

Доп. занятие 5.

1. Дано $n = n_1 + n_2 + n_3$ наблюдений переменных x и y . Известно, что для n_1 наблюдений $x = 1$ и $y = 1$, для n_2 наблюдений $y = 0$ и $x = 1$, для n_3 наблюдений $x = 0$ и $y = 0$. Покажите, что как для *logit*, так и для *probit* модели уравнения правдоподобия не имеет конечного решения.
2. (Универсиада по эконометрике, 2022) Виктор интересуется, как программы переобучения безработных в стране А влияют на их заработок после нового трудоустройства. В базе есть следующие показатели:
 - Y — заработная плата (в условных единицах) после того как человек нашёл новую работу;
 - D — бинарная переменная, равная 1, если безработный фактически прошёл программу переобучения, и 0, если нет;
 - X_1 — возраст (лет);
 - X_2 — бинарная переменная пола: 1 для мужчин и 0 для женщин;
 - Z — бинарная переменная, равная 1, если в службе занятости страны А безработному предложили пройти программу переобучения, и 0, если не предложили (далее он мог согласиться или не согласиться или же без предложения самостоятельно попросить записать его на переобучение);
 - n — число людей в однородных группах (с одинаковыми характеристиками).

Виктор считает, что пол и возраст влияют на то, прошёл ли безработный фактически программу переобучения. Он предложил оценить влияние переобучения безработных на заработную плату после трудоустройства следующим способом: Сначала с помощью модели бинарного выбора оценивается вероятность безработного фактически пройти программу переобучения в зависимости от пола и возраста. Затем берутся пары безработных (1 к 1) с одинаковой расчётной вероятностью такие, что один человек из пары фактически прошёл переобучение ($D = 1$), а другой — нет ($D = 0$). И среди получившихся пар надо посчитать оценку для выражения:

$$\alpha_1 = E[Y \mid D = 1] - E[Y \mid D = 0] = \frac{E[DY]}{P(D = 1)} - \frac{E[(1 - D)Y]}{P(D = 0)}$$

(Можно использовать выражение без доказательства.) Он оценил пробит-модель бинарного выбора с зависимой переменной D и показал следующие частичные результаты в таблицах 1 и 2 :

ТАБЛИЦА 1. Частичные результаты для тех, кто фактически прошёл переобучение ($D = 1$)

№ (название группы)	Y	$\hat{P}(D = 1 X)$	X_1	X_2	n
G1	10	0.50	25	0	20
G2	9	0.84	21	1	30

и коэффициенты для пробит-регрессии $\hat{\beta}_0 = -1.25, \hat{\beta}_1 = 0.05, \hat{\beta}_2 = 1.2$

ТАБЛИЦА 2. Частичные результаты для тех, кто не прошёл переобучение ($D = 0$)

№ (название группы)	Y	$\hat{\beta}_0 + \hat{\beta}_1 * X_1 + \hat{\beta}_2 * X_2$	n
A1	7	-1.000	200
A2	8	0.000	30
A3	2	1.000	40
A4	8	1.668	30
A5	4	1.815	30

- Найдите оценку влияния переобучения безработных на заработную плату после трудоустройства способом, который предлагает Виктор
- Можно ли сказать, что выводы из оценки Виктора можно применить ко всем безработным в стране А? Почему?
- Можно ли вместо способа Виктора посчитать оценку как разницу средних заработных плат между теми, кто прошёл и не прошёл переобучение? Почему?
- Коллега Виктора сомневается в результатах его расчётов и считает, что между фактическим переобучением и будущей заработной платой есть эндогенность: существуют такие ненаблюдаемые характеристики человека, как мотивация и способности, а они могли повлиять и на решение переобучаться, и на будущую зарплату. Предложите способ оценки, который решал бы эту проблему. Какие предпосылки об имеющихся данных нужно сделать, чтобы воспользоваться этим методом?