

Доп. занятие 1.

1. Рассмотрим НКЛММР $Y_i = \beta_1 + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i$ с неслучайными регрессорами. Дополнительно известно, что на самом деле $\beta_2 = \dots = \beta_k = 0$.

(a) Найдите $\mathbb{E}(R^2)$.

(b) Найдите $\mathbb{E}(R_{adj.}^2)$.

2. (Универсиада по эконометрике, 2022, очный тур). Консультант Антон оценил регрессию вида

$$y_i = x_i' \beta + \varepsilon_i,$$

где $x_i \in R^5$ и все стандартные предпосылки теоремы Гаусса-Маркова выполнены, и по заданию от менеджера протестировал гипотезы

$$H_0 : \beta_1 + \beta_2 = 1$$

$$H_1 : \beta_1 + \beta_2 > 1$$

Получившееся p -value оказалось равно 0.52. За 5 минут до дедлайна Антон заметил, что, сохраняя данные в Excel, он случайно скопировал их дважды. То есть вместо 124 наблюдений он использовал 248, так что каждое наблюдение повторяется ровно два раза.

- (a) Как это повлияло на его оценки $\hat{\beta}_1, \hat{\beta}_2$, их дисперсии и значимость?
- (b) Предположим, что менеджера волнует только ответ на вопрос, какую из гипотез следует принять. Стоит ли консультанту Антону исправлять свои выводы? Если ваш ответ да, объясните, как именно исправить, если нет, объясните почему.
3. (Универсиада по эконометрике, 2023). 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 $X_1 \leq c$, 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 .

- (a) 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 ?
- (b) 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: $p(x_1 \mid X_1 > c)$
- (c) Запишите формулу совместной плотности X_1 и X_2 при условии, что мы наблюдаем оба исследования.
- (d) 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. Приведите также формальное доказательство.

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