



Математические методы в экономике

исов Магомед-Рамзан Бувайсарович,
кандидат экономических наук, доцент

Лекция 9. Множественная регрессия



План

1. Типы данных
2. Классическая модель множественной регрессии:
предпосылки
3. Качество подгонки:
 - стандартная ошибка регрессии (SEE),
 - и скорректированный
4. Гипотезы и доверительные интервалы

Типы данных в статистических исследованиях

- Пространственные выборки
- Временные ряды
- Панельные данные



Классическая линейная модель множественной регрессии (КЛММР)

$$y_i = a_0 + a_1 x_i^{(1)} + a_2 x_i^{(2)} + \dots + a_p x_i^{(p)} + \varepsilon_i$$

y_i – значения зависимой переменной,

$x_i^{(p)}$ – значения независимых переменных (регрессоров),

p – число коэффициентов в модели,

ε_i – случайные ошибки.



Предпосылки КЛММР

1) модель линейна по параметрам и правильно специфицирована

$$y_i = a_0 + a_1 x_i^{(1)} + a_2 x_i^{(2)} + \dots + a_p x_i^{(p)} + \varepsilon_i$$

2) $x^{(1)} \dots x^{(p)}$ – детерминированные переменные, которые являются линейно независимыми

3) математическое ожидание случайных ошибок равно нулю: $E(\varepsilon_i) = 0$

Предпосылки КЛММР

4) случайные ошибки имеют постоянную дисперсию: **$V(\varepsilon_i) = \sigma^2 = \text{const}$**

5) случайные ошибки, соответствующие разным наблюдениям не зависят друг от друга (не коррелированы) **$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$** при

$i \neq j$ * случайные ошибки ε_i имеют нормальное распределение



Теорема Гаусса - Маркова



Если выполнены условия (1) – (5), **то** оценка коэффициентов модели, полученная по МНК, является:

- (а) несмещенной;
- (б) эффективной.

Качество подгонки модели



1. Стандартная ошибка регрессии
2. Коэффициент детерминации R^2
3. Скорректированный (нормированный) коэффициент детерминации R^2

Оценка дисперсии случайной ошибки



Если выполнены условия (1)–(6), то несмещенная оценка дисперсии случайных ошибок σ^2 имеет вид:

$$\hat{\sigma}^2 = S^2 = \frac{1}{n - p} \sum_{i=1}^n \varepsilon_i^2$$

S^2 – используется для расчета стандартных ошибок коэффициентов и стандартной ошибки регрессии.

Стандартная ошибка регрессии

(Standard Error of Estimate – не путать с ESS)

$$SEE = \sqrt{S^2} = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i^2}{n - p}}$$

Измеряет среднюю величину ошибки модели.

Используется для оценки качества подгонки модели: чем меньше SEE, тем точнее модель.

Можно использовать для сравнения нескольких однотипных уравнений регрессии.



Коэффициент детерминации R^2

Ранее нами показано, что:

$$y_i = \tilde{y}_i + e_i \quad \text{Var}(y) = \text{Var}(\epsilon) + \text{Var}(\tilde{y})$$

Это означает, что мы можем разложить $\text{Var}(y)$ на две части: $\text{Var}(\tilde{y})$ — часть, которая «объясняется» уравнением регрессии и $\text{Var}(\epsilon)$ — «необъяснённую» часть.

$$R^2 = \frac{\text{Var}(\tilde{y})}{\text{Var}(y)} = \frac{RSS}{TSS} = 1 - \frac{\text{Var}(e)}{\text{Var}(y)}$$

$$0 \leq R^2 \leq 1$$



Скорректированный (нормированный) R^2

R^2 с учетом штрафа за число переменных:

$$R^2_{adj} = R^2 - \frac{p - 1}{n - p} (1 - R^2)$$

R^2_{adj} можно использовать для сравнения моделей с одинаковой зависимой переменной, но разным числом независимых переменных



Использование R^2 и R^2_{adj}

Есть соблазн свести выбор уравнения к задаче максимизации R^2 или R^2_{adj} . Не стоит этого делать:

1. Высокий R^2 (или R^2_{adj}) **говорит** о том, что регрессоры предсказывают большую долю изменений y .

2. Высокий R^2 (или R^2_{adj}) **не говорит** о том, что вы верно выявили причинно-следственную связь между переменными.

3. Высокий R^2 (или R^2_{adj}) **не гарантирует**

Тестирование гипотез

1. Тестирование значимости коэффициента
2. Тестирование гипотезы $a_i = A$
3. Доверительный интервал для коэффициента
4. Тестирование значимости уравнения



Тестирование значимости коэффициента



Рассматриваемая модель:

$$y_i = a_0 + a_1 x_i^{(1)} + a_2 x_i^{(2)} + \dots + a_p x_i^{(p)} + \varepsilon_i$$

Тестируемая гипотеза H_0 : **$a_p = 0$**

«Переменная $x^{(p)}$ не оказывает значимого влияния на переменную y ».

Альтернативная гипотеза H_1 : **$a_p \neq 0$**

«Переменная $x^{(p)}$ оказывает значимое влияние на переменную y »

Тестирование значимости коэффициента

Шаг 1. Вычисляем расчетное значение t -статистики

$$t_{расч} = \frac{\tilde{a}_p}{SE(\tilde{a}_p)}$$

Шаг 2. Выбираем **уровень значимости**

Уровень значимости – **вероятность ошибки** первого рода, то есть вероятность отклонить гипотезу H_0 , если на самом деле гипотеза H_0 верна.

В эконометрике обычно используется уровень значимости $\alpha = 0,01 = 1\%$ или $\alpha = 0,05 = 5\%$.



Тестирование значимости коэффициента

Шаг 3. Из таблиц t -распределения

Стьюдента находим критическое значение t -статистики $t_{кр}$. Оно зависит от уровня значимости α и так называемого числа степеней свободы, которое в случае нашего теста равно $(p-n)$.

Шаг 4. Сравниваем расчетное и критическое значение t -статистик

Если

то гипотеза H_0 не отклоняется (принимается), то есть мы делаем вывод о том, что переменная $x^{(p)}$ не оказывает значимого влияния на переменную y . В этом случае коэффициент при переменной $x^{(p)}$ называют незначимым



Тестирование значимости уравнения



Рассматриваемая модель:

$$\mathbf{y}_i = \mathbf{a}_0 + \mathbf{a}_1 \mathbf{x}_i^{(1)} + \mathbf{a}_2 \mathbf{x}_i^{(2)} + \dots + \mathbf{a}_p \mathbf{x}_i^{(p)} + \varepsilon_i$$

Тестируемая гипотеза H_0 : $\mathbf{a}_0 \dots \mathbf{a}_p = \mathbf{0}$ «Все переменные $\mathbf{x}^{(1)} \dots \mathbf{x}^{(p)}$ не оказывают значимого влияния на переменную $\mathbf{y}$ ».

Альтернативная гипотеза H_1 : «Хотя бы одна из переменных $\mathbf{x}^{(1)} \dots \mathbf{x}^{(p)}$ оказывает значимое влияние на переменную $\mathbf{y}$ ».

Тестирование значимости уравнения



Алгоритм проведения теста:

Шаг 1. Вычисляем расчетное значение F -статистики.

$$F_{расч.} = \frac{R^2}{1 - R^2} \frac{n - p}{p - 1}$$

Шаг 2. Выбираем уровень значимости.

Шаг 3. Из таблиц F -распределения Фишера находим критическое значение F -статистики ($F_{кр}$).

Оно зависит от уровня значимости α и от

Тестирование значимости уравнения



Шаг 4. Сравниваем расчетное и критическое значение F-статистик.

Если $F_{расч.} < F_{кр.}$,

то гипотеза H_0 не отклоняется (принимается), то есть мы делаем вывод о том, что все переменные $x^{(1)}...x^{(p)}$ не оказывает значимого влияния на переменную y .

В этом случае уравнение называют незначимым. В противном случае гипотеза H_0 не принимается (отклоняется).

Литература к лекции

- 1.Магнус** Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учеб. – 6-е изд., перераб. и доп. – М.: Дело, 2004. **Глава 3**
- 2.Доугерти** К. Введение в эконометрику: Учебник. 3-е изд. / Пер. с англ. – М.: ИНФРА-М, 2009. **Глава 3**