

Микроэконометрика

Домашнее задание №1

I. Оформление и сроки

1. На титульном листе должны быть указаны: имя, фамилия, группа и номер домашнего задания.

2. Таблицы и графики должны быть подписаны и пронумерованы. Все оси подписываются на русском языке.

3. Все приведенные графики и таблицы должны быть проинтерпретированы со ссылкой на их номер.

4. Задание можно выполнять, используя любые программные средства (не обязательно R или Stata). К работе прикладываются файлы (R-файл, do-файл и т.д.), при помощи которых можно воспроизвести результаты вашего исследования на обычном компьютере с операционной системой Windows 11. Из содержания приложенного файла должно быть ясно, какая часть кода или выполненных операций относится к какому заданию. Например, в .R или .do файле можно расставить комментарии по поводу того, какая часть кода относится к тому или иному заданию.

5. Работу следует отправить на почту studypotantin@gmail.com **не позднее чем 20.10.2023 включительно**. Тема письма **обязательно** должна иметь следующий формат “Фамилия_Имя_Группа_Микроэконометрика”, например, “Потанин_Богдан_123_Микроэконометрика”. В противном случае отправленное вами задание скорее всего не будет получено.

6. Все пункты работы должны быть выполнены **последовательно** с указанием **текста задания (очень важно!)**. То есть оформление должно быть следующим:

Задание №1. Текст задания.

Ваш текст.

Задание №2. Текст задания.

Ваш текст.

...

Задание №N. Текст задания.

Ваш текст.

Где N – номер последнего задания.

7. Все задания выполняются индивидуально.

8. Все переменные в таблицах и на графиках должны быть понятным образом обозначены. Например, название переменной `ufj15` — не подходит, даже если в тексте указано, что именно данная переменная означает. Если эта переменная относится к, например, полу, то её следует назвать `male`, `gender` или `sex`, а также дать интерпретацию кодам, например, 0 – женщина, 1 – мужчина.

9. Работа отправляется **только в pdf** формате и **обязательно** должна иметь название “Фамилия_Имя_Группа_Микроэконометрика.pdf”, например, “Потанин_Богдан_123_Микроэконометрика.pdf”.

10. Во всех заданиях, где требуется оценить что-либо для индивида с произвольными характеристиками, необходимо, предварительно, выписать в форме таблицы характеристики соответствующего индивида. Если характеристики произвольного индивида на протяжении работы не меняются, то можно указать их в отдельной таблице в приложении и ссылаться на нее на протяжении работы.

11. При проведении **всех** тестов необходимо в тексте работы указать **нулевую гипотезу** и **тестовую статистику**, а также её **асимптотическое распределение при верной нулевой гипотезе**. Исключение составляют лишь обычные тесты о значимости одного регрессионного коэффициента, по поводу которых достаточно привести лишь краткую устную интерпретацию полученного результата.

12. Формулы можно как набирать на компьютере, так и вставлять картинкой с аккуратным рукописным тестом. Копировать формулы из других источников нельзя.

13. Необходимо подробно прояснять **все** используемые обозначения, включая индексы переменных.

14. При выполнении отдельных заданий можно менять набор используемых переменных, предварительно оговорив это в тексте.

II. Задания

Игровая студия разработала новый градостроительный симулятор про кошек и собак. Игроки были приглашены на закрытое бета-тестирование с целью описать свои впечатления от однократного прохождения половины игры. Вас наняли для того, чтобы проанализировать, как различные социо-демографические факторы влияют на удовлетворенность игроков от прохождения разработанной игры. Также, вас отдельно

попросили тщательно проанализировать, как на удовлетворенность от игры влияет выбор игрока проходить игру за кошек или за собак.

В качестве выборки используйте данные «homework.rds». Чтобы считать соответствующие данные в R необходимо воспользоваться функцией «readRDS».

Данные содержат информацию о следующих индивидуальных характеристиках:

- game — степень удовлетворенности игрой (0 — низкая, 1 — средняя, 2 — высокая)
- choice — за кого проходил игру игрок (1 — кошки, 0 — собаки)
- income — доход в рублях
- age — возраст в годах
- hours — среднее количество игровых часов в неделю
- health — здоровье (bad — плохое, medium — среднее, good — хорошее)
- marriage — семейный статус (1 — в браке, 0 — холостой)
- chl — факт наличия детей (1 — есть, 0 — нет)
- male — пол (1 — мужской, 0 — женский)
- residence — место проживания (capital — столица региона, city — город, village — деревня)
- cat — факт наличия кота (1 — есть кот, 0 — нет кота)
- dog — факт наличия собаки (1 — есть собака, 0 — нет собаки)
- bugs — число багов, в которыми пользователь столкнулся на протяжении игры
- price — стоимость компьютера игрока в рублях
- design — какой из двух видов дизайна был случайным образом предложен игроку в меню, в котором осуществлялся выбор в пользу прохождения игры за кошек или за собак (1 — первый дизайн, 0 — второй дизайн)
- educ — уровень образования (basic — среднее, vocational — среднее специальное, higher — высшее, phd — кандидат наук)

Часть 1. Теория и гипотезы.

1.1. Выберите независимые переменные. Кратко теоретически обоснуйте выбор каждой из них: не обязательно со ссылками на литературу, достаточно здравого смысла. Укажите и кратко обоснуйте предполагаемые направления эффектов. При этом вам понадобится как минимум одна непрерывная переменная (например, возраст или доход) и

одна дамми переменная (например, половая принадлежность или брак). Не рекомендуется брать больше трех различных независимых переменных, не считая их нелинейных преобразований: квадрат, логарифм, перемножение с целью получения переменной взаимодействия и т.д.

1.2. Сформулируйте по крайней мере одну гипотезу о наличии эффекта взаимодействия и **еще** одну о наличии нелинейного эффекта (например, квадратичного). Теоретически обоснуйте выдвигаемые вами гипотезы. Включите соответствующие переменные в вашу модель. При этом переменная, входящая нелинейно, должна иметь и линейную часть, например, $\beta_1 * X_1 + \beta_2 * X_1^2$.

1.3. Создайте новую бинарную переменную на удовлетворённость игрой, объединив любые две соседние категории переменной «game». Данная переменная выступает в качестве зависимой при оценивании моделей бинарного выбора. Для краткости далее в условии будем считать, что 1 значение соответствует тому, что индивид удовлетворен игрой, а 0 тому, что он не удовлетворен. Приведите и далее в тексте вашей работы используйте более подробную содержательную интерпретацию 1 и 0 значений созданной вами переменной, учитывающую то, что к одному из них относятся сразу две категории.

Часть 2. Линейно-вероятностная модель.

2.1. Оцените линейно-вероятностную модель бинарного выбора. Предварительно запишите регрессионное уравнение. Укажите оцениваемые параметры и метод получения оценок. Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R или python).

2.2. Перечислите основные недостатки линейно-вероятностной модели. Напишите, можно ли интерпретировать оценки коэффициентов, их значимость (с использованием обычной оценки ковариационной матрицы), коэффициент детерминации и F-статистику? Если да, то приведите интерпретацию, а если нет, то объясните (без непосредственной реализации), почему она в данном случае невозможна и предложите альтернативный способ оценки качества модели.

2.3. Оцените и проинтерпретируйте, независимо от значимости, предельные эффекты на вероятность подписки каждой из используемых вами независимых переменных, предварительно записав формулы, по которым осуществлялся расчет. Результат представьте в форме таблицы, где для переменных, входящих нелинейно, рассчитан средний предельный эффект. Также, для этих переменных должно быть указано,

при каких значениях независимой переменной их предельный эффект является положительным, а при каких — отрицательным.

2.4*. Протестируйте гипотезы о значимости коэффициентов с помощью бутстрапа. Результат представьте в форме таблицы. При этом предварительно (словами или самостоятельно нарисованной схемой) опишите алгоритм, который вы использовали для построения бутстрапированных доверительных интервалов.

2.5**. Протестируйте гипотезы о значимости коэффициентов используя состоятельную (скорректированную на гетероскедастичность) оценку асимптотической ковариационной матрицы. Результат представьте в форме таблицы (в том числе с p-value), предварительно выписав используемую для расчетов формулу.

Часть 3. Пробит модель.

3.1. Оцените пробит модель бинарного выбора. Предварительно запишите максимизируемую функцию правдоподобия (поясните все используемые обозначения), указав оцениваемые параметры и метод получения оценок, а также их основные свойства. Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R или python).

3.2. Проинтерпретируйте оценки коэффициентов для каждой независимой переменной. Поясните, как полученные результаты соотносятся с высказанными вами ранее предположениями.

3.3. Оцените вероятность того, что индивид удовлетворен игрой, для индивида с произвольными (например, вашими) характеристиками. Запишите формулу, по которой осуществлялся расчет (подставьте в нее полученные реализации оценок).

3.4. Для произвольных непрерывной и бинарной независимых переменных оцените средний предельный эффект на вероятность удовлетворенности игрой, предварительно записав формулы (с подставленными реализациями оценок), по которым осуществлялся расчет. Результат представьте в форме таблицы.

3.5. Посчитайте долю верных прогнозов и сопоставьте её с результатом наивного прогноза и линейно-вероятностной модели. Сделайте вывод о предсказательной силе пробит модели.

3.6. На уровне значимости 5% проверьте гипотезу о том, что предельный эффект на вероятность удовлетворенности игрой по произвольной (на ваш выбор) независимой переменной является значимым для индивида с произвольными характеристиками.

3.7*. Повторите предыдущий пункт для переменной, имеющей взаимодействие.

3.8**. Воспользуйтесь дельта-методом для того, чтобы вручную рассчитать асимптотическую дисперсию предельного эффекта любой непрерывной переменной, входящей нелинейно. Запишите использовавшееся вами выражение, подставив в него численные значения оценок пробит модели.

Часть 4. Тестирование корректности спецификации пробит модели.

4.1. При помощи LM-теста проверьте гипотезу о соблюдении допущения о нормальном распределении случайных ошибок в пробит модели. Укажите, к каким негативным последствиям может привести нарушение данного допущения.

4.2. Предположите, какие переменные могут влиять на дисперсию случайной ошибки. При этом по крайней мере одна переменная должна входить и в линейный индекс основного уравнения, и в линейный индекс уравнения дисперсии. При помощи LR теста проверьте гипотезу о гомоскедастичности случайных ошибок. Запишите, к каким негативным последствиям может привести нарушение данного допущения. Объясните преимущество LM теста над LR тестом в данном случае.

4.3. Для модели с гетероскедастичной случайной ошибкой рассчитайте предельный эффект на 1) вероятность удовлетворенности игрой и на 2) дисперсию случайной ошибки по переменной, входящей и в основное уравнение, и уравнение дисперсии. Предварительно запишите формулы, по которым осуществляется расчет.

4.4. Для переменной, входящей в линейный индекс нелинейно, при помощи LR теста проверьте гипотезы о том, что:

- 1) Коэффициент при линейной части равняется нулю
- 2) Оба коэффициента равняются нулю
- 3) Коэффициент при линейной части совпадает по знаку и в k раз больше (по модулю), чем при нелинейной, где $k \neq 0$ можно выбрать произвольным, указав выбранное значение.
- 4) Коэффициент при линейной части совпадает по знаку и в k раз больше, чем при нелинейной, а коэффициент при произвольной бинарной переменной равняется t , где $t \neq 0$ можно выбрать произвольным, указав выбранное значение.

4.5. При помощи LR теста проверьте, можно ли оценивать совместную модель для мужчин и для женщин, либо стоит оценить две различные модели.

4.6*. При помощи LR теста проверьте, можно ли оценивать совместную модель для людей, проживающих в населенных пунктах различного типа (рассмотрите все три возможных типа населенного пункта).

4.7**. Для индивида с произвольными характеристиками оцените асимптотическую дисперсию предельного эффекта любой непрерывной переменной, входящей сразу в оба уравнения, на вероятность удовлетворенности игрой. Запишите использовавшуюся вами формулу.

Часть 5. Логит модель.

5.1. Оцените логит модель, предварительно записав максимизируемую функцию правдоподобия и указав, чем логит модель отличается от пробит модели. Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R или python).

5.2. Проинтерпретируйте значения оценок изменений в отношениях шансов по каждой независимой переменной, входящей линейно.

5.3*. Запишите выражения для расчета изменений в отношениях шансов по каждой независимой переменной, входящей нелинейно. Рассчитайте соответствующие предельные эффекты для индивида с произвольными характеристиками. Результаты расчетов представьте в форме таблицы.

5.4**. Для произвольной независимой переменной, входящей нелинейно, проверьте гипотезу о том, что при её изменении на единицу отношение шансов возрастет на 10%. Запишите тестовую статистику, а также формулу, по которой рассчитывалась стандартная ошибка оценки изменения отношений шансов.

Часть 6. Система бинарных уравнений.

6.1. Оцените систему бинарных уравнений, одно из которых описывает вероятность удовлетворенности игрой, а второе — вероятность того, что индивид станет играть за кошек. При спецификации уравнений соблюдайте наличие ограничений исключения.

6.2. Проинтерпретируйте оценки коэффициентов при независимых переменных и коэффициента корреляции между случайными ошибками рассматриваемых уравнений.

6.3. При помощи LR теста проверьте, имеется ли необходимость в том, чтобы оценивать оба уравнения совместно.

6.4. Для индивида с произвольными характеристиками оцените:

- 1) Вероятность того, что игрок удовлетворен игрой
- 2) Вероятность того, что игрок проходил игру за кошек
- 3) Вероятность того, что игрок неудовлетворен игрой и играл за собак

4) Вероятность того, что игрок удовлетворен игрой, при условии, что он играл за кошек

6.5*. Для индивида с произвольными характеристиками оцените предельный эффект общей для двух уравнений переменной на вероятность того, что индивид удовлетворен игрой, при условии, что он играл за кошек. Запишите использовавшуюся для расчетов формулу. Протестируйте значимость этого предельного эффекта.

6.6*. Оцените средний эффект воздействия и средний эффект воздействия на подвергшихся воздействию прохождения игры за кошек на удовлетворенность игрой. Протестируйте значимость этих эффектов воздействия. Объясните разницу между рассчитанными вами эффектами воздействия, а также, чем они отличаются от оцененного в предыдущем пункте предельного эффекта.

6.7**. Замените в рассматриваемой системе бинарную переменную удовлетворенности игрой на порядковую «game». Оцените по крайней мере два различных средних эффекта воздействия (самостоятельно сформулируйте эти средние эффекты воздействия и дайте их интерпретацию) и протестируйте их значимость.

Часть 7. Сравнение моделей

7.1. Определите, какая из оцененных вами моделей лучше прогнозирует удовлетворенность игрой.

7.2. Выберите лучшую из оцененных вами моделей руководствуясь информационными критериями.

7.3**. Руководствуясь информационным критерием Акаике и качеством вневыборочного прогноза (разделите выборку на тестовую и тренировочную части) определите оптимальные маргинальные распределения для случайных ошибок и копулу. Укажите, совпадают ли лучшие модели, выбранные по AIC и вневыборочному прогнозу. В данном задании рассмотрите по крайней мере 5 различных копул и две различные комбинации маргинальных распределений случайных ошибок. Результат представьте в форме таблицы.

Часть 8. Модель бинарного выбора со случайными ошибками, имеющими распределение Стьюдента

Рассмотрим распределение Стьюдента с df степенями свободы, где df равняется сумме букв в ваших имени и фамилии.

8.1***. Используя воображение придумайте и кратко опишите экономическую задачу, для решения которой необходимо применить модель бинарного выбора. Например, можно рассмотреть влияние различных факторов на вероятность дефолта банка. Укажите зависимую переменную и по крайней мере три независимых, а также кратко опишите предполагаемый механизм влияния независимых переменных на зависимую.

8.2***. Симулируйте процесс генерации данных, соответствующий вашей задаче в логике бинарной модели со случайными ошибками, имеющими распределение Стьюдента (по аналогии с тем, как это делалось для обычной пробит модели для дефолта на семинаре). В тексте работы этот пункт отражать не нужно, достаточно реализовать его в коде. Все дальнейшие пункты выполняются на симулированных данных из выборки объемом 5000 наблюдений. Перед началом симуляций необходимо указать `set.seed(123)`.

8.3***. Оцените параметры вашей модели с использованием бинарной модели со случайными ошибками, имеющими распределение Стьюдента с df степенями свободы. Результат представьте в форме таблицы, содержащей оценки коэффициентов и p -value тестов на значимость.

8.4***. Напишите функцию, позволяющую рассчитывать предельный эффект на вероятность по произвольной непрерывной переменной в отношении индивида с произвольными характеристиками: для простоты переменными, входящими нелинейно, можно пренебречь. При помощи данной функции оцените предельный эффект произвольной непрерывной переменной на вероятность успеха для индивида с произвольными характеристиками.

8.5***. Симулируйте процесс генерации данных таким образом, чтобы дисперсия случайной ошибки зависела по крайней мере от двух независимых переменных, одна из которых не входит в основное уравнение. Выполните одно задание на выбор:

1) Напишите функцию, осуществляющую LM тест на гомоскедастичность случайных ошибок в рамках вашей модели. Воспользуйтесь вашей функцией и опишите полученный результат.

2) Напишите функцию, при помощи которой можно оценить параметры модели бинарного выбора с гетероскедастичной случайной ошибкой, имеющей распределение Стьюдента с df степенями свободы. При помощи вашей функции и LR теста проверьте гипотезу о гомоскедастичности случайных ошибок.

8.6***. Повторите предыдущие пункты, предполагая, что степени свободы df не фиксированы, а являются оцениваемым параметром. При этом для идентифицируемости параметров необходимо стандартизировать распределение Стьюдента таким образом, чтобы дисперсия случайной ошибки равнялась 1, независимо от числа степеней свободы df .

Подсказка: посмотрите функцию `pt0` в пакете `mporm` через `help(pt0)`.