

Министерство образования и науки Украины
Кировоградский национальный технический университет
Факультет учета и финансов
Кафедра экономической теории, маркетинга и экономической
кибернетики

ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ МЕТОДЫ И МОДЕЛИ

методические указания для студентов всех форм обучения
(часть II. Эконометрика)

Составители:

доктор физ.-мат. наук, проф. Гамалий В. Ф.

к.э.н., доцент Дмитришин Б. В.

к.т.н., доцент Сотников В.С.

Утверждено на заседании кафедры
экономической теории, маркетинга и
экономической кибернетики
протокол № 2014 г.

Кировоград-2014

Экономико-математические методы и модели. Методические указания для студентов всех форм обучения. Ч.2 – Эконометрика / Сост.: В.Ф. Гамалий, Б.В. Дмитришин, В.С. Сотников. – Кировоград: КНТУ, 2014. – 48 с.

Составители:

Гамалий Владимир Федорович – доктор физико-математических наук, профессор, заведующий кафедрой экономической теории, маркетинга и экономической кибернетики;

Дмитришин Богдан Васильевич – кандидат экономических наук, доцент кафедры экономической теории, маркетинга и экономической кибернетики;

Сотников Валентин Семенович – кандидат технических наук, доцент кафедры экономической теории, маркетинга и экономической кибернетики

Содержание

Введение	4
Глава 1. Парная регрессия (однофакторная эконометрическая модель)	5
1.1. Проблема взаимосвязи экономических показателей	5
1.2. Построение зависимости между показателями	6
1.3. Парная регрессия и её виды	8
Глава 2. Метод наименьших квадратов	13
2.1. Процедура метода наименьших квадратов	13
2.2. Предпосылки метода наименьших квадратов	16
Глава 3. Множественная регрессия (многофакторная эконометрическая модель)	19
3.1. Спецификация модели. Отбор факторов при построении уравнения множественной регрессии	19
3.2. Матричная форма МНК	20
Глава 4. Анализ адекватности эконометрической модели и показатели качества регрессии	23
4.1. Показатели тесноты связи	23
4.2. Оценка качества подбора вида функции	25
4.3. Оценка значимости уравнения регрессии в целом	26
4.3. Показатели качества отдельных параметров уравнения регрессии	27
Лабораторный практикум	29
Лабораторная работа № 1	29
Лабораторная работа № 2	38
Рекомендованная литература	47

ВВЕДЕНИЕ

Эконометрика — фундаментальная экономико-математическая дисциплина, которая изучает *методику построения экономико-математических моделей* на основе статистических данных о социально-экономических явлениях.

Общепринятого определения эконометрики в настоящее время нет. Сам термин «эконометрика» сложился для определения нового направления в экономической науке в 1930 г. и состоит из двух слов: «экономика» и «метрика» (от греческого *metron* – мера).

Эконометрические модели устанавливают конкретные количественные связи между экономическими факторами и количественно описывают динамику экономических процессов во времени.

Целью создания таких моделей являются:

- 1. Прогнозирование.*
- 2. Анализ взаимовлияния экономических факторов.*
- 3. Принятие оптимальных решений при планировании, распределении материальных, трудовых и финансовых ресурсов.*

Формирование эконометрии как сравнительно молодой науки осуществляется на стыке теоретической экономики, математики и статистики.

Последние 40 лет эконометрика как научная дисциплина стремительно развивается. Свидетельством всемирного признания и значимости эконометрики является присуждение ряда нобелевских премий в области экономики за выдающиеся исследования в области эконометрики. Растет число публикаций и исследований с применением эконометрических методов. В эконометрику все глубже проникают новейшие методы математики и информационные технологии.

ГЛАВА 1 . ПАРНАЯ РЕГРЕССИЯ

(ОДНОФАКТОРНАЯ ЭКОНОМЕТРИЧЕСКАЯ МОДЕЛЬ)

1.1. Проблема взаимосвязи экономических показателей

1.2 Построение зависимости между показателями

1.3. Парная регрессия и её виды

1.1. Проблема взаимосвязи экономических показателей

Проблема оценивания экономических переменных, проблема взаимосвязи экономических показателей является одной из важнейших проблем экономического анализа.

Любая экономическая политика заключается в регулировании экономических показателей, и эта политика должна основываться на знании того, как эти показатели влияют на другие переменные.

Вполне очевидно, что в рыночной экономике нельзя регулировать темп инфляции. В то же время на темп инфляции можно соответствующим образом воздействовать. Например, при помощи средств бюджетно-налоговой политики, кредитно-финансовой политики. Таким образом, становится вполне очевидно, что одни экономические переменные соответствующим образом воздействовать на другие.

Учитывая указанные обстоятельства, необходимо изучать функции предложения денег и уровня цен.

Можно отметить, что вся сфера экономических исследований в определенном смысле может быть охарактеризована как изучение взаимосвязей экономических переменных.

Инструментом для базового анализа взаимосвязи экономических переменных служат методы математической статистики и эконометрии.

Наиболее простой подход к изучению экономических переменных состоит в исследовании взаимовлияния **двух** переменных (x и y).

Такой подход, с одной стороны, несколько упрощает математические выкладки, а с другой стороны, позволяет в достаточно удобной форме получить соответствующие геометрические интерпретации.

Можно указать **два типа взаимосвязи x и y** :

1. В первом случае нельзя указать, какая из переменных является аргументом, а какая функцией. Тогда отмечают равноправность этих переменных и указывают статистическую взаимосвязь между ними корреляционного типа.

2. Во втором случае имеют ту ситуацию, когда переменные неравноправны, и при этом одна из них считается объясняющей (независимая, аргумент), а другая объясняемой (зависящая, функция).

Это тот случай, когда изменение одной переменной влечет за собой изменение другой.

Снижение процентной ставки ведет к росту инвестиций, повышение валютного курса ведет к уменьшению чистого экспорта, рост дохода ведет к увеличению спроса.

1.2. Построение зависимости между показателями

Пусть рассматриваются два экономических показателя X и Y . Целью является исследование зависимости между ними.

Необходимо выяснить, существует эта зависимость или нет. Если она существует, описать ее формулами, чтобы, зная значение одной, можно было вычислить, какое значение примет другая.

Для этого проводится серия наблюдений, в каждом из которых фиксируются значения обеих величин X и Y . Результатом таких наблюдений является выборка: $(x_1 ; y_1)$

$(x_2 ; y_2)$

$(x_3 ; y_3)$

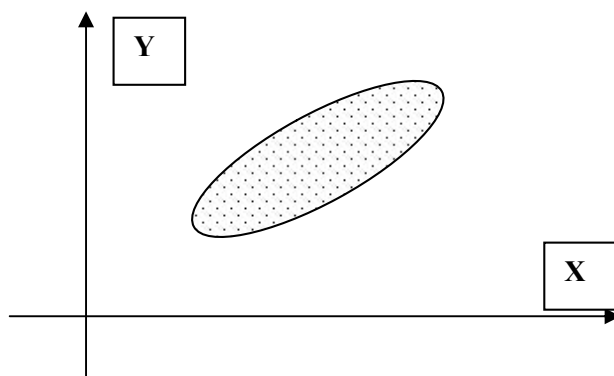
$(x_4 ; y_4)$

.....

$(x_n ; y_n)$

Всего проведено n наблюдений (объем выборки равен n).

Если данные наблюдений нанести на координатную плоскость, получим корреляционное поле:

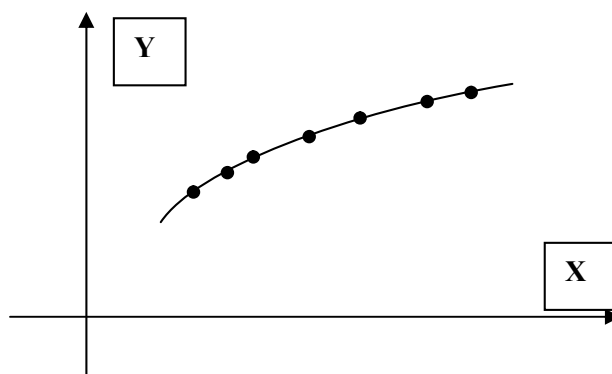


При исследовании двух факторов этот построенный график уже показывает, существует зависимость или нет, характер этой зависимости. В частности, на приведенном графике уже видно, что с ростом фактора X значение результата Y тоже увеличивается. Правда зависимость эта нечеткая, размытая, или, правильно говоря, статистическая.

Выделяют следующие типы зависимостей между показателями:

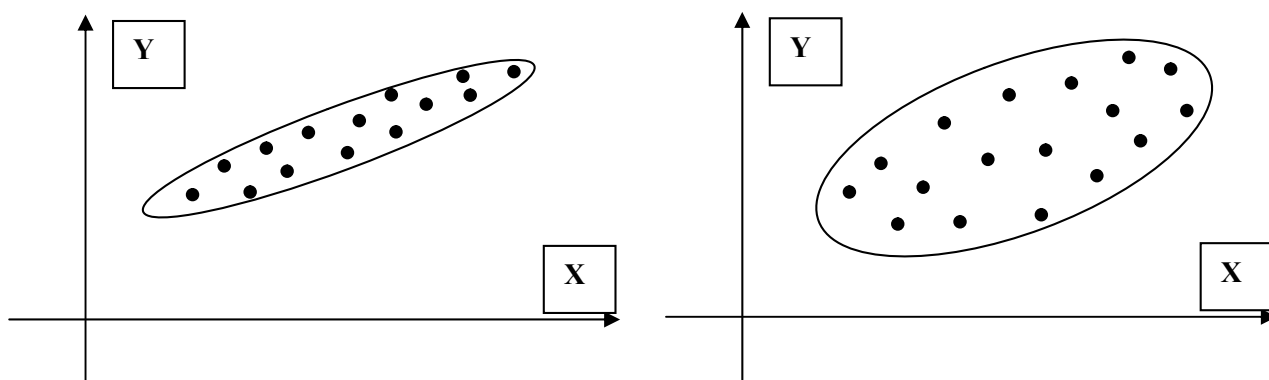
Функциональная зависимость: каждому значению фактора X соответствует только одно значение результативного признака Y .

На корреляционном поле в этом случае мы увидим, что данные наблюдений выстраиваются точно на некоторую линию или кривую.



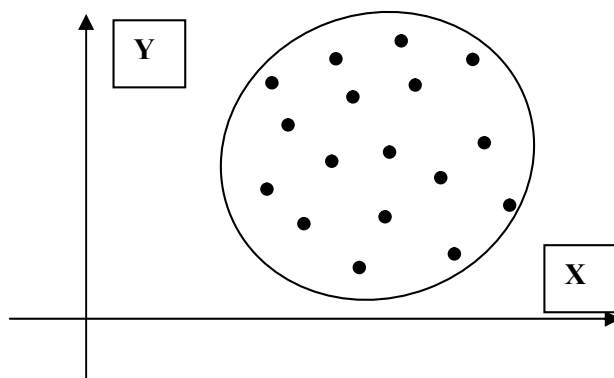
Статистическая зависимость: при одном и том значении фактора X результативный признак Y может принимать различные значения.

На корреляционном поле в этом случае мы увидим, что данные наблюдений размыты, имеется более или менее значительный разброс данных.



В этом случае принято говорить о тесноте статистической зависимости. На левом рисунке представлена тесная статистическая зависимость, на правом – слабая.

Отсутствие зависимости:



1.3. Парная регрессия и её виды

Парная регрессия представляет собой модель зависимости между двумя переменными – y и x , т. е. модель вида: $y = \hat{f}(x)$, где y – зависимая переменная (результативный признак); x – независимая, или объясняющая, переменная (признак-фактор). Знак « $\wedge$ » означает, что между переменными x и y нет строгой функциональной зависимости, поэтому практически в каждом отдельном случае величина y складывается из двух слагаемых: $y = \hat{y}_x + \varepsilon$, где y – фактическое значение результативного признака; $\hat{y}_x$ – теоретическое значение результативного признака, найденное исходя из уравнения регрессии; ε – случайная величина, характеризующая отклонения реального значения

результативного признака от теоретического, найденного по уравнению регрессии.

Модель парной регрессии также называют **однофакторной эконометрической моделью**, т.к. в неё входит только один признак-фактор.

Случайная величина ε называется также возмущением. Она включает влияние неучтенных в модели факторов, случайных ошибок и особенностей измерения. Ее присутствие в модели порождено тремя источниками: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

При построении корреляционной зависимости по данным наблюдений решаются две основные задачи:

1. Определить, какой формулой можно описать зависимость среднего значения y от x . Эту часть исследования называют **спецификацией** уравнения регрессии.

2. После того, как уравнение для описания подобрано, нужно **оценить параметры**, которые в него вошли. Данная процедура будет рассмотрена позже.

В парной регрессии выбор вида математической функции $y = \hat{f}(x)$, т.е. уравнения модели, может быть осуществлен тремя методами:

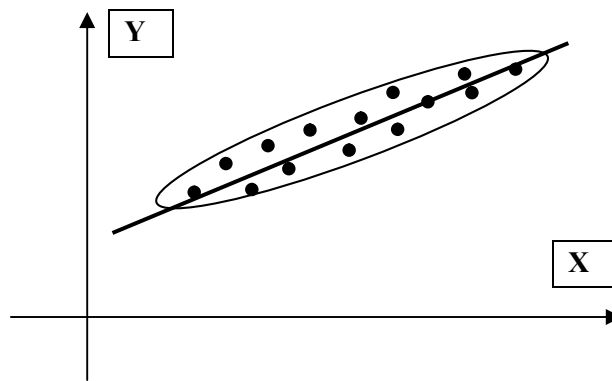
- 1) графическим;
- 2) аналитическим, т.е. исходя из теории изучаемой взаимосвязи;
- 3) экспериментальным.

При изучении зависимости между двумя признаками графический метод подбора вида уравнения регрессии достаточно нагляден. Он основан на поле корреляции.

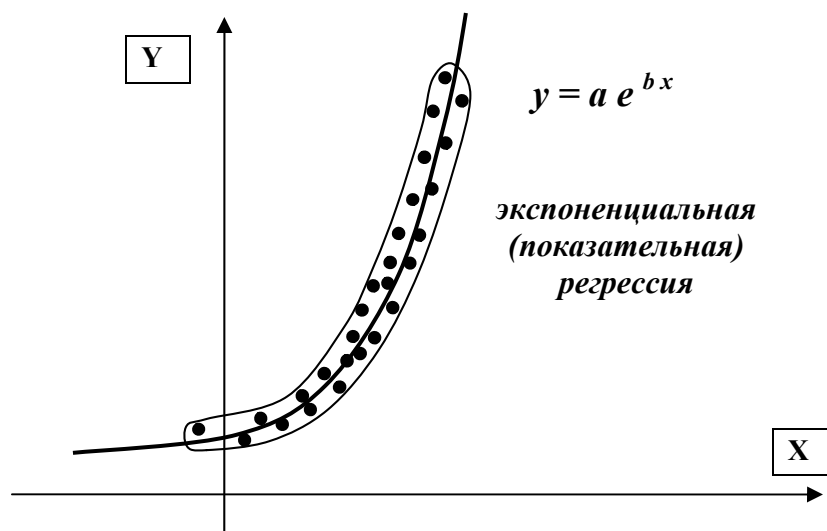
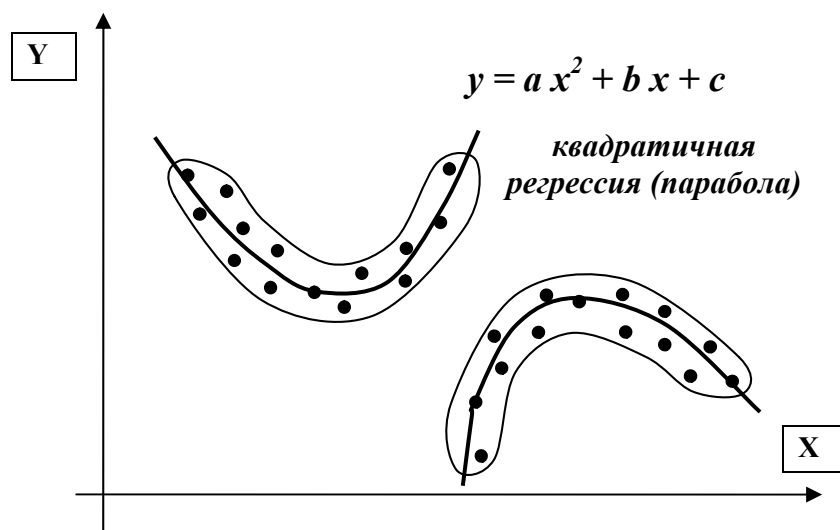
В простейшем случае зависимость среднего значения фактора Y от изменения фактора X можно описать **линейной функцией**:

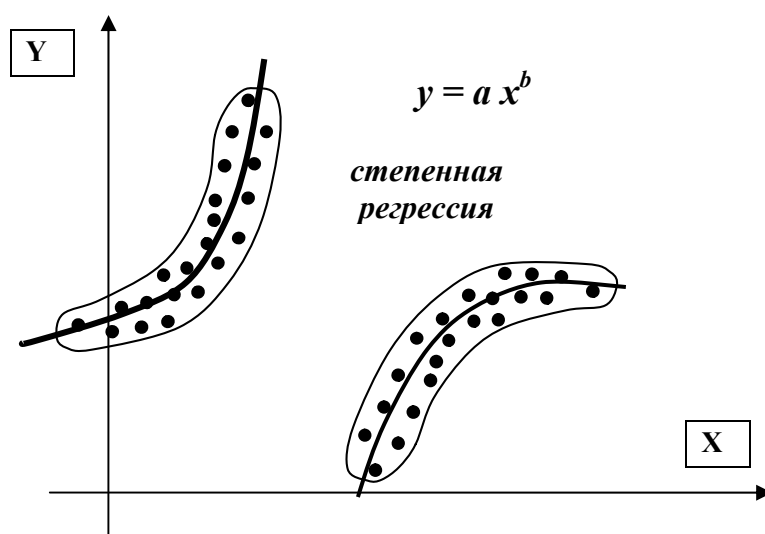
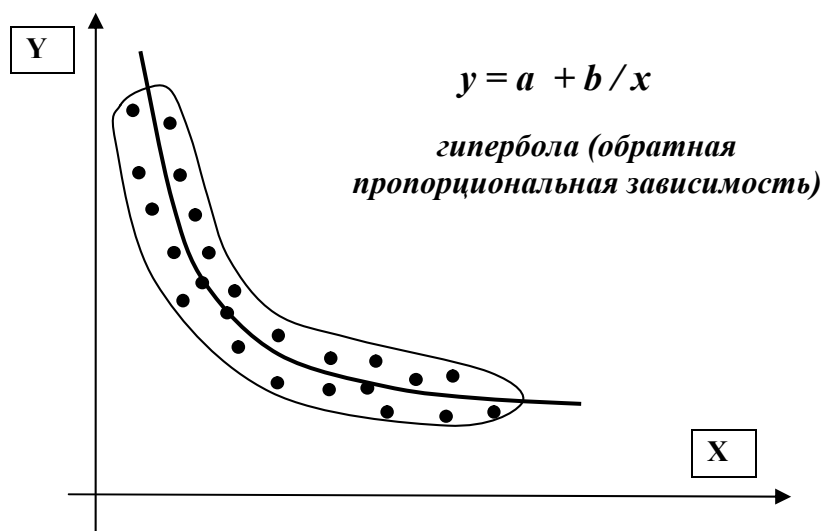
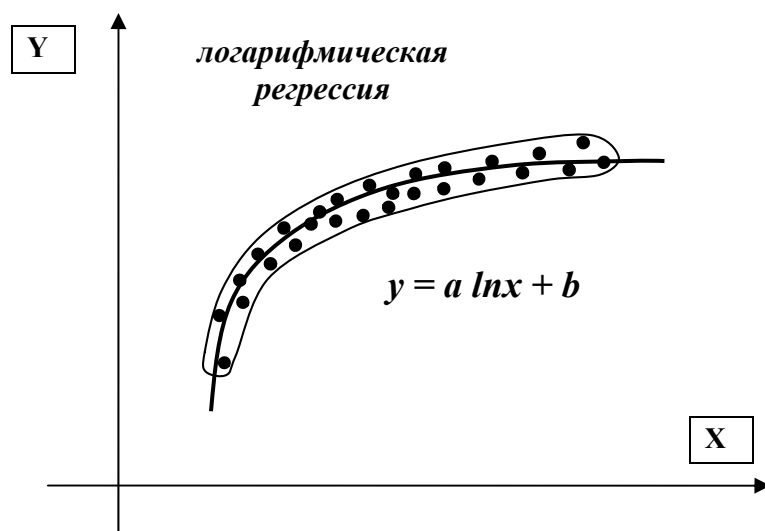
$$\hat{y} = \hat{a}x + \hat{b}$$

Т.е., линия регрессии – прямая линия.



В случаях, когда корреляционное поле указывает на то, что прямая линия не подходит для описания характера зависимости, применяются более сложные виды зависимости, т.е. **нелинейная регрессия**. Форма этой зависимости может быть разной. В зависимости от того, какое уравнение выбирается, регрессия получает соответствующее название.





Значительный интерес представляет *аналитический метод* выбора типа уравнения регрессии. Он основан на изучении материальной природы связи исследуемых признаков.

При обработке информации на компьютере выбор вида уравнения регрессии обычно осуществляется *экспериментальным методом*, т. е. путем сравнения величины остаточной дисперсии, рассчитанной при разных моделях.

Считается, что число наблюдений должно в 7-8 раз превышать число рассчитываемых параметров при переменной x . Это означает, что искать линейную регрессию, имея менее 7 наблюдений, вообще не имеет смысла. Если вид функции усложняется, то требуется увеличение объема наблюдений, ибо каждый параметр при x должен рассчитываться хотя бы по 7 наблюдениям. Значит, если мы выбираем параболу, то нам потребуется объем информации уже не менее 14 наблюдений.

ГЛАВА 2 . МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

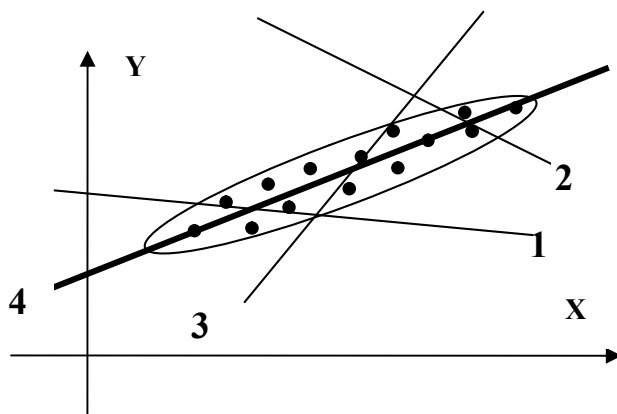
2.1. Процедура метода наименьших квадратов

2.2. Предпосылки метода наименьших квадратов

2.1. Процедура метода наименьших квадратов

В записанное уравнение обязательно входят коэффициенты (параметры), и эти коэффициенты нужно находить, определять. Причем сделать это таким образом, чтобы построенное уравнение наилучшим образом описывало данные наблюдений, чтобы линия регрессии проходила как можно ближе к точкам корреляционного поля. Этот этап называется: **статистическая оценка параметров регрессии**.

Например, изобразим случай, когда корреляционное поле показывает, что можно применить линейную регрессию.



В уравнение прямой ($y = ax + b$) входят два параметра: a и b . Меняя их, получаем различные прямые линии. Выбрать нужно только одну, с конкретными значениями параметров. Ту, которая на рисунке изображена под номером 4.

В общем случае задача ставится таким образом.

Выбрана формула для уравнения регрессии в виде прямой линии:

$$y = f(x, a, b) \text{ или } \hat{y}_x = a \cdot x + b.$$

Подобрать параметры регрессии таким образом, чтобы отклонения данных наблюдений от линии регрессии были бы минимальными.

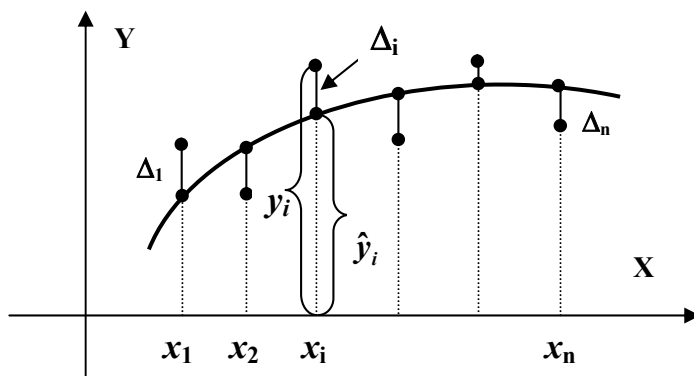
Для решения этой задачи имеется несколько методов. Самый популярный и наиболее часто используемый из них – **метод наименьших квадратов (МНК)**.

Чтобы измерить степень отклонения кривой от экспериментальных точек (или наоборот, экспериментальных точек от кривой, которую мы строим) введем следующее определение:

Отклонением (или остатком, обозначается в литературе по-разному ε_i , Δ_i либо u_i) назовем разность между теоретическим значением $\hat{y}_i$, подсчитанным по предполагаемой формуле и экспериментальным значением y_i :

$$\Delta_i = \hat{y}_i - y_i = f(x_i, a, b) - y_i$$

Здесь x_i , y_i – это данные наблюдений, известные числа. Покажем эти отклонения на графике:



Все они должны быть как можно меньше. Этого можно добиться, если взять сумму квадратов этих чисел и потребовать, чтобы она была минимальна:

$$\sum \Delta_i^2 \rightarrow \min$$

$$\sum \Delta_i^2 = \sum (\hat{y}_i - y_i)^2 = \sum [f(x_i, a, b) - y_i]^2 = S(a, b) \rightarrow \min$$

Т.е., эта сумма квадратов является функцией от коэффициентов регрессии, и нужно найти минимум функции нескольких переменных. Из курса высшей математики известно, что для этого нужно взять частные производные и приравнять их нулю:

$$\begin{cases} \frac{\partial S}{\partial a} = 0 \\ \frac{\partial S}{\partial b} = 0 \end{cases}$$

Получим систему уравнений для отыскания коэффициентов регрессии. Сколько коэффициентов присутствует в уравнении регрессии, столько будет и уравнений в этой системе.

Как известно из курса математического анализа, чтобы найти минимум функции, надо вычислить частные производные по каждому из параметров a и b и приравнять их к нулю. Обозначим $\sum \Delta_i^2$ через $S(a, b)$, тогда: $S(a, b) = \sum (y - a \cdot x - b)^2$.

$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum (y - a \cdot x - b) = 0 \\ \frac{\partial S}{\partial b} = -2 \sum x(y - a \cdot x - b) = 0 \end{cases}$$

Вычислим производные и выполним преобразования полученных выражений:

$$\begin{cases} \frac{\partial}{\partial a} \sum [(a x_i + b) - y_i]^2 = \sum 2 \cdot [(a x_i + b) - y_i] \cdot (x_i) = 2 \sum [a x_i^2 + b x_i - x_i \cdot y_i] = 0; \\ \frac{\partial}{\partial b} \sum [(a x_i + b) - y_i]^2 = \sum 2 \cdot [(a x_i + b) - y_i] \cdot (1) = 2 \sum [a x_i + b - y_i] = 0; \end{cases}$$

После преобразований, получим следующую систему линейных уравнений для оценки параметров a и b :

$$\begin{cases} n \cdot b + \sum_{i=1}^n x_i \cdot a = \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i \cdot b + \sum_{i=1}^n x_i^2 \cdot a = \sum_{i=1}^n x_i y_i \end{cases}$$

Эта система называется *нормальной системой*.

Если мы ее решим, то коэффициенты регрессии a, b будут найдены.

Свое название МНК получил, исходя из смыслового содержания критерия, которому должны удовлетворять полученные на его основе оценки параметров эконометрической модели: *сумма квадратов значений фактической ошибки модели должна быть минимальной*.

2.2. Предпосылки метода наименьших квадратов

Оценивание модели (уравнения) регрессии по методу наименьших квадратов предусматривает проверку выполнимости предпосылок МНК.

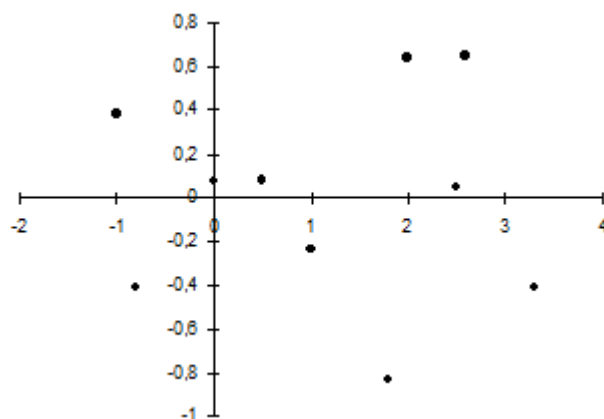
Первая предпосылка МНК

Математическое ожидание, т.е. средняя величина случайного отклонения ε_i (остатков) для всех наблюдений равна нулю. Это означает, что случайное отклонение в среднем не оказывает влияния на зависимую переменную. В каждом конкретном наблюдении случайный член может быть либо положительным, либо отрицательным, но он не должен иметь систематического смещения.

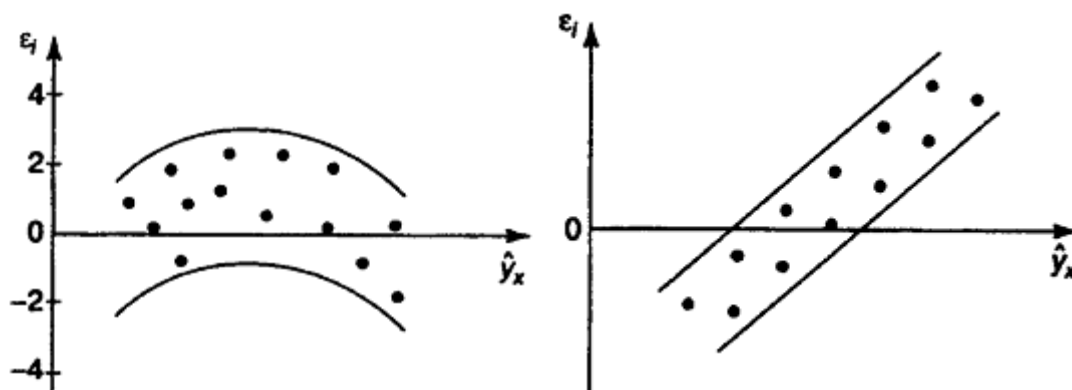
Если ошибка имеет не нулевое среднее значение (матожидание), то оценки по МНК окажутся смещенными. Другими словами, в ошибке осталась детерминированная (неслучайная) составляющая, которая может быть функцией входящих в модель регрессоров, что и означает, что функциональная форма выбрана неверно.

Вторая предпосылка МНК

Случайный характер остатков ε_i . Для проверки этого свойства определяют значения остатков ε_i и строится график зависимости остатков ε_i от теоретических значений результативного признака. На графике ниже приведены остатки ε_i , которые представляют собой случайные величины. В этом случае МНК оправдан, теоретические значения y_x независимы от ε_i .



При этом возможны следующие случаи, если остатки ε_i зависят от y_x или остатки ε_i не случайны.



Третья предпосылка МНК

Дисперсия случайных отклонений σ_ε^2 постоянна для любых наблюдений. Данное условие подразумевает, что несмотря на то, что при каждом конкретном наблюдении случайное отклонение может быть либо большим, либо меньшим, а следовательно не должно быть причины, вызывающей большое отклонение (ошибку). Выполнимость данной предпосылки называется **гомоскедастичностью** (постоянством дисперсии отклонений σ_ε^2). Невыполнимость данной предпосылки называется **гетероскедастичностью** (непостоянством дисперсий отклонений σ_ε^2).

Четвертая предпосылка МНК

Важной предпосылкой построения качественной регрессионной модели по МНК является независимость значений случайных отклонений от значений отклонений во всех других наблюдениях. Это гарантирует отсутствие

коррелированности между любыми отклонениями и, в частности, между соседними отклонениями.

Автокорреляция (последовательная корреляция) определяется как корреляция между наблюдаемыми показателями, упорядоченными во времени (временные ряды) или в пространстве (перекрестные ряды). Автокорреляция остатков (отклонений) обычно встречается в регрессионном анализе при использовании данных временных рядов и очень редко при использовании перекрестных данных.

При построении регрессионных моделей важно соблюдение отсутствия автокорреляции остатков, т.е. значения остатков должны быть распределены независимо друг от друга. Выполнимость данной предпосылки предполагает, что отсутствует систематическая связь между любыми случайными отклонениями. Если данное условие выполняется, то говорят об отсутствии автокорреляции.

Пятая предпосылка МНК

В случае построения многофакторной эконометрической модели также важным условием является ***отсутствие мультиколлинеарности***. Мультиколлинеарность – это наличие линейной корреляции (зависимости) объясняющих переменных (факторов) между собой.

Если все эти пять предпосылок МНК выполняются, то коэффициенты регрессии, найденные исходя из системы нормальных уравнений, будут несмещенными, эффективными и состоятельными.

При несоблюдении основных предпосылок МНК приходится корректировать модель, изменяя ее спецификацию, добавлять или исключать некоторые факторы, преобразовывать исходные данные, чтобы получить оценки коэффициентов регрессии, обладающих свойствами несмещенности, эффективности и состоятельности.

ГЛАВА 3. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

(МНОГОФАКТОРНАЯ ЭКОНОМЕТРИЧЕСКАЯ МОДЕЛЬ)

3.1. Спецификация модели. Отбор факторов при построении уравнения множественной регрессии

3.2. Матричная форма МНК

3.1. Спецификация модели. Отбор факторов при построении уравнения множественной регрессии

Парная регрессия может дать хороший результат при моделировании, если влиянием других факторов, воздействующих на объект исследования, можно пренебречь. Если же этим влиянием пренебречь нельзя, то в этом случае следует попытаться выявить влияние других факторов, введя их в модель, т.е. построить уравнение множественной регрессии

$$y = \hat{f}(x_1, x_2, \dots, x_m),$$

где y – зависимая переменная (результативный признак), x_i – независимые, или объясняющие, переменные (признаки-факторы).

Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели. Он включает в себя два круга вопросов: отбор факторов и выбор вида уравнения регрессии.

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

Отбор факторов производится на основе качественного теоретико-экономического анализа. Однако теоретический анализ часто не позволяет однозначно ответить на вопрос о количественной взаимосвязи рассматриваемых признаков и целесообразности включения фактора в модель. Поэтому отбор факторов обычно осуществляется в две стадии: на первой подбираются факторы исходя из сущности проблемы; на второй – на основе

матрицы показателей корреляции определяют статистики для параметров регрессии.

Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии мультиколлинеарности факторов, когда более чем два фактора связаны между собой линейной зависимостью, т.е. имеет место совокупное воздействие факторов друг на друга.

Существует ряд подходов преодоления сильной межфакторной корреляции. Самый простой путь устранения мультиколлинеарности состоит в исключении из модели одного или нескольких факторов. Другой подход связан с преобразованием факторов, при котором уменьшается корреляция между ними.

При отборе факторов также рекомендуется пользоваться следующим правилом: число включаемых факторов обычно в 6–7 раз меньше объема совокупности, по которой строится регрессия.

3.2. Матричная форма МНК

Возможны разные виды уравнений множественной регрессии: линейные и нелинейные.

Ввиду четкой интерпретации параметров наиболее широко используется линейная функция. В линейной множественной регрессии $\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \dots + \hat{a}_m x_m$ параметры при x называются коэффициентами «чистой» регрессии. Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

Рассмотрим линейную модель множественной регрессии:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m + u.$$

Классический подход к оцениванию параметров линейной модели множественной регрессии основан на уже известном нам МНК.

При этом линейная эконометрическая модель записывается в векторно-матричной форме следующего вида:

$$Y=X \cdot A+\varepsilon,$$

где Y – вектор-столбец, состоящий из n компонент зависимой (результативной) переменной;

X – матрица размера $n \times (m+1)$ (если в модели присутствует “свободный” коэффициент α_0);

$A=(\alpha_0, \alpha_1, \dots, \alpha_m)$ – вектор-столбец параметров, состоящий из $m+1$ -й компоненты;

ε – вектор-столбец ошибки модели (остатков или отклонений), состоящий, как и вектор Y , из n компонент.

Исходными данными при оценке параметров $\alpha_0, \alpha_1, \dots, \alpha_m$ являются измеренные (наблюдаемые) значения зависимой переменной, которые можно представить в виде вектора-столбца,

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$$

и наблюдаемые значения независимых переменных (факторов), которые объединены в матрицу следующего вида:

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{1m} \\ 1 & x_{12} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{nm} \end{pmatrix}.$$

В матрице X первый столбец состоит сугубо из значений “1” является необходимым для расчета значения свободного члена (параметра α_0) линейной эконометрической модели. Следует заметить, что расчет параметров данной модели можно производить и без столбца со значениями “1”, но в этом случае параметр α_0 нужно будет находить отдельно. С этой целью используется формула:

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}_1 - \hat{a}_2 \bar{x}_2 - \dots - \hat{a}_m \bar{x}_m,$$

где $\bar{y}$, $\bar{x}_1$, $\bar{x}_2$, ..., $\bar{x}_m$ – средние значения соответственно зависимой и независимых переменных (факторов).

Вектор оценок параметров A линейной эконометрической модели в матричной форме определяется на основе следующего выражения:

$$A = (X^T X)^{-1} \cdot (X^T Y).$$

Приведенное выше выражение ещё называют как **оператор оценивания**. При этом значения вектора A являются решением уже известной нам системы нормальных уравнений МНК. В векторно-матричной форме эта система записывается в виде:

$$(X^T X)A = X^T Y$$

Следует отметить, что оператор оценивания может быть использован исключительно для расчета параметров линейной эконометрической модели, причем не только множественной регрессии, но и парной. В случае, если модель является нелинейной нужно предварительно преобразовать её – свести к линейной и только потом использовать выражение $A = (X^T X)^{-1} \cdot (X^T Y)$.

ГЛАВА 4. АНАЛИЗ АДЕКВАТНОСТИ ЭКОНОМЕТРИЧЕСКОЙ МОДЕЛИ И ПОКАЗАТЕЛИ КАЧЕСТВА РЕГРЕССИИ

- 4.1. Показатели тесноты связи
- 4.2. Оценка качества подбора вида функции
- 4.3. Оценка значимости уравнения регрессии в целом
- 4.4. Показатели качества отдельных параметров уравнения регрессии

4.1. Показатели тесноты связи

Формы проявления взаимосвязей весьма разнообразны. В качестве самых общих их видов выделяют функциональную (полную) и корреляционную (неполную) связи.

Показатели тесноты связи дают возможность охарактеризовать зависимость вариации результативного признака от вариации признака-фактора.

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии ($\hat{y}_x = \hat{a} \cdot x + \hat{b}$) в качестве такого показателя выступает линейный **коэффициент корреляции** r_{xy} , который можно рассчитать по формуле:

$$r_{xy} = \hat{a} \cdot \frac{\sigma_x}{\sigma_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y},$$

где σ_x и σ_y – среднеквадратическое отклонение соответственно признака-фактора и результативного признака:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}; \sigma_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}.$$

Линейный коэффициент корреляции находится в пределах: $-1 \leq r_{xy} \leq 1$.

Чем ближе абсолютное значение r_{xy} к единице, тем сильнее линейная связь между факторами (при $r_{xy} = \pm 1$ имеем строгую функциональную зависимость).

Положительное значение коэффициента корреляции говорит о положительной связи между x и y , когда с ростом одной из переменных другая тоже растет. Отрицательное значение коэффициента корреляции означает, с ростом одной из переменных другая убывает, с убыванием одной из переменных другая растет.

Теснота линейной связи между переменными может быть оценена на основании шкалы Чеддока:

Теснота связи	Значение коэффициента корреляции при наличии:	
	Прямой связи	Обратной связи
Слабая	0,1–0,3	(–0,3)–(–0,1)
Умеренная	0,3–0,5	(–0,5)–(–0,3)
Заметная	0,5–0,7	(–0,7)–(–0,5)
Высокая	0,7–0,9	(–0,9)–(–0,7)
Очень высокая	0,9–1	(–1)–(–0,9)

Коэффициенты регрессии (в нашем примере a) нежелательно использовать для непосредственной оценки влияния факторов на результативный признак в том случае, если существует различие единиц измерения результативного показателя y и факторного признака x .

Для этих целей вычисляются коэффициенты эластичности. Средний **коэффициент эластичности** $E_{y/x}$ 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.

Коэффициент эластичности находится по формуле:

$$E_{y/x} = \hat{a} \cdot \frac{\bar{x}}{\bar{y}}$$

Для нелинейной или многофакторной формы зависимости теснота связи как правило определяется с помощью **множественного коэффициента корреляции**. Его можно вычислить как квадратный корень коэффициента детерминации, который будет рассмотрен ниже.

4.2. Оценка качества подбора вида функции

Для оценки качества подбора линейной функции в модели парной регрессии рассчитывается квадрат линейного коэффициента корреляции $D = r_{yx}^2$, называемый **коэффициентом детерминации**.

Коэффициент детерминации характеризует долю дисперсии результативного признака y , объясняемую регрессией, в общей дисперсии результативного признака.

Дисперсия (обозначается как σ^2) – статистический разброс чего-либо и численная характеристика такого разброса

Коэффициент детерминации, как правило, рассчитывают по формуле:

$$D = r_{yx}^2 = 1 - \frac{\sigma_{\varepsilon}^2}{\sigma_y^2}, \quad 0 \leq D \leq 1,$$

где $\sigma_{\varepsilon}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ – дисперсия отклонений;

$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ – дисперсия результативного признака.

Соответственно величина $1-D$ характеризует долю дисперсии результативного признака, вызванную влиянием остальных, неучтенных в модели, факторов.

Квадратный корень коэффициента детерминации дает нам показатель **множественного коэффициента корреляции**, который характеризует тесноту связи всех включенных в модель факторов с результативным признаком:

$$R = \sqrt{D}, \quad 0 \leq R \leq 1.$$

Итак коэффициент детерминации показывает долю вариации результативного признака, объясненную вариацией факторного признака.

Чаще всего, давая интерпретацию коэффициента детерминации, его выражают в процентах. Например, $D = 0,8023$. Значит в 80,23% случаев изменения x приводят к изменению y . Другими словами – точность подбора уравнения регрессии – высокая. Остальные 19,77 % изменения y объясняются факторами, неучтенными в модели.

Таким образом, чем ближе значение коэффициента детерминации D к единице (100%) тем более качественно построена эконометрическая модель, т.к. более правильно подобран вид функции. Для приемлемых моделей предполагается, что коэффициент детерминации должен быть хотя бы не меньше 50 % (в этом случае коэффициент множественной корреляции превышает 70 %). Модели с коэффициентом детерминации выше 80 % можно признать достаточно хорошими (коэффициент корреляции превышает 90 %). Значение коэффициента детерминации 1 означает функциональную зависимость между переменными, т.е. модельные значения результативного признака на 100% соответствуют его статистическим (экспериментальным, реальным) данным.

4.3. Оценка значимости уравнения регрессии в целом

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.

Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}_x}{y} \right| \cdot 100\%.$$

Средняя ошибка аппроксимации не должна превышать 8–10%.

Оценка значимости уравнения регрессии в целом производится на основе ***F-критерия Фишера***:

$$F_{\text{факт}} = \frac{D / m}{(1-D) / (n-m-1)},$$

где m – число факторов в эконометрической модели.

Фактическое значение F -критерия Фишера сравнивается с табличным (критическим) значением $F_{\text{табл}}$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$. При этом, если фактическое (расчетное) значение F -критерия больше табличного, то признается статистическая значимость уравнения в целом.

Для парной линейной регрессии $m = 1$, поэтому:

$$F_{\text{факт}} = \frac{D}{1-D} \cdot (n-2).$$

4.4. Показатели качества отдельных параметров уравнения регрессии

В любых эконометрических моделях оценивается значимость не только уравнения в целом, но и отдельных его параметров. Рассмотрим данную процедуру на примере парной линейной регрессии ($\hat{y}_x = \hat{a} \cdot x + \hat{b}$). С этой целью по каждому из параметров определяется его стандартная ошибка: S_a и S_b .

Стандартная ошибка коэффициента регрессии, который стоит перед фактором-признаком x определяется по формуле:

$$S_a = \frac{\sigma_{\text{остат}}^2}{\sigma_x \cdot \sqrt{n}},$$

где $\sigma_{\text{остат}}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}$ – остаточная дисперсия на одну степень свободы.

Стандартная ошибка параметра a определяется по формуле:

$$S_b = \sigma_{\text{остат}}^2 \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n}$$

Величина стандартной ошибки совместно с t -распределением Стьюдента при $n-2$ степенях свободы применяется для проверки существенности коэффициента регрессии и для расчета его доверительного интервала.

Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т.е. определяется **фактическое значение t -критерия Стьюдента** ($t_a = \frac{|a|}{S_a}$ для коэффициента a или $t_b = \frac{|b|}{S_b}$ для коэффициента b) которое затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы $(n-2)$.

Если фактическое значение t -критерия Стьюдента превышает табличное, то соответствующий параметр эконометрической модели является достоверным.

ЛАБОРАТОРНЫЙ ПРАКТИКУМ

Лабораторная работа № 1. Построение и анализ однофакторной эконометрической модели

Задание

По данным затрат на продвижение некоторого товара (x , тыс. дол. США) и размером полученной прибыли (y , тыс. дол. США) за шестнадцать месяцев построить однофакторную эконометрическую модель (модель парной регрессии). Оценить параметры модели с помощью метода наименьших квадратов (МНК) и провести анализ её адекватности.

Исходные данные представлены в таблице 1.1 ниже.

Примечание: индивидуальный номер варианта формируется с помощью значения N – это две последние цифры номера Вашей зачетной книжки.

В случае, если предпоследняя цифра номера Вашей зачетной книжки равняется нулю, она игнорируется и учитывается только последняя цифра.

Таблица 1.1

№ п/п	x_i	y_i
1	$48+N/10$	$1590+N/10$
2	$55+N/10$	$1760+N/10$
3	$61+N/10$	$1975+N/10$
4	$67+N/10$	$2205+N/10$
5	$78+N/10$	$2460+N/10$
6	$88+N/10$	$2725+N/10$
7	$91+N/10$	$3035+N/10$
8	$114+N/10$	$3320+N/10$
9	$122+N/10$	$3485+N/10$
10	$147+N/10$	$3575+N/10$
11	$163+N/10$	$3785+N/10$
12	$200+N/10$	$4025+N/10$
13	$240+N/10$	$4150+N/10$
14	$275+N/10$	$4275+N/10$
15	$300+N/10$	$4540+N/10$
16	$330+N/10$	$4730+N/10$

Пример выполнения

Пусть две последних цифры зачетной книжки **N=82**.

Будем считать, что между фактором (сумма расходов на продвижение товара) и результативной переменной (размер полученной прибыли) существует *линейная зависимость*. В таком случае имеем простейшую эконометрическую модель, а именно *линейную однофакторную модель* (модель парной регрессии) общего вида (см. § 1.3):

$$\hat{y} = \hat{a}x + \hat{b},$$

где $\hat{a}$ и $\hat{b}$ – неизвестные параметры модели (коэффициенты, которые нам нужно рассчитать или оценить).

Рассчитывать (оценивать) параметры нашей модели будем при помощи *метода наименьших квадратов* (см. § 2.1).

С этой целью выполним следующие шаги.

1. Все данные с таблицы 1.1 увеличиваем на 8,2 (82/10). Результаты заносим в программу Microsoft Excel. Получим следующую таблицу.

Таблица 1.2

<i>№ n/n</i>	<i>x_i</i>	<i>y_i</i>
1	56,2	1598,2
2	63,2	1768,2
3	69,2	1983,2
4	75,2	2213,2
5	86,2	2468,2
6	96,2	2733,2
7	99,2	3043,2
8	122,2	3328,2
9	130,2	3493,2
10	155,2	3583,2
11	171,2	3793,2
12	208,2	4033,2
13	248,2	4158,2
14	283,2	4283,2
15	308,2	4548,2
16	338,2	4738,2

2. Находим столбцы значений x_i^2 и $x_i \cdot y_i$, а также суммы по всем столбцам (для этого пользуемся функцией Excel СУММ либо значком автосуммы Σ). Получим следующую таблицу.

Таблица 1.3

№ n/n	x_i	y_i	x_i^2	$x_i y_i$
1	56,2	1598,2	3158,44	89818,84
2	63,2	1768,2	3994,24	111750,24
3	69,2	1983,2	4788,64	137237,44
4	75,2	2213,2	5655,04	166432,64
5	86,2	2468,2	7430,44	212758,84
6	96,2	2733,2	9254,44	262933,84
7	99,2	3043,2	9840,64	301885,44
8	122,2	3328,2	14932,84	406706,04
9	130,2	3493,2	16952,04	454814,64
10	155,2	3583,2	24087,04	556112,64
11	171,2	3793,2	29309,44	649395,84
12	208,2	4033,2	43347,24	839712,24
13	248,2	4158,2	61603,24	1032065,24
14	283,2	4283,2	80202,24	1213002,24
15	308,2	4548,2	94987,24	1401755,24
16	338,2	4738,2	114379,24	1602459,24
Сума	2510,2	51766,2	523922,44	9438840,64

2. Полученные значения сумм подставляем в систему нормальных уравнений, которая поможет нам найти неизвестные коэффициенты $\hat{a}$ и $\hat{b}$.

$$\begin{cases} n \cdot b + \sum_{i=1}^n x_i \cdot a = \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i \cdot b + \sum_{i=1}^n x_i^2 \cdot a = \sum_{i=1}^n x_i y_i \end{cases}$$

В нашем случае:

$$\begin{cases} 16b + 2510,2a = 51766,2 \\ 2510,2b + 523922,44a = 9438840,64 \end{cases}$$

3. Решаем систему уравнений одним из общеизвестных методов (методом Крамера, методом Гаусса либо матричным методом).

Примечание. Все вычисления рекомендуется выполнять строго в программе Microsoft Excel, что позволит уменьшить количество округлений и технических ошибок!

Например, согласно методу Гаусса разделим первое уравнение на 16, а второе – на 2510,2.

Получим следующую систему:

$$\begin{cases} b + 156,8875a = 3235,3875 \\ b + 208,7174a = 3760,1946 \end{cases}$$

Далее отнимем от второго уравнения первое (можно и наоборот – от первого уравнения второе, на итоговый результат это не повлияет).

Неизвестный параметр b сокращается, получим:

$$208,7174a - 156,8875a = 3760,1946 - 3235,3875;$$

$$51,8299a = 524,8071 \Rightarrow a = \frac{524,8071}{51,8299} = 10,1255.$$

Таким образом, *первый неизвестный параметр мы нашли:* $a = 10,1255$.

Для того, чтобы найти второй параметр нужно подставить уже известное нам значение коэффициента a в любое из уравнений предыдущей системы, например в первое. Получим:

$$b + 156,8875 \cdot 10,1255 = 3235,3875 \Rightarrow b = 3235,3875 - 156,8875 \cdot 10,1255 = 1646,9021.$$

Таким образом, *и второй неизвестный параметр мы тоже нашли:*

$$b = 1646,9021.$$

4. Записываем уравнение модели, подставив найденные значения параметров a и b в линейную форму $\hat{y} = \hat{a}x + \hat{b}$.

Аналитическая форма однофакторной эконометрической модели будет иметь следующий вид:

$$\hat{y} = 10,1255x + 1646,9021.$$

Данная модель показывает зависимость между фактором (x , сумма расходов на продвижение товара) и результативной переменной (y , размер полученной прибыли)

5. Находим колонку значений $\hat{y}$ (используем формулу $\hat{y} = 10,1255x + 1646,9021$ и значения фактора x со второй колонки) и добавляем её к таблице 1.3. Получим следующую таблицу.

Таблица 1.4

№ n/n	x_i	y_i	x_i^2	$x_i y_i$	$\hat{y}_i$
1	2	3	4	5	6
1	56,2	1598,2	3158,44	89818,84	2215,93
2	63,2	1768,2	3994,24	111750,24	2286,80
3	69,2	1983,2	4788,64	137237,44	2347,55
4	75,2	2213,2	5655,04	166432,64	2408,30
5	86,2	2468,2	7430,44	212758,84	2519,68
6	96,2	2733,2	9254,44	262933,84	2620,93
7	99,2	3043,2	9840,64	301885,44	2651,30
8	122,2	3328,2	14932,84	406706,04	2884,18
9	130,2	3493,2	16952,04	454814,64	2965,18
10	155,2	3583,2	24087,04	556112,64	3218,30
11	171,2	3793,2	29309,44	649395,84	3380,30
12	208,2	4033,2	43347,24	839712,24	3754,93
13	248,2	4158,2	61603,24	1032065,24	4159,93
14	283,2	4283,2	80202,24	1213002,24	4514,30
15	308,2	4548,2	94987,24	1401755,24	4767,43
16	338,2	4738,2	114379,24	1602459,24	5071,18
Сума	2510,2	51766,2	523922,44	9438840,64	51766,21

6. Чтобы получить *графическую интерпретацию модели*, построим в системе координат xOy прямую, которая соответствует полученному уравнению модели. Это будут теоретические (расчетные, модельные) значения зависимой переменной $\hat{y}$.

Нанесем также на той же координатной плоскости точки, которые соответствуют исходным, т.е. фактическим данным переменной y .

С этой целью используем Мастер диаграмм Microsoft Excel (пункт меню *Вставка→Диаграмма* либо пиктограмма ):

➤ Выбираем тип диаграммы – точечная, вид можно оставить по умолчанию, т.е. первый.

➤ Нажимаем на кнопку "Далее", переключаемся на вкладку "Ряд" (вкладки расположены вверху окна).

➤ В случае, если в поля "Ряд" мастером уже занесены какие-либо данные –очищаем их нажимая на кнопку "Удалить" до тех пор пока она станет неактивной.

➤ Нажимаем на кнопку "Добавить" чтобы создать первый ряд данных.

➤ В поле "Имя:" вводим с клавиатуры текст *фактические значения*.

➤ В поле "Значения X:" указываем столбец со значениями x со второй колонки таблицы 1.4 (выделяем диапазон значений мышкой, номер колонки и значение суммы в диапазон не входит!).

➤ В поле "Значения Y:" указываем столбец со значениями y с третьей колонки таблицы 1.4 (выделяем диапазон значений мышкой, номер колонки и значение суммы в диапазон не входит!).

➤ Нажимаем на кнопку "Добавить" чтобы создать второй ряд данных.

➤ В поле "Имя:" вводим с клавиатуры текст *модельные значения*.

➤ В поле "Значения X:" указываем столбец со значениями x со второй колонки таблицы 1.4 (выделяем диапазон значений мышкой, номер колонки и значение суммы в диапазон не входит!).

➤ В поле "Значения Y:" указываем столбец со значениями $\hat{y}$ с шестой колонки таблицы 1.4 (выделяем диапазон значений мышкой, номер колонки и значение суммы в диапазон не входит!).

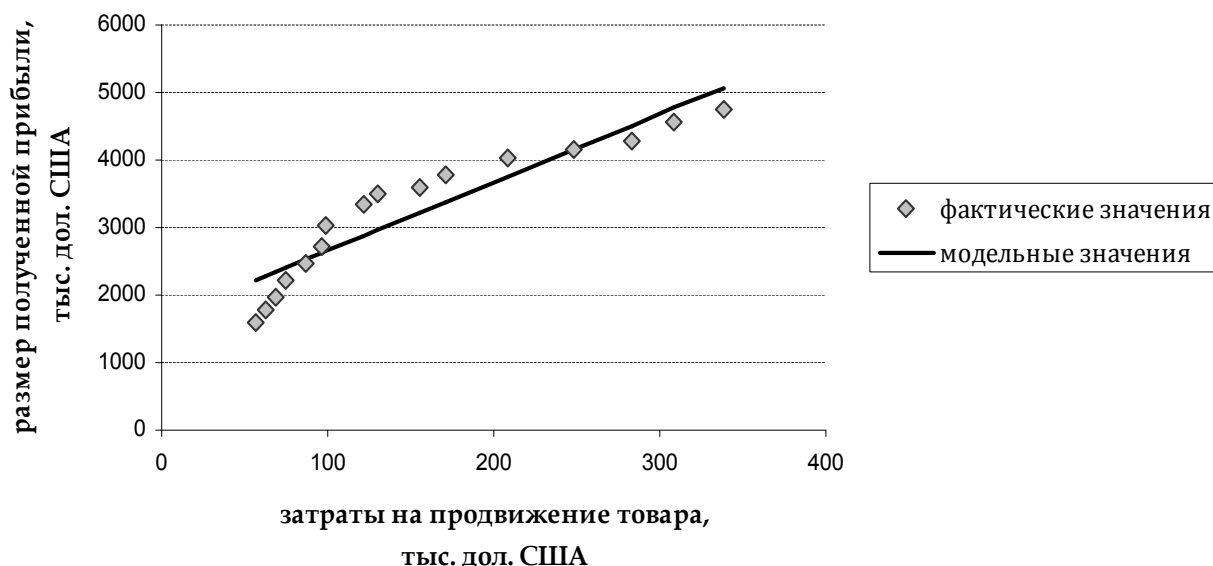
➤ Нажимаем на кнопку "Далее", вводим с клавиатуры названия для осей X (затраты на продвижение товара, тыс. дол. США) и Y (размер полученной прибыли, тыс. дол. США).

➤ Нажимаем на кнопку "Далее", потом на кнопку "Готово".

➤ Форматируем график: нажимаем мышкой на ряд данных с модельными значениями, заходим в меню *Формат* → *Выделенный ряд*, в появившемся окне в поле "Линия" ставим переключатель в положение "обычная", а в поле "Маркер" – в положение "отсутствует".

➤ Меняем цвета рядов данных, шрифты, заливку области построения, внешний вид линий сетки и осей по собственному усмотрению.

Результат представлен на рисунке ниже.



Как видно с рисунка теоретическая линия в целом достаточно хорошо отражает основные тенденции взаимосвязи между переменными x и y .

7. Находим *столбец отклонений* $\Delta_i = \hat{y}_i - y_i$ и добавляем его в таблицу.

Таблица 1.5

№ n/n	x_i	y_i	x_i^2	$x_i y_i$	$\hat{y}_i$	Δ_i
1	2	3	4	5	6	7
1	56,2	1598,2	3158,44	89818,84	2215,93	-617,73
2	63,2	1768,2	3994,24	111750,24	2286,80	-518,6
3	69,2	1983,2	4788,64	137237,44	2347,55	-364,35
4	75,2	2213,2	5655,04	166432,64	2408,30	-195,1
5	86,2	2468,2	7430,44	212758,84	2519,68	-51,48
6	96,2	2733,2	9254,44	262933,84	2620,93	112,27
7	99,2	3043,2	9840,64	301885,44	2651,30	391,9
8	122,2	3328,2	14932,84	406706,04	2884,18	444,02
9	130,2	3493,2	16952,04	454814,64	2965,18	528,02
10	155,2	3583,2	24087,04	556112,64	3218,30	364,9
11	171,2	3793,2	29309,44	649395,84	3380,30	412,9
12	208,2	4033,2	43347,24	839712,24	3754,93	278,27
13	248,2	4158,2	61603,24	1032065,24	4159,93	-1,73
14	283,2	4283,2	80202,24	1213002,24	4514,30	-231,1
15	308,2	4548,2	94987,24	1401755,24	4767,43	-219,23
16	338,2	4738,2	114379,24	1602459,24	5071,18	-332,98
Сума	2510,2	51766,2	523922,44	9438840,64	51766,21	-0,02

Примечание. Значение суммы по столбцу Δ_i должно быть как можно ближе к нулю и свидетельствует о правильности выполненных расчетов. По умолчанию программа Microsoft Excel очень маленькие числа записывает в экспоненциальном формате.

Например, запись -8,37E-12 в обычном формате (-0,00000000000837) будет занимать очень много места, поэтому можно оставить это число в экспоненциальной форме. В тоже время запись 4,368E-05 уже не такая длинная, поэтому её формат можно изменить на обычный (0,00004368).

Замена формата числа производится через пункт меню *Формат* → *Ячейки* во вкладке "Число". Предварительно требуемая ячейка (диапазон ячеек) должна быть выделена мышкой.

8. Находим дисперсии отклонений и результативного признака. При этом пользуемся функцией ДИСП категории "Статистические".

Дисперсия результативного признака рассчитывается по третьей колонке таблицы 1.5 (выделяем диапазон значений мышкой, номер колонки и значение суммы в диапазон не входит!) и в нашем случае равна $\sigma_y^2 = 1026579,896$.

Дисперсия отклонений рассчитывается по седьмой колонке таблицы 1.5 (выделяем диапазон значений мышкой, номер колонки и значение суммы в диапазон не входит!) и в нашем случае равна $\sigma_{\Delta}^2 = 137304,718$.

9. Рассчитываем коэффициент детерминации за формулой:

$$D = r_{yx}^2 = 1 - \frac{\sigma_{\Delta}^2}{\sigma_y^2}.$$

В нашем случае:

$$D = r_{yx}^2 = 1 - \frac{137304,718}{1026579,896} = 0,866.$$

Полученное значение коэффициента детерминации является достаточно большим и характеризует правильность подбора линейной функции для построения эконометрической модели.

В 86,6% случаев изменения x приводят к изменению y . Остальные 13,4% изменения y объясняются факторами, неучтенными в модели.

10. Находим *множественный коэффициент корреляции* как квадратный корень с коэффициента детерминации. Для этого используем функцию КОРЕНЬ категории "Математические".

В нашем случае, поскольку модель однофакторная, то множественный коэффициент корреляции совпадает с обычным линейным коэффициентом корреляции, который равен $r_{xy} = \sqrt{D} = \sqrt{0,866} = 0,931$ или 93,1%.

На основании шкалы Чеддока можно сделать вывод, что связь между x (затраты на продвижение товара) и y (размер полученной прибыли) является очень высокой.

Лабораторная работа № 2. Построение многофакторной эконометрической модели

Задание

Оценить параметры эконометрической модели, характеризующей зависимость между недельными затратами на питание, общими расходами и размером семьи.

Исходные данные представлены в таблице 2.1 ниже.

Примечание: индивидуальный номер варианта формируется с помощью значения N – это две последние цифры номера Вашей зачетной книжки.

В случае, если предпоследняя цифра номера Вашей зачетной книжки равняется нулю, она игнорируется и учитывается только последняя цифра.

Таблица 2.1

№ п/п	Затраты на питание (дол. США)	Общие расходы (дол. США)	Размер семьи (чел.)
	y	x_1	x_2
1	$22+N/10$	$45+N/10$	$1,5+N/10$
2	$34+N/10$	$75+N/10$	$1,6+N/10$
3	$50+N/10$	$125+N/10$	$1,9+N/10$
4	$67+N/10$	$223+N/10$	$1,8+N/10$
5	$47+N/10$	$92+N/10$	$3,4+N/10$
6	$66+N/10$	$146+N/10$	$3,6+N/10$
7	$81+N/10$	$227+N/10$	$3,4+N/10$
8	$106+N/10$	$358+N/10$	$3,5+N/10$
9	$70+N/10$	$135+N/10$	$5,5+N/10$
10	$95+N/10$	$218+N/10$	$5,4+N/10$
11	$119+N/10$	$331+N/10$	$5,4+N/10$
12	$147+N/10$	$490+N/10$	$5,3+N/10$
13	$93+N/10$	$175+N/10$	$8,5+N/10$
14	$133+N/10$	$305+N/10$	$8,3+N/10$
15	$169+N/10$	$468+N/10$	$8,1+N/10$
16	$197+N/10$	$749+N/10$	$7,3+N/10$

Пример выполнения

Пусть две последних цифры зачетной книжки $N=00$.

Будем считать, что между факторами (общие расходы и размер семьи) и результирующей переменной (размер полученной прибыли) существует *линейная зависимость*. В таком случае мы имеем *линейную многофакторную модель* (модель множественной регрессии) общего вида (см. § 3.2):

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2 + \dots + \hat{a}_m x_m,$$

где $\hat{a}_0, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_m$ – неизвестные параметры модели (коэффициенты, которые нам нужно рассчитать).

Конкретно для нашего задания формула эконометрической модели будет выглядеть так:

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2$$

Рассчитывать (оценивать) параметры нашей модели будем при помощи МНК в матричной форме следующего вида:

$$Y = X \cdot A + \varepsilon,$$

где Y – вектор-столбец, состоящий из n (в нашем случае – 16) компонент зависимой (результативной) переменной;

X – матрица размера $n \times (m+1)$ (в нашем случае 16×3), состоящая из компонент независимых переменных;

$A = (\alpha_0, \alpha_1, \dots, \alpha_m)$ – вектор-столбец параметров, состоящий из $m+1$ -й компоненты (в нашем случае – из 3);

ε – вектор-столбец ошибки модели (остатков или отклонений), состоящий, как и вектор Y , из n компонент.

Напомним, что оператор оценивания параметров линейной эконометрической модели в матричной форме имеет вид:

$$A = (X^T X)^{-1} \cdot (X^T Y).$$

Процедуру оценивания параметров эконометрической модели целесообразно разбить на следующие шаги.

1) Записываем матрицу независимых переменных X :

$$X = \begin{pmatrix} 1 & 45 & 1,5 \\ 1 & 75 & 1,6 \\ 1 & 125 & 1,9 \\ 1 & 223 & 1,8 \\ 1 & 92 & 3,4 \\ 1 & 146 & 3,6 \\ 1 & 227 & 3,4 \\ 1 & 358 & 3,5 \\ 1 & 135 & 5,5 \\ 1 & 218 & 5,4 \\ 1 & 331 & 5,4 \\ 1 & 490 & 5,3 \\ 1 & 175 & 8,5 \\ 1 & 305 & 8,3 \\ 1 & 468 & 8,1 \\ 1 & 749 & 7,3 \end{pmatrix}$$

В Microsoft Excel матрица записывается как обычная таблица – каждое число в отдельной ячейке:

19					
20		1	45	1,5	
21		1	75	1,6	
22	X=	1	125	1,9	
23		1	223	1,8	
24		1	92	3,4	
25		1	146	3,6	
26		1	227	3,4	
27		1	358	3,5	
28		1	135	5,5	
29		1	218	5,4	
30		1	331	5,4	
31		1	490	5,3	
32		1	175	8,5	
33		1	305	8,3	
34		1	468	8,1	
35		1	749	7,3	
36					
37					

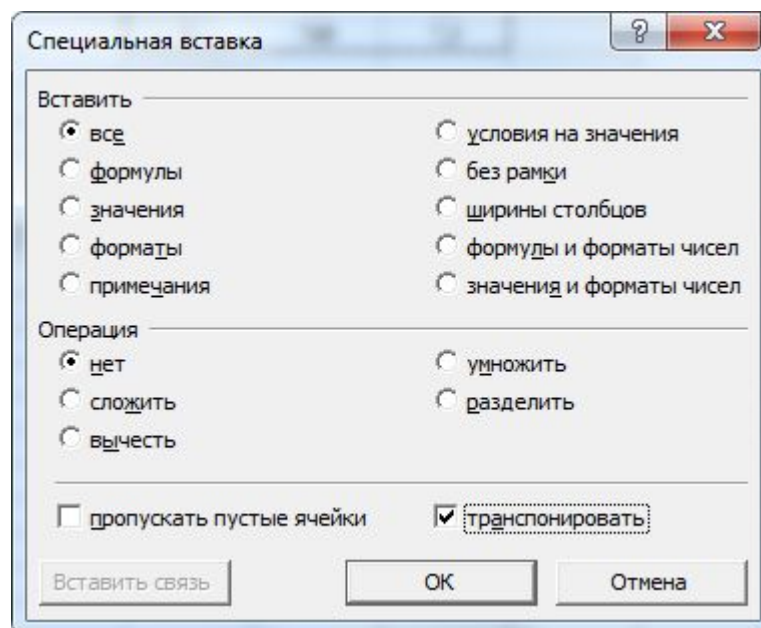
2) Записываем вектор-столбец зависимой (результативной) переменной Y :

$$Y = \begin{pmatrix} 22 \\ 34 \\ 50 \\ 67 \\ 47 \\ 66 \\ 81 \\ 106 \\ 70 \\ 95 \\ 119 \\ 147 \\ 93 \\ 133 \\ 169 \\ 197 \end{pmatrix}$$

	22	
	34	
Y=	50	
	67	
	47	
	66	
	81	
	106	
	70	
	95	
	119	
	147	
	93	
	133	
	169	
	197	

3) Находим транспонированную матрицу к матрице независимых переменных X , т.е. матрицу X^T .

При этом рекомендуется копирование с использованием пункта контекстного меню "Специальная вставка", проставив галочку напротив пункта "транспонировать" (предварительно нужно выделить цифры матрицы X , нажать пункт контекстного меню "копировать" и проставить курсор в незанятую ячейку, после чего вызывать меню "Специальная вставка").



4) Находим произведение матриц $X^T X$. При этом используем функцию Microsoft Excel МУМНОЖ категории "Математические".

Алгоритм использования функции МУМНОЖ следующий:

- Выделяем на листе блок свободных ячеек, в котором будет размещен результат произведения матриц (в нашем случае размеры этого блока 3x3);
- Вызываем функцию МУМНОЖ;
- В окне "Аргументы функции" ставим курсор в поле "Массив1" выделяем мышкой цифры матрицы X^T ;
- Ставим курсор в поле "Массив2" выделяем мышкой цифры матрицы X ;
- Одной рукою зажимаем на клавиатуре комбинацию клавиш Ctrl+Shift, другой рукой нажимаем мышкой на клавишу ОК (второй способ – нажатие на клавиатуре комбинации клавиш Ctrl+Shift+Enter, тогда клавишу ОК нажимать не требуется).

[illegible]

$$X^T X = \begin{pmatrix} 16 & 416,2 & 74,5 \\ 416,2 & 1601562 & 23271,2 \\ 74,5 & 23271,2 & 436,69 \end{pmatrix}$$

5) Находим произведение матриц $X^T Y$. Используем функцию МУМНОЖ. В этом случае диапазон свободных ячеек будет иметь размеры 3x1:

$$X^T Y = \begin{pmatrix} 1495 \\ 520090 \\ 8367,9 \end{pmatrix}$$

6) Находим матрицу, обратную к матрице $X^T X$. При этом пользуемся функцией МОБР категории "Математические". Диапазон свободных ячеек будет иметь такие же размеры, как и матрица $X^T X$. Не забываем об нажатии клавиш Ctrl+Shift+OK или Ctrl+Shift+Enter.

$$(X^T X)^{-1} = \begin{pmatrix} 0,314 & -0,00017 & -0,0446 \\ -0,00017 & 0,00003 & -0,00012 \\ -0,0446 & -0,00012 & 0,0165 \end{pmatrix}$$

7) Находим произведение матриц $(X^T X)^{-1} \cdot (X^T Y)$, т.е оператор оценивания $\hat{A}$. Используем функцию МУМНОЖ. Диапазон свободных ячеек будет иметь размеры 3x1. В поле "Массив1" указываем цифры матрицы $(X^T X)^{-1}$, а в поле "Массив2" – матрицу $(X^T Y)$.

$$\hat{A} = \begin{pmatrix} 9,111 \\ 0,200 \\ 6,931 \end{pmatrix}$$

Итак найденные в матричной форме оценки параметров следующие:
 $\hat{a}_0 = 9,1111$; $\hat{a}_1 = 0,2$; $\hat{a}_2 = 6,931$.

Таким образом, *эконометрическая модель* зависимости затрат на питание от общих расходов и размера семьи будет иметь вид:

$$\hat{y} = 9,111 + 0,2x_1 + 6,931x_2$$

8) Выполняем проверку правильности расчета параметров эконометрической модели. При этом используем функцию ЛИНЕЙН категории "Статистические". Алгоритм будет следующий:

- Выделяем мышкой на листе блок свободных ячеек размером 5x3 (число строк (5) – стандартное значение, т.е. одинаковое всегда, число столбцов (3) определяется количеством оцениваемых параметров);
- Вызываем функцию ЛИНЕЙН;

The screenshot shows the Excel interface with the 'Аргументы функции' (Function Arguments) dialog box for the 'ЛИНЕЙН' (LINEAR) function. The dialog box is open over a worksheet. The worksheet contains data for a regression analysis, including a matrix of independent variables (X) and a vector of dependent variables (Y). The dialog box fields are filled with the following values:

- Известные_значения_y**: G20:G35
- Известные_значения_x**: C20:D35
- Конст**: 1
- Статистика**: 1

The dialog box also displays the result of the function: $\{6,93088793886897;0,2\}$. Below the dialog box, the worksheet shows the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M
22	X=	1											
23		1											
24		1											
25		1											
26		1											
27		1											
28		1											
29		1											
30		1											
31		1											
32		1											
33		1											
34		1											
35		1											
36													
37													
38													
39		1	1	1	1	1	1	1	1	1	1	1	1
40		45	75	125	223	92	146	227	358	135	218	331	490
41	X ^T =	1,5	1,6	1,9	1,8	3,4	3,6	3,4	3,5	5,5	5,4	5,4	5,3
42													
43													
44													
45		16	4162	74,5									
46	X ^T X=	4162	1601562	23271,2									
47		74,5	23271,2	436,69									
48													
49													
50		1496											
51	X ^T Y=	520090											
52		8367,9											
53													
54	(X ^T X) ⁻¹ =	0,313647	-0,000166522	-0,044635									
55		-0,000167	2,85512E-06	-0,000124									
56		-0,044635	-0,00012374	0,016499									
57													
58													
59		9,111											
60	A	0,200											
61		6,931											
62													
63		D35;1;1)											
64													
65													
66													
67													
68													
69													

- В поле "Известные_значения_y" указываем цифровые значения вектора-столбца зависимой (результативной) переменной Y, а в поле "Известные_значения_x" – цифровые значения матрицы X (указываем только

значения, которые соответствуют переменным x_1 и x_2 , первый столбец (тот, что с единичками) не трогаем!);

- В поле "Конст" пишем 1 либо текст "Истина" (это позволит рассчитать свободный член $\hat{a}_0$);

- В поле "Статистика" тоже пишем 1 либо текст "Истина" (это позволит рассчитать дополнительные коэффициенты качественных характеристик нашей модели);

- Нажимаем клавиши Ctrl+Shift+OK или Ctrl+Shift+Enter.

Регрессионная статистика будет выводиться в порядке, указанном в следующей схеме:

Значение коэффициента $\hat{a}_2$	Значение коэффициента $\hat{a}_1$	Значение коэффициента $\hat{a}_0$
Среднеквадратическое отклонение $\hat{a}_2$	Среднеквадратическое отклонение $\hat{a}_1$	Среднеквадратическое отклонение $\hat{a}_0$
Коэффициент детерминации $D(R^2)$	Среднеквадратическое отклонение y	
F-статистика	Число степеней свободы	
Регрессионная сумма квадратов	Остаточная сумма квадратов	

В нашем случае получены такие значения:

62									
63		6,930888	0,200355379	9,11061					
64		0,846587	0,011136723	3,691184					
65		0,984536	6,590909186	#Н/Д					
66		413,8261	13	#Н/Д					
67		35953,28	564,7210906	#Н/Д					
68									

Таким образом, значения коэффициентов эконометрической модели $\hat{y} = 9,111 + 0,2x_1 + 6,931x_2$ рассчитаны правильно.

Выводы

1. С помощью МНК (в матричной форме) мы нашли уравнение эконометрической модели, которая показывает зависимость затрат на питание

от общих расходов и размера семьи и относится к классу линейных многофакторных эконометрических моделей.

2. Коэффициенты уравнения $\hat{y} = 9,111 + 0,2x_1 + 6,931x_2$ имеют следующее экономическое обоснование. Когда при всех других одинаковых условиях общие расходы увеличиваются (уменьшаются) на 1 дол. США, то результативная переменная (расходы на питание) также увеличивается (уменьшается) на 0,2 дол. США. Если размер семьи увеличивается (уменьшается) на 1 чел., то расходы на питание также увеличивается (уменьшается) на 6,93 дол. США.

Параметр $\hat{a}_0$ с формальной точки зрения показывает следующее: если значение двух независимых переменных (общие расходы и размер семьи) равно нулю, то расходы на питание = 9,11 дол. США). Очевидно, что в данной модели этот коэффициент экономического смысла не имеет и служит в роли начальной точки отсчета.

3. Полученное с помощью функции ЛИНЕЙН высокое значение коэффициента детерминации (0,9845 или 98,45%) характеризует правильность подбора линейной функции для построения эконометрической модели, т.е. модель в целом является адекватной.

4. Фактическое значение F -критерия Фишера (413,826) гораздо больше табличного значения, что подтверждает значимость уравнения регрессии в целом.

5. Полученные значения среднеквадратических отклонений (стандартных ошибок) коэффициентов эконометрической модели не являются слишком большими. За исключением коэффициента $\hat{a}_0$ среднеквадратические отклонения других параметров существенно меньше по сравнению с абсолютными значениями самых параметров, таким образом их относительная ошибка будет в пределах допустимых значений: для параметра $\hat{a}_1$ – в пределах 5,6% ($\frac{0,011}{0,200} \cdot 100\%$); для параметра $\hat{a}_2$ – в пределах 12,2% ($\frac{0,847}{6,931} \cdot 100\%$).

РЕКОМЕНДОВАННАЯ ЛИТЕРАТУРА

1. Бородич С.А. Вводный курс эконометрики: Учебное пособие. – Минск: БГУ, 2000. – 354 с.
2. Горбатков С.А. Эконометрика: авторский курс лекции. – Уфа: Филиал ВЗФЭИ в г. Уфа, 2008.
3. Замков О.О. Математические методы в экономике: Учебник / О.О. Замков, А.В. Толстопятенко, Ю.Н. Черемных / Под общ. ред. А.В. Сидоровича. – 4-е изд., стереотип. – М.: Издательство «Дело и Сервис», 2004. – 368 с.
4. Клебанова Т.С., Дубровина Н.А., Раевнева Е.В. Эконометрия: Учебное пособие. – 2-е изд., испр. – Харьков: ИД «ИНЖЭК», 2005. – 160 с.
5. Лугінін О.Є., Білоусова С.В., Білоусов О.М. Економетрія: Навч. посібник. – К.: Центр навчальної літератури, 2005. – 252 с.
6. Лук'яненко І.Г., Краснікова Л.І. Економетрика: Практикум з використанням комп'ютера. – К.: Товариство „Знання”, 1998. – 220 с.
7. Наконечний С.І., Терещенко Т.О., Романюк Т.П. Економетрія: Підручник. – Вид. 3-тє, доп. та перероб. – К.: КНЕУ, 2004. – 520 с.
8. Орлов А.И. Эконометрика: Учебное пособие для ВУЗов. – М.:ЭКЗАМЕН, 2007.
9. Практикум по эконометрике: Учебное пособие / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2006. – 344 с.
10. Шалабанов А.К., Роганов Д.А. Практикум по эконометрике с применением MS Excel. Линейные модели парной и множественной регрессии. – Казань: ТИСБИ, 2008. – 53 с.

