

# Микроэконометрика

## Домашнее задание

### 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. 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 [studypotandin@gmail.com](mailto:studypotandin@gmail.com) **не позднее чем 01.11.2021 до 8.00 утра 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**. 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.

## II. Задания

В данной работе вы будете изучать, как различные факторы влияют на вероятность того, что индивид оформит подписку на онлайн кинотеатр. Для этого используйте данные “homework.excel”. Чтобы считать соответствующие данные в R необходимо подключить библиотеку “xlsx” и воспользоваться функцией “read.xlsx”.

Данные содержат информацию о следующих индивидуальных характеристиках:

- `income` — доход
- `age` — возраст
- `internet` — доля свободного времени, проводимого в интернете
- `series` — количество просмотренных за год сериалов
- `health` — субъективная оценка здоровья

- male — половая принадлежность
- marriage — состоит в официальном браке
- residence — место проживания
- cat — факт наличия кота
- news — субъективная оценка степени, в которой индивид интересуется новостями
- sub — факт наличия подписки на онлайн кинотеатр (зависимая переменная)
- TV — индивид смотрит телевизор не реже раза в неделю

## **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 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,  $\beta_1 * X_1 + \beta_2 * X_1^2$ .

## **Часть 2. Линейно-вероятностная модель.**

2.1. Оцените линейно-вероятностную модель, предварительно записав 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.

### **Часть 3. Пробит модель.**

3.1. Оцените пробит модель, предварительно записав максимизируемую функцию правдоподобия, 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 наличия подписки 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 наличия подписки, предварительно записав формулы (с подставленными реализациями оценок), по которым осуществлялся расчет. Результат представьте в форме таблицы.

3.5. Посчитайте долю верных предсказаний 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 наличия подписки по произвольной (на ваш выбор) независимой переменной является значимым для индивида с произвольными характеристиками.

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

#### **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 вероятность занятости и на 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).

#### **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%.

## **Часть 6. Система бинарных уравнений.**

6.1. Оцените систему бинарных уравнений, одно из которых описывает вероятность подписки, а второе — вероятность того, что индивид смотрит телевизор не реже раза в неделю. При этом оба уравнения должны иметь по крайней мере одну общую и одну различающуюся независимую переменную. При необходимости спецификация уравнения подписки может отличаться от той, что использовалась в предыдущих разделах.

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 подписки
- 2) Вероятность того, что индивид смотрит телевизор по крайней мере раз в неделю
- 3) Вероятность того, что индивид и имеет подписку, и смотрит телевизор не реже раза в неделю
- 4) Вероятность того, что у индивида имеется подписка, при условии, что он смотрит телевизор реже раза в неделю

6.5\*\*. Для индивида с произвольными характеристиками оцените предельный эффект общей для двух уравнений переменной на вероятность того, что индивид имеет подписку, при условии, что во втором уравнении зависимая бинарная переменная примет единичное значение. Запишите использовавшуюся для расчетов формулу.

## **Часть 7. Сравнение моделей**

7.1. Определите, какая из оцененных вами моделей обладает наибольшей предсказательной силой.

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 Стюдента**

Рассмотрим распределение Стюдента 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 Стюдента (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 Стюдента 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 занятости 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.