

Данные для этого задания доступны в файле *lect*.

Преподавателя интересует ответ на следующий вопрос: *есть ли причинно-следственная связь между посещением лекций и результатами экзамена по математическому анализу?* Преподавателю доступны данные по переменным  $exam_i$  — количество баллов за экзамен, полученное  $i$ -м студентом (по столбальной шкале) и  $lect_i$  — количество лекций, посещенное  $i$ -м студентом. Преподаватель располагает данными о переменной  $female_i$ , равной единице для студенток и нулю для студентов.

Кроме того, преподавателю известно, что все студенты, участвующие в исследовании, живут в одном из двух корпусов общежития. Часть живет в корпусе совсем рядом с университетом, а часть — в другом корпусе, от которого до университета нужно 45 минут добираться на общественном транспорте (после поступления в университет студенты были расселены по корпусам случайным образом). Используя эти данные, преподаватель создает фиктивную переменную  $d_i$ , равную единице, если  $i$ -ый студент живет рядом с университетом и равную нулю в противном случае.

Лекции по математическому анализу проходили по пятницам. Изучив расписание студентов, преподаватель обнаружил, что у некоторых из них по пятницам были другие пары кроме его лекции, а у некоторых — нет. Преподаватель создал дополнительную фиктивную переменную  $friday_i$ , равную единице, если у  $i$ -го студента по пятницам были исключительно лекции по математическому анализу и ничего больше.

Преподаватель предполагает, что его студенты обладают разным уровнем математических способностей, причем более способные студенты чаще ходят на его лекции, чем менее способные. К сожалению, преподаватель не располагает никакими статистическими данными помимо переменных, описанных выше, и в результате не может включить уровень математических способностей в свою регрессию в качестве контрольной переменной.

**(a)** (2) Оцените регрессию переменной  $\ln(exam)$  на переменную  $lect$  при помощи обычного МНК. Запишите полученное уравнение в стандартной форме. Интерпретируйте результат.

Модель 1: МНК, использованы наблюдения 1-190

Зависимая переменная:  $\ln\_exam$

Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC1

|       | Коэффициент | Ст. ошибка | t-статистика | p-значение |     |
|-------|-------------|------------|--------------|------------|-----|
| const | 3,23017     | 0,118411   | 27,28        | <0,0001    | *** |
| lect  | 0,0662707   | 0,0114363  | 5,795        | <0,0001    | *** |

|                        |           |                          |          |
|------------------------|-----------|--------------------------|----------|
| Среднее завис. перемен | 3,827302  | Ст. откл. завис. перемен | 0,513060 |
| Сумма кв. остатков     | 42,86435  | Ст. ошибка модели        | 0,477495 |
| R-квадрат              | 0,138416  | Исправ. R-квадрат        | 0,133833 |
| F(1, 188)              | 33,57916  | P-значение (F)           | 2,84e-08 |
| Лог. правдоподобие     | -128,1449 | Крит. Акаике             | 260,2898 |
| Крит. Шварца           | 266,7838  | Крит. Хеннана-Куинна     | 262,9204 |

$$\ln(\widehat{exam})_i = \underset{(0,12)}{3,23} + \underset{(0,01)}{0,07} * lect_i$$

$$R^2 = 0,14$$

Так как р-значение для переменной  $lect < 0,001$  переменная значима на 1%. В среднем и прочих равных посещение одной дополнительной лекции увеличивает баллы за экзамен на 7%.

**(б) (3)** Детально объясните, почему эта регрессия скорее всего дает несостоятельный ответ на вопрос, интересующий преподавателя?

Переменная  $lect$  может быть **эндогенной из-за пропуска существенной ненаблюдаемой переменной таланта**: 1. талант коррелирует с переменной  $lect$  так как по предположению преподавателя более способные студенты чаще ходят на его лекции, чем менее способные 2. Талант влияет на результат  $exam$ .

Эндогенность приводит к несостоятельности оценки.

**(в) (5)** Предполагая, что опасения, сформулированные вами в предыдущем пункте, справедливы, постарайтесь решить проблему несостоятельности оценок. Выберите наилучшую модель, используя имеющиеся данные. Развернуто обоснуйте свой выбор.

Для состоятельного оценивания нужно применить 2МНК. Нами были рассмотрены в качестве инструментов различные комбинации переменных  $d$ ,  $female$  и  $friday$ . Они:

1. экзогенные: не влияют на шоки  $exam$
2. релевантные: коррелируют с переменной  $lect$  (чем дальше живёт тем реже посещает, если в пятницу нет других пар, может прогулять, студентки посещают обычно больше лекций чем студенты)

Наилучшая модель выбрана на основе тестирования наборов инструментов на экзогенность с помощью теста Саргана (возможно провести когда инструментов больше чем эндогенных регрессоров) и на слабые инструменты

Переменную пола не стали брать в качестве инструмента так как она не значимо коррелирует с переменной  $lect$ , кроме того модели где она добавлена в качестве инструмента и нет друг от друга не сильно отличаются результатами оценивания.

Коэффициенты корреляции, наблюдения 1 - 190  
 5% критические значения (двухсторонние) = 0,1424 для n = 190

|        |        |         |         |        |
|--------|--------|---------|---------|--------|
| d      | friday | female  | lect    |        |
| 1,0000 | 0,0002 | -0,0104 | 0,6816  | d      |
|        | 1,0000 | 0,0208  | -0,5425 | friday |
|        |        | 1,0000  | -0,0403 | female |
|        |        |         | 1,0000  | lect   |

Модель 17: 2МНК, использованы наблюдения 1-190

Зависимая переменная: l\_exam

Независимые переменные: lect

Инструменты: const d friday

Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC1

|                        | Коэффици-<br>ент | Ст. ошибка               | t-стати-<br>стика | p-значение |     |
|------------------------|------------------|--------------------------|-------------------|------------|-----|
| const                  | 3,73916          | 0,138976                 | 26,91             | <0,0001    | *** |
| lect                   | 0,00978239       | 0,0148392                | 0,6592            | 0,5106     |     |
| Среднее завис. перемен | 3,827302         | Ст. откл. завис. перемен |                   | 0,513060   |     |
| Сумма кв. остатков     | 47,86766         | Ст. ошибка модели        |                   | 0,504594   |     |
| R-квадрат              | 0,138416         | Исправ. R-квадрат        |                   | 0,133833   |     |
| Хи-квадрат(1)          | 0,434577         | p-значение               |                   | 0,509752   |     |

Тест Хаусмана (Hausman) -

Нулевая гипотеза: МНК оценки состоятельны

Асимптотическая тестовая статистика: Хи-квадрат(1) = 110,528

p-значение = 7,50616e-26

Тест на сверхидентификацию Саргана (Sargan) -

Нулевая гипотеза: все инструменты допустимы

Тестовая статистика: LM = 0,0512467

p-значение = P(Хи-квадрат(1) > 0,0512467) = 0,820908

Тест на слабые инструменты -

F-статистика для 1-го шага (2, 187) = 329,564

**(в) (3) Оцените выбранную модель. Запишите в стандартной форме все регрессии, которые вы оценили в рамках этой модели.**

1ШАГ

$$\widehat{lect}_i = 8,66 + 3,92 * d_i - 3,12 * friday_i$$

(0,18)
(0,21)
(0,21)

## 2ШАГ

$$\ln(\widehat{exam})_i = \underset{(0,14)}{3,74} + \underset{(0,015)}{0,01} * \widehat{lect}_i$$

**(г) (5)** Справедливы ли были ваши опасения относительно несостоятельности оценки в пункте (а)? Удалось ли решить проблему? Ответьте на эти вопросы, используя результаты соответствующих формальных тестов.

Тест Хаусмана (Hausman) -

Нулевая гипотеза: МНК оценки состоятельны

р-значение =  $7,50616e-26 < 0,01$  значит на 1% **мнк оценки действительно были несостоятельными.**

Тест на свехидентификацию Саргана (Sargan) -

Нулевая гипотеза: все инструменты допустимы

р-значение =  $0,820908 > 0,05$  значит на 5% **инструменты экзогенные**

Тест на слабые инструменты -

F-статистика для 1-го шага =  $329,564 > 10$  значит **инструменты сильные**

**(д) (2)** На основе полученных результатов, ответьте на вопрос, сформулированный в самом начале этого задания.

2мнк оценка вышла не значимой на любом общепринятом уровне значимости. Не установлено значимой *причинно-следственная связь между посещением лекций и результатами экзамена по математическому анализу.*