

Проблема изучения взаимосвязей экономических показателей является одной из важнейших в экономическом анализе. Эконометрика - наука, исследующая количественные закономерности и взаимосвязи в экономике при помощи методов математической статистики. Название «эконометрика» введено в 1926 г. норвежским экономистом и статистиком Р. Фришем. Буквальный перевод этого понятия - «измерения в экономике».

Исследуя природу, общество, экономику, необходимо считаться со взаимосвязью наблюдаемых процессов и явлений. При этом полнота описания так или иначе определяется количественными характеристиками причинно-следственных связей между ними. Оценка наиболее существенных из них, а также воздействия одних факторов на другие является одной из основных задач статистики.

По направлению связи бывают прямыми, когда зависимая переменная растет с увеличением факторного признака, и обратными, при которых рост последнего сопровождается уменьшением функции. Такие связи также можно назвать соответственно положительными и отрицательными.

Относительно своей аналитической формы связи бывают линейными и нелинейными. В первом случае между признаками в среднем проявляются линейные соотношения. Нелинейная взаимосвязь выражается нелинейной функцией, а переменные связаны между собой в среднем нелинейно.

Существует еще одна достаточно важная характеристика связей с точки зрения взаимодействующих факторов. Если характеризуется связь двух признаков, то ее принято называть парной, если изучаются более чем две переменные, -- множественной.

По силе различаются слабые и сильные связи. Эта формальная характеристика выражается конкретными величинами и интерпретируется в соответствии с общепринятыми критериями силы связи для конкретных показателей.

В наиболее общем виде задача эконометрики в области изучения взаимосвязей состоит в количественной оценке их наличия и направления, а также характеристике силы и формы влияния одних факторов на другие. Для ее решения применяются две группы методов, одна из которых включает в себя методы корреляционного анализа, а другая -- регрессионный анализ. Задачи собственно корреляционного анализа сводятся к измерению тесноты связи между варьирующими признаками, определению неизвестных причинных связей и оценке факторов, оказывающих наибольшее влияние на результативный признак.

Задачи регрессионного анализа лежат в сфере установления формы зависимости, определения функции регрессии, использования уравнения для оценки неизвестных значений зависимой переменной.

Решение названных задач опирается на соответствующие приемы, алгоритмы,

показатели, применение которых дает основание говорить о статистическом изучении взаимосвязей и использовать математический аппарат статистики. Основная идея эконометрики - взаимосвязь между экономическими переменными. Обычно рассматриваются переменные, величина которых формируется под воздействием некоторых факторов. В случае наличия взаимосвязи между переменными, одна из них является зависимой или объясняемой переменной (результатирующим показателем), и другие - объясняющими переменными. Установив характер взаимосвязи можно получить ожидаемое значение зависимой переменной при заданных значениях объясняющих переменных, то есть построить эконометрическую модель. Следует подчеркнуть, что конкретные, наблюдаемые значения зависимой переменной в большинстве случаев зависят также от случайных явлений, таких как инфраструктура рынка, характер продавца, наличие конкретной суммы и т.д.

Поэтому, как правило, в каждое соотношение приходится вводить несколько объясняющих переменных и остаточную случайную составляющую, отражающую влияние на результирующий показатель всех неучтенных факторов.

Участвующая в каждом из этих соотношений случайная составляющая, отражающая влияние на результирующий показатель всех неучтенных факторов, обуславливает стохастический (вероятностный) характер зависимости: даже фиксируя значения объясняющих переменных нельзя ожидать однозначно, каким будет значение объясняемой переменной.

Таким образом, общим моментом для любой эконометрической модели является разбиение зависимой переменной на две части - объясненную и случайную:

$$Y = f(X) + l$$

Задача моделирования состоит в том, чтобы на основании экспериментальных данных определить объясненную часть и, рассматривая случайную составляющую как случайную величину, получить (возможно после некоторых предположений) оценки параметров ее распределения.

Наиболее распространенной формализацией стохастической (случайной, вероятностной) зависимости между результирующим показателем y и объясняющими переменными x_1, x_2, \dots, x_n в экономике является аддитивная (получаемая путем сложения) линейная форма:

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + l_t \quad (1.1)$$

где a_0, a_1, \dots, a_n - некоторые параметры (обычно неизвестные до проведения соответствующего статического анализа), $+ l_t$ - случайная составляющая, характеризующая разницу между модельным и наблюдаемым значениями анализируемой результирующей переменной, обнаруженную в t -м измерении. Под модельным (ожидаемым) значением результирующей переменной понимаем ее значение, восстановленное по заданным величинам объясняющих переменных при условии, что коэффициенты a_0, a_1, \dots, a_n нам известны:

$$= a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$$

Случайную составляющую l можно считать случайной ошибкой прогноза y по заданным значениям x_1, \dots, x_n . Обычно полагают, что среднее значение l_t при всех t

равно нулю ($l_t = 0$).

Математическая модель - упрощенное, формализованное представление реальности (объекта, явления, процесса). Количество связей в модели зависит от условий, в которых она конструируется, от подробности объяснения, к которой стремится исследователь.

Все экономические модели имеют общие особенности:

- принимается, что модель улавливает главные характеристики изучаемого объекта;
- полагается, что на основе достигнутого с ее помощью понимания реальной системы удастся предсказать будущее движение экономических показателей.

Результирующая (зависимая) переменная y . Она характеризует результат или эффективность функционирования экономической системы. Значения ее формируются в процессе и внутри функционирования этой системы под воздействием ряда других переменных и факторов, часть из которых поддается регистрации, управлению и планированию (т.е. объясняющие переменные). В регрессионном анализе результирующая переменная играет роль функции, значение которой определяется значениями объясняющих переменных, выполняющих роль аргументов. По своей природе результирующая переменная всегда случайна (стохастична).

Объясняющие переменные X . Это - переменные, поддающиеся регистрации, описывающие условия функционирования реальной экономической системы. Они в значительной мере определяют значения результирующих переменных. Обычно, часть из них поддается регулированию и управлению. Значение этих переменных могут задаваться вне анализируемой системы. Поэтому их называют экзогенными. В регрессионном анализе это аргументы результирующей функции y .

Уравнения регрессионной связи между y и X . Математически уравнение регрессионной связи можно выразить следующим образом:

Здесь $1(X)$ - «остаточная» составляющая («регрессионные остатки»), присутствие которой обусловлено как тем, что она отражает влияние на формирование значений y факторов, не учтенных в наборе объясняющих переменных X , так и тем, что она может включить в себя случайную погрешность в измерении значений y . Второе соотношение в системе означает равенство нулю математического ожидания остаточной составляющей.

Парная регрессия характеризует связь между двумя признаками: результативным и факторным. Аналитическая связь между ними описывается уравнениями:

1. прямой
2. гиперболы
3. параболы

Определить тип уравнения можно, исследуя зависимость графически. Однако существуют более общие указания, позволяющие выявить уравнение связи, не прибегая к графическому изображению. Если результативный и факторный признаки возрастают одинаково, примерно в арифметической прогрессии, то это свидетельствует о том, что связь между ними линейная, а при обратной связи - гиперболическая. Если факторный признак увеличивается в арифметической

прогрессии, а результативный - значительно быстрее, то используется параболическая, или степенная регрессия.

1. Рассмотрим парную линейную регрессию. пусть некоторый экономический процесс характеризуется статистическими данными, в которых представлены соответствующие друг другу выборки n вариантов двух дискретных случайных величин Y и X , записанные в таблице:

i

1

2

...

i

...

n

x_i

x_1

x_2

...

x_i

...

x_n

y_i

y_1

y_2

...

y_i

...

y_n

2. Наглядным изображением корреляционной таблицы служит корреляционное поле. Оно представляет собой график, где на оси абсцисс откладываются значения X , по оси ординат -- Y , а точками показывается сочетание первичных наблюдений X и Y . По расположению точек, их концентрации в определенном направлении можно

судить о наличии связи.

Изобразим полученную зависимость графически точками координатной плоскости, то есть - корреляционное поле.

судя по рисунку, можно предположить, что между X и Y существует корреляционная связь и зависимость Y от X может быть изображена в виде прямой линии, максимально приближенной к точкам корреляционного поля, уравнение которой будет:

Значение называется модельным (теоретическим, расчетным) значением результирующей переменной.

3. Для количественной оценки тесноты связи между переменными Y и X служит линейный коэффициент парной корреляции r_{xy} , вычисляемый по формуле:

, где

, , - средние значения,

, - средние квадратичные отклонения,

, - дисперсия признаков,

В таблице приведены качественные оценки степени тесноты связи между переменными Y и X с помощью коэффициента корреляции.

Таблица. Теснота связи и величина коэффициента корреляции

Коэффициент корреляции

Теснота связи

± 1

Функциональная связь

от $\pm 0,91$ до $\pm 1,00$

Очень сильная

от $\pm 0,81$ до $\pm 0,90$

Весьма сильная

от $\pm 0,65$ до $\pm 0,80$

Сильная

от $\pm 0,45$ до $\pm 0,64$

Умеренная

от $\pm 0,25$ до $\pm 0,44$

Слабая

От 0 до $\pm 0,25$

Очень слабая

0

Связи нет

4. Оценка параметров уравнений регрессии осуществляется методом наименьших квадратов, в основе которого лежит предположение о независимости наблюдений исследуемой совокупности.

Основной принцип метода наименьших квадратов применим для случая, когда две величины (два показателя) X и Y взаимосвязаны между собой, причем Y находится в некоторой зависимости от X . Следовательно, Y будет зависимой, а X - независимой величинами.

Сущность метода наименьших квадратов заключается в нахождении параметров модели (a_0, a_1) . Неизвестные параметры a_0 и a_1 линейного уравнения парной регрессии выбираются таким образом, чтобы сумма квадратов отклонений эмпирических значений y_i от расчетных значений, найденных по уравнению регрессии была минимальной:

Рассматриваем S в качестве функции параметров a_0 , и a_1 , и проводим математические преобразования (дифференцирование). на основании необходимого условия экстремума функции двух переменных $S=S(a_0, a_1)$ приравняем к нулю ее частные производные и получаем:

Откуда система нормальных уравнений для нахождения параметров линейной парной регрессии методом наименьших квадратов имеет следующий вид:

где n - объем исследуемой совокупности (число единиц наблюдений).

Коэффициент a_1 называется выборочным коэффициентом регрессии u по x и характеризует влияние, которое оказывает изменение X на Y . Он показывает на сколько единиц в среднем изменяется переменная Y при увеличении переменной X на одну единицу. Если $a_1 > 0$, то связь - положительная, если $a_1 < 0$, то увеличение X на единицу влечет за собой уменьшение Y в среднем на a_1 .

Коэффициент a_0 - постоянная величина в уравнении регрессии, которую можно рассматривать как начальное значение Y , если в этом есть экономический смысл.

Решая систему найдем формулы для расчета коэффициентов a_0 и a_1 уравнения, а именно:

,

Качество полученной модели оценивается с помощью средней ошибки аппроксимации (приближения), которая показывает среднее отклонение расчетных значений от фактических и определяется по формуле:

,

При $V < 8 - 10\%$ качество модели считается весьма хорошим, а допустимый предел составляют значения в $12 - 15\%$.

При расчетах - исходные данные значений y , a - рассчитанные по формуле (т.е. по полученному уравнению регрессии). Причем одинаковый индекс указывает на то, что эти значения соответствуют одному и тому же значению x и находятся в одной строке.

Проверка адекватности всей модели производится на основе тестирования статистических гипотез.

Статистической гипотезой называется любое предположение о виде и параметре неизвестного закона распределения.

Проверяемую гипотезу обычно называют нулевой и обозначают H_0 . Наряду с нулевой гипотезой H_0 рассматривают альтернативную, или конкурирующую, гипотезу H_1 , являющуюся логическим отрицанием H_0 . Нулевая и альтернативная гипотезы представляют собой две возможности выбора, осуществляемого в задачах проверки статистических гипотез.

Правило, по которому гипотеза h_0 отвергается или принимается, называется статистическим критерием или статистическим тестом. Проверить значимость уравнения регрессии -- значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Проверка значимости уравнения регрессии производится на основе дисперсионного анализа, который применяется как вспомогательное средство для изучения качества регрессионной модели.

На практике F-тест - оценивание качества уравнения регрессии - состоит в проверке гипотезы H_0 о статистической незначимости уравнения регрессии и показателя тесноты связи. Для этого выполняется сравнение фактического $F_{\text{факт}}$ и критического (табличного) $F_{\text{табл}}$ значений F-критерия Фишера. $F_{\text{факт}}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы:

где p - число единиц совокупности; t - число параметров при переменных x .

$F_{\text{табл}}$ - это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости β . Уровень значимости β - вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно β принимается равной 0,05 или 0,01.

Если $F_{табл} < F_{факт}$, то H_0 - гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность. Если $F_{табл} > F_{факт}$, то гипотеза H_0 не отклоняется и признается статистическая незначимость, ненадежность уравнения регрессии.

Для оценки статистической значимости коэффициентов регрессии рассчитываются t -критерий Стьюдента и доверительные интервалы каждого из показателей.

Выдвигается гипотеза H_0 о случайной природе показателей, т.е. о незначимом их отличии от нуля. Оценка значимости коэффициентов регрессии и корреляции с помощью t -критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

Сравнивая фактическое и критическое (табличное) значения t -статистики - $t_{табл}$ и $t_{факт}$ - принимаем или отвергаем гипотезу H_0 .

Если $t_{табл} < t_{факт}$, то H_0 отклоняется, т.е. a_0, a_1 не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x . Если $t_{табл} > t_{факт}$, то гипотеза H_0 не отклоняется и признается случайная природа формирования a_0, a_1 и r_{xy} .

. Одной из наиболее эффективных оценок адекватности регрессионной модели, мерой качества уравнения регрессии, характеристикой прогностической силы анализируемой регрессионной модели является коэффициент детерминации K_d , который показывает измерение степени тесноты статистической связи между y и X .

Для линейных регрессионных связей типа $f(X) = a_0 + a_1x_1 + \dots + a_nx_n$ (где a_0, a_1, \dots, a_n - числовые параметры) коэффициент детерминации совпадает с квадратом множественного коэффициента корреляции $R^2_{y,x}$,

Чем ближе R^2 к единице, тем лучше регрессия аппроксимирует эмпирические данные, тем теснее наблюдения примыкают к линии регрессии. Если $R^2 = 1$, то эмпирические точки (X_j, y_i) лежат на линии регрессии и между переменными Y и X существует линейная функциональная зависимость. Если $R^2 = 0$, то вариация зависимой переменной полностью обусловлена воздействием неучтенных в модели переменных, и линия регрессии параллельна оси абсцисс.

Средний коэффициент эластичности показывает, на сколько процентов в среднем по совокупности изменится результат y от своей средней величины при изменении фактора x на 1% от своего среднего значения:

Множественная (многофакторная) регрессия
статистический гипотеза экономический модель

Изучение связи между тремя и более связанными между собой признаками носит название множественной (многофакторной) регрессии. При исследовании зависимостей методами множественной регрессии задача формулируется так же, как и при использовании парной регрессии, т.е. требуется определить аналитическое выражение связи между результативным признаком (Y) и факторными признаками ($x_1, x_2, x_3, \dots, x_k$), найти функцию: Построение моделей множественной регрессии включает несколько этапов:

* выбор формы связи (уравнения регрессии);

* обеспечение достаточного объема совокупности для получения несмещенных

оценок.

Практика построения многофакторных моделей взаимосвязи показывает, что все реально существующие зависимости между социально-экономическими явлениями можно описать, используя пять типов моделей:

1) линейная: 2) степенная: 3) показательная: 4) параболическая: 5)

гиперболическая: Основное значение имеют линейные модели в силу простоты и логичности их экономической интерпретации. Нелинейные формы зависимости приводятся к линейным путем линеаризации.

Важным этапом построения уже выбранного уравнения множественной регрессии является отбор и последующее включение факторных признаков.

Проблема размерности модели связи, т.е. определение оптимального числа факторных признаков, является одной из основных проблем построения множественного уравнения регрессии. С одной стороны, чем больше факторных признаков включено в уравнение, тем оно лучше описывает явление. Однако модель размерностью 100 и более факторных признаков сложно реализуема и требует больших затрат времени. Сокращение размерности модели за счет исключения второстепенных, экономически и статистически несущественных факторов способствует простоте и качеству ее реализации. В то же время построение модели регрессии малой размерности может привести к тому, что такая модель будет недостаточно адекватна исследуемым явлениям и процессам.

Наиболее приемлемым способом отбора факторных признаков является шаговая регрессия (шаговый регрессионный анализ). Сущность метода шаговой регрессии заключается в последовательном включении факторов в уравнение регрессии и последующей проверке их значимости. Факторы поочередно вводятся в уравнение так называемым "прямым методом". При проверке значимости введенного фактора определяется, насколько уменьшается сумма квадратов остатков и увеличивается величина множественного коэффициента корреляции (R). Одновременно используется и обратный метод, т.е. исключение факторов, ставших незначимыми на основе t -критерия Стьюдента. Фактор является незначимым, если его включение в уравнение регрессии только изменяет значение коэффициентов регрессии, не уменьшая суммы квадратов остатков и не увеличивая их значения. Если при включении в модель соответствующего факторного признака величина множественного коэффициента корреляции увеличивается, а коэффициент регрессии не изменяется (или меняется несущественно), то данный признак существенен и его включение в уравнение регрессии необходимо.

Если же при включении в модель факторного признака коэффициенты регрессии меняют не только величину, но и знаки, а множественный коэффициент корреляции не возрастает, то данный факторный признак признается нецелесообразным для включения в модель связи.

Качество уравнения регрессии зависит от степени достоверности и надежности исходных данных и объема совокупности. Исследователь должен стремиться к увеличению числа наблюдений, так как большой объем наблюдений является одной из предпосылок построения адекватных статистических моделей.

Аналитическая форма выражения связи результативного признака и ряда факторных называется многофакторным (множественным) уравнением регрессии, или моделью связи.