بسم الله الرحمن الرحیم

آموزش بررسی فروض کلاسیک داده‌های Panel در نرم افزار R-studio

حسین خاندانی

مدرس نرم افزارهای اقتصادسنجی

بسم الله الرحمن الرحیم

سخن مدیر

در فایل‌های گذشته آموزشی نحوه ورود داده و تخمین رگرسیون در داده‌های مقطعی و پانل دیتا آموزش داده شده است. همچنین در فایل آموزش پانل دیتا بخش ۱ آزمون‌های F لیمر و هاسمن برای مدل‌های پانل یکرفه و دوطرفه معرفی گردید.
از این رو در این فایل نحوه انجام آزمون فروض کلاسیک به تفکیک مدلهای اثرات ثابت و تصادفی یک و دو طرفه آموزش داده می‌شود./

http://econometrics.blog.ir

برای راحتی کار با نرم افزار R به دوستان توصیه میشود از نرم افزار کمکی R بنام Rstudio استفاده نمایند.

نکته ۱: دیگر این مسأله است که تمام دستورات در نرم افزار R با رنگ آبی می باشد و نتایج با رنگ مشکی نمایش داده می شوند.

نکته ۲: نرم افزار R به حروف بزرگ و کوچک حساس است.

Console ~/

```
> PANEL DATA in R    -- BY: Hossein Khandani
```

```
arch.test(re)
arch.test(fe)
sigsq <- VAR(data1[,1], p = 2, type = "const")
var.2c <- VAR(arima, p = 2, type = "const")
arima<-arima(gdp, order=c(2,1,1))
arima
var.2c <- VAR(arima, p = 2, type = "const")
library(vars)
var.2c <- VAR(arima, p = 2, type = "const")
sidsq <- VAR(data1, p = 1, type = "const")
sigsq <- VAR(data1, p = 1, type = "const")
arch(test)
archtest(sigsq)
sigsq <- VAR(data1, p = 2, type = "const")
archtest(sigsq)
arch.test(sigsq)
summary(re)
AMozesh 2:
```

Files | Plots | Packages | Help | Viewer

The R Language | Find in Topic

Statistical Data Analysis

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function

Project: (None)

x | lm-tests.R | khandani | panel | Untitled1* | Source Viewer

110 observations of 8 variables

	row.names	UE	TB	REG	GEI	CCI	INF	PGDP
1	1-1380	14.3	23.540640	1.4442422	1.6971966	1.7533991	3.00317105	20071.966
2	1-1381	14.3	24.314924	1.5840256	1.8148083	1.9449878	2.77073524	23455.842
3	1-1382	14.1	24.311413	1.7343420	2.0388211	2.0969678	2.34361233	30464.004
4	1-1383	13.9	24.832071	1.5979233	1.7656513	1.9742009	2.66873278	34011.739

Environment History

To Console To Source

```
summary(re)
fe<-plm(PGDP~INF+UR, data=panel, model="within")
offset <- as.vector(model.offset(mf))
library(vars)
library(vars)
arch.test(re)
arch.test(fe)
sigsq <- VAR(data1[,1], p = 2, type = "const")
var.2c <- VAR(arima, p = 2, type = "const")
arima<-arima(gdp, order=c(2,1,1))
arima
var.2c <- VAR(arima, p = 2, type = "const")
library(vars)
var.2c <- VAR(arima, p = 2, type = "const")
sidsq <- VAR(data1, p = 1, type = "const")
sigsq <- VAR(data1, p = 1, type = "const")
arch(test)
archtest(sigsq)
sigsq <- VAR(data1, p = 2, type = "const")
archtest(sigsq)
arch.test(sigsq)
summary(re)
AMozesh 2:
PANEL DATA in R -- BY: Hossein Khandani
```

چند اسلاید اول یادآوری آموزش گذشته در رابطه با ورود دیتا و معرفی ساختار پانل به نرم افزار است و پس از آن به آموزش مربوطه پرداخته میشود.

Console ~/

قبل از هر چیز ابتدا باید پکیج پانل دیتا "PLM" را در نرم افزار R فراخوانی نمایید. در صورتی که پکیج مربوطه یافت نگردید ابتدا آنرا از منوی tools نصب و سپس دستور ذیل را اجرا نمایید.

> دستور زیر را در Console تایپ نمایید و اینتر را بزنید تا پکیج مربوطه فراخوانی شود.

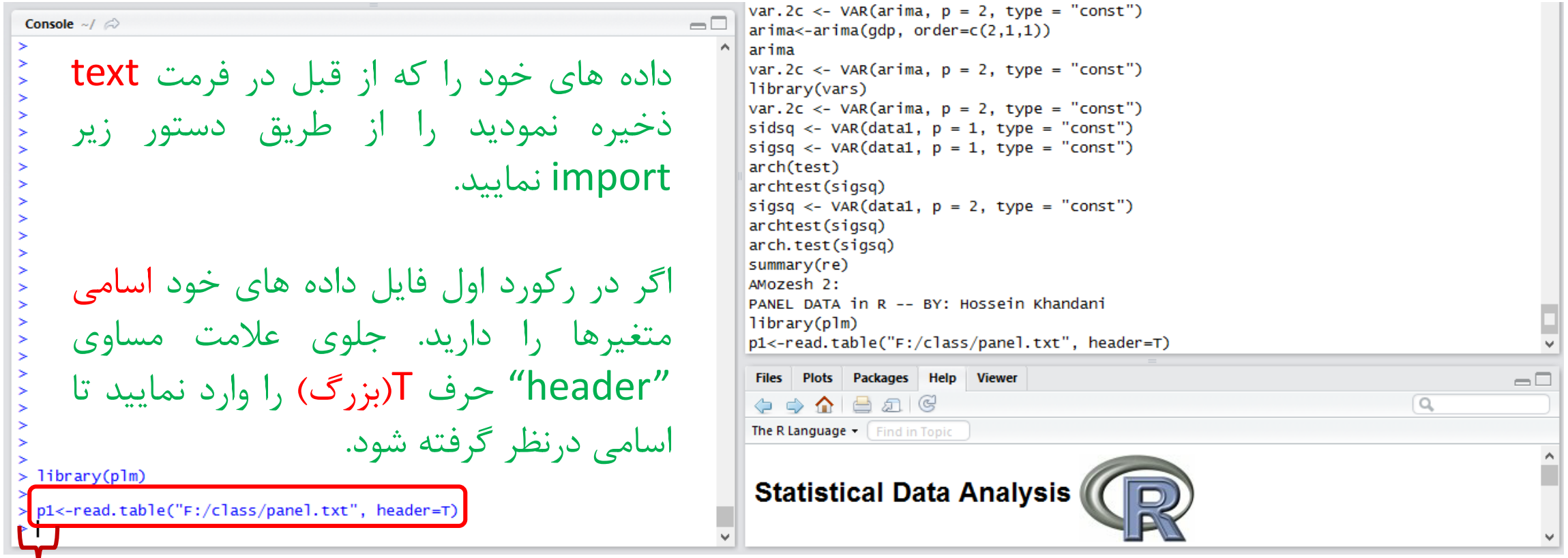
```
> library(plm)
```

```
> library(plm)
```

Help Viewer

The R Language Find in Topic

Statistical Data Analysis



توجه: برای نحوه صحیح وارد نمودن داده ها و نکات مرتبط با آن از جمله فرمت به وبلاگ مراجعه نمایید.

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function

Project: (None)

x | lm-tests.R | khandani | panel | Untitled1* | Source Viewer

110 observations of 8 variables

	row.names	UE	TB	REG	GEI	CCI	INF	PGDP
1	1-1380	14.3	23.540640	1.4442422	1.6971966	1.7533991	3.00317105	20071.966
2	1-1381	14.3	24.314924	1.5840256	1.8148083	1.9449878	2.77073524	23455.842
3	1-1382	14.1	24.311413	1.7343420	2.0388211	2.0969678	2.34361233	30464.004
4	1-1383	13.9	24.832071	1.5979233	1.7656513	1.9742009	2.66873278	34011.739

Console ~/

```
>
> برای مشاهده داده هایتان اسم دلخواهی را که انتخاب نموده اید بار
>
> دیگر تایپ نمایید و اینتر را بزنید. فایل داده هایتان حتما باید شامل
>
>
> دو ستون بازه زمانی و مقاطع باشد.
>
> library(plm)
>
> p1<-read.table("F:/class/panel.txt", header=T)
> p1
    cross year   UE        TB       REG       GEI       CCI          INF
PGDP
1       1 1380 14.3 23.540640 1.4442422 1.6971966 1.7533991   3.00317105 200
71.966
2       1 1381 14.3 24.314924 1.5840256 1.8148083 1.9449878   2.77073524 234
55.842
3       1 1382 14.1 24.311413 1.7343420 2.0388211 2.0969678   2.34361233 304
64.004
4       1 1383 13.9 24.832071 1.5979233 1.7656513 1.9742009   2.66873278 340
11.739
5       1 1384 13.7 24.563263 1.6183741 1.7559556 1.9914446   3.53848734 361
13.002
6       1 1385 13.7 24.088805 1.6750186 1.8240310 2.0415982   2.33236152 409
96.318
7       1 1386 13.7 24.229909 1.7577902 1.7823105 2.0717480   4.35264324 496
72.748
8       1 1387 13.7 22.087087 1.8169362 1.6981726 2.0783314   1.82011224 427
21.885
9       1 1388 13.2 20.609122 1.6917941 1.7699274 2.0442386   2.84522568 518
```

Environment | History

To Console | To Source

```
offset <- as.vector(model.offset(mf))
library(vars)
library(vars)
arch.test(re)
arch.test(fe)
sigsq <- VAR(data1[,1], p = 2, type = "const")
var.2c <- VAR(arima, p = 2, type = "const")
arima<-arima(gdp, order=c(2,1,1))
arima
var.2c <- VAR(arima, p = 2, type = "const")
library(vars)
var.2c <- VAR(arima, p = 2, type = "const")
sidsq <- VAR(data1, p = 1, type = "const")
sigsq <- VAR(data1, p = 1, type = "const")
arch(test)
archtest(sigsq)
sigsq <- VAR(data1, p = 2, type = "const")
archtest(sigsq)
arch.test(sigsq)
summary(re)
AMozesh 2:
PANEL DATA in R -- BY: Hossein Khandani
library(plm)
p1<-read.table("F:/class/panel.txt", header=T)
p1
```

Files | Plots | Packages | Help | Viewer

The R Language | Find in Topic

Statistical Data Analysis

بعد از وارد نمودن داده ها، حال نوبت به معرفی ساختار داده ها (سری زمانی، مقطعی یا پانل) به نرم افزار است. برای این منظور با استفاده از دستور اسلاید بعد دو ستون زمان و مقطع را به نرم افزار معرفی می نماییم. که برای اطمینان از درستی انجام آن؛ اسم دلخواهی که در این مرحله به داده هایتان داده اید، را بار دیگر تایپ نمایید و اینتر را بزنید نرم افزار باید دو ستون مقاطع و زمان را با هم ترکیب نموده باشد.

RStudio

File Edit Code View Plots Session Build Debug Tools Help

Go to file/function | Project: (None)

x | lm-tests.R | khandani | panel | Untitled1* | Source Viewer

110 observations of 8 variables

	row.names	UE	TB	REG	GEI	CCI	INF	PGDP
1	1-1380	14.3	23.540640	1.4442422	1.6971966	1.7533991	3.00317105	20071.966
2	1-1381	14.3	24.314924	1.5840256	1.8148083	1.9449878	2.77073524	23455.842
3	1-1382	14.1	24.311413	1.7343420	2.0388211	2.0969678	2.34361233	30464.004
4	1-1383	13.9	24.832071	1.5979233	1.7656513	1.9742009	2.66873278	34011.739

Console ~/

```
107 10.0
108 11.2
109 10.9
110 10.9
> panel<-pdata.frame(p1, index=c("cross", "year"), drop.index=TRUE)
> panel
          UE       TB       REG       GEI       CCI        INF      PGDP
1-1380  14.3 23.540640 1.4442422 1.6971966 1.7533991  3.00317105 20071.966
6.4
1-1381  14.3 24.314924 1.5840256 1.8148083 1.9449878  2.77073524 23455.842
5.9
1-1382  14.1 24.311413 1.7343420 2.0388211 2.0969678  2.34361233 30464.004
5.4
1-1383  13.9 24.832071 1.5979233 1.7656513 1.9742009  2.66873278 34011.739
5.0
1-1384  13.7 24.563263 1.6183741 1.7559556 1.9914446  3.53848734 36113.002
4.8
1-1385  13.7 24.088805 1.6750186 1.8240310 2.0415982  2.33236152 40996.318
4.4
1-1386  13.7 24.229909 1.7577902 1.7823105 2.0717480  4.35264324 49672.748
4.2
1-1387  13.7 22.087087 1.8169362 1.6981726 2.0783314  1.82011224 42721.885
5.6
1-1388  13.2 20.609122 1.6917941 1.7699274 2.0442386  2.84522568 51824.798
5.2
1-1389  13.5 20.478563 1.8523644 1.6979556 2.0800290  3.38933951 62080.982
5.1
1-1390  13.4 21.389557 1.7695536 1.6121427 1.9952996  1.76278016 67435.954
5.2
```

* دستور خواندن ساختار داده های پانل دیتا

ترکیب مقطع و زمان

Environment | History

To Console | To Source

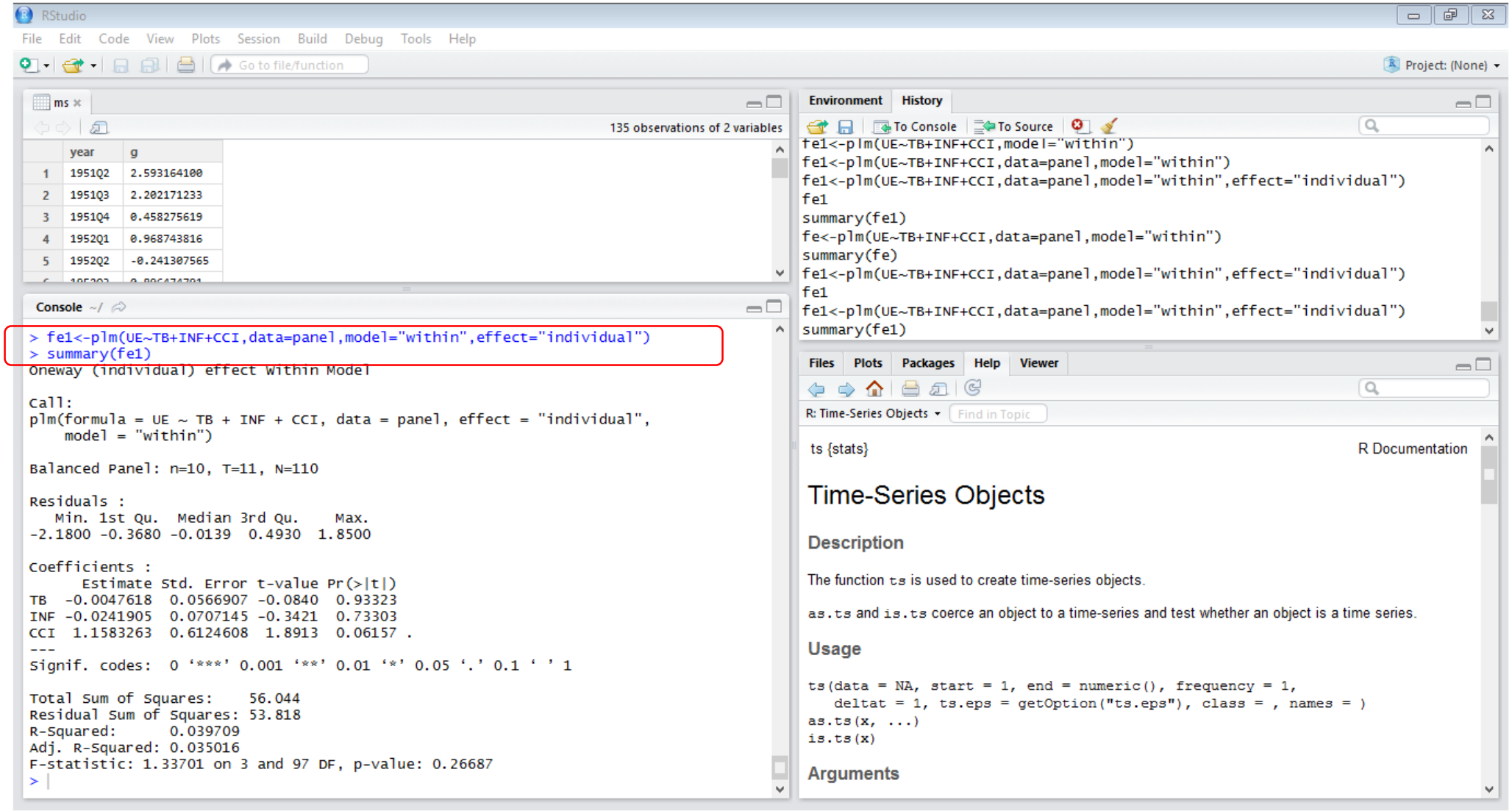
```
library(vars)
arch.test(re)
arch.test(fe)
sigsq <- VAR(data1[,1], p = 2, type = "const")
var.2c <- VAR(arima, p = 2, type = "const")
arima<-arima(gdp, order=c(2,1,1))
arima
var.2c <- VAR(arima, p = 2, type = "const")
library(vars)
var.2c <- VAR(arima, p = 2, type = "const")
sidsq <- VAR(data1, p = 1, type = "const")
sigsq <- VAR(data1, p = 1, type = "const")
arch(test)
archtest(sigsq)
sigsq <- VAR(data1, p = 2, type = "const")
archtest(sigsq)
arch.test(sigsq)
summary(re)
AMozesh 2:
PANEL DATA in R -- BY: Hossein Khandani
library(plm)
p1<-read.table("F:/class/panel.txt", header=T)
p1
panel<-pdata.frame(p1, index=c("cross", "year"), drop.index=TRUE)
panel
```

Files | Plots | Packages | Help | Viewer

The R Language | Find in Topic

Statistical Data Analysis

برای آشنایی با چگونگی تخمین مدل‌های Pool و پانل (اثرات ثابت و تصادفی یک و دو طرفه) به فایل آموزشی قبل مراجعه نمایید.



در اسلایدهای بعد نحوه آزمون خودهمبستگی مرتبه اول، سریالی و ناهمسانی واریانس در نرم افزار R-Studio آموزش داده می‌شود.

در این آموزش آزمون‌های اختصاصی هر متد(ثابت و تصادفی) به تفکیک اثرات یکطرفه و دوطرفه آموزش داده شده است؛ تا کارشناسان و دانشجویان توانایی این را داشته باشند که به صورت تخصصی و دقیق‌تر به تحلیل مدل‌های پانل دیتا بپردازند.

آزمون خودهمبستگی مرتبه اول دوربین–واتسون (D-W) مختص اثرات ثابت و تصادفی

- برای تست خودهمبستگی مرتبه اول AR(1) در مدل اثرات ثابت یکطرفه (fe1) و اثرات ثابت دوطرفه (fe2) از دستور زیر استفاده می‌شود.

دستور در برنامه: pdwtest(model`s name, data=data name)

```
> pdwtest(fe1,data=panel)

 Durbin-Watson test for serial correlation in panel models

data:  UE ~ TB + INF + CCI
DW = 0.6103, p-value = 1.029e-13
alternative hypothesis: serial correlation in idiosyncratic errors

> pdwtest(fe2,data=panel)

 Durbin-Watson test for serial correlation in panel models

data:  UE ~ TB + INF
DW = 0.5793, p-value = 3.902e-14
alternative hypothesis: serial correlation in idiosyncratic errors
```

فرمان

تفسیر نتایج آزمون:

با توجه به اینکه احتمال آماره آزمون زیر ۵ صدم است درنتیجه فرضیه عدم خودهمبستگی مرتبه اول رد می‌شود و دو مدل دارای خودهمبستگی مرتبه اول می‌باشند.

نکته:

این آزمون در مدل اثرات تصادفی نیز قابل استفاده است.

تذکر:

fe1 نام مدل تخمینی اثرات ثابت یکطرفه است.

fe2 نام مدل تخمینی اثرات ثابت دوطرفه است.

آزمون خودهمبستگی سریالی از مراتب بالا مختص اثرات ثابت یکطرفه و دو طرفه (آزمون ولدریج)

- برای تست خودهمبستگی سریالی پسماندها در مدل اثرات ثابت یکطرفه (fe1) و اثرات ثابت دوطرفه (fe2) برای مواردی که مقدار زمان (T) کوچک است می‌توان از آزمون اختصاصی این متد استفاده نمود. دستور آزمون خودهمبستگی سریالی ولدریج در ذیل آورده شده است.

pwartest(model`s name, data=data name)

نام مدل تخمینی نام دیتای مورد استفاده

```
> pwartest(fe1,data=panel)

 Wooldridge's test for serial correlation in FE panels

data:  fe1
chisq = 317.9617, p-value < 2.2e-16
alternative hypothesis: serial correlation

> pwartest(fe2,data=panel)

 Wooldridge's test for serial correlation in FE panels

data:  fe2
chisq = 81.217, p-value < 2.2e-16
alternative hypothesis: serial correlation
```

تفسیر نتایج آزمون:

با توجه به اینکه احتمال آماره آزمون زیر ۵ صدم است، درنتیجه فرضیه عدم خودهمبستگی سریالی رد می‌شود و دو مدل مذکوردارای مشکل خودهمبستگی می‌باشند.

نکته:

-این آزمون نسبت به مشکل ناهمسانی واریانس مقاوم (Robust) است.

-این آزمون مختص مدل اثرات ثابت (FE) است.

آزمون خودهمبستگی مرتبه اول مختص اثرات تصادفی یکطرفه و دو طرفه (آزمون PBSYTest)

- برای تست خودهمبستگی مرتبه اول در مدل اثرات تصادفی یکطرفه (re1) و اثرات ثابت دوطرفه (re2) می‌توان از آزمون اختصاصی این متد استفاده نمود. دستور این آزمون در ذیل آورده شده است.

pbsytest(model`s name, data=data name)

first-order serial correlation (ar), random effects (re)

```
> pbsytest(re1,data=panel)

 Bera, Sosa-Escudero and Yoon locally robust test

data:  formula
chisq = 124.5075, df = 1, p-value < 2.2e-16
alternative hypothesis: AR(1) errors sub random effects

> pbsytest(re2,data=panel)

 Bera, Sosa-Escudero and Yoon locally robust test

data:  formula
chisq = 96.5526, df = 1, p-value < 2.2e-16
alternative hypothesis: AR(1) errors sub random effects
```

تفسیر نتایج آزمون:

با توجه به اینکه احتمال آماره آزمون زیر ۵ صدم است درنتیجه فرضیه عدم خودهمبستگی مرتبه اول رد می‌شود و دو مدل دارای خودهمبستگی مرتبه اول می‌باشند.

نکته:

این آزمون نسب به مشکل ناهمسانی مقاوم می‌باشد.

تذکر۱:

re1 نام مدل تخمینی اثرات تصادفی یکطرفه است.
re2 نام مدل تخمینی اثرات تصادفی دوطرفه است.
این آزمون مختص اثرات تصادفی است.

آزمون خودهمبستگی مرتبه اول مختص اثرات تصادفی به تفکیک یکطرفه و دو طرفه (آزمون بالتاجی و لی)

Baltagi and Li (1995)'s Lagrange multiplier test for AR(1) or MA(1) idiosyncratic errors in panel models with random effects.

آزمون دیگری که برای بررسی خودهمبستگی در مدل اثرات تصادفی می‌تواند مورد استفاده قرار گیرد، آزمون بالتاجی و لی است که دو ورژن مختلف برای اثرات تصادفی یکطرفه و دوطرفه قابل آزمون است (**این آزمون مختص مدل اثرات تصادفی (RE) است**).

```
> pbltest(UE~TB+INF+CCI,data=p1,alternative="onesided")

 Baltagi and Li one-sided LM test

data:  UE ~ TB + INF + CCI
z = 9.083, p-value < 2.2e-16
alternative hypothesis: AR(1)/MA(1) errors in RE panel models
```

تفسیر نتایج آزمون (مدل یکطرفه):

با توجه به اینکه احتمال آماره آزمون زیر ۵ صدم است، درنتیجه فرضیه عدم خودهمبستگی رد می‌شود و مدل مذکور دارای مشکل خودهمبستگی می‌باشد.

onesided LM test FOR "RE"

```
> pbltest(UE~TB+INF+CCI,data=p1,alternative="twosided")

 Baltagi and Li two-sided LM test

data:  UE ~ TB + INF + CCI
chisq = 82.5002, df = 1, p-value < 2.2e-16
alternative hypothesis: AR(1)/MA(1) errors in RE panel models
```

تفسیر نتایج آزمون (مدل دوطرفه):

با توجه به اینکه احتمال آماره آزمون زیر ۵ صدم است، درنتیجه فرضیه عدم خودهمبستگی رد می‌شود و مدل مذکور دارای مشکل خودهمبستگی می‌باشد.

twosided LM test FOR "RE"

آزمون عمومی خودهمبستگی سریالی بروش–گادفری/ولدریج (اثرات ثابت، تصادفی و pool)

Breusch-Godfrey/Wooldridge test for serial correlation in panel models (General tests)

```
>pbgtest(fe1, order=2)

 Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  UE ~ TB + INF + CCI
chisq = 53.9281, df = 2, p-value = 1.948e-12
alternative hypothesis: serial correlation in idiosyncratic errors

>pbgtest(re1, order=2)

 Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  UE ~ TB + INF + CCI
chisq = 54.4745, df = 2, p-value = 1.483e-12
alternative hypothesis: serial correlation in idiosyncratic errors

>pbgtest(fe2, order=2)

 Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  UE ~ TB + INF
chisq = 59.7579, df = 2, p-value = 1.056e-13
alternative hypothesis: serial correlation in idiosyncratic errors

>pbgtest(re2, order=2)

 Breusch-Godfrey/Wooldridge test for serial correlation in panel models

data:  UE ~ TB + INF
chisq = 58.6974, df = 2, p-value = 1.795e-13
alternative hypothesis: serial correlation in idiosyncratic errors
```

همانطور که مشاهده می‌شود این آزمون برای تمام حالات اثرات ثابت و تصادفی مورد استفاده قرار گرفته است؛ دقت نمایید که تنها آزمون بروش-گادفری/ولدریج است که می‌تواند برای مدل اثرات ثابت و تصادفی و حتی آمیخته مورد استفاده قرار گیرد.

تفسیر نتایج آزمون:

با توجه به اینکه احتمال آماره آزمون در تمام مدل‌ها زیر ۵ صدم است، درنتیجه فرضیه عدم خودهمبستگی رد می‌شود و مدل‌های مذکور دارای مشکل خودهمبستگی سریالی می‌باشند.

نکته:

برای انجام آزمون خودهمبستگی سریالی از مراتب بالاتر از ۲ می‌توان **order** را به مراتب بالاتر تغییر داد و نتیجه آزمون را مشاهده کرد.

آزمون ناهمسانی واریانس بروش پاگان (داده‌های آمیخته)

```
> bptest(pool,data = Panel, studentize=F)

 Breusch-Pagan test

data:  pool
BP = 1.1695, df = 2, p-value = 0.5572
```

فرضیه صفر این آزمون همسانی واریانس(عدم ناهمسانی واریانس) است. با توجه به اینکه احتمال آماره آزمون بالای ۵ صدم است در نتیجه فرصیه صفر را نمی‌توان رد نمود، درنتیجه مدل دارای همسانی واریانس است.

آزمون همسانی واریانس بر روی دیتاهای پانل

```
> fligner.test(panel)

 Fligner-Killeen test of homogeneity of variances

data:  panel
Fligner-Killeen:med chi-squared = 638.6378, df = 7, p-value < 2.2e-16

> bartlett.test(panel)

 Bartlett test of homogeneity of variances

data:  panel
Bartlett's K-squared = 12169.75, df = 7, p-value < 2.2e-16
```

برای تست همسانی واریانس داده‌های پانل می‌توان از دو آزمون روبرو استفاده نمود. فرضیه صفر این آزمون همسانی واریانس می‌باشد. با توجه به اینکه آماره آزمون زیر ۵ صدم است درنتیجه داده‌های دارای ناهمگنی یا ناهمسانی واریانس هستند.

تذکر۱:
این آزمون بر روی دیتا انجام می‌شود و نه نتایج مدل.

برطرف نمودن نقض فروض کلاسیک

general feasible generalized least squares

درصورتی که در اثر ناهمسانی واریانس یا خود همبستگی فروض کلاسیک نقض گردید می‌توان از حالت عمومی FGLS (General) استفاده نمود که در این روش با استفاده از ماتریس واریانس و کواریانس مشکل نقض شده رفع می‌گردد. در فایل آموزشی بعدی به صورت تخصصی‌تر چگونگی استفاده از متدهای مربوطه برای رفع نقض فروض آموزش داده می‌شود (مانند رگرسیون مقاوم، wls، gls و ...).

```
> fe_FGLS<-pggls(UE~TB+INF,data=panel,model="within")
> summary(fe_FGLS)
 Within model

Call:
pggls(formula = UE ~ TB + INF, data = panel, model = "within")

Balanced Panel: n=10, T=11, N=110

Residuals
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-2.26500 -0.37220  0.03826  0.00000  0.44830  1.76000

Coefficients
      Estimate Std. Error  z-value Pr(>|z|)
TB   0.0026619  0.0042567   0.6253   0.5317
INF -0.0416055  0.0018634 -22.3282   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Total Sum of Squares: 1543.6
Residual Sum of Squares: 55.807
Multiple R-squared: 0.96385
```

دستور تخمین FGLS

دستور تخمین FGLS:

pggls()

پایان بخش سوم از مجموعه آموزش های نرم افزار R و R-Studio

مدیریت وب و تهیه کننده:
حسین خاندانی

Econometrics.blog.ir