ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

В экономике поведение показателей зависит от множества разных факторов.

Наличие взаимосвязи между двумя случайными переменными показывает коэффициент корреляции. Но он не выявляет того, как они связаны друг с другом. Возможность прогнозировать поведение одной зависимой переменной по другим объясняющим переменным на основании выявления зависимости между ними предоставляет регрессионный анализ.

В парной регрессии есть только две переменные: зависимая переменная y и объясняющая переменная x.

Например, доход влияет на потребительские расходы, но не задает их однозначно, поскольку существуют другие факторы, такие как вкусы, предпочтения, от которых зависит потребление.

Модель парной линейной регрессии предполагает линейную зависимость:

$$y = \alpha + \beta x + \varepsilon, \qquad (1)$$

где α, β – коэффициенты модели, x – объясняющая переменная, регрессор, y – зависимая переменная, ε – случайный член.

Если на диаграмме рассеяния точки наблюдений $M_i(x_i, y_i)$, $i = 1,2,..., n$, распределяются случайным образом примерно вблизи некоторой прямой линии, то можно предполагать, что между переменными x, y существует линейная статистическая связь.

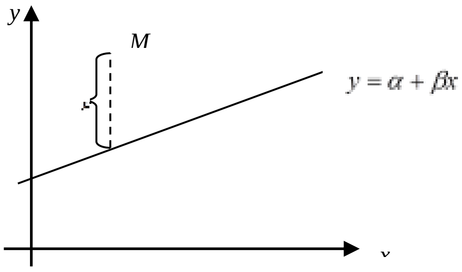

Рис. 4.1. Отклонение от линейной зависимости

Включение случайного члена ε в уравнение связано с возмущениями, которые не учтены в данной модели. Ими могут быть невключение других объясняющихся переменных, возможная нелинейность модели, неправильный выбор объясняющей переменной, ошибки измерений, агрегирование переменных и других факторы.

2. Оценивание параметров модели

Параметры $\alpha, \beta,$ которые определяют модель парной линейной регрессии (1), на практике не известны, и их можно оценивать по математическим данным, опираясь на различие принципы.

Пусть $\hat{y} = a + bx$ – оцененное уравнение регрессии, где a, b – некоторые коэффициенты. Обозначим через $e_i = y_i - \hat{y}_i$ остатки, где $\hat{y}_i = a + bx_i$, $i = 1,2,..., n.$

Для оценки коэффициентов модели наиболее широко используется метод наименьших квадратов (МНК). Согласно этому принципу значения коэффициентов a, b оцененного уравнения регрессии определяются из условия минимума суммы квадратов остатков:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \to \min.$$

Необходимое условие минимума функции двух переменных заключается в обращении в нуль частных производных по a и b:

$$2an - 2\sum_{i=1}^{n} y_i + 2b\sum_{i=1}^{n} x_i = 0, \tag{2}$$

$$2b\sum_{i=1}^{n} x_i^2 + 2a\sum_{i=1}^{n} x_i - 2\sum_{i=1}^{n} x_i y_i = 0. \tag{3}$$

Это система двух алгебраических уравнений относительно двух неизвестных a и b. Решим ее. Обе части уравнения (2) разделим на $2n$ и перепишем его в виде:

$$a = \bar{y} - b\bar{x}\,. \tag{4}$$

Найденное в (4) значение a подставим в уравнение (3). Разделим его на $2n$ и, группируя слагаемые, приведем уравнение (3) к виду:

$$\left[\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2\right]b=\left[\frac{1}{n}\sum_{i=1}^{n}x_i y_i-\bar{x}\cdot\bar{y}\right].$$

Если вспомнить формулы для вычисления выборочной вариации и выборочной ковариации, то это равенство можно переписать так: $bVar(x)=Cov(x,y)$.

Таким образом, решение системы линейных уравнений (2)-(3) находится по формулам:

$$b=\frac{Cov(x,y)}{Var(x)}, \qquad a=\bar{y}-b\bar{x}. \tag{5}$$

Учитывая формулы для выборочной ковариации и вариации, формулу для оценки коэффициента при регрессоре x можно переписать в виде:

$$b=\frac{\overline{xy}-\bar{x}\bar{y}}{\overline{x^2}-\bar{x}^2}.$$

Здесь r_{xy} – выборочный коэффициент корреляции между переменными x и y. Отметим, что уравнение $\hat{y}=a+bx$ является лишь оценкой для истинной линейной модели (1). Оцененная линия регрессии, вообще говоря, не совпадает с истинной линией регрессии (рис. 2).

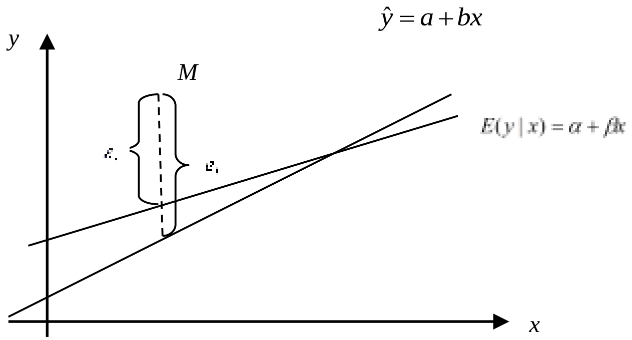

Рис. 2. Истинная и оцененная линии регрессии

$$r_{xy} = \frac{Cov(x,y)}{\sqrt{Var(x)Var(y)}} = \frac{Cov(x,y)\sqrt{Var(x)}}{Var(x)\sqrt{Var(y)}} = b\frac{s_x}{s_y}.$$

Поскольку регрессионная модель отражает статистическую связь между экономическими показателями, для ее параметров возможна содержательная интерпретация. Величина b представляет собой наклон, т.е. угловой коэффициент оцененной линии регрессии (рис. 2). Увеличение x на единицу приведет к увеличению y на b единиц. И это утверждение настолько верно, насколько b близко к β. А параметр a формально определяет оценку зависимой переменной y при $x = 0$.

Например, если построена зависимость потребления C от валового внутреннего продукта Y страны, т.е. оценена регрессия

$$\hat{C} = a + bY,$$

то по экономическому смыслу величина b равна предельной склонности к потреблению. А величина a не имеет экономического смысла, так как невозможно, чтобы выпуск в стране был равен нулю.