Лекция 2 по микроэконометрике

Buchko Daniil, BEC175

16 октября 2020 г.

В прошлый раз мы остановились на использовании ММП для оценки параметров β модели линейной вероятности. Кратко напомним основные сюжеты. Функция максимального правдоподобия имела вид:

$$\max_{\beta} \mathcal{L} = \prod_{i=1}^{n} P(Y_i = y_i) = \prod_{i=1}^{n} \left(F(x_i'\beta)\right)^{y_i} \left(1 - F(x_i'\beta)\right)^{1-y_i} \to \max_{\beta_0,\ldots,\beta_{k-1}}$$

Напомним, что оценки полученные ММП являются случайными величинами, потому что для разных частей обучающей выборки мы получаем их разные значения. Кроме того, всё, что зависит от полученных оценок $\hat{\beta}$ (например, предсказание вероятности для нового объекта) тоже является случайными величинами, поэтому даже для предельных эффектов можно строить доверительные интервалы, потому что они связаны с оценками следующим способом:

$$\frac{\partial \hat{P}(\bar{Y} = 1)}{\partial x_j} = \frac{\partial F(\overline{x}'\hat{\beta})}{\partial x_j} = f(\overline{x}'\hat{\beta})\hat{\beta}_j, \quad for\ x_j \in R$$

Как найти дисперсию функции от случайной величины (оценки)? При помощи дельта-метода. **Как делаются предсказания по оцененным моделям?** Рассмотрим пример. Мы обучили модель:

$$\hat{P}(x) = F\left(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_{k-1} x_{k-1}\right)$$

Теперь, подставим какое-то значение x_0, чтобы получить оценку вероятности, события, что x_0 принадлежит классу 1:

$$\hat{P}(x_0) = F\left(\hat{\beta}_0 + \hat{\beta}_1 x_{1_0} + \hat{\beta}_2 x_{2_0} + \cdots + \hat{\beta}_{k-1} x_{k-1_0}\right)$$

При каких значениях оценки вероятности мы будем считать, что объект x_0 принадлежит классу 1? Можно рассматривать различные пороги, но обычно, при сбалансированных классах и в случае, если стоимость ошибки классификации одинакова, берут порог в 0.5:

$$\hat{Y} = \begin{cases} 1, & \text{Если } F(x_0'\beta) > 0.5 \\ 0, & \text{Если } F(x_0'\beta) \leqslant 0.5 \end{cases}$$

Супер крутой пример про МакФаддена (1974)

МакФадден в 1974 пытался оценить, насколько будет выгодно строительство метрополитена (BART). Для этого он собрал данные 160 человек, которые были следующего рода:

$$Y_i = \begin{cases} 1, & \text{человек выбирает автомобиль } (n_1 = 78) \\ 0, & \text{человек выбирает автобус } (n_0 = 82) \end{cases}$$

В качестве объясняющих переменных он выбрал следующие:

1. I - Доход семьи (\$10 000)
2. wage - Заработная плата в час (\$)

3. cost(k) - Стоимость поездки для выбранного вида транспорта (центы)
4. time(k) - Время потраченное на поездку в один конец для выбранного типа транспорта (минуты)
5. timeA - Время пешком до остановки (минуты)

Обратим внимание, что единицы измерения различны, чтобы факторы находились приблизительно в одном масштабе.

Далее он обозначил потенциально возможные зависимости интересным образом. Он определил модель отдельно для автомобилистов и отдельно для пользователей автобусов. В своём исследовании МакФадден пользовался линейными вероятностными моделями, поэтому целевую переменную следовало трактовать в терминах полезности. Модель для автолюбителей имела вид:

$$u_1 = \gamma_{01} + \gamma_{11} I + \gamma_2 cost_1 + \gamma_3 time_1 \times wage + \varepsilon_1$$

Соответственно, модель для пользователей автобусами имела следующий вид:

$$u_0 = \gamma_{00} + \gamma_{10} I + \gamma_2 cost_0 + \gamma_3 time_0 \times wage + \gamma_4 timeA \times wage + \varepsilon_0$$

Задав такие модели, он, фактически, предположил, что 1) у 0 и 1 различное влияние бюджета на желание выбрать транспорт, 2) у 0 и 1 одинаковое влияние стоимости поездки на выбор транспорта, 3) у 0 и 1 одинаковое влияние альтернативной стоимости дороги. Кроме того, он добавил в модель с автобусом еще влияние стоимости времени дойти до остановки.

Следующим шагом, автор вывел разницу в полезностях автомобилистов и автобусников следующим образом:

$$y_i^* = u_1 - u_0 = (\gamma_{01} - \gamma_{00}) + (\gamma_{11} - \gamma_{10}) I + \gamma_2 \Delta cost + \gamma_3 \Delta time \times wage - \gamma_4 timeA \times wage + (\varepsilon_1 - \varepsilon_0)$$

В результате оценки последней модели, оказались значимыми коэффициенты разности расходов и разницы в альтернативной стоимости времени. То есть выходит, что если автомобиль действительно дороже, чем автобус, то желание ездить на автомобиле уменьшается. Если альтернативная стоимость времени, затраченного на поездку в машине растёт, то желание поехать на машине уменьшается.

МакФадден ставит интересный вопрос: если стоимость поездки на автобусе увеличится на 1 цент, то насколько больше людей выберут автомобиль? Чтобы ответить на него, посчиатем предельный эффект:

$$\frac{\partial \hat{P}(Y=1)}{\partial cost} = \frac{\sum_{i=1}^n Y_i}{n} \left(1 - \frac{\sum_{i=1}^n Y_i}{n}\right) \gamma_2 \approx 0.0025$$

Можно ли посчитать предельный эффект от изменения на 100 долларов? Нет, потому что предельный эффект создан для малых величин. Затем МакФадден задается вопросом, а сколько готов заплатить человек, чтобы время в пути сократилось на час? Для этого, разница полезностей должна остаться прежней при одновременном изменении времени в поездке:

$$0 = \Delta y^* = 0.0148 \times 30 \times wage - 0.0102 \Delta(\Delta cost)$$

0.0148 - время, которое человек тратит на поездку в один конец. Если решить это уравнение, получим, что

$$\Delta(\Delta cost) \approx 43 wage$$

Получается, что человек готов потратить 43% своей часовой заработной платы, чтобы время в пути сократилось на час.

Тестирование гипотез. Классическая тройка тестов.

Общая линейная гипотеза. В эконометрике только одна гипотеза. Это гипотеза об ограничении на коэффициенты (общая линейная гипотеза).

$$H_0\colon R\beta = r$$

$R \in \mathbb{R}^{m\times k}$ – матрица некоторых ограничений, где m – количество ограничений
$\beta \in \mathbb{R}^{k\times 1}$ – вектор оценок
r – вектор ограничений

Пример: Гипотезу о незначимости коэффициента β_s можно записать следующим образом:

$$\begin{bmatrix} 0 & \dots & 1 & \dots & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_s \\ \vdots \\ \beta_k \end{bmatrix} = 0 \iff \beta_s = 0$$

Пример: Гипотезу об неадекватности модели в целом (все коэффициенты кроме свободного члена равны нолю) можно записать следующим образом:

$$\begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_s \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} \iff \begin{bmatrix} \beta_2 \\ \vdots \\ \beta_s \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Чтобы проверить общую линейную гипотезу, используется: тест Вальда. Логика такая: если линейное ограничение верно (а мы оценивали модель без учета ограничения), то при подстановке $\hat{\beta}$ в уравнение $R\beta = r$, равенство должно почти соблюдаться.

Wald:

$$W = (R\hat{\beta} - r)^T (As\hat{V}ar(\hat{\beta}))^{-1} (R\hat{\beta} - r) \sim^{as} \chi^2_m, \quad \text{если } H_0 \text{ верна}$$

LR:

$$LR = -2(lnL_R - lnL_{UR}) \sim \chi^2_m, \quad \text{если } H_0 \text{ верна}$$

LM (тест множителей Лагранжа):

$$LM = S^T(\hat{I}^{-1}(\beta))S \sim \chi^2_m, \quad \text{если } H_0 \text{ верна}$$

Во всех тестах логика одинаковая: при выполнении линейных ограничений, разница между $R\beta$ и r не должна быть слишком большой. Для LM теста статистика считается в точке $\beta = \hat{\beta}_R$. $S-$ вектор первых производных функций логарифма правдоподобия по каждому из весов.