
**УНИВЕРСИАДА ПО ЭКОНОМЕТРИКЕ
ЭКОНОМИЧЕСКОГО ФАКУЛЬТЕТА МГУ ИМЕНИ М. В. ЛОМОНОСОВА
ПРИ ПОДДЕРЖКЕ ДЕПАРТАМЕНТА ЭКОНОМИЧЕСКОЙ ПОЛИТИКИ И
РАЗВИТИЯ ГОРОДА МОСКВЫ
18 АПРЕЛЯ 2026**

**100 БАЛЛОВ, КАЖДАЯ ЗАДАЧА ПО 25 БАЛЛОВ
Удачи!**

Задача 1

Иван Сергеевич работает в Аналитическом центре при Департаменте экономической политики и развития города Москвы и изучает, как меры государственной поддержки влияют на рентабельность активов (y) малых и средних предприятий (МСП) в Москве. Иван Сергеевич располагает данными о нескольких показателях МСП (матрица X), в том числе получении компаниями финансовой поддержки в виде субсидий и грантов, и хочет оценить линейную регрессию y в зависимости от X . У Ивана Сергеевича проходят стажировку два студента Илья и Эвелина, и Иван Сергеевич решает разделить выборку на малые и средние предприятия соответственно: X_1, y_1 – малые (анализирует Илья), и X_2, y_2 – средние (анализирует Эвелина). Отметим, что две выборки не пересекаются, статистически независимы, состоят из n_1 и n_2 наблюдений соответственно. Считаем, что выполнены все предпосылки теоремы Гаусса-Маркова.

Справочно: табличные значения стандартного нормального распределения: 1.69 для уровня значимости 10%, 1.96 для 5% и 2.58 для 1%. Размер выборки $n = n_1 + n_2$ достаточно велик: на 1 марта 2026 года число МСП [составляло](#) 926,3 тыс.

Оба стажёра оценили линейные МНК-регрессии с обычными стандартными ошибками (каждый – на своей подвыборке) и сообщили Ивану Сергеевичу в качестве результатов анализа получившиеся в их моделях суммы квадратов остатков (ESS), вектор оценок $\hat{\beta}$ и оцененные ковариационные матрицы коэффициентов $\hat{V}_{\hat{\beta}}$.

К сожалению, при этом сами данные X_1, y_1 и X_2, y_2 для Ивана Сергеевича временно недоступны.

1. В результате Илья получил оценку коэффициента МНК-регрессии для переменной «получение финансовой поддержки» для малых предприятий равную +1.65. Эвелина же для средних предприятий +0,55. Что можно сказать о статистической значимости влияния финансовой поддержки для малых предприятий? для средних предприятий? Дайте пояснения. Оценки ковариационных матриц оценок коэффициентов представлены в таблицах:

Малые предприятия						
	Факт оказания финансовой поддержки	Стоимость совокупных активов	Численность работников	Доля нематериальных активов в активах	Уровень долга	Коэффициент текущей ликвидности
Факт оказания финансовой поддержки	0,090	0,150	0,220	0,110	0,170	0,550
Стоимость совокупных активов	0,150	0,008	0,330	0,147	0,720	0,316
Численность работников	0,220	0,330	0,014	0,230	0,900	0,320
Доля нематериальных активов в активах	0,110	0,147	0,230	0,002	0,320	0,532
Уровень долга	0,170	0,720	0,900	0,320	0,000	0,134
Коэффициент текущей ликвидности	0,550	0,316	0,320	0,532	0,134	0,001

Средние предприятия						
	Факт оказания финансовой поддержки	Стоимость совокупных активов	Численность работников	Доля нематериальных активов в активах	Уровень долга	Коэффициент текущей ликвидности
Факт оказания финансовой поддержки	0,063	0,100	0,210	0,150	0,110	0,550
Стоимость совокупных активов	0,100	0,005	0,311	0,151	0,690	0,330
Численность работников	0,210	0,311	0,009	0,210	0,890	0,423
Доля нематериальных активов в активах	0,150	0,151	0,210	0,001	0,315	0,511
Уровень долга	0,110	0,690	0,890	0,315	0,000	0,132
Коэффициент текущей ликвидности	0,550	0,330	0,423	0,511	0,132	0,001

Все следующие вопросы имеют теоретический характер и не предполагают использование таблиц из пункта 1.

2. Может ли Иван Сергеевич восстановить коэффициенты линейной регрессии рентабельности активов (y) в зависимости от характеристик МСП (X), как если бы модель оценивалась на данных обо всех МСП Москвы без деления на малые и средние? Обоснуйте Ваш ответ выкладками.
3. Может ли Иван Сергеевич проверить гипотезу о (не)значимости коэффициентов и делать тесты на линейное ограничение, как если бы модель оценивалась на данных обо всех МСП Москвы без деления на малые и средние? Обоснуйте Ваш ответ выкладками.

Предположим теперь, что вместо оценок ковариационных матриц $\hat{\nabla}_{\hat{\beta}}$ стажёры сообщили Ивану Сергеевичу только регрессионные таблицы с коэффициентами, их t -статистиками и соответствующими суммами квадратов остатков (ESS).

4. Может ли Иван Сергеевич теперь проверить гипотезу о равенстве оценок регрессий для малых и средних предприятий? Обоснуйте Ваш ответ выкладками.
5. Иван Сергеевич собирается конструировать интегральную оценку коэффициентов в виде линейной комбинации оценок, сообщенных ему стажёрами. Из каких соображений и как ему следует выбирать коэффициенты в этой комбинации?

Задача 2

Предположим, что для каждого работника наблюдаются следующие случайные переменные: 1) Зарботная плата Y и 2) бинарная переменная D , где $D = 1$ для работника с высшим образованием и $D = 0$ в противном случае. На протяжении задачи предполагается, что для зафиксированного распределения (Y, D) имеем $\mathbb{P}[D = 1] \in (0, 1)$, а математическое ожидание и дисперсия Y конечные.

1. Для модели

$$Y = \beta_0 + \beta_1 D + \varepsilon, \quad \mathbb{E}[\varepsilon | D] = 0, \beta_0, \beta_1 \in \mathbb{R} \quad (1)$$

выразите β_0 и β_1 через моменты условного распределения $Y | D$.

Предположим, что также определены случайные переменные $Y(1)$ и $Y(0)$, которые описывают заработную плату, которую данный работник мог бы получать, если бы имел образование $D = 1$ или $D = 0$ соответственно. Таким образом, наблюдаемая заработная плата удовлетворяет уравнению

$$Y = DY(1) + (1 - D)Y(0). \quad (2)$$

Заметьте, что для работника с $D = 1$ наблюдается только $Y(1) = Y$, но не $Y(0)$, а для работника с $D = 0$ – только $Y(0) = Y$, но не $Y(1)$.

2. Докажите, что β_1 , в общем случае, не равна τ , определенному как

$$\tau = \mathbb{E}[Y(1) - Y(0)],$$

и объясните содержательно, почему эмпирически вероятно, что $\beta_1 \geq \tau$. Можете использовать обозначение $p = P(D = 1) = \mathbb{E}[D]$.

Предположим, что $Y, Y(1), Y(0) \in [0, 1]$ (нормализованы). Кроме того, исследователь наблюдает случайную выборку $(Y_i, D_i)_{i=1}^n$, где наблюдения имеют то же распределение, что и (Y, D) .

3. При заданном распределении (Y, D) докажите, что τ ограничена отрезком $\underline{\beta} \leq \tau \leq \bar{\beta}$, где

$$\underline{\beta} = \mathbb{E}[(2D - 1)Y - D], \quad \bar{\beta} = \mathbb{E}[(2D - 1)Y + (1 - D)].$$

Используйте тот факт, что $Y(1), Y(0) \in [0, 1]$. И что при заданном распределении вероятность $p = P(D = 1)$ фиксированная.

4. На основании имеющейся случайной выборки определите 95% односторонний доверительный интервал вида $I_n = [c_n, \infty)$, где $c_n = c_n((Y_i, D_i)_{i=1}^n)$, такой что

$$\lim_{n \rightarrow \infty} \mathbb{P}[\underline{\beta} \geq c_n] = 0.95.$$

(Заметьте, что, если $Z \sim N(0, 1)$, то $\mathbb{P}[Z \leq 1.645] \approx 0.95$.)

Допустим, что дополнительно известно, что $\mathbb{E}[Y(1) | D = 1] \geq \mathbb{E}[Y(1) | D = 0]$ и $\mathbb{E}[Y(0) | D = 1] \geq \mathbb{E}[Y(0) | D = 0]$.

5. При заданном распределении (Y, D) докажите, что τ ограничена отрезком $\underline{\beta} \leq \tau \leq \beta_1$. (Используйте тот факт, что $Y(0), Y(1) \in [0, 1]$ и $\mathbb{E}[Y(1) | D = 1] \geq \mathbb{E}[Y(1) | D = 0]$ и $\mathbb{E}[Y(0) | D = 1] \geq \mathbb{E}[Y(0) | D = 0]$.)
6. Пусть на основании имеющейся случайной выборки определена МНК-оценка $\hat{\beta}_{1,n}$ для β_1 , и имеется $\hat{\sigma}_n = \hat{\sigma}_n((Y_i, D_i)_{i=1}^n)$, такая, что, если для $\sigma > 0$,

$$\sqrt{n}(\hat{\beta}_{1,n} - \beta_1)/\sigma \xrightarrow{d} N(0, 1),$$

то $\hat{\sigma}_n \xrightarrow{p} \sigma$. Определите 95% односторонний доверительный интервал $I_n^* = (-\infty, c_n^*]$, где $c_n^* = c_n^*((Y_i, D_i)_{i=1}^n)$, такой что

$$\lim_{n \rightarrow \infty} \mathbb{P}[\beta_1 \leq c_n^*] = 0.95.$$

Задача 3

Знакомый Вам по заданию отборочного тура Сергей Т. исследует факторы заработной платы. Есть данные о 540 работниках (270 мужчин и 270 женщин): $\ln(\text{EARNINGS})$ — логарифм часового заработка в долларах США, S — число полных лет обучения, EXP — стаж работы после окончания учебы, FEMALE — пол респондента (0 — для мужчин, 1 — для женщин). Табличные значения стандартного нормального распределения: 1.69 для уровня значимости 10%, 1.96 для 5% и 2.58 для 1%. Ниже представлены результаты оценки модели №1:

Модель 1: МНК, использованы наблюдения 1–540
 Зависимая переменная: $\ln(\text{EARNINGS})$
 Робастные оценки стандартных ошибок (с поправкой на гетероскедастичность), вариант HC1

	Коэффициент	Ст. ошибка	t-статистика	p-значение
константа	0,7863	0,2204	3,567	0,0004
EXP	A	0,0081	3,637	0,0003
S	0,1200	0,0093	12,78	<0,0001
FEMALE	-0,2258	0,1736	-1,301	0,1938
EXP*FEMALE	-0,0258	B	C	0,0100

Среднее значение зависимой переменной	2,79	Стандартное отклонение зависимой переменной	0,588
Сумма кв. остатков	126,24	Стандартная ошибка модели	0,485757
R-квадрат	D	Исправленный R-квадрат	E

- (10 баллов) По имеющимся данным заполните пропуски A, B, C, D, E, если это возможно. Обоснуйте Ваш ответ.
- (2 балла) Дайте содержательную количественную интерпретацию коэффициента при переменной EXP в модели №1.
- (4 балла) Для модели 1 p-значение теста Уайта 0.00218. Также Сергей провёл тест Голдфелда-Квандта с упорядочиванием по переменной S, табличное значение равно 1,2832. Каким может быть расчётное значение статистики из теста Голдфелда-Квандта, чтобы в результатах двух тестов не было противоречий?
- (5 баллов) Сергей решил проверить, есть ли на данном рынке дискриминация женщин/мужчин. Для этого он оценил еще одно уравнение: $\ln(\text{EARNINGS})_i = \alpha_1 + \alpha_2 \text{EXP}_i + \alpha_3 S_i + \varepsilon_i$ В новом уравнении $R^2 = 0,25$. Используя данную вам информацию дайте ответ, есть ли на данном рынке дискриминация женщин/мужчин. Не забудьте сформулировать тестируемую и альтернативную гипотезу и описать тест. Табличное значение статистики равно 3.01.
- (4 балла) Для модели №2 Сергей рассчитал коэффициенты VIF для определения наличия мультиколлинеарности. Для переменной S $\text{VIF} = 1.050$. Чему равен VIF для переменной EXP? Обоснуйте Ваш ответ. Сделайте вывод из полученных результатов.

Задача 4

Предположим, у вас есть данные о доходах людей из одного поколения, а также данные о доходах их родителей (неотрицательное число, ограниченное снизу нулём).

Ваша цель — оценить, в какой степени положение детей в распределении доходов коррелирует с положением родителей. Это соотношение — один из способов описания межпоколенческой мобильности. Предположим следующее:

- доходы (X, Y) являются независимыми и одинаково распределёнными (iid) из совместного распределения с непрерывной функцией совместного распределения $F(x, y)$, с маргинальными (частными) функциями распределения, обозначенными F_X, F_Y ;
 - каждому X присваивается перцентильный ранг $R_i = 100 * F_X(X)$;
 - соответственно, каждому Y присваивается перцентильный ранг $P_i = 100 * F_Y(Y)$;
 - оценивается линейная модель методом наименьших квадратов (МНК): $R_i = \alpha + \beta P_i + \varepsilon_i$.
1. Покажите, что $P_i/100$ имеет маргинальное (частное) равномерное распределение $\text{Uni}(0,1)$ (то же самое будет выполняться для $R_i/100$).
 2. Используя результат предыдущего пункта, покажите, что коэффициент МНК равен $\rho = \text{corr}(R_i, P_i)$.
 3. Докажите, что для определения свободного члена α достаточно знать коэффициент наклона β (подсказка: используйте формулу МНК для α). Каков ожидаемый ранг дохода ребёнка из домохозяйства с медианным доходом согласно линейной модели?
 4. Предложите дизайн эконометрического исследования для оценки вероятности события, что ребёнок будет иметь более высокий ранг в распределении доходов своего поколения, чем имели его родители в их поколении. Гипотеза: «Государственные расходы на образование ослабляют связь между образованием родителей и относительным ростом дохода детей (более высоким рангом), особенно на нижней части распределения доходов родителей». Ваш дизайн должен включать:
 - спецификацию эконометрической модели с расшифровкой используемых статистических показателей,
 - обсуждение возможной эндогенности,
 - предложение по идентификационной стратегии,
 - перечень необходимых статистических тестов и критериев (с пояснением, зачем используется каждый из них).