

**Домашнее задание №1.**

Срок сдачи: 23:59, 20 октября 2023 г.

Домашнее задание должно быть аккуратно оформлено, к каждой задаче необходимо привести подробное решение. Сканы решения должны быть объединены в единый pdf-документ и загружены в систему Smart LMS в соответствии с указанным дедлайном. Домашние задания, не удовлетворяющие требованиям к оформлению, выполненные не самостоятельно или сданные позже срока, получают 0 баллов.

**Задача 1. (20 баллов)** Рассмотрим классическую линейную модель множественной регрессии (КЛММР) вида  $Y_i = \alpha + \beta X_i + \varepsilon_i$ , удовлетворяющую предпосылкам теоремы Гаусса-Маркова. Обычные МНК-оценки  $\hat{\alpha}$  и  $\hat{\beta}$  являются несмещенными для соответствующих параметров модели. Пусть  $\tilde{\beta}$  — оценка  $\beta$ , полученная в предположении, что свободный член  $\alpha$  равен нулю.

- (а) **(5 баллов)** Выразите  $E(\tilde{\beta})$  через  $X_i$ ,  $\alpha$  и  $\beta$ . Убедитесь, что  $\tilde{\beta}$  является несмещенной оценкой для  $\beta$ , когда свободный член  $\alpha$  равен нулю. Существуют ли другие случаи, когда  $\tilde{\beta}$  будет являться несмещенной оценкой параметра  $\beta$ ?
- (б) **(5 баллов)** Найдите дисперсию  $\tilde{\beta}$ . [Подсказка: дисперсия не зависит от  $\alpha$ ].
- (в) **(5 баллов)** Покажите, что  $Var(\tilde{\beta}) \leq Var(\hat{\beta})$ . [Подсказка: Для любой выборки справедливо выполнение неравенства  $\sum_{i=1}^n X_i^2 \geq \sum_{i=1}^n (X_i - \bar{X})^2$  со строгим знаком тогда и только тогда, когда  $\bar{X} \neq 0$ .]
- (г) **(5 баллов)** Прокомментируйте компромисс между смещением оценки и дисперсией при выборе между  $\tilde{\beta}$  и  $\hat{\beta}$ .

**Задача 2. (20 баллов)** Пусть Вы хотите оценить с помощью МНК регрессию  $y = X\beta + \varepsilon$  (всего  $k$  регрессоров). При этом константа не включена в матрицу регрессоров  $X$ . Однако эконометрический пакет, которым Вы располагаете, автоматически добавляет константу к регрессорам.

Обозначим

$$y^* = \begin{bmatrix} y \\ -y \end{bmatrix}, X^* = \begin{bmatrix} l & X \\ l & -X \end{bmatrix},$$

где  $l$  —  $(n \times 1)$ -вектор, состоящий из единиц.

- (а) **(10 баллов)** Покажите, что регрессия  $y^*$  на  $X^*$  дает желаемый результат.
- (б) **(10 баллов)** Покажите, что полученные в предыдущем пункте оценки стандартных ошибок оценок коэффициентов, должны быть скорректированы на некоторый множитель, и укажите его.

**Задача 3 (30 баллов).** Ассистент Никита выставляет оценку за очередную работу по дополнительным главам макроэкономики. У Никиты не выходит прочесть эту работу из-за неразборчивого почерка, поэтому вместо этого он оценивает четыре регрессии с разными наборами детерминант, принимая балл за работу в качестве зависимой переменной:

1. Hard и Easy
2. Respect
3. Respect и Delta
4. Easy и Respect

В качестве наблюдений выступают проверенные ранее работы. Обозначения для регрессоров следующие: *Hard* — количество слов, написанное при ответе на самый сложный вопрос, *Easy* — количество слов, написанное при ответе на самый легкий вопрос, *Respect* — их сумма, *Delta* — их разность. Сложность вопросов воспринимается Никитой всегда одинаково, независимо от конкретной работы. Кроме того, Никита всегда оценивает регрессии с константой, но не любит выводить оценки констант в таблице с результатами.

В качестве балла за работу с неразборчивым почерком Никита всегда выставляет минимальное из предсказанных значений в четырёх регрессиях, но не более 8 баллов.

- (а) **(26 баллов)** К сожалению, Никита случайно стёр некоторые числа из финальной таблицы с результатами оценки регрессий, но он уверен, что их все можно восстановить. Помогите Никите сделать это, либо объясните ему, что он не прав.

Таблица 1 представлена ниже. В скобках указаны стандартные ошибки. На месте букв должны стоять числа.

Table 1. Результаты оценки регрессий

	(1)	(2)	(3)	(4)
Hard	0.03 (0.02)	-	-	-
Easy	0.04 (0.02)	-	-	<i>H</i> ( <i>I</i> )
Respect	-	0.036 (0.01)	<i>B</i> ( <i>C</i> )	<i>K</i> ( <i>L</i> )
Delta	-	-	<i>D</i> ( <i>E</i> )	-
$R^2$	0.3	<i>A</i>	<i>F</i>	<i>M</i>
$ESS$	150	170	<i>G</i>	<i>N</i>

- (б) (4 балла) Рассчитайте, какую оценку Никита поставит студенту с неразборчивым почерком, если этот студент написал ровно по 100 слов в каждом из вопросов работы, при этом оценка константы совпадает во всех регрессиях и равна 0.5.

**Задача 4. (10 баллов)** Пусть выборка состоит из  $n$  наблюдений с матрицей объектов-признаков  $\mathbf{X}_n$  и вектором значений зависимой переменной  $y_n$ . Молодой эконометрист Вася получил вектор МНК-оценок:  $\mathbf{b}_n = (\mathbf{X}_n' \mathbf{X}_n)^{-1} \mathbf{X}_n' y_n$ . Однако затем стало известно новое наблюдение  $(\mathbf{x}_s, y_s)$ . Покажите, что МНК-оценка, рассчитанная с использованием дополнительного наблюдения, будет иметь вид:

$$\mathbf{b}_{n,s} = \mathbf{b}_n + \frac{1}{1 + \mathbf{x}_s' (\mathbf{X}_n' \mathbf{X}_n)^{-1} \mathbf{x}_s} (\mathbf{X}_n' \mathbf{X}_n)^{-1} \mathbf{x}_s (y_s - \mathbf{x}_s' \mathbf{b}_n)$$

**Подсказка:** воспользуйтесь следующей формулой:

$$(\mathbf{A} \pm \mathbf{b}\mathbf{b}')^{-1} = \mathbf{A}^{-1} \mp \frac{1}{1 \pm \mathbf{b}'\mathbf{A}^{-1}\mathbf{b}} \cdot \mathbf{A}^{-1}\mathbf{b}\mathbf{b}'\mathbf{A}^{-1},$$

где  $\mathbf{A}$  — симметричная обратимая матрица размера  $n \times n$ ,  $\mathbf{b}$  — вектор-столбец размера  $n \times 1$ .

**Задача 5. (20 баллов)** В данном задании вам предстоит работа с симулированными (сгенерированными) данными. В качестве инструмента для выполнения рекомендуется использовать R или Python. К работе, соответственно, нужно будет приложить R-файл или Jupyter Notebook, в котором помимо расчетов будут комментарии к коду и развернутые ответы на вопросы.

- (а) (2 балла) Рассмотрим модель линейной регрессии, имеющую следующую спецификацию:

$$Y_i = \beta_0 + \beta_1 X1_i + \beta_2 X2_i + \varepsilon_i.$$

Самостоятельно задайте значения для  $\beta_0, \beta_1$  и  $\beta_2$ . Пусть для начала имеется выборка объемом 100 наблюдений. Сгенерируйте непрерывные признаки  $X1$  и  $X2$ , укажите использованные распределения. Сгенерируйте случайную ошибку  $\varepsilon$  из нормального распределения с нулевым математическим ожиданием и дисперсией, корень из которой равен номеру первой буквы вашей фамилии в алфавите. Запишите получившуюся спецификацию.

- (б) (6 баллов) Оцените значения коэффициентов тремя способами:

1. минимизируйте сумму квадратов остатков и аналитически найдите формулы оценок;
2. используйте известную с лекций и семинаров матричную формулу;
3. рассчитайте оценки численно. (В R это можно сделать с помощью функции `optim`, в Python — с помощью `optimize` из библиотеки `scipy`).

Убедитесь, что значения оценок во всех трех методах совпадают.

- (в) **(4 балла)** Сгенерируйте новые выборки с теми же распределениями и значениями коэффициентов размером 500 и 1000 наблюдений. Посчитайте оценки коэффициентов одним из способов, использованных в предыдущем задании. Сравните их с теоретическими, заданными вами, значениями. Какое свойство МНК-оценок вы можете наблюдать?
- (г) **(2 балла)** Для исходной выборки размером 100 наблюдений рассчитайте оценку ковариационной матрицы коэффициентов. Сравните ее с теоретической ковариационной матрицей.
- (д) **(3 балла)** Для  $\beta_1$  или  $\beta_2$  на уровне значимости 5% проверьте гипотезу о равенстве значения коэффициента числу  $k \neq 0$ , где  $k$  можно выбрать произвольным, указав выбранное значение. Укажите значение тестовой статистики, ее распределение при верной нулевой гипотезе и p-value.
- (е) **(3 балла)** К имеющимся регрессорам добавьте бинарный признак, принимающий два значения. На уровне значимости 5% проверьте гипотезу о неоднородности выборки по этому признаку с помощью теста Чоу. Укажите значение тестовой статистики, ее распределение при верной нулевой гипотезе и  $p - value$ .