

## Демонстрационная версия проверочной работы 1

**Задание 1.** Ниже представлены результаты оценивания линейной регрессионной модели процента детей, получающих образование, (*eduger*), на социально-экономические показатели. *polityIV* — индекс демократии Polity IV (от -10 до 10); *gini\_8090* — средний индекс имущественного неравенства в 1980 и 1990 г. (от 0 до 100); *ssw* — консолидированные бюджетные расходы на социальное обеспечение (% ВВП); *prot80* — процент протестантов от численности населения в 1980 г.

```
            coef    std. error     t   Pr>|t|   [0.025; 0.975]
(Intercept) 59.0982      9.2578
polityIV     1.0462      0.4522
gini_8090    0.2556      0.1952
ssw          1.4918      0.3066
prot80       0.1016      0.0576
---
```

```
              ANOVA
-----
           sum_sq      df   mean_sq      f   PR(>F)
-----
Regression
Residual    5665.644
Total        14178.832      57
```

1. Постройте 99%-ый доверительный интервал для коэффициента при *ssw*, используя нормальную аппроксимацию для расчета процентной точки (однако в ответах запишите обязательно количество степеней свободы по распределению Стьюдента), и проинтерпретируйте его
2. Проинтерпретируйте оценку коэффициента при предикторе *ssw*
3. Рассчитайте коэффициент детерминации и проинтерпретируйте полученное значение
4. Проверьте гипотезу, что регрессия на константу не хуже модели с предикторами, на фиксированном уровне значимости 0.01. Запишите значение статистики и ее промежуточные расчеты, а также выберите необходимую критическую точку – квантиль. Сделайте вывод.
  - (a) квантиль хи-квадрат распределения уровня 0.99,  $df = 57$ : **84.733**
  - (b) квантиль хи-квадрат распределения уровня 0.01,  $df = 57$ : **35.131**
  - (c) квантиль распределения Фишера уровня 0.99,  $df_1 = 5$ ,  $df_2 = 53$ : **3.384**
  - (d) квантиль распределения Фишера уровня 0.99,  $df_1 = 4$ ,  $df_2 = 53$ : **3.695**
  - (e) квантиль распределения Фишера уровня 0.01,  $df_1 = 5$ ,  $df_2 = 53$ : **0.108**
  - (f) квантиль распределения Фишера уровня 0.01,  $df_1 = 4$ ,  $df_2 = 53$ : **0.073**

**Задание 2.** По результатам анализа данных, посвящённых рынку труда в США, с использованием подхода разность разностей (*difference-in-difference*) была получена следующая таблица:

	До	После	Разность
Группа воздействия	20.44	21.03	?
Контрольная группа	23.33	21.17	?
DiD			?

Исследователи сравнивали средний уровень занятости в ресторанах быстрого питания до повышения минимальной заработной платы в штате и после. В качестве группы воздействия был выбран штат Нью-Джерси, в котором произошло повышение минимальной заработной платы, а в качестве контрольной группы был выбран штат Пенсильвания, так как его повышение заработной платы не затронуло.

**Восстановите оценки коэффициентов в классической модели разность разностей (DiD) без контрольных переменных и проинтерпретируйте коэффициент при переменной взаимодействия.**

**Задание 3.**

Ниже представлены оценки регрессионной модели. Зависимая переменная – успеваемость школьника по литературе (рассчитан средний балл по предмету за год). В качестве предикторов используется показатель количества книг дома у школьника (*Books*), наличие доступа к интернету (*Internet* – дамми-переменная, при этом 1 соответствует наличию доступа, 0 – доступ отсутствует) и переменная взаимодействия.

Grade (literature)	
Books	1.06** (0.424)
Internet	0.72*** (0.176)
Books × Internet	-0.61*** (0.141)
Intercept	3.5*** (0.367)

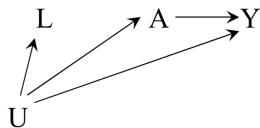
Standard errors are given in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

1. Проинтерпретируйте оценку коэффициента при переменной *Internet*
2. Рассчитайте значение предельного эффекта *Books* в случае, если доступ к интернету есть
3. Проинтерпретируйте оценку коэффициента при переменной взаимодействия

**Задание 4.** Определите, стоит ли включать ту или иную обозначенную переменную как контрольную в модели. Свой ответ обоснуйте на основе «критерия черного хода» (backdoor criterion). В качестве ключевого предиктора в графе используется A, в качестве отклика – Y.

Рис. 1: U – латентная переменная. Стоит ли контролировать L?



**Задание 5.** Ответьте на вопросы ниже:

1. Приведите пример строгой мультиколлинеарности. Каковы последствия строгой мультиколлинеарности? Свой ответ поясните
2. Прочтайте нижеприведенный отрывок из статьи и продолжите его, аргументировав позицию авторов:

Much of the concern about multicollinearity arises when the analyst observes that the coefficients from a linear-additive model change when an interaction term is introduced. In the linear-additive world, the sensitivity of results to the inclusion of an additional variable is often taken as a sign of multicollinearity. However, this need not be the case with interaction models.

**Задание 6.** Ниже в таблице представлены значения переменных: Y и X.

Y	-5	0	-1	-2	2
X	3	0	0	-4	1

Получите оценки коэффициентов в регрессии Y на X с помощью общей векторно-матричной формулы получения оценок коэффициентов. Представьте промежуточные расчеты, выпишите полученный вектор оценок коэффициентов и запишите спецификацию модели, подставив эти оценки в уравнение.

**Задание 7.** Выберите ВСЕ верные утверждения из списка ниже. Если верные утверждения отсутствуют, обязательно в ответе напишите "НЕТ":

1. По построению линейной регрессионной модели верно, что сумма квадратов остатков равна 0
2. В результате сильной мультиколлинеарности стандартные ошибки оценок коэффициентов становятся, как правило, ниже
3. Согласно допущениям Гаусса-Маркова, распределение ошибок в линейной регрессионной модели является нормальным
4. Модель DiD основывается на допущении о параллельности трендов