

Практическое занятие 3

Решить следующую задачу:

В следующей таблице приводятся данные для 10 фермерских хозяйств.

Хозяйства	Качество почвы (x_1), балл бонитета	Количество внесенных удобрений (x_2)	Урожайност ь (Y), ц/га
1	55	75	18,1+k
2	50	72	21,1+k
3	68	79	22,9+k
4	48	81	18,9+k
5	87	83	18,6+k
6	60	81	30,5+k
7	75	90	23,4+k
8	80	95	27,6+k
9	66	100	20,9+k
10	58	95	18,2+k

Здесь k – порядковый номер студента по журналу.

1. Найти коэффициенты ковариации, корреляции и детерминации.
2. Построить уравнение регрессии.
3. Провести анализ полученных результатов.

Здесь к – порядковый номер студента по журналу.

Методические указания к выполнению задания

Множественный регрессионный анализ

В этой лекции регрессионный анализ по методу наименьших квадратов обобщается для случая, когда в модели регрессии вместо одной независимой переменной используется несколько независимых переменных. Рассматриваются два новых вопроса. Один из них касается проблемы разграничения эффектов различных независимых переменных. Эта проблема в случае ее обострения известна под названием мультиколлинеарности. Другой вопрос состоит в оценке объединенной объясняющей способности независимых переменных в противоположность их отдельным эффектам.

6.1. Иллюстрация: модель с двумя независимыми переменными

Множественный регрессионный анализ является развитием парного регрессионного анализа применительно к случаям, когда зависимая переменная гипотетически связана с более чем одной независимой переменной. Большая часть анализа будет непосредственным расширением парной регрессионной модели, но здесь мы сталкиваемся с двумя новыми проблемами. Во-первых, при оценке влияния данной независимой переменной на зависимую переменную нам придется решать проблему разграничения ее воздействия и воздействия других независимых переменных. Во-вторых, мы должны будем решить проблему спецификации модели. Часто предполагается, что несколько переменных могут оказывать влияние на зависимую переменную, с другой стороны некоторые переменные могут не подходить для модели. Мы должны решить, какие из них следует включить в уравнение регрессии, а какие – исключить из него. В данной главе мы полагаем, что спецификация модели правильна. В большинстве ситуаций мы ограничимся основным случаем, где используются только две независимые переменные.

Начнем с рассмотрения примера, в котором определяются факторы совокупного спроса на продукты питания. Расширим первоначальную модель, включив учет влияния ценовых изменений на спрос, и допустим, что истинную зависимость можно выразить следующим образом:

$$y = \alpha + \beta_1 x + \beta_2 P + u, \quad (6.1)$$

где Y – общая величина расходов на питание, x – располагаемый личный доход, а P – цена продуктов питания. Это, разумеется, является значительным упрощением как с точки зрения состава независимых переменных, включенных в зависимость, так и с точки зрения математической формулы связи. Кроме того, мы неявно предполагаем наличие лишь прямой связи за счет допущения о том, что расходы на питание не влияют на доход и цену. Это могло бы быть в том случае, если бы цены определялись на мировом рынке, но в большинстве ситуаций более реально допустить, что расходы на продукты и их цены определяются совместно в результате взаимодействия предположения и спроса.



Рис. 6.1. Истинная модель с двумя независимыми переменными: расход как функция дохода и цены

Для геометрической иллюстрации этой зависимости необходима трехмерная диаграмма с отдельными осями для Y , x и P (рис. 6.1). Основание диаграммы содержит оси для x и P , и если пренебречь текущим влиянием случайного члена, то наклонная плоскость над ним показывает величину Y , соответствующую любому сочетанию x и P , измеренную расстоянием по вертикали от данной точки до этой плоскости. Так как расходы на питание могут увеличиваться с ростом доходов и уменьшаться с

увеличением цены, изображение на диаграмме было построено на основе допущения о том, что величина β_1 является положительной, а β_2 – отрицательной. Конечно, нереально было бы предполагать, что одна из величин x и P могла бы быть равной нулю, и структуру диаграммы можно описать следующим образом. Если бы обе величины x и P оказались равными нулю, то величина y равнялась бы α . При сохранении $P = 0$ уравнение (5.1) означает, что для любого положительного дохода величина y будет равна $(\alpha + \beta_1 x)$, и на рисунке приращение $\beta_1 x$ обозначено как «чистый эффект дохода». При сохранении $x = 0$ уравнение означает, что для любой положительной цены величина y будет равной $(\alpha + \beta_2 x)$, приращение $\beta_2 x$ на рисунке обозначено как «чистый эффект цены». Поскольку β_2 на практике является отрицательной величиной, отрицательным будет и этот эффект. Показан также комбинированный эффект дохода и цены $(\beta_1 x + \beta_2 P)$.

Итак, до сих пор мы пренебрегали случайным членом. Если он отсутствует на данный момент в уравнении (6.1), то значения y в выборке наблюдений для y , x и P будут находиться точно на наклонной плоскости и будет довольно просто вывести точные значения β_1 и β_2 (это не так просто сделать геометрически, если вы не имеете достаточно большого опыта построения трехмерных моделей, однако это довольно просто сделать алгебраическим путем).

Учет случайного члена приводит к тому, что фактические значения y будут лежать несколько выше или несколько ниже значений, соответствующих наклонной плоскости. Следовательно, теперь мы имеем трехмерный аналог для двухмерной задачи, показанной на рис. 2.2. Вместо нахождения линии, соответствующей двумерному рассеиванию точек, мы теперь должны расположить плоскость так, чтобы она соответствовала трехмерному рассеиванию. Уравнение для выбранной плоскости будет иметь вид:

$$y = a + b_1 x + b_2 P, \quad (6.2)$$

и ее расположение будет зависеть от выбора величин a , b_1 и b_2 , являющихся, соответственно, оценками α , β_1 и β_2 .

Используя данные для США за 1959-1983 гг. из табл. Б.1 и Б.2 по затратам на питание, располагаемому личному доходу и ценам, мы получим уравнение регрессии:

$$y = 116,7 + 0,112x - 0,739P, \quad R^2 = 0,99$$

(6.3)

$$(с.о.) \quad (9,6) \quad (0,003) \quad (0,114)$$

где y и x измерены в долларах США в постоянных ценах 1972 г., а P является индексом относительной цены, вычисленным путем деления неявного дефлятора цен продуктов питания на неявный дефлятор общих расходов (равный 100 в 1972 г.) и умноженным на 100.

Полученное уравнение следует интерпретировать следующим образом. При каждом увеличении располагаемого личного дохода на 1 млрд. сум (при сохранении постоянных цен) расходы на питание увеличатся на 112 млн. сум. На каждую единицу увеличения индекса цен (при сохранении постоянных доходов) эти расходы уменьшатся на 739 млн. долл. Чистый эффект в любой момент времени будет зависеть не только от этих коэффициентов, но также от размеров изменений x и P .

Например, в период 1975-1980 гг. располагаемый личный доход увеличился на 145,8 млрд. долл., и, согласно, уравнению (6.3), это привело к увеличению расходов на питание на 16,3 млрд. долл. В течение указанного периода индекс цен упал со 111,9 до 109,7, т.е. на 2,2 пункта, и это привело к дальнейшему увеличению y на 1,6 млрд. долл. Совместный эффект, прогнозируемый уравнением (6.3), таким образом, составил увеличение затрат на питание в размере 17,9 млрд. долл. Как видно из табл. Б.1 фактическое увеличение несколько больше, а именно 20,3 млрд. долл.

Даже если бы спецификация модели оказалась правильной (разумеется, это является большим упрощением), то между прогнозируемым изменением и полученным результатом будет наблюдаться расхождение. Прежде всего, оценки β_1 и β_2 подвержены влиянию ошибки выборки. Кроме того, фактические уровни затрат на питание в 1975 и 1980 гг. определялись не только экономической зависимостью, но и случайным членом и в тот и другой годы, а следовательно, измеренное приращение в течение этого периода имеет, наряду с экономической составляющей, также и случайную составляющую.

