Двенадцатая универсиада по эконометрике
Экономический факультет МГУ имени М. В. Ломоносова
8 апреля 2023

У Д А Ч И !

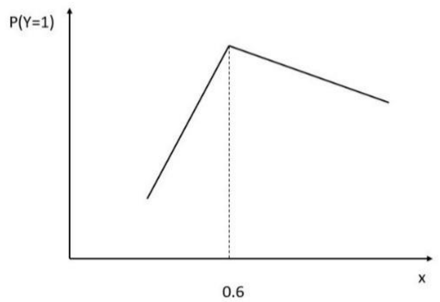
Задача 1. О градообразующих предприятиях (25 баллов)

Необходимо понять, как характеристики моногорода влияют на вероятность того, что в нём будет создана Территория опережающего развития (ТОР). Для этого на выборке из 200 городов оценивается пробит-модель, в которой вероятность создания ТОР в моногороде зависит от доли занятых на градообразующем предприятии среди всех работающих (X), квадрата этой доли (X2) и некоторых контрольных переменных.
В таблице приведены частичные результаты оценивания коэффициентов модели:

Регрессор	Оценка коэффициента
константа	0.360
X	0.187
X2	-0.17
Псевдо-R2	0.218

Фрагмент оцененной ковариационной матрицы оценок коэффициентов:

0.151	0.063	0.001
0.063	0.075	0.024
0.001	0.024	0.051



1. Квадратичная форма зависимости выбрана в предположении, что существует доля занятых на градообразующем предприятии среди всех работающих, при которой максимизируется вероятность создания ТОРа в городе. По вышеприведённым данным рассчитайте эту долю.
2. Проверьте гипотезу о том, что эта оптимальная доля равна 0.6. Укажите нулевую и альтернативную гипотезу, расчётную статистику и критерий отвержения нулевой гипотезы, сделайте вывод.
3. Ваш оппонент уверен, что 0.6 - это «переломная точка», но сомневается в правильности выбранной спецификации модели (симметрии относительно вершины параболы). Поэтому он предлагает кусочно-линейную модель с максимумом в точке 0.6 (см. рисунок). Запишите спецификацию, соответствую этому рисунку, с учётом непрерывности кусочно-линейной регрессии.
4. Для построенной Вами модели задайте критерий проверки гипотезы о том, что 0.6 – точка максимума.
5. Допустим, эта гипотеза отвергнута. В этом случае возникает задача поиска точки максимума для непрерывной кусочно-линейной регрессии. Предложите соответствующую процедуру.

Табличные значения стандартного нормального распределения:
1.69 - 10%, 1.96 - 5%, 2.58 – 1%

Задача 2. О пропавших оценках (20 баллов)

Знакомая вам по отборочному туру Маша продолжает исследовать характер цикличности госрасходов и оценивает регрессию $g_{it} = \alpha_0 + \alpha_1 * y_{it} + \varepsilon_{it}$, где g_{it} -- циклическая компонента государственных расходов страны i в году t (отклонения в % от тренда), y_{it} -- циклическая компонента ВВП страны i в году t (отклонения в % от тренда), ε_{it} -- случайный шок. Оценка коэффициента наклона равна 0.675 и является статистически значимой. Маша делает вывод о процикличности госрасходов.
Также Маша для каждой отдельной страны из 43 стран оценивает парную регрессию g_{it} от y_{it} и получает «коэффициент процикличности госрасходов» -- значение коэффициента наклона.

А) Маша замечает, что примерно половина стран имеет положительный коэффициент, вторая половина – отрицательный. Среднее арифметическое из этих коэффициентов равно 0.097, что сильно отличается от оценки 0.675 и скорее говорит об ациклической динамике госрасходов. Нет ли противоречия в результатах Маши? Дайте пояснения.

Б) Маша также по данным об этих 43 странах выясняет, как коэффициент процикличности госрасходов зависит от открытости экономики, которую Маша измеряет несколькими показателями в % от ВВП: экспорт (X), импорт (IMP), их сумма (Sum) и чистый экспорт (Nx).
Оцениваются 5 регрессий коэффициента процикличности госрасходов от:

1. X и IMP
2. Sum
3. Sum и Nx
4. IMP и Sum
5. X, IMP, Sum, Nx

В таблице показаны результаты: оценки коэффициентов, в скобках под ними – стандартные ошибки, коэффициенты детерминации, ESS – сумма квадратов остатков. Буквы стоят на месте пропущенных чисел.

	(1)	(2)	(3)	(4)	(5)
X	0.6 (0.04)	-	-	-	**Т** **(Ф)**
IMP	0.8 (0.04)	-	-	**И** **(К)**	**Ц** **(Ч)**
Sum	-	0.72 (0.02)	**Б** **(В)**	**Л** **(М)**	**Ш** **(Ы)**
Nx	-	-	**Г** **(Д)**	**-**	**Э** **(Ю)**
R2	0.6	**А**	**Е**	**П**	
ESS	200	220	**Ж**	**С**	

Восстановите числа, на месте которых стоят буквы (где это возможно). Приведите подробные пояснения.

Задача 3. О нелинейной регрессии (30 баллов)

Рассматривается следующая нелинейная модель регрессии с неизвестными параметрами α и β:

$$y_i = \beta x_i^{\alpha} + u_i, \tag{1}$$

где $(x_1, y_1), \dots , (x_n, y_n)$ — выборка независимых, одинаково распределённых двумерных наблюдений; x_i имеют невырожденное распределение; $E(u_i|x_i) = 0$.
Исследователь может оценивать только линейные модели, а также пользоваться таблицей стандартного нормального распределения.
Считая, что $\beta \neq 0$, он хочет предложить процедуру проверки гипотезы H_0: $\alpha = 1$:

- Вычисляется b – МНК-оценка коэффициента β из модели:
$$y_i = \beta x_i + u_i \tag{2}$$
- Составляется статистика $$H = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(y_i - bx_i) \tag{3}$$

Задание:
Объясните, почему статистика H может быть использована для конструирования теста проверки гипотезы H_0. Для этого выполните нижеуказанные шаги I-VIII:
(Если Вам не удалось обосновать очередной шаг, то на следующем шаге пользуйтесь предыдущим как доказанным фактом.)

I. Покажите, что при верной нулевой гипотезе

$$H = -\sqrt{n}(b - \beta)\frac{1}{n}\sum_{i=1}^{n} x_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} u_i \tag{4}$$

II. Покажите, что при верной нулевой гипотезе статистику H можно представить в виде:

$$H = \begin{pmatrix} -\frac{\frac{1}{n}\sum_{i=1}^{n} x_i}{\frac{1}{n}\sum_{i=1}^{n} x_i^2} & 1 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i}^{n} x_i u_i \\ \frac{1}{\sqrt{n}}\sum_{i}^{n} u_i \end{pmatrix} \tag{5}$$

III. Покажите, что асимптотическое распределение статистики H при верной нулевой гипотезе нормальное; найдите математическое ожидание $E(H)$ и дисперсию $V(H)$. (Вспомните, что $V(A\varepsilon) = A^T V(\varepsilon) A$).

IV. Чтобы сконструировать тест, необходимо найти распределение H при верной альтернативной гипотезе H_1: $\alpha \neq 1$.
Покажите, что при верной альтернативной гипотезе H_1 МНК-оценка b представима в виде:

$$b = \beta\frac{\sum_{i=1}^{n} x_i^{1+\alpha}}{\sum_{i=1}^{n} x_i^2} + \frac{\sum_{i}^{n} x_i u_i}{\sum_{i=1}^{n} x_i^2} \tag{6}$$

V. Покажите, что при верной альтернативной гипотезе статистику H можно представить в виде суммы $H = \beta\sqrt{n}\,A_n + B_n$, где B_n – правая часть уравнения (5), а A_n вычисляется по формуле:

$$A_n = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^{\alpha} - \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^{1+\alpha}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\frac{1}{n}\sum_{i=1}^{n} x_i\right) \tag{7}$$

VI. Используя закон больших чисел и ЦПТ, найдите предел по вероятности для A_n.

VII. Пользуясь фактом, что найденный Вами предел равен нулю только при $\alpha = 1$ (и не равен нулю в остальных случаях), покажите, что в условиях верной альтернативной гипотезы $\underset{n\to\infty}{\operatorname{plim}}\left(\beta\sqrt{n}A_n\right) = \infty$ и $\underset{n\to\infty}{\operatorname{plim}}(H) = \infty$.

VIII. На основе полученных результатов постройте правило отвержения нулевой гипотезы.

Задача 4. О репликации исследования (25 баллов)

Нередко бывает, что исследователи склонны публиковать результаты своих исследований только в том случае, если получен статистически значимый результат. В то же время «репликации» исследований (т.е. повторение логики исходного исследования на новых данных) могут публиковаться как со статистически значимыми, так и с незначимыми результатами (последние – для «опровержения выводов первопроходцев»).

Представьте, что в исходной работе для проверки нулевой гипотезы используется некоторая статистика X_1, имеющая нормальное распределение с параметрами $(\mu, 1)$, а в «репликации» используется X_2 с тем же самым распределением, причём случайные величины X_1 и X_2 независимы.

Первое исследование публикуется, если оно даёт значимый результат, т.е. если $X_1 > c$, где c – некоторая константа.
Если $X_1 \le c$, то первое исследование не публикуется, и значит, нет и его репликации. Соответственно, мы либо наблюдаем оба исследования при условии значимости результатов первого исследования, либо ничего не наблюдаем.

Необходимо оценить параметр μ.

1. *Предложите какую-нибудь функцию от* X_1 и X_2 *в качестве оценки для μ, так чтобы она была несмещенной при условии, что мы наблюдаем оба исследования. Можно ли в качестве такой оценки использовать* X_2?

2. *Можно ли с той же целью использовать* X_1?

Указание: предварительно найдите условную функцию плотности для X_1 *при условии, что мы наблюдаем оба исследования:* $p(x_1 \mid X_1 > c)$

3. *Запишите формулу совместной плотности* X_1 и X_2 *при условии, что мы наблюдаем оба исследования.*

4. *А) Дайте содержательное (интуитивное) объяснение, почему оценка* $\bar{X} = \frac{X_1+X_2}{2}$ *параметра* μ *является завышенной при условии, что мы наблюдаем оба исследования.*
Б) Приведите также формальное доказательство.

В решении используйте обозначения $\phi(x)$ и $\Phi(x)$ для функции плотности и функции распределения стандартной нормальной величины соответственно.