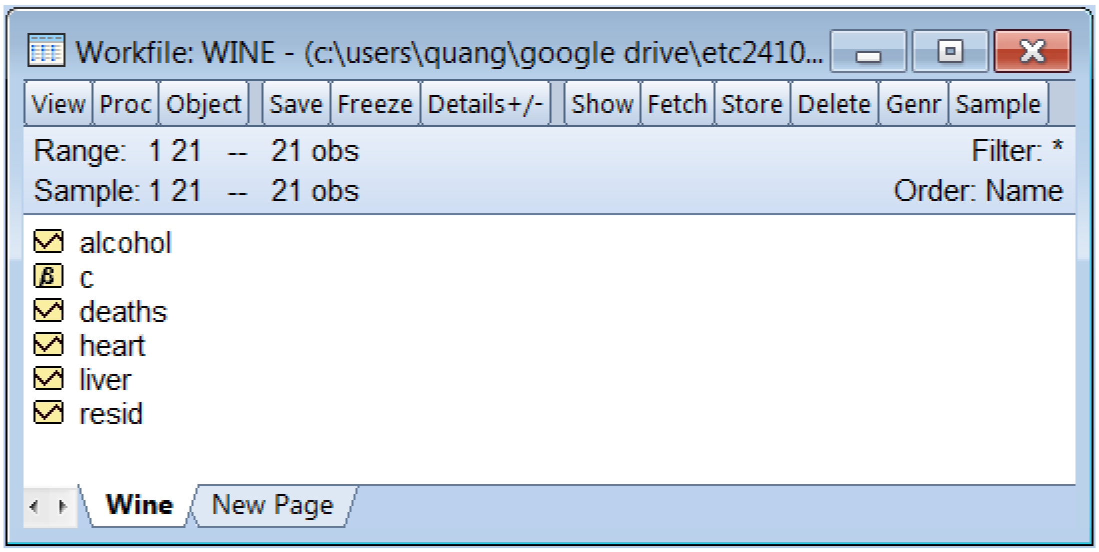
Quang Bui

July 24, 2018

Question 1

<u>Exploring EViews with *wine.wf1*</u>

EViews is a computer software designed specifically for econometric analysis. In today's class, we use EViews to perform basic econometric analysis on the workfile, $wine.wf1$.

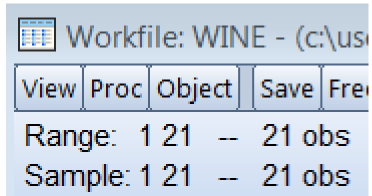


Objects

The elements in the whitespace of the *wine.wf1* workfile are called EViews **objects**. There are many types of EViews objects but in this unit we mostly focus on **series**, **groups**, **graphs**, **equations**, **scalar** and **coef** objects. Each object type is identified by a unique icon,

Icon	Object
	series
	group
	graph
	equation
	scalar
	coef

Data set



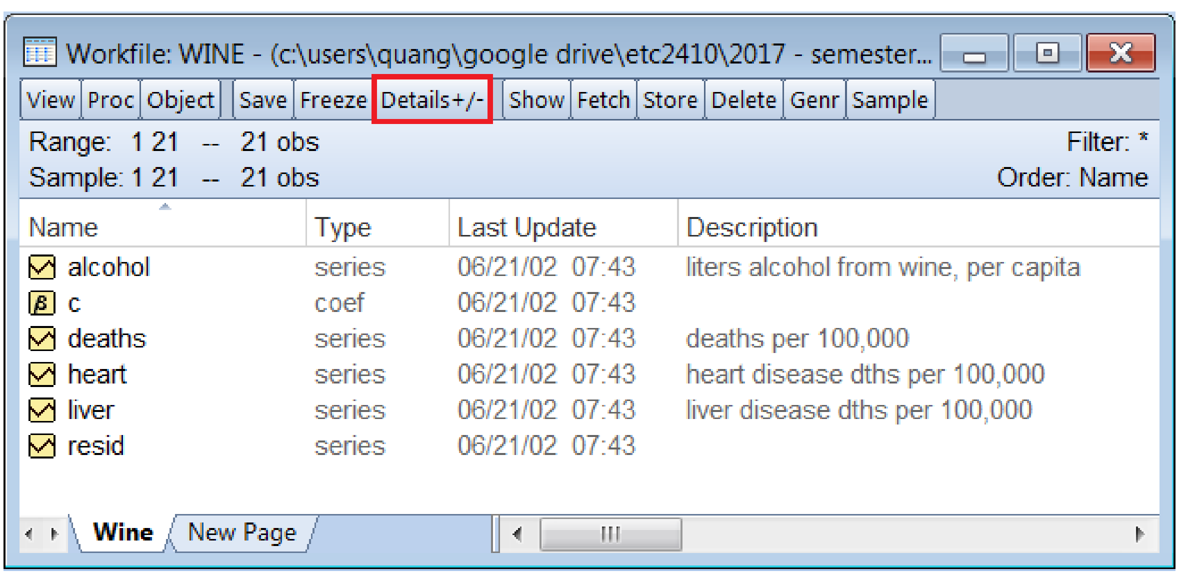
- *Range: 1 21 – 21 obs* indicates that the data set in *wine.wf1* contains 21 observations starting at observation number 1 to 21.
- *Sample: 1 21 – 21 obs* denotes the sample which EViews will use to produce graphs, run regressions, and obtain other EViews outputs. The current sample contains all observations in the data set.

Variables

The *wine.wf1* workfile contains data on 21 countries. This data set holds information about each country through the following variables:

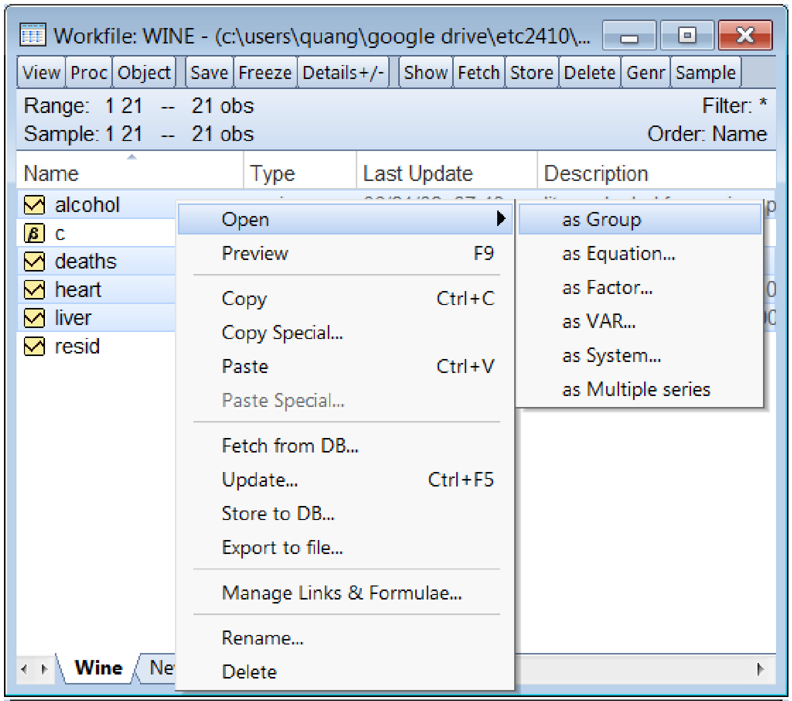
- *alcohol* - litres of wine consumed per capita per annum
- *deaths* - number of deaths per 100,000 of population
- *heart* - number of deaths from heart disease per 100,000 of population
- *liver* - number of deaths from liver disease per 100,000 of population

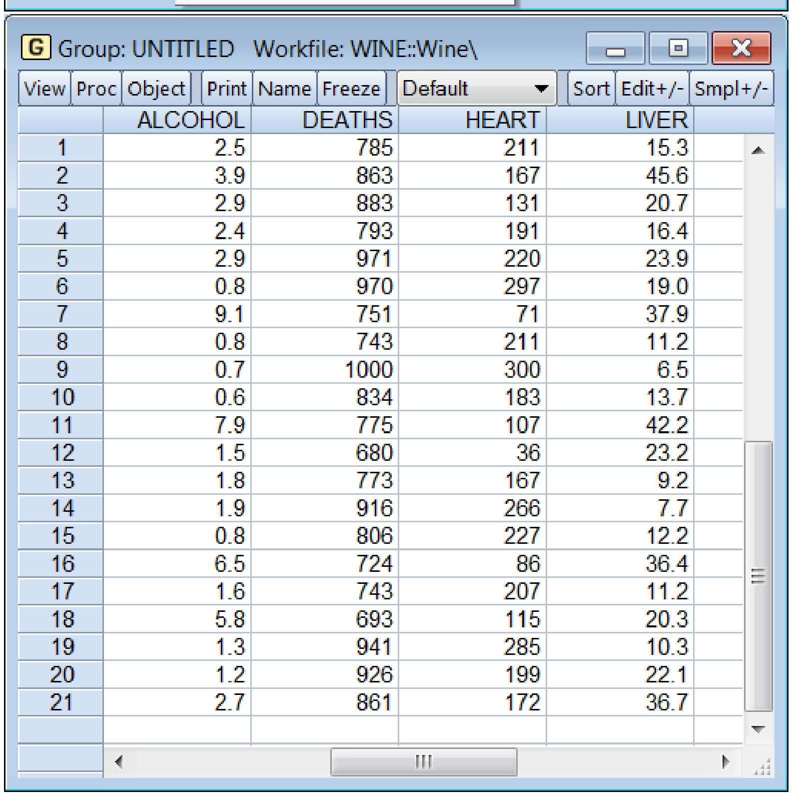
A description of each variable appears by clicking **Details +/-** on the workfile windows.



To view the data in a single variable, double click the variable. To view the data in multiple variables, select and highlight the variables of interest then,

$$\textit{Right click} \rightarrow \textit{Open} \rightarrow \textit{as Group}$$



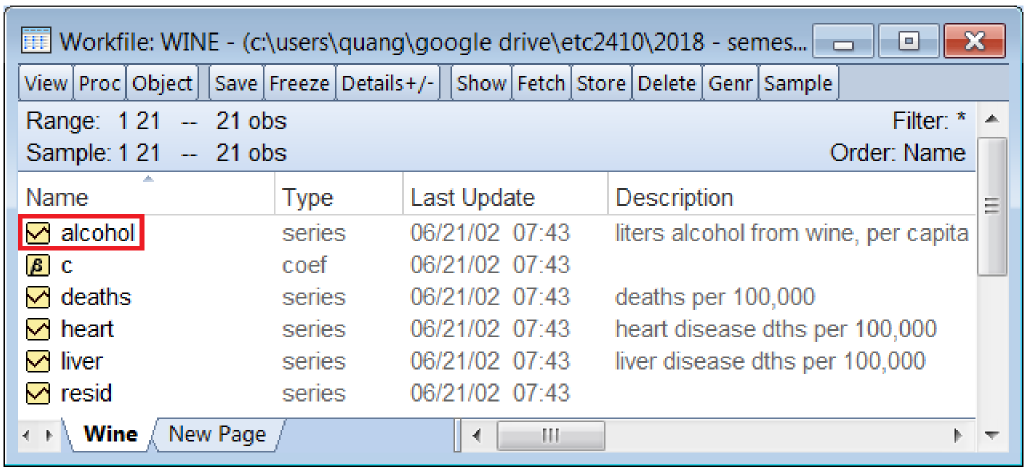


	ALCOHOL	DEATHS	HEART	LIVER
1	2.5	785	211	15.3
2	3.9	863	167	45.6
3	2.9	883	131	20.7
4	2.4	793	191	16.4
5	2.9	971	220	23.9
6	0.8	970	297	19.0
7	9.1	751	71	37.9
8	0.8	743	211	11.2
9	0.7	1000	300	6.5
10	0.6	834	183	13.7
11	7.9	775	107	42.2
12	1.5	680	36	23.2
13	1.8	773	167	9.2
14	1.9	916	266	7.7
15	0.8	806	227	12.2
16	6.5	724	86	36.4
17	1.6	743	207	11.2
18	5.8	693	115	20.3
19	1.3	941	285	10.3
20	1.2	926	199	22.1
21	2.7	861	172	36.7

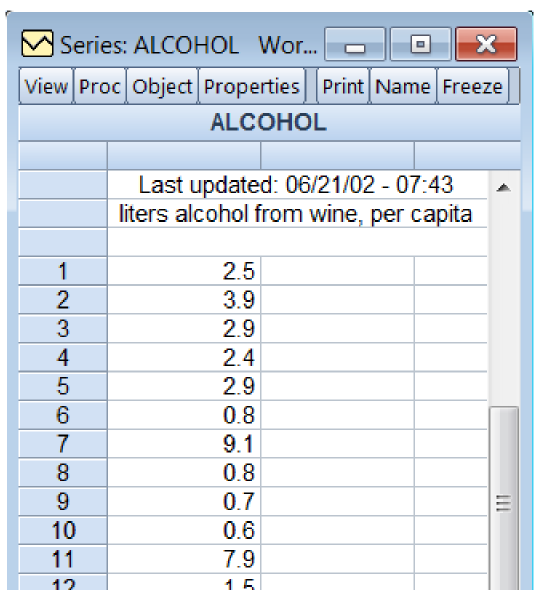
Question 2

<u>Descriptive analytics with histograms and descriptive statistics</u>

To obtain the histogram and descriptive statistics of the variable *alcohol*,

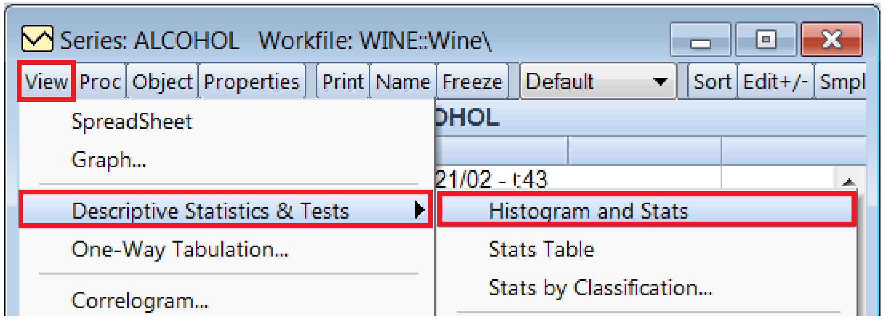


double-click on the variable *alcohol*,



then,

$$View \rightarrow Descriptive\ Statistics\ \&\ Tests \rightarrow Histogram\ and\ Stats$$



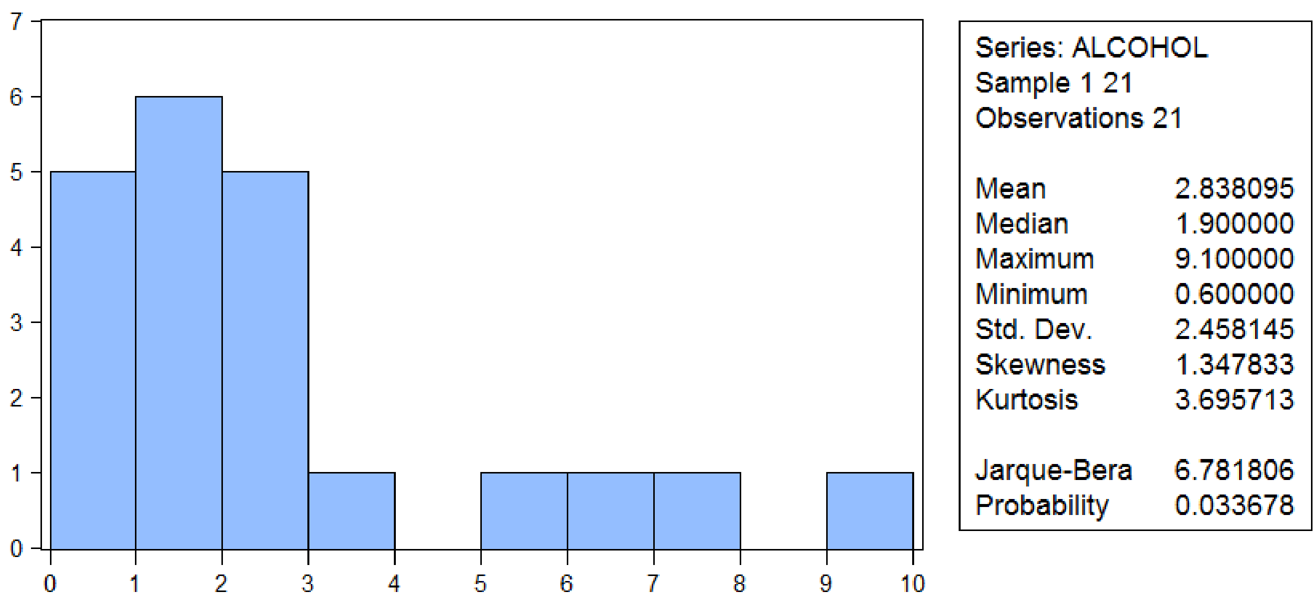


Figure 1: Histogram and descriptive statistics of litres of wine consumed per capita per annum from a sample of 21 countries.

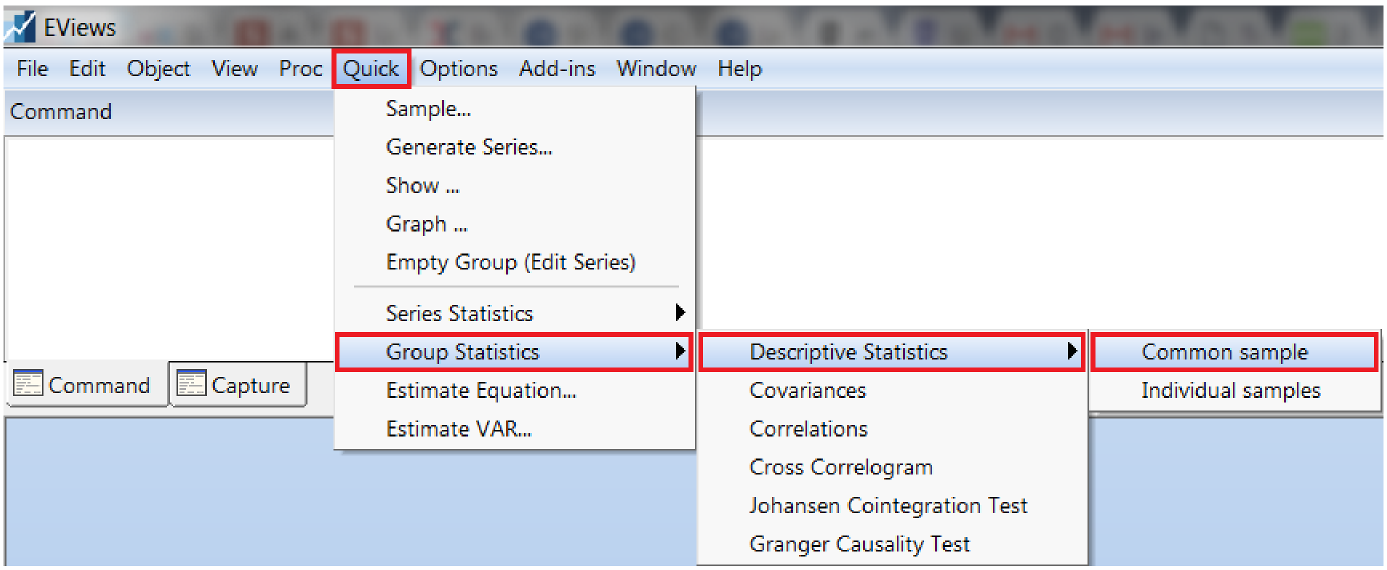
As we can see from Figure 1, wine consumption per capita per annum is positively skewed (right-tailed). On average, countries consume approximately 2.838 litres of wine per capita per annum. The mean wine consumption is greater than the median wine consumption,

$$2.838 > 1.90$$

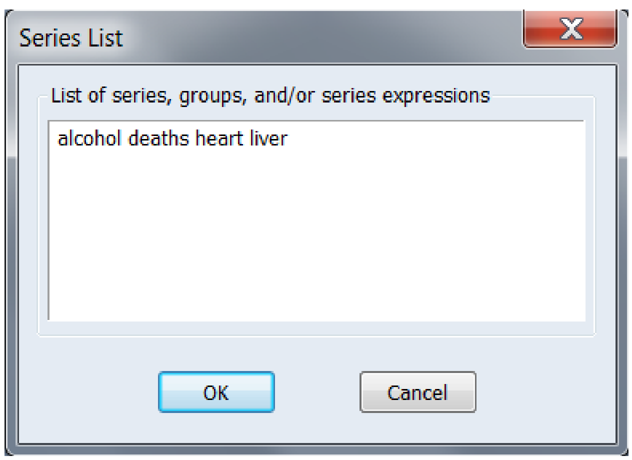
which also indicates that wine consumption is positively skewed.

We can also obtain descriptive statistics for multiple variables in a single spreadsheet. One way to do this is by selecting *Quick* from the EViews workfile menu,

$$Quick \rightarrow Group\ Statistics \rightarrow Descriptive\ Statistics \rightarrow Common\ sample$$



then type in the variables of interest into the *Series List* window,



Group: UNTITLED Workfile: WINE::Wine\

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

	ALCOHOL	DEATHS	HEART	LIVER
Mean	2.838095	830.0476	183.2857	21.03333
Median	1.900000	806.0000	191.0000	19.00000
Maximum	9.100000	1000.000	300.0000	45.60000
Minimum	0.600000	680.0000	36.00000	6.500000
Std. Dev.	2.458145	96.51864	73.15883	11.95869
Skewness	1.347833	0.232603	-0.206096	0.748989
Kurtosis	3.695713	1.862984	2.375217	2.311134
Jarque-Bera	6.781806	1.320569	0.490224	2.378666
Probability	0.033678	0.516704	0.782617	0.304424
Sum	59.60000	17431.00	3849.000	441.7000
Sum Sq. Dev.	120.8495	186317.0	107044.3	2860.207
Observations	21	21	21	21

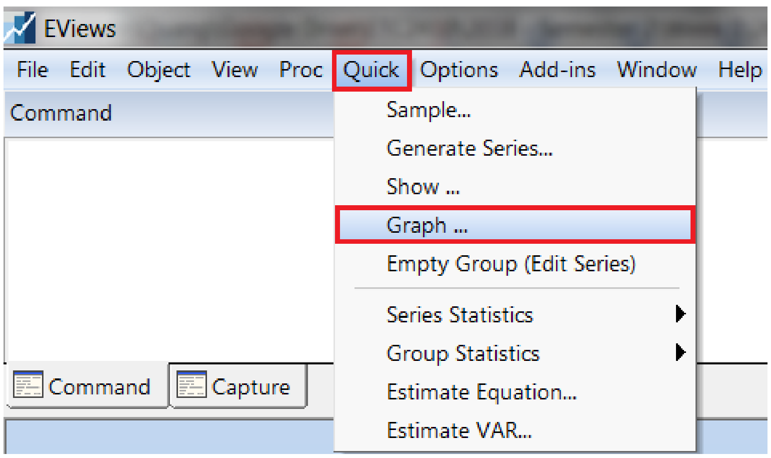
Note: The *Freeze* button takes a screenshot, while the *Name* button names and saves the object into the workfile.

Question 3

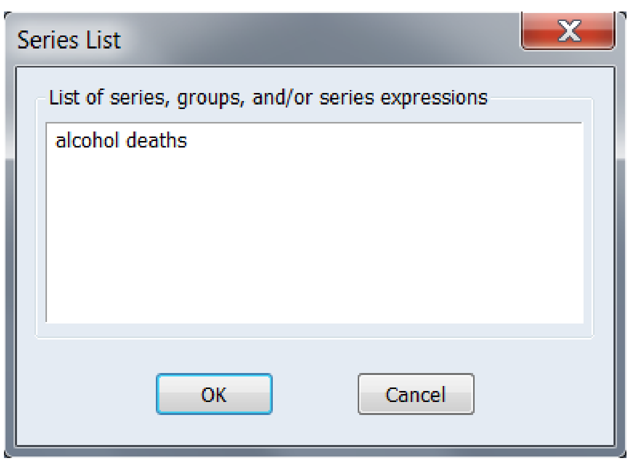
Scatter plots and correlations

Scatter plots are graphs used to visually explore the relationship between two variables. To obtain a scatterplot of *deaths* (y-axis) against *alcohol* (x-axis),

$$Quick \rightarrow Graph \ldots$$

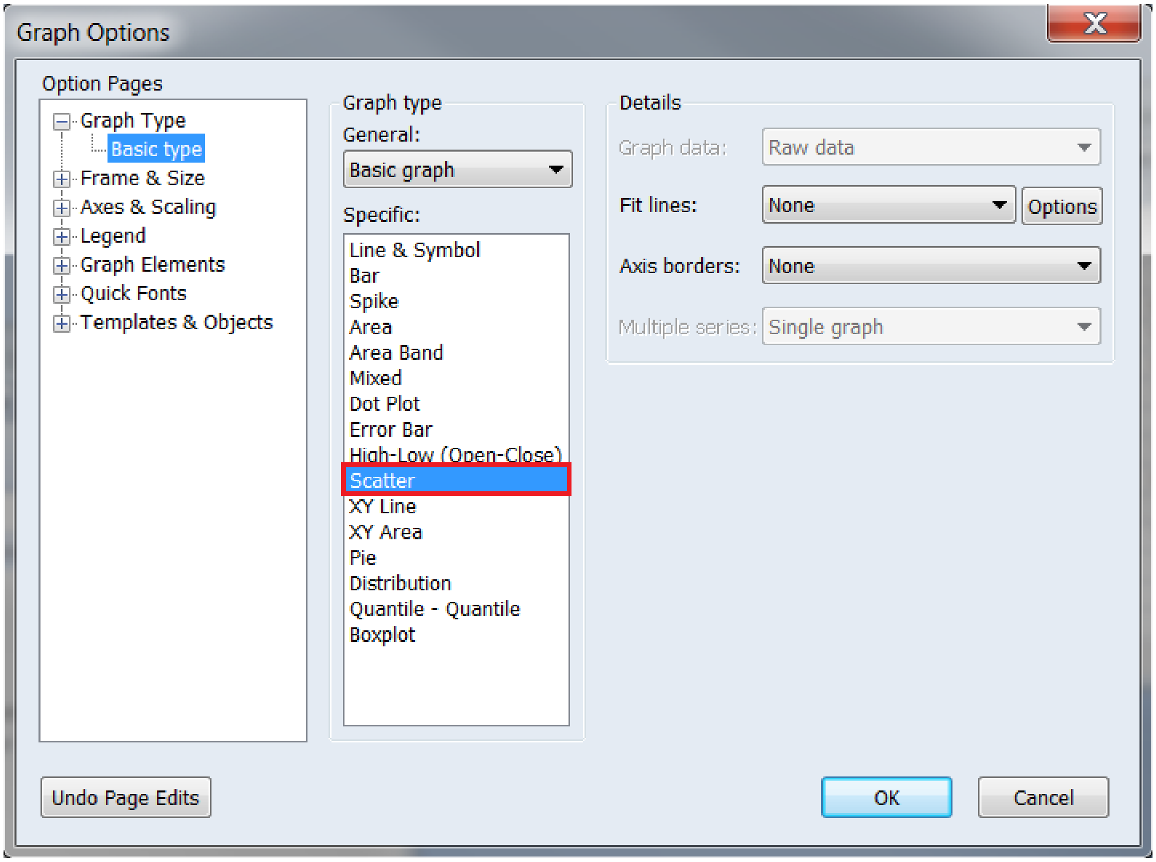


in the *Series List* window, type the x-variable followed by the y-variable,

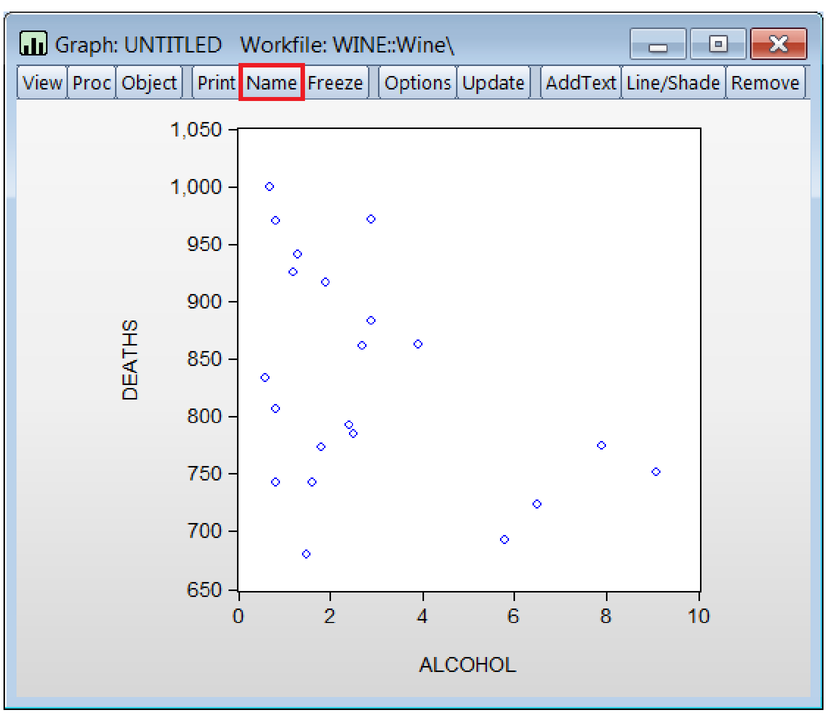


Under *Specific*, select *Scatter* and press *OK*,

$$Specific : Scatter \rightarrow OK$$

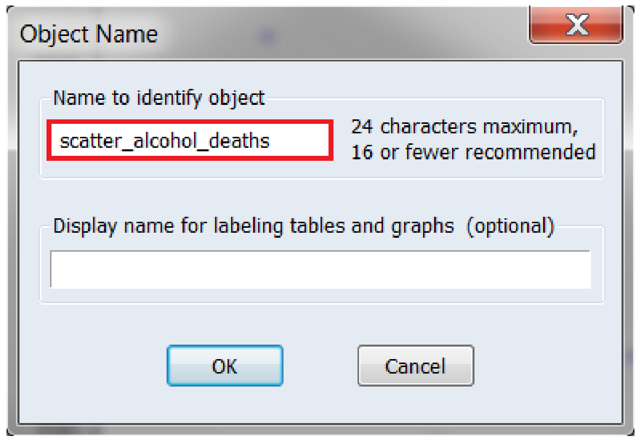


To give this graph the name *scatter_alcohol_deaths*, select *Name*,



then,

$$Name\ to\ identify\ object : scatter\_alcohol\_deaths$$



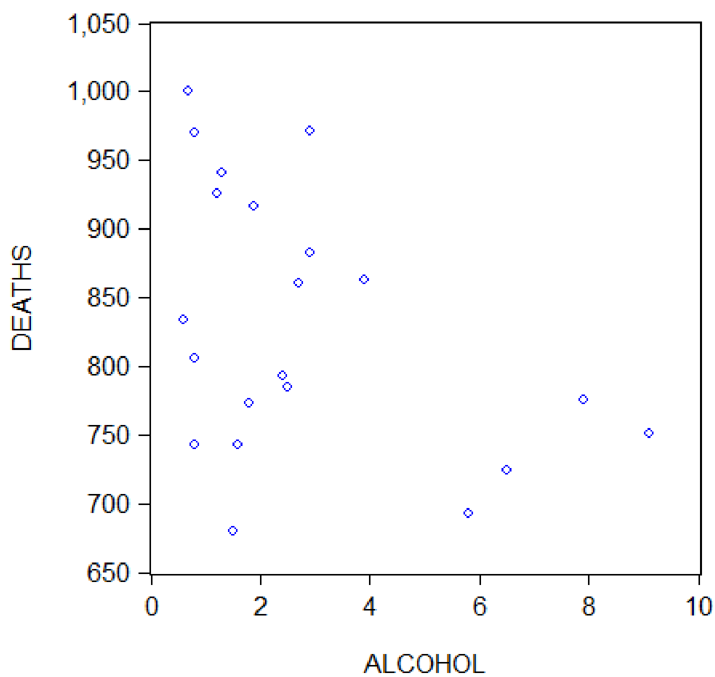
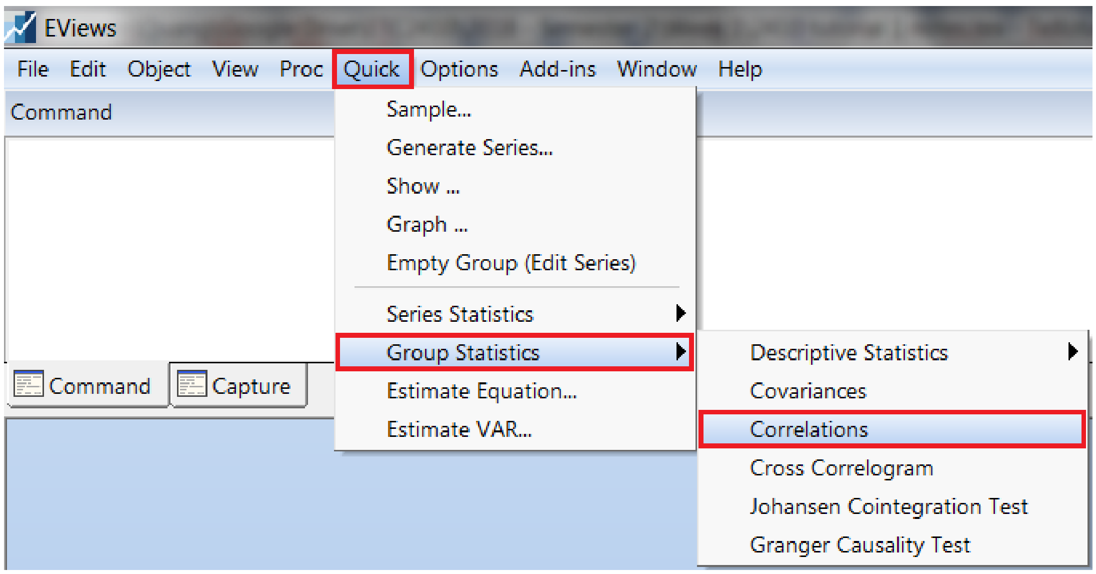


Figure 2: Scatter plot of deaths per 100,000 of population against litres of wine consumption per capita per annum.

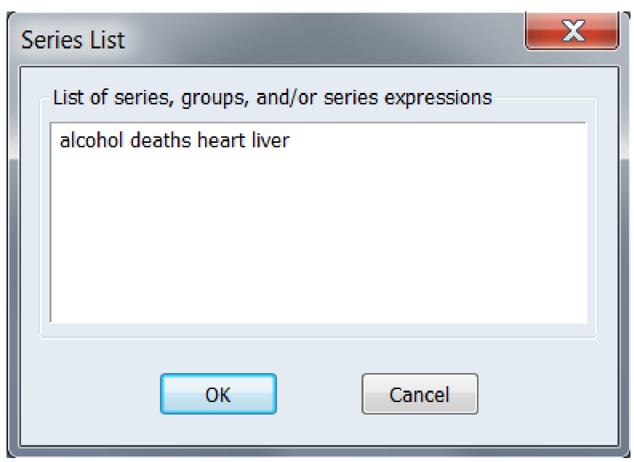
From the scatter plot, we observe a moderate negative linear relationship between wine consumption per capita per annum and deaths per 100,000 of population.

To obtain a table of sample correlation coefficients,

Quick → *Group Statistics* → *Correlations*



then type in the variables of interest into the *Series List* windows,



Correlation				
	ALCOHOL	DEATHS	HEART	LIVER
ALCOHOL	1.000000	-0.414201	-0.661336	0.737193
DEATHS	-0.414201	1.000000	0.749801	-0.201888
HEART	-0.661336	0.749801	1.000000	-0.612870
LIVER	0.737193	-0.201888	-0.612870	1.000000

The numbers in the table above represent the following sample correlation coefficients:

$$\widehat{corr}(alcohol, deaths) = -0.41420$$
$$\widehat{corr}(alcohol, heart) = -0.66134$$
$$\widehat{corr}(alcohol, liver) = 0.73719$$
$$\widehat{corr}(deaths, heart) = 0.74980$$
$$\widehat{corr}(deaths, liver) = -0.20189$$

The sample correlation coefficient measures strength and direction of the linear relationship between 2 variables and takes on a value between -1 and 1.

- If the sample correlation coefficient equal to -1, then the two variables have a perfect negative linear relationship.
- If the sample correlation coefficient equal to 1, then the two variables have a perfect positive linear relationship.
- If the sample correlation coefficient equal to 0, then the two variables have no linear relationship.

Therefore, the magnitude of the sample correlation coefficient measures the strength of the linear relationship between two variables and the size of the sample correlation coefficient measures the direction of the linear relationship.

Question 4

Estimating a simple linear regression model

Background

Simple Linear Regression Model

Simple - one independent variable
Linear - linear in the parameters

Suppose we specify a simple linear regression model of y on a constant (intercept) and x,

$$y = \beta_0 + \beta_1 x + u \tag{1}$$

where u is the error term that captures unobserved factors that affect y (factors other than x that affect y). It is included in the model because no matter how we specify the model, we cannot guarantee that the x input will perfectly output y i.e. there will always be some error. β_0 and β_1 are the intercept and slope coefficients respectively. They are population parameters and are unknown in practice.

To quantify the relationship between y and x, we need to estimate these population parameters. With data on y and x, we can use an estimator (more on this in subsequent weeks) to estimate (1), giving us estimates of β_0 and β_1 and the following estimated regression model,

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \tag{2}$$

Suppose that the true population relationship between deaths per 100,000 in the population and wine consumption per capita per annum is represented by the following simple regression model,

$$deaths = \beta_0 + \beta_1 alcohol + u \tag{3}$$

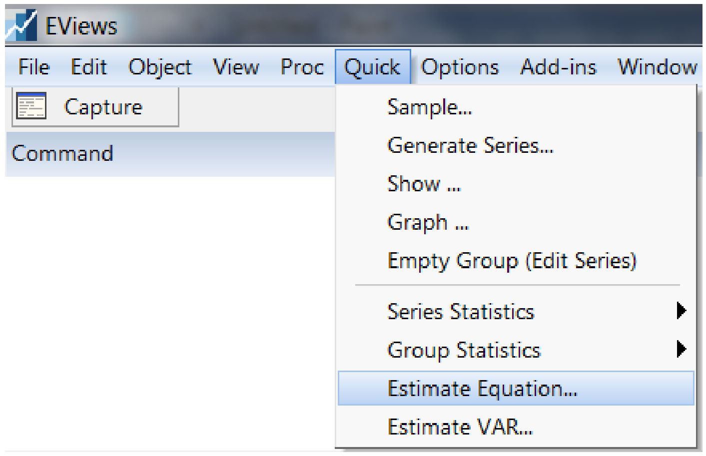
To estimate our model, we need data on *deaths* and *alcohol*. The EViews workfile *wine.wf1* contains data on *deaths* and *alcohol* from 21 countries $i = 1, 2, \ldots, 21$ ($n = 21$). We can express our model in terms of each country i,

$$deaths_i = \beta_0 + \beta_1 alcohol_i + u_i \tag{4}$$

$$i = 1, 2, \ldots, 21$$

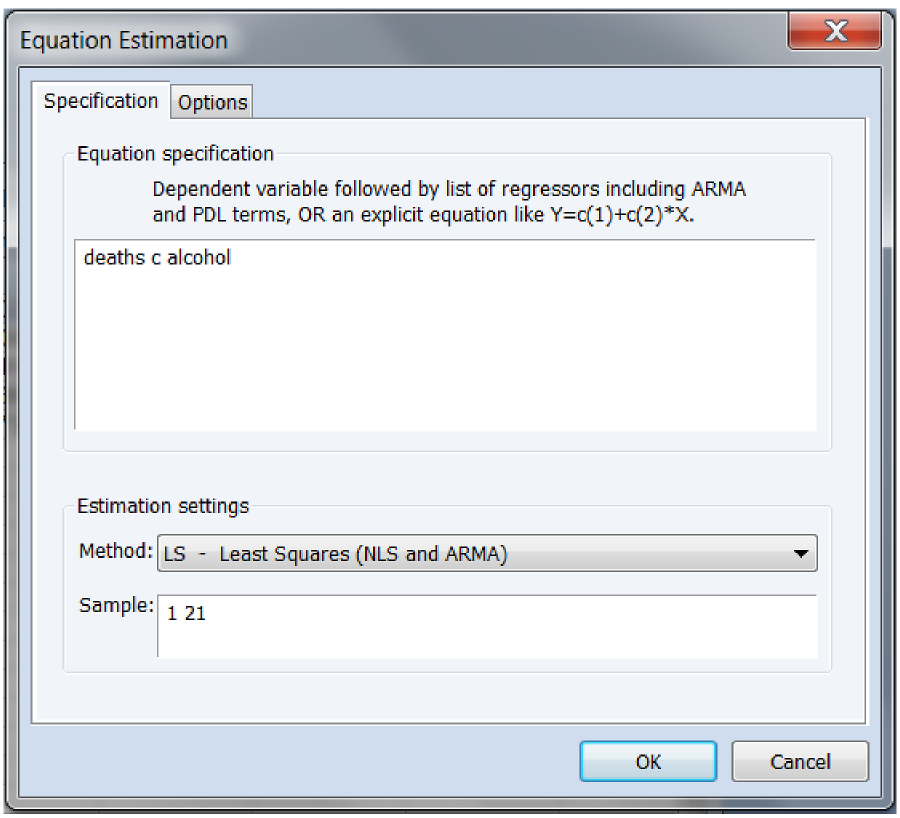
To estimate (4) in EViews,

$$Quick \rightarrow Estimate\ Equation \ldots$$



then in the *Equation Estimation* windows type in,

$$deaths\ c\ alcohol$$



Dependent Variable: DEATHS
Method: Least Squares
Sample: 1 21
Included observations: 21

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	876.2050	30.46816	28.75805	0.0000
ALCOHOL	−16.26352	8.198924	−1.983616	0.0619

R-squared	0.171562	Mean dependent var	830.0476
Adjusted R-squared	0.127960	S.D. dependent var	96.51864
S.E. of regression	90.13207	Akaike info criterion	11.93082
Sum squared resid	154352.0	Schwarz criterion	12.03030
Log likelihood	−123.2736	Hannan-Quinn criter.	11.95241
F-statistic	3.934731	Durbin-Watson stat	1.964148
Prob(F-statistic)	0.061939		

Table 1: Regression output of *deaths* on a constant and *alcohol*

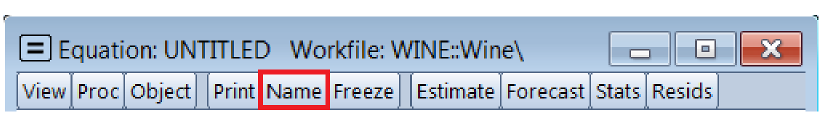
We report the estimated model by placing a 'hat' above the dependent variable and the standard error of the estimated coefficient in parenthesis underneath the estimated coefficient,

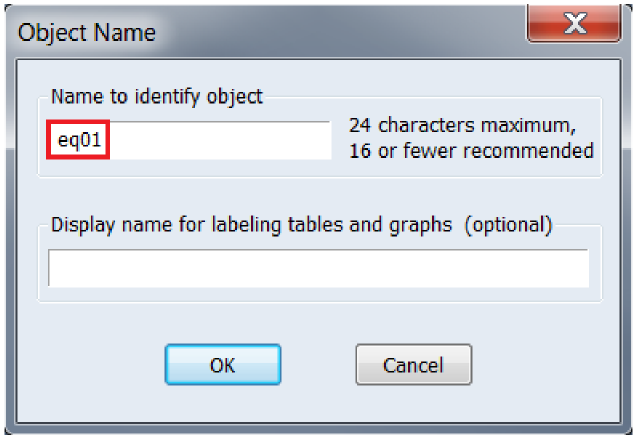
$$\widehat{deaths}_i = \underset{(30.4682)}{876.2050} - \underset{(8.1989)}{16.2635} alcohol_i$$

$$i = 1, 2, \ldots, 21$$

To name/save this regression output into our workfile,

$$Name \rightarrow Name\ to\ indentify\ object : eq01 \rightarrow OK$$

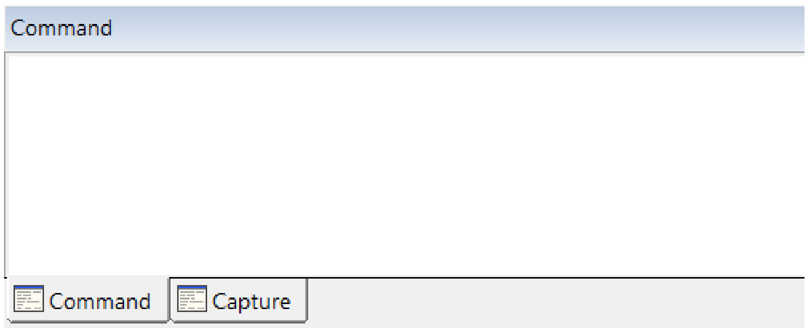



Object Name
Name to identify object
eq01
24 characters maximum,
16 or fewer recommended
Display name for labeling tables and graphs (optional)
OK
Cancel

Question 5

<u>Predictive analytics</u>

So far, we have generated outputs in EViews by 'pointing-and-clicking'. We can also obtain these outputs by executing lines of code through the **Command** window,



The Command window can also functions as a calculator. For example, we can use our estimated regression model,

$$\widehat{deaths}_i = \underset{(30.4682)}{876.2050} - \underset{(8.1989)}{16.2635}alcohol_i$$

$$i = 1, 2, \ldots, 21$$

to calculate the number of death per 100,000 of the population for a country that consumes <u>3</u> litres of wine per capita per annum,

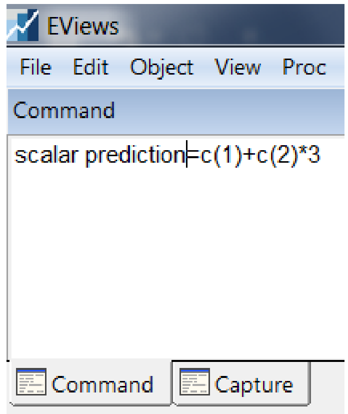
$$\widehat{deaths} = 876.2050 - 16.2635 \times 3$$

This is an example of predictive analytics. We have estimated a regression model of *deaths* and used this model to predict the number of death per 100,000 when a country consumes 3 litres of wine per capita per annum.

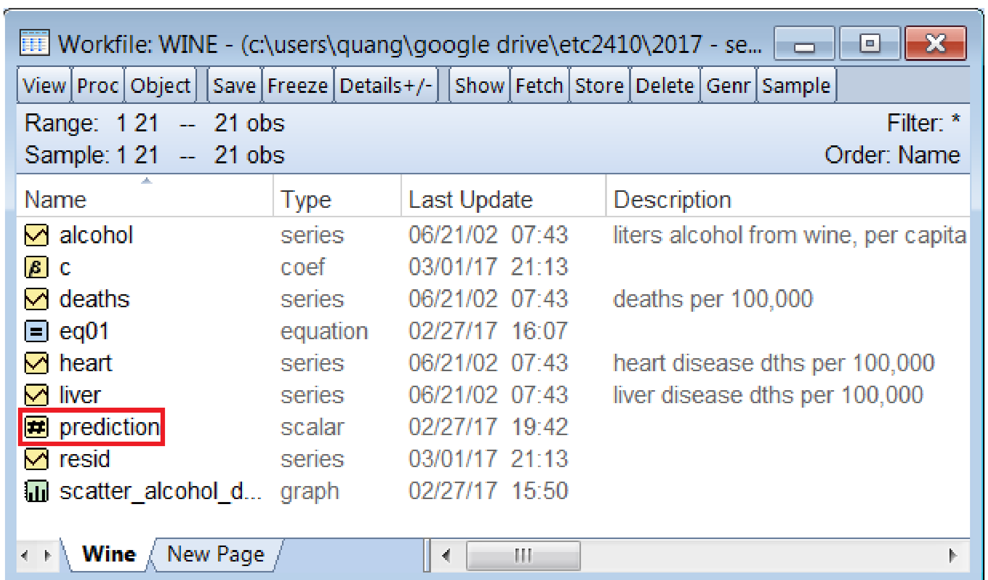
From the Command window, executing the code

$$scalar\ prediction = c(1) + c(2)^*3$$

will store the calculation into a **scalar** object named *prediction*.



Note: $c(1)$ and $c(2)$ represents the 1^{st} and 2^{nd} estimated coefficient of the most recent estimated regression model i.e. the values 876.2050 and -16.2635



Question 6

Presenting results

(discuss in class)

Question 7

Discuss each of the outputs you obtained above. What do you learn about the impact of alcohol consumption on the death rate? How could you improve your analysis?

(discuss in class)