

Практическое занятие 3. Множественный регрессионный анализ

Решить следующую задачу:

Заданы следующие данные опытов:

№	Качество почвы, балл бонитета (x_1)	Количество внесенных удобрений (x_2)	Урожайность (y)
1	31	75+k	22+k
2	34	72+k	23+k
3	40	79+k	23+k
4	44	81+k	24+k
5	51	83+k	25+k
6	56	81+k	25+k
7	62	90+k	27+k
8	64	95+k	29+k
9	69	100+k	31+k
10	75	95+k	33+k
11	81	110+k	34+k
12	83	115+k	36+k
13	88	110+k	35+k
14	95	120+k	36+k
15	98	130+k	37+k

Здесь k – порядковый номер студента по журналу.

Выполнить следующее:

1. Построить уравнения регрессии:

$$Y = a + bx_1$$

$$Y = a + bx_2$$

$$Y = a + b_1 x_1 + b_2 x_2$$

2. Найти коэффициенты ковариации, корреляции и детерминации.

3. Построить графики.

4. Провести экономический анализ полученных результатов.

Анализ полученных результатов:

1) $r = 0,98$;

$D = 0,96$;

$Y = 13,4 + 0,25x_1$.

Взаимосвязь между качеством почвы и урожайностью **прямая и полная**.

Изменения значений урожайности на **96%** зависят от качества почвы и на **4%** от других факторов.

При увеличении качества почвы на 1 балл урожайность увеличится на **0,25 ц/га**. Если $x_1 = 0$, то прогнозируемый уровень урожайности равняется **13,4 ц/га**.

3) $R = 0,98$;

$D = 0,96$;

$Y = 9,4 + 0,17x_1 + 0,09x_2$.

Взаимосвязь между качеством почвы, количеством вносимых удобрений и урожайностью **прямая и полная**.

Изменения значений урожайности на **96%** зависят от качества почвы и количества вносимых удобрений на **4%** от других факторов.

При увеличении качества почвы на 1 балл урожайность увеличится на **0,17 ц/га**, при увеличении количества вносимых удобрений на 1 единицу урожайность увеличится на **0,09 ц/га**. Если $x_1 = 0$ и $x_2 = 0$, то прогнозируемый уровень урожайности равняется **9,4 ц/га**.

Методические указания к выполнению заданий

Вывод формул для нахождения коэффициентов.

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 предельным эффектам.

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 воздействий 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. Вторая проблема будет рассмотрена в главе 6. 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:

$$y = \alpha + \beta_1 x + \beta_2 p + u,$$

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 предложения и спроса. Проблемы, которые возникают в таких моделях, будут рассмотрены в главе 11.

Для геометрической иллюстрации этой зависимости необходима трехмерная диаграмма с отдельными осями для y , x и p (рис. 5.1). 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 p , является положительной, а p^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 a . При сохранении $p = 0$ уравнение (5.1) означает, что для любого положительного дохода величина y будет равна $(a + p^*)$, и на рисунке приращение p, x обозначено как «чистый эффект дохода». При сохранении $x = 0$ уравнение означает, что для любой положительной цены величина y будет равной $(a + p^\#)$, приращение p^\wedge на рисунке обозначено как «чистый эффект цены». Поскольку p^2 на практике является отрицательной величиной, отрицательным будет и этот эффект.