

Эконометрика, 2018-2019, 2 модуль
Семинар 4
26.11.18 для
Группы Э_Б2016_Э_3
Семинарист О.А.Демидова

Проверка гипотез

1) Зависимость длительности обучения индивида от его способностей и длительности обучения родителей

Для выполнения приведенных ниже упражнений используются данные файла Dougherty.dta.

1.1 Проверка гипотезы об адекватности модели

Переменные:

S – длительность обучения индивида, SM - длительность обучения мамы индивида, SF – длительность обучения отца индивида, $ASVABC$ – обобщенный показатель способностей индивида, рассчитанный по результатам тестов.

- 1) Оцените параметры уравнения множественной регрессии

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + \beta_4 SF + \varepsilon$$

- 2) Сформулируйте гипотезу об адекватности регрессии. Является ли построенная регрессия адекватной?

- 3) Дайте экономическую интерпретацию полученным результатам.

1.2 Зависимость длительности обучения индивида от его способностей

- 1) Влияет ли на длительности обучения индивидуума длительность обучения его родителей (или только его способности)? Для ответа на этот вопрос оцените параметры уравнения множественной регрессии

$$S = \beta_1 + \beta_2 ASVABC + \beta_3 SM + \beta_4 SF + \varepsilon .$$

- 2) Проверьте гипотезу об одновременном равенстве коэффициентов β_3 и β_4 нулю:

$$H_0 : \beta_3 = \beta_4 = 0$$

при альтернативной гипотезе

$$H_1 : \beta_3^2 + \beta_4^2 \neq 0 .$$

1.3 В равной ли степени родители влияют на длительность обучения индивида?

Для ответа на этот вопрос проверьте гипотезу о равенстве коэффициентов β_3 и β_4 в предыдущей модели:

$$H_0 : \beta_3 = \beta_4$$

при альтернативной гипотезе

$$H_1 : \beta_3 \neq \beta_4$$

1.4 Правильно ли выбраны веса в показателе, характеризующем способности индивида?

Значения переменной, характеризующей способности индивида, рассчитывались следующим образом: $ASVABC = 0.5ASVAB02 + 0.25ASVAB03 + 0.25ASVAB04$, где

ASVAB02 – результаты теста по арифметике,

ASVAB03 – результаты теста по правописанию,

ASVAB04 – результаты теста по пониманию прочитанного материала.

Оценив параметры уравнения множественной регрессии

$$S = \beta_1 + \beta_2 ASVAB0_2 + \beta_3 ASVAB0_3 + \beta_4 ASVAB0_4 + \beta_5 SM + \beta_6 SF + \varepsilon,$$

проверьте, правильно ли были выбраны веса в переменной ASVABC?

Для этого проверьте гипотезу:

$$H_0 : \beta_2 = 2\beta_3 = 2\beta_4$$

при альтернативной гипотезе

H_1 : гипотеза H_0 не имеет места.

Методические рекомендации по выполнению упражнения 1

Воспользуйтесь соответствующими указаниями в разделе «Оценка регрессий в пакете STATA»

1) Наберите в командном окне

reg S ASVABC SM SF

Найдите в появившейся таблице F-статистику, сравните p-value для F-статистики с выбранным уровнем значимости. Если p-value меньше выбранного уровня значимости, например, 0.05, то регрессия является адекватной.

2) Для проверки гипотезы о равенстве коэффициентов при переменных SM и SF одновременно 0 в командном окне следует набрать:

test SM SF

и сделать вывод с помощью p-value для F-статистики. Если p-value меньше выбранного уровня значимости, то гипотеза H_0 отвергается.

3) Гипотеза о равенстве коэффициентов при переменных SM и SF может быть проверена аналогично предыдущей, только в последнем окне следует набрать команду

test (SM= SF)

4) Оцените регрессию, набрав в командном окне

reg S ASVAB02 ASVAB03 ASVAB04 SM SF

Гипотеза о правильности выбора весов может быть проверена с помощью команды

test (ASVAB02=2*ASVAB03) (ASVAB03= ASVAB04)

2) Проверка гипотезы о постоянной отдаче от масштаба для функции Кобба – Дугласа

В файле CobbDouglas.dta содержатся данные о 569 бельгийских фирмах.

Переменные:

Output – добавленная стоимость (в млн. евро),

Capital – общая стоимость основных фондов (в млн. евро),

Labor – количество рабочих,

Wage – затраты на заработную плату в расчете на одного рабочего (в тыс. евро).

Оцените регрессию $\ln Q = \beta_1 + \beta_2 \ln L + \beta_3 \ln K + \varepsilon$

по данным файла CobbDouglas.dta и

проверьте гипотезу о постоянной отдаче от масштаба

$$H_0 : \beta_2 + \beta_3 = 1$$

при альтернативной гипотезе

$$H_1 : \beta_2 + \beta_3 \neq 1$$

Методические рекомендации по выполнению упражнения 2

1) Создайте новые переменные, набрав в командном окне поочередно

gen lnQ = log(output)

gen lnK = log(capital)

gen lnL = log(labor)

2) Оцените параметры уравнения регрессии с помощью команды

reg lnQ lnL lnK

3) Для проверки гипотезы в командном окне следует набрать:

test (lnL+lnK=1)

и сделать вывод с помощью p -value для F -статистики. Если p -value меньше выбранного уровня значимости, то гипотеза H_0 отвергается.

Дамми (фиктивные) переменные и тест Чоу

Материалы из учебника О.Демидовой и Д.Малахова «Эконометрика. Учебник и практикум»

1) Различия в потреблении россиянами основных типов продуктов

Используйте данные файла Chow.dta.

- 1) Оцените зависимость потребления одного из видов товаров Y от его цены P и дохода домохозяйства I с помощью модели:

$$Y = \beta_1 + \beta_2 P + \beta_3 I + \varepsilon$$

$$\text{или } \ln Y = \beta_1 + \beta_2 \ln P + \beta_3 \ln I + \varepsilon.$$

- 2) Оцените коэффициенты регрессии по двум выборкам, например,

А) для жителей двух различных типов населенных пунктов (2 – город, 3 –поселок городского типа),

Б) для жителей двух различных административных образований (например, Москвы и Московской области)

В) Центрального и Северо-Западного округов и общую.

- 3) С помощью теста Чоу ответьте на вопрос, можно ли считать, что для двух выделенных видов домохозяйств имеет место единая зависимость?

Методические рекомендации по выполнению упражнения 7.5

Предположим, Вы хотите оценить функцию спроса на картофель в виде линейной в логарифмах модели.

Воспользуйтесь соответствующими указаниями в разделе «Оценка регрессий в пакете STATA».

- 1) Создайте новую переменную, набрав в командном окне

gen lnpotat_c = log(buypotat_c)

- 2) Аналогично создайте логарифмы остальных необходимых переменных

- 3) Если Вы хотите оценить регрессию только для домохозяйств Центрального округа, то следует набрать команду:

reg lnpotat_c lnpr_potat lninc if fed_okr ==1

(сохраните RSS с помощью команды *scalar rss1=e(rss)*),

если для домохозяйств Северо-Западного округа, то:

reg lnpotat_c lnpr_potat lninc if fed_okr ==2

(сохраните RSS с помощью команды *scalar rss2=e(rss)*),

для оценки объединенной по двум округам регрессии

reg lnpotat_c lnpr_potat lninc if fed_okr ==1|fed_okr ==2

(сохраните RSS с помощью команды **scalar rssp=e(rss)**),

4) Используя RSS из оцененных регрессий, следует рассчитать тестовую F – статистику

scalar F=((rssp-rss1-rss2)/3)/((rss1+rss2)/(306-2*3))

display F

и p -value для этой статистики:

display Ftail(3, 300, F)

Если рассчитанное p -value не превышает выбранного уровня значимости, то основная гипотеза отвергается, зависимость нельзя считать единой для двух наборов данных.

Выбор функциональной формы модели

I. По данным файла `dougherty.dta`, используя тест Бокса-Кокса, с помощью статистического пакета STATA, оцените параметры модели

$$EARNINGS^{(\theta)} = \beta_0 + \beta_1 S^{(\lambda)} + \beta_2 ASVABC^{(\lambda)} + \varepsilon$$

1) Когда переменные левой и правой части преобразуются не одинаково:
Это можно сделать с помощью команды:

```
boxcox EARNINGS S ASVABC, model (theta)
```

2) Когда переменные в обеих частях модели преобразуются одинаково.
Это можно сделать с помощью команды:

```
boxcox EARNINGS S ASVABC, model (lambda),
```

3) Когда преобразуется только зависимая переменная
Это можно сделать с помощью команды

```
boxcox EARNINGS S ASVABC, model (lhsonly),
```

4) Когда преобразуются только независимые переменные
Это можно сделать с помощью команды

```
boxcox EARNINGS S ASVABC, model (rhsonly),
```

5) поэкспериментируйте с проведением теста Бокса-Кокса в модели с другим набором переменных (если Вы не хотите преобразовывать какие-то переменные, например, `dumtmy`, то это можно сделать с помощью команды, аналогичной следующей:

```
boxcox EARNINGS S ASVABC, notrans(MALE) model (theta)
boxcox EARNINGS S ASVABC, notrans(MALE) model (lambda)
```

II. По данным файла `dougherty.dta` выберите между линейной, полулогарифмической и линейной в логарифмах моделями с помощью теста

- А) РЕ теста Дэвидсона, Уайта и МакКиннона
б) Бера и МакАлера.

а) РЕ теста Дэвидсона, Уайта и МакКиннона

Необходимые команды

Выбор между линейной и линейной в логарифмах моделью

Сначала оценим линейную модель с помощью команды:

```
reg EARNINGS S ASVABC,
```

Сохранить предсказанные значения зависимой переменной можно с помощью команды:

```
predict y_hat
```

Предварительно создав необходимые дополнительные переменные,

аналогично оценим линейную в логарифмах модель с помощью команды,

```
gen lnEARNINGS = ln(EARNINGS)
```

```
gen lnS = ln(S)
```

```
gen lnASVABC = ln(ASVABC)
```

```
reg lnEARNINGS lnS lnASVABC
```

и сохраним предсказанные значения зависимой переменной с помощью команды:

```
predict ln_y_hat.
```

Теперь переходим к шагу 2, оценим дополнительные модели.

Сначала создадим дополнительную разность для линейной модели:

```
gen lin_add= ln(y_hat)- ln_y_hat
```

и оценим эту модель:

```
reg EARNINGS S ASVABC lin_add
```

Создадим также дополнительную разность для линейной в логарифмах модели:

```
gen log_add=y_hat-exp(ln_y_hat)
```

оценим эту модель:

```
reg lnEARNINGS lnS lnASVABC log_add
```

Если

- 1) оба коэффициента при дополнительных разностях значимы или оба незначимы, то выбрать посредством теста Дэвидсона, Уайта и МакКиннона невозможно,
- 2) если незначим только коэффициент при дополнительной разности в линейной модели, то лучше линейная модель,
- 3) если незначим только коэффициент при дополнительной разности в линейной в логарифмах модели, то лучше линейная в логарифмах модель.

Выбор между линейной и полулогарифмической моделями

Сначала оценим линейную модель с помощью команды:

```
reg EARNINGS S ASVABC
```

Сохранить предсказанные значения зависимой переменной можно с помощью команды:

```
predict y_hat
```

Аналогично оценим полулогарифмическую модель с помощью команды:

```
reg lnEARNINGS S ASVABC
```

и сохраним предсказанные значения зависимой переменной с помощью команды:

```
predict semiln_y_hat.
```

Теперь переходим к шагу 2, оценим дополнительные модели.

Сначала создадим дополнительную разность для линейной модели:

```
gen lin_adds= ln(y_hat)- semiln_y_hat
```

и оценим эту модель:

```
reg EARNINGS S ASVABC lin_adds
```

Создадим также дополнительную разность для полулогарифмической модели:

```
gen semilog_add=y_hat-exp(ln_y_hat)
```

оценим эту модель:

```
reg lnEARNINGS S ASVABC semilog_add
```

Если

- 1) оба коэффициента при дополнительных разностях значимы или оба незначимы, то выбрать посредством теста Дэвидсона, Уайта и МакКиннона невозможно,
- 2) если незначим только коэффициент при дополнительной разности в линейной модели, то лучше линейная модель,
- 3) если незначим только коэффициент при дополнительной разности в полулогарифмической модели, то лучше полулогарифмическая модель.

б) Тест Бера и МакАлера

Необходимые команды

Выбор между линейной и полулогарифмической моделями

Сначала оценим полулогарифмическую модель с помощью команды:

```
reg lnEARNINGS S ASVABC
```

и сохраним предсказанные значения зависимой переменной с помощью команды:

```
predict semiln_y_hat.
```

Аналогично оценим линейную модель с помощью команды:

```
reg EARNINGS S ASVABC
```

Сохранить предсказанные значения зависимой переменной можно с помощью команды:

```
predict y_hat
```

Теперь переходим к шагу 2, оценим дополнительные модели.

```
1) gen y1= exp(semiln_y_hat)
```

```
reg y1 S ASVABC
```

Сохранить остатки регрессии можно с помощью команды

```
predict res1, resid
```

```
2) gen y2= ln(y_hat)
```

```
reg y2 S ASVABC
```

Сохранить остатки регрессии можно с помощью команды

```
predict res2, resid
```

Теперь переходим к шагу 3, оценив еще 2 дополнительные модели.

```
reg lnEARNINGS S ASVABC res1
```

```
reg EARNINGS S ASVABC res2
```

Если

- 1) оба коэффициента при res1 и res2 незначимы или оба значимы, то выбрать посредством теста Бера и МакАлера невозможно,
- 2) если незначим только коэффициент при res1, то лучше полулогарифмическая модель,
- 3) если незначим только коэффициент при res2, то лучше линейная модель.

III. По данным файла `dougherty.dta` выберите между линейной и полулогарифмической моделями с помощью теста Зарембки.

Необходимые команды

Выбор между линейной и полулогарифмической моделями

```
reg EARNINGS S ASVABC
```

```
gen lnEARNINGS = ln(EARNINGS)
```

```
reg lnEARNINGS S ASVABC
```

means EARNINGS					
Variable	Type	Obs	Mean	[95% Conf. Interval]	
EARNINGS	Arithmetic	540	19.71924	18.48493	20.95355
	Geometric	540	16.3442	15.54379	17.18584
	Harmonic	540	13.77391	13.05586	14.57555

```
gen EARNINGSstar= EARNINGS/16.3442
```

```
gen lnEARNINGSstar = ln(EARNINGSstar)
```

```
reg EARNINGSstar S ASVABC
```

сохраните RSS с помощью команды `scalar rss3=e(rss)`

```
reg lnEARNINGSstar S ASVABC
```

сохраните RSS с помощью команды `scalar rss4=e(rss)`

Используя RSS из оцененных регрессий, следует рассчитать тестовую F – статистику

$$scalar xi2 = 0.5*540*abs(ln(rss4/rss3))$$

$$display xi2$$

и p-value для этой статистики:

display chi2tail(1, xi2)

Если рассчитанное p-value не превышает выбранного уровня значимости, то основная гипотеза отвергается, есть разница между исходными линейной и полулогарифмическими моделями.