

Дорогие студенты, вам предстоит выполнить компьютерное домашнее задание #2 (проект). Представляем вам инструкцию, как приступить к его выполнению.

1. Домашнее задание выполняется в мини-группах по 2–3 человека (если есть необходимость и сильное желание, работу можно выполнять самостоятельно). Мини-группы могут состоять только из студентов, закрепленных за одним семинаристом!

2. Вам необходимо провести эконометрическое исследование для выявления и описания влияния различных факторов на выбранный показатель (*цель – НЕ прогнозирование, а описание влияния!*<sup>1</sup>). Для этого найдите данные по интересующей вас тематике, соответствующие следующим требованиям:

- Данные должны принадлежать кросс-секционному типу
- На основе данных можно построить множественную линейную регрессию (то есть в данных есть одна непрерывная переменная, поведение которой хотя бы предположительно может быть объяснено другими переменными из данных)
- Количество наблюдений: минимум 50
- Количество переменных: минимум 5 независимых переменных с обязательным включением хотя бы одной дамми-переменной

Возможные источники данных: Kaggle, сайт Росстата, сайты Всемирного банка, ОЭСР, других международных организаций. Также, вы можете использовать данные Вашей курсовой работы или вручную собрать данные с любого сайта (можно использовать программу-парсер).

*Пример: исследование зависимости цены мобильного телефона от его характеристик. Данные получены путем парсинга с сайта магазина «Эльдорадо».*

*Переменные:*

- *зависимая переменная: цена определенной модели телефона*
- *независимые переменные: диагональ экрана, встроенная память, количество камер, разрешение основной камеры, количество ядер процессора; наличие беспроводной зарядки (дамми: 1 если есть, 0 - иначе), возможность использования более одной сим-карты (дамми: 1 если есть два и более слота для сим-карт, 0 - иначе)*

3. После того, как вы выбрали данные, **ссылку** на них (а также информацию о выбранном временном периоде и переменных для работы, ну и о своей команде :) ) необходимо внести в гугл-таблицу

([https://docs.google.com/spreadsheets/d/13kbpO6Vo1yscfdP0bYOrujLmnOHA2\\_Bzh281ob3kqE/edit?usp=sharing](https://docs.google.com/spreadsheets/d/13kbpO6Vo1yscfdP0bYOrujLmnOHA2_Bzh281ob3kqE/edit?usp=sharing)).

Перед тем, как вы добавите свои данные в таблицу, пожалуйста, **убедитесь, что такие данные не используют ваши сокурсники**, иначе оценка за работу обеих групп, использующих одинаковые данные, будет 0.

**Дедлайн по внесению информации о группах, данных и переменных – 1 декабря 23:59**

---

<sup>1</sup> **Пример ситуации, когда целью является прогнозирование, а не описание влияния:**

Предположим, что в качестве зависимой переменной в модели используется возраст краба, а в качестве объясняющих - размеры панциря и конечностей. Ожидаемо, что размеры краба скорее всего коррелируют с его возрастом, поэтому в модели получится большой  $R^2$  и значимые коэффициенты. Однако в подобной задаче интерпретация коэффициентов при переменных, связанных с размерами краба, не является разумной, здесь нет механизма влияния (скорее с увеличением возраста краба панцирь и конечности становятся больше)! При этом приведённую модель вполне можно использовать для прогнозирования возраста краба, что не является целью вашей работы.

4. Ура! Вы выбрали данные, вписали информацию о них в таблицу, и теперь можно приступить непосредственно к эконометрическому исследованию (см. ниже).

**Дедлайн** по сдаче домашнего задания – **13 декабря 23:59**. Начиная с 00:00 14 декабря работа штрафует на 1 балл. За каждый последующий час просрочки еще по минус 1 баллу.

Один участник команды присылает своему семинаристу на почту (каждый семинарист сообщит в чат группы, на какую почту присылать работу) три отдельных файла:

**1) Текст работы. Формат ПДФ!!!** Примерный план работы представлен ниже. Помимо этого, в текст работы необходимо включить основные графики и таблицы. Текст должен быть законченным и читаться самостоятельно, без обращения к вашему коду, таблицам с данными и другим внешним источникам. **Текст работы может быть сдан только в формате pdf!!!** Текст, сданный в любом другом формате, проверяться не будет и будет оценён в 0 баллов.

**2) Код** (в любом софте, можно несколько скриптов из разных софтов / на разных языках программирования. Если работа выполнена в Excel – файл с расчётами)

**3) Данные**

Внимание! Ниже представлен ПРИМЕРНЫЙ комплексный план работы. Это отнюдь не точная схема, которой необходимо пользоваться. Вы можете добавлять/опускать пункты данного плана – если это осмысленно, то оценка за это снижаться не будет. **Исключение составляют красные комментарии – они обязательные.** Главное, чтобы ваше исследование носило законченный характер, и чтобы проверяющим было понятно, на основе каких манипуляций получены те или иные выводы. Творчество приветствуется!

#### **Требования:**

- Не забывать про то, что красные комментарии обязательные!
- В работе должно быть не менее 3 ссылок на научные статьи (не менее 2 ссылок, если вы делаете работу самостоятельно, без команды). Подробнее про то, где и зачем ссылаться на статьи, см. ниже.
- Объем не важен, важно качество работы.

### 1) Титульный лист – **обязательный элемент!**

Формат титульного листа свободный, но он должен содержать два обязательных элемента: **название работы** и **полный список всех участников команды**. Оценка за проект выставляется только тем студентам, которые указаны в списке на титульном листе работы.

### 2) Введение

Сформулируйте цель и задачи исследования, а также поясните его актуальность. В данной части работы очень рекомендуется ссылаться на научную литературу.

*Где-то в тексте работы (мы рекомендуем во введении или описании модели) обязательно необходимо описать ваши данные: откуда они (источник, ссылка), кем/как/когда собирались, количество наблюдений.*

### 3) Экономическая модель

В данной части работы очень рекомендуется ссылаться на научную литературу, использовать свои знания экономической теории или, как минимум, логику.

- a. Список объясняющих переменных – **обязательно с единицами измерения!**
- b. Почему был выбран именно этот набор объясняющих переменных?
- c. Предположительно, каким образом независимые переменные будут оказывать влияние на зависимую?
- d. Какие содержательные гипотезы могут быть протестированы в ходе исследования?
  - *Содержательные гипотезы должны выходить за рамки значимости / незначимости. Гипотезы должны иметь под собой обоснование (например, «Мы хотим проверить закон Энгеля для  $d/x$  – при увеличении доходов должна сокращаться доля расходов на потребление, поэтому коэффициент при переменной доходов должен быть отрицательным»).*
  - *Не нужно формулировать гипотезы для каждой из объясняющих переменных.*
  - *Достаточно 2–3 содержательных гипотез.*

### 4) Предварительный анализ данных

Анализ выборки на наличие статистических выбросов, интерпретация описательных статистик, корреляционных матриц, визуальный анализ.

### 5) Оценка модели

Представьте результаты оценки модели (моделей) и прокомментируйте основные результаты. *Внимание: вставлять таблицы с результатами оценивания в виде скриншотов или скопированного текста – дурной тон! Необходимо воспользоваться специальными библиотеками, позволяющими выгружать результаты оценивания в виде красивых таблиц в word/tex (например, в R это stargazer).*

### 6) Выводы

НЕ нужно просто перечислять полученные значения коэффициентов, **необходимо содержательно их проинтерпретировать**, то есть пояснить, что полученные значения значат экономически, есть ли в них смысл, соответствуют ли они теории по данной теме и т.д. Также в данном разделе **необходимо вернуться к гипотезам** и сделать вывод относительно сформулированных в начале исследования утверждений. Если гипотезы не подтвердились (это нормально, не пугайтесь этого!), то попробуйте порассуждать, почему – возможно, вы найдете идеи в научной литературе.