

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное  
учреждение высшего образования «Казанский (Приволжский)  
федеральный университет»

Набережночелнинский институт (филиал)

**Кафедра Бизнес-информатики и математических методов в  
экономике**

**Эконометрическое моделирование  
(методические указания к занятиям)**

*Учебно-методическое пособие*

Набережные Челны  
2019 г.

УДК 519.712.1  
ББК 22.162

Печатается по решению учебно-методической комиссии экономического отделения Набережночелнинского института (филиала) федерального государственного автономного образовательного учреждения высшего образования «Казанский (Приволжский) федеральный университет», от «26» февраля 2019г. (протокол №6)

Рецензенты:

Доктор физ.-мат. наук, профессор А.Г. Исавнин

Доктор экономических наук, профессор А.Н. Макаров

Розенцвайг А.К., Исавнин А.Г. Эконометрическое моделирование (методические указания к занятиям): учебно-методическое пособие / А.К. Розенцвайг, А.Г. Исавнин – Набережные Челны: Изд-во Набережночелнинского института КФУ, 2019. – 54 с.

Учебно-методическое пособие содержит последовательное изложение базовых понятий теории методических указаний к практическим занятиям в эконометрическом моделировании. Подробно изложены: особенности построения эконометрической модели, множественная линейная регрессия, анализ качества уравнения регрессии, стандартизованная форма уравнения множественной линейной регрессии, возможности экономического анализа на основе многофакторной регрессионной модели, порядок выполнения и оформления типового примера.

Учебно-методическое пособие предназначено для использования в учебном процессе студентами технических направлений в экономике и экономического отделения дневной, заочной и дистанционной форм обучения.

© Розенцвайг А.К., Исавнин А.Г., 2019

© НЧИ КФУ, 2019

© Кафедра Бизнес-информатики и  
математических методов в экономике, 2019 г.

## Введение

Постоянно усложняющиеся процессы и явления реальной экономической жизни привели к необходимости создания и совершенствования особых методов их изучения и анализа. При этом широкое распространение получило использование моделирования и количественного анализа. На базе последних выделилось и сформировалось одно из направлений экономических исследований — эконометрика.

Формально «эконометрика» означает «измерения в экономике». Однако область исследований данной дисциплины гораздо шире. *Эконометрика* — это наука, в которой на базе реальных статистических данных строятся, анализируются и совершенствуются математические модели реальных экономических явлений. Обоснование и интерпретация самих моделей является задачей экономической теории, а эконометрика позволяет найти количественное подтверждение либо опровержение того или иного существующего экономического закона либо новой теоретической гипотезы. Одним из важнейших направлений эконометрики является построение прогнозов по различным экономическим показателям.

Эконометрика как научная дисциплина зародилась и получила развитие на основе слияния экономической теории, математической экономики, экономической и математической статистики. Действительно, предмет ее исследования — экономические явления. Но в отличие от экономической теории эконометрика делает упор на количественные, а не на качественные аспекты этих явлений.

Например, экономическая теория утверждает, что спрос на товар с ростом его цены убывает. Но при этом практически неисследованным остается вопрос, как быстро и по какому закону происходит это убывание. Эконометрика отвечает на этот вопрос для каждого конкретного вида товара, места и времени его реализации.

Изучение экономических процессов (взаимосвязей) в эконометрике

осуществляется через математические (эконометрические) модели. В этом состоит ее родство с математической экономикой. Но если математическая экономика строит и анализирует обобщенные модели без использования реальных числовых значений, то эконометрика концентрируется на изучении моделей, обоснованных экономической теорией, и сопоставлении их с реальными статистическими данными.

Одной из основных задач экономической статистики является сбор, обработка и представление экономических данных в наглядной форме: в виде таблиц, графиков, диаграмм. Эконометрика также активно пользуется этим инструментарием, но идет дальше, применяя его для анализа теоретически обоснованных экономических взаимосвязей и прогнозирования.

Мощным инструментом эконометрических исследований является аппарат математической статистики. Действительно, большинство экономических показателей носит характер случайных величин, предсказать точные значения которых практически невозможно. Например, весьма сложно предвидеть доход или потребление какого-либо индивидуума, объемы экспорта и импорта страны в течение следующего года и т.д.

Связи между экономическими показателями также не могут носить строгий функциональный характер, а допускают наличие случайных неконтролируемых отклонений (особенно это касается макроэкономических данных). Вследствие этого использование методов математической статистики в эконометрике является естественным и вполне обоснованным. Однако в силу специфики получения статистических данных в экономике (например, в экономике невозможно проведение управляемого эксперимента) эконометрике приходится создавать свои собственные наработки и специальные приемы анализа, которые в математической статистике обычно не рассматривают.

При эконометрическом исследовании имеют место две стороны проблемы обеспечения необходимого качества его результатов – качественная и количественная. Качественная сторона заключается в установлении

соответствия между построенной эконометрической моделью и лежащими в ее основе положениями экономической теории. Другая – количественная - состоит в обеспечении наиболее полного соответствия между количественными характеристиками модели и статистической информацией, характерной для поведения изучаемых социально – экономических явлений и процессов в реальных условиях.

## 1. Особенности построения эконометрических моделей

### 1.1. Основные этапы эконометрического исследования

Построение эконометрической модели — центральная проблема любого эконометрического исследования, поскольку ее «качество» определяет достоверность и обоснованность результатов анализа тенденций развития, прогнозов рассматриваемых социально-экономических процессов, а также вытекающих из них выводов, в том числе и по вопросам разработки необходимых управленческих мероприятий.

В эконометрических исследованиях обычно предполагается, что закономерности моделируемого процесса складываются под влиянием ряда других явлений, факторов. Обобщенную форму эконометрической модели, описывающей закономерности развития такого процесса, обозначенного переменной  $y$ , в зависимости от уровней, воздействующих на него внешних явлений, факторов  $X_i$ ,  $i = 1, 2, \dots, p$ , можно представить следующим уравнением:

$$y = f(b_0, b_1, \dots, b_p, X_1, X_2, \dots, X_p) + \varepsilon, \quad (1.1)$$

выражающий вид и структуру взаимосвязей между уровнями переменных  $y$ , и  $X_i$ ;  $\varepsilon$ , — случайная ошибка модели, в отношении свойств и характеристик которой, как это будет показано далее, выдвигается ряд дополнительных предположений.

Эконометрика допускает различные предположения относительно «статистического» содержания факторных переменных  $X_i$ , в то время как переменная  $y$  согласно (1.1) всегда рассматривается как случайная величина.

Использование той или иной интерпретации значений переменных эконометрических моделей, как правило, не вносит принципиальные изменения в процедуры их построения, в методы оценки их параметров, но часто сказывается на свойствах полученных результатов.

Выражение (1.1) определяет лишь общий вид эконометрической модели. В конкретных эконометрических исследованиях могут использоваться специальные типы моделей, каждый из которых имеет свои характерные особенности. Эти типы обычно можно классифицировать на основе двух признаков. Во-первых, по виду факторов  $X_i$ , во-вторых, по свойствам ошибки модели.

В частности, в моделях регрессии классического типа обычно используются факторы, независимые между собой. Также предполагается, что ошибка модели имеет свойства «белого шума» — случайного процесса с нулевым математическим ожиданием, постоянной конечной дисперсией и нулевой корреляцией между ее разновременными значениями. Это означает, что в ряду ошибки  $\varepsilon$  отсутствуют автокорреляционные связи.

Модели могут различаться и по характеру связей факторов с переменной  $y$ . По этому признаку их разделяют на линейные и нелинейные модели. Эконометрические модели могут различаться и по свойствам своих параметров (модели с постоянной и переменной структурой), и по целому ряду других признаков.

Характерная особенность эконометрического исследования заключается в том, что зачастую априорно наиболее подходящий для рассматриваемого процесса тип модели определить не представляется возможным. Но при этом, как правило, на основе содержательного анализа рассматриваемого явления обычно удается выделить приемлемые альтернативные варианты модели и сформировать их исходные предпосылки. По результатам этапов эконометрического исследования эти варианты уточняются, и среди них выбирается тот, который в большей степени соответствует рассматриваемому процессу, явлению.

В общем случае процедуру построения эконометрической модели можно разделить на несколько взаимосвязанных между собой этапов. На каждом из них последовательно решаются задачи, которые имеют следующее содержание:

1. *Спецификация модели* - анализ специфических свойств рассматриваемых явлений и процессов (предметной области) и обоснование типа моделей, наиболее подходящих для их описания. Отметим, что в общем случае целями этого этапа являются:

1.1. Обоснование формы модели, выражаемой общим математическим уравнением (системой уравнений), связывающим включенные в модель переменные и содержащим неизвестные параметры (коэффициенты).

1.2. Выбор рационального состава включаемых в модель переменных и определение количественных характеристик, отражающих их уровни в прошлые периоды времени (на однородных объектах некоторой совокупности — территориях, предприятиях и т.п.).

2. *Параметризация модели* - оценка параметров выбранного варианта модели на основании исходных статистических данных, выражающих уровни показателей (переменных) на пространственной совокупности однородных объектов или в различные моменты времени.

3. *Верификация модели* - проверка качества построенной модели и обоснование вывода о целесообразности ее использования в ходе дальнейшего эконометрического исследования, а также для объяснения поведения изучаемых экономических показателей, прогнозирования и предсказания их поведения.

4. При выводе о нецелесообразности использования построенной эконометрической модели в дальнейших исследованиях следует вернуться к первому (или какому-либо другому этапу) и попытаться выбрать более качественную модификацию модели (другой вариант модели).

Выделенные этапы построения моделей достаточно условны и отражают циклический характер современных экономических исследований: от экономической теории к моделированию; от моделирования к

совершенствованию теории и более глубокому пониманию сути происходящих процессов; от понимания сути к осуществлению продуманной и целенаправленной экономической политики. Состав используемых на них процедур, приемов и методов, их очередность зависят от типа разрабатываемой эконометрической модели, особенностей исследуемых экономических процессов, свойств исходных данных и т.п.

## 1.2. Обоснование формы эконометрической модели

Основные подходы к решению проблем первого этапа исследования в значительной степени базируются на методах содержательного анализа закономерностей рассматриваемых процессов, подкрепляемых по мере необходимости методами общей и математической статистики. Дело в том, что обычно в практических исследованиях функциональный вид эконометрической модели может быть не известным. Часто рассматривают несколько альтернативных их вариантов, среди которых необходимо выбрать наиболее подходящий как с точки зрения требований экономической теории, так и необходимой точности аппроксимации функциональным выражением (1.1) исходного ряда измеренных значений зависимой переменной  $y$ .

В этой связи, прежде чем подойти к решению задач первого этапа, необходимо сформировать хотя бы предварительные исходные предпосылки экономического и математического содержания в отношении вида функциональной зависимости. Здесь можно отметить два возможных подхода. В одних случаях составом переменных  $X_i$ ,  $i = 1, 2, \dots, p$  и формой зависимости (1.1), отражают общепринятую экономическую концепцию. В других - выявляют эмпирические взаимосвязи между ними в ходе конкретных исследований статистических данных.

Примером первого подхода может служить двухфакторная модель Кобба-Дугласа

$$y = b_0 \cdot X_1^{b_1} \cdot X_2^{b_2} + \varepsilon, \quad (1.2)$$

которую применяют в макроэкономических исследованиях взаимосвязи между объемом полученного валового национального продукта ( $y$ ) и используемыми производственными ресурсами ( $X_1$  – основные фонды,  $X_2$  – затраты живого труда). Для параметров этой модели известна содержательная экономическая интерпретация, их можно сопоставлять с имеющимися результатами аналогичных исследований.

Второй подход связан с привлечением исходных статистических данных для построения эконометрической модели. В общем случае ими являются известные наборы (массивы) значений зависимой переменной  $y$  и независимых факторов  $X_i$ . Различают два принципиально различных типа исходных информационных массивов — статический и динамический.

Статический массив представляет собой значения результирующей (зависимой, объясняемой и т.п.) переменной  $y$  и влияющих на нее факторов (независимых, объясняющих переменных)  $X_i$ , имевших место у объектов однородной совокупности в определенный период времени. Пример таких объектов — однотипные промышленные предприятия (заводы одной отраслевой направленности). В качестве  $y$  в практических исследованиях часто рассматривают показатели производительности труда, объемов выпускаемой продукции и некоторые другие. В качестве  $X_i$  — влияющие на уровень этих показателей факторы — объемы используемых фондов, численность и квалификация рабочей силы и т.п.

Приведем другой пример статической информации, характерной для социальных исследований. В качестве  $y$  часто рассматривают показатели заболеваемости (смертности) населения в регионах страны. Их уровень в каждом из регионов определяют значения независимых факторов, отражающих достигнутый материальный уровень жизни, климатические условия, состояние окружающей среды и т.п. В этом случае необходимая для построения эконометрической модели информация собирается по совокупности регионов страны за определенный фиксированный промежуток времени, например год.

В общем случае будем считать, что необходимая для построения

эконометрической модели базового типа (1.1) статическая информация выражается следующими массивами взаимосвязанных (взаимосоответствующих) наборов данных: Такие наборы называют *пространственными данными (cross-sectional data)* или пространственной выборкой:

$$\begin{array}{ccccccc}
 y_1 & x_{11} & x_{21} & \cdots & x_{i1} & \cdots & x_{p1} \\
 y_2 & x_{12} & x_{22} & \cdots & x_{i2} & \cdots & x_{p2} \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 y_j & x_{1j} & x_{2j} & \cdots & x_{ij} & \cdots & x_{pj} \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 y_n & x_{1n} & x_{2n} & \cdots & x_{in} & \cdots & x_{pn}
 \end{array} ,$$

где  $y_j$  — уровень зависимой переменной на  $j$ -ом объекте совокупности;  $x_{ij}$  — уровень  $i$ -го фактора на  $j$ -ом объекте статистической совокупности;  $j = 1, 2, \dots, n$ ;  $i = 1, 2, \dots, p$ .

Динамическую информацию, которая связывает значения некоторой зависимой переменной  $y$  в моменты времени  $t$ , называют *временным (динамическим) рядом (time-series data)*.

Исходная информация для построения эконометрических моделей может быть и смешанного типа. Если выборка выражает уровни интересующих нас переменных  $x_{it}$  по группе заводов за ряд лет, ее называют *панельными данными (panel data)*.

В случае, когда периодов времени наблюдений  $T$  больше, чем число объектов  $p$ , панельные данные называют *объединенным временным рядом (pooled time-series)*. Они связывают значения некоторой зависимой переменной  $y$  в моменты времени  $t$  со значениями независимых переменных (факторов)  $x_{it}$ , рассматриваемых в те же моменты времени,  $t = 1, 2, \dots, T$ .

В ходе содержательного анализа явление часто рассматривается на качественном уровне, и оперируют достаточно обобщенными понятиями, например, заболеваемость, уровень медицинского обслуживания, качество и уровень жизни, климат, качество рабочей силы и т.п. Часто эконометрическая модель строится именно для выражения закономерности, существующей между

явлениями. При построении модели используется исходная информация, наборы показателей, которые уже выражают эти зависимости, их свойства, тенденции в виде количественных характеристик. Следовательно, модельное выражение должно представлять эти реально существующие взаимосвязи.

Для традиционных направлений исследований проблема обоснования состава показателей обычно считается решенной. Например, в макроэкономических исследованиях производительности труда, обычно рассматриваются регламентированные, уже устоявшиеся наборы показателей, значения которых публикуются в статистических сборниках, научных отчетах и других официальных изданиях Госкомстата. Такие как выработка на одного работающего как показатель, выражающий явление «производительность труда», объемы ВВП - показатель результативности экономики), объемы основных фондов - показатель уровня материальной обеспеченности производственного процесса, экономики) и ряд других.

Вместе с тем в ряде областей эконометрических исследований такие системы показателей не могут быть сформированы столь однозначно. Часто одно и то же явление может быть выражено альтернативными вариантами показателей.

В отсутствие объективных данных в эконометрических исследованиях допускается замена одного показателя другим, косвенно отражающим то же явление. Например, среднедушевой доход как показатель материального уровня жизни может быть заменен среднегодовым товарооборотом на одного жителя региона. Неправильный выбор показателя, представляющего рассматриваемое явление в модели, может существенно повлиять на качество эконометрической модели.

### 1.3. Выбор факторов эконометрической модели

Проблема обоснования «оптимального» набора факторов обычно решается на основе как содержательного (теоретического), так и

количественного (статистического) анализа тенденций рассматриваемых процессов.

На этапе содержательного анализа решается вопрос о целесообразности включения в модель тех или иных факторов, исходя из их экономического смысла. В макроэкономических исследованиях состав факторов, как правило, определяется на основании допущений экономической теории. Пример — двухфакторные производственные функции типа Кобба-Дугласа, которые строятся в предположении, что объем выпуска (производства) экономической системы в основном зависит от размеров используемых основных фондов и количества затраченного труда. Функция типа Кобба-Дугласа учитывает теоретическое предположение о постоянной эластичности выпуска по каждому из производственных факторов.

На этапе содержательного анализа обычно решается проблема установления самого факта наличия взаимосвязей между явлениями. Однако каждое из явлений может быть представлено разными наборами факторами и даже их комбинациями. Поэтому в ряде исследований на основании содержательного анализа однозначно состав независимых переменных модели определить практически невозможно. Могут существовать их альтернативные наборы.

Например, для исследования закономерностей динамики производительности труда на заводе могут быть отобраны следующие факторы: объем основных фондов, энерговооруженность труда, фондовооруженность труда, численность рабочей силы, ее квалификация. При этом квалификация как явление может выражаться разными показателями, например, средним уровнем образования работников, их усредненным квалификационным разрядом и т.п. Кроме того, можно ожидать, что показатели энерговооруженности, фондовооруженности труда, объема основных фондов характеризуют одно и то же явление — уровень материально-технической оснащенности производственного процесса. Таким образом, некоторые из рассматриваемых в таком исследовании показателей, выражающих

количественные характеристики независимых переменных, относятся к сходным явлениям.

Факторы, выражающие одну и ту же причину, могут быть тесно взаимосвязаны между собой. Так, уровень розничного товарооборота в основном зависит от среднедушевого дохода; концентрация загрязняющих веществ — от объемов их выбросов. Вследствие этого одновременное включение таких факторов в модель вряд ли целесообразно, поскольку, таким образом, одна и та же причина будет учтена дважды. В результате в общем случае на этапе обоснования эконометрической модели решается задача выбора наиболее предпочтительного состава независимых факторов среди ряда альтернативных вариантов.

Можно выделить два основных подхода к решению этой проблемы. Первый предполагает априорное (до построения модели) исследование характера и силы взаимосвязей между рассматриваемыми переменными, по результатам которого в модель включаются факторы, наиболее значимые по своему «непосредственному» влиянию на зависимую переменную  $y$ . И, наоборот, из модели исключаются факторы, которые, либо малозначимы с точки зрения силы своего влияния на эту переменную, либо их сильное влияние на нее обусловлено индуцированными взаимосвязями с другими переменными.

В основе «априорного» подхода лежат следующие предположения.

1. Сильное влияние фактора на зависимую переменную должно подтверждаться определенными количественными характеристиками, важнейшей является их парный линейный коэффициент корреляции, выборочное значение которого рассчитывается на основании имеющейся информации.

Логика использования коэффициента парной корреляции при отборе значимых факторов на практике состоит в следующем. Если значение коэффициента корреляции достаточно велико, т.е. превосходит некоторый эмпирический рубеж (на практике 0,5-0,6), то можно говорить о наличии существенной линейной связи между переменными  $y$  и  $X_i$ , или о достаточно

сильном влиянии  $X_i$  на  $y$ . Чем больше абсолютное значение  $r_{yxi}$ , тем сильнее это влияние (положительное или отрицательное, в зависимости от знака коэффициента парной корреляции).

2. Если два и более факторов выражают одно и то же явление, то, как правило, между ними также должна существовать достаточно сильная взаимосвязь. На это может указать выборочное значение их парного коэффициента корреляции. На практике взаимосвязь между факторами признается существенной, если их коэффициент корреляции достигает величины 0,8-0,9. В таких ситуациях один из этих факторов целесообразно исключить из модели, чтобы одна и та же причина не учитывалась дважды. Однако такое исключение следует проводить только в тех случаях, когда факторы выражают одно и то же явление.

Приведенные рубежные значения (в первом случае — 0,5-0,6; во втором — 0,8-0,9) достаточно условны. В каждом конкретном случае они устанавливаются индивидуально. Значительно усложняет проблему отбора факторов явление *ложной корреляции*, которое характеризуется достаточно высокими по абсолютной величине значениями коэффициентов парной корреляции с содержательной точки зрения между собой никак не связанных факторов. Иными словами, большие значения парных коэффициентов корреляции могут иметь место и в тех случаях, когда тенденции рассматриваемых процессов совпали случайно, при отсутствии между ними взаимосвязи, обоснованной представлениями соответствующей экономической теории.

Ложная корреляция может помешать при построении «правильной» модели по двум причинам. Во-первых, в модель случайно могут быть введены незначимые с содержательной точки зрения факторы, характеризующиеся значимыми величинами коэффициента парной корреляции. Во-вторых, из модели могут быть исключены значимые с точки зрения влияния на  $y$  факторы, в отношении которых ошибочно признана гипотеза о том, что они выражают то же явление, что и другой фактор (факторы), уже включенный в эту модель.

Среди основных причин включения в модель переменных с ложной корреляцией часто называют ненадежность информации, используемой при определении значений факторов в различные моменты времени, трудности формализации факторов, имеющих качественный характер, неустойчивость тенденций изменения рассматриваемых переменных, неправильную форму взаимосвязи между ними и т.п. Основным путем, придерживаясь которого можно избежать ошибок, связанных с понятием «ложной корреляции», связан с проведением качественного анализа проблемы, направленного на обоснование адекватного ей содержания и формы модели.

Второй подход к отбору независимых факторов — можно назвать апостериорным — предполагает первоначально включить в модель все отобранные на этапе содержательного анализа факторы. Уточнение их состава в этом случае производится на основе анализа характеристик качества построенной модели и силы влияния каждого из факторов на зависимую переменную.

Если фактор  $X_i$  признается незначимым, его целесообразно удалить из модели. Эта операция приводит к уменьшению общего количества независимых переменных в модели. Таким образом, на практике используют следующую поэтапную процедуру построения окончательного варианта модели на основе апостериорного подхода:

1. В исходный вариант модели включаются все факторы, отобранные в ходе содержательного анализа проблемы. Рассчитывают значения оценок коэффициентов модели, их среднеквадратические ошибки и значения критериев Стьюдента.

2. Из модели удаляют незначимый фактор, характеризующийся наименьшим значением критерия Стьюдента, при условии, что он статистически незначим и формируют новый вариант модели с уменьшенным на один числом факторов.

Заметим, что в модели может быть несколько незначимых факторов. Однако все их одновременно удалять не следует. Возможно, что недостаточная

значимость большинства факторов обусловлена влиянием «наихудшего» из незначимых факторов и на следующем шаге расчетов они окажутся значимыми.

3. Процесс отбора факторов считают законченным, когда остающиеся в модели факторы являются значимыми, если полученный вариант модели удовлетворяет и другим критериям ее качества, то процесс построения модели можно считать завершенным в целом.

В противном случае попытаются сформировать другой альтернативный вариант модели, отличающийся от предыдущего либо составом факторов, либо формой их взаимосвязи с зависимой переменной.

Каждый из этих подходов имеет свои преимущества и недостатки. «Априорный» путь отбора факторов не обладает достаточной обоснованностью. Он в большей степени использует «прямые» количественные индикаторы «силы» взаимосвязей между рассматриваемыми величинами и не принимает во внимание в полной мере особенности комплексного влияния независимых факторов на переменную  $y$  т.е. своеобразные эффекты «эмерджентности» такого влияния.

Этот эффект выражается в том, что совокупное воздействие нескольких факторов на переменную  $y$ , может значительно отличаться от суммы воздействий каждого из них именно в силу наличия внутренних взаимосвязей между независимыми переменными. Вместе с тем использование априорного подхода часто позволяет уточнить некоторые предварительные альтернативные варианты наборов независимых факторов, проверить исходные предпосылки модели относительно правильности выбора формы взаимосвязей между ними.

«Апостериорный» подход к отбору факторов, на первый взгляд, предпочтительнее из-за того, что целесообразность включения в модель каждого из факторов определяется на основании всего комплекса взаимосвязей между переменными. Однако когда общее количество факторов достаточно велико, нет никаких гарантий того, что множество несущественных, а то и ложных взаимосвязей между ними не будет превалировать над основными связями. В результате может оказаться, что в числе первых кандидатов на

исключение будут «названы» наиболее важные, значимые с точки зрения влияния на переменную  $y$ , факторы. Поэтому в сложных случаях, т.е. при наличии большого числа отобранных для включения в модель на этапе содержательного анализа факторов, полезно сочетать при обосновании их «оптимального» состава оба подхода, как априорный, так и апостериорный.

## 2. Множественная линейная регрессия

Множественный регрессионный анализ является расширением парного регрессионного анализа. Он применяется в тех случаях, когда поведение объясняемой, зависимой переменной необходимо связать с влиянием более чем одной факторной, независимой переменной. Хотя определенная часть многофакторного анализа представляет собой непосредственное обобщение понятий парной регрессионной модели, при выполнении его может возникнуть ряд принципиально новых задач.

Так, при оценке влияния каждой независимой переменной необходимо уметь разграничивать ее воздействие на объясняемую переменную от воздействия других независимых переменных. При этом множественный корреляционный анализ сводится к анализу парных, частных корреляций. На практике обычно ограничиваются определением их обобщенных числовых характеристик, таких как частные коэффициенты эластичности, частные коэффициенты корреляции, стандартизованные коэффициенты множественной регрессии.

Затем решаются задачи спецификации регрессионной модели, одна из которых состоит в определении объема и состава совокупности независимых переменных, которые могут оказывать влияние на объясняемую переменную. Хотя это часто делается из априорных соображений или на основании соответствующей экономической (качественной) теории, некоторые переменные могут в силу индивидуальных особенностей изучаемых объектов

не подходить для модели. В качестве наиболее характерных из них можно назвать *мультиколлинеарность* или *автокоррелированность* факторных переменных.

## 2.1. Анализ множественной линейной регрессии с помощью метода наименьших квадратов (МНК)

В данном разделе полагается, что рассматривается модель регрессии, которая специфицирована правильно. Обратное, если исходные предположения оказались неверными, можно установить только на основании качества полученной модели. Следовательно, этот этап является исходным для проведения множественного регрессионного анализа даже в самом сложном случае, поскольку только он, а точнее его результаты могут дать основания для дальнейшего уточнения модельных представлений. В таком случае выполняются необходимые изменения и дополнения в спецификации модели, и анализ повторяется после уточнения модели до тех пор, пока не будут получены удовлетворительные результаты.

На любой экономический показатель в реальных условиях обычно оказывает влияние не один, а несколько и не всегда независимых факторов. Например, спрос на некоторый вид товара определяется не только ценой данного товара, но и ценами на замещающие и дополняющие товары, доходом потребителей и многими другими факторами. В этом случае вместо парной регрессии  $M(Y/X=x) = f(x)$  рассматривается множественная регрессия

$$M(Y/X_1 = x_1, X_2 = x_2, \dots, X_p = x_p) = f(x_1, x_2, \dots, x_p) \quad (2.1)$$

Задача оценки статистической взаимосвязи переменных  $Y$  и  $X_1, X_2, \dots, X_p$  формулируется аналогично случаю парной регрессии. Уравнение множественной регрессии может быть представлено в виде

$$Y = f(\mathbf{B}, \mathbf{X}) + \varepsilon \quad (2.2)$$

где  $\mathbf{X}$  — вектор независимых (объясняющих) переменных;  $\mathbf{B}$  — вектор

параметров уравнения (подлежащих определению);  $\varepsilon$  - случайная ошибка (отклонение);  $Y$  — зависимая (объясняемая) переменная.

Предполагается, что для данной генеральной совокупности именно функция  $f$  связывает исследуемую переменную  $Y$  с вектором независимых переменных  $X$ .

Рассмотрим самую употребляемую и наиболее простую для статистического анализа и экономической интерпретации модель множественной линейной регрессии. Для этого имеются, по крайней мере, две существенные причины.

Во-первых, уравнение регрессии является линейным, если система случайных величин  $(X_1, X_2, \dots, X_p, Y)$  имеет совместный нормальный закон распределения. Предположение о нормальном распределении может быть в ряде случаев обосновано с помощью предельных теорем теории вероятностей. Часто такое предположение принимается в качестве гипотезы, когда при последующем анализе и интерпретации его результатов не возникает явных противоречий.

Вторая причина, по которой линейная регрессионная модель предпочтительней других, состоит в том, что при использовании ее для прогноза риск значительной ошибки оказывается минимальным.

Теоретическое линейное уравнение регрессии имеет вид:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_p X_p + \varepsilon, \quad (2.3)$$

или для индивидуальных наблюдений с номером  $i$ :

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip} + \varepsilon_i \quad (2.4)$$

где  $i = 1, 2, \dots, n$ .

Здесь  $\mathbf{B} = (b_0, b_1, \dots, b_p)$  — вектор размерности  $(p+1)$  неизвестных параметров  $b_j, j = 0, 1, 2, \dots, p$ , называется  $j$ -ым теоретическим коэффициентом регрессии (частичным коэффициентом регрессии). Он характеризует чувствительность величины  $Y$  к изменению  $X_j$ . Другими словами, он отражает влияние на условное математическое ожидание  $M(Y/X_1 = x_1, X_2 = x_2, \dots, X_p = x_p)$  зависимой переменной  $Y$  объясняющей переменной  $X_j$  при условии, что все

другие объясняющие переменные модели остаются постоянными.  $b_0$  — свободный член, определяющий значение  $Y$  в случае, когда все объясняющие переменные  $X_j$  равны нулю.

После выбора линейной функции в качестве модели зависимости необходимо оценить параметры регрессии.

Пусть имеется  $n$  наблюдений вектора объясняющих переменных  $X = (1, X_1, X_2, \dots, X_p)$  и зависимой переменной  $Y$ :

$$(1, x_{i1}, x_{i2}, \dots, x_{ip}, y_i), \quad i = 1, 2, \dots, n.$$

Для того чтобы однозначно можно было бы решить задачу отыскания параметров  $b_0, b_1, \dots, b_p$  (т.е. найти некоторый наилучший вектор  $\mathbf{B}$ ), должно выполняться неравенство  $n > p + 1$ . Если это неравенство не будет выполняться, то существует бесконечно много различных векторов параметров, при которых линейная формула связи между  $X$  и  $Y$  будет абсолютно точно соответствовать имеющимся наблюдениям. При этом, если  $n = p + 1$ , то оценки коэффициентов вектора  $\mathbf{B}$  рассчитываются единственным образом — путем решения системы  $p + 1$  линейного уравнения:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip}, \quad (2.5)$$

где  $i = 1, 2, \dots, n$ .

Например, для однозначного определения оценок параметров уравнения регрессии  $Y = b_0 + b_1 X_1 + b_2 X_2$  достаточно иметь выборку из трех наблюдений  $(1, x_{i1}, x_{i2}, y_i), i = 1, 2, 3$ . В этом случае найденные значения параметров  $b_0, b_1, b_2$  определяют такую плоскость  $Y = b_0 + b_1 X_1 + b_2 X_2$  в трехмерном пространстве, которая пройдет именно через имеющиеся три точки.

С другой стороны, добавление в выборку к имеющимся трем наблюдениям еще одного приведет к тому, что четвертая точка  $(x_{41}, x_{42}, x_{43}, y_4)$  практически всегда будет лежать вне построенной плоскости (и, возможно, достаточно далеко). Это потребует определенной переоценки параметров.

Таким образом, вполне логичен следующий вывод: если число наблюдений больше минимально необходимой величины, т.е.  $n > p + 1$ , то уже нельзя подобрать линейную форму, в точности удовлетворяющую всем

наблюдениям. Поэтому возникает необходимость оптимизации, т.е. оценивания параметров  $b_0, b_1, \dots, b_p$ , при которых формула регрессии дает наилучшее приближение одновременно для всех имеющихся наблюдений.

В данном случае число  $\nu = n - p - 1$  называется числом степеней свободы. Нетрудно заметить, что если число степеней свободы невелико, то статистическая надежность оцениваемой формулы невысока. Например, вероятность надежного вывода (получения наиболее реалистичных оценок) по трем наблюдениям существенно ниже, чем по тридцати. Считается, что при оценивании множественной линейной регрессии для обеспечения статистической надежности требуется, чтобы число наблюдений превосходило число оцениваемых параметров, по крайней мере, в 3 раза.

Прежде чем перейти к описанию алгоритма нахождения оценок коэффициентов регрессии, отметим желательность выполнимости ряда предпосылок МНК, которые позволят обосновать характерные особенности регрессионного анализа в рамках классической линейной многофакторной модели.

## 2.2. Теоретические предпосылки МНК

1°. *Математическое ожидание случайного отклонения  $\varepsilon_i$  равно нулю для всех наблюдений:*

$$M(\varepsilon_i) = 0, \quad i = 1, 2, \dots, n.$$

2°. *Наличие гомоскедастичности (постоянство дисперсии случайных отклонений).* Дисперсия случайных отклонений  $\varepsilon_i$  должна быть постоянной:

$$D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2 \text{ для любых наблюдений с номером } i \text{ и } j.$$

3°. *Отсутствие автокорреляции.* Случайные отклонения  $\varepsilon_i$  и  $\varepsilon_j$  не должны зависеть друг от друга для всех  $i \neq j$ .

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, & \text{àñëë } i \neq j, \\ \sigma^2, & \text{àñëë } i = j. \end{cases}$$

4°. Случайное отклонение должно быть независимым от объясняющих

переменных:

$$\text{Cov}(\varepsilon_i, X_{ij}) = 0.$$

5°. Модель эмпирической регрессии должна являться линейной относительно параметров. Это ограничение не распространяется на факторные переменные.

6°. *Отсутствие мультиколлинеарности.* Между объясняющими переменными должна отсутствовать строгая (сильная) линейная зависимость.

7°. *Случайные величины -ошибки*  $\varepsilon_i$ ,  $i = 1, 2, \dots, n$ , должны иметь нормальный закон распределения ( $\varepsilon_i \sim N(0, \sigma_\varepsilon)$ ).

Выполнимость данной предпосылки важна для проверки статистических гипотез и построения интервальных оценок.

Как и в случае парной регрессии, истинные значения параметров  $b_j$  с помощью случайной выборки получить невозможно. В этом случае вместо теоретического уравнения регрессии (2.3) оценивается так называемое *эмпирическое уравнение регрессии*. Эмпирическое уравнение регрессии представим в виде:

$$Y = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \dots + \hat{b}_p X_p + e \quad (2.6)$$

Здесь  $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p$  — оценки теоретических значений  $b_0, b_1, \dots, b_p$  коэффициентов регрессии (*эмпирические коэффициенты регрессий*);  $e$  — эмпирическая оценка неизвестного случайного отклонения  $\varepsilon$ . Для индивидуальных наблюдений имеем:

$$y_i = \hat{b}_0 + \hat{b}_1 x_{i1} + \hat{b}_2 x_{i2} + \dots + \hat{b}_p x_{ip} + e_i \quad (2.7)$$

Оцененное уравнение в первую очередь должно описывать общую закономерную тенденцию изменения зависимой переменной  $Y$ . При этом необходимо иметь возможность оценить случайные отклонения измеренных значений  $y_i$  от таких неслучайных расчетных значений.

По данным выборки объема  $n$ :  $(1, x_{i1}, x_{i2}, \dots, x_{ip}, y_i)$ ,  $i = 1, 2, \dots, n$ , требуется оценить значения параметров  $b_j$  вектора  $\mathbf{B}$ , т.е. провести параметризацию выбранной модели (здесь  $x_{ij}$ ,  $j = 0, 1, 2, \dots, p$  значение переменной  $X_j$  в  $i$ -ом

наблюдении).

При выполнении перечисленных выше предпосылок МНК относительно ошибок  $\varepsilon_i$  оценки  $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p$  коэффициентов  $b_0, b_1, \dots, b_p$  множественной линейной регрессии с помощью МНК являются несмещенными, эффективными и состоятельными (т.е. BLUE-оценками).

На основании (5.7) отклонение  $e_i$  значения зависимой переменной  $Y$  от модельного значения  $\hat{y}_i$ , соответствующего уравнению регрессии в  $i$ -ом наблюдении ( $i = 1, 2, \dots, n$ ), рассчитывается по формуле:

$$e_i = y_i - \hat{y}_i, \quad \hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{i1} + \hat{b}_2 x_{i2} + \dots + \hat{b}_p x_{ip}. \quad (2.8)$$

Наиболее распространенным методом оценки параметров уравнения множественной линейной регрессии является метод наименьших квадратов (МНК). Его суть состоит в минимизации суммы квадратов отклонений наблюдаемых значений зависимой переменной  $Y$  от ее расчетных значений  $\hat{Y}$ , получаемых с помощью модельного уравнения регрессии:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min.$$

По МНК для нахождения оценок минимизируется следующая функция, квадратичная относительно коэффициентов регрессии  $b_0, b_1, \dots, b_p$ :

$$S(\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left( y_i - \left( \hat{b}_0 + \sum_{j=1}^p \hat{b}_j x_{ij} \right) \right)^2. \quad (2.9)$$

Данная функция является квадратичной относительно неизвестных величин  $b_j, j = 0, 1, \dots, p$ . Она ограничена снизу, следовательно, имеет минимум. Необходимым условием минимума функции  $S(b_0, b_1, \dots, b_p)$  является равенство нулю всех ее частных производных по  $b_j$ . Частные производные квадратичной функции (2.9) являются линейными функциям относительно искомых оценок коэффициентов регрессии:

$$\frac{\partial S(\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p)}{\partial b_0} = 2 \sum_{i=1}^n \left( y_i - \left( \hat{b}_0 + \sum_{j=1}^p \hat{b}_j x_{ij} \right) \right) (-1),$$

$$\frac{\partial S(\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p)}{\partial b_j} = 2 \sum_{i=1}^n \left( y_i - \left( \hat{b}_0 + \sum_{j=1}^p \hat{b}_j x_{ij} \right) \right) (-x_{ij}), \quad (2.10)$$

где  $j = 1, 2, \dots, p$ .

Приравнявая их к нулю, получаем нормальную систему  $p + 1$  линейных уравнений с  $p + 1$  неизвестными оценками коэффициентов регрессии, что является одним из достоинств метода МНК. Такая система имеет обычно единственное решение:

$$\sum_{i=1}^n \left( y_i - \left( \hat{b}_0 + \sum_{j=1}^p \hat{b}_j x_{ij} \right) \right) = 0,$$

$$\sum_{i=1}^n \left( y_i - \left( \hat{b}_0 + \sum_{j=1}^p \hat{b}_j x_{ij} \right) \right) (x_{ij}) = 0, \quad (2.11)$$

где  $j = 1, 2, \dots, p$ .

В исключительных случаях, когда столбцы системы линейных уравнений линейно зависимы, она имеет бесконечно много решений или не имеет решения вовсе. Однако данные реальных статистических наблюдений к таким исключительным случаям практически никогда не приводят.

Система линейных уравнений относительно неизвестных оценок параметров линейной модели имеет следующий вид:

$$\begin{cases} \hat{b}_0 \cdot n + \hat{b}_1 \sum_{i=1}^n x_{i1} + \hat{b}_2 \sum_{i=1}^n x_{i2} + \dots + \hat{b}_p \sum_{i=1}^n x_{ip} = \sum_{i=1}^n y_i \\ \hat{b}_0 \sum_{i=1}^n x_{i1} + \hat{b}_1 \sum_{i=1}^n x_{i1}^2 + \hat{b}_2 \sum_{i=1}^n x_{i2} x_{i1} + \dots + \hat{b}_p \sum_{i=1}^n x_{ip} x_{i1} = \sum_{i=1}^n y_i x_{i1} \\ \hat{b}_0 \sum_{i=1}^n x_{i2} + \hat{b}_1 \sum_{i=1}^n x_{i2} x_{i1} + \hat{b}_2 \sum_{i=1}^n x_{i2}^2 + \dots + \hat{b}_p \sum_{i=1}^n x_{ip} x_{i2} = \sum_{i=1}^n y_i x_{i2} \\ \dots \\ \hat{b}_0 \sum_{i=1}^n x_{ip} + \hat{b}_1 \sum_{i=1}^n x_{ip} x_{i1} + \hat{b}_2 \sum_{i=1}^n x_{ip} x_{i2} + \dots + \hat{b}_p \sum_{i=1}^n x_{ip}^2 = \sum_{i=1}^n y_i x_{ip} \end{cases}$$

После деления всех уравнений системы на объем выборки  $n$  все суммарные величины преобразуются в соответствующие средние величины:

$$\begin{cases} \hat{b}_0 + \hat{b}_1 \bar{X}_1 + \hat{b}_2 \bar{X}_2 + \dots + \hat{b}_p \bar{X}_p = \bar{Y} \\ \hat{b}_0 \bar{X}_1 + \hat{b}_1 \bar{X}_1^2 + \hat{b}_2 \bar{X}_2 \bar{X}_1 + \dots + \hat{b}_p \bar{X}_p \bar{X}_1 = \overline{YX}_1 \\ \hat{b}_0 \bar{X}_2 + \hat{b}_1 \bar{X}_2 \bar{X}_1 + \hat{b}_2 \bar{X}_2^2 + \dots + \hat{b}_p \bar{X}_p \bar{X}_2 = \overline{YX}_2 \\ \dots \\ \hat{b}_0 \bar{X}_p + \hat{b}_1 \bar{X}_p \bar{X}_1 + \hat{b}_2 \bar{X}_p \bar{X}_2 + \dots + \hat{b}_p \bar{X}_p^2 = \overline{YX}_p. \end{cases} \quad (2.12)$$

Из первого уравнения можно определить величину коэффициента регрессии  $\hat{b}_0$ :

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}_1 - \hat{b}_2 \bar{X}_2 - \dots - \hat{b}_p \bar{X}_p.$$

Подставляя его в уравнение (2.8), получим следующую форму записи эмпирического линейного уравнения множественной регрессии:

$$\hat{y}_i = \bar{Y} + \hat{b}_1(x_{i1} - \bar{X}_1) + \hat{b}_2(x_{i2} - \bar{X}_2) + \dots + \hat{b}_p(x_{ip} - \bar{X}_p). \quad (2.13)$$

Нормальную систему линейных уравнений МНК (2.11) наиболее наглядно можно представить с помощью векторно-матричной формы записи.

### 2.3. Оценивание коэффициентов множественной линейной регрессии

Представим данные наблюдений и соответствующие коэффициенты в матричной форме.

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}, \quad B = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \dots \\ \hat{b}_p \end{pmatrix}, \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}.$$

Здесь  $Y$  —  $n$ -мерный вектор-столбец наблюдений зависимой переменной  $Y$ ;  $X$  — матрица размерности  $n \times (p + 1)$ , в которой  $i$ -я строка ( $i = 1, 2, \dots, n$ ) представляет наблюдение вектора значений независимых переменных  $X_1, X_2, \dots, X_p$ ; единица соответствует переменной при свободном члене  $b_0$ ,  $B$  — вектор-столбец размерности  $(p+1)$  параметров уравнения регрессии (2.8);  $e$  — вектор-столбец отклонений выборочных (реальных) значений  $y_i$  зависимой переменной

$Y$  (2.7) от значений  $\hat{y}_i$ , размерности  $n$ , получаемых из модельного уравнения регрессии:

$$\hat{y}_i = X \cdot \hat{B} = \hat{b}_0 + \hat{b}_1 x_{i1} + \hat{b}_2 x_{i2} + \dots + \hat{b}_p x_{ip}. \quad (2.14)$$

Сумма квадратов отклонений МНК в матричном виде запишется следующим образом:

$$e^T \cdot e \rightarrow \min \quad \Leftrightarrow \quad S(\hat{B}) = (Y - X \cdot \hat{B})^T (Y - X \cdot \hat{B}) \rightarrow \min.$$

$$\text{Условие экстремума: } S'(\hat{B}) = 0. \quad (2.15)$$

Частные производные по параметрам в матричной форме вычисляются следующим образом:

$$\begin{aligned} S'_B(\hat{B}) &= (Y - X \cdot \hat{B})^T (Y - X \cdot \hat{B}) = (Y^T - \hat{B}^T X^T)(Y - X \hat{B}) = \\ &= Y^T Y - Y^T X \hat{B} - \hat{B}^T X^T Y + \hat{B}^T X^T X \hat{B} = \\ &= Y^T Y - 2 \cdot \hat{B}^T X^T Y + \hat{B}^T X^T X \hat{B}. \\ S'_B(\hat{B}) &= -2 \cdot X^T Y + 2 \cdot X^T X \hat{B} = 0 \Rightarrow (X^T X) \hat{B} = (X^T Y) \end{aligned}$$

Таким образом, получим в матричной форме оценки параметров линейной регрессии:

$$\hat{B} = (X^T X)^{-1} (X^T Y). \quad (2.16)$$

Для вычисления их не нужно составлять и решать нормальную систему линейных уравнений МНК. Достаточно выполнить указанные в формуле (2.16) алгебраические операции в матричной форме над результатами исходных, выборочных наблюдений  $X$  и  $Y$ . В частности, при выполнении контрольной работы такие расчеты достаточно легко выполняются с помощью Мастера функций табличного процессора Microsoft Excel.

#### 2.4. Интерпретация оценок параметров и уравнения множественной линейной регрессии

Интерпретация – содержательное объяснение – результатов анализа

экономического явления или объекта, представленного статистическими (выборочными данными), является одной из самых важных задач регрессионного анализа. Так, рассматривая полученные оценки параметров уравнения регрессии, можно сказать, изменение фактора  $X_i$  на одну единицу своего измерения ведет к изменению объясняемой переменной  $Y$  на  $\hat{b}_i$  единиц измерения этой переменной. Направление ее изменения определяется знаком коэффициента перед фактором  $X_i$ .

При этом единицы, в которых измерены выборочные значения переменных  $Y$  и  $X_i$ , влияют на величину оценок параметров регрессии  $\hat{b}_i$ . Нужно обязательно фиксировать, в каких единицах измерены значения всех переменных, прежде чем заменять слово «единица» конкретными названиями: тонны, рубли и т.п. Отсюда следует, что коэффициенты регрессии перед различными факторами нельзя сравнивать друг с другом.

Все другие более общие показатели характера влияния факторов на объясняемую переменную, не зависящие от масштаба их измерения, такие как стандартизованные коэффициенты регрессии и коэффициенты эластичности, получают на основе этих оценок параметров  $\hat{b}_i$ .

Параметр  $\hat{b}_0$  представляет оценку значения объясняемой переменной  $Y$  при нулевых значениях факторных переменных. Она может иметь или не иметь экономический смысл в зависимости от характера конкретной ситуации.

При интерпретации модельного уравнения регрессии важно отмечать его следующие характерные особенности. Во-первых,  $\hat{b}_0$  и  $\hat{b}_i$  являются только оценками неизвестных констант  $b_0$  и  $b_i$  истинной, теоретической регрессии, которая к тому же не обязательно является линейной. Кроме того, величина и качество статистических оценок зависят от правильности выбора самого метода оценивания. В частности предпосылки МНК обосновывают условия, в которых можно получить «лучшие» оценки параметров модельной регрессии.

Во-вторых, эмпирическое уравнение регрессии отражает общую

закономерную тенденцию, представленную выборочными данными, тогда как каждое отдельное наблюдение подвержено случайным воздействиям со стороны неконтролируемых факторов. Следовательно расчетные значения объясняемой переменной не могут быть детерминированными и их нужно дополнять характеристиками вариации, например, стандартными ошибками или доверительными интервалами.

И, наконец, в-третьих, правильность интерпретации зависит от правильного выбора и полноты модельного представления статистической связи. Это связано с включением в уравнение всех статистически значимых объясняющих переменных, а также выбором формы уравнения эмпирической функции регрессии. Если форма уравнения является линейной, то можно использовать такие характеристики линейной статистической связи, такие как коэффициенты парной, частной и множественной корреляции. Но в случае, когда реальная функция регрессии нелинейная, они не могут отражать силы влияния факторов.

### 3. Анализ качества уравнения регрессии

Выявление лучшего варианта эконометрической модели обычно осуществляется сравнением соответствующих им качественных характеристик, которые можно рассчитать на основе исходной статистической информации, содержащейся в векторе  $Y$ , матрице  $X$ , и новой расчетной информации, появляющейся после построения каждого из вариантов модели. Основным условием высокого «качества» модели является обоснованность математической формы уравнения эмпирической регрессии. Важную роль при этом играет как состав включенных в него независимых переменных, так и характер их взаимосвязей с зависимой переменной  $y$ , которые в совокупности определяют причины ее изменчивости.

Сопоставление новой расчетной информации, полученной после оценки параметров модельной регрессии с исходной статистической информацией

позволяет установить, насколько удалось реализовать это условие на практике.

### 3.1. Характеристики и критерии качества эконометрических моделей

Ведущая роль при определении характеристик качества эконометрической модели принадлежит ряду ее «выборочной» ошибки  $e_i$ ,  $i = 1, 2, \dots, n$ , которая формируется с использованием найденных оценок ее параметров как

$$e_i = y_i - \hat{y}_i,$$

где  $\hat{y}_i$  — расчетное значение переменной  $y_i$  при известных значениях независимых переменных  $X_j - x_{ij}$ ,  $i = 1, 2, \dots, n$ ;  $j = 0, 1, 2, \dots, p$ . Так для линейной модели (2.7) значения  $\hat{y}_i$  определяются на основании следующего выражения:

$$\hat{y}_i = \hat{b}_0 + \hat{b}_1 x_{i1} + \hat{b}_2 x_{i2} + \dots + \hat{b}_p x_{ip}$$

Для каждого набора оценок параметров  $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p$  того или иного варианта модели, описывающей рассматриваемый процесс, рассчитывается «свой» ряд ошибки  $e_i$ , который можно интерпретировать как ряд оценок ее истинных, но неизвестных значений  $\varepsilon$ , теоретической регрессии (2.4).

В общем случае «качество» эконометрической модели оценивается с помощью различных характеристик. Самой простой из них является *средняя ошибка аппроксимации*, которая вычисляется как среднее отклонение расчетных значений от результатов фактических измерений. Совокупность отклонений  $e_i = y_i - \hat{y}_i$  можно рассматривать как абсолютные ошибки аппроксимации, а их абсолютные относительные величины

$$\left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%$$

как относительные ошибки аппроксимации.

Чтобы получить общее представление о качестве модели, из

относительных отклонений по каждому наблюдению вычисляют среднюю ошибку аппроксимации  $\bar{A}$  как простую среднюю арифметическую:

$$\bar{A} = \frac{1}{n} \cdot \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%. \quad (3.1)$$

Считается, что допустимый предел ошибки не должен превышать 8 – 10%.

Другой характеристикой качества модельного уравнения регрессии является несмещенная оценка дисперсии случайных отклонений  $S_e^2$ :

$$S_e^2 = \frac{\sum_{i=1}^n e_i^2}{n-p-1} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p-1}, \quad (3.2)$$

где  $p$  – число объясняющих переменных, факторов. Корень квадратный из оценки дисперсии обозначается как  $S_e$  и называется *стандартной ошибкой регрессии*.

Ошибка модельной регрессии во многом предопределена тем, что оценки  $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p$  рассчитывают по данным случайных измерений, и они являются случайными значениями величин  $b_0, b_1, \dots, b_p$  – неизвестных коэффициентов регрессии. Насколько хорошим оказывается соответствие между ними, насколько приемлемым можно считать «качество» полученной модели регрессии.

Надежность случайных оценок устанавливают также с помощью определения оценок их дисперсий (стандартных ошибок). Кроме того, строят доверительные интервалы для теоретических значений и проверяют статистические гипотезы о значимости отличия их эмпирических величин от ожидаемых, теоретических значений.

### 3.2. Дисперсии и стандартные ошибки параметров линейной регрессии

Оценки коэффициентов множественной линейной регрессии в матричной форме (2.2) определяются следующим образом:

$$\hat{B} = (X^T X)^{-1} (X^T Y).$$

Чтобы оценить ошибку оценки матрицы коэффициентов регрессии  $\hat{B}$  подставим в правую часть формулы теоретические значения объясняемой переменной  $Y = XB + \varepsilon$ :

$$\begin{aligned} \hat{B} &= (X^T X)^{-1} (X^T Y) = (X^T X)^{-1} X^T (XB + \varepsilon) = (X^T X)^{-1} (X^T X) B + \\ &+ (X^T X)^{-1} X^T \varepsilon = B + (X^T X)^{-1} X^T \varepsilon. \end{aligned}$$

Таким образом, ошибка полученной оценки  $\Delta \hat{B}$  имеет вид:

$$\Delta \hat{B} = \hat{B} - B = (X^T X)^{-1} X^T \varepsilon.$$

Дисперсия многомерной случайной величины  $\hat{B}$  определяется с помощью ковариационной матрицы  $\Sigma_{\hat{B}}$ :

$$\begin{aligned} \Sigma_{\hat{B}} &= M((\hat{B} - B)(\hat{B} - B)^T) = \\ &= M[(X^T X)^{-1} X^T \varepsilon][(X^T X)^{-1} X^T \varepsilon]^T] = M[(X^T X)^{-1} X^T \varepsilon \varepsilon^T X (X^T X)^{-1}]. \end{aligned}$$

В силу того, что объясняющие переменные  $X_j$  не являются случайными величинами, их можно вынести за знак математического ожидания:

$$\Sigma_{\hat{B}} = (X^T X)^{-1} X^T M(\varepsilon \cdot \varepsilon^T) X (X^T X)^{-1}. \quad (3.4)$$

Матрица  $M(\varepsilon \cdot \varepsilon^T)$  представляет собой ковариационную матрицу неизвестных случайных отклонений  $\varepsilon$  теоретической регрессии:

$$M(\varepsilon \varepsilon^T) = \Sigma_{\varepsilon} = \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1 \varepsilon_2) & \cdots & M(\varepsilon_1 \varepsilon_n) \\ M(\varepsilon_2 \varepsilon_1) & M(\varepsilon_2^2) & \cdots & M(\varepsilon_2 \varepsilon_n) \\ \cdot & \cdot & \cdot & \cdot \\ M(\varepsilon_n \varepsilon_1) & M(\varepsilon_n \varepsilon_2) & \cdots & M(\varepsilon_n^2) \end{pmatrix}.$$

В силу предпосылки МНК 2° все диагональные элементы одинаковы  $M(\varepsilon_i^2) = D(\varepsilon_i) = \sigma_{\varepsilon}^2$ , а все остальные равны нулю в силу предпосылки 3°. Таким образом, ковариационная матрица случайных ошибок  $\Sigma_{\varepsilon} = \sigma_{\varepsilon}^2 \cdot E_n$ , а выражение (3.4) принимает следующий вид:

$$\Sigma_{\hat{B}} = \sigma_{\varepsilon}^2 (X^T X)^{-1} X^T E_n X (X^T X)^{-1} = \sigma_{\varepsilon}^2 (X^T X)^{-1}. \quad (3.5)$$

Неизвестное значение дисперсии случайного отклонения теоретической

регрессии  $\sigma_e^2$  заменяется соответствующей несмещенной выборочной оценкой  $S_e^2$  (3.2). Следовательно, по выборке мы можем определить только выборочные оценки дисперсий коэффициентов эмпирической регрессии  $\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p$ , которые являются диагональными элементами матрицы  $\Sigma_{\hat{B}}$ :

$$S_{\hat{b}_j}^2 = S_e^2 (X^T X)_{jj}^{-1}, \quad (3.6)$$

где через  $(X^T X)_{jj}^{-1}$  обозначены диагональные элементы обратной матрицы  $(X^T X)^{-1}$ ,  $j = 0, 1, 2, \dots, p$ .

Стандартные ошибки коэффициентов регрессии вычисляются по формулам:

$$S_{\hat{b}_j} = S_e \sqrt{(X^T X)_{jj}^{-1}}, \quad (3.7)$$

где  $j = 0, 1, \dots, p$ .

### 3.3. Доверительные интервалы коэффициентов регрессии

Для построения интервальной оценки неизвестных коэффициентов регрессии  $b_i$  вводится случайная величина  $\frac{\hat{b}_i - b_i}{S_{\hat{b}_i}} = t_{bi}$  - стандартизованный коэффициент регрессии, имеющая распределение Стьюдента с числом степеней свободы  $\nu = n - p - 1$ .

При заданном уровне значимости  $\alpha$  доверительный интервал записывается следующим образом:

$$\left| \frac{\hat{b}_i - b_i}{S_{\hat{b}_i}} \right| = |t_{\hat{b}_i}| < t_{\alpha, n-p-1}$$

где  $t_{\alpha, n-p-1}$  - табличное значение  $t$ -критерия Стьюдента

Из данного неравенства следует:

$$\begin{aligned} \hat{b}_i - S_{\hat{b}_i} \cdot t_{\alpha, n-p-1} < b_i < \hat{b}_i + S_{\hat{b}_i} \cdot t_{\alpha, n-p-1} \\ \hat{b}_i - \Delta \hat{b}_i < b_i < \hat{b}_i + \Delta \hat{b}_i \end{aligned} \quad (3.8)$$

где  $S_{\hat{b}_i}$  и  $\Delta \hat{b}_i = S_{\hat{b}_i} \cdot t_{\alpha, n-p-1}$  - стандартная и предельная ошибки выборочных оценок  $\hat{b}_i$  соответственно.

### 3.4. Стандартная ошибка и доверительные интервалы уравнения регрессии

Дисперсия многомерной случайной величины  $\hat{Y} = X \cdot \hat{B}$  определяется с помощью ковариационной матрицы  $\Sigma_{\hat{B}}$ :

$$\begin{aligned} \Sigma_{\hat{Y}} &= M((X \hat{B} - XB)(X \hat{B} - XB)^T) = \\ &= M[X(\hat{B} - B)(\hat{B} - B)^T X^T] = X \cdot M[(\hat{B} - B)(\hat{B} - B)] \cdot X^T = X \cdot \Sigma_{\hat{B}} X^T, \end{aligned}$$

где матрица  $\Sigma_{\hat{B}}$  определена формулой (3.6).

Таким образом, получаем окончательное выражение в матричной форме:

$$\Sigma_{\hat{Y}} = X \cdot S_e^2 (X^T X)^{-1} X^T = S_e^2 \cdot X (X^T X)^{-1} X^T. \quad (3.9)$$

Выборочные оценки дисперсий  $i$ -го значения эмпирической регрессии  $S_{\hat{y}_i}^2$ , соответствующего  $i$ -му набору значений факторов  $(1, x_{i1}, x_{i2}, \dots, x_{ip})$  в  $i$ -ой строке матрицы  $X$  исходных данных, располагаются по диагонали матрицы  $\Sigma_{\hat{Y}}$ :

$$S_{\hat{y}_i}^2 = S_e^2 \cdot X_i (X^T X)^{-1} X_i^T. \quad (3.10)$$

Стандартные ошибки оценок значений регрессии вычисляются по формулам:

$$S_{\hat{y}_i} = S_e \cdot \sqrt{X_i (X^T X)^{-1} X_i^T}, \quad (3.11)$$

где  $i = 1, 2, \dots, n$ .

Доверительные интервалы для неизвестной функции регрессии  $\bar{y}_i$  строятся также с помощью случайной величины  $\frac{\hat{y}_i - \bar{y}_i}{S_{\hat{y}_i}} = t_{\hat{y}_i}$  - стандартизованная переменная, имеющей распределение Стьюдента с числом степеней свободы  $\nu = n - p - 1$ ,  $i$  - номер измерения.

$$\left| \frac{\hat{y}_i - \bar{y}_i}{S_{\hat{y}_i}} \right| = \left| t_{\hat{y}_i} \right| \leq t_{\alpha, n-p-1}$$

Из данного неравенства следует:

$$\hat{y}_i - \Delta \hat{y}_i < \bar{y}_i < \hat{y}_i + \Delta \hat{y}_i, \quad (3.12)$$

где  $S_{\hat{y}_i}$  и  $\Delta \hat{y}_i = S_{\hat{y}_i} \cdot t_{\alpha, n-p-1}$  - стандартная и предельная ошибки расчетного значения  $\hat{y}_i$  соответственно.

### 3.5. Статистическая значимость уравнения регрессии

Проверить статистическую значимость уравнения множественной регрессии означает установить, соответствует ли регрессионная модель, принятая для объяснения взаимосвязи между переменными, исходным статистическим данным. Или, другими словами, достаточно ли включенных в уравнение регрессии факторов для описания поведения объясняемой переменной на основе имеющихся выборочных данных.

Проверка значимости уравнения регрессии производится с помощью метода статистического анализа – дисперсионного анализа. Оценивание качества уравнения регрессии с помощью  $F$ -критерия Фишера состоит в проверке гипотезы  $H_0$  о статистической значимости уравнения регрессии или показателя тесноты связи. В случае, когда нулевая гипотеза отвергается, влияние включенных в регрессию факторов на объясняемую переменную преобладает над ее изменениями в силу других причин.

Для этого сравнивают фактическое значение критерия  $F_{эмл}$  с критическим, табличным значением  $F_{табл}$ .

$$F_{\hat{y}_i} = \frac{Q_{\hat{y}_i} / p}{Q_e / (n - p - 1)} = \frac{R_{yx1x2}^2}{1 - R_{yx1x2}^2} \cdot \frac{n - p - 1}{p},$$

где  $p$  - число объясняющих переменных, или факторов, включенных в модель.

$$F_{\hat{\alpha}\hat{\alpha}\hat{\alpha}} = F_{\alpha, \nu_1, \nu_2} = F(\alpha, p, n - p - 1),$$

где  $\alpha$  - уровень значимости;  $\nu_1$  и  $\nu_2$  - число степеней свободы большей (числителя) и меньшей (знаменателя) дисперсий соответственно.

Если  $F_{\hat{y}\hat{y}} > F_{\hat{\alpha}\hat{\alpha}\hat{\alpha}}(\alpha, n - p - 1)$ , то гипотеза  $H_0$  о случайной природе статистической связи отклоняется. С вероятностью  $(1 - \alpha)$  делается заключение о статистической значимости уравнения регрессии в целом и показателя тесноты связи  $R_{y_1x_2}$ , которые сформировались под совместным воздействием факторов  $X_1$  и  $X_2$ , которое нет оснований считать случайным. В противном случае оснований для отклонения гипотезы  $H_0$  нет и данная статистическая связь статистически незначима.

#### 4. Уравнение регрессии в стандартизованной форме

##### 4.1. Стандартизованные переменные

В эконометрике часто используется иной подход к определению параметров множественной регрессии (2.13) с исключенным коэффициентом  $\hat{b}_0$ :

$$\hat{y} = \bar{Y} + \hat{b}_1(x_1 - \bar{X}_1) + \hat{b}_2(x_2 - \bar{X}_2) + \dots + \hat{b}_p(x_p - \bar{X}_p).$$

Разделим обе части уравнения на стандартное отклонение объясняемой переменной  $S_Y$  и представим его в виде:

$$\frac{\hat{y} - \bar{Y}}{S_Y} = \frac{\hat{b}_1}{S_Y}(x_1 - \bar{X}_1) + \frac{\hat{b}_2}{S_Y}(x_2 - \bar{X}_2) + \dots + \frac{\hat{b}_p}{S_Y}(x_p - \bar{X}_p).$$

Разделим и умножим каждое слагаемое на стандартное отклонение соответствующей факторной переменной, чтобы перейти к стандартизованным (центрированным и нормированным) переменным:

$$\frac{\hat{y} - \bar{Y}}{S_Y} = \frac{\hat{b}_1 S_{X_1}}{S_Y} \frac{(x_1 - \bar{X}_1)}{S_{X_1}} + \frac{\hat{b}_2 S_{X_2}}{S_Y} \frac{(x_2 - \bar{X}_2)}{S_{X_2}} + \dots + \frac{\hat{b}_p S_{X_p}}{S_Y} \frac{(x_p - \bar{X}_p)}{S_{X_p}}$$

где новые переменные обозначены как

$$\hat{t}_y = \frac{\hat{y} - \bar{Y}}{S_Y}, t_{x1} = \frac{x_1 - \bar{X}_1}{S_{X1}}, \dots, t_{xp} = \frac{x_p - \bar{X}_p}{S_{Xp}}.$$

Все стандартизованные переменные имеют нулевую среднюю величину и одинаковую дисперсию, равную единице.

Уравнение регрессии в стандартизованной форме имеет вид:

$$\hat{t}_y = \hat{\beta}_1 \cdot t_{x1} + \hat{\beta}_2 \cdot t_{x2} + \dots + \hat{\beta}_p \cdot t_{xp}, \quad (3.1)$$

где  $\hat{\beta}_1 = \frac{\hat{b}_1 S_{X1}}{S_Y}, \hat{\beta}_2 = \frac{\hat{b}_2 S_{X2}}{S_Y}, \dots, \hat{\beta}_p = \frac{\hat{b}_p S_{Xp}}{S_Y}$  - стандартизованные коэффициенты регрессии.

Стандартизованные коэффициенты регрессии  $\hat{\beta}_i$  отличаются от коэффициентов  $\hat{b}_i$  обычной, естественной формы тем, что их величина не зависит масштаба измерения объясняемой и объясняющих переменных модели. Кроме того, между ними существует простая взаимосвязь:

$$\hat{b}_1 = \frac{\hat{\beta}_1 S_Y}{S_{X1}}, \hat{b}_2 = \frac{\hat{\beta}_2 S_Y}{S_{X2}}, \dots, \hat{b}_p = \frac{\hat{\beta}_p S_Y}{S_{Xp}}, \quad (3.2)$$

которая дает другой способ вычисления коэффициентов  $\hat{b}_i$  по известным значениям  $\hat{\beta}_i$ , более удобный в случае, например, двухфакторной регрессионной модели.

#### 4.2. Нормальная система уравнений МНК в стандартизованных переменных

Оказывается, что для вычисления коэффициентов стандартизованной регрессии нужно знать только парные коэффициенты линейной корреляции. Чтобы показать каким образом это делается, исключим из нормальной системы уравнений МНК неизвестную  $\hat{b}_0$  с помощью первого уравнения. Умножая первое уравнение на  $(-\bar{X}_1)$  и складывая его почленно со вторым уравнением,

получим:

$$\begin{aligned} \hat{b}_1(\overline{X_1^2} - \overline{X_1}^2) + \hat{b}_2(\overline{X_2 X_1} - \overline{X_2} \overline{X_1}) + \dots + \hat{b}_p(\overline{X_p X_1} - \overline{X_p} \overline{X_1}) = \\ = \overline{Y X_1} - \overline{Y} \cdot \overline{X_1}. \end{aligned}$$

Заменяя обозначениями дисперсии и ковариаций выражения в скобках

$$\begin{aligned} (\overline{X_1^2} - \overline{X_1}^2) = S_{X_1}^2, \quad (\overline{X_2 X_1} - \overline{X_2} \overline{X_1}) = \text{cov}(X_2, X_1) = S_{X_1 X_2}, \\ (\overline{X_p X_1} - \overline{X_p} \overline{X_1}) = S_{X_p X_1}, \quad \overline{Y X_1} - \overline{Y} \cdot \overline{X_1} = S_{Y X_1}, \end{aligned}$$

перепишем второе уравнение в удобном для дальнейшего упрощения

виде:

$$\hat{b}_1 S_{X_1}^2 + \hat{b}_2 S_{X_2 X_1} + \dots + \hat{b}_p S_{X_p X_1} = S_{Y X_1}.$$

Разделим обе части этого уравнения на стандартное отклонение переменных  $S_Y$  и  $S_{X_1}$ , а каждое слагаемое разделим и умножим на стандартное отклонение переменной, соответствующей номеру слагаемого:

$$\frac{\hat{b}_1 S_{X_1}}{S_Y} \frac{S_{X_1}^2}{S_{X_1} S_{X_1}} + \frac{\hat{b}_2 S_{X_2}}{S_Y} \frac{S_{X_2 X_1}}{S_{X_2} S_{X_1}} + \dots + \frac{\hat{b}_p S_{X_p}}{S_Y} \frac{S_{X_p X_1}}{S_{X_p} S_{X_1}} = \frac{S_{Y X_1}}{S_Y S_{X_1}}.$$

Вводя характеристики линейной статистической связи:

$$r_{X_1 X_2} = \frac{S_{X_1 X_2}}{S_{X_1} S_{X_2}}, \quad \dots \quad r_{X_1 X_p} = \frac{S_{X_1 X_p}}{S_{X_1} S_{X_p}}, \quad r_{X_1 Y} = \frac{S_{X_1 Y}}{S_{X_1} S_Y},$$

и стандартизованные коэффициенты регрессии

$$\hat{\beta}_1 = \frac{\hat{b}_1 S_{X_1}}{S_Y}, \hat{\beta}_2 = \frac{\hat{b}_2 S_{X_2}}{S_Y}, \dots, \hat{\beta}_p = \frac{\hat{b}_p S_{X_p}}{S_Y},$$

получаем:

$$\hat{\beta}_1 + \hat{\beta}_2 r_{X_1 X_2} + \hat{\beta}_3 r_{X_1 X_3} + \dots + \hat{\beta}_p r_{X_1 X_p} = r_{X_1 Y}.$$

После аналогичных преобразований всех остальных уравнений ,нормальная система линейных уравнений МНК (2.12) принимает следующий, более простой вид:



## 5. Возможности экономического анализа на основе многофакторной регрессионной модели

### 5.1. Коэффициенты $\beta_i$ стандартизованной регрессии

Стандартизованные коэффициенты регрессии показывают, на сколько стандартных отклонений  $S_Y$  изменится в среднем объясняемая переменная  $Y$ , если соответствующая объясняющая переменная  $X_i$  изменится на величину  $S_{X_i}$  одного ее стандартного отклонения при сохранении неизменным среднего уровня всех остальных факторов.

В силу того, что в стандартизованной регрессии все переменные заданы как центрированные и нормированные случайные величины, коэффициенты  $\beta_i$  сравнимы между собой. Сравнивая их друг с другом, можно ранжировать соответствующие им факторы  $X_i$  по силе воздействия на объясняемую переменную  $Y$ . В этом состоит основное преимущество стандартизованных коэффициентов регрессии от коэффициентов  $b_i$  регрессии в естественной форме, которые несравнимы между собой.

Эта особенность стандартизованных коэффициентов регрессии позволяет использовать при отсеке наименее значимых факторов  $X_i$  с близкими к нулю значениями их выборочных оценок  $\hat{\beta}_i$ . Решение об исключении их из модельного уравнения линейной регрессии принимается после проверки статистических гипотез о равенстве нулю его средней величины, равной.

### 5.2. Средние и частные коэффициенты эластичности

В экономических исследованиях для количественного представления результатов множественного анализа также часто применяют коэффициенты эластичности. Коэффициент эластичности характеризует изменение в процентах объясняемой переменной при изменении факторной переменной на

1% и определяется для зависимости  $\hat{y} = f(x)$  как

$$\dot{Y}_{y/x} = \frac{d\hat{y}/\hat{y}}{dx/x} = f'(x) \frac{x}{\hat{y}}.$$

Он имеет четкую экономическую интерпретацию для степенной функции  $\hat{y} = a \cdot x^b$ . Производная этой функции  $f'(x) = a \cdot b \cdot x^{b-1}$ . Оказывается, что коэффициент эластичности при этом равен

$$\dot{Y}_{y/x} = a \cdot b \cdot x^{b-1} \frac{x}{a \cdot x^b} = \frac{a \cdot b \cdot x^b}{a \cdot x^b} = b$$

и имеет постоянную величину.

Обобщение понятия эластичности на другие формы регрессионной связи приводит к тому, что  $\dot{Y}_{y/x}$  оказывается зависящим от величины факторной переменной. Так для линейной регрессии  $\hat{y} = a + b \cdot x$  эластичность принимает вид:

$$\dot{Y}_{y/x} = b \cdot \frac{x}{\hat{y}} = b \cdot \frac{x}{a + bx}.$$

Поскольку коэффициент эластичности для линейной функции не является постоянной величиной, то вводят средний показатель эластичности:

$$\bar{Y}_{y/x} = b \cdot \frac{\bar{X}}{\bar{Y}}.$$

Для характеристики относительной силы влияния каждого фактора в многофакторной регрессии используют частные коэффициенты эластичности:

$$\dot{Y}_{Y/X_i} = \frac{\partial \hat{Y}/\hat{y}}{\partial X_i / X_i} = f'(X_1, X_2, \dots, X_p) \frac{X_i}{\hat{Y}}$$

В частном случае двухфакторной линейной зависимости

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 \cdot X_1 + \hat{b}_2 \cdot X_2$$

для характеристики влияния  $X_1$  и  $X_2$  на  $Y$  рассчитывают частные коэффициенты эластичности:

$$\bar{Y}\dot{e}_{y/x1} = \hat{b}_1 \cdot \frac{\bar{X}_1}{\bar{Y}},$$

$$\bar{Y}\dot{e}_{y/x2} = \hat{b}_2 \cdot \frac{\bar{X}_2}{\bar{Y}}.$$

Экономическая интерпретация этой статистической взаимосвязи является следующей. При увеличении фактора  $X_1$  на 1% от его среднего уровня объясняемая переменная  $Y$  изменяется на  $\hat{b}_1/\bar{Y}$ , % своего от среднего уровня, а при увеличении фактора  $X_2$  на 1% от его среднего уровня объясняемая переменная  $Y$  изменяется на  $\hat{b}_2/\bar{Y}$ , % от своего среднего уровня.

Направление изменения объясняемой переменной зависит от знака коэффициента при соответствующем факторе.

### 5.3. Линейные коэффициенты парной, частной и множественной корреляции

Наиболее обоснованной характеристикой статистической связи является выборочный коэффициент линейной корреляции. При исследовании двухфакторной модели  $\hat{Y} = \hat{b}_0 + \hat{b}_1 \cdot X_1 + \hat{b}_2 \cdot X_2$  линейные коэффициенты парной корреляции определяются следующим:

$$r_{x_1y} = \frac{\overline{x_1 y} - \bar{x}_1 \cdot \bar{y}}{S_{x_1} \cdot S_y},$$

$$r_{x_2y} = \frac{\overline{x_2 y} - \bar{x}_2 \cdot \bar{y}}{S_{x_2} \cdot S_y},$$

$$r_{x_1x_2} = \frac{\overline{x_1 x_2} - \bar{x}_1 \cdot \bar{x}_2}{S_{x_1} \cdot S_{x_2}}.$$

Анализируя значения коэффициентов парной корреляции, приходим к выводам относительно силы и направления взаимосвязи между каждой парой переменных. Однако очевидно, что влияние каждой из факторных переменных на объясняемую переменную не может выражаться также просто, как в случае парной регрессии, в силу их взаимосвязи между собой.

Для характеристики влияния каждого фактора множественной регрессии, аналогичной стандартизованным коэффициентам регрессии, на основе парных коэффициентов линейной корреляции рассчитывают частные коэффициенты

корреляции. Они также используются при отборе факторов: целесообразность включения того или иного фактора в модель можно обосновывать величиной показателя частной корреляции.

Частные коэффициенты корреляции характеризуют влияние фактора  $X_i$  на объясняемую переменную  $Y$  при устранении влияния других факторов, включенных в уравнение регрессии. В зависимости от их числа вычисляют частные корреляции различного порядка.

В случае двухфакторной модели частные коэффициенты корреляции первого порядка вычисляются непосредственно с помощью парных коэффициентов корреляции по формулам:

$$r_{x_1y/x_2} = \frac{r_{x_1y} - r_{x_1x_2} \cdot r_{x_2y}}{\sqrt{(1 - r_{x_1x_2}^2)(1 - r_{x_2y}^2)}},$$

$$r_{x_2y/x_1} = \frac{r_{x_2y} - r_{x_1x_2} \cdot r_{x_1y}}{\sqrt{(1 - r_{x_1x_2}^2)(1 - r_{x_1y}^2)}},$$

$$r_{x_1x_2/y} = \frac{r_{x_1x_2} - r_{x_1y} \cdot r_{x_2y}}{\sqrt{(1 - r_{x_1y}^2)(1 - r_{x_2y}^2)}}.$$

Если сравнить значения коэффициентов парной и частной корреляции, то приходим к выводу, что из-за сильной связи  $x_1$  и  $y$ , умеренной связи  $x_2$  и  $y$  коэффициенты  $r_{x_1y}$  и  $r_{x_1y/x_2}$  отличаются незначительно,  $r_{x_2y}$  и  $r_{x_2y/x_1}$  - значительно. Выводы о направлении связи на основе этих коэффициентов совпадают.

Когда общее число факторов больше двух и равно  $p$ , то возможны частные коэффициенты корреляции не только первого, но и второго, третьего, ...,  $(p-1)$  порядка. Так влияние фактора  $X_1$  на переменную  $Y$  можно оценить при разных способах исключения влияния других факторов: по одному, по два, ..., всех  $(p-2)$  - ух факторов из числа входящих в регрессию. Хотя частная корреляция различного порядка и может представлять интерес при изучении различных взаимосвязей факторных переменных, на практике ограничиваются вычислением частных корреляций самого высокого порядка. Этими показателями обычно принято дополнять эмпирические модели множественной

регрессии

Практическая значимость эмпирического уравнения множественной регрессии оценивается помощью коэффициента множественной корреляции  $R_{YX1X2...Xp}$  или его квадрата  $R_{YX1X2...Xp}^2$  – коэффициента множественной детерминации.

Коэффициент множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с объясняемой переменной, или оценивает тесноту совместного влияния факторов на результат. Его величина не может быть меньше, чем величина наибольшего коэффициента парной корреляции:

$$R_{YX1X2...Xp} \geq \max_{i=1,2,\dots,p} (r_{XiY}).$$

При правильном выборе факторов величина коэффициента множественной корреляции должен существенно отличаться от коэффициентов парной корреляции. Целесообразность включения в состав регрессии нового фактора можно обосновать, сравнивая коэффициенты множественной и парной корреляции с его участием.

Коэффициент множественной корреляции можно рассчитать несколькими способами, это полезно делать для контроля правильности вычислений:

1) наиболее просто это сделать для уравнения линейной регрессии в стандартизованной форме:

$$R_{YX1X2...Xp} = \sqrt{\beta_1 \cdot r_{X1Y} + \beta_2 \cdot r_{X2Y} + \dots + \beta_p r_{XpY}}.$$

2) используя матрицу парных коэффициентов корреляции:

$$R_{YX1X2} = \sqrt{1 - \frac{|R'|}{|R|}},$$

где  $|R'| = \begin{vmatrix} 1 & r_{yx1} & r_{yx2} & \dots & r_{yxp} \\ r_{x1y} & 1 & r_{x1x2} & \dots & r_{x1xp} \\ r_{x2y} & r_{x2x1} & 1 & \dots & r_{x2xp} \\ \dots & \dots & \dots & \dots & \dots \\ r_{xpy} & r_{xpx1} & r_{xpx2} & \dots & 1 \end{vmatrix}$  – определитель матрицы

коэффициентов парной корреляции;

$$|R| = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix} - \text{определитель матрицы коэффициентов парной}$$

корреляции между факторами, или межфакторного взаимодействия.

3) как корень квадратный из коэффициента множественной детерминации:

$$\begin{aligned} R_{YY_1Y_2\dots X_p} &= \sqrt{R_{YX_1X_2\dots X_p}^2} = \sqrt{\frac{S_{\hat{y}}^2}{S_y^2}} = \sqrt{\frac{S_y^2 - S_e^2}{S_y^2}} = \\ &= \sqrt{1 - \frac{S_e^2}{S_y^2}} = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{(X_1X_2\dots X_p)_i})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}. \end{aligned}$$

На основании значения коэффициента множественной регрессии делается вывод о том, что зависимостью переменной  $y$  от факторных переменных объясняется  $R^2$  процентов вариации, представленной исходными данными. Прочие факторы, не включенные в модель, составляют соответственно  $(1 - R^2)\%$  от общей вариации объясняемой переменной  $y$ .

## 6. Порядок выполнения и оформления типового примера.

Имеются следующие выборочные данные:

№ п/п	X <sub>1</sub>	X <sub>2</sub>	Y
1	28	-69	298
2	30	-67	310
3	29	-62	295
4	30	-77	316
5	31	-78	311
6	29	-63	308
7	30	-72	309
8	30	-79	306
9	29	-74	307
10	29	-68	293
11	30	-67	303
12	29	-72	300
13	30	-73	311
14	29	-66	302
15	29	-71	299
16	29	-66	304
17	29	-76	309
18	29	-73	297
19	30	-67	315
20	29	-69	297
21	29	-59	308
22	30	-76	304
23	31	-78	318
24	32	-82	337
25	28	-69	298
26	30	-67	310
27	29	-62	295
28	30	-77	316
29	31	-78	311
30	29	-63	308
31	30	-72	309
32	30	-79	306
33	29	-74	307
34	29	-68	293
35	30	-67	303
36	29	-72	300
37	30	-73	311
38	29	-66	302
39	29	-71	299
40	29	-66	304
41	29	-76	309
42	29	-73	297
43	30	-67	315
44	29	-69	297
45	29	-59	308
46	30	-76	304
47	31	-78	318
48	32	-82	337

где  $X_1$  и  $X_2$  - объясняющие или факторные переменные;  $Y$  - объясняемая переменная;  $n = 48$  - число выполненных измерений, или объем выборки.

Исследуем зависимость объясняемой переменной  $Y$  от факторных переменных  $X_1$  и  $X_2$ , используя множественный регрессионный анализ. Для этого необходимо выполнить расчеты:

1. Вычисление осредненных характеристик выборки (допускается выполнение на компьютере, остальные разделы – с помощью калькулятора).
2. Вычисление коэффициентов парной корреляции.
3. Вычисление стандартизованных коэффициентов

множественной линейной регрессии и ранжирование с их помощью факторных переменных.

4. Обоснование формы и оценка параметров линейной множественной регрессии.
5. Построение множественной линейной регрессии в естественной форме.
6. Вычисление стандартной ошибки регрессии.
7. Вычисление частных коэффициентов корреляции и частных коэффициентов эластичности.
8. Вычисление коэффициента множественной регрессии и индекса множественной корреляции.
9. Проверка гипотезы о статистической значимости полученного уравнения множественной регрессии
10. Вычисление доверительных интервалов параметров регрессии при уровне значимости  $\alpha = 0,1$  и  $0,05$ .

Каждый из них должен включать:

- а) изложение теоретических предпосылок и обоснование расчетных формул;
- б) выполнение расчетов;
- в) обсуждение полученных результатов.

Исследование должно включать в себя наряду с получением оценок статистических характеристик многомерной пространственной выборки обоснование применяемых методов их вычисления, а также элементы экономического анализа.

Выполнение численных расчетов необходимо сопровождать обсуждением экономического содержания полученных результатов. В конце работы на отдельном листе прилагается список использованной методической и учебной литературы.

В таблице 1. показаны вычисления всех средних величин, которые необходимы для проведения этого анализа и вычисления по данным выборки

следующих показателей:

1. Статистические дисперсии:

$$S_Y^2 = \overline{Y^2} - \bar{Y}^2 = 93953,00 - 306,21^2 = 189,457;$$

$$S_{X_1}^2 = \overline{X_1^2} - \bar{X}_1^2 = 882,77 - 29,69^2 = 1,423;$$

$$S_{X_2}^2 = \overline{X_2^2} - \bar{X}_2^2 = 49,2360 - 69,90^2 = 38,177;$$

Стандартные (среднеквадратические) отклонения:

$$S_{X_1} = \sqrt{1,423} = 1,19; \quad S_{X_2} = \sqrt{38,177} = 6,18;$$

$$S_Y = \sqrt{189,457} = 13,76.$$

Ковариации:

$$S_{YX_1} = \overline{Y \cdot X_1} - \bar{Y} \cdot \bar{X}_1 = 9105,23 - 29,69 \cdot 306,21 = 14,67;$$

$$S_{YX_2} = \overline{Y \cdot X_2} - \bar{Y} \cdot \bar{X}_2 = -21453,52 - 306,21 \cdot (-69,90) = -50,83;$$

$$S_{X_1X_2} = \overline{X_1 X_2} - \bar{X}_1 \cdot \bar{X}_2 = -2079,54 - 29,69 \cdot (-69,90) = -4,51.$$

2. Линейные коэффициенты парной корреляции:

$$r_{YX_1} = \frac{S_{YX_1}}{S_{X_1} \cdot S_Y} = \frac{14,67}{1,19 \cdot 13,76} = 0,893;$$

$$r_{YX_2} = \frac{S_{YX_2}}{S_{X_2} \cdot S_Y} = \frac{-50,83}{6,18 \cdot 13,76} = -0,598;$$

$$r_{X_1X_2} = \frac{S_{X_1X_2}}{S_{X_1} \cdot S_{X_2}} = \frac{-4,51}{1,19 \cdot 6,18} = -0,612.$$

Анализируя значения коэффициентов парной корреляции, приходим к следующему выводу:

Связанные переменные	Теснота связи	Направление связи
$x_1$ и $y$	сильная	прямая
$x_2$ и $y$	умеренная	обратная
$x_1$ и $x_2$	умеренная	обратная

3. Стандартизованные коэффициенты регрессии:

$$\hat{\beta}_1 = \frac{r_{YX_1} - r_{X_1X_2} \cdot r_{YX_2}}{1 - r_{X_1X_2}^2} = \frac{0,8934 - (-0,6117) \cdot (0,5977)}{1 - 0,6117^2} = 0,843;$$

$$\hat{\beta}_2 = \frac{r_{YX_2} - r_{X_1X_2} \cdot r_{YX_1}}{1 - r_{X_1X_2}^2} = \frac{-0,5977 - (-0,6117) \cdot (0,9934)}{1 - 0,6117^2} = -0,082.$$

Сравнивая модули значений  $\hat{\beta}_1$  и  $\hat{\beta}_2$ , приходим к выводу, что сила влияния фактора  $X_1$  на объясняемую переменную  $Y$  намного больше, чем сила влияния фактора  $X_2$ .

4. Оценки коэффициенты множественной линейной регрессии:

$$\hat{b}_1 = \beta_1 \cdot \frac{S_y}{S_{X_1}} = 0,8433 \cdot \frac{13,76}{1,19} = 9,73;$$

$$\hat{b}_2 = \beta_2 \cdot \frac{S_y}{S_{X_2}} = -0,0819 \cdot \frac{13,76}{6,18} = -0,18;$$

$$\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}_1 - \hat{b}_2 \bar{X}_2 = 306,21 - 9,73 \cdot 29,69 - 0,18 \cdot 69,90 = 4,61.$$

5. Уравнение множественной линейной регрессии в естественной форме:

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 \cdot X_1 + \hat{b}_2 \cdot X_2 = 4,61 + 9,73 \cdot X_1 - 0,18 \cdot X_2$$

6. Вычисление стандартной ошибки регрессии  $S_e$  :

Таблица 2

№ п/п	$x_{i1}$	$x_{i2}$	$Y_i$	$\hat{Y}_i$	$e_i = Y_i - \hat{Y}_i$	$e_i^2$
1	28	-69	298	289,63	8,37	70,12
2	30	-67	310	308,72	1,28	1,64
3	29	-62	295	298,08	-3,08	9,48
4	30	-77	316	310,54	5,46	29,76
5	31	-78	311	320,46	-9,46	89,42
6	29	-63	308	298,26	9,74	94,84
7	30	-72	309	309,63	-0,63	0,40
8	30	-79	306	310,91	-4,91	24,10
9	29	-74	307	300,27	6,73	45,32
10	29	-68	293	299,17	-6,17	38,11
11	30	-67	303	308,72	-5,72	32,73
12	29	-72	300	299,90	0,10	0,01

№ п/п	$x_{i1}$	$x_{i2}$	$Y_i$	$\hat{Y}_i$	$e_i = Y_i - \hat{Y}_i$	$e_i^2$
13	30	-73	311	309,81	1,19	1,40
14	29	-66	302	298,81	3,19	10,18
15	29	-71	299	299,72	-0,72	0,52
16	29	-66	304	298,81	5,19	26,95
17	29	-76	309	300,63	8,37	70,02
18	29	-73	297	300,09	-3,09	9,52
19	30	-67	315	308,72	6,28	39,43
20	29	-69	297	299,36	-2,36	5,55
21	29	-59	308	297,53	10,47	109,58
22	30	-76	304	310,36	-6,36	40,48
23	31	-78	318	320,46	-2,46	6,03
24	32	-82	337	330,92	6,08	37,02
25	30	-79	316	310,91	5,09	25,92
26	30	-64	302	308,17	-6,17	38,11
27	31	-77	320	320,27	-0,27	0,08
28	29	-64	291	298,44	-7,44	55,41
29	29	-71	290	299,72	-9,72	94,49
30	32	-74	324	329,46	-5,46	29,77
31	29	-63	304	298,26	5,74	32,93
32	29	-68	292	299,17	-7,17	51,46
33	29	-63	309	298,26	10,74	115,31
34	30	-61	303	307,63	-4,63	21,40
35	30	-67	307	308,72	-1,72	2,96
36	31	-76	329	320,09	8,91	79,36
37	32	-73	336	329,27	6,73	45,24
38	31	-77	310	320,27	-10,27	105,56
39	30	-64	301	308,17	-7,17	51,46
40	27	-63	270	278,80	-8,80	77,48
41	32	-73	338	329,27	8,73	76,14
42	28	-60	283	287,98	-4,98	24,85
43	30	-63	305	307,99	-2,99	8,95
44	32	-82	332	330,92	1,08	1,18
45	30	-70	304	309,27	-5,27	27,75
46	28	-60	293	287,98	5,02	25,15
47	29	-73	299	300,09	-1,09	1,18
48	27	-66	283	279,35	3,65	13,33
□	1425	-3355	14698	14698,00	0,00	1798,07
□/n	29,69	-69,90	306,21	306,21	0,00	37,46

$$S_e^2 = \overline{e^2} - \bar{e}^2 = 37,46 - 0 = 37,46; \quad S_e = \sqrt{37,46} = 6,12.$$

## 7. Частные коэффициенты линейной корреляции:

$$r_{x_1y/x_2} = \frac{r_{x_1y} - r_{x_1x_2} \cdot r_{x_2y}}{\sqrt{(1-r_{x_1x_2}^2)(1-r_{x_2y}^2)}} = \frac{0,893 - (-0,598) \cdot (-0,612)}{\sqrt{(1-(-0,598)^2)(1-(-0,612)^2)}} = 0,832$$

$$r_{x_2y/x_1} = \frac{r_{x_2y} - r_{x_1x_2} \cdot r_{x_1y}}{\sqrt{(1-r_{x_1x_2}^2)(1-r_{x_1y}^2)}} = \frac{-0,598 - (-0,612) \cdot 0,893}{\sqrt{(1-(-0,612)^2)(1-0,893^2)}} = -0,144$$

$$r_{x_1x_2/y} = \frac{r_{x_1x_2} - r_{x_1y} \cdot r_{x_2y}}{\sqrt{(1-r_{x_1y}^2)(1-r_{x_2y}^2)}} = \frac{-0,612 - 0,893 \cdot (-0,598)}{\sqrt{(1-0,893^2)(1-(-0,598)^2)}} = -0,216$$

Сравнение с парными коэффициентами корреляции показывает, что существует слабая взаимосвязь между факторными переменными. Но она сильно проявилась на связи  $Y$  и  $X_2$ , уменьшив характеристику связи с  $-0,598$  до  $-0,144$ . На связь  $Y$  и  $X_1$  она повлияла незначительно ( $0,893$  и  $0,832$  соответственно).

## Средние частные коэффициенты эластичности:

$$\bar{Y}_{\hat{y}/x_1} = \hat{b}_1 \cdot \frac{x_1}{y} = 9,73 \cdot \frac{29,69}{306,21} = 0,944,$$

$$\bar{Y}_{\hat{y}/x_2} = \hat{b}_2 \cdot \frac{x_2}{y} = -0,18 \cdot \frac{-69,90}{306,21} = 0,041.$$

При увеличении фактора  $X_1$  на 1% от его среднего уровня объясняемая переменная  $Y$  возрастает на 0,944 % от среднего уровня, а при увеличении фактора  $X_2$  на 1% от его среднего уровня объясняемая переменная  $Y$  возрастает на 0,041 % от своего среднего уровня.

8. Характеристика совокупного влияния всех факторов на объясняемую переменную - коэффициент множественной корреляции  $R_{yx_1x_2}$  можно рассчитать несколькими способами:

1) для уравнения регрессии в стандартизованной форме:

$$R_{yx_1x_2} = \sqrt{\beta_1 \cdot r_{x_1y} + \beta_2 \cdot r_{x_2y}} =$$

$$= \sqrt{0,8433 \cdot 0,8934 + (-0,0819) \cdot (-0,5977)} = \sqrt{0,8023} = 0,8957.$$

2) используя матрицы парных коэффициентов корреляции  $R'$  и  $R$ :

$$R_{yx1x2} = \sqrt{1 - \frac{|R'|}{|R|}},$$

$$\text{Где } |R'| = \begin{vmatrix} 1 & r_{x1y} & r_{x2y} \\ r_{x1y} & 1 & r_{x1x2} \\ r_{x2y} & r_{x1x2} & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0,8934 & -0,5977 \\ 0,8934 & 1 & -0,6117 \\ -0,5977 & -0,6117 & 1 \end{vmatrix} = 0,124 \text{ - определитель}$$

матрицы коэффициентов парной корреляции;

$$|R| = \begin{vmatrix} 1 & r_{x1x2} \\ r_{x1x2} & 1 \end{vmatrix} = \begin{vmatrix} 1 & -0,6117 \\ -0,6117 & 1 \end{vmatrix} = 0,626 \text{ - определитель матрицы}$$

межфакторной корреляции, или межфакторного взаимодействия.

$$\text{Таким образом: } R_{yx1x2} = \sqrt{1 - \frac{0,1237}{0,6258}} = \sqrt{0,8023} = 0,8957.$$

3) как корень квадратный из коэффициента множественной детерминации:

$$\begin{aligned} R_{yx1x2} &= \sqrt{R_{yx1x2}^2} = \sqrt{\frac{S_{\hat{y}}^2}{S_y^2}} = \sqrt{\frac{S_y^2 - S_e^2}{S_y^2}} = \sqrt{1 - \frac{S_e^2}{S_y^2}} = \\ &= \sqrt{1 - \frac{\sum_{i=1}^{48} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{48} (y_i - \bar{y})^2}} = \sqrt{1 - \frac{1798,07}{9093,93}} = \sqrt{0,8023} = 0,8957. \end{aligned}$$

На основании значения коэффициента множественной регрессии делается вывод, что зависимость объясняемой переменной  $y$  от факторов  $x_1$  и  $x_2$  характеризуется как тесная, в которой 80,2% вариации  $y$  определяются вариацией данных факторов. Прочие факторы, не включенные в модель, составляют соответственно только 19,8 % от общей вариации  $y$ .

9. Оценивание качества уравнения регрессии с помощью  $F$ -критерия Фишера состоит в проверке гипотезы  $H_0$  о статистической значимости уравнения регрессии или показателя тесноты связи. Для этого сравнивают фактическое значение критерия  $F_{эмн}$  с критическим, табличным значением  $F_{табл}$ .

$$F_{\hat{y}\hat{y}} = \frac{R_{yx1x2}^2}{1 - R_{yx1x2}^2} \cdot \frac{n - p - 1}{p} = \frac{0,8957}{1 - 0,8957} \cdot \frac{48 - 2 - 1}{2} = 193,22,$$

где  $p$  - число объясняющих переменных, или факторов, включенных в модель.

$$F_{\hat{\delta}\hat{\alpha}\hat{\alpha}\hat{\epsilon}} = F_{\alpha, \nu_1, \nu_2} = F(\alpha, p, n - p - 1) = F(0,05, 2, 45) = 3,25,$$

где  $\alpha = 0,05$  - уровень значимости;  $\nu_1 = 2$  и  $\nu_2 = 45$  - число степеней свободы большей (числителя) и меньшей (знаменателя) дисперсий соответственно.

Так как в нашем случае  $F_{\hat{y}\hat{i}\hat{i}} > F_{\hat{\delta}\hat{\alpha}\hat{\alpha}\hat{\epsilon}}$  ( $193,22 > 3,25$ ), то гипотеза  $H_0$  о случайной природе статистической связи отклоняется. Имеющиеся статистические данные свидетельствуют о том, что в 95 % случаев связь обусловлена влиянием факторов регрессии, а остальные не включенные в нее факторы являются статистически не значимыми. С вероятностью 0,95 делаем заключение о статистической значимости уравнения регрессии в целом и показателя тесноты связи  $R_{y \times 1 \times 2}$ , которые сформировались под неслучайным воздействием факторов  $X_1$  и  $X_2$ .

10. Найдем доверительные интервалы для коэффициентов регрессии  $\hat{b}_0$ ,  $\hat{b}_1$  и  $\hat{b}_2$ .

$$(X^T \cdot X)^{-1} = \begin{pmatrix} 13,4317 & -0,5013 & -0,021072 \\ 0,5014 & 0,0234 & 0,002763 \\ 0,0211 & 0,0028 & 0,000872 \end{pmatrix}.$$

Табличная величина критерия Стьюдента:

$$t_{\alpha, n-p-1} = t(0,05, 45) = 2,027.$$

Дисперсия стандартной ошибки оценки коэффициентов  $\hat{b}_i$ :

$$1) S_{\hat{b}_0}^2 = S_e^2 \cdot ((X^T \cdot X)^{-1})_{11} = 37,46 \cdot 13,4317 = 503,151.$$

Предельная ошибка выборки оценки коэффициента  $\hat{b}_0$ :

$$\Delta \hat{b}_0 = S_{\hat{b}_0} \cdot t_{\alpha, n-p-1} = 22,431 \cdot 2,027 = 45,468$$

Таким образом, интервальная оценка коэффициента  $\hat{b}_0$ :

$$10,401 - 45,468 < \hat{b}_0 < 10,401 + 45,468 .$$

$$- 35,067 < \hat{b}_0 < 55,869$$

$$2) S_{\hat{b}_1}^2 = S_e^2 \cdot ((X^T \cdot X)^{-1})_{22} = 37,46 \cdot 0,0234 = 0,877 ,$$

$$\Delta \hat{b}_1 = S_{\hat{b}_1} \cdot t_{\alpha, n-p-1} = 0,936 \cdot 2,027 = 1,898 .$$

Таким образом:  $9,687 - 1,898 < \hat{b}_1 < 9,687 + 1,898 ,$

$$7,789 < b_1 < 12,285 .$$

$$3) S_{\hat{b}_2}^2 = S_e^2 \cdot ((X^T \cdot X)^{-1})_{33} = 37,46 \cdot 0,00087 = 0,0326$$

$$\Delta \hat{b}_2 = S_{\hat{b}_2} \cdot t_{\alpha, n-p-1} = 0,181 \cdot 2,027 = 0,366$$

Таким образом:  $-0,114 - 0,366 < \hat{b}_2 < -0,114 + 0,366$

$$-0,480 < \hat{b}_2 < 0,252 .$$

Розенцвайг А.К., Исавнин А.Г.

**Эконометрическое моделирование**  
(методические указания к занятиям)

Учебно-методическое пособие

Подписано в печать 22.04.2019.  
Формат 60x84/16. Печать ризографическая.  
Бумага офсетная. Гарнитура «Times New Roman».  
Усл.п.л. 3,375 Уч.-изд. л. 3,24  
Тираж 100 экз. Заказ № 1237

Отпечатано в Издательско-полиграфическом центре  
Набережночелнинского института  
Казанского (Приволжского) федерального университета

---

423810, г. Набережные Челны, Новый город, пр.Мира, 68/19  
тел./факс (8552) 39-65-99 e-mail: [ic-nchi-kpfu@mail.ru](mailto:ic-nchi-kpfu@mail.ru)