На любой экономический показатель чаще всего оказывает влияние не один, а несколько факторов. Например, спрос на некоторое благо определяется не только ценой данного блага, но и ценами на замещающие и дополняющие блага, доходом потребителей и многими другими факторами. В этом случае вместо парной регрессии рассматривается множественная регрессия

 (1)

Множественная регрессия широко используется в решении проблем спроса, доходности акций, при изучении функции издержек производства, в макроэкономических расчетах и в ряде других вопросов экономики. В настоящее время множественная регрессия – один из наиболее распространенных методов в эконометрике. Основной целью множественной регрессии является построение модели с большим числом факторов, а также определение влияния каждого фактора в отдельности и совокупного их воздействия на моделируемый показатель.

Множественный регрессионный анализ является развитием парного регрессионного анализа в случаях, когда зависимая переменная связана более чем с одной независимой переменной. Большая часть анализа является непосредственным расширением парной регрессионной модели, но здесь также появляются и некоторые новые проблемы, из которых следует выделить две. Первая проблема касается исследования влияния конкретной независимой переменной на зависимую переменную, а также разграничения её воздействия и воздействий других независимых переменных. Второй важной проблемой является спецификация модели, которая состоит в том, что необходимо ответить на вопрос, какие факторы следует включить в регрессию (1), а какие – исключить из неё. В дальнейшем изложение общих вопросов множественного регрессионного анализа будем вести, разграничивая эти проблемы. Поэтому вначале будем полагать, что спецификация модели правильна.

Самой употребляемой и наиболее простой из моделей множественной регрессии является линейная модель множественной регрессии:

 (2)

По математическому смыслу коэффициенты  в уравнении (2) равны частным производным результативного признака *y* по соответствующим факторам:

,,…,.

Параметр *α* называется свободным членом и определяет значение *y* в случае, когда все объясняющие переменные равны нулю. Однако, как и в случае парной регрессии, факторы по своему экономическому содержанию часто не могут принимать нулевых значений, и значение свободного члена не имеет экономического смысла. При этом, в отличие от парной регрессии, значение каждого регрессионного коэффициента  равно среднему изменению *y* при увеличении *xj* на одну единицу лишь при условии, что все остальные факторы остались неизменными. Величина *ε* представляет собой случайную ошибку регрессионной зависимости.

Попутно отметим, что наиболее просто можно определять оценки параметров , изменяя только один фактор *xj*, оставляя при этом значения других факторов неизменными. Тогда задача оценки параметров сводилась бы к последовательности задач парного регрессионного анализа по каждому фактору. Однако такой подход, широко используемый в естественнонаучных исследованиях, (физических, химических, биологических), в экономике является неприемлемым. Экономист, в отличие от экспериментатора – естественника, лишен возможности регулировать отдельные факторы, поскольку не удаётся обеспечить равенство всех прочих условий для оценки влияния одного исследуемого фактора.

Получение оценок параметров  уравнения регрессии (2) – одна из важнейших задач множественного регрессионного анализа. Самым распространенным методом решения этой задачи является метод наименьших квадратов (МНК). Его суть состоит в минимизации суммы квадратов отклонений наблюдаемых значений зависимой переменной *y* от её значений , получаемых по уравнению регрессии. Поскольку параметры  являются случайными величинами, определить их истинные значения по выборке невозможно. Поэтому вместо теоретического уравнения регрессии (2) оценивается так называемое **эмпирическое уравнение регрессии**, которое можно представить в виде:

 (3)

Здесь  - оценки теоретических значений , или эмпирические коэффициенты регрессии, *е* – оценка отклонения *ε*. Тогда расчетное выражение имеет вид:

 (4)

Пусть имеется *n* наблюдений объясняющих переменных и соответствующих им значений результативного признака:

 (5)

Для однозначного определения значений параметров уравнения (4) объем выборки *n* должен быть не меньше количества параметров, т.е. . В противном случае значения параметров не могут быть определены однозначно. Если *n=p*+1, оценки параметров рассчитываются единственным образом без МНК простой подстановкой значений (5) в выражение (4). Получается система (*p*+1) уравнений с таким же количеством неизвестных, которая решается любым способом, применяемым к системам линейных алгебраических уравнений (СЛАУ). Однако с точки зрения статистического подхода такое решение задачи является ненадежным, поскольку измеренные значения переменных (5) содержат различные виды погрешностей. Поэтому для получения надежных оценок параметров уравнения (4) объём выборки должен значительно превышать количество определяемых по нему параметров. Практически, как было сказано ранее, объём выборки должен превышать количество параметров при *xj* в уравнении (4) в 6-7 раз.

Для проведения анализа в рамках линейной модели множественной регрессии необходимо выполнение ряда предпосылок МНК. В основном это те же предпосылки, что и для парной регрессии, однако здесь нужно добавить предположения, специфичные для множественной регрессии:

50.Спецификация модели имеет вид (2).

60.Отсутствие мультиколлинеарности: между объясняющими переменными отсутствует строгая линейная зависимость, что играет важную роль в отборе факторов при решении проблемы спецификации модели.

70.Ошибки  имеют нормальное распределение . Выполнимость этого условия нужна для проверки статистических гипотез и построения интервальных оценок.

При выполнимости всех этих предпосылок имеет место многомерный аналог теоремы Гаусса – Маркова: оценки , полученные по МНК, являются наиболее эффективными (в смысле наименьшей дисперсии) в классе линейных несмещенных оценок.

Оценка параметров линейного уравнения множественной регрессии

Рассмотрим три метода расчета параметров множественной линейной регрессии.

1. **Матричный метод**. Представим данные наблюдений и параметры модели в матричной форме.

** -  *n* – мерный вектор – столбец наблюдений зависимой переменной;

** - (*p*+1) – мерный вектор – столбец параметров уравнения регрессии (3);

** - *n* – мерный вектор – столбец отклонений выборочных значений *yi* от значений , получаемых по уравнению (4).

Для удобства записи столбцы записаны как строки и поэтому снабжены штрихом для обозначения операции транспонирования.

Наконец, значения независимых переменных запишем в виде прямоугольной матрицы размерности :



Каждому столбцу этой матрицы отвечает набор из *n* значений одного из факторов, а первый столбец состоит из единиц, которые соответствуют значениям переменной при свободном члене.

В этих обозначениях эмпирическое уравнение регрессии выглядит так:

 (6)

Отсюда вектор остатков регрессии можно выразить таким образом:

 (7)

Таким образом, функционал , который, собственно, и минимизируется по МНК, можно записать как произведение вектора – строки ***е’*** на вектор – столбец ***е***:

 (8)

В соответствии с МНК дифференцирование ***Q*** по вектору ***В*** приводит к выражению:

 (9)

которое для нахождения экстремума следует приравнять к нулю. В результате преобразований получаем выражение для вектора параметров регрессии:

 10)

Здесь  - матрица, обратная к .

**Пример.** Бюджетное обследование пяти случайно выбранных семей дало следующие результаты (в тыс. руб.):

|  |  |  |  |
| --- | --- | --- | --- |
| Семья | Накопления, ***S*** | Доход, ***Y*** | Имущество, ***W*** |
| 1 | 3 | 40 | 60 |
| 2 | 6 | 55 | 36 |
| 3 | 5 | 45 | 36 |
| 4 | 3,5 | 30 | 15 |
| 5 | 1,5 | 30 | 90 |

Оценить регрессию ***S*** на ***Y*** и ***W***.

Введем обозначения:

***S***=[3;6;5;3,5;1,5]’ – вектор наблюдений зависимой переменной;

***B***=[*a;b1;b2*]’ – вектор параметров уравнения регрессии;



- матрица значений независимых переменных.

Далее с помощью матричных операций вычисляем (используем табличный процессор MS Excel и функции ТРАНСП, МУМНОЖ и МОБР в нем):





Регрессионная модель в скалярном виде:



**2. Скалярный метод**. При его применении строится система нормальных уравнений, решение которой и позволяет получить оценки параметров регрессии:

(11)

Решить эту систему можно любым подходящим способом, например, методом определителей или методом Гаусса. При небольшом количестве определяемых параметров использование определителей предпочтительнее.

Рассмотрим пример, приведенный выше. Здесь для двух факторов, *Y* и *W*, система нормальных уравнений запишется так:



Рассчитываем значения сумм, получаем:



Рассчитаем значения определителей этой системы, используем функцию МОПРЕД в Excel:



Отсюда получим оценки параметров модели:



Обратите внимание, что коэффициенты в левой части системы нормальных уравнений совпадают с соответствующими элементами матрицы .