

Stata Kullanımı (Stata 14)

Arş. Gör. Selçuk GEMİCİOĞLU

Eylül 2019

İçindekiler

1 Giriş	2
2 Stata Nedir?	2
3 Niçin Stata?	2
4 Stata Arayüzünün Tanıtımı	3
5 Çalışma Klasörü Tanımlama	4
6 Veri Girişi	5
7 Veri'nin İncelenmesi	10
8 Veri'nin Kaydedilmesi	12
9 Veri Setine Ait Tanımlayıcı İstatistikler	15
10 Veri Seti ve Değişkenler ile Yapılan Temel İşlemler	18
11 Grafik Çizme	26
12 Do ve Log Dosyaları	29
13 Yardım	32
14 Kaynaklar	34

1 Giriş

Ekonometri teori ve uygulamanın iç içe olduğu bir alandır. Uygulamasız bir ekonometri dersi yakıtsız bir araba gibidir. Bunun içindir ki birçok araştırmacı teorik bilgilerin geçerliliğinin gerçek veriler kullanılarak test edilmesine oldukça önem vermektedir. Bu doğrultuda dersin amacı teorik bilgilerinizin uygulamayla nasıl entegre edileceğini öğretmektir¹.

Ekonometrik analizler Stata programı kullanılarak yapılacaktır. Ekonometrik analizlere geçmeden önce Stata'nın kullanımı ile ilgili temel bilgiler sizlere verilecektir. Bu bilgiler analiz yeteneğinizi artırmanın yanı sıra ilerde karşınıza çıkacak olası ödev, tez, proje vb. çalışmalarda da işinizi oldukça kolaylaştıracaktır. Stata kullanım becerisini daha fazla geliştirmek isteyenler yalnızca derste anlatılanlarla kendilerini sınırlamamalı, önerilecek olan ek kaynaklarla çalışmaya devam etmelidirler.

2 Stata Nedir?

StataCorp'a (2016) göre Stata veri analizi, veri yönetimi ve grafikler için ihtiyacınız olan her şeyi sağlayan eksiksiz ve entegre bir istatistiksel yazılım paketidir. Temel olarak Stata, verileri (büyük ve küçük veri setlerini) saklamanıza ve yönetmenize, verileriniz üzerinde istatistiksel analizler yapmanıza ve anlaşılır görselliğe sahip grafikler oluşturmanıza olanak sağlayan bir yazılımdır.

3 Niçin Stata?

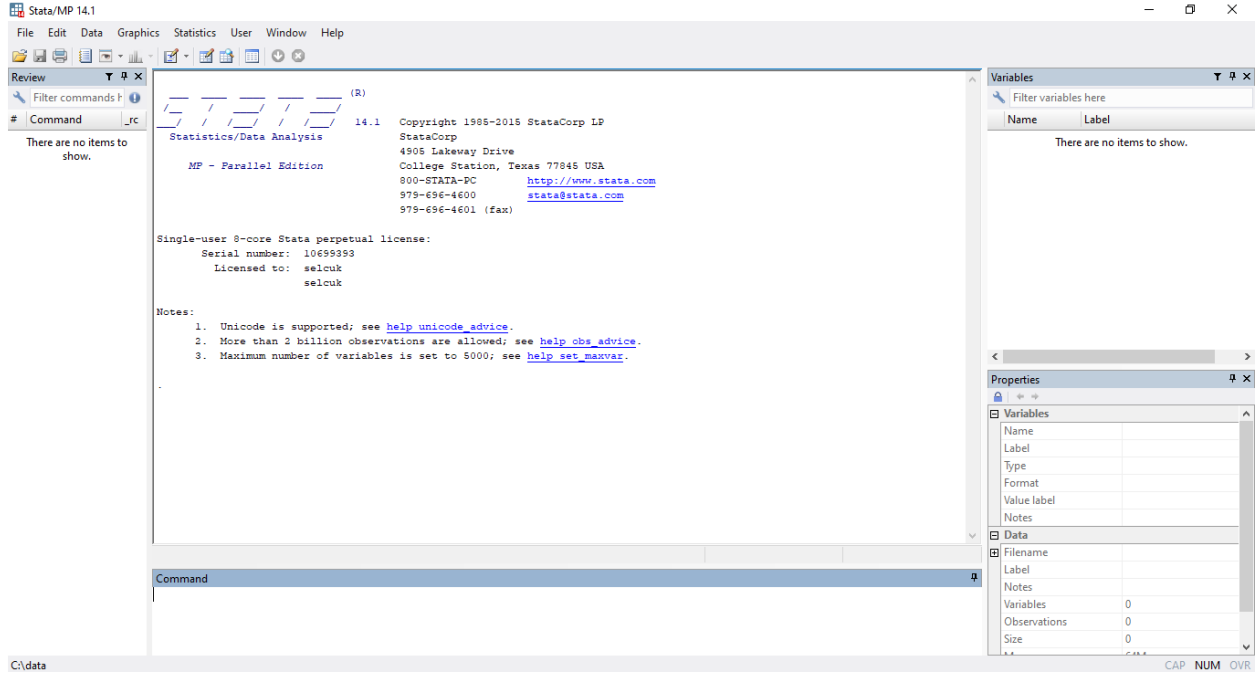
Ekonometrik analizleri gerçekleştirmek için Stata kullanan çok sayıda araştırmacı vardır. Fakat diğer istatistik programlarını (SPSS, SAS, Eviews, R, Matlab ve Gauss) kullanan pek çok insan olduğu da bilinmektedir. Bu durumda aklınıza gelen ilk soru neden Stata'yı tercih ediyoruz olacaktır.

Stata'yı tercih etmemizin başlıca nedenleri arasında büyük veri setlerinin (örneğin milyonlarca gözlem) kullanımına ve düzenlenmesine izin vermesi ve panel ve zaman serisi regresyon analizleri için sürekli gelişme göstermesidir. Ayrıca Stata belirli bir ihtiyacı olan kullanıcılar tarafından sürekli olarak güncellenmekte veya geliştirilmektedir. Bu güncellemeler kendi yazılımımızla kolayca entegre edilebilmektedir.

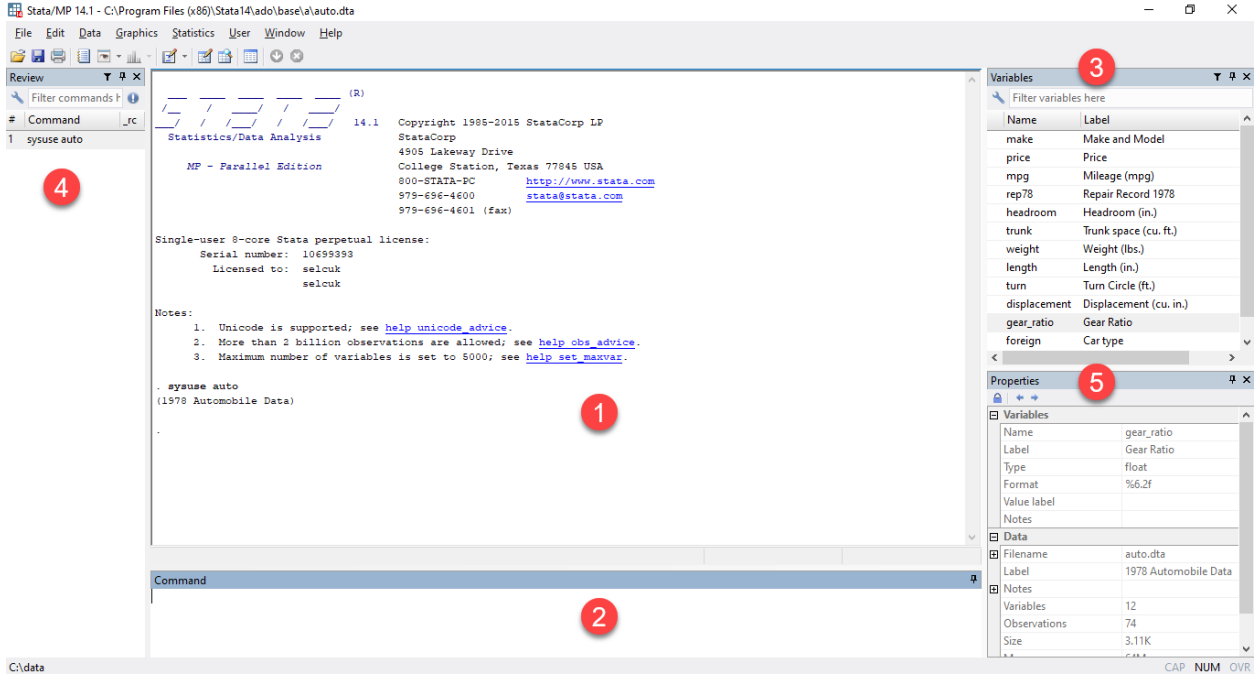
¹Bu ders notu SBF Maliye Bölümü son sınıf öğrencileri için hazırlanmıştır. Ders notunda yer alan tüm hata ve yanlışların sorumluluğu bana aittir. Sorularınız ve değerli görüşleriniz için email adresi <sgemicioğlu@ankara.edu.tr>. Görüşmeler için adres, SBF Fazıl Kafadar Spor Salonu 2. kat 145 nolu oda.

4 Stata Arayüzünün Tanıtımı

Bilgisayarımızda Stata'yı (Stata 14 sürümü için) açtığımızda varsayılan ekran görüntüsü aşağıdaki gibidir.



Stata'nın arayüzünde 1- Sonuç Penceresi, 2- Komut Penceresi, 3- Değişkenler Penceresi, 4- Geçmiş İşlemler Penceresi ve 5- Özellikler Penceresi yer almaktadır.



1- Sonuç Penceresi

Stata'ya yazdığımız komutların sonucunu gösterir. Yeni bir dosya açtığımızda bile, bu pencerede bir “sonuç” görülmektedir. Sonuç penceresi, özellikle bir komutun çalışıp çalışmadığını anlamak için kullanışlıdır. Örneğin, Stata'ya yeni bir değişken oluşturması komutunu verirseniz, sonuç penceresi size yeni bir değişken oluşturup oluşturmadığını söyleyecektir.

2- Komut Penceresi

Stata'ya verilerinizle ne yapmak istediğinizi anlatmak için komutları (Stata'nın tanıdığı ve belirli bir işleyle ilişkilendirdiği kelimeler) girdiğiniz alandır.

3- Değişkenler Penceresi

Stata'ya aktarılan veri setindeki değişken isimlerinin ve açıklamalarının yer aldığı alandır.

4- Geçmiş İşlemler Penceresi

Stata'da çalışan ve çalışmayan tüm komutların kaydının tutulduğu yerdir. Bu alan, yazdığımız tüm komutları izlemenize yardımcı olur. Bunun yanı sıra komutları bu pencereye kopyalayıp yapıştırabilir ve bu pencerede yeniden çalıştırabilirsiniz.

5- Özellikler Penceresi

Stata'ya aktarılan veri setinde yer alan değişkenlerin ve veri setinin özelliklerini görüntüler. Özellikler penceresini kullanarak değişken ve veri seti özelliklerini değiştirebilirsiniz.

5 Çalışma Klasörü Tanımlama

Açık bir Stata dosyasıyla herhangi bir işlem yapmadan önce, daima çalışma klasörünüzün konumunu tanımlamanız gerekir. “**Çalışma klasörü**”, belirli bir proje veya analiz için tüm verilerinizi sakladığımız ve bilgisayarımızın herhangi bir sürücüsünde yer alan bir klasördür.

Birden fazla veri seti ile çalışılması, dosyaların başka dosyalarla karışması ya da çalışılan dosyaların ayrı klasörlerde yer alması oldukça kafa karıştırıcı bir durumdur. Bu yüzden tüm veri setlerinin ve dosyaların içerildiği çalışma klasörünün oluşturulması ve tanımlanması bu kafa karıştırıcı durumları ortadan kaldıracaktır.

Bunu yapmak için öncelikle bilgisayarınızın “**D**” sürücüsünde **Stata_workshop** adında boş bir klasör oluşturunuz ve ardından Stata'yı açınız. Stata'da çalışma klasörünün yerini tanımlamak için aşağıdaki kodu komut penceresine yazıp enter'a basınız.

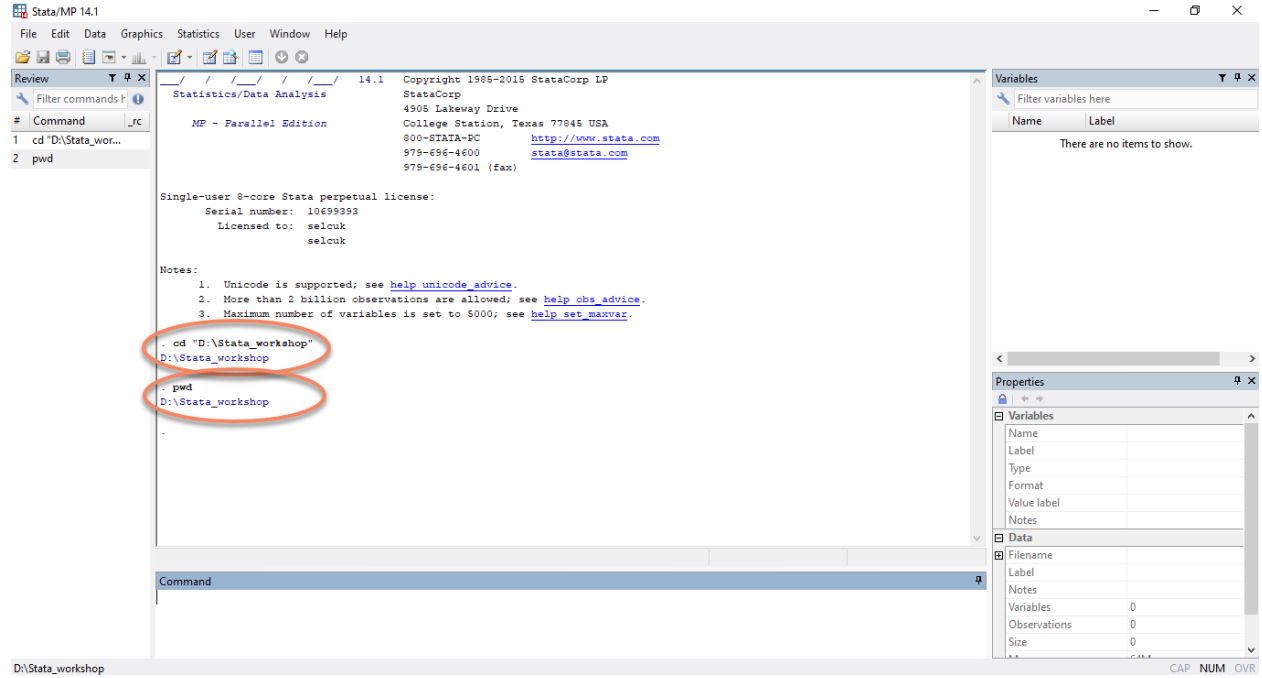
Not: Burada çalışma dosyası “**D**” sürücüsünde tanımlanmıştır. Fakat bilgisayarınızın hangi sürücüsünde çalışacağımız tamamen size kalmıştır isterseniz “**C**” veya başka bir sürücüde de çalışabilirsiniz. Ayrıca klasör ve dosya isimlerinde Türkçe karakter ve boşluk bulunmamasına dikkat ediniz.

```
. cd "D:\Stata_workshop" /* Çalışma klasörünüzü tanımlar */
```

Çalışma klasörünün yerininin doğru tanımlanıp tanımlanmadığını kontrol etmek için aşağıdaki kodu komut penceresine yazınız ve enter'a basınız.

```
. pwd /* Şu anki çalışma alanınızı gösterir */
```

Not: Yukarıdaki kodlarda görüleceği üzere kodların başında “**”** vardır. “**”**, kodların parçası olmadığı için kodları yazarken “**”**’ı kullanmayınız. Ayrıca yukarıdaki gibi bazı kodların yanında yorumlara **/* YORUM */** yer verilmiştir. Yine bu yorumlar kodların bir parçası değildir yalnızca açıklama amaçlı kullanılmıştır.



Sonuç penceresinde görüleceği üzere “D” sürücüsünde **Stata_workshop** adlı klasörde çalışmaktasınız. Şu anki çalışma alanınızı gösteren komut da aynı sonucu vermiştir.

6 Veri Girişi

Bu bölümde veri setinin Stata’ya nasıl girileceği anlatılacaktır. Veri seti Stata menüsü kullanılarak ya da komut yazılarak Stata’ya aktarılabilir. Stata **xls**, **xlsx**, **csv**, **dta**, **txt** ve bazı istatistik programlarının kendi uzantılı dosyalarında yer alan veri setlerini çekebilmektedir. Burada dikkat edilmesi gereken durum ise farklı uzantılı dosyaları Stata’da açabilmek için yazılacak kodun farklılaşmasıdır. Bu derste **xls**, **xlsx** ve **dta** uzantılı dosyalarda yer alan veri setleri kullanılacaktır. Bu yüzden yalnızca bu dosyaların Stata’ya nasıl aktarılacağı öğretilenektir.

Veri seti Stata formatında yani **dta** uzantılı bir dosyadaysa, **use** komutu kullanılarak aşağıdaki kodlar yardımıyla veri seti Stata’ya aktarılır.

```
. use "D:\Stata_workshop\auto.dta", clear
. use ".\auto.dta", clear
```

Not: İkinci komutta “D:/Stata_workshop” yerine yalnızca “.” kullanılmıştır. Çünkü “.”, Stata’ya tanımladığımız çalışma klasörü içinde çalıştığımızı söyler.

The screenshot shows the Stata software interface. The main window displays the command window with the following text:

```

Stata/MP 14.1 - D:\Stata_workshop\auto.dta
File Edit Data Graphics Statistics User Window Help
Review
Filter commands
# Command _rc
1 cd "D:\Stata_wor...
2 use "D:\Stata_wo...
3 use "/auto.dta", ...

MP - Parallel Edition
14.1 Copyright 1985-2015 StataCorp LP
StataCorp
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 8-core Stata perpetual license:
Serial number: 10699393
Licensed to: selcuk
selcuk

Notes:
1. Unicode is supported; see help unicode advice.
2. More than 2 billion observations are allowed; see help obs advice.
3. Maximum number of variables is set to 5000; see help set maxvar.

. cd "D:\Stata_workshop"
D:\Stata_workshop
. use "D:\Stata_workshop\auto.dta", clear
(1978 Automobile Data)
. use "/auto.dta", clear
(1978 Automobile Data)

```

The right-hand side of the interface shows the 'Variables' panel with a list of variables and their labels. The variables listed are:

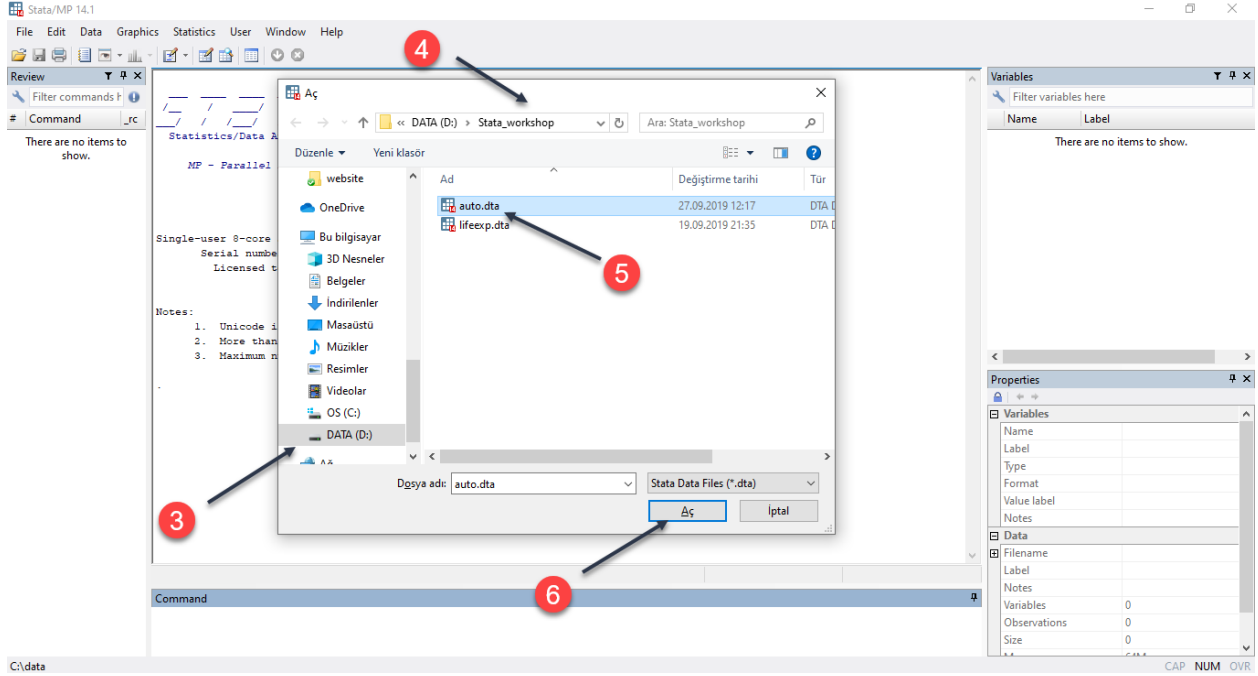
Name	Label
make	Make and Model
price	Price
mpg	Mileage (mpg)
rep78	Repair Record 1978
headroom	Headroom (in.)
trunk	Trunk space (cu. ft.)
weight	Weight (lbs.)
length	Length (in.)
turn	Turn Circle (ft.)
displacement	Displacement (cu. in.)
gear_ratio	Gear Ratio
foreign	Car type

The 'Properties' panel on the right shows the following information:

Variable	Value
Name	make
Label	Make and Model
Type	str18
Format	%-18s
Value label	
Notes	
Data	
Filename	auto.dta
Label	1978 Automobile Data
Notes	
Variables	12
Observations	74
Size	3.11K

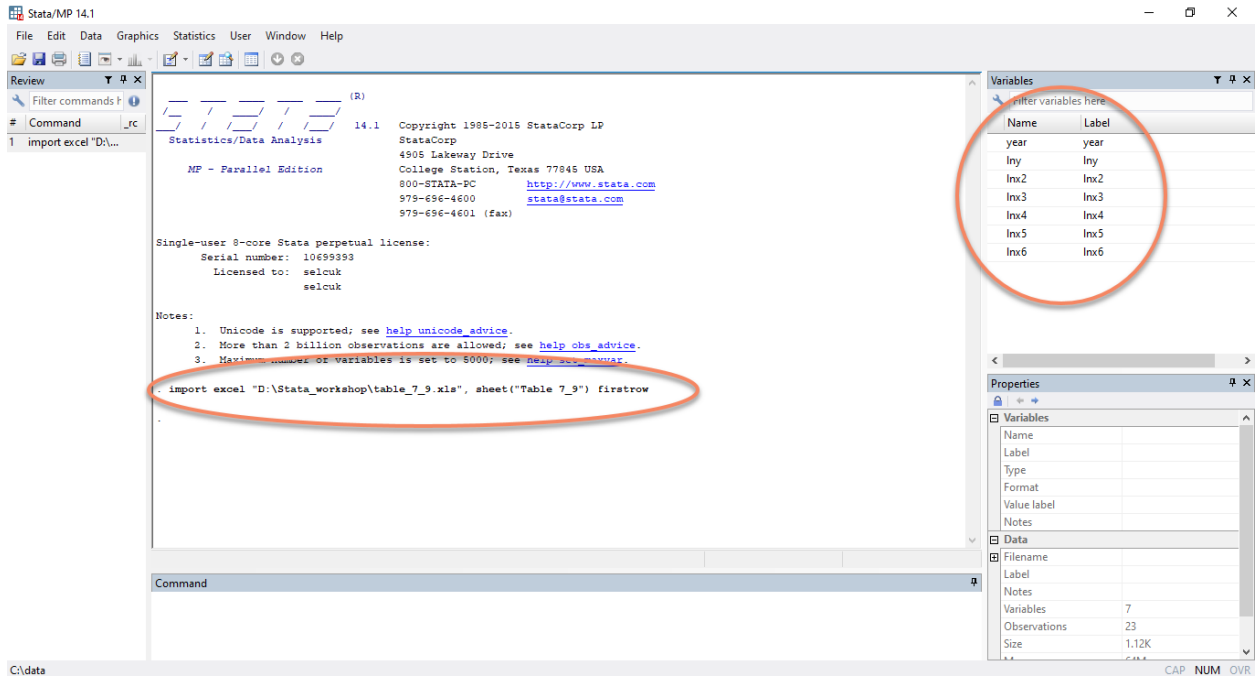
Stata menüsünden ise **File > Open > D > Stata_workshop > Auto** dosyasını seç > **Aç** adımlarını izleyerek aynı veri setini açabilirsiniz.

The screenshot shows the Stata software interface with the 'File' menu open. The 'Open...' option is highlighted, and a red circle with the number '1' is placed over it. The 'Open...' option is also highlighted in the menu, and a red circle with the number '2' is placed over it. The main window displays the same text as the previous screenshot.

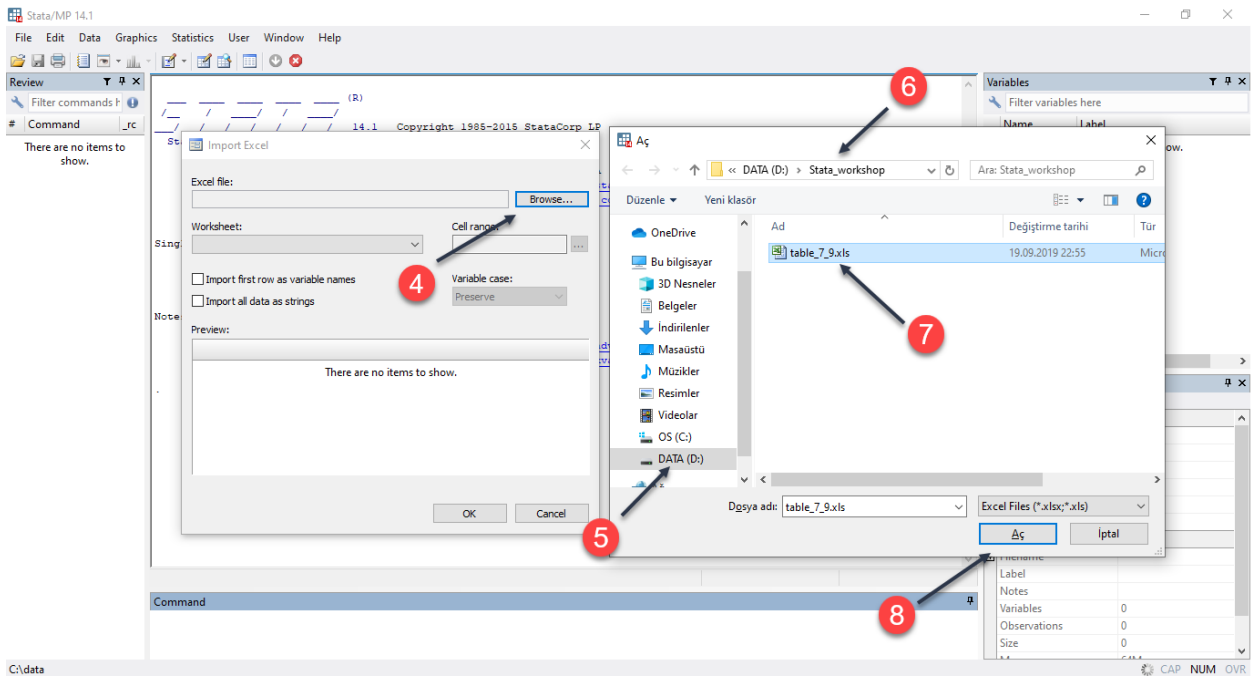
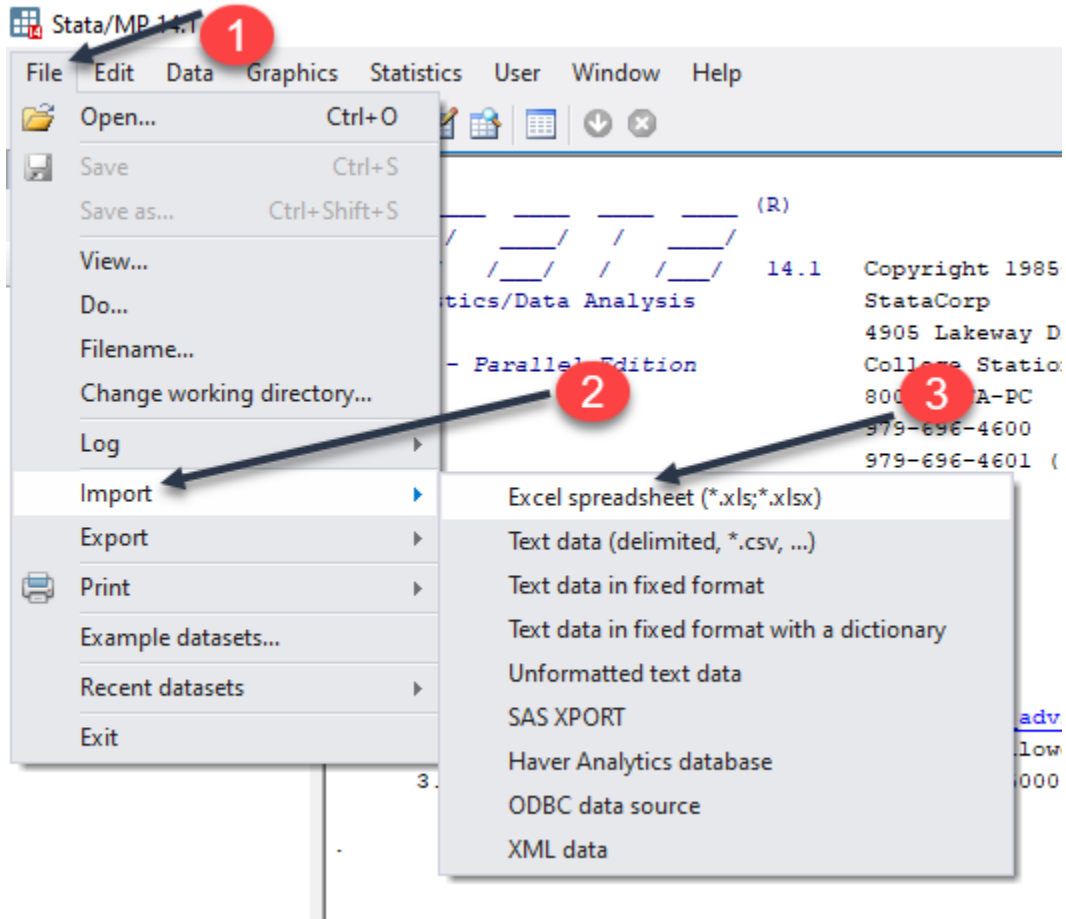


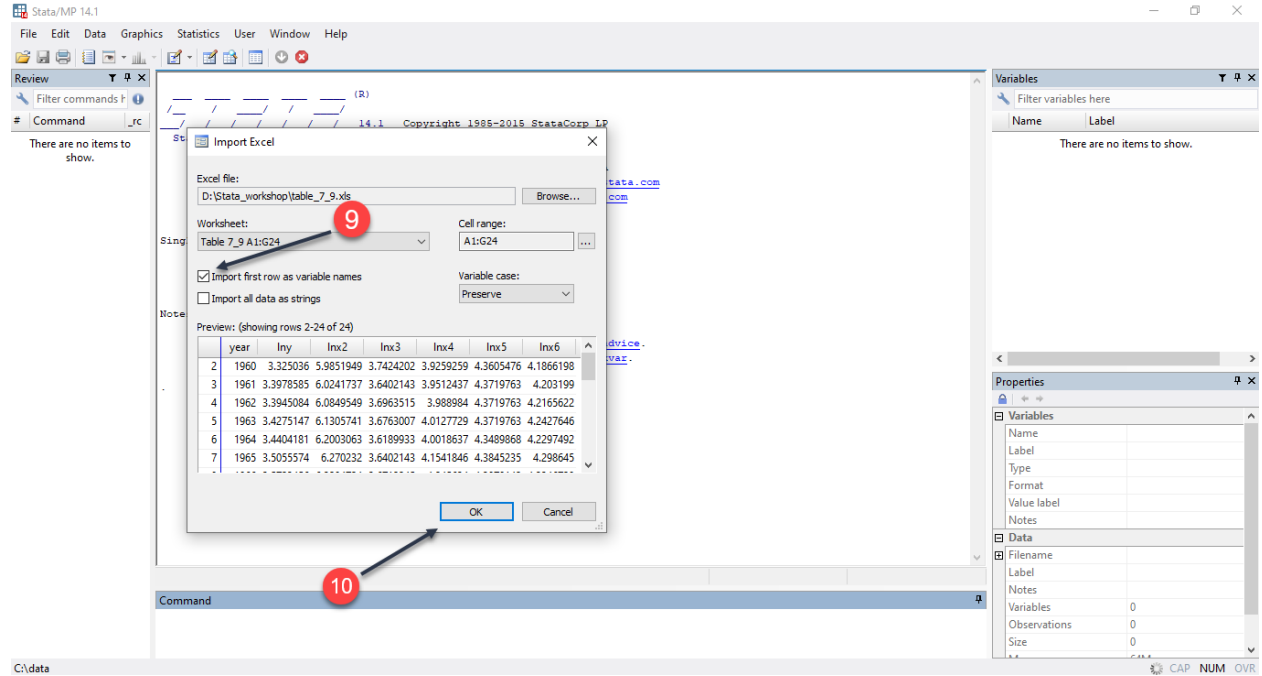
Veri seti xls uzantılı yani bir excel dosyasındaysa aşağıdaki kod ile Stata'ya aktarabilirsiniz.

```
. import excel "D:\Stata_workshop\table_7_9.xls", sheet("Table 7_9") firstrow clear
```

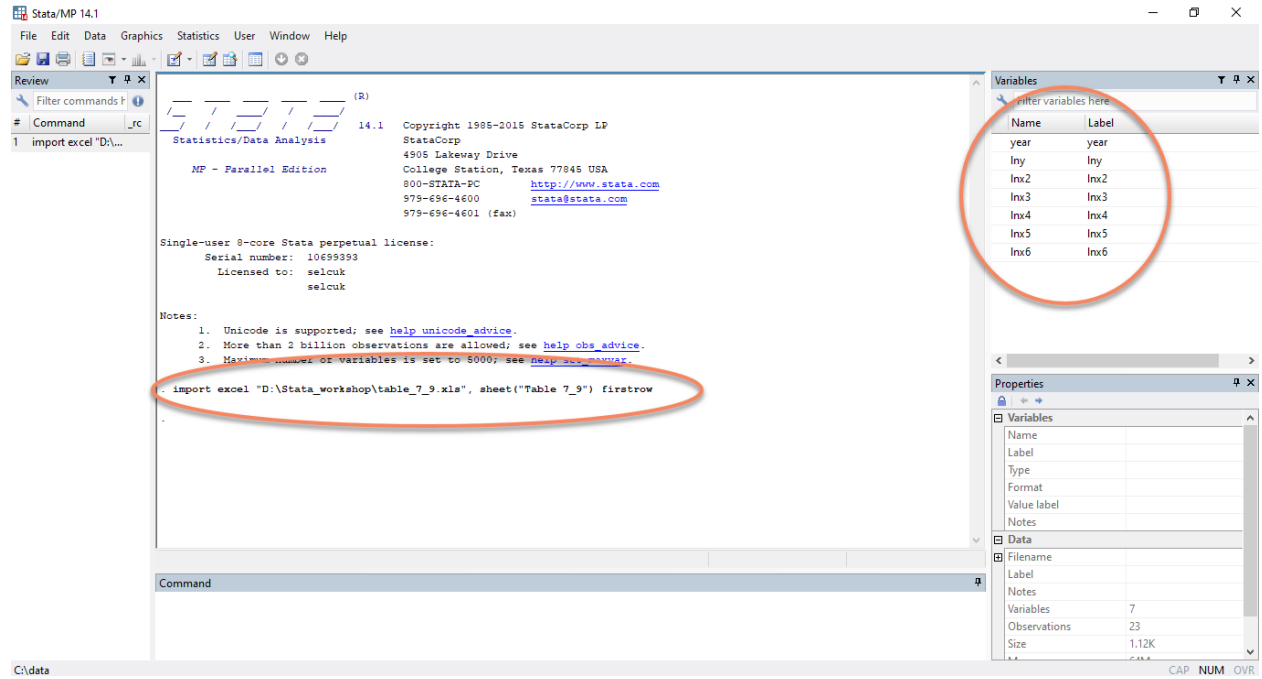


Yine Stata menüsünden **File > Import > Excel spreadsheet > Browse > D > Stata_workshop > table_7_9** adlı dosyayı seç > **Aç > Import first row as variable names** kutucuğuna tıkla > **Ok** adımlarını izleyerek aynı veri setini açabilirsiniz.





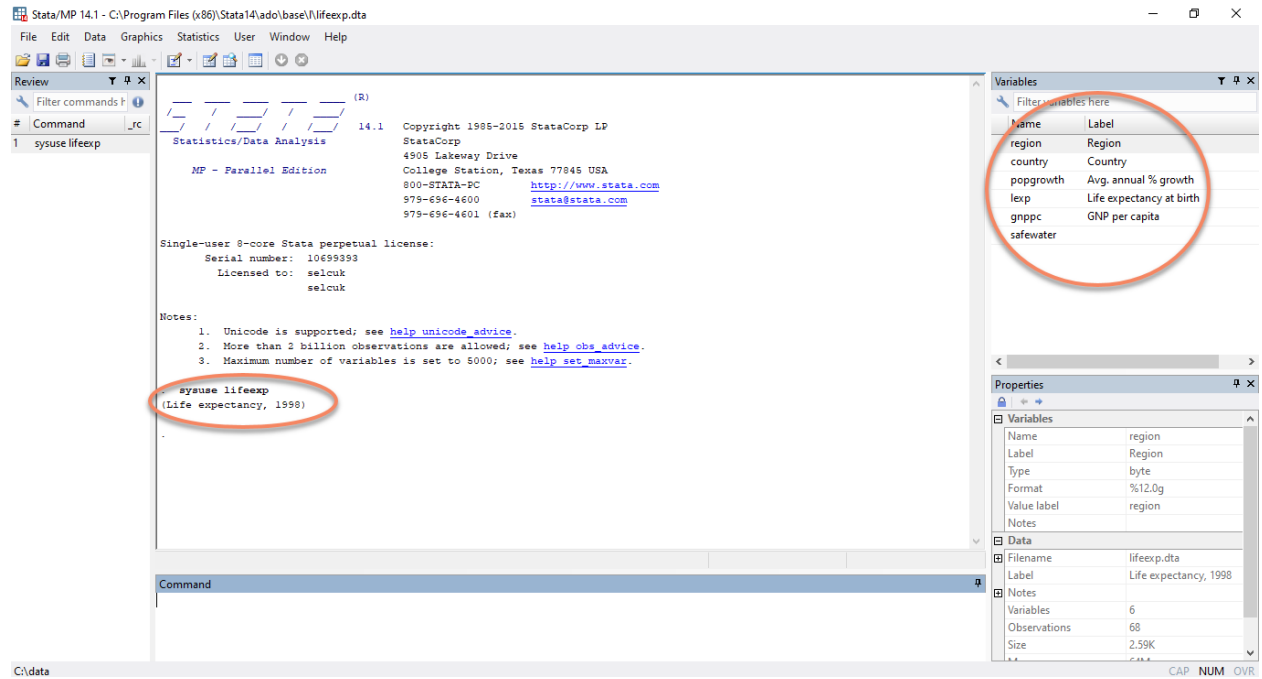
Gerekli adımlar takip edildikten sonra aşağıda görüleceği üzere veri seti Stata'ya aktarılmıştır.



Şimdi Stata programının kurulumu ile birlikte gelen ve farklı ülkelerdeki beklenen yaşam süresi hakkında bilgiler içeren bir veri setini açıp kullanacağız. Bu veri setine **lifeexp.dta** adı verilmiştir ve Stata programıyla aynı yerde saklanmaktadır. Kodunuzda yer alan dosya yoluna ve dosya uzantısına (.dta) yazmaya gerek yoktur (**use** komutuyla yaptığımız gibi), çünkü **sysuse** komutu Stata'ya kendi veri dosyasını kullandığını söylemektedir.

Not: Stata'nın içindeki diğer veri setlerini görmek için **sysuse dir** komutunu kullanabilirsiniz.

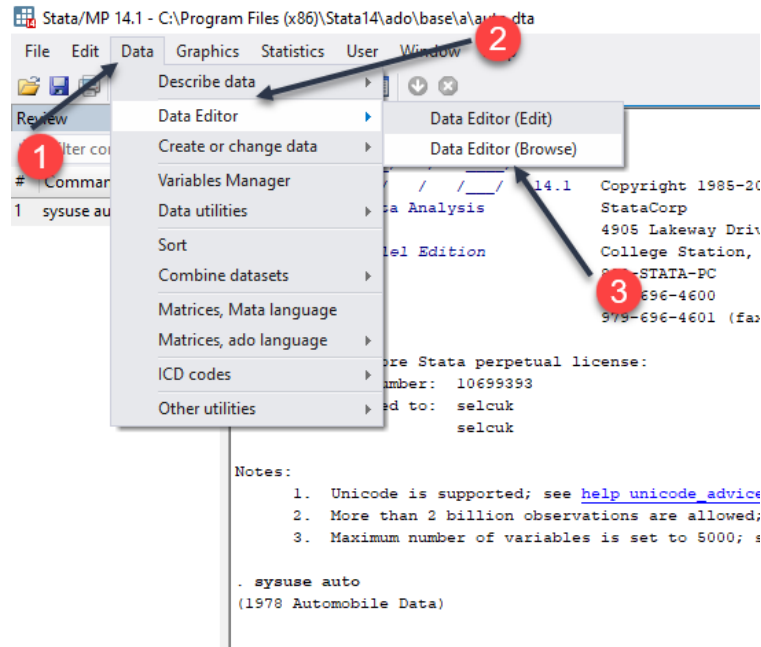
```
. sysuse lifeexp, clear
```



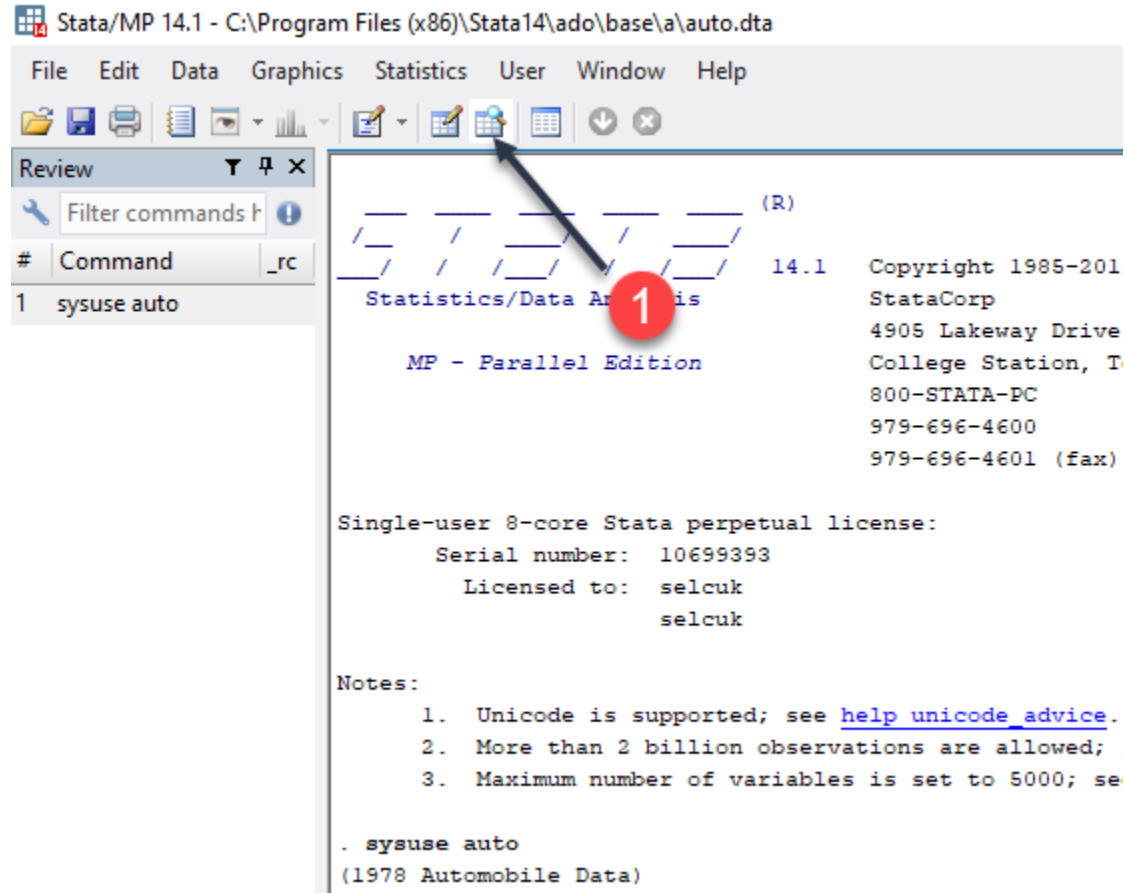
7 Veri'nin İncelenmesi

Stata'ya aktardığımız veri setine göz atmak isteyebilirsiniz. Bunu, Stata menüsünden **Data > Data Editor > Data Editor (Browse)** seçeneğine giderek, menüden gözet sekmesine tıklayarak, **browse** komutunu ya da **CTRL + 8** kısayolunu kullanarak yapabilirsiniz.

Data > Data Editor > Data Editor (Browse) seçeneği ile;



Menüden gözet sekmesine tıklayarak;



Stata/MP 14.1 - C:\Program Files (x86)\Stata14\ado\base\auto.dta

File Edit Data Graphics Statistics User Window Help

Review Filter commands

#	Command	_rc
1	sysuse auto	

```

----- (R)
Statistics/Data Analysis 14.1 Copyright 1985-201
MP - Parallel Edition StataCorp
4905 Lakeway Drive
College Station, TX 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)

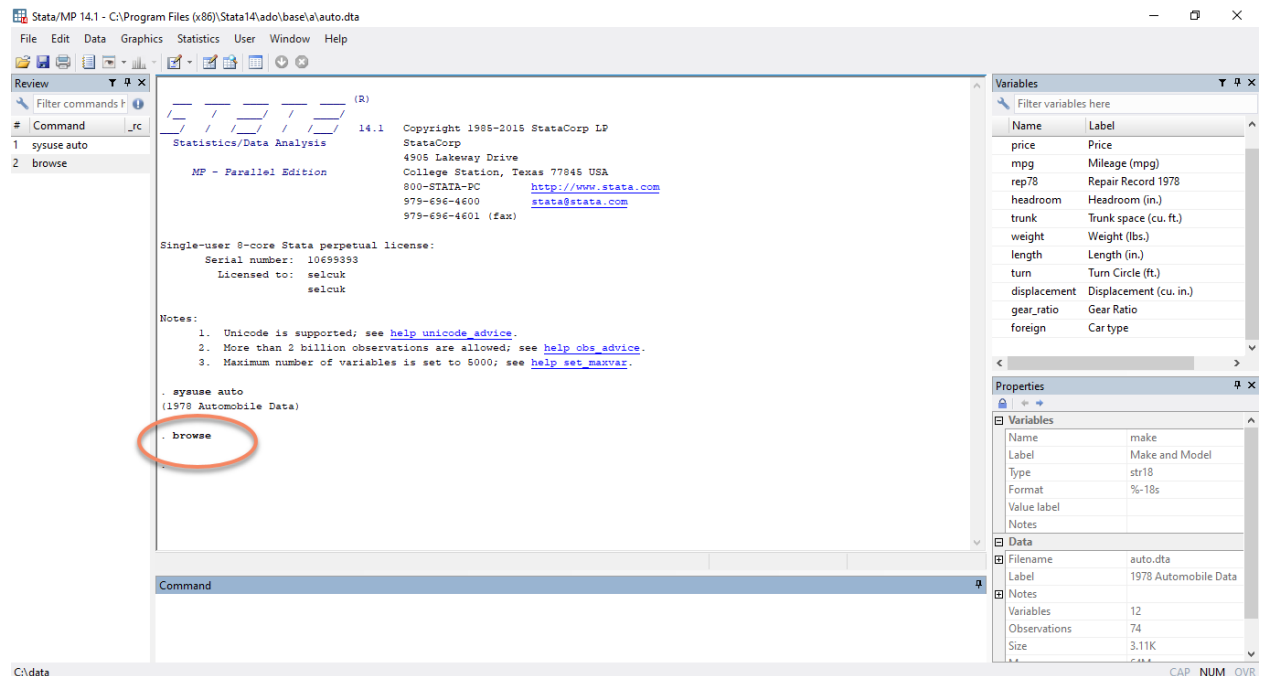
Single-user 8-core Stata perpetual license:
Serial number: 10699393
Licensed to: selcuk
selcuk

Notes:
1. Unicode is supported; see help unicode advice.
2. More than 2 billion observations are allowed; see help obs advice.
3. Maximum number of variables is set to 5000; see help set maxvar.

. sysuse auto
(1978 Automobile Data)

```

. browse



Stata/MP 14.1 - C:\Program Files (x86)\Stata14\ado\base\auto.dta

File Edit Data Graphics Statistics User Window Help

Review Filter commands

#	Command	_rc
1	sysuse auto	
2	browse	

```

----- (R)
Statistics/Data Analysis 14.1 Copyright 1985-2015 StataCorp LP
MP - Parallel Edition StataCorp
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 8-core Stata perpetual license:
Serial number: 10699393
Licensed to: selcuk
selcuk

Notes:
1. Unicode is supported; see help unicode advice.
2. More than 2 billion observations are allowed; see help obs advice.
3. Maximum number of variables is set to 5000; see help set maxvar.

. sysuse auto
(1978 Automobile Data)

. browse

```

Name	Label
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. in.)
gear_ratio	Gear Ratio
foreign	Car type

Name	Label
make	Make and Model
str18	%-18s

Filename	Label
auto.dta	1978 Automobile Data
Notes	
Variables	12
Observations	74
Size	3.11K

C:\data CAP NUM OVR

Dört farklı şekilde de veri düzenleyicisi açılır ve aşağıdaki ekran görüntüsüne ulaşılır.

	make	price	mpg	rep78	headroom	trunk	weight	length	turn	displacement	gear_ratio	foreign
1	AMC Concord	4,099	22	3	2.5	11	2,930	186	40	121	3.58	Domestic
2	AMC Pacer	4,749	17	3	3.0	11	3,350	173	40	258	2.53	Domestic
3	AMC Spirit	3,799	22	.	3.0	12	2,640	168	35	121	3.08	Domestic
4	Buick Century	4,816	20	3	4.5	16	3,250	196	40	196	2.93	Domestic
5	Buick Electra	7,827	15	4	4.0	20	4,080	222	43	350	2.41	Domestic
6	Buick LeSabre	8,788	18	3	4.0	21	3,670	218	43	231	2.73	Domestic
7	Buick Opel	4,453	26	.	3.0	10	2,230	170	34	304	2.87	Domestic
8	Buick Regal	5,189	20	3	2.0	16	3,280	200	42	196	2.93	Domestic
9	Buick Riviera	10,372	16	3	3.5	17	3,980	207	43	231	2.93	Domestic
10	Buick Skylark	4,082	19	3	3.5	13	3,400	200	42	231	3.08	Domestic
11	Cad. Deville	11,385	14	3	4.0	20	4,330	221	44	425	2.28	Domestic
12	Cad. Eldorado	14,500	14	2	3.5	16	3,900	204	43	350	2.19	Domestic
13	Cad. Seville	15,506	21	3	3.0	13	4,250	204	45	350	2.24	Domestic
14	Chev. Chevette	3,299	29	3	2.5	9	2,110	163	34	231	2.93	Domestic
15	Chev. Impala	5,705	16	4	4.0	20	3,690	212	43	250	2.56	Domestic
16	Chev. Malibu	4,504	22	3	3.5	17	3,180	183	31	200	2.73	Domestic
17	Chev. Monte Carlo	5,104	22	2	2.0	16	3,220	200	41	200	2.73	Domestic
18	Chev. Monza	3,667	24	2	2.0	7	2,750	175	40	151	2.73	Domestic
19	Chev. Nova	3,555	19	3	3.5	13	3,430	197	43	250	2.56	Domestic
20	Dodge Colt	3,584	30	5	2.0	8	2,120	163	35	98	3.54	Domestic
21	Dodge Diplomat	4,010	18	2	4.0	17	3,600	206	46	318	2.47	Domestic
22	Dodge Magnum	5,886	16	2	4.0	17	3,600	206	46	318	2.47	Domestic
23	Dodge St. Regis	6,342	17	2	4.5	21	3,740	220	46	225	2.94	Domestic
24	Ford Fiesta	4,389	28	4	1.5	9	1,800	147	33	98	3.15	Domestic
25	Ford Mustang	4,187	21	3	2.0	10	2,650	175	43	140	3.08	Domestic
26	Linc. Continental	11,497	12	3	3.5	22	4,840	233	51	400	2.47	Domestic

Veri düzenleyicisinde görüleceği üzere her bir sütunda değişkenler ve onların değerleri yer almaktadır.

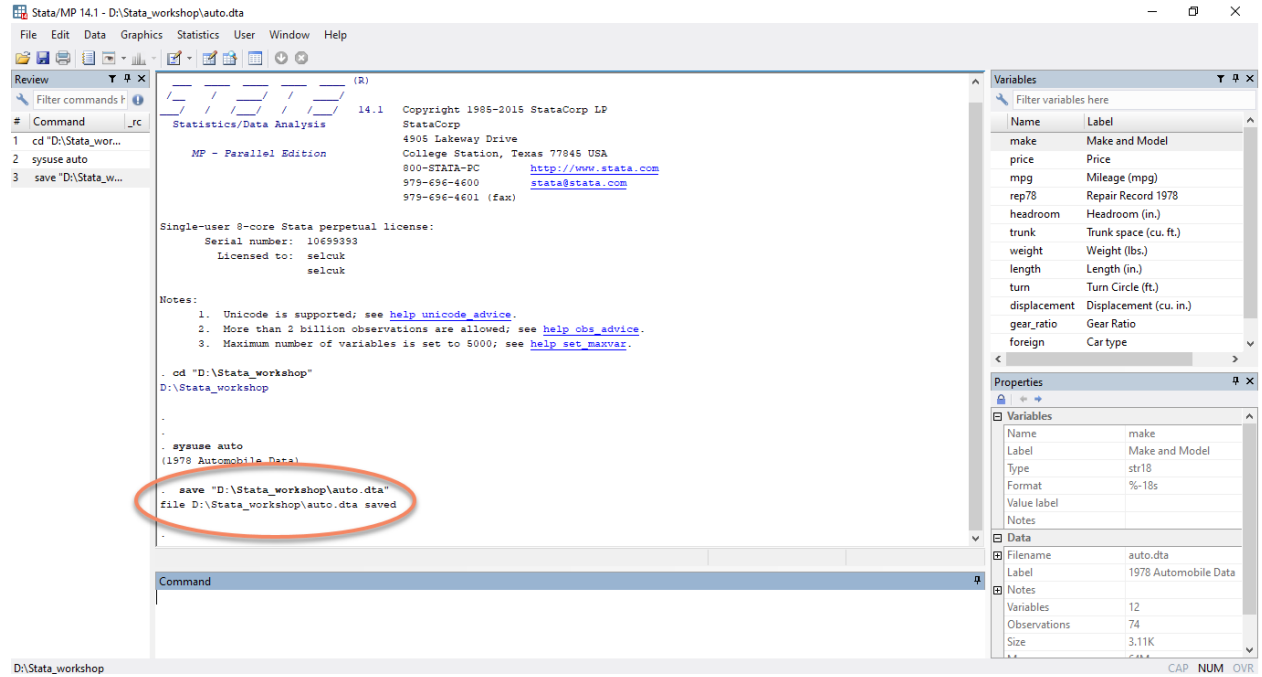
8 Veri'nin Kaydedilmesi

Ekonometrik bir analiz için veri seti oluşturduğunuzda veya size bir veri seti verildiğinde, bu dosyaları indirilenler klasörünüzden veya masaüstünüzden kullanmak gerçekten kötü bir uygulamadır. Bu dosyaları kaydetmeye karar verdikten sonra dosyaların kaydedilmesi için en makul yer çalışma klasörünüzdür. Dosyalarınızı çalışma klasörünüze kaydetmek için ya Stata menüsünden faydalanabilir ya da **save** komutunu kullanabilirsiniz.

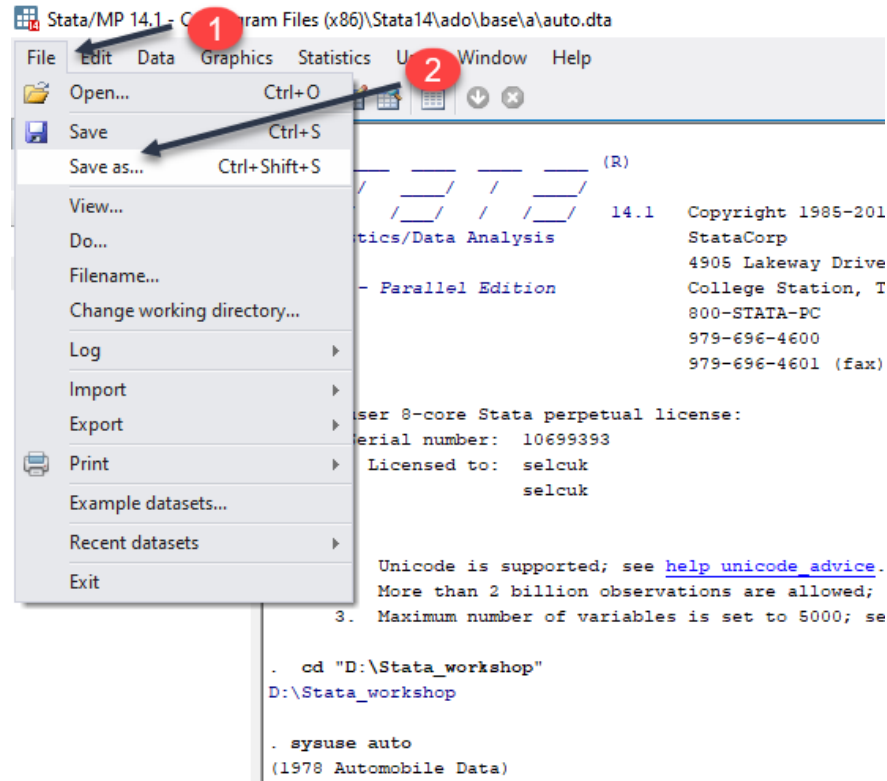
Örneğin Stata içinde yer alan **auto.dta** veri setini Stata'ya aktarınız (hatırlarsanız bu kısmın nasıl yapılacağını veri girişi başlığı altında öğrenmişsiniz) ve bu veri setini çalışma klasörünüzün içine aşağıdaki kodu kullanarak kaydediniz.

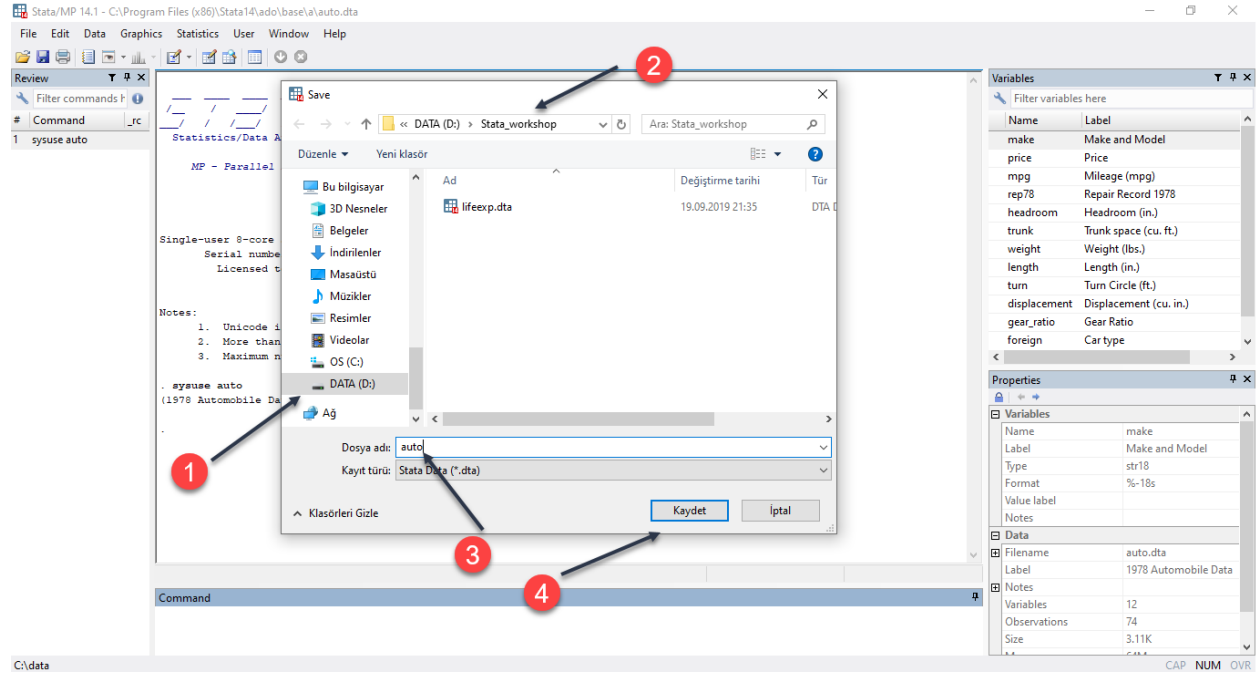
Not: **auto** isimli veri seti Amerika'da satılan 1978 model araçlarla ilgili bilgilerin yer aldığı bir veri setidir.

. save "D:\Stata_workshop\auto.dta"

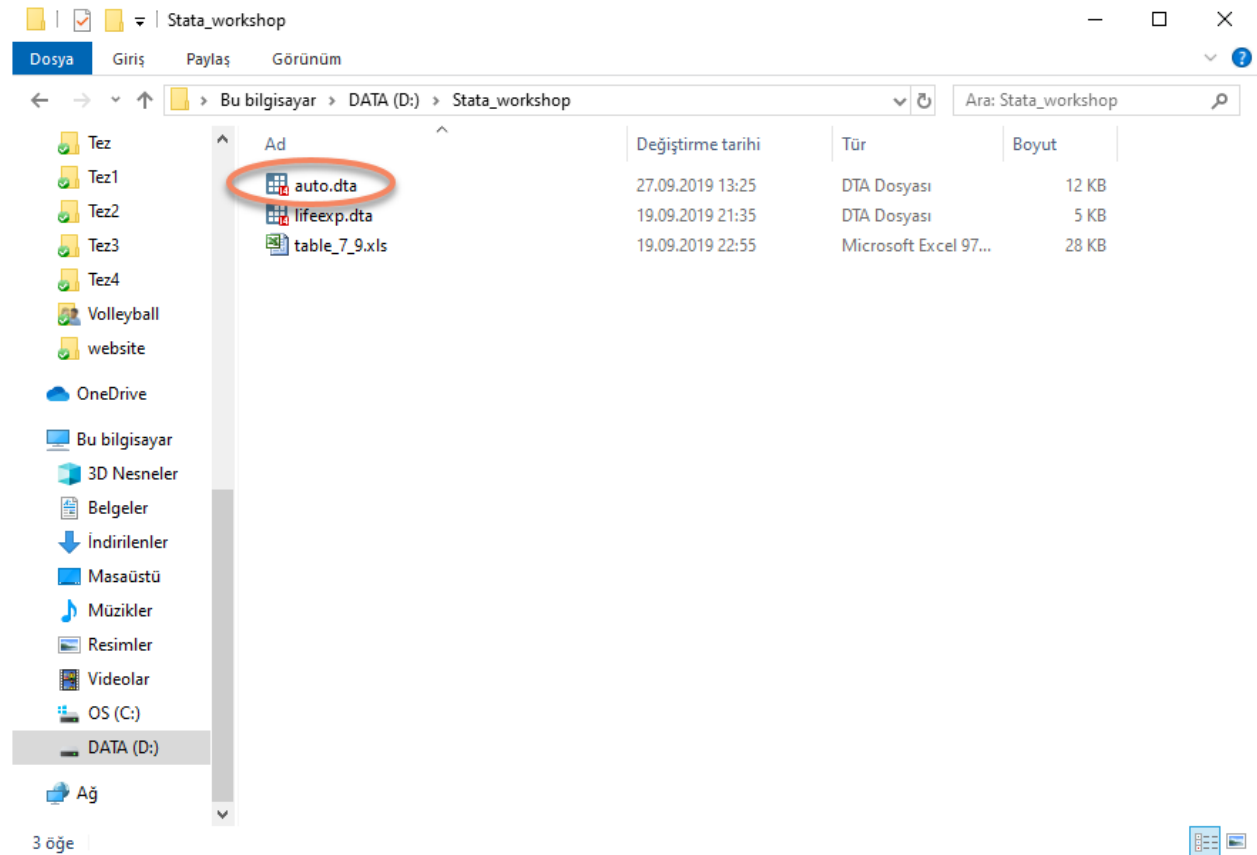


Stata menüsünden ise **File > Save as > D > Stata_workshop > Dosya adı yazınız (auto) > Kaydet** adımlarını izleyerek veri setinizi kaydedebilirsiniz.





İki farklı şekilde de veri setini kaydettikten sonra “D” sürücüsünde **Stata_workshop** klasörünün içinde **auto.dta** isimli veri setinin yer aldığı aşağıdaki ekran görüntüsünde görülmektedir.



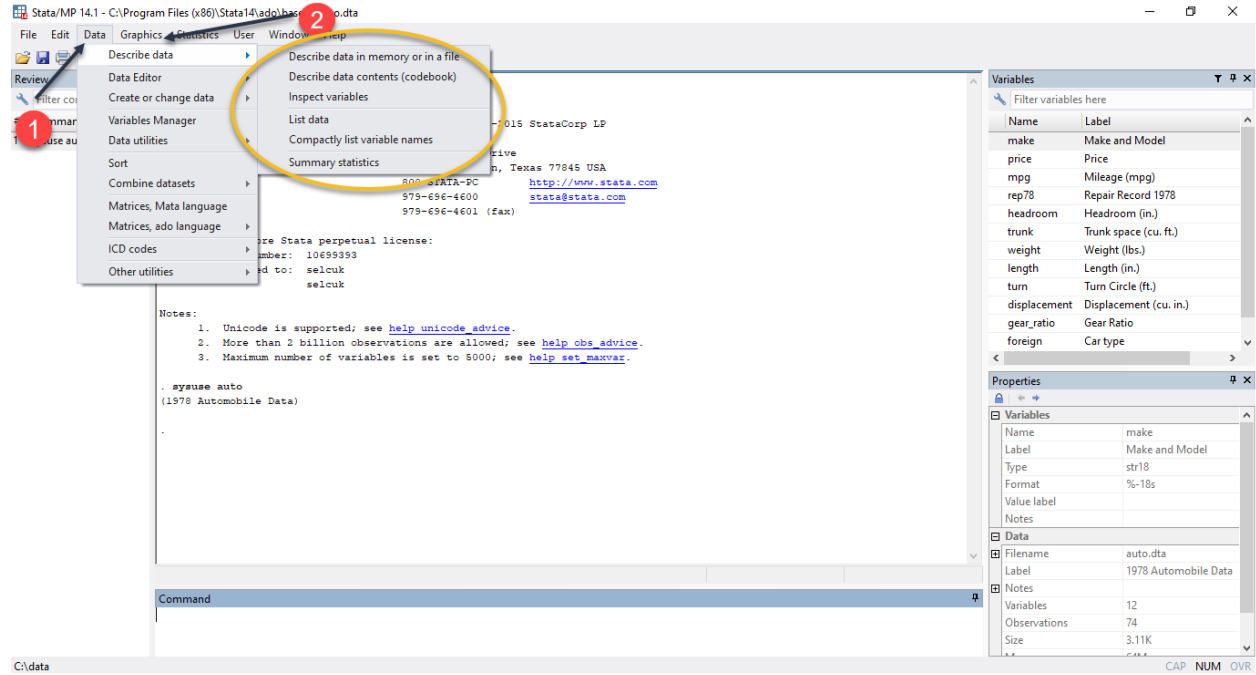
Çoğu zaman elinizdeki veri seti üzerinde bazı değişiklikler yapmak ve bu değişiklikleri kayıtlı veri seti üzerine yazdırmak isteyebilirsiniz. Böyle bir durumda aşağıdaki kodu komut penceresine yazarak bu işlemi gerçekleştirebilirsiniz.

```
. save "D:\Stata_workshop\auto.dta", replace
```

9 Veri Setine Ait Tanımlayıcı İstatistikler

Ekonometrik analize başlamadan önce, elinizdeki veri setini tanımanız gereklidir. Veri setinizin özelliklerini anlamaya yardımcı olacak birkaç komut vardır. Bu komutların bazıları benzer bilgiler sunar ve hangisini kullanmayı tercih edeceğinize size kalmıştır. Yine, ekonometrik analize başlamadan önce, her zaman tüm değişkenlerin frekanslarını, kategorik değişkenlerin nasıl kodlandığını, minimum, maksimum değerleri ve kayıp gözlem sayısını kontrol etmelisiniz. Bu, veri setindeki uç noktaları ve olası hataları ortaya çıkarmanın en iyi yoludur.

Veri setine ait tanımlayıcı istatistiklere **codebook**, **list**, **tab**, **count**, **describe** ve **summarize** komutlarını kullanarak ulaşabilirsiniz. Bu komutların çoğunu Stata menüsünden **Data > Describe Data > Açılan Sekmelerden İstedğiniz Sekmeye Tıklayın** takip ederek çalıştırabilirsiniz.



Komutları **auto.dta** veri setini üzerinde çalıştırarak aşağıdaki özet istatistiklere ulaşabilirsiniz.

`codebook` komutu ile değişkenlerin özelliklerini özetleyebilirsiniz.

. codebook

Stata/MP 14.1 - C:\Program Files (x86)\Stata14\ado\base\auto\auto.dta

File Edit Data Graphics Statistics User Window Help

Review Filter commands t

Command _rc . codebook

1 sysuse auto

2 codebook

make Make and Model

type: string (str18), but longest is str17

unique values: 74 missing "": 0/74

examples: "Cad. Deville"
"Dodge Magnum"
"Merc. XR-7"
"Pont. Catalina"

warning: variable has embedded blanks

price Price

type: numeric (int)

range: [3291,15906] units: 1

unique values: 74 missing .: 0/74

mean: 6165.26

std. dev: 2549.5

Command

Variables

Filter variables here

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. in.)
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. in.)
gear_ratio	Gear Ratio
foreign	Car type

Data

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

C:\data CAP_NUM_OVR

`list` komutu ile gözlemleri listeleyebilirsiniz.

. list

Stata/MP 14.1 - C:\Program Files (x86)\Stata14\ado\base\auto\auto.dta

File Edit Data Graphics Statistics User Window Help

Review Filter commands t

Command _rc

1 sysuse auto

2 list

Serial number: 10699393

Licensed to: selcuk selcuk

Notes:

1. Unicode is supported; see [help unicode_advice](#).
2. More than 2 billion observations are allowed; see [help obs_advice](#).
3. Maximum number of variables is set to 5000; see [help set_maxvar](#).

. sysuse auto

(1978 Automobile Data)

. list

	make	price	mpg	rep78	headroom	trunk	weight	length	turn	displacement	gear_ratio	foreign
1.	AMC Concord	4,099	22	3	2.5	11	2,930	186	40	121	3.58	Domestic
2.	AMC Pacer	4,749	17	3	3.0	11	3,850	173	40	258	2.53	Domestic
3.	AMC Spirit	3,799	22	.	3.0	12	2,640	168	35	121	3.08	Domestic
4.	Buick Century	4,816	20	3	4.5	16	3,250	196	40	196	2.93	Domestic
5.	Buick Electra	7,827	15	4	4.0	20	4,080	222	43	350	2.41	Domestic
6.	Buick LeSabre	5,788	18	3	4.0	21	3,670	218	43	231	2.73	Domestic
7.	Buick Opel	4,463	26	.	3.0	10	2,230	170	34	304	2.87	Domestic
8.	Buick Regal	5,189	20	3	2.0	16	3,280	200	42	196	2.93	Domestic
9.	Buick Riviera	10,372	16	3	3.5	17	3,880	207	43	231	2.93	Domestic
10.	Buick Skylark	4,082	19	3	3.5	13	3,400	200	42	231	3.08	Domestic
11.	Cad. Deville	11,385	14	3	4.0	20	4,330	211	44	425	2.28	Domestic
12.	Cad. Eldorado	14,500	14	2	3.5	16	3,900	204	43	350	2.19	Domestic
13.	Cad. Seville	15,906	21	3	3.0	13	4,250	204	45	350	2.24	Domestic
14.	Chev. Chevette	3,299	29	3	2.5	9	2,110	163	34	231	2.93	Domestic
15.	Chev. Tmala	5,705	16	4	4.0	20	3,690	212	43	250	2.56	Domestic

Command

Variables

Filter variables here

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. in.)
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. in.)
gear_ratio	Gear Ratio
foreign	Car type

Data

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

C:\data CAP_NUM_OVR

Not: `list` komutunun devamına **in 1/10** yazarsanız 1. satırdan 10. satıra kadar olan verileri listelersiniz.

tab komutu ile seçili değişken ve değişkenler için frekans tabloları oluşturabilirsiniz.

. tab price

Stata/MP 14.1 - C:\Program Files (x86)\Stata14\ado\base\auto.dta

File Edit Data Graphics Statistics User Window Help

Review Filter commands Filter commands | Command _rc

1 sysuse auto
2 tab price

Notes:

1. Unicode is supported; see [help unicode advice](#).
2. More than 2 billion observations are allowed; see [help obs advice](#).
3. Maximum number of variables is set to 5000; see [help set maxvar](#).

```
. sysuse auto
(1978 Automobile Data)

. tab price
```

Price	Freq.	Percent	Cum.
3,291	1	1.35	1.35
3,295	1	1.35	2.70
3,667	1	1.35	4.05
3,748	1	1.35	5.41
3,759	1	1.35	6.76
3,799	1	1.35	8.11
3,825	1	1.35	9.46
3,895	1	1.35	10.81
3,955	1	1.35	12.16
3,984	1	1.35	13.51
3,995	1	1.35	14.86
4,010	1	1.35	16.22
4,060	1	1.35	17.57
4,082	1	1.35	18.92
4,095	1	1.35	20.27
4,172	1	1.35	21.62
4,181	1	1.35	22.97
4,187	1	1.35	24.32
4,195	1	1.35	25.68
4,296	1	1.35	27.03
4,389	1	1.35	28.38
4,424	1	1.35	29.73

Command

C:\data CAP NUM OVR

describe komutu size yine değişkenlerin özelliklerini verir.

. describe

Stata/MP 14.1 - C:\Program Files (x86)\Stata14\ado\base\auto.dta

File Edit Data Graphics Statistics User Window Help

Review Filter commands Filter commands | Command _rc

1 sysuse auto
2 describe

Notes:

1. Unicode is supported; see [help unicode advice](#).
2. More than 2 billion observations are allowed; see [help obs advice](#).
3. Maximum number of variables is set to 5000; see [help set maxvar](#).

```
. sysuse auto
(1978 Automobile Data)

. describe
```

Contains data from C:\Program Files (x86)\Stata14\ado\base\auto.dta

obs: 74 1978 Automobile Data
vars: 12 18 Apr 2014 17:45
size: 3,182 (_dta has notes)

variable name	storage type	display format	value label	variable label
make	strl8	%-18s		Make and Model
price	int	%8.0gc		Price
mpg	int	%8.0g		Mileage (mpg)
rep78	int	%8.0g		Repair Record 1978
headroom	float	%8.1f		Headroom (in.)
trunk	int	%8.0g		Trunk space (cu. ft.)
weight	int	%8.0gc		Weight (lbs.)
length	int	%8.0g		Length (in.)
turn	int	%8.0g		Turn Circle (ft.)
displacement	int	%8.0g		Displacement (cu. in.)
gear_ratio	float	%6.2f		Gear Ratio
foreign	byte	%8.0g	origin	Car type

Sorted by: foreign

Command

C:\data CAP NUM OVR

summarize komutu ile değişkenlerin maksimum, minimum, ortalama, standart sapma ve gözlem sayısı gibi özelliklerine ulaşabilirsiniz.

Not: **codebook** komutu **compact** opsiyonu ile kullanılırsa summarize komutunun sonucuna benzer bir çıktı elde edilir.

. summarize

Single-user 8-core Stata perpetual license:
Serial number: 10659393
Licensed to: selcuk
selcuk

Notes:
1. Unicode is supported; see [help unicode advice](#).
2. More than 2 billion observations are allowed; see [help obs advice](#).
3. Maximum number of variables is set to 5000; see [help set maxvar](#).

```
. sysuse auto
(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	5.798503	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	4940
length	74	187.9324	22.26634	142	233
turn	74	35.64865	4.395354	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

10 Veri Seti ve Değişkenler ile Yapılan Temel İşlemler

Veri Seti ile Yapılan Temel İşlemler

Stata, veriyi **numerik** veya **metin (string)** olmak üzere iki şekilde saklar. Stata'nın metin olarak sakladığı veriler belirli bilgiler vermek için kullanılır. Örneğin **auto** veri setinde araba markaları metin veriler olarak tutulmaktadır. Veri düzenleyicisinde araba markalarının üzerine tıklandığında hiçbir sayısal değer almadığı görülecektir. Bunun yanı sıra Stata'da metin verilerinin varsayılan rengi **kırmızıdır**.

Not: Metin verileri görünürde sayısal değerler alabilir fakat analizlerde kullanılamazlar. Çünkü Stata bu tür verilere özünde sayısal değerler atamaz.

Aşağıda görüleceği üzere **make** değişkeni içinde 8. satırdaki hücrenin üzerine tıklandığında veri tarayıcısında bu hücrenin sayısal bir değer almadığı **Buick Regal** adında bir metin olduğu görülmektedir.

The screenshot shows the Stata Data Editor window for the file [auto.dta]. The 'make' variable is selected, and the 'Buick Regal' row is highlighted in yellow. The 'price' variable is also highlighted in yellow. The 'Variables' panel on the right shows the list of variables and their properties.

make	price	mpg	rep78	headroom	trunk
1 AMC Concord	4,099	22	3	2.5	11
2 AMC Pacer	4,749	17	3	3.0	11
3 AMC Spirit	3,799	22	.	3.0	12
4 Buick Century	4,816	20	3	4.5	16
5 Buick Electra	7,827	15	4	4.0	20
6 Buick LeSabre	5,788	18	3	4.0	21
7 Buick Opel	4,453	26	.	3.0	10
8 Buick Regal	5,189	20	3	2.0	16
9 Buick Riviera	10,372	16	3	3.5	17
10 Buick Skylark	4,082	19	3	3.5	13
11 Cad. Deville	11,385	14	3	4.0	20
12 Cad. Eldorado	14,500	14	2	3.5	16
13 Cad. Seville	15,906	21	3	3.0	13
14 Chev. Chevette	3,299	29	3	2.5	9
15 Chev. Impala	5,705	16	4	4.0	20
16 Chev. Malibu	4,504	22	3	3.5	17
17 Chev. Monte Carlo	5,104	22	2	2.0	16
18 Chev. Monza	3,667	24	2	2.0	7
19 Chev. Nova	3,955	19	3	3.5	13
20 Dodge Colt	3,984	30	5	2.0	8
21 Dodge Diplomat	4,010	18	2	4.0	17
22 Dodge Magnum	5,886	16	2	4.0	17
23 Dodge St. Regis	6,342	17	2	4.5	21
24 Ford Fiesta	4,389	28	4	1.5	9
25 Ford Mustang	4,199	21	3	2.0	10

Stata'da eğer bir veri metin olarak saklanmıyorsa o veri mutlaka numerik olarak saklanmaktadır. **auto** veri setinde **make** değişkeni dışındaki diğer tüm değişkenler numerik verilerden oluşmaktadır. Örneğin **price** değişkeni içinde 4. satırdaki hücrenin üzerine tıkladığında veri tarayıcısında bu hücrenin sayısal değer aldığı görülmektedir.

The screenshot shows the Stata Data Editor window for the file [auto.dta]. The 'price' variable is selected, and the '4816' value in the 'price' column for the 'Buick Century' row is highlighted in yellow. The 'Variables' panel on the right shows the list of variables and their properties.

make	price	mpg	rep78	headroom	trunk
1 AMC Concord	4,099	22	3	2.5	11
2 AMC Pacer	4,749	17	3	3.0	11
3 AMC Spirit	3,799	22	.	3.0	12
4 Buick Century	4,816	20	3	4.5	16
5 Buick Electra	7,827	15	4	4.0	20
6 Buick LeSabre	5,788	18	3	4.0	21
7 Buick Opel	4,453	26	.	3.0	10
8 Buick Regal	5,189	20	3	2.0	16
9 Buick Riviera	10,372	16	3	3.5	17
10 Buick Skylark	4,082	19	3	3.5	13
11 Cad. Deville	11,385	14	3	4.0	20
12 Cad. Eldorado	14,500	14	2	3.5	16
13 Cad. Seville	15,906	21	3	3.0	13
14 Chev. Chevette	3,299	29	3	2.5	9
15 Chev. Impala	5,705	16	4	4.0	20
16 Chev. Malibu	4,504	22	3	3.5	17
17 Chev. Monte Carlo	5,104	22	2	2.0	16
18 Chev. Monza	3,667	24	2	2.0	7
19 Chev. Nova	3,955	19	3	3.5	13
20 Dodge Colt	3,984	30	5	2.0	8
21 Dodge Diplomat	4,010	18	2	4.0	17
22 Dodge Magnum	5,886	16	2	4.0	17
23 Dodge St. Regis	6,342	17	2	4.5	21
24 Ford Fiesta	4,389	28	4	1.5	9
25 Ford Mustang	4,199	21	3	2.0	10

Herhangi bir veri setinde karşınıza çıkabilecek olası bir durum **kayıp (missing)** verilerin olmasıdır. Stata kayıp verileri pozitif büyük değerler olarak algılar ve veri düzenleyicisinde kayıp veriler “.” olarak görünür. Aşağıda görüleceği üzere **auto** veri setindeki **rep78** değişkeni içinde kayıp veriler bulunmaktadır.

The screenshot displays the Stata Data Editor window for the file 'auto.dta'. The main window shows a list of 24 automobiles with columns for 'make', 'price', 'mpg', 'rep78', 'headroom', and 'trunk'. The 'rep78' column is highlighted with a yellow circle and an arrow. The 'Variables' panel on the right shows the properties of the 'rep78' variable, including its label 'Repair Record 1978', type 'int', and format '%8.0g'.

make	price	mpg	rep78	headroom	trunk
1 AMC Concord	4,099	22	3	2.5	11
2 AMC Pacer	4,749	17	3	3.0	11
3 AMC Spirit	3,799	22	.	3.0	12
4 Buick Century	4,816	24	3	4.5	16
5 Buick Electra	7,827	15	4	4.0	20
6 Buick LeSabre	5,788	18	3	4.0	21
7 Buick Opel	4,453	16	.	3.0	10
8 Buick Regal	5,189	10	3	2.0	16
9 Buick Riviera	10,372	16	3	3.5	17
10 Buick Skylark	4,082	19	3	3.5	13
11 Cad. Deville	11,385	14	3	4.0	20
12 Cad. Eldorado	14,500	14	2	3.5	16
13 Cad. Seville	15,906	11	3	3.0	13
14 Chev. Chevette	3,299	39	3	2.5	9
15 Chev. Impala	5,705	16	4	4.0	20
16 Chev. Malibu	4,504	12	3	3.5	17
17 Chev. Monte Carlo	5,104	22	2	2.0	16
18 Chev. Monza	3,667	24	2	2.0	7
19 Chev. Nova	3,555	11	3	3.5	13
20 Dodge Colt	3,984	30	5	2.0	8
21 Dodge Diplomat	4,010	18	2	4.0	17
22 Dodge Magnum	5,886	16	2	4.0	17
23 Dodge St. Regis	6,342	17	2	4.5	21
24 Ford Fiesta	4,389	28	1	1.5	9

Veri seti üzerinde yapılabilecek düzenlemelerin başında elinizdeki veri setinin istediğiniz bölümünü tutmak veya atmak gelmektedir. Veri setinin belirli bir bölümünü tutmak için **keep**, atmak içinse **drop** komutu kullanılır. Aynı zamanda bu komutlar değişkenlerin veri setinde tutulması ya da atılması için de kullanılır.

Örneğin **auto.dta** veri setinde yalnızca fiyatı (price) 6000'e eşit ya da daha büyük olan araçları tutmak istiyoruz. Başka bir deyişle fiyatı 6000'den daha düşük olan araçları veri setinden atmak istiyoruz. Bunu yapmak için aşağıdaki Stata komutlarını kullanabilirsiniz.

```
. keep if price >= 6000 /* fiyatı 6000'ne eşit veya büyük olan araçları tutar */
. drop if price < 6000 /* fiyatı 6000'den küçük olan araçları atar */
```

Bu işlemlerin sonunda veri düzenleyicisini açtığımızda aşağıdaki ekran görüntüsü elde edilir.

Stata/MP 14.1 - C:\Program Files (x86)\Stata14\ado\base\auto\auto.dta

File Edit Data Graphics Statistics User Window Help

Review Filter commands t

Command _rc

1 sysuse auto

2 keep if price >= ...

MP - Parallel Edition

14.1 Copyright 1985-2015 StataCorp LP
StataCorp
4905 Lakeview 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 8-core Stata perpetual license:
Serial number: 106959393
Licensed to: selcuk
selcuk

Notes:

1. Unicode is supported; see [help unicode advice](#).
2. More than 2 billion observations are allowed; see [help obs advice](#).
3. Maximum number of variables is set to 5000; see [help set maxvar](#).

. sysuse auto
(1978 Automobile Data)

. keep if price >= 6000
(51 observations deleted)

help obs_advice

Command

C:\data

Data Editor (Edit) - [auto.dta]

File Edit View Data Tools

make[1] Buick Electra

make	price	mpg	rep78
1 Buick Electra	7,827	16	4
2 Buick Riviera	10,372	16	3
3 Cad. Deville	11,385	14	3
4 Cad. Eldorado	14,500	14	2
5 Cad. Seville	18,906	21	3
6 Dodge St. Regis	6,342	17	2
7 Linc. Continental	11,497	12	3
8 Linc. Mark V	13,554	12	2
9 Linc. Versailles	13,466	14	3
10 Merc. Marquis	6,165	16	3
11 Merc. XR-7	6,303	14	4
12 Olds 98	8,814	21	4
13 Olds Toronado	10,371	16	3
14 Plym. Sapporo	6,486	26	.
15 Audi 5000	9,650	17	5
16 Audi Fox	6,295	23	3
17 BMW 320i	9,735	26	4
18 Datsun 200	6,229	23	4
19 Datsun 810	8,129	21	4
20 Peugeot 604	12,990	14	.
21 VW Dasher	7,140	23	4
22 VW Scirocco	6,850	25	4
23 Volvo 260	11,955	17	5

Variables

Filter variables here

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. in.)

gear_ratio Gear Ratio

Variables % Snapshots

Properties

Variables

Name make

Label Make and Model

Type str18

Format %-18s

Value label

Notes

Data

Filename auto.dta

Label 1978 Automot

Notes

Variables 12

Observations 23

Size 989

Length: 18 Vars: 12 Order: Dataset Obs: 23 Filter: Off Mode: E

Görüleceği üzere bu komutun çalışması ile birlikte 51 gözlem silinmiştir. Yani 51 tane arabanın fiyatı 6000'den daha küçük olduğu için veri setinde tutulamamış atılmıştır. Elinizdeki veri seti küçük gözlem sayısına sahip olduğu için veri tarayıcısından arabaların fiyatlarını kontrol etmek oldukça kolaydır. Fakat elinizdeki veri setindeki gözlem sayısı çok büyük ise veri tarayıcısından değişken değerlerini kontrol etmek oldukça zorlaşacaktır. Bunun için **keep if price >= 6000** komutu çalıştırıldıktan sonra (hatırlarsanız daha önce öğrendiğiniz) **summarize** komutunu kullanarak tuttuğunuz veri setindeki en düşük araba fiyatını görebilirsiniz.

Stata/MP 14.1 - C:\Program Files (x86)\Stata14\ado\base\auto\auto.dta

File Edit Data Graphics Statistics User Window Help

Review Filter commands t

Command _rc

1 sysuse auto

2 keep if price >= ...

3 summarize

Notes:

1. Unicode is supported; see [help unicode advice](#).
2. More than 2 billion observations are allowed; see [help obs advice](#).
3. Maximum number of variables is set to 5000; see [help set maxvar](#).

. sysuse auto
(1978 Automobile Data)

. keep if price >= 6000
(51 observations deleted)

. summarize

Variable	Obs	Mean	Std. Dev.	Min	Max
make	0				
price	23	9655.696	3049.89	6165	15506
mpg	23	18.04348	4.497254	12	26
rep78	21	3.47619	.8135753	2	5
headroom	23	3.043478	.7964844	1.5	4.5
trunk	23	15.3913	4.570052	6	23
weight	23	2455.652	868.8019	1590	4840
length	23	159.3913	21.20936	156	233
turn	23	41.04348	4.587317	34	51
displacement	23	243.0435	115.7809	97	425
gear_ratio	23	2.94913	.6036549	2.19	3.89
foreign	23	.3913043	.4990109	0	1

Command

C:\data

Variables

Filter variables here

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. in.)

gear_ratio Gear Ratio

foreign Car type

Properties

Variables

Name make

Label Make and Model

Type str18

Format %-18s

Value label

Notes

Data

Filename auto.dta

Label 1978 Automobile Data

Notes

Variables 12

Observations 23

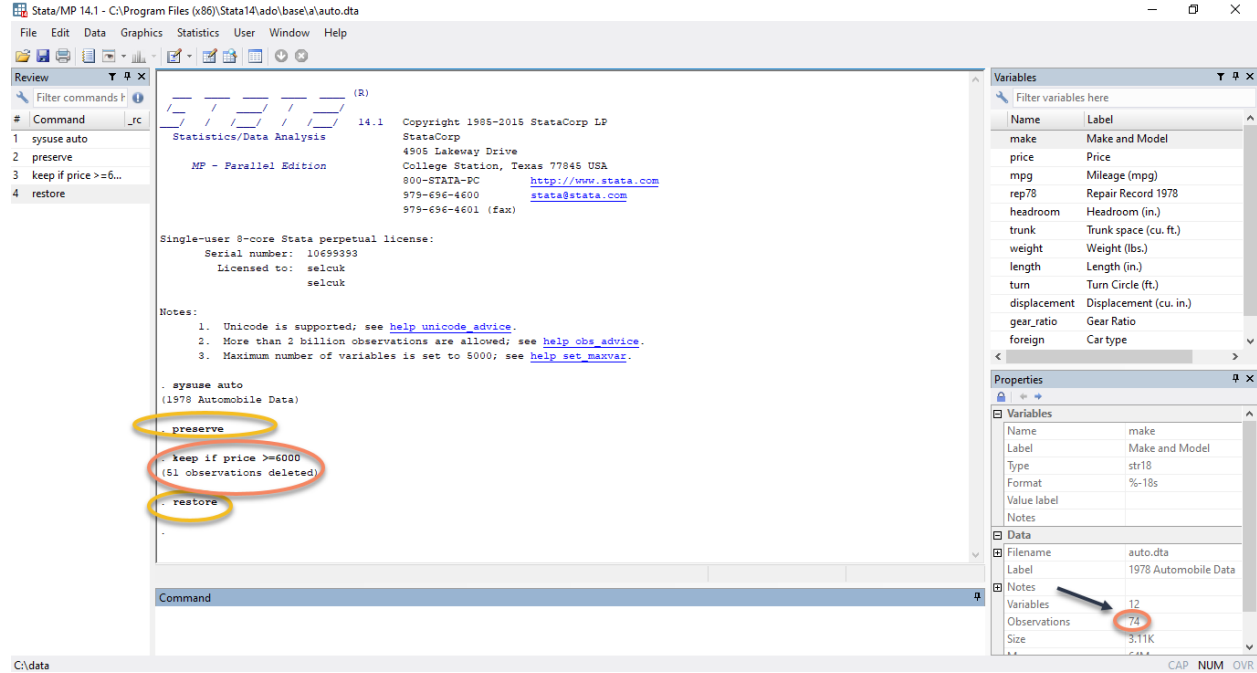
Size 989

CAP NUM OVR

Hem veri düzenleyicisinden hem de summarize komutundan sonra çıkan sonuç ekranında göreceğiniz üzere fiyatı 6000'den düşük arabalar veri setinden atılmış olup en düşük araba fiyatı 6000'den yüksektir.

Orjinal veri setine daha sonra ihtiyacınız olabileceği için **keep** ve **drop** komutlarını kullanarak yaptığımız değişikliklerden sonra yeni oluşan veri setinizi farklı bir isimle kaydetmeniz (hatırlarsanız daha önce veri setini nasıl kaydedeceğinizi öğrenmişsiniz) gerekmektedir. Fakat **preserve** ve **restore** komutları veri seti üzerinde yapılan değişikliklerden sonra yeni bir veri seti kaydetmenize gerek kalmaksızın size orjinal veri setinize ulaşımı sağlamaktadır. **preserve** komutu orjinal veri setinizin anlık görüntüsünü alır ve **restore** komutu bu orjinal veri setinize geri dönmenizi sağlar. Bunu aşağıdaki kodu kullanarak yapabilirsiniz.

```
. preserve
. keep if price >= 6000
. restore
```



Stata sonuç ekranında **keep** komutundan sonra 51 tane gözlemin silindiği görülmektedir. Fakat özellikler penceresinde görüleceği üzere başlangıçta 74 olan gözlem sayımız bu komut dizini kullanılarak yeniden elde edilmiştir.

Yukardaki örnekte olduğu gibi veri seti üzerinde düzenleme yaparken bir takım mantıksal operatörlerin kullanımına gerek duyulmaktadır. **Tablo 1**'de bu operatörler ve açıklamaları yer almaktadır.

Tablo 1: Mantıksal Operatörler

Operatörler	Açıklamalar
==	eşit
!=	eşit değil
>	den büyük
>=	den büyük ya da eşit
<	den küçük
<=	den küçük ya da eşit
&	ve
	veya
~	değil
!	değil

Değişkenler ile Yapılan Temel İşlemler

Analiz yaparken genellikle yeni değişkenler oluşturmaya ihtiyaç duyulmakta ve bunu yapmak için **generate** komutu kullanılmaktadır.

auto.dta veri setini kullanarak **generate** komutu için aşağıda birkaç örnek verilmiştir.

```
. gen lnprice = ln(price)      /* fiyat değişkeninin logaritması */
. gen pricesquare = price^2    /* fiyat değişkeninin karesi */
. gen yirmisekiz = 28         /* sabit bir değer alan değişken */
```

Not: **gen** komutu **generate** komutunun kısaltılmış halidir.

The screenshot shows the Stata Data Editor window with a table of car data. The table has columns for 'make', 'price', 'lnprice', 'pricesquare', and 'yirmisekiz'. The 'make' column lists various car models like AMC Concord, AMC Pacer, AMC Spirit, Buick Century, Buick Electra, Buick LeSabre, Buick Opel, Buick Regal, Buick Riviera, Buick Skylark, Cad. Deville, Cad. Eldorado, Cad. Seville, Chev. Chevette, Chev. Impala, Chev. Malibu, Chev. Monte Carlo, Chev. Monza, Chev. Nova, Dodge Colt, Dodge Diplomat, Dodge Magnum, Dodge St. Regis, Ford Fiesta, Ford Mustang, and Linc. Continental. The 'price' column shows values from 4,099 to 11,497. The 'lnprice' column shows values from 8.318499 to 9.349841. The 'pricesquare' column shows values from 1.68e+07 to 1.32e+08. The 'yirmisekiz' column shows the value 28 for all rows.

The Command window shows the following commands:

```
. sysuse auto
. gen lnprice = ln(price)
. gen pricesquare = price^2
. gen yirmisekiz = 28
```

The 'generate' commands are highlighted with a red circle. The Variables window on the right shows the list of variables and their properties.

Yeni değişken oluşturmak için kullanılan bir başka komut ise **egen** komutudur. Bu komut minimum, maksimum, ortalama ve toplam gibi özet istatistiklere bağlı olarak değişken oluşturmanıza izin verir.

Yine **auto** veri setini kullanarak yurtiçi ve yurtdışında üretilen araçların toplam, maksimum, minimum ve ortalama fiyatlarını **egen** komutu ile aşağıdaki şekilde bulabilirsiniz.

```
. egen totalprice = sum(price), by(foreign) /* fiyat değişkeninin toplamı */
. egen minprice = min(price), by(foreign) /* fiyat değişkeninin minimumu */
. egen maxprice = max(price), by(foreign) /* fiyat değişkeninin maksimumu */
. egen avgprice = mean(price), by(foreign) /* fiyat değişkeninin ortalaması */
```

Aşağıda görüleceği üzere bu dört yeni değişken oluşturulmuştur.

The screenshot shows the Stata 14.1 interface. The Data Editor window displays a table with columns: foreign, totalprice, minprice, maxprice, and avgprice. The rows show data for various car models, with 'make' values like Domestic and Foreign. The Command window shows the following commands:

```

. sysuse auto
. egen totalprice = sum(price), by(foreign)
. egen minprice = min(price), by(foreign)
. egen maxprice = max(price), by(foreign)
. egen avgprice = mean(price), by(foreign)

```

The 'make' variable in the Data Editor is highlighted in red, and the 'price' variable in the Command window is highlighted in yellow.

Oluşturduğunuz değişkenin ismini yanlış yazdığınızı farketmeniz ve bunu düzeltmek istiyorsanız bunun için **rename** komutunu kullanabilirsiniz.

```
. rename price fiyat
```

The screenshot shows the Stata 14.1 interface. The Command window shows the following commands:

```

. sysuse auto
. rename price fiyat

```

The 'fiyat' variable in the Variables window is highlighted in red.

Aynı zamanda en yaygın kullanılan komutlar arasında **recode** ve **replace** komutları da yer almaktadır. Bu iki komut var olan bir değişkenin değerlerinin yeniden kodlanmasını ya da değiştirilmesini sağlamaktadır. Bu komutlar kullanılarak var olan değişkenin eski değerleri üzerine yeni değerleri yazdırılabilir veya değiştirilen değerleri içeren yeni bir değişken oluşturulabilir.

Örneğin `lifeexp.dta` veri setindeki **bölge (region)** değişkeni incelendiğinde bu değişkenin **1 = “Eur & C.Asia”, 2 = “N.A.” ve 3 = “S.A.”** değerlerinden yani üç farklı kategoriden oluştuğu görülmektedir. Şimdi **2 = “N.A.” ve 3 = “S.A.”** Kuzey ve Güney Amerika bölgelerinin tek bir bölge olacak şekilde birleştirdiğiniz ve **1 = “Eur & C.Asia”** Asya ve Avrupa bölgesinin aynı değeri aldığı **region2** adlı yeni bir değişken oluşturunuz. Daha sonra bu oluşturduğunuz yeni değişkende birleştirdiğiniz Kuzey ve Güney Amerika bölgelerine **America** ismini vererek etiketlendiriniz.

Not: Değişkenlerin aldığı değerlerin neyi temsil ettiğini bilmek için genellikle bu değerler etiketlenilmektedir. Bölge değişkeninde olduğu gibi **“1”, Avrupa ve Asya’yı “2”, Kuzey Amerika ve “3”, Güney Amerika’yı** temsil etmektedir. Bu bağlamda kodlama yapılırken bölgelerin isimleri değil aldığı değerler göz önünde bulundurulurken işlemler gerçekleştirilmektedir.

Yukarıda verilen örneği aşağıdaki komutları kullanarak yapabilirsiniz.

```
. recode region (1 =1) (2/3 =2), generate(region2)
. label define region2 1 "Eur & C.Asia" 2 "America"
. label values region2 region2
```

The screenshot displays the Stata 14.1 interface. On the left, the Command window shows the following commands being executed:

```
. sysuse lifeexp
. recode region (1 =1) (2/3 =2), generate(region2)
. label define region2 1 "Eur & C.Asia" 2 "America"
. label values region2 region2
```

On the right, the Data Editor window shows the resulting dataset. The 'region' variable is highlighted in red, and the 'region2' variable is highlighted in yellow. The 'region2' variable has values 1 and 2, and labels 'Eur & C.Asia' and 'America' respectively. The 'region' variable has values 1, 2, and 3, and labels 'Eur & C.Asia', 'N.A.', and 'S.A.' respectively.

Görüleceği üzere Güney ve Kuzey Amerika tek bir bölge olacak şekilde **region2** adlı yeni bir değişken içinde oluşturulmuş ve **America** ismi ile etiketlenilmiştir.

Aynı işlemi **replace** komutunu kullanarak aşağıdaki şekilde yapabilirsiniz.

```
. gen region2 = .
. replace region2 =1 if region ==1
. replace region2 =2 if region ==2 | region ==3
. label define region2 1 "Eur & C.Asia" 2 "America"
. label values region2 region2
```

The screenshot displays the Stata software interface. On the left, the Command window shows the following commands:

```

1 sysuse lifeexp
2 gen region2 = .
3 replace region2=1 if region==1
4 replace region2=2 if region==2 | region==3
5 label define region2_label 1 "Eur & C.Asia" 2 "America"
6 label values region2 region2_label

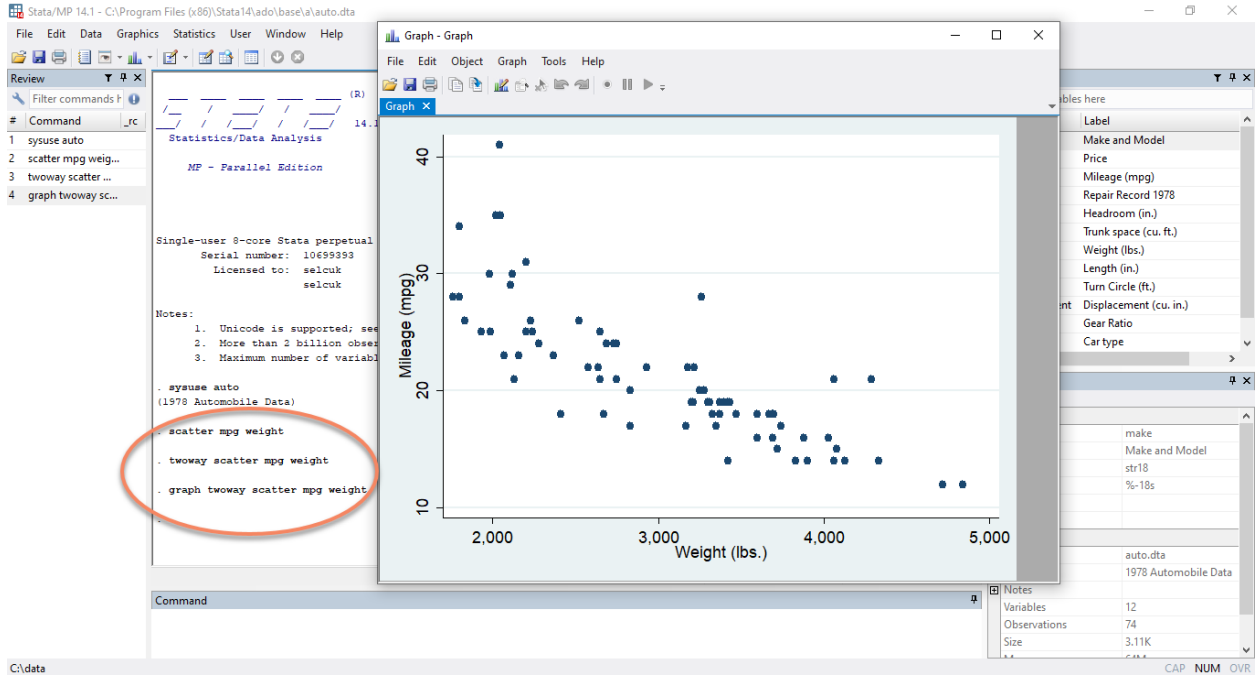
```

On the right, the Data Editor (Browse) window shows a table with two columns: 'region' and 'region2'. The 'region' column contains values 'N.A.' and 'S.A.', and the 'region2' column contains values 'America'. A yellow oval highlights the command window, and a red oval highlights the 'region' column in the data editor.

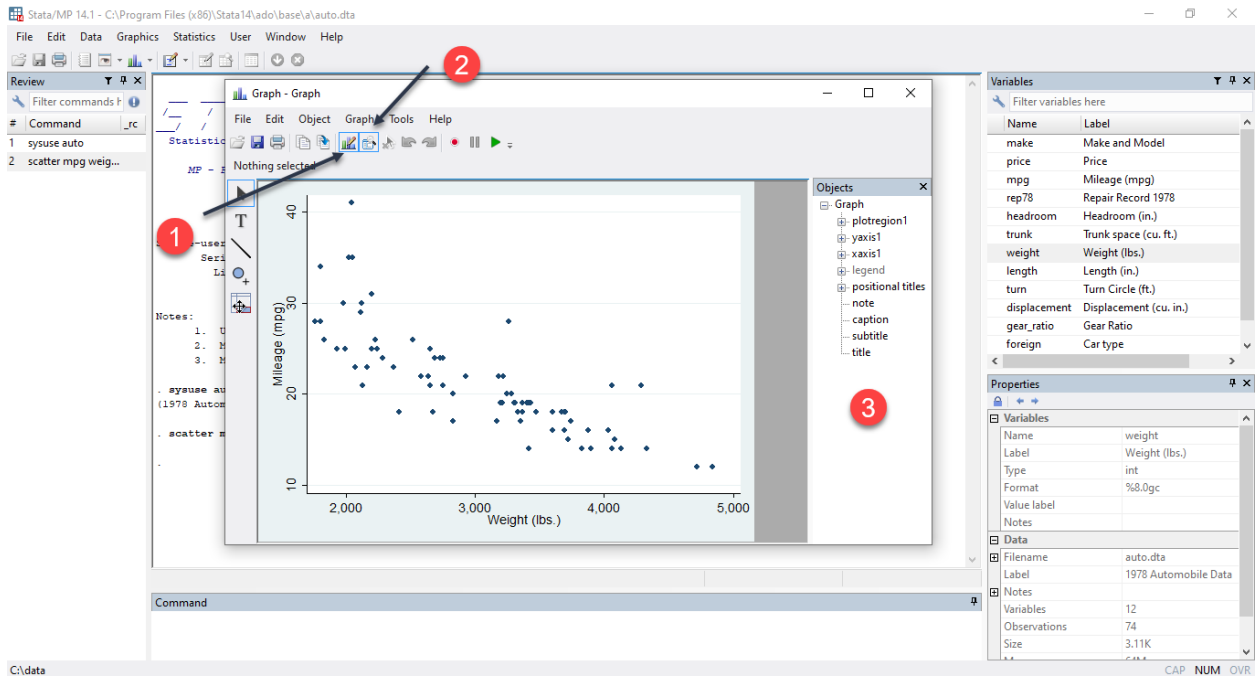
11 Grafik Çizme

Değişkenler arasında nasıl bir ilişki olduğunu görmek için tercih edilen yollardan biri grafik çizmektir. Stata'da bar, nokta, kutu, pasta ve dağılım grafikleri çizilebilmektedir. Stata çizilen grafikler üzerinde değişiklik yapmak, not ve başlık eklemek açısından oldukça esneklerdir. Örneğin çizilen grafiğin x ve y ekseninin isimleri, bu eksenlerdeki aralıklar ve grafik alanının rengi değiştirilebilir. Grafiğin sağ, sol, üst ve alt taraflarına açıklama ve grafiğe not ve başlıklar eklenebilir. Burada yalnızca derste kullanacağımız dağılım grafiği anlatılacaktır. **auto** veri setini kullanarak y ekseninde mpg, x ekseninde weight değişkeni olacak şekilde iki boyutlu dağılım grafiği aşağıdaki üç farklı kod kullanılarak çizilebilir.

- . scatter mpg weight
- . twoway scatter mpg weight
- . graph twoway scatter mpg weight



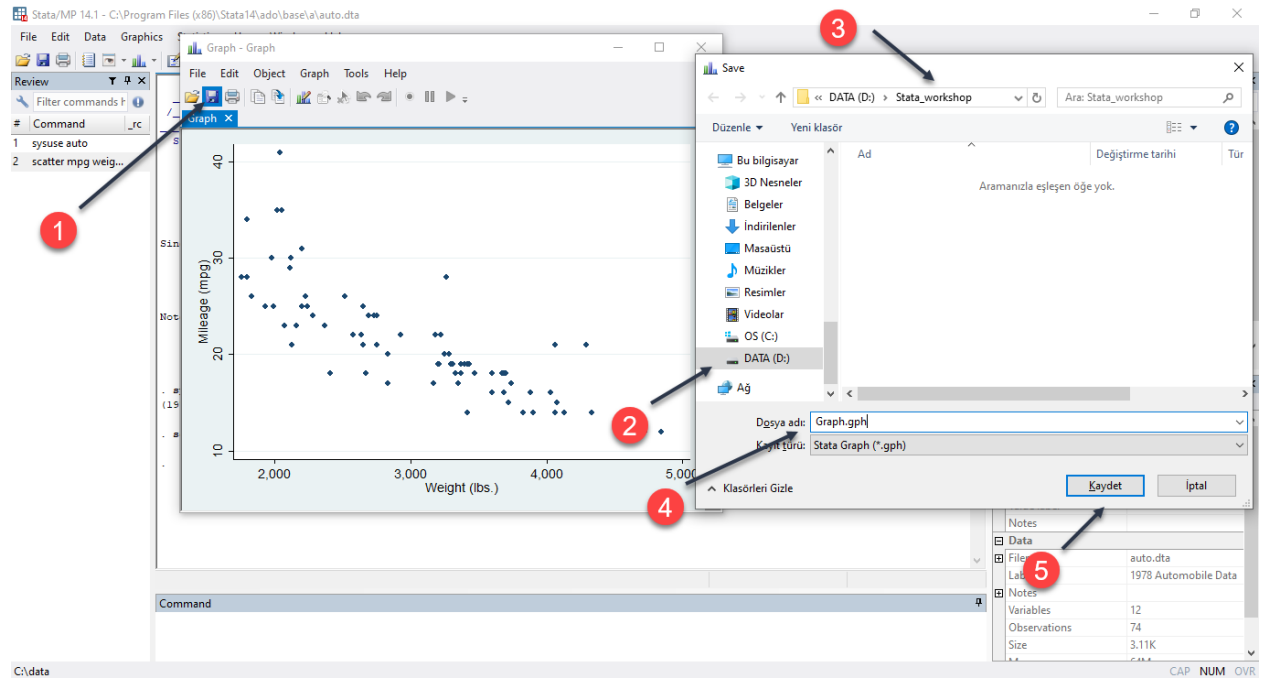
Yukarıdaki grafikte görüleceği üzere aracın ağırlığı ve aldığı yol mesafesi arasında negatif yönlü ve doğrusal olmayan bir ilişki vardır.



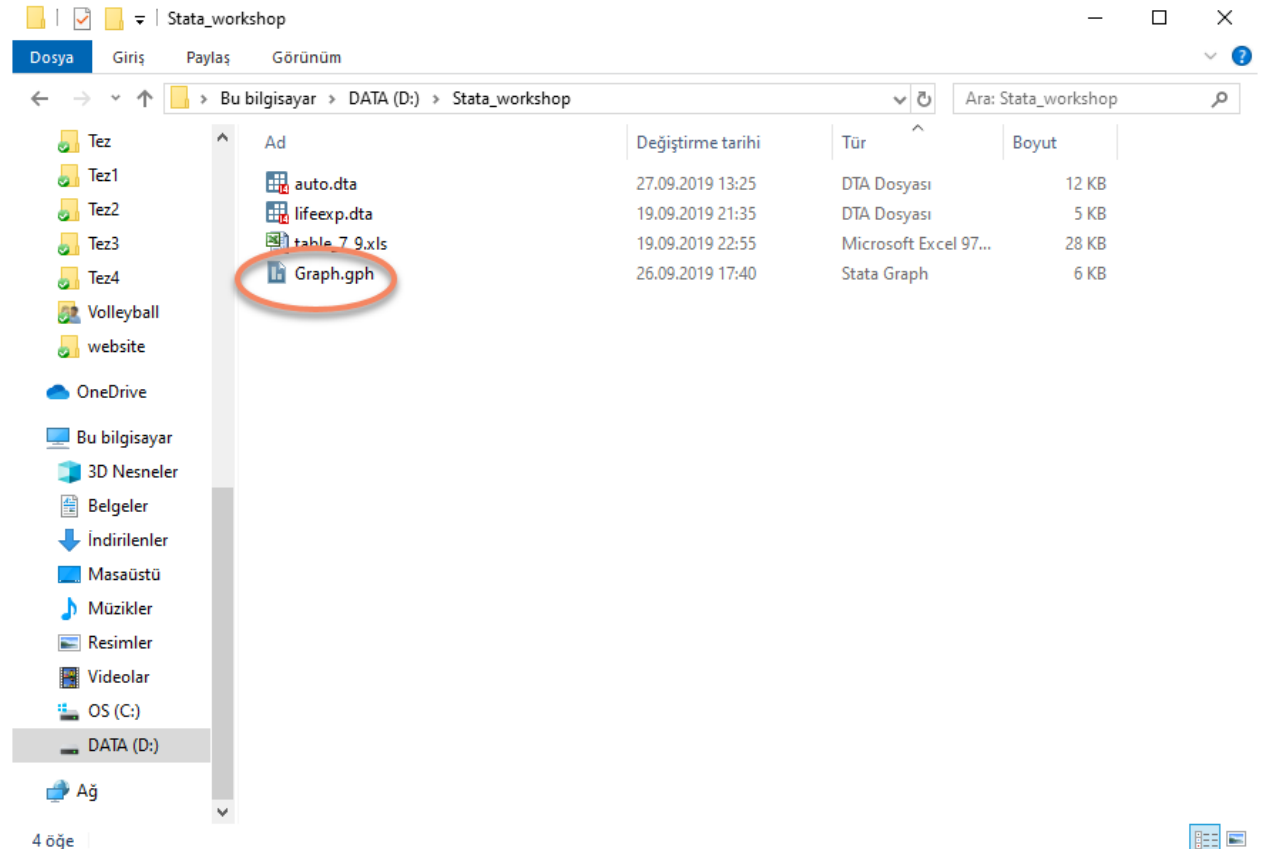
Sırası ile 1 ve 2 nolu sekmelere tıkladıktan sonra 3 nolu alandaki opsiyonları kullanarak grafik üzerinde gerekli değişiklikleri yapabilirsiniz.

Tüm değişikliklerin ardından grafiği aşağıdaki adımları izleyerek ya da kodu kullanarak kaydedebilirsiniz.

. graph save Graph "D:\Stata_workshop\Graph.gph"



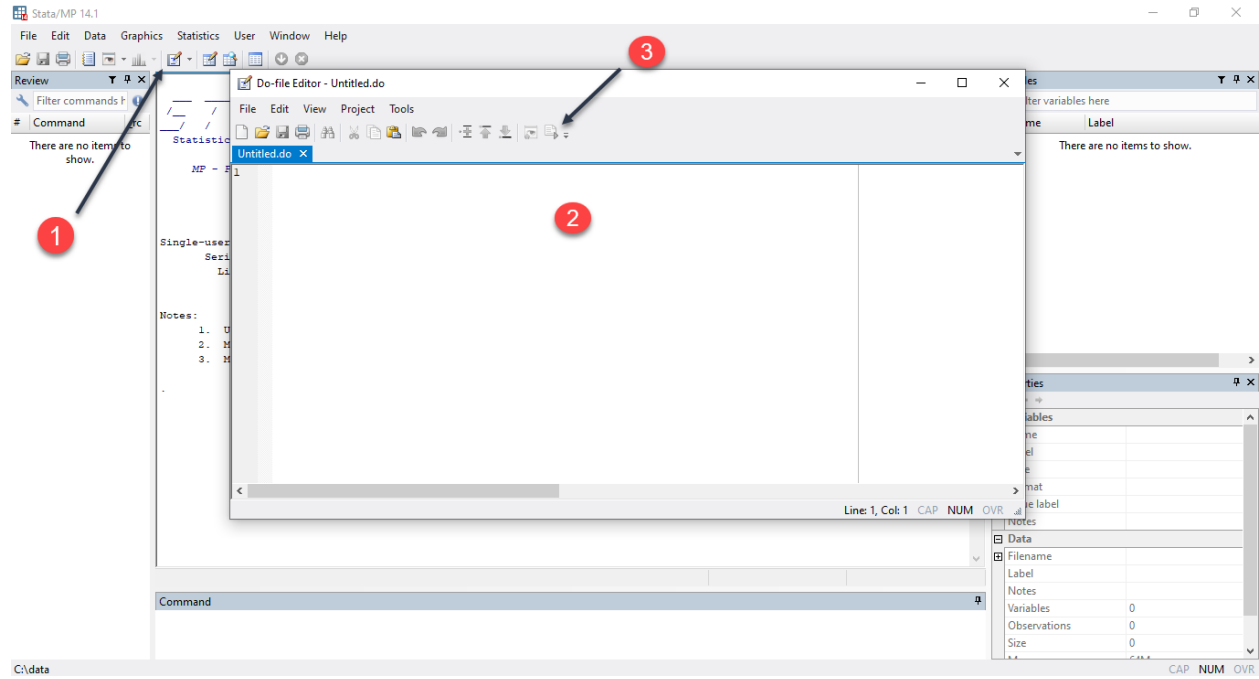
Aşağıda görüleceği üzere çalışma klasörü içinde **Graph** adında bir grafik kaydedilmiştir.



12 Do ve Log Dosyaları

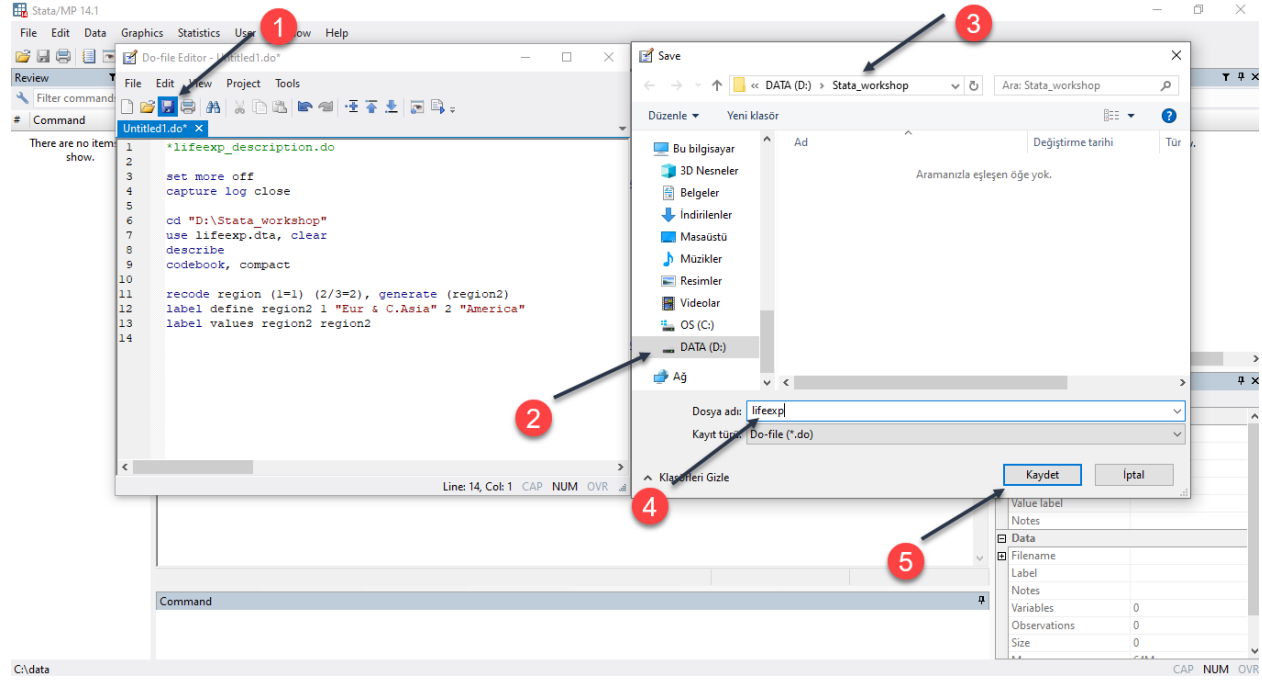
Do Dosyası

do dosyası yazılan kodların komut penceresi kullanılmadan çalıştırıldığı ve bu kodların saklandığı bir dosyadır. Stata'da analiz yaparken **do** dosyasını kullanmanın en büyük avantajı yazdığımız kodlarda bir hata varsa onu belirlemenizi ve kaldığımız yerden yazdığımız kodlara devam etmenizi sağlamasıdır. Komut penceresi kullanıldığında yazdığımız tüm kodlar geçmiş işlemler penceresinde kayıtlı olsa bile Stata'yı kapattığımızda tüm kodlar kaybolmaktadır. Bu yüzden Stata'da analiz yaparken **do** dosyası ile çalışmanız çok daha iyi olacaktır.

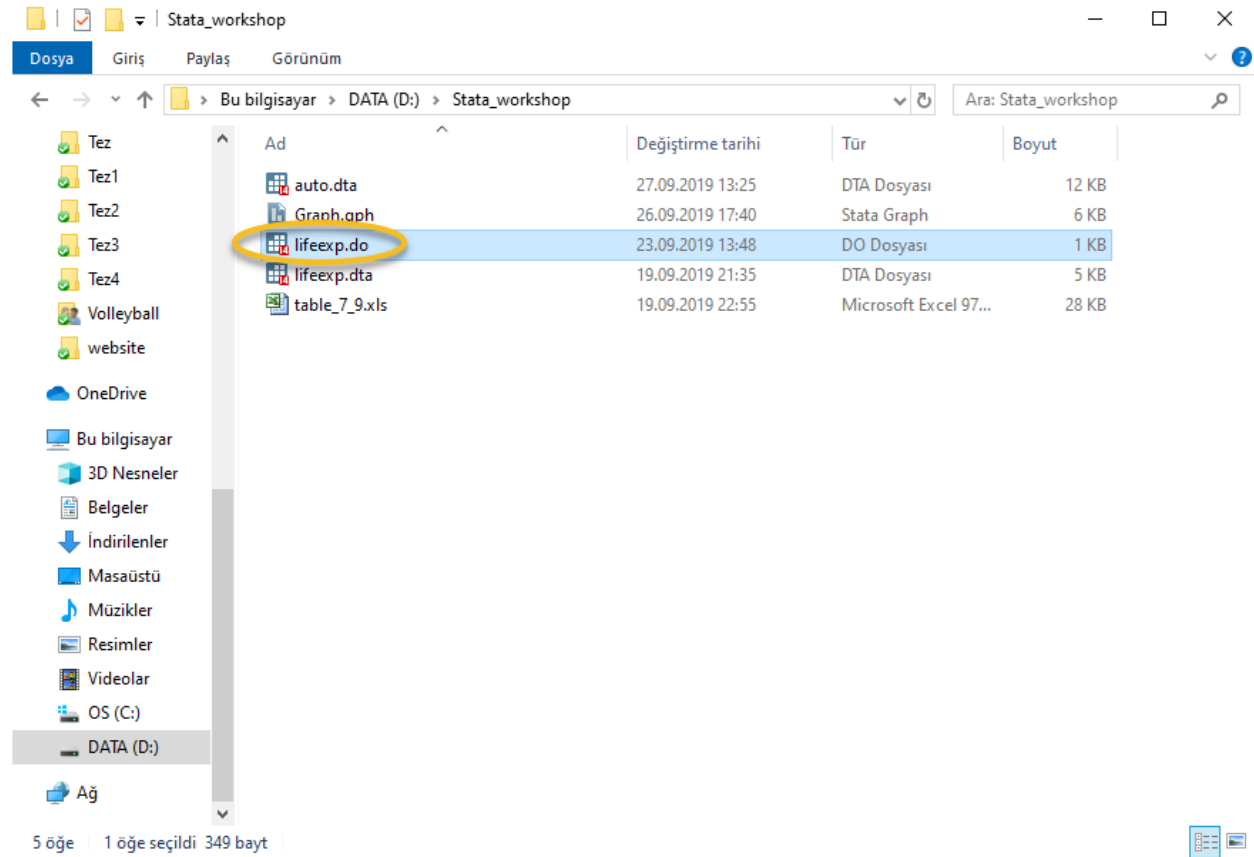


Stata menüsünden **do** dosyası **1** nolu sekmeye tıklanarak açılır. **2** nolu alan kodlarımızın yazıldığı alandır. **3** nolu sekme ise kodlarımızı çalıştırmak için tıklamanız gereken sekmedir. **do** dosyası **CTRL + 9** kısa yolu kullanılarak da açılabilir. **do** dosyasında kodlarımızı çalıştırmamızın bir başka yolu ise **CTRL + D** kısa yolunu kullanmaktır.

do dosyasında yapılmasını istediğiniz işlemlerin kodlarını yazdıktan sonra kaydetmek için aşağıdaki adımları takip edebilirsiniz.



Aşağıda görüleceği üzere çalışma klasörü içinde, **lifeexp** adlı **do** dosyası kaydedilmiştir.

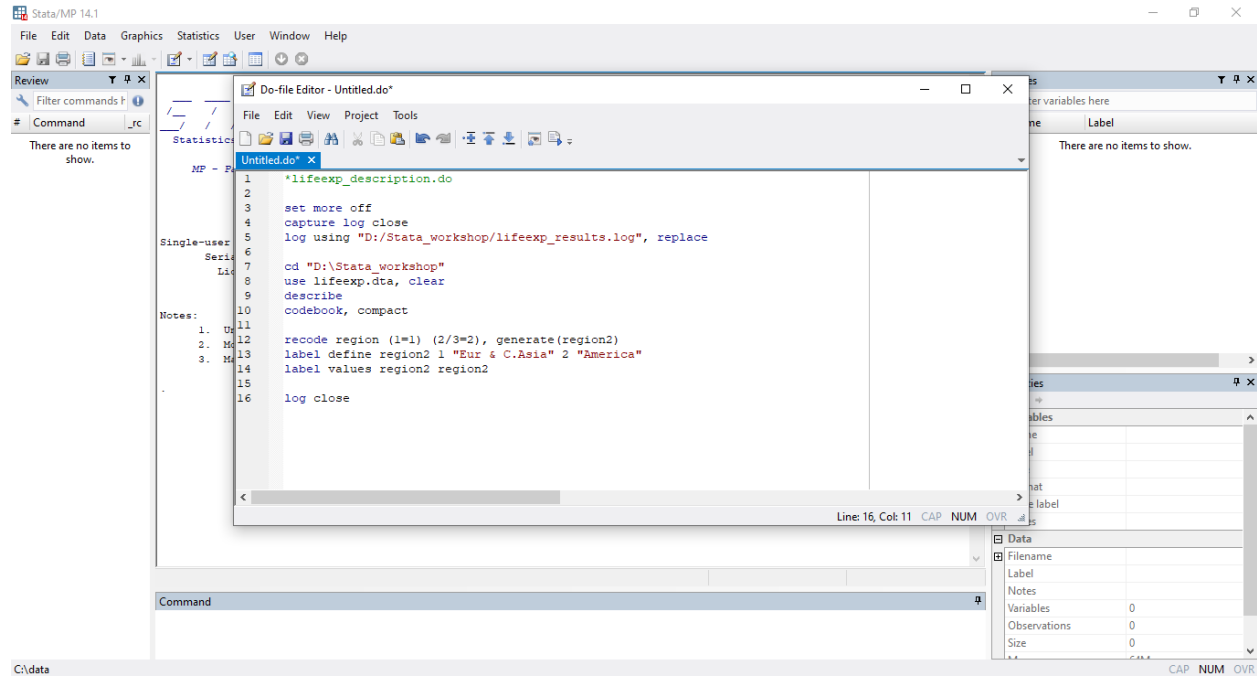


Log Dosyası

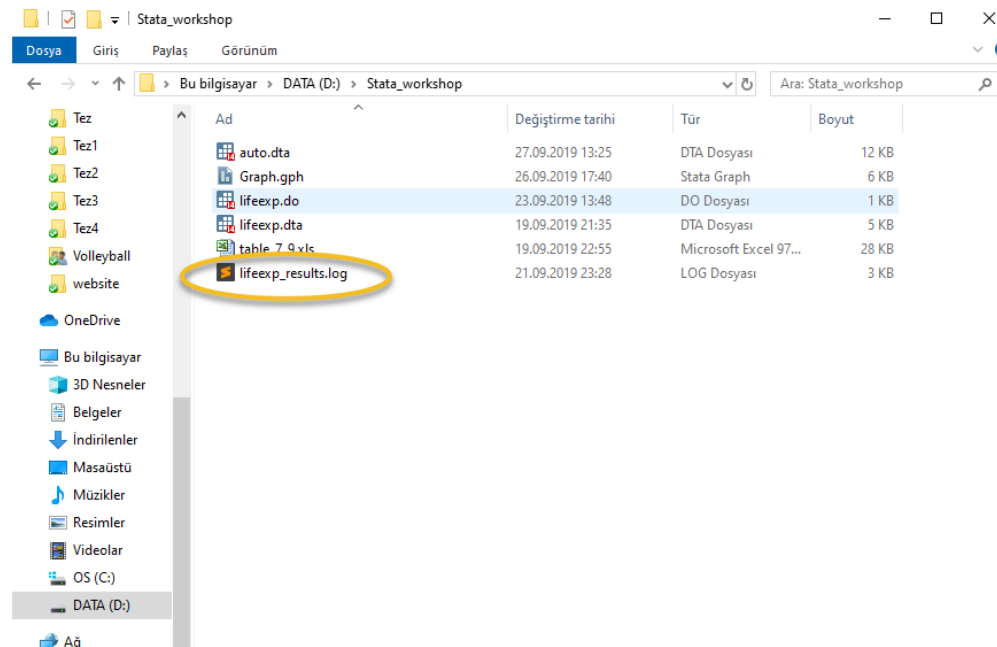
Log dosyası Stata'da yapılan tüm işlemlerin sonucunun kaydedildiği bir çıktı dosyasıdır. Log dosyasını çalışma klasörüne kaydedebilir ve yaptığımız tüm işlemlerin sonucuna kolayca ulaşabilirsiniz. Log dosyasında yapılan işlemlerin kaydına başlanması ve sonlandırılması için **do** dosyasında aşağıdaki kodu yazabilirsiniz.

```
. log using "D:/Stata_workshop/lifeexp_results.log", replace
. log close
```

Şimdiye kadar yazdığımız kodlarla do ve log dosyalarının kullanıldığı bir örnek oluşturunuz.



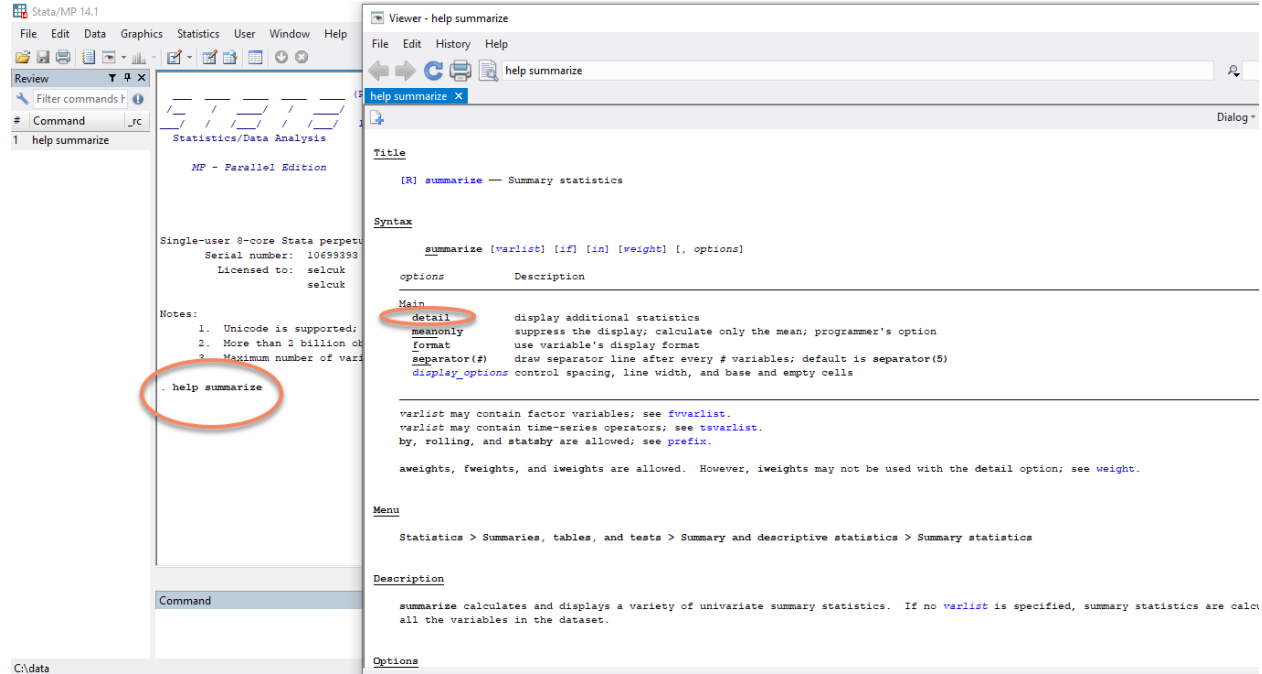
Aşağıda görüleceği üzere çalışma klasörü içinde, yapılan işlemlerin kaydedildiği bir **log** dosyası yer almaktadır.



13 Yardım

Stata komutlarının hangi opsiyonlarla çalıştığı ve ne amaçla kullanıldığını akılda tutmak sizlere başlangıçta zor gelebilir. Stata bu anlamda komutların nasıl ve ne amaçla kullanıldığını öğrenmeniz ya da hatırlamanız için sizlere bir yardım imkanı sunmaktadır. Bunun yanı sıra unutulmamalıdır ki Google her zaman yardım alacağınız iyi bir seçenektir! Stata'da yardım almak için **help** komutu kullanılmaktadır. Herhangi bir stata komutu hakkında yardım almak için Stata menüsü ya da komut penceresi kullanılabilir.

```
. help summarize /* summarize komutunun nasıl çalıştığı ile ilgili bilgiler verir */
```



Görüleceği üzere **summarize** komutunun nasıl ve hangi opsiyonlarla kullanılacağını anlatıldığı bir pencere karşınıza çıkmıştır. Opsiyonlar, komutun ardındaki “,” den sonra yazılan sözcüklerdir. Örneğin **summarize** komutunun opsiyonlarından biri **detail**'dir. Bu opsiyon **summarize** komutunun daha detaylı bilgiler vermesini sağlar. Bu opsiyonla beraber kod aşağıdaki gibi yazılır.

```
. summarize price, detail
```

summarize komutu normalde değişkenlerin yüzdelik dilimleri, varyansı, basıklığı ve çarpıklığı hakkında bilgi vermemektedir. **detail** opsiyonu ile birlikte kullanıldığında ise bu ilave bilgileri de sağlamaktadır.

Not: **sum**, **summarize** komutunun kısaltılmış halidir.

Stata/MP 14.1 - C:\Program Files (x86)\Stata14\ado\base\auto.dta

File Edit Data Graphics Statistics User Window Help

Review 979-656-4601 (fax)

Filter commands | Command | _rc

```

1 sysuse auto
2 sum price, detail

```

Single-user 8-core Stata perpetual license:
Serial number: 10699393
Licensed to: selcuk selcuk

Notes:

1. Unicode is supported; see [help unicode advice](#).
2. More than 2 billion observations are allowed; see [help obs advice](#).
3. Maximum number of variables is set to 5000; see [help set maxvar](#).

```

. sysuse auto
(1978 Automobile Data)

. sum price, detail

```

Price				
Percentiles	Smallest		Obs	Sum of Wgt.
1%	3291	3291		
5%	3740	3299	74	74
10%	3956	3667		
25%	4195	3740		
50%	5006.5		Mean	6165.257
			Std. Dev.	2949.496
75%	6342	13466	Variance	8699526
90%	11385	13594	Skewness	1.663434
95%	13466	14500	Kurtosis	4.819188
99%	18906	18906		

Command

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. in.)
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. in.)
gear_ratio	Gear Ratio
foreign	Car type

Data

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

C:\data

CAP NUM OVR

14 Kaynaklar

- Baum, C. F., & Christopher, F. (2006). An introduction to modern econometrics using Stata. Stata press.
- Cameron, A. C., & Trivedi, P. K. (2009). Microeconometrics Using Stata STATA Press. College Station, Texas.
- StataCorp 2016, Stata: data analysis and statistical software, StataCorp, Erişim: 27 Eylül 2019, <http://www.stata.com/>
- Stata, "GETTING STARTED WITH STATA FOR WINDOWS RELEASE 16, Erişim: 27 Eylül 2019, <https://www.stata.com/bookstore/getting-started-windows/>