

# 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. 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 [bogdanpotanin@gmail.com](mailto:bogdanpotanin@gmail.com) до 01.12.2019 включительно. Тема письма **обязательно** должна иметь следующий формат “Фамилия\_Имя\_Группа\_Микроэконометрика\_1”. В противном случае отправленное вами задание не будет получено.

6. Все пункты работы должны быть выполнены **последовательно**. 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.

## II. Задания

В данной работе вы будете изучать, как различные факторы влияют на вероятность того, что индивид состоит в браке. В данных РМЭЗ эта переменная закодирована как “u\_marst”.

Скачайте базу данных РМЭЗ за 2016 год (25 волна) по индивидам. Это те же самые данные, которые использовались нами на семинаре.

**Обратите внимание**, что некоторые задания имеют от одной до трех звездочек: в зависимости от уровня сложности. Выполнение заданий со звездочками не обязательно для получения оценки до восьми включительно. Для получения оценки девять необходимо выполнить по крайней мере одно задание со звездочкой. Получить оценку десять можно при условии выполнения заданий, суммарное количество звездочек в которых равняется, по крайней мере, трем: например, можно выполнить три задания с одной звездочкой либо одно задание с тремя звездочками. Верное выполнение заданий со звездочками также позволит получить дополнительные баллы, тем самым компенсируя возможные потери баллов в других заданиях вследствие допущенных неточностей.

### Часть 1. Теория и гипотезы.

1.1. Выберите независимые переменные. Теоретически обоснуйте выбор каждой из них. Укажите предполагаемые направления эффектов. При этом вам понадобится как минимум одна непрерывная переменная и одна дамми переменная (не рекомендуется брать больше трех различных переменных, не считая их нелинейных преобразований: квадрат, логарифм, перемножение с целью получения переменной взаимодействия и т.д.).

Приведите по крайней мере две ссылки на научные работы (желательно из Q1 журналов<sup>1</sup>), изучавшие влияние различных факторов на вероятность брака. Кратко опишите методологию и основные результаты этих исследований, а также как ваш выбор переменных и сформулированные гипотезы соотносятся с данными результатами.

---

<sup>1</sup> Чтобы узнать квартиль журнала, следует написать в google “Название журнала scimago” и перейти по первой ссылке.

Постарайтесь подбирать независимые переменные таким образом, чтобы избежать проблемы обратной причинности. То есть ваши независимые переменные должны, в теоретическом смысле, влиять на зависимую, а не наоборот. Например, уместно предположить, что возраст влияет на вероятность брака. В то же время, скорее всего, не количество детей влияет на вероятность брака, а, вполне возможно, наоборот. Поэтому возраст можно включить в число независимых переменных, а число детей — не желательно.

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.

1.3. Определитесь с тем, будете ли вы предсказывать вероятность брака для мужчин и женщин в рамках единой модели, либо остановитесь лишь на одном из полов. Обоснуйте свой выбор теоретически.

## **Часть 2. Обработка данных.**

2.1. Приведите ваши данные в порядок, удалив пропуски и очистив используемые переменные от неточностей (осуществляется в коде, в работе это описывать не обязательно).

2.2. Рассмотрите доли холостых людей и состоящих в браке. Укажите в таблице описательные статистики для ваших независимых переменных: отдельно для холостых и состоящих в браке. Сделайте выводы о различии в распределении независимых переменных среди индивидов в зависимости от семейного статуса.

2.3. Найдите характеристики среднего и медианного индивидов. Результат представьте в форме таблицы и опишите его словами. Сделайте выводы о различии в характеристиках между средним и медианным индивидами.

2.4. Отобразите при помощи двух гистограмм различия в распределении одной из ваших независимых переменных в зависимости от семейного статуса индивида. Сделайте выводы о различиях в распределении, а также приведите возможное теоретическое обоснование данных различий.

### Часть 3. Линейно-вероятностная модель.

3.1. Оцените линейно-вероятностную модель, предварительно записав регрессионное уравнение. Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R или python).

3.2. Скорректируйте оценку ковариационной матрицы оценок регрессионных коэффициентов при помощи бутстрапа. Объясните причины, по которым необходимо осуществлять данную корректировку и к каким негативным последствиям может привести её отсутствие. Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R, python).

3.3. Протестируйте гипотезы о значимости коэффициентов при помощи перцентильных бутстрапированных доверительных интервалов и сравните полученный вами результат с тем, что был получен при изначальном оценивании (без корректировки). Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R, python). Поясните, как полученные результаты соотносятся с высказанными вами ранее предположениями.

3.4. Запишите формулу, по которой можно рассчитать предельные эффекты в линейно-вероятностной модели для:

- А) Непрерывной переменной, входящей в регрессионное уравнение линейно.
- Б) Дамми-переменной.
- В) Непрерывной переменной, входящей в регрессионное уравнение кубически.
- Г) Непрерывной переменной, имеющей взаимодействие с другой переменной.

3.5. Проинтерпретируйте полученные значения оценок предельных эффектов для каждой независимой переменной. Значения оценок предельных эффектов должны быть представлены в форме таблицы с указанием соответствующих значений оценок стандартных отклонений (стандартных ошибок). Исключение составляют лишь независимые переменные, входящие в регрессионное уравнение нелинейно: для них стандартные ошибки считать не обязательно.

3.6. Посчитайте средний предельный эффект, предельный эффект для среднего индивида, предельный эффект для медианного индивида и предельный эффект для индивида с вашими характеристиками по каждой из независимых переменных, представив результат в форме таблицы.

3.7\*. Посчитайте стандартные ошибки оценок предельных эффектов независимых переменных, входящих в регрессионное уравнение нелинейно (не считая переменные

взаимодействия), для индивида с вашими характеристиками, предварительно пояснив методологию расчета.

3.8. Посчитайте долю верных предсказаний и сопоставьте её с результатом наивного прогноза. Сделайте вывод о предсказательной силе полученной модели.

3.9. Перечислите основные недостатки этой модели. Посчитайте долю предсказанных вероятностей, которые меньше 0 или больше 1. Напишите, можно ли интерпретировать коэффициент детерминации и F-статистику? Если да, то приведите интерпретацию, а если нет, то объясните почему она в данном случае невозможна.

3.10. Оцените вероятность брака для индивида с вашими характеристиками. Если вы оцениваете модель по представителям противоположного пола, то оцените соответствующую вероятность для индивида с вашими характеристиками, но противоположного пола.

#### **Часть 4. Пробит модель.**

4.1. Оцените пробит модель, предварительно записав максимизируемую функцию правдоподобия. Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R или python).

4.2. Проинтерпретируйте полученные значения оценок для каждой независимой переменной. Поясните, как полученные результаты соотносятся с высказанными вами ранее предположениями.

4.3. Запишите формулу, по которой можно рассчитать предельные эффекты в пробит модели для:

- А) Непрерывной переменной, входящей в регрессионное уравнение линейно.
- Б) Дамми-переменной.
- В) Непрерывной переменной, входящей в регрессионное уравнение кубически.
- Г) Непрерывной переменной, имеющей взаимодействие с другой переменной.

4.4. Рассчитайте значения оценок предельных эффектов для каждой из независимых переменных для каждого индивида в выборке (результат достаточно представить в коде, в тексте описывать это не обязательно). Для переменной, входящей в регрессионное уравнение нелинейно, постройте гистограмму, отражающую распределение значений оценок её предельного эффекта. Запишите, какие значения оценок предельных эффектов для данной переменной встречаются чаще: положительные или отрицательные.

4.5\*. Для переменной из предыдущего пункта посчитайте долю значимых на уровне значимости 5% предельных эффектов.

4.6. Посчитайте значения оценок среднего предельного эффекта, предельного эффекта для среднего индивида, предельного эффекта для медианного индивида и предельного эффекта для индивида с вашими характеристиками для каждой из независимых переменных, представив результат в форме таблицы.

4.7\*. Посчитайте стандартные ошибки оценок предельных эффектов независимых переменных, входящих в регрессионное уравнение нелинейно, для индивида с вашими характеристиками, предварительно пояснив методологию расчета.

4.8. Посчитайте долю верных предсказаний и сопоставьте её с результатом наивного прогноза. Сделайте вывод о предсказательной силе полученной модели.

4.9. Посчитайте значение оценки вероятности брака для индивида с вашими характеристиками. Если вы оцениваете модель по представителям противоположного пола, то оцените соответствующую вероятность для индивида с вашими характеристиками, но противоположного пола.

## **Часть 5. Тестирование верности спецификации пробит модели.**

5.1. При помощи LM-теста протестируйте гипотезу о соблюдении допущения о распределении случайных ошибок: при этом необходимо формально записать нулевую и альтернативную гипотезы, статистику теста и её распределение. Запишите, к каким негативным последствиям может привести нарушение данного допущения.

5.2. Предположите, какие переменные могут влиять на дисперсию случайной ошибки. При помощи LR и LM тестов протестируйте гипотезу о гомоскедастичности случайных ошибок: при этом необходимо формально записать нулевую и альтернативную гипотезы, статистику теста и её распределение. Запишите, к каким негативным последствиям может привести нарушение данного допущения. Объясните преимущество LM теста над LR тестом в данном случае.

5.3. Предпримите действия для борьбы с гетероскедастичностью, оценив учитывающую её модель. Сравните знаки и значимости коэффициентов с теми, что были получены при оценивании модели без учета гетероскедастичности.

5.4. Сравните долю верных прогнозов пробит модели с гетероскедастичной случайной ошибкой с долей верных прогнозов, полученных при помощи обычной пробит модели (для тех, кто работает в STATA, это задание со звездочкой \*).

5.5\*. Для модели с гетероскедастичной случайной ошибкой запишите формулу, по которой можно рассчитать предельные эффекты в гетпробит модели для:

- А) Непрерывной переменной, входящей в регрессионное уравнение линейно.
- Б) Дамми-переменной.
- В) Непрерывной переменной, входящей в регрессионное уравнение кубически.
- Г) Непрерывной переменной, имеющей взаимодействие с другой переменной.

Для данной модели представьте в форме таблицы результаты расчетов значений оценок предельных эффектов каждой из независимых переменных для индивида с вашими характеристиками.

5.6\*\*. Для модели с гетероскедастичной случайной ошибкой при помощи дельта-метода посчитайте значение оценки стандартного отклонения оценки предельного эффекта любой независимой переменной для индивида с вашими характеристиками. Поясните методологию расчета.

5.7. При помощи LR теста проверьте, можно ли оценивать совместную модель для людей с различными уровням образования, либо стоит оценить три различные модели для индивидов с 1) высшим, 2) средним или средним специальным 3) иным уровнем образования.

5.8. Выберите любые две, желательно незначимых на уровне значимости 10% переменные, и при помощи LR теста или теста Вальда проверьте, можно ли исключить их из модели: при этом необходимо формально записать нулевую и альтернативную гипотезы, статистику теста и её распределение.

## **Часть 6. Логит модель.**

6.1. Оцените логит модель, предварительно записав максимизируемую функцию правдоподобия. Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R или python).

6.2. Проинтерпретируйте значения оценок изменений в отношениях шанса по каждой независимой переменной.

6.3\*. Включите в модель куб возраста, а затем запишите формулу изменения отношений шансов при изменении возраста на единицу. Рассчитайте значение соответствующей оценки.

6.4\*\*. При помощи дельта-метода рассчитайте стандартную ошибку оценки изменения отношения шансов при изменении возраста на единицу для индивида с вашими характеристиками.

## **Часть 7. Система бинарных уравнений.**

7.1. Оцените систему бинарных уравнений на брак и факт употребления алкоголя, предварительно определив, какая из этих переменных влияет на другую. Результат представьте в форме таблицы (можно, например, использовать выдачу из stata, R или python). Дайте интерпретацию изменениям в оценках коэффициентов по сравнению с обычной пробит моделью.

7.2. Проинтерпретируйте значение оценки коэффициента корреляции между случайными ошибками рассматриваемых уравнений.

7.3\*\*. Для индивида с вашими характеристиками рассчитайте значение оценки вероятности того, что вы вступите в брак при условии, что вы не будете употреблять алкоголь. Затем, посчитайте значение оценки безусловной вероятности вступления в брак.

7.4\*\*\*. При помощи дельта-метода посчитайте стандартные ошибки для оценок из предыдущего пункта. Протестируйте гипотезу об их значимом отличии от нуля.

## **Часть 8. Сравнение моделей.**

8.1. Постройте Рок-Кривые для пробит и логит моделей, а также сравните их предсказательную силу. Какую из этих моделей лучше использовать для получения уровня специфичности, равного 0.8? Запишите любой уровень специфичности, при котором лучше использовать пробит модель, чем логит модель, или поясните, почему такого уровня специфичности не существует, по крайней мере если в качестве критерия использовать точность предсказания внутри выборки, на которой осуществлялось оценивание моделей.

8.2. Сравните пробит, логит и модель с гетероскедастичной случайной ошибкой по критериям AIC и BIC, выбрав лучшую из них. Поясните, почему вы не можете сравнить эти модели по данному критерию с линейно-вероятностной, по крайней мере используя стандартную выдачу из stata или R.

8.3\*\*. Придумайте, как можно сравнить перечисленные выше модели с системой бинарных уравнений по критерию AIC. Осуществите сравнение и выберите лучшую модель.

8.4\*. Сравните оцененные модели по точности предсказания по тестовой выборке (out of sample prediction accuracy).

8.5\*\*\*. Оцените модель бинарного выбора при допущении о том, что случайные ошибки имеют распределение Стьюдента, обеспечив самостоятельную программную



реализацию расчета функции правдоподобия, значений оценок коэффициентов и их стандартных ошибок. Сравните оцененную вами модель с предыдущими по критерию AIC.