

Эконометрика — 2022

Контрольная работа

Пробный вариант

Ноябрь 2022

Дорогие студенты! Контрольная работа - это 100% самостоятельная письменная проверочная работа, которая проводится в течение 80 минут в формате "closed-book". Вам не разрешается использовать какие-либо материалы или устройства, кроме персонального калькулятора, который не находится в телефоне или ином коммуникационном или носимом смарт-устройстве или т.п. Общение с кем-либо в или вне аудитории не разрешено. Если у вас есть вопрос, задайте его преподавателю или ассистенту.

Каждый вопрос задания имеет фиксированную стоимость в баллах, которая указана в квадратных скобках после номера вопроса. Например, "3. [8] ..." означает, что ответ на вопрос номер 3 может принести вам до 8 баллов. Отвечайте на вопросы аккуратно и максимально кратко, по сути. Примите к сведению и помните, что во всех вопросах правильный ответ является кратким и ясным. При этом запутанные ответы или ответы, содержащие не относящуюся к вопросу информацию, приведут к штрафам (будут оценены отрицательными баллами).

Оценка за работу определяется по сумме набранных баллов согласно шкале перевода:

Баллы	до 20	[20;30)	[30;40)	[40;50)	[50;60)	[60;70)	[70;75)	[75;80)	[80;85)	85 и более
Оценка	1	2	3	4	5	6	7	8	9	10

Распределяйте своё время и усилия разумно. Удачи!

Представьтесь

0. [1] Фамилия Имя Группа

Да-нетки

В этой части нужно согласиться или нет с утверждением или ответить на вопрос ДА или НЕТ, вычеркнув неверный ответ рядом с вопросом.

1. [1] Эконометрика ограничена использованием вероятностного параметрического подхода ДА | НЕТ
2. [1] Выборочные аналоги существуют только в условиях случайной выборки ДА | НЕТ
3. [1] Прогнозирование будущего не является эконометрической задачей ДА | НЕТ
4. [1] Выборочная величина $n^{-1} \sum_{i=1}^n z_i x_i^2$ – это выборочный аналог для $\mathbb{E} z x^2$ ДА | НЕТ
5. [1] Ошибка регрессии Y на X – это разница $Y - \mathbb{E}[Y|X]$ ДА | НЕТ
6. [2] Если $\bar{x} \xrightarrow{P} 0$, и $\bar{y} \xrightarrow{d} \mathcal{N}(0, 1)$, то $\bar{x} \cdot \bar{y} \xrightarrow{P} 0$ ДА | НЕТ
7. [2] Модель $\mathbb{E} y = \mu$ – это пример линейной регрессии ДА | НЕТ

8. [2] Оценки методом аналогии всегда состоятельны ДА | НЕТ

9. [3] Если $\sqrt{n}(\hat{\mu} - \mu) \rightarrow \mathcal{N}(0, 1)$, то $SE(\hat{\mu}) = 1$ ДА | НЕТ

10. [3] В модели $E[\log(y_i)|x_i = x] = \beta_0 + \beta_1 x$ при изменении x на 1 ед. y в среднем меняется на $\beta_1\%$ ДА | НЕТ

Задача 1. Mind the gap

В этой части нужно ответить на вопросы о гипотетическом исследовании. Старайтесь максимально придерживаться курса. Ответы, требующие формул, должны быть записаны в стандартных выражениях. Ответы, требующие пояснений, должны быть выполнены с использованием обсуждаемого исследования и модели. Ответы, требующих примеров, должны быть конкретными примерами. Ответы должны вмещаться в рамки после вопроса.

В последнее время достаточно активно обсуждается идея, что мужчины получают более высокие зарплаты, чем женщины такой же квалификации. Вас заинтересовал этот вопрос, и вы провели опрос на улицах Санкт-Петербурга, опрашивая мужчин и женщин об их зарплате, возрасте, образовании и опыте работы. Опросив $n = 300$ человек, вы получили данные в следующем формате:

- w_i – зарплата индивида, руб/мес, в среднем за 2020 год,
- a_i – возраст индивида, полных лет,
- x_i – общий рабочий стаж индивида, полных лет,
- e_i – образование индивида, полных лет, потраченных на образование выше среднего,
- $f_i = 1$, если индивид – женщина, $= 0$ иначе,

при этом выборка достаточно сбалансирована по этим переменным, в ней нет пропусков и сильных выбросов.

Вы решаете, что этих переменных достаточно для постановки и успешного применения ММЛР для ответа на вопрос **"Действительно ли женщины получают в среднем зарплату ниже, чем мужчины с таким же образованием и опытом (и при прочих равных)?"**

11. [2] Запишите основное регрессионное уравнение, соответствующее этому вопросу и переменным. Уравнение в линейно-линейной форме. Возраст, как дополнительная контрольная переменная.

12. [4] Вы считаете, что это уравнение может быть основной ММЛР, и не будет проблемы пропущенных переменных. Какие основные предположения о модели нужно сделать, чтобы продолжить работу? *Выписывайте предположения в обозначениях вашей модели.*

13. [1] Какой метод оценивания параметров вы будете использовать?

14. [1] А как стоит оценить матрицу ковариаций оценок параметров?

15. [4] Запишите гипотезы, соответствующие вашему основному вопросу, в модельной и нормальной форме.

16. [2] Какой тест стоит использовать для проверки этих гипотез?

17. [2] Запишите формулу тестовой статистики.

18. [4] Вы провели оценивание и получили значение статистики (из предыдущих двух вопросов) -3.82. Ваша интерпретация результатов теста и предварительный ответ на вопрос:

19. [6] Приведите три *разных и конкретных* примера угроз *внутренней* валидности этого исследования.

20. [6] Приведите три *разных и конкретных* примера угроз *внешней* валидности этого исследования.

Задача 2. Attendance matters

В этой части приводится пример небольшого эмпирического исследования. Необходимо ответить на вопросы об исследовании, по возможности подкрепляя их расчетами на основе предоставленных данных (Таблицы 1 и 2). Если расчеты занимают слишком много времени, постарайтесь описать хотя бы их ход и логику.

Исследователи обратили внимание на вопрос о влиянии посещения занятий на успеваемость студентов в отдельно взятом вузе галактики. Был проведен сбор данных за весенний семестр одного вполне типичного учебного года. Зафиксирована следующая информация об $n = 680$ студентах:

- ΔGPA – насколько отличается средний балл за семестр от накопленного среднего балла на начало семестра (положительные значения – средний балл улучшился, отрицательные – ухудшился),
- $InitialGPA$ – накопленный средний балл на начало семестра,
- $Final$ – стандартизированная средняя оценка за финальные экзамены в семестре (положительное значение – студент сдал экзамены лучше среднего на своем потоке, отрицательные – хуже),
- $Missed$ – пропущено занятий (от 0 – посетил все, до 30 – пропустил все),
- $HWRate$ – % сданных домашних (от 0 – ничего не сдал, до 100 – сдал все),
- $First, Second$ – первокурсник, второкурсник (дамми, не образуют полную группу, базовая группа – "старшекурсники").

То, что студенты, которые меньше пропускают занятия, больше занимаются дома и больше готовятся к экзаменам, имеют более высокую успеваемость, для исследователей довольно очевидно. Однако им хотелось проверить, как эффект пропуска занятий сказывается на студентах второго курса, т.к. именно в этот год студенты, с одной стороны, пропускают занятия больше, с другой стороны к ним часто предъявляются не такие высокие требования, как в первый год учебы. Таким образом, исследователей интересует, как отличается влияние пропущенных занятий и интенсивности самостоятельной работы на успеваемости студентов второго курса. Однако исследователи до конца не уверены в возможных направлениях этих эффектов. Они оценили две спецификации ММЛР с изменением среднего балла за семестр в качестве зависимой переменной (см Таблицу 1 на стр. 7). Ответьте на следующие вопросы.

21. [2] Выпишите основные уравнения регрессии, соответствующие спецификациям (1) и (2) из Таблицы 1

22. [4] В контексте исследовательского вопроса, в чем основная проблема спецификации (1)?

23. [6] Объясните роль переменных $InitialGPA$ и $Final$ в спецификациях. Не создают ли они проблему взаимной причинности или т.п.?

24. [6] Исходя из исследовательского вопроса, сформулируйте основную гипотезу исследования в содержательной, модельной и нормальной формах (*не забудьте про корректные альтернативные гипотезы*¹)

25. [3] Как выразить основную гипотезу в терминах предельных эффектов регрессоров?

26. [4] Что можно предварительно сказать об ответе на основной вопрос и гипотезе, основываясь только на оценках коэффициентов в Таблице 1?

27. [4] Прокомментируйте оценку перекрестного эффекта $Final \times Second$ в спецификации (2). Не стоит ли удалить этот эффект и спецификации?

28. [4] Запишите формулу тестовой статистики для проверки основной гипотезы. Каково ее предельное распределение при верной нулевой гипотезе?

¹напоминалка, которой за просто не будет в настоящем варианте КР

29. [6] Проверьте основную гипотезу.

Напоминание: критические значения для статистики хи-квадрат для $\alpha = 0.05$:

m	1	2	3	4	5	6
qchisq(0.95,m)	3.84	5.99	7.81	9.49	11.07	12.59

30. [6] Приведите три *разных и конкретных* примера угроз *внутренней* валидности этого исследования.

31. [3] Для одного из примеров в предыдущем вопросе предложите вариант решения

Таблица 1: Результаты оценивания

Зависимая переменная:		
	ΔGPA	
	(1)	(2)
1. Константа	0.562*** (0.188)	0.197 (0.244)
2. InitialGPA	-0.461*** (0.042)	-0.456*** (0.042)
3. HWRate	0.008*** (0.001)	0.012*** (0.002)
4. Missed	-0.025*** (0.005)	-0.013* (0.008)
5. Final	0.232*** (0.020)	0.262*** (0.027)
6. First	0.110* (0.059)	0.132** (0.058)
7. Second	0.042 (0.048)	0.727** (0.293)
8. Missed×Second		-0.021** (0.010)
9. HWRate×Second		-0.006** (0.003)
10. Final×Second		-0.053 (0.035)

Примечание: *p<0.1; **p<0.05; ***p<0.01
робастные стандартные ошибки в скобках

Таблица 2: Оценка матрицы ковариаций оценок параметров регрессии (2), оценка методом Уайта

	1	2	3	4	5	6	7	8	9	10
1	40.538	-3.329	-0.288	-0.910	0.854	-1.614	-31.767	0.790	0.300	0.148
2	-3.329	1.202	-0.003	0.020	-0.169	0.566	0.155	0.022	-0.003	-0.192
3	-0.288	-0.003	0.002	0.007	-0.006	-0.012	0.293	-0.008	-0.003	0.005
4	-0.910	0.020	0.007	0.042	-0.001	-0.040	0.860	-0.042	-0.008	-0.009
5	0.854	-0.169	-0.006	-0.001	0.512	0.107	-0.401	-0.006	0.005	-0.466
6	-1.614	0.566	-0.012	-0.040	0.107	2.318	0.073	0.059	0.009	-0.257
7	-31.767	0.155	0.293	0.860	-0.401	0.073	58.277	-1.468	-0.554	0.138
8	0.790	0.022	-0.008	-0.042	-0.006	0.059	-1.468	0.071	0.012	0.018
9	0.300	-0.003	-0.003	-0.008	0.005	0.009	-0.554	0.012	0.005	-0.004
10	0.148	-0.192	0.005	-0.009	-0.466	-0.257	0.138	0.018	-0.004	0.819

Примечание: нумерация строк и столбцов по порядку регрессоров в таблице 1

