БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ГОСУДАРСТВЕННЫЙ ИНСТИТУТ УПРАВЛЕНИЯ
И СОЦИАЛЬНЫХ ТЕХНОЛОГИЙ

Кафедра управления финансами

СОГЛАСОВАНО Заведующий
кафедрой управления финансами
_____________М.Л. Зеленкевич
«31» 08 2017 г.

СОГЛАСОВАНО
Директор ГИУСТ БГУ
_____________П.И. Бригадин
«__8__» сентября 2017 г.
Регистрационный номер № 44

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС
ПО УЧЕБНОЙ ДИСЦИПЛИНЕ

ПРИКЛАДНАЯ ЭКОНОМЕТРИКА В БИЗНЕСЕ

для специальности 1-26 02 02 Менеджмент (по направлениям)
направление специальности 1-26 02 02-04 «Менеджмент недвижимости»

Составитель: Рачковский Н.Н., кандидат физ.-мат. наук, доцент

Одобрено и рекомендовано к утверждению учебно-методической комиссией ГИУСТ БГУ, (протокол № 1 от 7.09.2017 г.)

Рассмотрено и утверждено
на заседании совета 8.09.2017 г.,протокол № _17__

ОГЛАВЛЕНИЕ

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Целью УМК «Прикладная эконометрика в бизнесе» является оказание помощи студентам первой ступени высшего образования специальностей 1-26 02 02 Менеджмент (по направлениям), направление специальности 1-26 02 02-04 «Менеджмент недвижимости» в изучении соответствующей дисциплины по выбору.

УМК содержит теоретический раздел (конспект лекций), практический раздел (материалы к практическим занятиям (задачи с разбивкой по темам)), раздел контроля знаний (задачи для контрольных работ, перечень вопросов для зачета) и вспомогательный раздел (учебно-методическая карта дисциплины и список рекомендуемой литературы). УМК «Прикладная эконометрика в бизнесе» базируется современных достижениях экономической и математической наук и адаптирован к условиям существующей экономической ситуации Республики Беларусь.

Программа дисциплины «Прикладная эконометрика в бизнесе» входит в состав комплекса программ непрерывной подготовки студентов первой ступени высшего образования в области эффективного применения эконометрических методов и ЭВМ. Она определяет содержание базовой математической подготовки, обеспечивает связь обучения методам решения эконометрических задач с общеэкономической подготовкой специалистов.

Цель дисциплины – обучение студентов общим вопросам теории математического моделирования процессов экономики и управления, методам построения эконометрических моделей этих процессов.

Задачи дисциплины:

- ознакомить студентов с основными понятиями эконометрического моделирования, теоретическими положениями и экспериментальными данными, используемыми при построении эконометрических моделей в будущей профессиональной деятельности студентов;
- обучить методам построения эконометрических моделей и их качественным исследованиям, численным методам реализации моделей, методам постановки и проведения вычислительного эксперимента с эконометрическими моделями, анализа результатов эксперимента;
- обучить применению эконометрических методов и моделей для решения задач экономического содержания по будущей специальности.

В результате прохождения дисциплины будущий специалист должен уметь:

- строить эконометрические модели простых объектов;

- выполнять качественный анализ моделей;
- проводить вычислительные эксперименты и анализировать их результаты;
- ставить экономические задачи и находить оптимальные условия функционирования эконометрических моделей.

На основании полученных знаний студент должен приобрести практические навыки формирования эконометрических моделей для исследования процессов экономики и управления, реализовывать эти модели на ЭВМ с применением современных математических методов, алгоритмов и программ, анализировать и обобщать полученные результаты.

ТЕОРЕТИЧЕСКИЙ РАЗДЕЛ

1. Основные понятия эконометрики

Основная задача эконометрики формулируется следующим образом. Предположим, что имеются наблюдаемые значения $(x_1^1, x_2^1, \ldots, x_m^1, y_1)$, $(x_1^2, x_2^2, \ldots, x_m^2, y_2)$, …, $(x_1^n, x_2^n, \ldots, x_m^n, y_n)$ некоторых количественных признаков X_1, X_2, …, X_m, Y; требуется по результатам наблюдений спрогнозировать значение y_0 признака Y, соответствующее ранее не встречавшимся значениям $(x_1^0, x_2^0, \ldots, x_m^0)$ признаков X_1, X_2, …, X_m. Для этого нужно найти функцию $y = f(x_1, x_2, \ldots, x_m)$, которая с достаточной точностью отражала бы зависимость значений признака Y от значений признаков X_1, X_2, …, X_m, и тогда искомое прогнозное значение признака Y можно найти, подставив в эту функцию указанные значения $(x_1^0, x_2^0, \ldots, x_m^0)$: $y_0 = f(x_1^0, x_2^0, \ldots, x_m^0)$. Такая функция называется **регрессией признака $\boldsymbol{Y}$ на признаки $\boldsymbol{X_1}$, $\boldsymbol{X_2}$, …, $\boldsymbol{X_m}$**, а соответствующее уравнение $y = f(x_1, x_2, \ldots, x_m)$, составленное по результатам наблюдений, называется **эмпирическим уравнением регрессии**.

Процесс создания эмпирического уравнения регрессии разбивается на следующие этапы.

1. **Спецификация**. На этом этапе определяют вид регрессии $y = f(x_1, x_2, \ldots, x_m)$.
2. **Параметризация**. Здесь по результатам наблюдений вычисляют значения (или, точнее, оценки) параметров регрессии, выбранной на этапе спецификации; при этом чаще всего используется метод наименьших квадратов.
3. **Верификация**. На этом этапе происходит проверка адекватности созданного эмпирического уравнения регрессии $y = f(x_1, x_2, \ldots, x_m)$.

Если рассматриваются только два признака X и Yто говорят о **парной регрессии**; в противном случае имеем **множественную регрессию**. Для парной регрессии этап спецификации производится следующим образом. На координатной плоскости строится **корреляционное поле**, т.е. точки (x_1, y_1), (x_2, y_2), …, (x_n, y_n), координатами которых являются наблюдаемые пары значений признаков X и Y. Тогда:

а) Если корреляционное поле расположено вдоль некоторой (невертикальной) прямой, то регрессию считают линейной, т.е. полагают $y = a_0 + a_1 \cdot x$.

б) Если корреляционное поле расположено вдоль некоторой возрастающей выпуклой вниз кривой, то полагают $y = a_0 \cdot x^{a_1}$, или $y = a_0 \cdot e^{a_1 x}$, или $y = a_0 + a_1 \cdot x^2$.

в) Если корреляционное поле расположено вдоль некоторой возрастающей выпуклой вверх кривой, то полагают $y = a_0 \cdot x^{a_1}$ или $y = a_0 + a_1 \cdot \ln x$.

г) Если корреляционное поле расположено вдоль некоторой убывающей выпуклой вниз кривой, то полагают $y = a_0 + \frac{a_1}{x}$ или $y = a_0 \cdot e^{a_1 x}$.

д) Если корреляционное поле расположено вдоль некоторой убывающей выпуклой вверх кривой, то полагают $y = a_0 \cdot x^{a_1}$.

Здесь a_0 и a_1 – это параметры, значения которых нужно оценить на этапе параметризации.

Если же рассматривается несколько признаков X_1, X_2, …, X_m, то на этапе спецификации, как правило, выбирают множественную линейную регрессию $y = a_0 + a_1x_1 + a_2x_2 + \cdots + a_mx_m$.

Основным методом вычисления оценок параметров a_i выбранной регрессии является **метод наименьших квадратов** (МНК), суть которого раскроем ниже. Верификация найденного эмпирического уравнения регрессии основывается на анализе его качества (более подробно об этом также будет рассказано ниже).

2. Общая идея метода наименьших квадратов

Предположим сначала, что в общем случае для выбранной на этапе спецификации функции $y = f(x_1, x_2, \dots, x_m, a_0, a_1, \dots, a_k)$, зависящей от объясняющих переменных $x_1, x_2, \dots, x_m$ и определяемой числовыми параметрами $a_0, a_1, \dots, a_k$, требуется по наблюдаемым значениям x_i^j и y_j, $i = 1,2, \dots m$, $j = 1,2, \dots n$, найти наиболее подходящие значения параметров $a_0, a_1, \dots, a_k$. Другими словами, нужно найти такие значения $a_0^*, a_1^*, \dots, a_k^*$ параметров $a_0, a_1, \dots, a_k$, при которых отклонения

$$\varepsilon_j = f\left(x_1^j, x_2^j, \dots, x_m^j, a_0^*, a_1^*, \dots, a_k^*\right) - y_j$$

значений $f\left(x_1^j, x_2^j, \dots, x_m^j, a_0^*, a_1^*, \dots, a_k^*\right)$ рассматриваемой функции в наблюдаемых значениях $x_1^j, x_2^j, \dots, x_m^j$ объясняющих переменных от наблюдаемых значений y_j объясняемой переменной были минимальными в совокупности. Казалось бы, для достижения этой цели достаточно решить задачу нахождения точек минимума суммы всех отклонений ε_j; однако, заметим, что отклонения ε_j могут быть разных знаков (как положительные, так и отрицательные), и поэтому их сумма может быть мала при достаточно больших по абсолютной величине, но разных по знаку слагаемых. Поэтому на самом деле нужно минимизировать сумму модулей отклонений ε_j или их квадратов. Заметим далее, что решать указанную задачу минимизации отклонений ε_j удобно с помощью частных производных; следовательно, предпочтительнее рассматривать сумму квадратов величин ε_j (как известно, функция «модуль» в отличие от функции «возведение в квадрат» не имеет производной в своей точке минимума). Таким образом, искомые значения $a_0^*, a_1^*, \dots, a_k^*$ являются точкой минимума функции

$$Q(a_0, a_1, \dots, a_k) = \varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2.$$

Следовательно, эти значения найдем, приравняв к нулю все частные производные функции $Q = Q(a_0, a_1, \dots, a_k)$ и решив систему соответствующих алгебраических уравнений:

$$\begin{cases} \dfrac{\partial Q}{\partial a_0} = 0, \\ \dfrac{\partial Q}{\partial a_1} = 0, \\ \dots \\ \dfrac{\partial Q}{\partial a_k} = 0. \end{cases}$$

В частности, если на этапе спецификации была выбрана линейная функция нескольких переменных $y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m$, то в этом случае $\varepsilon_j = a_0 + a_1x_1^j + a_2x_2^j + \dots + a_mx_m^j - y_j$, $j = 1,2, \dots n$. Тогда

$$Q(a_0, a_1, \dots, a_m) = \varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2 =$$
$$= \sum_{j=1}^{n} (a_0 + a_1x_1^j + a_2x_2^j + \dots + a_mx_m^j - y_j)^2.$$

Следовательно, $\frac{\partial Q}{\partial a_0} = 2\sum_{j=1}^{n}(a_0 + a_1x_1^j + a_2x_2^j + \dots + a_mx_m^j - y_j)$, $\frac{\partial Q}{\partial a_i} = 2\sum_{j=1}^{n} x_i^j(a_0 + a_1x_1^j + a_2x_2^j + \dots + a_mx_m^j - y_j)$, $i = 1,2, \dots m$. Приравняв эти частные производные к нулю, получим следующую систему линейных алгебраических уравнений относительно неизвестных $a_0, a_1, \dots, a_m$:

$$\begin{cases} na_0 + a_1 \sum_{j=1}^{n} x_1^j + a_2 \sum_{j=1}^{n} x_2^j + \cdots + a_m \sum_{j=1}^{n} x_m^j = \sum_{j=1}^{n} y_j\,, \\ a_0 \sum_{j=1}^{n} x_1^j + a_1 \sum_{j=1}^{n} (x_1^j)^2 + a_2 \sum_{j=1}^{n} x_1^j x_2^j + \cdots + a_m \sum_{j=1}^{n} x_1^j x_m^j = \sum_{j=1}^{n} x_1^j y_j\,, \\ \cdots \\ a_0 \sum_{j=1}^{n} x_m^j + a_1 \sum_{j=1}^{n} x_m^j x_1^j + a_2 \sum_{j=1}^{n} x_m^j x_2^j + \cdots + a_m \sum_{j=1}^{n} (x_m^j)^2 = \sum_{j=1}^{n} x_m^j y_j\,. \end{cases}$$

Эту систему удобно решать методом обратной матрицы (подробнее об этом см. ниже).

Если же объясняющая переменная является единственной (т.е. x), и на этапе спецификации было выбрана линейная функция $y = a_0 + a_1 x$, то аналогичным образом получаем систему линейных алгебраических уравнений для вычисления искомых значений a_0^*, a_1^*:

$$\begin{cases} na_0 + a_1 \sum_{j=1}^{n} x_j = \sum_{j=1}^{n} y_j\,, \\ a_0 \sum_{j=1}^{n} x_j + a_1 \sum_{j=1}^{n} (x_j)^2 = \sum_{j=1}^{n} x_j y_j\,. \end{cases}$$

Эту систему можно упростить, разделив обе части каждого уравнения на n и введя в рассмотрение средние величины:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n};$$

$$\bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n};$$

$$\overline{xy} = \frac{x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n}{n};$$

$$\overline{x^2} = \frac{{x_1}^2 + {x_2}^2 + \cdots + {x_n}^2}{n}.$$

В результате получим следующую систему:

$$\begin{cases} a_0 + a_1 \bar{x} = \bar{y}, \\ a_0 \bar{x} + a_1 \overline{x^2} = \overline{xy}. \end{cases}$$

Ниже мы подробно покажем, как решать последнюю систему.

Подобным образом получаются системы уравнений и тогда, когда на этапе спецификации выбрана нелинейная функция. Но в этом случае нужно иметь в виду, что уравнения в системе, скорее всего, будут нелинейными относительно параметров, а значит, усложнится процесс решения этой системы, а также усложнится анализ качества полученного эмпирического уравнения регрессии. В такой ситуации, как правило, с помощью специальных приемов нелинейную модель приводят к линейной и только после этого применяют метод наименьших квадратов (подробнее об этом см. ниже).

3. Парная линейная регрессия

Предположим, что рассматриваются только два количественных признака X и Y, и требуется по имеющимся наблюдаемым значениям (x_1, y_1), (x_2, y_2), …, (x_n, y_n) этих признаков найти эмпирическое уравнение регрессии; на этапе спецификации была выбрана линейная модель $y = a_0 + a_1 \cdot x$. Перейдем теперь к этапу параметризации.

Метод наименьших квадратов для оценки параметров a_0 и a_1 парной линейной регрессии $y = a_0 + a_1 \cdot x$ реализуется **посредством следующего алгоритма**.

1. По результатам наблюдений вычислим средние:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n};$$

$$\bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n};$$

$$\overline{xy} = \frac{x_1 \cdot y_1 + x_2 \cdot y_2 + \cdots + x_n \cdot y_n}{n}.$$

2. Найдем выборочные средние квадратические отклонения:

$$\sigma_x = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n} - \overline{x}^2};$$

$$\sigma_y = \sqrt{\frac{y_1^2 + y_2^2 + \cdots + y_n^2}{n} - \overline{y}^2}.$$

3. Вычислим выборочный коэффициент корреляции:

$$r_{\text{в}} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}.$$

4. Проанализируем значение выборочного коэффициента корреляции:
   а) если $|r_{\text{в}}| \geq 0{,}7$, то рассматриваемая линейная модель достаточно адекватна;
   б) если $0{,}3 \leq |r_{\text{в}}| < 0{,}7$, то желательно найти более адекватную (нелинейную) модель;
   в) если $|r_{\text{в}}| < 0{,}3$, то рассматриваемая линейная модель неадекватна; следовательно, нужно найти адекватную нелинейную модель.
5. В случае адекватности линейной модели вычислим оценки параметров a_0 и a_1 парной линейной регрессии:

$$a_1 = r_{\text{в}} \cdot \frac{\sigma_y}{\sigma_x};$$

$$a_0 = \bar{y} - a_1 \cdot \bar{x}.$$

После вычисления оценок параметров a_0 и a_1 приступим к этапу **верификации**. Частично проверка адекватности выбранной модели была проведена на этапе параметризации (см. пункт 4 приведенного выше алгоритма). Закончим эту проверку, используя **коэффициент детерминации**

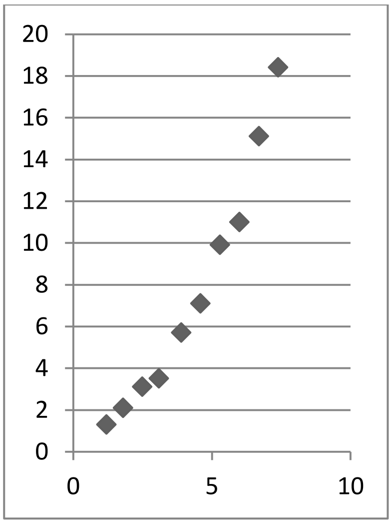
$$R^2 = 1 - \frac{(y(x_1) - y_1)^2 + (y(x_2) - y_2)^2 + \dots + (y(x_n) - y_n)^2}{(y_1 - \overline{y})^2 + (y_2 - \overline{y})^2 + \dots + (y_n - \overline{y})^2},$$

который показывает, насколько расчетные значения $y(x_1)$, $y(x_2)$, …, $y(x_n)$ признака Y отличаются от его наблюдаемых значений y_1, y_2, …, y_n: чем ближе значение коэффициента детерминации R^2 к 1, тем это отличие меньше. Здесь $y(x_i) = f(x_i, a_0, a_1) = a_0 + a_1 x_i$ – значения выбранной на этапе спецификации функции $y = f(x, a_0, a_1) = a_0 + a_1 \cdot x$ в наблюдаемых значениях $x_1, x_2, \dots, x_n$ признака X с учетом вычисленных на этапе параметризации оценок параметров a_0 и a_1. Считается, что построенное эмпирическое уравнение регрессии адекватно, если $R^2 > 0{,}7$.

Пример 1. По результатам наблюдений значений признаков X и Y (см. таблицу) найти эмпирическое уравнение парной линейной регрессии.

x_i	1,2	1,8	2,5	3,1	3,9	4,9	5,3	6,0	6,7	7,4
y_i	1,3	2,1	3,1	3,5	5,7	7,1	9,9	11,0	15,1	18,4

Решение. 1. Построим корреляционное поле.



2. Нетрудно видеть, что построенное корреляционное поле можно считать расположенным вдоль некоторой прямой; следовательно, можно рассмотреть парную линейную регрессию $y = a_0 + a_1 \cdot x$.

3. Вычислим выборочные средние:

$$\bar{x} = \frac{1{,}2 + 1{,}8 + 2{,}5 + 3{,}1 + 3{,}9 + 4{,}9 + 5{,}3 + 6{,}0 + 6{,}7 + 7{,}4}{10} = 4{,}25;$$

$$\bar{y} = \frac{1{,}3 + 2{,}1 + 3{,}1 + 3{,}5 + 5{,}7 + 7{,}1 + 9{,}9 + 11{,}0 + 15{,}1 + 18{,}4}{10} = 7{,}72;$$

$$\overline{xy} = \frac{1{,}2 \cdot 1{,}3 + 1{,}8 \cdot 2{,}1 + 2{,}5 \cdot 3{,}1 + 3{,}1 \cdot 3{,}5 + 3{,}9 \cdot 5{,}7}{10} +$$

$$+ \frac{4{,}9 \cdot 7{,}1 + 5{,}3 \cdot 9{,}9 + 6{,}0 \cdot 11{,}0 + 6{,}7 \cdot 15{,}1 + 7{,}4 \cdot 18{,}4}{10} = 43{,}463.$$

4. Найдем выборочные средние квадратические отклонения:

$$\sigma_x = \left(\frac{1{,}2^2 + 1{,}8^2 + 2{,}5^2 + 3{,}1^2 + 3{,}9^2 + 4{,}9^2 + 5{,}3^2 + 6{,}0^2 + 6{,}7^2 + 7{,}4^2}{10}\right.$$

$$\left. - 4{,}25^2\right)^{\frac{1}{2}} = 2{,}001;$$

$$\sigma_y =$$

$$= \left(\frac{1{,}3^2 + 2{,}1^2 + 3{,}1^2 + 3{,}5^2 + 5{,}7^2 + 7{,}1^2 + 9{,}9^2 + 11{,}0^2 + 15{,}1^2 + 18{,}4^2}{10}\right.$$

$$\left. - 7{,}72^2\right)^{\frac{1}{2}} = 5{,}481.$$

5. Вычислим выборочный коэффициент корреляции:

$$r_{\text{в}} = \frac{43{,}463 - 4{,}25 \cdot 7{,}72}{2{,}001 \cdot 5{,}481} = 0{,}971.$$

6. Проанализируем значение выборочного коэффициента корреляции. Поскольку этот коэффициент больше числа 0,7, то рассматриваемая линейная модель вполне адекватна, и мы можем вычислить оценки ее параметров:

$$a_1 = 0{,}971 \cdot \frac{5{,}481}{2{,}001} = 2{,}662;$$

$$a_0 = 7{,}72 - 2{,}662 \cdot 4{,}25 = -3{,}592.$$

7. Запишем эмпирическое уравнение регрессии:

$$y = 2{,}662x - 3{,}592.$$

8. Проверим адекватность найденного эмпирического уравнения парной линейной регрессии. Для этого вычислим коэффициент детерминации. С этой целью предварительно найдем значения $y(x_i) - y_i$, а также $y_i - \overline{y}$, где $y(x_i) = 2{,}662x_i - 3{,}592$:

$y_i - \overline{y}$	−6,42	−5,62	−4,62	−4,22	−2,02	−0,62	2,18	3,28	7,38	10,68
$y(x_i) - y_i$	−1,70	−0,90	−0,04	1,16	1,09	1,55	0,61	1,38	−0,86	−2,30

Тогда:

$$R^2 = 1 -$$

$$-\frac{(-1{,}70)^2 + (-0{,}90)^2 + (-0{,}04)^2 + 1{,}16^2 + 1{,}09^2 + 1{,}55^2 + 0{,}61^2 + 1{,}38^2 + (-0{,}86)^2 + (-2{,}30)^2}{(-6{,}42)^2 + (-5{,}62)^2 + (-4{,}62)^2 + (-4{,}22)^2 + (-2{,}02)^2 + (-0{,}62)^2 + 2{,}18^2 + 3{,}28^2 + 7{,}38^2 + 10{,}68^2} =$$

$$= 0{,}94.$$

Поскольку значение коэффициента детерминации $R^2 = 0{,}94$ очень близко к 1, заключаем, что найденное эмпирическое уравнение парной линейной регрессии $y = 2{,}662x - 3{,}592$ достаточно адекватно.

4. Множественная линейная регрессия

Предположим теперь, что рассматриваются несколько количественных признаков X_1, X_2, …, X_m, Y, и требуется по имеющимся наблюдаемым значениям $(x_1^1, x_2^1, \dots, x_m^1, y_1)$, $(x_1^2, x_2^2, \dots, x_m^2, y_2)$, …, $(x_1^n, x_2^n, \dots, x_m^n, y_n)$ этих признаков найти эмпирическое уравнение регрессии. Как указывалось выше, в этом случае на этапе спецификации обычно выбирается множественная линейная модель $y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m$. Перейдем теперь к этапу параметризации.

Но прежде чем перейти к описанию алгоритма нахождения оценок коэффициентов множественной линейной регрессии, отметим желательность выполнимости ряда предпосылок МНК:

1) Для каждого наблюдения (с номером i) математическое ожидание случайного отклонения $\varepsilon_i = a_0 + a_1x_1^i + a_2x_2^i + \dots + a_mx_m^i - y_i$ должно быть равно нулю.
2) Гомоскедастичность – дисперсии всех случайных отклонений ε_i должны быть равны друг другу.
3) Отсутствие автокорреляции – все случайные отклонения ε_i должны быть попарно независимыми случайными величинами.
4) Каждое случайное отклонение ε_i должно быть независимо от переменных x_1, x_2,…, x_m.
5) Отсутствие мультиколлинеарности – между переменными x_1, x_2,…, x_m – должна отсутствовать строгая (сильная) линейная зависимость.
6) Все случайные отклонения ε_i должны иметь нормальное распределение.
7) Выбранная модель должна быть линейной относительно своих коэффициентов.

При выполнении этих предпосылок применение МНК обеспечивает получение несмещенных, эффективных и состоятельных оценок параметров $a_0, a_1, a_2, \ldots, a_m$.

Расчет оценок параметров a_0, a_1, $a_2, \ldots$, a_m удобно производить в матричной форме. Для этого введем в рассмотрение матрицы

$$A = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{pmatrix}, \; Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \; X = \begin{pmatrix} 1 & x_1^1 & x_2^1 & \ldots & x_m^1 \\ 1 & x_1^2 & x_2^2 & \ldots & x_m^2 \\ & & \ldots & & \\ 1 & x_1^n & x_2^n & \ldots & x_m^n \end{pmatrix}.$$

Тогда столбец оценок коэффициентов множественной линейной регрессии можно вычислить следующим образом:

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

где X^T обозначает матрицу, транспонированную к матрице X, а $(X^T \cdot X)^{-1}$ обозначает матрицу, обратную к матрице $X^T \cdot X$.

На этапе верификации вычисляется коэффициент детерминации по указанной выше формуле

$$R^2 = 1 - \frac{(y(x_1) - y_1)^2 + (y(x_2) - y_2)^2 + \cdots + (y(x_n) - y_n)^2}{(y_1 - \overline{y})^2 + (y_2 - \overline{y})^2 + \cdots + (y_n - \overline{y})^2},$$

только теперь значения $y(x_i), i = 1,2, \ldots, n$, вычисляются следующим образом: $y(x_i) = a_0 + a_1 x_1^i + a_2 x_2^i + \cdots + a_m x_m^i$, т.е. все значения $y(x_i)$ являются элементами матрицы-столбца $X \cdot A$. Как и выше, считается, что построенное эмпирическое уравнение множественной линейной регрессии адекватно, если $R^2 > 0{,}7$.

Пример 2. По результатам наблюдений значений признаков X_1, X_2 и Y (см. таблицу) найти эмпирическое уравнение множественной линейной регрессии.

x_1^i	2,1	2,5	2,8	3,2	3,6	3,9	4,1	4,3	4,8
x_2^i	7,4	7,1	6,9	6,5	6,2	5,8	5,6	5,1	4,9
y_i	11,5	11,5	11,5	11,3	11,1	10,7	10,5	9,9	9,8

Решение. 1. Запишем матрицы X и Y:

$$X = \begin{pmatrix} 1 & 2,1 & 7,4 \\ 1 & 2,5 & 7,1 \\ 1 & 2,8 & 6,9 \\ 1 & 3,2 & 6,5 \\ 1 & 3,6 & 6,2 \\ 1 & 3,9 & 5,8 \\ 1 & 4,1 & 5,6 \\ 1 & 4,3 & 5,1 \\ 1 & 4,8 & 4,9 \end{pmatrix}, \ Y = \begin{pmatrix} 11,5 \\ 11,5 \\ 11,5 \\ 11,3 \\ 11,1 \\ 10,7 \\ 10,5 \\ 9,9 \\ 9,8 \end{pmatrix}.$$

2. Вычислим матрицы X^T, $X^T \cdot X$:

$$X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,1 & 2,5 & 2,8 & 3,2 & 3,6 & 3,9 & 4,1 & 4,3 & 4,8 \\ 7,4 & 7,1 & 6,9 & 6,5 & 6,2 & 5,8 & 5,6 & 5,1 & 4,9 \end{pmatrix},$$

$$X^T \cdot X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,1 & 2,5 & 2,8 & 3,2 & 3,6 & 3,9 & 4,1 & 4,3 & 4,8 \\ 7,4 & 7,1 & 6,9 & 6,5 & 6,2 & 5,8 & 5,6 & 5,1 & 4,9 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2,1 & 7,4 \\ 1 & 2,5 & 7,1 \\ 1 & 2,8 & 6,9 \\ 1 & 3,2 & 6,5 \\ 1 & 3,6 & 6,2 \\ 1 & 3,9 & 5,8 \\ 1 & 4,1 & 5,6 \\ 1 & 4,3 & 5,1 \\ 1 & 4,8 & 4,9 \end{pmatrix}$$

$$= \begin{pmatrix} 9 & 31,3 & 35,5 \\ 31,3 & 115,25 & 186,76 \\ 35,5 & 186,76 & 348,49 \end{pmatrix},$$

3. Вычислим матрицу $(X^T \cdot X)^{-1}$:

$$(X^T \cdot X)^{-1} = \begin{pmatrix} 770,119 & -79,073 & -80,272 \\ -79,073 & 8,185 & 8,207 \\ -80,272 & 8,207 & 8,389 \end{pmatrix}.$$

4. Вычислим матрицу $(X^T \cdot X)^{-1} \cdot X^T$:

$$(X^T \cdot X)^{-1} \cdot X^T = \begin{pmatrix} 770,119 & -79,073 & -80,272 \\ -79,073 & 8,185 & 8,207 \\ -80,272 & 8,207 & 8,389 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,1 & 2,5 & 2,8 & 3,2 & 3,6 & 3,9 & 4,1 & 4,3 & 4,8 \\ 7,4 & 7,1 & 6,9 & 6,5 & 6,2 & 5,8 & 5,6 & 5,1 & 4,9 \end{pmatrix}$$

$$= \begin{pmatrix} 10,053 & 2,506 & -5,161 & -4,682 & -12,229 & -3,842 & -3,602 & 20,719 & -2,763 \\ -1,155 & -0,343 & 0,471 & 0,462 & 1,274 & 0,447 & 0,442 & -2,024 & 0,427 \\ -0,961 & -0,195 & 0,59 & 0,517 & 1,283 & 0,389 & 0,353 & -2,2 & 0,225 \end{pmatrix}.$$

5. Наконец, найдем оценки коэффициентов множественной линейной регрессии:

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

$$= \begin{pmatrix} 10{,}053 & 2{,}506 & -5{,}161 & -4{,}682 & -12{,}229 & -3{,}842 & -3{,}602 & 20{,}719 & -2{,}763 \\ -1{,}155 & -0{,}343 & 0{,}471 & 0{,}462 & 1{,}274 & 0{,}447 & 0{,}442 & -2{,}024 & 0{,}427 \\ -0{,}961 & -0{,}195 & 0{,}59 & 0{,}517 & 1{,}283 & 0{,}389 & 0{,}353 & -2{,}2 & 0{,}225 \end{pmatrix} \cdot \begin{pmatrix} 11{,}5 \\ 11{,}5 \\ 11{,}5 \\ 11{,}3 \\ 11{,}1 \\ 10{,}7 \\ 10{,}5 \\ 9{,}9 \\ 9{,}8 \end{pmatrix}$$

$$= \begin{pmatrix} -5{,}556 \\ 1{,}213 \\ 1{,}979 \end{pmatrix}.$$

6. Запишем эмпирическое уравнение множественной линейной регрессии: $y = -5{,}556 + 1{,}213x_1 + 1{,}979x_2$.

7. Для проверки адекватности построенного уравнения вычислим коэффициент детерминации; с этой целью предварительно найдем значения $\overline{y}$, $y(x_i) - y_i$, а также $y_i - \overline{y}$, где $y(x_i) = -5{,}556 + 1{,}213x_1^i + 1{,}979x_2^i$:

$$\overline{y} = \frac{11{,}5 + 11{,}5 + 11{,}5 + 11{,}3 + 11{,}1 + 10{,}7 + 10{,}5 + 9{,}9 + 9{,}8}{9} = 10{,}9,$$

$y(x_i) - y_i$	0,133	0,004	-0,026	-0,092	-0,04	-0,06	-0,021	-0,103	0,206
$y_i - \overline{y}$	0,6	0,7	0,7	0,4	0,3	-0,2	-0,3	-1	-1,1

Тогда

$$R^2 = 1 - \frac{0{,}133^2 + 0{,}004^2 + (-0{,}026)^2 + (-0{,}092)^2 + (-0{,}04)^2 + (-0{,}06)^2 + (-0{,}021)^2 + (-0{,}103)^2 + 0{,}206^2}{0{,}6^2 + 0{,}7^2 + 0{,}7^2 + 0{,}4^2 + 0{,}3^2 + (-0{,}2)^2 + (-0{,}3)^2 + (-1)^2 + (-1{,}1)^2} = 0{,}978.$$

Поскольку значение коэффициента детерминации R^2 очень близко к 1, то найденное эмпирическое уравнение множественной линейной регрессии $y = -5{,}556 + 1{,}213x_1 + 1{,}979x_2$ адекватно.

5. Парная нелинейная регрессия

Если выяснится, что линейная модель в рассматриваемой ситуации неадекватна (например, на этапе спецификации замечено, что корреляционное поле слабо напоминает прямую линию, или при выборе линейной модели оказалось, что значение коэффициента корреляции (или детерминации) достаточно далеко от 1), то следует рассмотреть нелинейную модель. Наиболее часто применяют следующие парные нелинейные модели:

- **степенную модель** $y = a_0 \cdot x^{a_1}$;
- **показательную модель** $y = a_0 \cdot e^{a_1 x}$;

- **квадратичную модель** $y = a_0 + a_1 \cdot x^2$;
- **логарифмическую модель** $y = a_0 + a_1 \cdot \ln x$;
- **обратную модель** $y = a_0 + \frac{a_1}{x}$.

Если на этапе спецификации выбрана какая-либо нелинейная модель, то нужно прежде всего эту нелинейную модель преобразовать к линейной и затем применить к полученной линейной модели метод наименьших квадратов для нахождения оценок параметров.

В частности, **степенную модель**

$$y = a_0 \cdot x^{a_1}$$

преобразуют к линейной, прологарифмировав ее: $\ln y = \ln a_0 + a_1 \cdot \ln x$ и после этого произведя замену: $x^* = \ln x$; $y^* = \ln y$; $a_0^* = \ln a_0$. В результате получим линейную модель $y^* = a_0^* + a_1 \cdot x^*$. Для этой линейной модели с помощью метода наименьших квадратов вычислим оценки параметров a_0^* и a_1, а затем найдем оценку исходного параметра $a_0 = e^{a_0^*}$.

Показательная модель

$$y = a_0 \cdot e^{a_1 x}$$

преобразуется к линейной также логарифмированием: $\ln y = \ln a_0 + a_1 \cdot x$; тогда замена производится следующая: $y^* = \ln y$; $a_0^* = \ln a_0$. В результате получим линейную модель $y^* = a_0^* + a_1 \cdot x$. Для этой линейной модели с помощью метода наименьших квадратов вычислим оценки параметров a_0^* и a_1, а затем найдем оценку исходного параметра $a_0 = e^{a_0^*}$.

Квадратичная модель

$$y = a_0 + a_1 \cdot x^2$$

преобразуется к линейной модели $y = a_0 + a_1 \cdot x^*$ заменой $x^* = x^2$; и уже для этой линейной модели с помощью метода наименьших квадратов вычислим искомые оценки параметров a_0 и a_1.

Логарифмическая модель

$$y = a_0 + a_1 \cdot \ln x$$

преобразуется к линейной модели $y = a_0 + a_1 \cdot x^*$ заменой $x^* = \ln x$; и уже для этой линейной модели с помощью метода наименьших квадратов вычислим искомые оценки параметров a_0 и a_1.

Обратная модель

$$y = a_0 + \frac{a_1}{x}$$

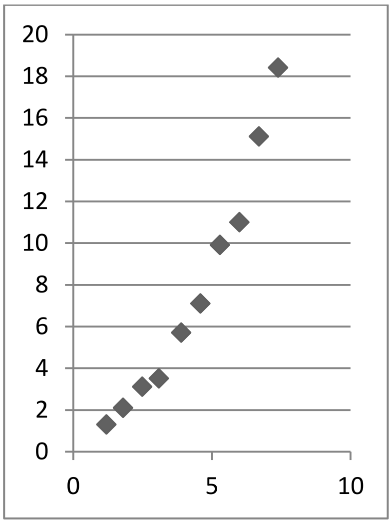
преобразуется к линейной модели $y = a_0 + a_1 \cdot x^*$ заменой $x^* = \frac{1}{x}$; и уже для этой линейной модели с помощью метода наименьших квадратов вычислим искомые оценки параметров a_0 и a_1.

Пример 3. В условиях примера 1 построим эмпирические уравнения парной нелинейной регрессии различных типов и выберем среди них самое адекватное.

Решение. Напомним, что в примере 1 наблюдаемые значения факторов X и Y были приведены в таблице:

x_i	1,2	1,8	2,5	3,1	3,9	4,9	5,3	6,0	6,7	7,4
y_i	1,3	2,1	3,1	3,5	5,7	7,1	9,9	11,0	15,1	18,4

1. Корреляционное поле имеет вид:



2. Нетрудно видеть, что построенное корреляционное поле можно считать расположенным вдоль некоторой возрастающей выпуклой вниз линии. Поэтому следует рассмотреть степенную, или показательную, или квадратическую модель. Мы рассмотрим все эти модели. Для степенной модели вычислим наблюдаемые значения $x_i^* = \ln x_i$ и $y_i^* = \ln y_i^*$; для показательной модели вычислим наблюдаемые значения $y_i^* = \ln y_i^*$; для квадратической модели вычислим наблюдаемые значения $x_i^* = x_i^2$:

$x_i^* = \ln x_i$	0,182	0,588	0,916	1,131	1,361	1,526	1,668	1,792	1,902	2,001
$y_i^* = \ln y_i^*$	0,262	0,742	1,131	1,253	1,740	1,960	2,293	2,398	2,715	2,913
$x_i^* = x_i^2$	1,44	3,24	6,25	9,61	15,21	21,16	28,09	36,00	44,89	54,76

3. Вычислим выборочные средние.

Степенная модель:

$$\overline{x^*} = \overline{\ln x} = \frac{0{,}182 + 0{,}588 + 0{,}916 + 1{,}131 + 1{,}361}{10} +$$

$$+\frac{1{,}526+1{,}668+1{,}792+1{,}902+2{,}001}{10}=1{,}307;$$

$$\overline{y^*}=\overline{\ln y}=\frac{0{,}262+0{,}742+1{,}131+1{,}253+1{,}740}{10}+$$

$$+\frac{1{,}960+2{,}293+2{,}398+2{,}715+2{,}912}{10}=1{,}741;$$

$$\overline{x^*y^*}=\frac{0{,}182\cdot 0{,}262+0{,}588\cdot 0{,}742+0{,}916\cdot 1{,}131+1{,}131\cdot 1{,}253}{10}+$$

$$+\frac{1{,}361\cdot 1{,}740+1{,}526\cdot 1{,}960+1{,}668\cdot 2{,}293+1{,}792\cdot 2{,}398}{10}+$$

$$+\frac{1{,}902\cdot 2{,}715+2{,}001\cdot 2{,}912}{10}=2{,}741.$$

Показательная модель:

$$\bar{x}=\frac{1{,}2+1{,}8+2{,}5+3{,}1+3{,}9+4{,}9+5{,}3+6{,}0+6{,}7+7{,}4}{10}=4{,}25;$$

$$\overline{y^*}=\overline{\ln y}=\frac{0{,}262+0{,}742+1{,}131+1{,}253+1{,}740}{10}+$$

$$+\frac{1{,}960+2{,}293+2{,}398+2{,}715+2{,}912}{10}=1{,}741;$$

$$\overline{xy^*}=\frac{1{,}2\cdot 0{,}262+1{,}8\cdot 0{,}742+2{,}5\cdot 1{,}131+3{,}1\cdot 1{,}253+3{,}9\cdot 1{,}740}{10}+$$

$$+\frac{4{,}9\cdot 1{,}960+5{,}3\cdot 2{,}293+6{,}0\cdot 2{,}398+6{,}7\cdot 2{,}715+7{,}4\cdot 2{,}912}{10}=9{,}044.$$

Квадратическая модель:

$$\overline{x^*}=\overline{x^2}=\frac{1{,}44+3{,}24+6{,}25+9{,}61+15{,}21+21{,}16}{10}+$$

$$+\frac{28{,}09+36{,}00+44{,}89+54{,}76}{10}=22{,}065;$$

$$\bar{y}=\frac{1{,}3+2{,}1+3{,}1+3{,}5+5{,}7+7{,}1+9{,}9+11{,}0+15{,}1+18{,}4}{10}=7{,}72;$$

$$\overline{x^*y}=\frac{1{,}44\cdot 1{,}3+3{,}24\cdot 2{,}1+6{,}25\cdot 3{,}1+9{,}61\cdot 3{,}5+15{,}21\cdot 5{,}7}{10}+$$

$$+\frac{21{,}16\cdot 7{,}1+28{,}09\cdot 9{,}9+36{,}00\cdot 11{,}0+44{,}89\cdot 15{,}1+54{,}76\cdot 18{,}4}{10}$$

$$=265{,}8133.$$

5. Найдем выборочные средние квадратические отклонения.

Степенная модель:

$$\sigma_{x^*} = (\frac{0{,}182^2 + 0{,}588^2 + 0{,}916^2 + 1{,}131^2 + 1{,}361^2}{10} +$$

$$+\frac{1{,}526^2 + 1{,}668^2 + 1{,}792^2 + 1{,}902^2 + 2{,}001^2}{10} - 1{,}307^2)^{\frac{1}{2}} = 0{,}567;$$

$$\sigma_{y^*} = (\frac{0{,}262^2 + 0{,}742^2 + 1{,}131^2 + 1{,}253^2 + 1{,}740^2}{10} +$$

$$+\frac{1{,}960^2 + 2{,}293^2 + 2{,}398^2 + 2{,}715^2 + 2{,}912^2}{10} - 1{,}741^2)^{\frac{1}{2}} = 0{,}831.$$

Показательная модель:

$$\sigma_x = \left(\frac{1{,}2^2 + 1{,}8^2 + 2{,}5^2 + 3{,}1^2 + 3{,}9^2 + 4{,}9^2 + 5{,}3^2 + 6{,}0^2 + 6{,}7^2 + 7{,}4^2}{10}\right.$$

$$\left. - 4{,}25^2\right)^{\frac{1}{2}} = 2{,}001;$$

$$\sigma_{y^*} = (\frac{0{,}262^2 + 0{,}742^2 + 1{,}131^2 + 1{,}253^2 + 1{,}740^2}{10} +$$

$$+\frac{1{,}960^2 + 2{,}293^2 + 2{,}398^2 + 2{,}715^2 + 2{,}912^2}{10} - 1{,}741^2)^{\frac{1}{2}} = 0{,}831.$$

Квадратическая модель:

$$\sigma_{x^*} = (\frac{1{,}44^2 + 3{,}24^2 + 6{,}25^2 + 9{,}61^2 + 15{,}21^2 + 21{,}16^2}{10} +$$

$$+\frac{28{,}09^2 + 36{,}00^2 + 44{,}89^2 + 54{,}76^2}{10} - 22{,}065^2)^{\frac{1}{2}} = 17{,}474;$$

$$\sigma_y = \left(\frac{1{,}3^2 + 2{,}1^2 + 3{,}1^2 + 3{,}5^2 + 5{,}7^2 + 7{,}1^2 + 9{,}9^2 + 11{,}0^2 + 15{,}1^2 + 18{,}4^2}{10}\right.$$

$$\left. - 7{,}72^2\right)^{\frac{1}{2}} = 5{,}481.$$

6. Вычислим выборочный коэффициент корреляции. **Степенная модель**:

$$r_{\text{в}} = \frac{2{,}741 - 1{,}307 \cdot 1{,}741}{0{,}567 \cdot 0{,}831} = 0{,}990.$$

Показательная модель:

$$r_{\text{в}} = \frac{9{,}044 - 4{,}25 \cdot 1{,}741}{2{,}001 \cdot 0{,}831} = 0{,}991.$$

Квадратическая модель:

$$r_{\text{в}} = \frac{265{,}8133 - 22{,}065 \cdot 7{,}72}{17{,}474 \cdot 5{,}481} = 0{,}997.$$

7. Проанализируем значение выборочного коэффициента корреляции. Заметим, что для всех трех выбранных моделей этот коэффициент больше числа 0,7; следовательно, каждая из этих моделей вполне адекватна, и мы можем вычислить оценки их параметров.

   Степенная модель:
$$a_1 = 0{,}990 \cdot \frac{0{,}831}{0{,}567} = 1{,}449;$$
$$a_0^* = 1{,}741 - 1{,}449 \cdot 1{,}307 = -0{,}153;$$
$$a_0 = e^{-0{,}153} = 0{,}858.$$

   Показательная модель:
$$a_1 = 0{,}991 \cdot \frac{0{,}831}{2{,}001} = 0{,}411;$$
$$a_0^* = 1{,}741 - 0{,}411 \cdot 4{,}25 = -0{,}008;$$
$$a_0 = e^{-0{,}008} = 0{,}992.$$

   Квадратическая модель:
$$a_1 = 0{,}997 \cdot \frac{5{,}481}{17{,}474} = 0{,}313;$$
$$a_0 = 7{,}72 - 0{,}313 \cdot 22{,}065 = 0{,}821.$$

8. Запишем эмпирические уравнения регрессии.

   Степенная модель:
$$y = 0{,}858 \cdot x^{1{,}449}.$$
   Показательная модель:
$$y = 0{,}992 \cdot e^{0{,}411x}.$$
   Квадратическая модель:
$$y = 0{,}821 + 0{,}313 \cdot x^2.$$

9. Выясним, какая из рассмотренных моделей адекватнее. Для этого вычислим и сравним соответствующие коэффициенты детерминации. С этой целью предварительно найдем значения $y(x_i) - y_i$, а также $y_i - \overline{y}$; при этом заметим, что для степенной модели $y(x_i) = 0{,}858 \cdot {x_i}^{1{,}449}$, для показательной модели $y(x_i) = 0{,}992 \cdot e^{0{,}411x_i}$, а для квадратической модели $y(x_i) = 0{,}821 + 0{,}313 \cdot {x_i}^2$.

$y_i - \overline{y}$	−6,42	−5,62	−4,62	−4,22	−2,02	−0,62	2,18	3,28	7,38	10,68
$y(x_i) - y_i$ (степенная модель)	−0,18	−0,09	0,14	0,92	0,47	0,73	−0,28	0,51	−1,59	−2,80
$y(x_i) - y_i$ (показательная модель)	0,33	−0,02	−0,32	0,05	−0,76	−0,52	−1,12	0,71	0,52	2,43
$y(x_i) - y_i$ (квадратическая	−0,03	−0,27	−0,33	0,33	−0,12	0,34	−0,30	1,08	−0,24	−0,46

модель)											

Степенная модель:

$$R^2 = 1 - \frac{(-0{,}18)^2 + (-0{,}09)^2 + 0{,}14^2 + 0{,}92^2 + 0{,}47^2 + 0{,}73^2 + (-0{,}28)^2 + 0{,}51^2 + (-1{,}59)^2 + (-2{,}80)^2}{(-6{,}42)^2 + (-5{,}62)^2 + (-4{,}62)^2 + (-4{,}22)^2 + (-2{,}02)^2 + (-0{,}62)^2 + 2{,}18^2 + 3{,}28^2 + 7{,}38^2 + 10{,}68^2} = 0{,}96.$$

Показательная модель:

$$R^2 = 1 - \frac{0{,}33^2 + (-0{,}02)^2 + (-0{,}32)^2 + 0{,}05^2 + (-0{,}76)^2 + (-0{,}52)^2 + (-1{,}12)^2 + 0{,}71^2 + 0{,}52^2 + 2{,}43^2}{(-6{,}42)^2 + (-5{,}62)^2 + (-4{,}62)^2 + (-4{,}22)^2 + (-2{,}02)^2 + (-0{,}62)^2 + 2{,}18^2 + 3{,}28^2 + 7{,}38^2 + 10{,}68^2} = 0{,}97.$$

Квадратическая модель:

$$R^2 = 1 - \frac{(-0{,}03)^2 + (-0{,}27)^2 + (-0{,}33)^2 + 0{,}33^2 + (-0{,}12)^2 + 0{,}34^2 + (-0{,}30)^2 + 1{,}08^2 + (-0{,}24)^2 + (-0{,}46)^2}{(-6{,}42)^2 + (-5{,}62)^2 + (-4{,}62)^2 + (-4{,}22)^2 + (-2{,}02)^2 + (-0{,}62)^2 + 2{,}18^2 + 3{,}28^2 + 7{,}38^2 + 10{,}68^2} = 0{,}99.$$

Напомним также, что коэффициент детерминации для рассмотренной в примере 1 парной линейной регрессии оказался равным 0,94. Таким образом, поскольку коэффициент детерминации квадратической модели среди всех рассмотренных наиболее близок к 1 (0,99), то именно квадратическая модель $y = 0{,}821 + 0{,}313 \cdot x^2$ наиболее адекватна.

6. Проверка выполнения условий применения МНК

Как уже указывалось выше, МНК обеспечивает хорошие результаты в случае выполнения определенных условий. Если же какие-нибудь из этих условий не выполняются, то полученные с помощью МНК оценки параметров построенного эмпирического уравнения регрессии перестанут быть надежными, и тогда это уравнение не будет достаточно адекватным. В этом аспекте особенно значимыми являются условия **гомоскедастичности**, **отсутствия автокорреляции** и **отсутствия мультиколлинеарности**. Таким образом, для проверки качества (адекватности) построенного эмпирического уравнения регрессии на этапе верификации кроме вычисления коэффициента детерминации нужно также выяснить, все ли условия применения МНК выполнены, и в случае невыполнения каких-либо из этих условий попытаться исправить ситуацию.

6.1. Гетероскедастичность

Гетероскедастичностью называется нарушение условия гомоскедастичности. Напомним, что гомоскедастичность состоит в том, что дисперсия отклонений $\varepsilon_i = y(x_i) - y_i$ постоянна, т.е. ее значение не зависит от номера наблюдения i. Поскольку дисперсия любой случайной величины, в том числе ε_i, характеризует степень рассеяния значений этой величины вокруг ее среднего значения, выполнение или нарушение условия гомоскедастичности можно проверить графически. Для этого на координатную плоскость нанесем точки (x_i, ε_i); в случае гомоскедастичности область заполнения этих точек будет напоминать полосу, расположенную симметрично вдоль горизонтальной координатной оси (см. Рисунок 5.1); любая другая конфигурация точек (x_i, ε_i) будет свидетельствовать о гетероскедастичности. Заметим, что графический анализ отклонений ε_i является самым простым, но не единственным методом обнаружения гетероскедастичности; для этого используются также тест ранговой корреляции Спирмена, тест Парка, тест Глейзера, тест Голдфелда-Квандта (более подробно об этом см. [1]).

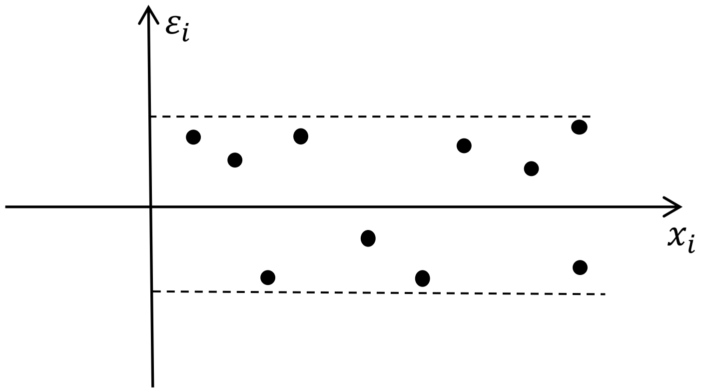


Рисунок 6.1 – Графический признак гомоскедастичности

Учитывая, что даже при выборе нелинейной парной модели для того чтобы использовать МНК, мы с помощью специальных приемов эту модель приводим к линейной, всюду ниже в данном разделе мы будем рассматривать именно линейную парную модель.

Естественным образом возникает вопрос, что нужно делать при обнаружении гетероскедастичности. Рассмотрим сначала случай, когда дисперсии σ_i^2 отклонений ε_i известны. Тогда для преодоления гетероскедастичности каждую пару наблюдаемых значений (x_i, y_i) нужно разделить на соответствующее значение σ_i; после этого введем в рассмотрение новые переменные $x_i^* = x_i/\sigma_i$, $y_i^* = y_i/\sigma_i$, а также дополнительную объясняющую переменную $z_i = 1/\sigma_i$. Эти новые переменные уже будут удовлетворять условию гомоскедастичности; для них с помощью МНК построим эмпирическое уравнение множественной линейной регрессии без свободного члена. Такой подход получил название **метода взвешенных наименьших квадратов** (МВНК).

Заметим, что на практике значения σ_i^2 отклонений ε_i известны редко. В этой ситуации можно предположить, что либо дисперсии σ_i^2 пропорциональны значениям x_i, либо средние квадратические отклонения σ_i пропорциональны значениям x_i. В первом случае множество точек (x_i, ε_i) расположено внутри некоторой параболы, осью которой является горизонтальная ось (см. Рисунок 5.2); во втором случае множество точек (x_i, ε_i) расположено внутри некоторого угла, биссектрисой которого является горизонтальная ось (см. Рисунок 5.3).

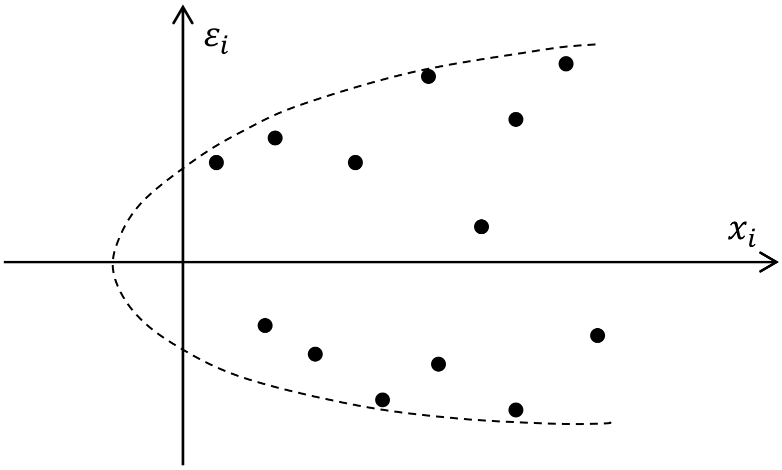


Рисунок 6.2 – Пропорциональность σ_i^2 и объясняющей переменной

Если дисперсии σ_i^2 пропорциональны значениям x_i, то каждую пару наблюдаемых значений (x_i, y_i) нужно разделить на соответствующее значение $\sqrt{x_i}$; после этого введем в рассмотрение новые переменные $x_i^* = \sqrt{x_i}$, $y_i^* = y_i/\sqrt{x_i}$, а также дополнительную объясняющую переменную $z_i = 1/\sqrt{x_i}$. Эти новые переменные уже будут удовлетворять условию гомоскедастичности; для них с помощью МНК построим эмпирическое уравнение множественной линейной регрессии без свободного члена $y^* = a_0 z + a_1 x^*$. По найденным коэффициентам a_0 и a_1 составим искомое эмпирическое уравнение парной линейной регрессии $y = a_0 + a_1 \cdot x$.

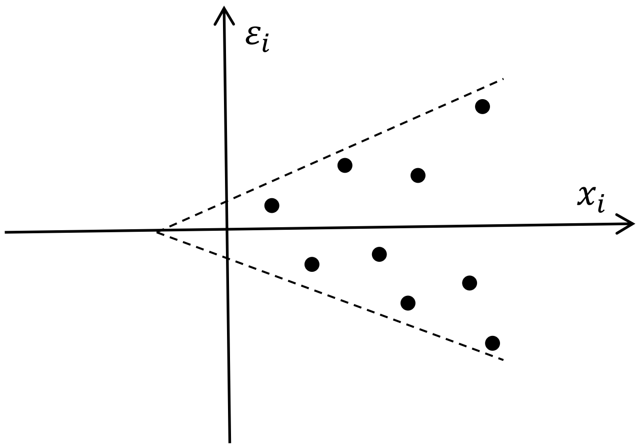


Рисунок 6.3 – Пропорциональность σ_i и объясняющей переменной

Если же средние квадратические отклонения σ_i пропорциональны значениям x_i, то каждую пару наблюдаемых значений (x_i, y_i) нужно разделить на соответствующее значениеx_i; после этого введем в рассмотрение новые переменные, $y_i^* = y_i/x_i$ и $z_i = 1/x_i$. Эти новые переменные уже будут удовлетворять условию гомоскедастичности; для них с помощью МНК построим эмпирическое уравнение парной линейной регрессии $y^* = a_0 z + a_1$. По найденным коэффициентам a_0 и a_1 составим искомое эмпирическое уравнение парной линейной регрессии $y = a_0 + a_1 \cdot x$.

6.2. Автокорреляция

Как отмечалось выше, одной из предпосылок применимости МНК является отсутствие автокорреляции остатков, т.е. отсутствие зависимости значений отклонений ε_i в i-ом наблюдении от значений отклонений во всех остальных наблюдениях. Среди негативных последствий автокорреляции при применении МНК чаще всего выделяют следующие. Во-первых, полученные оценки параметров, оставаясь несмещенными, перестают быть эффективными, что ухудшает их качество. Во-вторых, дисперсии оценок параметров являются смещенными, чаще всего заниженными, что может привести к признанию статистически значимыми объясняющие переменные, на самом деле таковыми не являющиеся. В силу вышесказанного прогнозные качества модели ухудшаются.

Выделяют следующие основные причины автокорреляции.

- Ошибки спецификации – неучет в модели какой-либо важной объясняющей переменной или же неправильный выбор формы зависимости объясняемой переменной от объясняющих переменных.
- Игнорирование эффекта инерции, который состоит в том, что многие экономические показатели, отражаемые объясняемой переменной, реагируют на изменение экономических условий, отражаемых объясняющими переменными, не мгновенно, а с некоторым запаздыванием.
- Сглаживание данных – часто данные по некоторому достаточно продолжительному временному периоду усредняют за весь этот

период, что приводит к сглаживанию колебаний, имевшихся внутри рассматриваемого периода.

Различают положительную и отрицательную автокорреляцию. Отрицательная автокорреляция означает, что за положительным отклонением ε_i следует отрицательное и наоборот. Автокорреляция, не являющаяся отрицательной, называется положительной.

В связи с вышесказанным особую значимость обретает проблема обнаружения автокорреляции. Здесь можно выделить три основных способа решения этой проблемы. Во-первых, **графический способ**, согласно которому на координатную плоскость наносят точки с координатами (i, ε_i); если нанесенные точки располагаются вдоль некоторой линии, то можно предположить наличие автокорреляции; если же эти точки располагаются хаотично, то нет оснований предполагать наличие автокорреляции.

Во-вторых, **метод рядов**, согласно которому последовательно определяются знаки отклонений ε_i («+» и «–»); после этого определяется количество рядов k (под рядом понимают непрерывную последовательность одинаковых знаков), а также длины рядов (количество знаков в каждом ряду). Пусть n – объем выборки, n_1 – общее количество знаков «+», n_2 – общее количество знаков «–»; вычислим значения

$$M(k) = \frac{2n_1n_2}{n_1+n_2} + 1; \quad D(k) = \frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2-1)}.$$

Тогда при $n_1 > 10$ и $n_2 > 10$ в случае выполнения двойного неравенства

$$M(k) - u_{\alpha/2} \cdot D(k) < k < M(k) + u_{\alpha/2} \cdot D(k)$$

гипотеза об отсутствии автокорреляции принимается; в противном случае эта гипотеза принимается. Здесь значение $u_{\alpha/2}$ определяется по таблице значений функции Лапласа Φ (см. [1, с. 335]), исходя из условия $\Phi\left(u_{\alpha/2}\right) = \alpha/2$, где α – уровень значимости. При небольшом количестве наблюдений ($n_1 < 20$ и $n_2 < 20$) из таблиц критических значений количества рядов по имеющимся значениям n_1 и n_2 определяют нижнее критическое значение k_1 и верхнее критическое значение k_2 количества рядов при уровне значимости $\alpha = 0{,}05$ (см. [1, с. 348]); тогда если $k \leq k_1$, то говорят о положительной автокорреляции остатков; если $k \geq k_2$, то говорят об отрицательной автокорреляции; если же $k_1 < k < k_2$, то говорят об отсутствии автокорреляции.

В-третьих, **критерий Дарбина-Уотсона**, согласно которому вычисляется статистика Дарбина-Уотсона

$$DW = \frac{(\varepsilon_2 - \varepsilon_1)^2 + (\varepsilon_3 - \varepsilon_2)^2 + \cdots + (\varepsilon_n - \varepsilon_{n-1})^2}{{\varepsilon_1}^2 + {\varepsilon_2}^2 + \cdots + {\varepsilon_n}^2}.$$

Из таблицы критических точек Дарбина-Уотсона по объему выборки n и количеству объясняющих переменных m при заданном уровне значимости α определяют нижнюю d_1 и верхнюю d_u границы области допустимости статистики DW (см. [1, с. 346-347]). Тогда:

- если $DW < d_1$, то существует положительная автокорреляция;
- если $d_1 < DW < d_u$, то ничего определенного о наличии автокорреляции сказать нельзя;
- если $d_u < DW < 4 - d_u$, то автокорреляция отсутствует;
- если $4 - d_u < DW < 4 - d_1$, то ничего определенного о наличии автокорреляции сказать нельзя;
- если $4 - d_1 < DW < 4$, то существует отрицательная автокорреляция.

Следует заметить, что использование критерия Дарбина-Уотсона целесообразно при следующих ограничениях.

1. Рассматриваемая модель должна содержать свободный член (числовое слагаемое, не содержащее переменных).
2. Отклонения ε_i определяются так называемой авторегрессионной схемой первого порядка AR(1), т.е. должно выполняться рекуррентное соотношение $\varepsilon_i = \rho \cdot \varepsilon_{i-1} + \vartheta_i$, где ρ – некоторая постоянная величина, а ϑ_i – некоторая случайная величина.
3. Статистические данные должны иметь одинаковую периодичность, т.е. не должно быть пропусков в наблюдениях.
4. В рассматриваемой модели объясняемая переменная y_i не должна зависеть от y_{i-1}.

Отметим также, что статистика Дарбина-Уотсона приводится во всех эконометрических и статистических пакетах прикладных программ.

Естественным образом возникает вопрос о преодолении существующей автокорреляции остатков. При обнаружении автокорреляции прежде всего нужно попробовать изменить модель (либо сам вид модели, либо состав объясняющих переменных). Если после всех возможных изменений модели автокорреляция не исчезает, то, возможно, ее причина кроется во внутренних свойствах ряда отклонений $\{\varepsilon_i\}$. В этом случае можно воспользоваться авторегрессионной схемой первого порядка AR(1) $\varepsilon_i = \rho \cdot \varepsilon_{i-1} + \vartheta_i$ (см. [1, с. 236-239]).

6.2. Мультиколлинеарность

Прежде всего отметим, что мультиколлинеарность может возникнуть лишь в моделях с несколькими объясняющими переменными и состоит она в наличии достаточно сильной взаимозависимости между некоторыми объясняющими переменными. Выделяют следующие последствия мультиколлинеарности. Во-первых, мультиколлинеарность приводит к ухудшению точности оценок истинных значений определяемых в данной модели величин. Во-вторых, завышается значимость объясняющих переменных, на самом деле не являющихся достаточно значимыми. В-третьих, оценки коэффициентов модели, полученные с помощью МНК, становятся неустойчивыми, т.е. слишком чувствительными к малейшим изменениям данных. В-четвертых, искажается степень вклада каждой объясняющей переменной в значение объясняемой переменной. В-пятых, возможно даже получение неверного знака у коэффициентов регрессии.

По поводу проблемы обнаружения мультиколлинеарности можно сказать следующее. Если объясняющих переменных только две, то признаком наличия мультиколлинеарности является достаточно высокий коэффициент корреляции между объясняющими переменными

$$r_{12} = \frac{\overline{x_1 x_2} - \overline{x_1} \cdot \overline{x_2}}{\sigma_1 \cdot \sigma_2}.$$

Здесь σ_1 и σ_2 обозначают выборочные средние квадратические отклонения для объясняющих переменных x_1 и x_2, соответственно. Если же объясняющих переменных больше двух, то нужно рассматривать так называемые **частные коэффициенты корреляции** между каждыми двумя объясняющими переменными. Для вычисления частных коэффициентов корреляции нужно сначала вычислить (обычные) коэффициенты корреляции r_{ij} между каждыми двумя объясняющими переменными x_i и x_j и составить из них матрицу

$$R = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & 1 \end{pmatrix}.$$

Затем для матрицы R вычисляют обратную матрицу $C^* = R^{-1}$; тогда частный коэффициент корреляции между объясняющими переменными x_i и x_j находится по формуле

$$r_{ij}^p = \frac{-c_{ij}^*}{\sqrt{c_{ii}^* \cdot c_{jj}^*}},$$

где символами c_{ij}^* обозначены элементы матрицы C^*. Если окажется, что какой-то из этих коэффициентов по абсолютной величине достаточно близок к 1, то между соответствующими объясняющими переменными существует достаточно сильная связь, и, значит, одну из этих переменных нужно исключить из модели. Заметим, что таким же образом можно вычислить частный коэффициент корреляции r_i^p между объясняемой переменной y и любой объясняющей переменной x_i; тогда квадрат этого коэффициента $r_i^{p^2}$, называемый частным коэффициентом детерминации, указывает процент дисперсии объясняемой переменной y, определяемый соответствующей объясняющей переменной x_i. Другими словами, коэффициента $r_i^{p^2}$ позволяет оценить вклад объясняющей переменной x_i в рассеивание значений объясняемой переменной y.

Касаясь проблемы устранения мультиколлинеарности, прежде всего отметим, что не всегда оправданы большие затраты на такое устранение. В частности, если основной задачей построенной модели является прогноз будущих значений объясняемой переменной, то при достаточно большом значении коэффициента детерминации $(R^2 \geq 0{,}9)$ наличие мультиколлинеарности, как правило, незначительно сказывается на качестве модели. Если же задачей рассматриваемой модели является определение степени влияния каждой объясняющей переменной на зависимую переменную, то наличие мультиколлинеарности становится серьезной проблемой. Заметим также, что единого метода устранения мультиколлинеарности, пригодного в любой ситуации, не существует. Наиболее часто используют способы устранения мультиколлинеарности.

1. Исключение из модели одной или нескольких коррелированных объясняющих переменных.
2. Увеличение объема выборки.
3. Изменение спецификации модели – либо ее формы, либо состава объясняющих переменных.
4. Использование предварительной информации о значениях некоторых параметров модели.

ПРАКТИЧЕСКИЙ РАЗДЕЛ

Линейная регрессия

1. По следующим статистическим данным с помощью МНК построить эмпирические уравнения регрессии:

а) $y = a_0 + a_1 x_1$;

б) $y = b_0 + b_1 x_2$;

в) $y = c_0 + c_1 x_1 + c_2 x_2$.

Для каждого из этих уравнений вычислить коэффициент детерминации и выяснить, какое из них (уравнений) адекватнее:

x_1	0	1	3	4
x_2	3	1	0	–2
y	1	3	5	11

2. Объем поставок Q товара, производимого некоторой фирмой, зависит линейно как от цены P на этот товар, так и от заработной платы W сотрудников этой фирмы. На основе приведенных ниже статистических данных с помощью МНК построить эмпирическое уравнение множественной регрессии, отражающее эту зависимость:

Q	20	35	30	45	60	69	75	90	105	110	120	130	130	130	135	140
P	10	15	20	25	40	37	43	35	38	55	50	35	40	55	45	65
W	12	10	9	9	8	8	6	4	4	5	3	1	2	3	1	2

3. Для объяснения изменения ВНП (млрд руб.) за 10 лет строится регрессионная модель с объясняющими переменными C (потребление, млрд руб.) и I (инвестиции, млрд руб.). На основе приведенных ниже статистических данных с помощью МНК построить эмпирическое уравнение множественной линейной регрессии, отражающее зависимость ВНП от объясняющих переменных:

ВНП	8	9,5	11	12	13	14	15	16,5	17	18
C	1,65	1,8	2,0	2,1	2,2	2,4	2,65	2,85	3,2	3,55
I	14	16	18	20	23	23,5	25	26,5	28,5	30,5

Парная нелинейная регрессия

На основе приведенных ниже (задания 1 –4) статистических данных за несколько лет построить эмпирические уравнения парной регрессии (линейной и подходящих нелинейных) и выяснить, какое из них наиболее адекватно отражает реальную зависимость объясняемой переменной Y от объясняющей переменной X.

1. Y – прибыль предприятия, X – расходы на рекламу.

Y	5	7	13	15	20	25	22	20	17
X	0,8	1,0	1,8	2,5	4,0	5,7	7,5	8,3	8,8

2. Y – индекс потребительских цен, X – объем денежной массы.

Y	72,5	77,5	82,0	85,5	88,5	91,0	95,0	100,0	106,5	112,0	115,5	118,5	120,0	120,5	121,0
X	132	137	160	177	192	215	235	240	245	250	275	285	295	320	344

3. Y – темп прироста заработной платы, X – уровень безработицы.

Y	1,61	1,66	1,80	1,95	2,05	2,12	2,25	2,45	2,55	2,67	2,73	2,80	2,93	3,02	3,15
X	1,0	1,38	1,15	1,50	1,55	1,20	1,10	1,0	1,35	1,80	1,90	1,45	1,85	1,20	1,50

4. Y – объем экспорта, X – год.

X	1	2	3	4	5	6	7	8	9	10	11	12
Y	54,1	35,4	56,6	46,6	47,6	52,1	56,6	44,8	68,3	36,3	65,0	57,2

5. В таблице приведены данные по объему выпуска продукции Q, а также затрат капитала K и труда L за 15 лет. Используя эти данные, оценить производственную функцию Кобба-Дугласа $Q = a \cdot K^{\alpha} \cdot L^{\beta}$ (с помощью логарифмирования привести данную нелинейную модель к множественной линейной и затем применить МНК).

Q	46	59	37,5	107	130	128	154	227	147	31,5	70,5	70,5	108	90,5	74
K	2	5,6	5,6	2	10,4	5,6	10,4	10,4	2	2	5,6	5,6	2	10,4	5,6
L	2	2	4	4	6	4	6	24	2	4	6	2	6	4	2

Гетероскедастичность

1. В таблице приведены данные в условных единицах по доходам X и расходам на непродовольственные товары Y для 30 домохозяйств. По этим данным с помощью МНК построить эмпирическое уравнение парной линейной регрессии, отражающее зависимость величины Y от величины X; выяснить, имеет ли место гетероскедастичность; если да, то смягчить ее.

X	26,2	33,1	42,5	47,0	48,5	49,0	49,1	50,9	52,4	53,2	54,0	54,8	59,0	61,3	62,5
Y	10,0	11,2	15,0	20,5	21,2	19,5	23,0	19,0	19,5	18,0	24,5	21,5	35,4	25,0	17,3

X	63,1	64,0	66,2	70,0	71,5	73,2	75,4	76,0	80,6	81,2	83,3	92,0	95,5	103,2	110,4
Y	21,6	15,3	32,6	34,0	23,8	22,5	27,4	40,0	23,5	20,0	40,1	15,5	39,0	47,4	21,3

2. Для предприятий некоторой отрасли анализируют зависимость заработной платы Y сотрудников от количества X сотрудников предприятия. Данные по 30 случайно отобранным предприятиям приведены в таблице. По этим данным с помощью МНК построить эмпирическое уравнение парной регрессии; выяснить, имеет ли место гетероскедастичность; если да, то смягчить ее.

Y						X
75,5	75,5	77,5	78,5	80,0	81,0	100
80,5	82,0	84,5	85,0	85,5	86,5	200
85,5	88,5	90,0	91,0	95,0	96,0	300
93,0	93,5	97,5	99,0	102,5	105,0	400
102,0	105,5	107,0	110,5	115,0	118,5	500

3. Проверяется предположение, что заработная плата наемных рабочих пропорциональна их трудовому стажу. Для этого обследуются по 20 рабочих восьми категорий стажа. Соответствующие данные приведены в таблице. По этим данным с помощью МНК построить эмпирическое уравнение парной регрессии, отражающее зависимость величины зарплаты от величины стажа; выяснить, имеет ли место гетероскедастичность; если да, то смягчить ее.

Стаж	[0;5)	[5;10)	[10;15)	[15;20)	[20;25)	[25;30)	[30;35)	[35;40)
З/п	10,0	12,5	14,3	18,7	25,4	29,0	32,0	34,3

4. Проводится анализ зависимости средней заработной платы от средней производительности на предприятиях различного масштаба. Соответствующие данные приведены в таблице. По этим данным с помощью МНК построить эмпирическое уравнение парной линейной регрессии,

отражающее зависимость величины средней зарплаты от величины средней производительности; выяснить, имеет ли место гетероскедастичность; если да, то смягчить ее (используя метод взвешенных наименьших квадратов).

Кол-во сотрудников предприятия	Средняя производительность	Средняя заработная плата	Стандартное отклонение заработной платы
1 – 4	9320	3320	740
5 – 9	8630	3640	850
10 – 19	8050	3900	730
20 – 49	8320	4120	820
50 – 99	8600	4090	950
100 – 199	9120	4200	1100
200 – 499	9540	4380	1250
500 –999	9730	4500	1290
1000 – 1999	10120	4610	1350
2000 – 4999	10740	4800	1100
>5000	1120	5000	1520

Автокорреляция

1. По квартальным данным за 9 лет анализируют зависимость между экспортом EX и импортом IM; соответствующие данные приведены в таблице. По этим данным построить эмпирическое уравнение регрессии, отражающее зависимость величины IM от величины EX; проверить качество этого уравнения (используя коэффициенты корреляции и детерминации); с помощью графического метода, метода рядов и метода Дарбина-Уотсона выяснить наличие автокорреляции.

EX	12,47	12,65	12,89	12,97	13,00	13,31	13,25	12,65	14,49	14,47	14,74	14,62
IM	11,07	11,50	12,01	12,28	13,16	13,43	13,28	13,50	15,32	15,62	17,44	16,14

EX	17,60	17,70	16,60	15,26	19,49	19,08	18,69	18,65	19,33	19,11	18,62	18,40
IM	16,13	16,08	16,55	15,00	18,72	17,80	16,64	17,39	18,70	18,02	17,46	16,96

EX	16,15	16,58	17,60	18,48	15,36	15,25	15,61	15,93	14,38	14,30	14,75	15,58
IM	15,06	16,01	16,63	17,86	14,56	15,64	16,45	17,42	14,30	14,59	14,66	14,95

2. В таблице приведены данные за 25 лет по темпам прироста заработной платы Y и производительности труда X_1, а также по годовым темпам инфляции X_2. По этим данным построить эмпирическое уравнение множественной линейной регрессии, отражающее зависимость величины Y от величин X_1 и X_2; проверить качество этого уравнения, используя коэффициент детерминации; выяснить наличие гетероскедастичности и автокорреляции.

Год	1	2	3	4	5	6	7	8	9	10	11	12	13
X_1	3,5	2,8	6,3	4,5	3,1	1,5	7,6	6,7	4,2	2,7	4,5	3,5	5,0
X_2	4,5	3,0	3,1	3,8	3,8	1,1	2,3	3,6	7,5	8,0	3,9	4,7	6,1
Y	9,0	6,0	8,9	9,0	7,1	3,2	6,5	9,1	14,6	11,9	9,2	8,8	12,0

Год	14	15	16	17	18	19	20	21	22	23	24	25
X_1	2,3	2,8	1,5	6,0	2,9	2,8	2,6	1,5	0,9	0,6	0,7	3,1
X_2	6,9	3,5	7,1	3,1	3,7	3,9	4,0	4,8	4,8	4,2	4,9	3,2
Y	12,5	6,7	8,5	5,9	6,8	5,6	4,8	4,5	6,7	5,5	4,0	3,3

3. В таблице приведены данные за 25 лет по годовым темпам инфляции Y и уровню безработицы X. По этим данным построить эмпирическое уравнение парной нелинейной регрессии, отражающее зависимость величины Y от величины X (использовать степенную модель). Для

линеаризованной модели с помощью графического метода, метода рядов и метода Дарбина-Уотсона выяснить наличие автокорреляции.

Y	3,07	0,70	4,08	2,20	2,38	0,90	1,10	5,12	0,93	2,54	1,55	3,45	1,09
X	3,69	9,10	3,92	6,50	4,63	8,50	9,55	3,71	5,80	3,60	5,63	4,32	9,20

Y	2,15	5,14	1,72	0,74	4,16	0,93	1,79	1,24	1,12	1,28	7,36	5,30
X	5,75	3,65	7,30	9,65	3,65	9,80	6,28	7,80	8,75	7,22	3,60	3,65

4. По результатам 30-летних наблюдений строится функция инвестиций $I = a_0 + a_1 Y + a_2 R$, где I – годовой объем инвестиций, Y – годовой объем валового национального продукта, R – ключевая процентная ставка в рассматриваемом году. С помощью МНК оценить коэффициенты этой функции; с помощью коэффициента детерминации оценить качество полученного эмпирического уравнения регрессии; с помощью метода Дарбина-Уотсона выяснить наличие автокорреляции.

Y	8,6	10,5	8,4	10,7	9,7	12,0	13,5	14,2	14,5	13,9	16,6	18,0	18,4	20,4	21,0
R	18,1	11,1	9,0	17,0	16,3	13,8	20,0	18,8	13,8	9,6	19,3	15,2	12,4	16,5	6,0
I	11,6	13,3	11,0	10,5	15,1	17,5	17,8	16,1	10,6	10,7	9,3	11,0	15,1	15,1	22,7

Y	23,8	25,8	24,2	25,2	26,2	28,6	30,6	31,3	26,0	26,9	32,1	33,0	33,3	33,9	35,6
R	17,5	16,4	7,4	15,5	19,2	5,5	9,5	8,0	7,5	19,9	8,7	21,4	11,1	15,8	21,7
I	22,0	21,3	25,7	26,2	25,6	28,1	24,2	32,3	21,5	23,0	30,5	24,6	32,5	31,2	29,5

Мультиколлинеарность

1. Имеется выборка из 10 наблюдений за переменными X_1, X_2 и Y. По этим данным выяснить наличие мультиколлинеарности. Построить эмпирическое уравнение множественной линейной регрессии Y на X_1 и X_2, а также уравнения парной линейной регрессии Y на X_1 и Y на X_2; для каждого из этих уравнений вычислить коэффициент детерминации и проанализировать полученные результаты.

X_1	1	2	3	4	5	6	7	8	9	10
X_2	1	1,6	2,2	2,8	3,4	4	4,6	5,2	5,8	6,4
Y	0	3	6	9	12	15	18	21	24	27

2. По выборке объема 50 для объясняющих переменных X_1, X_2, X_3 построена следующая корреляционная матрица:

$$R = \begin{pmatrix} 1{,}0 & 0{,}45 & -0{,}35 \\ 0{,}45 & 1{,}0 & 0{,}52 \\ -0{,}35 & 0{,}52 & 1{,}0 \end{pmatrix}.$$

Вычислить все частные коэффициенты корреляции и на основании их значений указать переменные, которые нужно включить в уравнение линейной регрессии.

3. В таблице приведены данные по валовому национальному продукту Y, объему потребления X_1 и объему инвестиций X_2. По этим данным построить эмпирическое уравнение множественной линейной регрессии, отражающее зависимость величины Y от величин X_1 и X_2; оценить качество этого уравнения; выяснить наличие мультиколлинеарности; построить эмпирические уравнения парной линейной регрессии, отражающие зависимость величины Y от величины X_1, а также величины Y от величины X_2; построить эмпирическое уравнение парной регрессии, отражающие зависимость величины X_1 от величины X_2; проанализировать полученные результаты.

Y	240	248	261	274	273	269	283	296	312	319
X_1	149	154	162	169	167	171	180	188	196	200
X_2	38,2	41,9	46,5	52,1	48,1	38,3	45,4	52,1	56,8	57,5

Y	318	325	317	327	350	361	372	385	402	412
X_1	200	202	205	215	225	235	245	252	261	266
X_2	50,9	54,5	44,7	50,4	65,8	63,7	64,0	76,4	71,6	71,8

РАЗДЕЛ КОНТРОЛЯ ЗНАНИЙ

I. По таблице статистических данных:

а) Построить эмпирическое уравнение парной линейной регрессии;
б) Построить эмпирическое уравнение парной нелинейной регрессии;
в) С помощью коэффициента детерминации выяснить, какое из этих уравнений адекватнее;
г) Выяснить наличие гетероскедастичности и в случае необходимости устранить ее;
д) Выяснить наличие автокорреляции.

1)

Y	1	1,5	2,1	2,7	3,2	3,6	4	4,6	5,1	5,5
X	4,1	4,9	4,9	5,6	6,2	5,9	6,3	6,8	6,5	6,9

2)

Y	2,1	2,4	2,8	3	3,3	3,5	3,8	4,1	4,4	4,9
X	4,7	5,3	5,2	5,6	6,2	5,9	6,5	7,2	7,1	8,0

3)

Y	0,7	1	1,5	1,9	2,3	2,8	3,4	3,9	4,2	4,6
X	4,4	5,8	6,6	7,5	8,4	8,5	9,3	10,0	9,7	10,2

4)

Y	2,2	2,5	2,9	3,2	3,4	3,8	4,1	4,5	4,8	5,2
X	11,4	11,2	10,2	10,1	10,2	9,4	9,4	9,5	8,9	8,9

5)

Y	0,3	0,6	1	1,4	1,8	2,1	2,5	2,7	3	3,4
X	7,3	8,8	10,3	12,8	15,9	18,0	22,2	24,9	28,4	34,9

6)

Y	2,1	2,5	2,9	3,2	3,5	3,8	4	4,3	4,7	5
X	16,1	22,7	29,5	35,9	43,1	50,0	55,5	64,3	76,0	86,1

7)

Y	1,1	1,6	2,2	2,8	3,3	3,7	4,1	4,7	5,2	5,6
X	5,3	6,2	6,3	7,0	7,6	7,4	7,8	8,0	8,7	8,6

8)

Y	2,2	2,5	2,9	3,1	3,4	3,6	3,9	4,2	4,5	5
X	6,3	7,0	7,0	7,5	8,2	8,0	8,7	9,1	10,1	10,8

9)

Y	0,8	1,1	1,6	2	2,4	2,9	3,5	4	4,3	4,7
X	5,0	6,0	6,3	7,0	7,6	7,5	8,1	8,2	8,9	8,7

10)

Y	2,3	2,6	3	3,3	3,5	3,9	4,2	4,6	4,9	5,3
X	10,8	10,7	9,7	9,7	9,8	9,1	9,1	8,7	9,1	8,7

11)

Y	0,4	0,7	1,1	1,5	1,9	2,2	2,6	2,8	3,1	3,5
X	6,5	7,9	9,1	11,3	14,2	15,9	19,7	21,6	25,6	31,0

12)

Y	2,2	2,6	3	3,3	3,6	3,9	4,1	4,4	4,8	5,1
X	12,7	17,6	22,5	27,2	32,5	37,4	41,4	47,4	56,7	63,5

13)

Y	0,8	1,3	1,9	2,5	3	3,4	3,8	4,4	4,9	5,3
X	5,4	6,6	6,9	7,7	8,4	8,2	8,7	9,4	9,2	9,6

14)

Y	1,9	2,2	2,6	2,8	3,1	3,3	3,6	3,9	4,2	4,7
X	6,9	7,7	7,7	8,3	9,0	8,9	9,6	10,5	10,6	11,9

15)

Y	0,5	0,8	1,3	1,7	2,1	2,6	3,2	3,7	4	4,4
X	3,8	5,2	6,0	6,8	7,7	7,7	8,4	9,1	8,8	9,2

16)

Y	2,0	2,3	2,7	3,0	3,2	3,6	3,9	4,3	4,6	5,0
X	11,0	10,8	9,8	9,7	9,8	9,1	9,1	9,2	8,6	8,7

17)

Y	0,2	0,5	0,9	1,3	1,7	2	2,4	2,6	2,9	3,3
X	5,1	6,2	7,1	8,9	11,2	12,5	15,5	17,4	19,7	24,4

18)

Y	1,9	2,3	2,7	3	3,3	3,6	3,8	4,1	4,5	4,8
X	7,2	10,4	13,3	16,4	19,9	22,9	25,6	30,0	35,3	40,3

I I. По таблице статистических данных:

а) Построить эмпирическое уравнение множественной линейной регрессии;

б) С помощью коэффициента детерминации оценить качество этого уравнения;

в) С помощью частных коэффициентов корреляции выяснить наличие мультиколлинеарности и в случае необходимости устранить ее;

г) Построить скорректированное эмпирическое уравнение регрессии.

1)

Y	8,0	9,7	8,0	7,7	8,8	8,2	9,7	9,1	6,6	4,2	7,1	6,6
X_1	3,1	3,5	4	4,2	4,9	5,3	6,1	6,4	6,8	7,1	7,6	7,9
X_2	4,3	5,1	4,7	4,4	5,3	5	6,2	5,8	4,9	3,7	5,5	5,1

2)

Y	8,4	10,0	8,5	8,3	9,4	8,9	10,3	9,9	7,7	5,6	8,2	7,8
X_1	2,9	3,3	3,8	4	4,7	5,1	5,9	6,2	6,6	6,9	7,4	7,7
X_2	4,5	5,3	4,9	4,6	5,5	5,2	6,4	6	5,1	3,9	5,7	5,3

3)

Y	8,6	10,1	8,7	8,6	9,6	9,2	10,6	10,2	8,2	6,2	8,7	8,4
X_1	2,8	3,2	3,7	3,9	4,6	5	5,8	6,1	6,5	6,8	7,3	7,6
X_2	4,6	5,4	5	4,7	5,6	5,3	6,5	6,1	5,2	4	5,8	5,4

4)

Y	8,7	10,2	8,9	8,8	9,8	9,5	10,8	10,5	8,6	6,8	9,2	8,9
X_1	2,7	3,1	3,6	3,8	4,5	4,9	5,7	6	6,4	6,7	7,2	7,5
X_2	4,7	5,5	5,1	4,8	5,7	5,4	6,6	6,2	5,3	4,1	5,9	5,5

5)

Y	8,8	10,3	9,0	9,0	10,0	9,7	11,0	10,8	9,0	7,3	9,6	9,4
X_1	2,6	3	3,5	3,7	4,4	4,8	5,6	5,9	6,3	6,6	7,1	7,4
X_2	4,8	5,6	5,2	4,9	5,8	5,5	6,7	6,3	5,4	4,2	6	5,6

6)

Y	8,8	10,3	9,1	9,1	10,1	9,9	11,2	11,0	9,3	7,8	10,0	9,8
X_1	2,5	2,9	3,4	3,6	4,3	4,7	5,5	5,8	6,2	6,5	7	7,3
X_2	4,9	5,7	5,3	5	5,9	5,6	6,8	6,4	5,5	4,3	6,1	5,7

7)

Y	8,8	10,2	9,2	9,2	10,2	10,1	11,3	11,2	9,6	8,3	10,3	10,2
X_1	2,4	2,8	3,3	3,5	4,2	4,6	5,4	5,7	6,1	6,4	6,9	7,2
X_2	5	5,8	5,4	5,1	6	5,7	6,9	6,5	5,6	4,4	6,2	5,8

8)

Y	8,8	10,1	9,2	9,3	10,2	10,2	11,3	11,3	9,9	8,7	10,6	10,6
X_1	2,3	2,7	3,2	3,4	4,1	4,5	5,3	5,6	6	6,3	6,8	7,1
X_2	5,1	5,9	5,5	5,2	6,1	5,8	7	6,6	5,7	4,5	6,3	5,9

9)

Y	8,7	10,0	9,1	9,3	10,2	10,3	11,4	11,4	10,1	9,1	10,8	10,9
X_1	2,2	2,6	3,1	3,3	4	4,4	5,2	5,5	5,9	6,2	6,7	7
X_2	5,2	6	5,6	5,3	6,2	5,9	7,1	6,7	5,8	4,6	6,4	6

10)

Y	8,6	9,9	9,1	9,3	10,2	10,3	11,4	11,5	10,3	9,4	11,0	11,2
X_1	2,1	2,5	3	3,2	3,9	4,3	5,1	5,4	5,8	6,1	6,6	6,9
X_2	5,3	6,1	5,7	5,4	6,3	6	7,2	6,8	5,9	4,7	6,5	6,1

11)

Y	8,4	9,7	9,0	9,2	10,1	10,3	11,3	11,5	10,5	9,7	11,2	11,4
X_1	2	2,4	2,9	3,1	3,8	4,2	5	5,3	5,7	6	6,5	6,8
X_2	5,4	6,2	5,8	5,5	6,4	6,1	7,3	6,9	6	4,8	6,6	6,2

12)

Y	8,2	9,4	8,8	9,1	10,0	10,2	11,2	11,5	10,6	10,0	11,3	11,6
X_1	1,9	2,3	2,8	3	3,7	4,1	4,9	5,2	5,6	5,9	6,4	6,7
X_2	5,5	6,3	5,9	5,6	6,5	6,2	7,4	7	6,1	4,9	6,7	6,3

13)

Y	8,0	9,1	8,6	9,0	9,8	10,1	11,1	11,4	10,7	10,2	11,4	11,8
X_1	1,8	2,2	2,7	2,9	3,6	4	4,8	5,1	5,5	5,8	6,3	6,6
X_2	5,6	6,4	6	5,7	6,6	6,3	7,5	7,1	6,2	5	6,8	6,4

14)

Y	7,7	8,8	8,4	8,8	9,6	10,0	10,9	11,3	10,7	10,4	11,5	11,9
X_1	1,7	2,1	2,6	2,8	3,5	3,9	4,7	5	5,4	5,7	6,2	6,5
X_2	5,7	6,5	6,1	5,8	6,7	6,4	7,6	7,2	6,3	5,1	6,9	6,5

15)

Y	7,4	8,5	8,1	8,6	9,4	9,8	10,7	11,2	10,7	10,5	11,5	12,0
X_1	1,6	2	2,5	2,7	3,4	3,8	4,6	4,9	5,3	5,6	6,1	6,4
X_2	5,8	6,6	6,2	5,9	6,8	6,5	7,7	7,3	6,4	5,2	7	6,6

16)

Y	7,0	8,1	7,8	8,3	9,1	9,6	10,5	11,0	10,6	10,6	11,5	12,0
X_1	1,5	1,9	2,4	2,6	3,3	3,7	4,5	4,8	5,2	5,5	6	6,3
X_2	5,9	6,7	6,3	6	6,9	6,6	7,8	7,4	6,5	5,3	7,1	6,7

17)

Y	6,6	7,6	7,5	8,0	8,8	9,4	10,2	10,8	10,5	10,7	11,4	12,0
X_1	1,4	1,8	2,3	2,5	3,2	3,6	4,4	4,7	5,1	5,4	5,9	6,2
X_2	6	6,8	6,4	6,1	7	6,7	7,9	7,5	6,6	5,4	7,2	6,8

18)

Y	6,2	7,1	7,1	7,7	8,4	9,1	9,8	10,5	10,4	10,7	11,3	12,0
X_1	1,3	1,7	2,2	2,4	3,1	3,5	4,3	4,6	5	5,3	5,8	6,1
X_2	6,1	6,9	6,5	6,2	7,1	6,8	8	7,6	6,7	5,5	7,3	6,9

ВСПОМОГАТЕЛЬНЫЙ РАЗДЕЛ

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Тема 1. Основные понятия регрессионного анализа

Функциональная и статистическая зависимость экономических переменных. Корреляционное поле. Парная и множественная регрессии. Этапы построения уравнения регрессии.

Тема 2. Метод наименьших квадратов

Общая идея метода наименьших квадратов. Применение метода наименьших квадратов при нахождении оценок параметров функций различных видов (линейной, квадратичной, показательной, логарифмической).

Тема 3. Парная линейная регрессия

Степень линейной зависимости между экономическими переменными. Коэффициент корреляции. Уравнение парной линейной регрессии. Проверка качества уравнения парной линейной регрессии.

Тема 4. Множественная линейная регрессия

Расчет коэффициентов множественной линейной регрессии. Интервальные оценки коэффициентов теоретического уравнения множественной линейной регрессии. Анализ качества эмпирического уравнения множественной линейной регрессии. Проверка статистической значимости коэффициентов уравнения регрессии. Проверка выполнимости предпосылок метода наименьших квадратов. Статистика Дарбина-Уотсона.

Тема 5. Нелинейная регрессия

Логарифмические модели. Полулогарифмические модели. Обратная модель. Степенная модель. Показательная модель. Преобразование случайного отклонения. Выбор формы модели. Проблемы спецификации.

Тема 6. Гетероскедастичность

Суть гетероскедастичности. Последствия гетероскедастичности. Обнаружение гетероскедастичности. Методы смягчения проблемы гетероскедастичности.

Тема 7. Автокорреляция

Суть и причины автокорреляции. Последствия автокорреляции. Обнаружение автокорреляции. Методы устранения автокорреляции.

Тема 8. Мультиколлинеарность

Суть мультиколлинеарности. Последствия мультиколлинеарности. Определение мультиколлинеарности. Методы устранения мультиколлинеарности.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА ДИСЦИПЛИНЫ
«Прикладная эконометрика в бизнесе»

Номер раздела, темы	Название раздела, темы, занятия; перечень изучаемых вопросов	Количество аудиторных часов				Управляемая самостоятельная работа	Иное	Формы контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия			
1	2	3	4	5	6	7	8	9
1	Основные понятия регрессионного анализа.	2	2					Решение задач у доски.
2	Метод наименьших квадратов.	5	4					Решение задач у доски.
3	Парная линейная регрессия.	5	4		1			Решение задач у доски.
4	Множественная линейная регрессия.	6	4		1			Решение задач у доски.
5	Нелинейная регрессия.	6	4		2			Решение задач у доски.
6	Гетероскедастичность.	4	2		2			Решение задач у доски.
7	Автокорреляция.	4	2		1			Решение задач у доски.
8	Мультиколлинеарность.	4	2		1			Решение задач у доски.
	ВСЕГО	**36**	**24**		**8**			**Зачет**

ЛИТЕРАТУРА

1. Бородич, С.А. Эконометрика. : Учеб. пособие / С.А. Бородич.– 2-е изд., испр. –Минск: Новое знание, 2004. – 416 с. – (Экономическое образование).
2. Доугерти, К. Введение в эконометрику. – М, 1997. – 402 с.
3. Замков О.О. и др. Математические методы в экономике. – М.: ДИС, 1997. – 248 с.
4. Булдык, Г.М. Статистическое моделирование и прогнозирование: Учеб./ Г.М. Булдык. – Мн.: НО ООО «БИП-С», 2003 г. – 150 с.
5. Эконометрика: Учеб. / под ред. И.И. Елисеевой. 2-е изд., перераб. и доп. М.: Финансы и статистика, 2005 г. – 360 с.