**Модель множественной регрессии**

**1 Понятие множественной регрессии**

Множественная регрессия представляет собой уравнение связи с несколькими независимыми переменными:

***y = f (x1,x2,...,xр),***

где у – зависимая переменная (результативный признак); х1,х2,…,хр – независимые переменные (факторы).

Множественная регрессия применяется в ситуациях, когда из множества факторов, влияющих на результативный признак, нельзя выделить один доминирующий фактор и необходимо учитывать влияние нескольких факторов.

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

Постановка задачи множественной регрессии. По имеющимся данным n наблюдений за совместным изменением p+1 переменной y и xj и {(yi, xj,i) j=1,2,...,p; i=1,2,...,n} (табл.1) необходимо определить аналитическую зависимость ŷ = f (x1,x2,...,xp), наилучшим образом описывающую данные наблюдений.

Таблица 1

Результаты наблюдений

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | y | x1 | x2 | … | xp |
| 1 | *y1* | *x11* | *x21* | *…* | *xp1* |
| 2 | *y2* | *x12* | *x22* | *…* | *xp2* |
| … | *…* | *…* | *…* | *…* | *…* |
| n | *yn* | *x1n* | *x2n* | *…* | *xpn* |

***Критерий качества*** ***выбранной зависимости:***

**S=**$\sum\_{}^{}(y\_{i}-ŷ)$**2 –>min**

Как и в случае парной регрессии, построение уравнения множественной регрессии осуществляется в два этапа:

– спецификация модели;

– оценка параметров выбранной модели.

Спецификация модели включает в себя решение двух задач:

– отбор p факторов xj, наиболее влияющих на величину y;

– выбор вида уравнения регрессии ŷ = f (x1,x2,...,xp);.

**2 Отбор факторов при построении множественной регрессии**

Включение в уравнение множественной регрессии того или иного набора факторов связано, прежде всего, с представлением исследователя о природе взаимосвязи моделируемого показателя с другими экономическими явлениями.

Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям:

1. Факторы не должны быть взаимно коррелированы и, тем более, находиться в точной функциональной связи. Если между факторами существует высокая корреляция, то нельзя определить их изолированное влияние на результативный показатель, и параметры уравнения регрессии оказываются не интерпретируемыми.

2. Включаемые во множественную регрессию факторы должны существенно влиять на вариацию независимой переменной. Т. е. включаемые в модель факторы должны быть статистически значимыми и существенно улучшать показатель качества модели (например, коэффициент детерминации R2).

Отбор факторов производится на основе качественного теоретико-экономического анализа и обычно осуществляется в две стадии:

– на первой стадии факторы подбираются исходя из сущности проблемы;

– на второй стадии применяются формальные статистические критерии, например, значения t-статистики для соответствующих коэффициентов регрессии.

Наличие высокой корреляции выявляется по значению линейного коэффициента корреляции rxixj . Если выполняется условие

rxixj≥0,8,

то факторные переменные xi, xj находятся в линейной зависимости между собой, а сами переменные xi, xj называются явно коллинеарными.

Значения линейных коэффициентов корреляции rxixj для всевозможных комбинаций переменные xi, xj составляют корреляционную матрицу { rxixj }.

Для трех факторов матрица { rxixj } принимает вид:

{ rxixj }=$\left|\begin{matrix}r\_{x1x1}\\r\_{x1x2}\\r\_{x1x3}\end{matrix}\begin{matrix}r\_{x2x1}\\r\_{x2x2}\\r\_{x2x3}\end{matrix}\begin{matrix}r\_{x3x1}\\r\_{x3x2}\\r\_{x3x3}\end{matrix}\right|$

В уравнение регрессии включается только один из коллинеарных факторов, при этом предпочтение отдается тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами.

Для преодоления сильной межфакторной корреляции используется ряд подходов:

– исключение из модели одного или нескольких факторов;

– преобразование факторов, при котором уменьшается корреляция между ними;

– переход к совмещенным уравнениям регрессии, т. е. к уравнениям, которые отражают не только влияние факторов, но и их взаимодействие, например y=a+b1x1+b2x2+b3x3+b12x1x2+b13x1x3+ε, где члены b12x1x2,b13x1x3 выражают взаимодействие факторов.

После исключения коллинеарных факторов осуществляется процедура отбора факторов, наиболее влияющих на изменение результативного признака (факторов, включаемых в регрессию). Наиболее широкое применение получили:

* метод исключения;
* метод включения.

**3 Парная коллинеарность и мультиколлинеарность**

Две переменные считаются явно коллинеарными, т.е. находятся между собой в линейной зависимости, если коэффициент интеркорреляции (корреляции между двумя объясняющими переменными) ≥ 0,7.

Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из уравнения.

Предпочтение в эконометрике отдается не фактору, более сильно связанному с результатом, а фактору, который при сильной связи с результатом имеет наименьшую тесноту связи с другими факторами т.е. коэффициент корреляции между факторами меньше 0,3 или, в идеале, близок к нулю. В этом условии проявляется специфика множественной регрессии как метода исследования комплексного влияния факторов на результат в условиях их независимости друг от друга.

Мультиколлинеарность – линейная зависимость между более чем двумя переменными, т.е. совокупное воздействие факторов друг на друга.

По величине парных коэффициентов корреляции обнаруживается лишь явная коллинеарность факторов. Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии мультиколлинеарности факторов. Чем сильнее мультиколлинеарность факторов, тем менее надежна оценка распределения суммы объясненной вариации по отдельным факторам с помощью метода наименьших квадратов.

Включение в модель мультиколлинеарных факторов нежелательно
по следующим причинам:

* затрудняется интерпретация параметров множественной регрессии; параметры линейной регрессии теряют экономический смысл;
* оценки параметров не надежны, имеют большие стандартные ошибки и меняются с изменением количества наблюдений (не только по величине, но и по знаку), что делает модель непригодной для анализа и прогнозирования.

Для оценки мультиколлинеарности используется определитель матрицы парных коэффициентов интеркорреляции.

Если факторы не коррелируют между собой, то матрица коэффициентов интеркорреляции является единичной, поскольку в этом случае все недиагональные элементы равны 0.

Например, для уравнения с тремя переменными 



*Если между факторами существует полная линейная зависимость* и все коэффициенты корреляции равны 1, то определитель такой матрицы равен 0.



Чем ближе к 0 определитель матрицы коэффициентов интеркорреляции, тем сильнее мультиколлинеарность и ненадежнее результаты множественной регрессии.

Чем ближе к 1определитель матрицы коэффициентов интеркорреляции, тем меньше мультиколлинеарность факторов.

Способы преодоления мультиколлинеарности факторов:

* исключение из модели одного или нескольких факторов;
* переход к совмещенным уравнениям регрессии, т.е. к уравнениям, которые отражают не только влияние факторов, но и их взаимодействие. Например, если ***y = f (x1,x2,...,xр)***, то можно построить следующее совмещенное уравнение:



* переход к уравнениям приведенной формы (в уравнение регрессии подставляется рассматриваемый фактор, выраженный из другого уравнения).

**4 Выбор формы уравнения регрессии**

Кроме точности модели для исследователя наиболее важными качествами модели являются простота модели и возможность наглядной интерпретации параметров модели. По этой причине наиболее широко используются линейная и степенная модели.

В уравнении линейной множественной регрессии:



параметры bi при хi называются коэффициентами «чистой» регрессии и интерпретируется следующим образом. Параметры bi характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизмененном значении других факторов, закрепленных на среднем уровне.

В уравнении степенной множественной регрессии 

показатели степеней bj являются коэффициентами эластичности. Они показывают, на сколько процентов изменяется в среднем результат с изменением соответствующего фактора на 1% при неизменности действия других факторов. Этот вид уравнения регрессии получил наибольшее распространение в производственных функциях, в исследованиях спроса и потребления.

Так же существуют **э**кспоненциальная модель 

и гиперболическая .

**5 Оценка параметров уравнения множественной регрессии**

Для оценки параметров уравнения множественной регрессии применяют метод наименьших квадратов (МНК). Для линейных уравнений регрессии строится система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии.

 или 





В случае линейной множественной регрессии система нормальных уравнений имеет следующий вид:



Решение системы уравнений с помощью метода определителей:

α=$\frac{∆α}{∆}$, b1=$\frac{∆b\_{1}}{∆}$,…, bp=$\frac{∆b\_{p}}{∆}$

где ∆ – определитель системы: 

∆a, ∆b1, ∆bp – частные определители (∆j) , которые получаются из основного определителя путем замены j-го столбца на столбец свободных членов 

**Метод оценки параметров через стандартизованные коэффициенты β**

Уравнение регрессии в стандартизованном (нормированном) масштабе:

, где

,  – стандартизованные переменные;

*β* - стандартизованные коэффициенты регрессии.

β-коэффициенты показывают, на сколько сигм (средних квадратических отклонений) изменится в среднем результат за счет изменения соответствующего фактора xi на одну сигму при неизменном среднем уровне других факторов.

Связь коэффициентов «чистой» регрессии *bi* с коэффициентами *βi* описывается соотношением:

 или 

Коэффициенты *β* определяются при помощи МНК из следующей системы уравнений методом определителей:



Параметр a определяется как: 

**6 Проверка качества уравнения регрессии**

Дисперсионный анализ – самостоятельный инструмент (метод) математической статистики. Кратко рассмотрим схему дисперсионного анализа, представленную в виде таблицы.

|  |  |  |  |
| --- | --- | --- | --- |
| Компоненты дисперсии | Сумма квадратов | Число степеней свободы | Дисперсия |
| Регрессия | http://edu.znate.ru/tw_files2/urls_10/623/d-622741/622741_html_21ab20a2.gif | http://edu.znate.ru/tw_files2/urls_10/623/d-622741/622741_html_md147f6e.gif | http://edu.znate.ru/tw_files2/urls_10/623/d-622741/622741_html_4e430ab8.gif |
| Остаточная | http://edu.znate.ru/tw_files2/urls_10/623/d-622741/622741_html_m53a2ffaf.gif | http://edu.znate.ru/tw_files2/urls_10/623/d-622741/622741_html_7aa3c00f.gif | http://edu.znate.ru/tw_files2/urls_10/623/d-622741/622741_html_m70c6287b.gif |
| Общая | http://edu.znate.ru/tw_files2/urls_10/623/d-622741/622741_html_56091892.gif | http://edu.znate.ru/tw_files2/urls_10/623/d-622741/622741_html_3977b99d.gif |  |

Для проверки значимости модели регрессии используется F-критерий Фишера.



где *m* – число независимых переменных в уравнении

 регрессии;

 *n* – число единиц совокупности.

Если Fфакт > Fтабл, то *Н0* о случайной природе связи отклоняется и признается статистическая значимость и надежность уравнения.

Если Fфакт < Fтабл, то *Н0* не отклоняется и признается статистическая незначимость уравнения регрессии.

Частный F-критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении:

