Множественная линейная регрессионная модель

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$$

Темы лекции

- Множественная линейная регрессионная модель
- Метод наименьших квадратов оценки коэффициентов МЛРМ.
- Матричное выражение МНК-оценок коэффициентов МЛРМ.

Множественные регрессионные модели

Независимая переменная Y характеризует состояние или поведение экономического объекта. Набор переменных $X_1,\ldots,X_k$ характеризуют этот экономический объект качественно или количественно.

МЛРМ $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$

Пример

$$Q^D = \beta_0 + \beta_1 P + \beta_2 X + \beta_3 PM + \varepsilon$$

где Q^D – объем спроса на масло,
X – доход,
P – цена на масло,
PM – цена на мягкое масло.

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \varepsilon$$

Здесь нам неизвестны коэффициенты β и параметры распределения ε.

Для их оценки имеется выборка из N наблюдений над переменными Y и $X_1, \ldots, X_k$.

Для каждого наблюдения должно выполнятся следующее равенство:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_k X_{ki} + \varepsilon_i$$

Матричная форма записи МЛРМ

$$Y = X\beta + \varepsilon$$

где

$$Y = \begin{bmatrix} Y_1 \\ \dots \\ Y_N \end{bmatrix} \quad X = \begin{bmatrix} 1\,X_{11} & \dots & X_{k1.} \\ . & . & . \\ 1\,X_{N1} & \dots & X_{kN} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \dots \\ \beta_k \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \dots \\ \varepsilon_N \end{bmatrix}$$

Векторная форма записи МЛРМ

$$y_i = x'_i \beta + \varepsilon_i$$

где

$$x'_i = (1, X_{1i}, X_{2i}, ..., X_{ki})$$

$$x'_i \beta = (1, X_{1i}, X_{2i}, ..., X_{ki}) \begin{pmatrix} \beta_0 \\ ... \\ \beta_k \end{pmatrix} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + ... \beta_k X_{ki}$$

Метод наименьших квадратов

$$\sum_{i=1}^{N} e_i^2 \to \min_{(\hat{\beta})}$$

Среди всех возможных гиперплоскостей выбираем ту, для которой сумма квадратов остатков минимальна

Что будем минимизировать

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \ldots + \hat{\beta}_k X_{ki}$$

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \ldots - \hat{\beta}_k X_{ki}$$

$$RSS = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2 =$$

$$\sum_{i=1}^{N} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \ldots - \hat{\beta}_k X_{ki})^2 \to \min_{(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)}$$

Минимизация

$$\begin{cases} \dfrac{\partial RSS}{\partial \hat{\beta}_0} = 0 \\ \dfrac{\partial RSS}{\partial \hat{\beta}_1} = 0 \\ \dfrac{\partial RSS}{\partial \hat{\beta}_2} = 0 \\ \ldots \\ \dfrac{\partial RSS}{\partial \hat{\beta}_k} = 0 \end{cases} \quad \text{или} \quad \begin{cases} -2\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \ldots - \hat{\beta}_k X_{ki}) = 0 \\ -2\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \ldots - \hat{\beta}_k X_{ki}) X_{1i} = 0 \\ -2\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \ldots - \hat{\beta}_k X_{ki}) X_{2i} = 0 \\ \ldots \\ -2\sum_{i=1}^{N}(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1i} - \ldots - \hat{\beta}_k X_{ki}) X_{ki} = 0 \end{cases}$$

Система нормальных уравнений

$$
\begin{cases}
N\hat{\beta}_0 + \hat{\beta}_1 \sum\limits_{i=1}^{N} X_{1i} + \hat{\beta}_2 \sum\limits_{i=1}^{N} X_{2i} + \ldots + \hat{\beta}_k \sum\limits_{i=1}^{N} X_{ki} = \sum\limits_{i=1}^{N} Y_{1i} \\
\hat{\beta}_0 \sum\limits_{i=1}^{N} X_{1i} + \hat{\beta}_1 \sum\limits_{i=1}^{N} X_{1i}^2 + \hat{\beta}_2 \sum\limits_{i=1}^{N} X_{1i} X_{2i} + \ldots + \hat{\beta}_k \sum\limits_{i=1}^{N} X_{1i} X_{ki} = \sum\limits_{i=1}^{N} Y_{1i} X_{1i} \\
\hat{\beta}_0 \sum\limits_{i=1}^{N} X_{2i} + \hat{\beta}_1 \sum\limits_{i=1}^{N} X_{1i} X_{2i} + \hat{\beta}_2 \sum\limits_{i=1}^{N} X_{2i}^2 + \ldots + \hat{\beta}_k \sum\limits_{i=1}^{N} X_{2i} X_{ki} = \sum\limits_{i=1}^{N} Y_{1i} X_{2i} \\
\ldots \\
\hat{\beta}_0 \sum\limits_{i=1}^{N} X_{ki} + \hat{\beta}_1 \sum\limits_{i=1}^{N} X_{1i} X_{ki} + \hat{\beta}_2 \sum\limits_{i=1}^{N} X_{2i} X_{ki} + \ldots + \hat{\beta}_k \sum\limits_{i=1}^{N} X_{ki}^2 = \sum\limits_{i=1}^{N} Y_{1i} X_{ki}
\end{cases}
$$

Система линейных уравнений

Система в матричном виде

$$
\begin{pmatrix}
N & \sum X_{1i} & \sum X_{2i} & \dots & \sum X_{ki} \\
\sum X_{1i} & \sum X_{1i}^2 & \sum X_{1i}X_{2i} & \dots & \sum X_{1i}X_{ki} \\
\sum X_{2i} & \sum X_{1i}X_{2i} & \sum X_{2i}^2 & \dots & \sum X_{2i}X_{ki} \\
 & & \dots & & \\
\sum X_{ki} & \sum X_{1i}X_{ki} & \sum X_{2i}X_{ki} & \dots & \sum X_{ki}^2
\end{pmatrix}
\begin{pmatrix}
\hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_k
\end{pmatrix}
=
\begin{pmatrix}
\sum Y_i \\ \sum Y_i X_{1i} \\ \sum Y_i X_{2i} \\ \dots \\ \sum Y_i X_{ki}
\end{pmatrix}
$$

или

$$(X'X)\hat{\beta} = X'Y$$

Итог

$$\hat{\beta} = (X'X)^{-1}X'Y$$

МНК оценки коэффициентов МЛРМ

Полная мультиколлинеарность

Коэффициенты по методу наименьших квадратов существуют не всегда, а только в том случае, когда определитель матрицы ($X'X$) отличен от нуля.

Определитель будет равен нулю в случае, если столбцы матрицы X линейно зависимы. Такое может произойти, если между независимыми переменными существует точное линейное соотношение.

Пример

$$Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 W + \varepsilon$$

где

Y - средняя оценка на экзамене состоящую из трех объясняющих переменных:

I – доход родителей,

D – среднее число часов, затраченных на обучение в день,

W – среднее число часов, затраченных на обучение в неделю.

Очевидно, что $W=7D$.

Устранение полной мультиколлинеарности

Случай полной мультиколлинеарности отследить легко, поскольку в этом случае невозможно построить оценки по методу наименьших квадратов. Если в модели присутствует полная мультиколлинеарность, следует удалить из регрессионного уравнения одну из переменных, которые входят в линейное соотношение.

DUMMY TRAP

Дамми-переменная – переменная, принимающая только два значения: 0 и 1.

С помощью таких переменных учитывается влияние качественных переменных, принимающих несколько значений.

Вопросы для самопроверки

- Система нормальных уравнений для нахождения коэффициентов по МНК.
- В каком случае линии регрессии по методу наименьших квадратов не существует
- Приведите примет модели, в которой присутствует полная мультиколлинеарность.
- Укажите размерности матриц, участвующих в формуле МНК-коэффициентов.
- .Как устранить проблему полной мультиколлинеарности.
- Выведите систему нормальных уравнений.
- Выведите матричную формулу МНК коэффициентов.
- Приведите пример ситуации, когда линейной зависимости между объясняющими переменными нет, а коэффииценты МЛРМ не существуют.
- Как влияют выбросы на результаты оценивания.
- Как исследовать устойчивость результатов оценивания.