

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УО Белорусский государственный университет

Л.Ф.Дежурко

ЭКОНОМЕТРИКА

Учебно-методическое пособие  
Для студентов всех специальностей

Минск 2009

## Содержание

Содержание курса.....	3
Тема 1. Основные понятия эконометрики.....	5
Тема 2. Парная линейная регрессия.....	7
Тема 3. Нелинейная регрессия.....	21
Тема 4. Множественная регрессия.....	27
Тема 5. Временные ряды.....	32
Тема 6 Эконометрический анализ при нарушении предпосылок метода наименьших квадратов.....	35
Литература.....	41

## Содержание курса

### Тема 1. Основные понятия эконометрики

1. Определение эконометрики и ее задачи.
2. Типы данных.
3. Терминология
4. Классификация экономических моделей.
5. Этапы экономического моделирования.
6. Виды зависимостей.

### Тема 2. Парная линейная регрессия

1. Истинное и выборочное уравнения регрессии.
2. Метод наименьших квадратов.
3. Геометрическая интерпретация метода наименьших квадратов.
4. Экономическая интерпретация коэффициентов парной линейной регрессии.
5. Основные предпосылки регрессионного анализа. Теорема Гаусса-Маркова.
6. Расчет стандартных ошибок коэффициентов регрессии.
7. Проверка значимости коэффициентов регрессии.
8. Построение доверительных интервалов для параметров теоретической регрессии.
9. Проверка общего качества уровня регрессии. Коэффициент детерминации.
10. Проверка значимости коэффициента детерминации.
11. Оценка тесноты связи между переменными. Коэффициент корреляции.
12. Проверка значимости коэффициента корреляции.
13. Прогнозирование.

### Тема 3. Нелинейная регрессия

1. Регрессии, нелинейные по переменным.
2. Регрессии, нелинейные по параметрам.
3. Индекс корреляции и индекс детерминации.
4. Эластичность функции.

### Тема 4. Множественная регрессия

1. Оценка параметров линейной модели множественной регрессии.
2. Оценка качества множественной линейной регрессии.
3. Анализ и прогнозирование на основе многофакторных моделей.

### Тема 5. Временные ряды

1. Основные понятия временных рядов.
2. Основная тенденция развития – тренд.

3. Построение аддитивной модели.

Тема 6 Эконометрический анализ при нарушении предпосылок метода наименьших квадратов

1. Понятие мультиколлинеарности.
2. Понятие автокорреляции.
3. Понятие гетероскедастичности.

**Тема1:**  
**Основные понятия эконометрики.**

*Вопросы:*

7. Определение эконометрики и ее задачи.
8. Типы данных.
9. Терминология
10. Классификация экономических моделей.
11. Этапы экономического моделирования.
12. Виды зависимостей.

**1.**

**Эконометрика** – это наука, изучающая количественные закономерности и взаимосвязи в экономике.

Она зародилась и получила свое развитие на основе слияния экономической теории, математической экономики, экономической и математической статистики. В современной эконометрике широко используются информатика, статистические пакеты прикладных программ.

**Объект** – экономика, различные экономические явления и взаимосвязи.

**Предмет** – их количественные характеристики.

Задачи: 1. построение эконометрических моделей и оценивание их параметров.

2. проверка гипотез, о свойствах показателей и формах их связей.

**Эконометрический анализ** - основа для экономического анализа и прогнозирования.

**2.**

Эконометрика базируется на реальных экономических данных.

2 типа данных:

1. **пространственные данные** – данные о каком-либо экономическом показателе, полученные от однотипных объектов и относящиеся к одному моменту (периоду времени). Модели, построенные по пространственным данным, называются **пространственными моделями**.

2. **временные ряды** – данные об экономическом показателе, характеризующем какой-либо объект в различные моменты времени. Модели, построенные на временных рядах, называются **моделями временных рядов**.

**3.**

Исследуемый экономический показатель называют **результативным, объясняемым, зависимым экономическим показателем**. Соответствующую переменную – **объясняемой или зависимой**. Экономические показатели, воздействие которых на исследуемый экономический показатель изучается,

называют *факторами, объясняющими или независимыми показателями (переменными)*.

#### 4.

В эконометрике выделяют следующие основные 3 класса моделей:

1. Модели временных рядов:

1. *Модели тренда* (описывают устойчивые изменения экономического показателя в течение длительного времени).

2. *Модели сезонности* (описывают устойчивые внутригодовые колебания).

3. *Модели авторегрессии* (в них описываются влияния значения объясняемого экономического показателя в прошедший момент времени на его значение в текущий момент времени).

2. Регрессионные модели с одним уравнением. В них объясняемый экономический показатель представляется в виде функции от объясняющих экономических показателей (факторов). В зависимости от вида функции эти модели бывают: линейные и нелинейные.

3. *Системы одновременных уравнений* – это системы регрессионных уравнений, в которых в качестве объясняющих переменных используются объясняемые переменные из других уравнений системы.

#### 5.

1 этап: постановочный. Формулируется цель исследования. Целью может служить анализ возможного развития экономического явления, прогноз экономических показателей, выработка на этой основе управленческих решений).

2 этап: априорный. Проводится анализ связей экономических переменных, выделяются зависимые и независимые переменные.

3 этап: информационный. Осуществляется сбор необходимой статистической информации о значениях экономических переменных.

4 этап: спецификация моделей. Для описания выявленных между экономическими показателями связей, подбирается математическая функция.

5 этап: параметризация. На основе собранных статистических данных об экономических переменных оцениваются параметры (коэффициенты) математических функций.

6 этап: верификация. Проводится проверка адекватности модели, т.е. насколько построенная модель соответствует реальному экономическому явлению.

#### 6.

Все зависимости между экономическими переменными можно разделить на 2 вида:

1. *Функциональные*. Если каждому значению независимой переменной или нескольким независимых переменных соответствует одно строго определенное значение зависимой переменной, то такая

зависимость называется функциональной. В ней отсутствует воздействие случайных факторов, поэтому в экономике функциональная зависимость встречается редко.

2. **Статистические.** В экономике каждому значению независимых переменных может соответствовать несколько значений зависимой переменной в зависимости от воздействия неучтенных и случайных факторов. Например, пусть исследуется зависимость прибыли предприятия от объема производства и цены за единицу продукции. При одном и том же объеме производства и цене за единицу продукции прибыль предприятия может быть различна, т.к. на нее воздействуют множество других факторов, в том числе случайных.

Зависимость между переменными, на которую накладывается воздействие случайных факторов, называется статистической. Для нее характерно то, что изменение независимой переменной приводит к изменению математического ожидания зависимой переменной. **Уравнение регрессии** – математическая формула, описывающая статистическую зависимость между переменными. Если формула описывается линейной функцией, то регрессия называется линейной. Если нелинейной функцией – нелинейной регрессией. Если регрессия связывает одну зависимую и одну независимую переменную, то такая регрессия называется парной (простой). Если рассматривается зависимость экономической переменной от нескольких экономических переменных, то такая регрессия называется множественной.

## Тема 2: Парная линейная регрессия

Вопросы:

14. Истинное и выборочное уравнения регрессии.
15. Метод наименьших квадратов.
16. Геометрическая интерпретация метода наименьших квадратов.
17. Экономическая интерпретация коэффициентов парной линейной регрессии.
18. Основные предпосылки регрессионного анализа. Теорема Гаусса-Маркова.
19. Расчет стандартных ошибок коэффициентов регрессии.
20. Проверка значимости коэффициентов регрессии.
21. Построение доверительных интервалов для параметров теоретической регрессии.
22. Проверка общего качества уровня регрессии. Коэффициент детерминации.
23. Проверка значимости коэффициента детерминации.
24. Оценка тесноты связи между переменными. Коэффициент корреляции.
25. Проверка значимости коэффициента корреляции.
26. Прогнозирование.

### 1.

Пусть исследуется статистическая зависимость экономического показателя  $Y$  (объясняемая зависимая переменная) от экономического показателя  $X$  (фактора, объясняющей или независимой переменной). Предположим, что зависимость носит линейный характер, тогда ее можно описать уравнением.

$$Y = \alpha + \beta X + E \quad (1),$$

где  $X$  – неслучайная величина,  $Y$  и  $E$  – случайные величины.

Случайная величина  $E$  отражает воздействие на зависимую переменную  $Y$  неучтенных и случайных факторов и называется ошибкой регрессии. Уравнение (1) называют **истинным (теоретическим) уравнением** регрессии или линейной регрессионной моделью. На основе реальных статистических данных об экономических показателях  $X$  и  $Y$  (выборке данных из генеральной совокупности) оцениваются параметры регрессии  $\alpha$  и  $\beta$  и строится выборочное уравнение регрессии

$$\hat{y} = a + vx, \quad (2)$$

$a$ ,  $v$ , - коэффициенты регрессии. Уравнение (2) называют еще эмпирическим уравнением регрессии.

Одним из методов нахождения коэффициентов регрессии  $a$  и  $v$  является **метод наименьших квадратов (МНК)**.

### 2.



Пусть из генеральной совокупности выбраны данные об экономических показателях  $Y: (y_1, y_2, \dots, y_n)$  и  $X: (x_1, x_2, \dots, x_n)$ . Если в (2) подставить наблюдаемое (выборочное значение  $x_i$ , то получим расчетное значение  $\hat{y}_i$  зависимой переменной  $y$ :

$$\hat{y}_i = a + bx_i \quad (3)$$

Разность между фактическими и расчетными значениями зависимой переменной обозначим  $e_i$  и назовем остатком, т.е.:

$$e_i = y_i - \hat{y}_i \quad (4)$$

Суть МНК заключается в следующем: коэффициенты  $a$  и  $b$  должны быть такими, чтобы сумма квадратов остатков была минимальна

$$S = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 (\min) \quad (5)$$

в (5)  $y_i$  и  $x_i$  – известные величины, а  $a$  и  $b$  – неизвестные.

Запишем необходимые условия экстремума функции  $S$  относительно  $a$  и  $b$ :

$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum (y_i - a - bx_i) = 0 \\ \frac{\partial S}{\partial b} = -2 \sum (y_i - a - bx_i)x_i = 0 \end{cases} \quad (6)$$

Система (6) является системой двух уравнений относительно двух неизвестных  $a$  и  $b$ . Она легко преобразовывается в систему (7):

$$\begin{cases} na + b \sum x_i = \sum y_i \\ a \sum x_i + b \sum x_i^2 = \sum x_i y_i \end{cases} \quad (7)$$

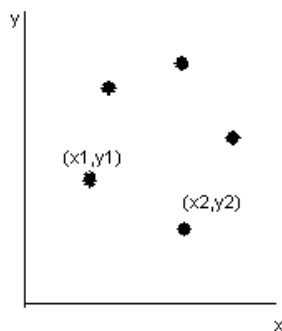
Разделим оба уравнения системы на  $n$ :

$$\begin{cases} a + b\bar{x} = \bar{y} \\ a\bar{x} + b\bar{x}^2 = \overline{xy} \\ b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - (\bar{x})^2} \\ a = \bar{y} - b\bar{x} \end{cases} \quad (8)$$

$$\bar{x} = \frac{1}{n} \sum x_i, \bar{x}^2 = \frac{1}{n} \sum x_i^2, \bar{y} = \frac{1}{n} \sum y_i, \overline{xy} = \frac{1}{n} \sum x_i y_i$$

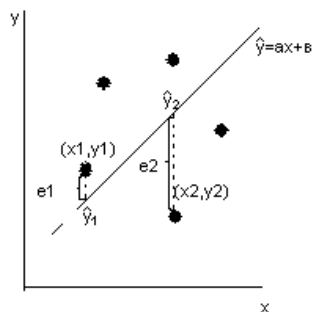
### 3.

начертим оси координат  $X, Y$  и изобразим в первой четверти точки  $(x_i, y_i)$



Полученное изображение называется диаграммой рассеяния или полем корреляции.

Проведем линию регрессии  $\hat{y} = ax + b$



Согласно МНК,  $a$  и  $b$  должны быть такими, чтобы построенная линия была ближайшей к точкам поля корреляции по их совокупности.

Сумма квадратов расстояний от точек поля корреляции до линии регрессии должна быть минимальной.

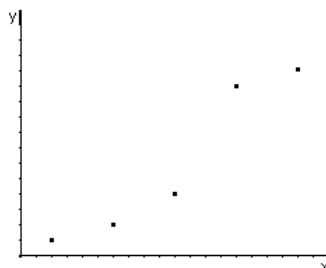
Пример 1: исследуется зависимость прибыли предприятия от затрат на приобретение нового оборудования и техники. Собраны статистические данные по пяти однотипным предприятиям. Данные в млн. ден.ед. представлены в таблице 1.

Таблица 1

№ предприятия	Затраты на новое оборудование, $x_i$	Прибыль, $y_i$
1	2	1
2	6	2
3	10	4
4	14	11
5	18	12

Построить уравнение регрессии.

Данные таблицы представим графически, т.е. построим поле корреляции:



Из полученной диаграммы рассеяния видно, что зависимость статистическая и ее можно представить линейной регрессией  $\hat{y} = ax + b$ . Для оценки коэффициентов регрессии  $a$  и  $b$  воспользуемся формулами (8), для этого построим рабочую таблицу 2.

Таблица2

№ предприятия	$\bar{x}$	$\bar{y}$	$\bar{x}^2$	$\overline{xy}$	$\bar{y}^2$
1	2	1	4	2	1
2	6	2	36	12	4
3	10	4	100	40	16
4	14	11	196	154	121
5	18	12	324	216	144
Итого:	50	30	660	424	286
Среднее	10	6	132	84,8	57,2
	$\bar{x}$	$\bar{y}$	$\bar{x}^2$	$\overline{xy}$	$\bar{y}^2$

Подставим результаты, полученные в таблице 2 в формулы (8): испр.

$$b = \frac{\overline{xy} - \bar{x}\bar{y}}{\bar{x}^2 - \bar{x}^2} = \frac{84,6 - 10 \cdot 6}{132 - 10^2} = 0,775$$

$$a = \bar{y} - b\bar{x} = 6 - 0,775 \cdot 10 = -1,75$$

Таким образом, уравнение регрессии, описывающее зависимость прибыли предприятия от затрат на новое оборудование и технику имеет вид:

$$\hat{y} = -1,75 + 0,775x$$

Выбрав с помощью диаграммы рассеяния для описания зависимости линейную регрессию мы выполнили этап спецификации (подбора функции), а рассчитав коэффициенты  $a$  и  $b$ , т.е. оценив параметры теоретической регрессии, мы выполнили этап параметризации.

#### 4.

Коэффициент парной линейной регрессии  $b$  показывает, как в среднем изменяется зависимый экономический показатель  $y$  с изменением независимого фактора  $x$  на единицу. Так в примере 1 коэффициент  $b=0,775$  показывает, что при увеличении расходов на приобретение нового оборудования и техники на 1 ден.ед. прибыль предприятия в среднем увеличится на 0,775 ден. ед.

Коэффициент  $a$  парной линейной регрессии экономического смысла не имеет.

#### 5.

Для того, чтобы оценки параметров теоретической регрессии, полученные на основе МНК были лучшими по сравнению с оценками, найденными с помощью других методов, должны выполняться определенные условия, которые называются **основными предпосылками регрессионного анализа**.

Для того, чтобы их сформулировать, вспомним что теоретическая регрессия описывается уравнением

$$Y = \alpha + \beta x + \varepsilon,$$

или для  $i$ -го наблюдения

$$Y = \alpha + \beta x_i + \varepsilon_i$$

Предпосылки:

1. Математическое ожидание случайного члена  $\varepsilon$  в любом наблюдении должно быть равно 0:

$$M(\varepsilon_i) = 0, (i = \overline{1, n})$$

2. Дисперсия случайного члена  $\varepsilon$  должна быть постоянной для всех наблюдений:

$$D(\varepsilon_i) = \sigma^2, (i = \overline{1, n})$$

3. Случайные члены должны быть статистически независимы друг от друга:

$$M(\varepsilon_i \varepsilon_j) = 0, (i \neq j)$$

4. Объясняющая переменная  $x_i$  – неслучайная величина

Теорема Гаусса-Маркова:

*Если выполняются предпосылки 1-4 регрессионного анализа, то оценки параметров теоретической регрессии  $a$  и  $b$  есть наилучшие линейные оценки, обладающие следующими свойствами:*

1. Они являются несмещенными:

$$M(a) = \alpha, M(b) = \beta$$

2. Они являются эффективными, т.е. имеют наименьшую дисперсию в классе всех несмещенных оценок.

$$D(a) = \frac{\overline{x^2} \sigma^2}{n * \text{Var}(x)}; \tag{9}$$

$$D(b) = \frac{\sigma^2}{n * \text{Var}(x)}.$$

3. Они являются состоятельными, т.е.

$$\lim_{n \rightarrow \infty} D(a) = 0,$$

$$\lim_{n \leftarrow \infty} D(b) = 0$$

Это значит, что при достаточно большом объеме выборки  $n$ , оценки  $a$  и  $b$  вблизи к истинным параметрам линейной регрессионной модели  $\alpha$  и  $\beta$ .

## 6.

Для расчета дисперсий  $D(a)$  и  $D(b)$  коэффициентов регрессии  $a$  и  $b$  в формулах (9) использовалась дисперсия  $\sigma^2$  случайного члена  $\varepsilon$ . Эта дисперсия неизвестна, но ее можно оценить, используя выборочные данные. Можно доказать, что несмещенной оценкой дисперсии  $\sigma^2$  является величина  $S^2$ , где:

$$S^2 = \frac{1}{n-2} \sum e_i^2 = \frac{n}{n-2} \text{Var}(e), \tag{10}$$

$$e = y - \hat{y}$$

Величина  $S$  называется стандартной ошибкой регрессии. Она служит мерой разброса зависимой переменной около линии регрессии. Запишем в формулах (9) дисперсию  $\sigma^2$  ее оценкой  $S^2$ :

$$S_a^2 = \frac{\overline{x^2} S^2}{n * \text{Var}(x)}, \quad (11)$$

$$S_b^2 = \frac{S_2}{n * \text{Var}(x)}$$

$S_a^2$  и  $S_b^2$  называют оценками дисперсии коэффициентов регрессии, а величина  $S_a$  и  $S_b$  – стандартными ошибками коэффициентов регрессии. Они используются для построения доверительных интервалов, которым принадлежат параметры истинной регрессии и для проверки значимости коэффициентов регрессии.

Вернемся в Примеру 1 и рассчитаем стандартные ошибки коэффициентов регрессии:

$$n = 5, \bar{x} = 10, \overline{x^2} = 132$$

$$\text{Var}(x) = \overline{x^2} - (\bar{x})^2 = 132 - 100 = 32$$

$$\text{Var}(e) = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 = 1,98$$

$$S^2 = \frac{n}{n-2} \text{Var}(e) = \frac{5}{3} * 1,98 = 3,3$$

$$S = \sqrt{3,3} = 1,816$$

$$S_a = S \sqrt{\frac{\overline{x^2}}{n * \text{Var}(x)}} = 1,816 \sqrt{\frac{132}{5 * 32}} = 1,65$$

$$S_b = S \sqrt{\frac{1}{n * \text{Var}(x)}} = 1,816 \sqrt{\frac{1}{5 * 32}} = 0,143$$

## 7.

Коэффициента регрессии получены на основании выборочных данных, отобранных случайным образом. Следовательно, коэффициенты регрессии  $a$  и  $b$  являются случайными числами и их значение может быть лишь случайно оказались отличными от нуля. Поэтому проводят проверку значимости коэффициентов регрессии, т.е. проверку того, значимо ли они отличны от нуля. Для этого используют процедуру проверки гипотез. Проверим значимость коэффициента  $b$ . Для этого:

1. Сформулируем гипотезу  $H_0$ :

$$H_0 : \beta = 0.$$

Она состоит в том, что истинный коэффициент  $\beta=0$ ,

2. В качестве критерия проверки гипотезы принимают случайную величину  $t$ :

$$t_b = \frac{b}{S_b}. \quad (12)$$

Эта случайная величина имеет распределение Стьюдента с  $v = n-2$  степенями свободы. Подставим в формулу (12) оцененное по выборке значение коэффициента  $b$  и его стандартную ошибку  $S_b$ , получим наблюдаемое или расчетное значение  $t$ -критерия  $t_{\text{расч.}}$ .

3. Выбирают уровень значимости проверки гипотезы. Как правило  $\alpha = 0,05$  или  $\alpha = 0,01$ , т.е. пятипроцентный или однопроцентный уровень значимости.

4. По таблице распределения Стьюдента для выборочного уровня значимости  $\alpha/2$  и  $v = n-2$  находят  $t_{\text{кр.}}$  (критическое).

5. Если  $|t_{\text{расч.}}| > t_{\text{кр.}}$ , то гипотеза  $H_0$  о равенстве параметра  $\beta = 0$  отвергается, параметр  $\beta$  существенно отличен от нуля, коэффициент  $b$  значим, а переменная  $x$  оказывает существенное влияние на зависимую  $y$  ( $H_0$  считается неверной с вероятностью  $1 - \alpha$ )

6. Если  $|t_{\text{расч.}}| < t_{\text{кр.}}$ , гипотеза  $H_0$  принимается, коэффициент  $b$  незначим и переменная  $x$  не оказывает существенного влияния на зависимую переменную  $y$ .

Замечание: аналогично проверяется значимость коэффициента  $a$  в уравнении регрессии, однако проверка значимости коэффициента  $b$  имеет гораздо большее значение в регрессионном анализе.

Вернемся в примере 1 и проверим значимость коэффициента  $b$ . Зависимость прибыли предприятия от расходов на новое оборудование и технику описывается регрессией:

$$\hat{y} = -1,75 + 0,775x$$

$$(1,65) \quad (0,143).$$

1. Формулируем гипотезу  $H_0$ , состоящую в том, что истинный коэффициент  $\beta = 0$ ,  $H_0 : \beta = 0$ .

2. Определим  $t_{\text{расч.}}$ .

$$t_{\text{расч.}} = \frac{0,775}{0,143} = 5,4.$$

3. Выбираем уровень значимости проверки гипотезы  $\alpha = 0,05$ .

4. По таблице распределения Стьюдента для  $\alpha/2 = 0,025$  и числа степеней свободы

$$v = 5 - 2 = 3$$

определим  $t_{\text{кр.}} = 3,182$ .

5.  $|t_{\text{расч.}}| = 5,4 > t_{\text{кр.}} = 3,182$ , поэтому гипотеза  $H_0$  не верна с вероятностью  $1 - \alpha = 1 - 0,05 = 0,95$ , параметр  $\beta$  существенно отличен от нуля, коэффициент  $b$  значим и затраты на новое оборудование и технику оказывают существенное влияние на прибыль предприятия.

## 8.

Вспомним, что линейная регрессионная модель (истинная или теоретическая регрессия) имеет вид:

$$Y = \alpha + \beta x + \varepsilon \quad (13)$$

На основании выборки строится выборочное уравнение регрессии:

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

Также на основании выборки рассчитываются стандартные ошибки регрессии  $S_a$  и  $S_b$ .

Можно доказать, что с вероятностью  $1-\alpha$  ( $\alpha$  – выбранный уровень значимости) значения параметра  $\beta$  лежат внутри интервала:

$$b - t_{кр.} * S_b < \beta < b + t_{кр.} * S_b \quad (14)$$

и с вероятностью  $1-\alpha$  ( $\alpha$  – выбранный уровень значимости) значение параметра  $\alpha$  истинной регрессии лежит внутри интервала:

$$a - t_{кр.} * S_a < \alpha < a + t_{кр.} * S_a \quad (15)$$

Вернемся к Примеру 1 и построим доверительный интервал для параметра  $\beta$  в регрессионной модели, описывающей зависимость прибыли предприятия от затрат на новое оборудование и технику. Выберем уровень значимости  $\alpha = 0,05$ . т.к. в данном примере  $v = 5-2=3$ , то  $t_{кр.} = 3,182$ ,  $b = 0,775$ ,  $S_b = 0,143$ .

Тогда с вероятностью  $1-\alpha = 1-0,05 = 0,95$  параметр  $\beta$  истинной регрессии попадает в интервал

$$0,775 - 3,18 * 0,143 < b < 0,775 + 3,18 * 0,143$$

или  $0,32 < b < 1,23$  с вероятностью 95%.

## 9.

Выборочное уравнение регрессии имеет вид:

$$y = a + \beta x + e,$$

$$e = y - \hat{y}, \text{ или}$$

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

тогда

$$y = \hat{y} + e$$

Рассчитаем выборочную дисперсию (вариацию)  $Var(y)$ :

$$Var(y) = Var(\hat{y} + e) = Var(\hat{y}) + Var(e) + 2Cov(\hat{y}, e).$$

Из основных предпосылок регрессионного анализа следует, что  $Cov(\hat{y}, e) = 0$ , следовательно

$$Var(y) = Var(\hat{y}) + Var(e)$$

т.е. дисперсия зависимой переменной  $y$  ( $Var(y)$ ) распадается на 2 части:

$Var(\hat{y})$  - часть, объясняемая уравнением регрессии, и часть  $Var(e)$  - необъяснимая часть, зависящая от неученых и случайных факторов.

Коэффициентом детерминации называют отношение  $R^2$ :

$$R^2 = \frac{Var(\hat{y})}{Var(y)} = 1 - \frac{Var(e)}{Var(y)}, \quad (16)$$

которое характеризует долю вариации зависимой переменной, объясненную уравнением регрессии. Из (16) следует что  $R^2$  меняется от 0 до 1:

$$0 \leq R^2 \leq 1,$$

чем ближе  $R^2$  к единице, тем меньше  $Var(e)$ , т.е. доля вариации зависимой переменной, объясняемая случайными и неучеными факторами, тем лучше качество уравнения регрессии. Если  $Var(e)=0$ , то  $R^2=1$ , имеем функциональную зависимость. Чем ближе  $R^2$  к 0, тем больше  $Var(e)$ , т.е. больше доля вариации, объясненная случайными и неучеными факторами, тем хуже качество регрессии. Т.к.

$$\begin{aligned} Var(\hat{y}) &= \frac{\sum (\hat{y}_i - \bar{y})^2}{n}, \\ Var(y) &= \frac{\sum (y_i - \bar{y})^2}{n}, \\ R^2 &= \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \end{aligned} \quad (17)$$

Вернемся к примеру 1, можно посчитать, что:

$$\begin{aligned} Var(\hat{y}) &= 19,22 \\ Var(y) &= 21,2 \\ R^2 &= \frac{19,22}{21,2} = 0,907 \end{aligned}$$

Коэффициент детерминации близок к 1, качество регрессии хорошее.

Можно утверждать, что вариация (изменчивость) прибыли предприятия на 90,7% объясняется затратами на новое оборудование и технику и на 9,3% - прочими неучтенными и случайными факторами.

## 10.

Т.к.  $R^2$  оценивается на основании выборочных данных, то его отличие от 0 может оказаться случайным. Поэтому проводят проверку его значимости:

1. Формулируется гипотеза  $H^0: R^2=0$ , состоящая в том, что истинный коэффициент детерминации равен 0.

2. В качестве критерия проверки гипотезы применяют случайную величину F:

$$F = \frac{R^2(n-2)}{1-R^2}. \quad (18)$$

Величина F имеет распределение Фишера с двумя степенями свободы  $v_1=1, v_2=n-2$ .

3. Выберем уровень значимости проверки гипотезы значимости:

$$\alpha = 0,05 \vee 0,01.$$

4. На основании  $\alpha, v_1, v_2$  в таблице распределения Фишера выбираем  $F_{кр.}$  (критическое)

5. Сравниваем  $F_{расч}$  и  $F_{кр.}$ . если  $F_{расч} > F_{кр.}$ , то с вероятностью  $1-\alpha$  гипотезу  $H_0$  считаем неверной, т.е. истинный коэффициент детерминации существенно отличен от нуля, уравнение регрессии значимо и переменные, включенные в уравнение регрессии достаточно объясняют



поведение зависимой переменной. Если  $F_{расч} < F_{кр.}$ , то принимаемая гипотеза  $H_0$ , уравнение регрессии считается незначимым.

Проверим значимость коэффициента детерминации в примере 1:

1. Формулируем гипотезу  $H_0: R^2=0$ .
2. Находим  $F_{расч.}$ . В (18) подставим значение коэффициента детерминации, оцененное по выборке:

$$F_{расч.} = \frac{0,907 * 3}{1 - 0,907} = 29,2 .$$

3. Выбираем уровень значимости  $\alpha=0,005$ .
4. В таблице распределения Фишера на основании  $\alpha=0,05$  и для степеней свободы  $v_1=1, v_2=5-2=3$  найдем  $F_{кр.}$

$$F_{кр.} = 10,13 .$$

6.  $F_{расч} = 29,2 > F_{кр.} = 10,13$ , поэтому  $H_0$  не верна в вероятностью  $1 - 0,05 = 0,95$ , коэффициент детерминации значим, значимо построенное в Примере 1 уравнение регрессии.

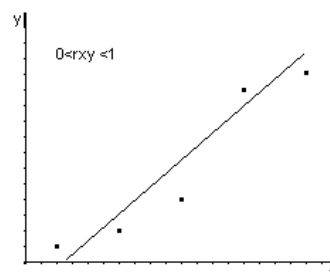
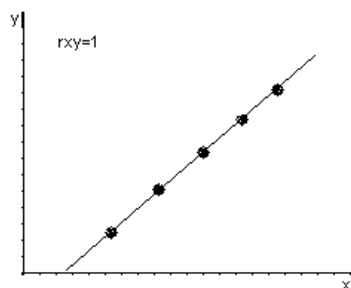
## 11.

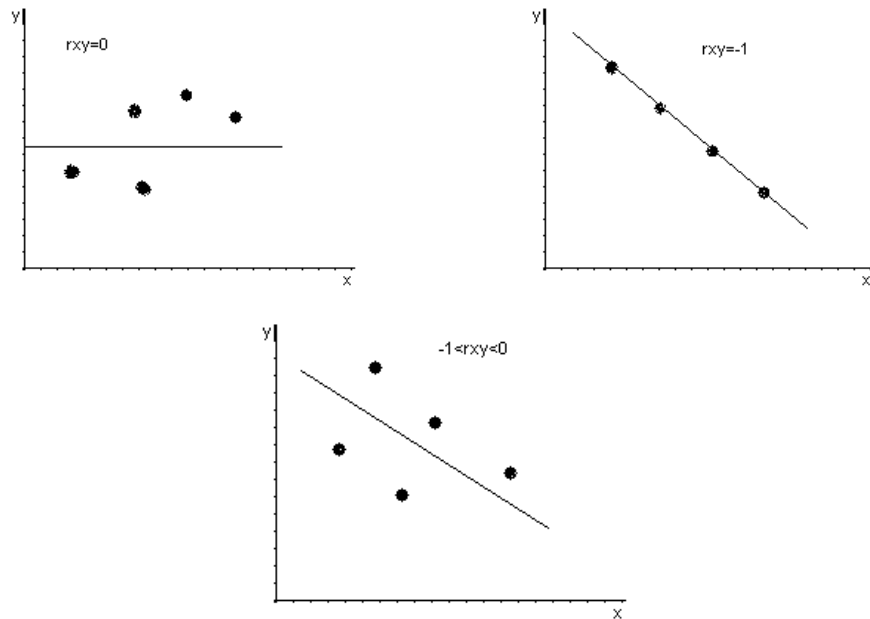
Уравнение регрессии всегда дополняется показателем тесноты связи между переменными. При использовании линейной регрессии в качестве такого показателя выступает линейный коэффициент корреляции:

$$r_{xy} = \frac{Cov(x, y)}{\sigma_x * \sigma_y} = \frac{Cov(x, y)}{\sqrt{Var(x) * Var(y)}} , \quad (19)$$

$$-1 \leq r_{xy} \leq 1,$$

$r_{xy}$  – безразмерная величина, показывает степень линейной зависимости между переменными. Чем ближе  $r_{xy}$  к  $\pm 1$ , тем сильнее линейная зависимость. Чем ближе  $r_{xy}$  к 0, тем линейная зависимость слабее. Если  $r_{xy} = \pm 1$ , то имеет место функциональная линейная зависимость. Если  $r_{xy} = 0$ , то линейная зависимость отсутствует. Если  $r_{xy} > 0$ , то связь между переменными положительная, если  $r_{xy} < 0$  – отрицательная.





Рассчитаем коэффициент корреляции в примере 1:

$$Cov(x, y) = \overline{xy} - \bar{x} * \bar{y} = 84,8 - 10 * 6 = 24,8$$

$$Var(x) = \overline{x^2} - (\bar{x})^2 = 132 - 10^2 = 32$$

$$Var(y) = \overline{y^2} - (\bar{y})^2 = 57,2 * 6^2 = 21,2$$

$$r_{xy} = \frac{24,8}{\sqrt{32 * 21,2}} = 0,752$$

$r_{xy} > 0$  и близок к 1 следовательно линейная зависимость между прибылью предприятия и затратами на новое оборудование – положительная и тесная.

## 12.

Осуществляется аналогично проверки значимости коэффициентов регрессии и детерминации, используется t-статистика:

$$t = \frac{r_{xy} \sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \quad (20)$$

Проведем проверку значимости коэффициента корреляции в примере 1:

1. Формулируем гипотезу, состоящую в том, что истинный коэффициент корреляции равен нулю:

$$H_0 : \rho_{xy} = 0.$$

2. Подставим значение коэффициента корреляции, вычисленное по выборке в (20):

$$t_{расч.} = \frac{0,952\sqrt{3}}{\sqrt{1-(0,952)^2}} = 5,5$$

3. Выбираем уровень значимости  $\alpha=0,05$ .
4. Для  $\alpha/2=0,025$  и для  $\nu=n-2=3$  в таблице распределения Стьюдента находим  $t_{кр.}$ :

$$t_{кр.} = 3,18,$$

$$|t_{расч.}| = 5,5 > t_{кр.} = 3,18.$$

Следовательно, истинный коэффициент корреляции существенно отличается от 0, линейная зависимость между прибылью предприятия и затратами на новое оборудование и технику действительно тесная..

Замечание 1:

В парном линейном регрессионном анализе проверка значимости коэффициента  $b$ , коэффициента корреляции и коэффициента детерминации являются эквивалентными.

Замечание 2:

Легко показать, что коэффициент детерминации равен квадрату коэффициента корреляции,  $r_{xy}^2 = R^2$ ,

### 13.

Прогнозирование на основе эконометрических моделей является одной из основных задач эконометрики.

Под прогнозированием в эконометрике понимают построение оценки зависимой переменной для таких значений независимых переменных, которых нет в исходных наблюдениях.

Различают точечное прогнозирование и интервальное.

Точечный прогноз это число, значение зависимой переменной, вычисляемое для заданных значений независимых переменных.

Интервальный прогноз это интервал, в котором с заданным уровнем значимости (с заданной вероятностью) находится истинное значение зависимой переменной для заданных значений независимых переменных.

Рассмотрим парную линейную регрессионную модель  $Y = \alpha + \beta x + \varepsilon$  и соответствующее выборочное уравнение регрессии  $\hat{y} = a + bx$ . Обозначим через  $y_p$  истинное значение переменной  $y$  для заданного значения независимой переменной  $x_p$ , т.е.  $y_p = \alpha + \beta x_p + \varepsilon$ .

Точечным прогнозом для  $y_p$  является  $\hat{y}_p = a + bx_p$ , т.е. чтобы получить точечный прогноз нужно в построенное уравнение регрессии подставить заданное значение независимой переменной.

**Ошибкой предсказания** ( $\Delta_p$ ) называют разность между прогнозным и истинным значениями независимой переменной.

$$\Delta_p = \hat{y}_p - y_p$$

Можно доказать, что дисперсия ошибки предсказания

$$D(\Delta_p) = \left(1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n * \text{Var}(x)}\right) * \sigma^2. \quad (21)$$

Из (21) следует, что чем ближе заданное значение независимой переменной  $x_p$  к  $\bar{x}$  тем меньше дисперсия прогноза и чем больше объем выборки  $n$ , тем меньше дисперсия прогноза.

Заменив в (21) дисперсию  $\sigma^2$  на ее оценку  $S^2$ , извлечем, квадратный корень и получим стандартную ошибку предсказания  $S_p$ .

$$S_p = S \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n * Var(x)}} \quad (22)$$

Выберем уровень значимости  $\alpha$  и по таблице распределения Стьюдента найдем  $t_{кр}$ . Тогда с вероятностью  $1 - \alpha$  истинное значение переменной  $y_p$  будет находится внутри интервала:

$$\hat{y}_p - t_{кр} * S_p < y_p < \hat{y}_p + t_{кр} * S_p \quad (23)$$

Очевидно, что чем ближе  $x_p$  к  $\bar{x}$  и чем больше  $n$ , тем уже доверительный интервал (тем точнее прогноз). Это надо учитывать, выбирая прогнозные значения для независимой переменной.

Вернемся в Примеру 1 и найдем точечный и интервальный прогнозы для прибыли предприятия для затрат на новое оборудование и технику в размере 20 млн. денежных единиц.

$$\hat{y}_p = -1,75 + 0,775 * 20 = 13,75,$$

$$S = 1,816, n = 5, x_p = 20, \bar{x} = 10, Var(x) = 32,$$

$$S_p = 1,816 \sqrt{1 + \frac{1}{5} + \frac{(20 - 10)^2}{5 * 32}} = 2,453$$

$$\alpha = 0,05, \nu = 5 - 2 = 3, t_{кр} = 3,18$$

$$13,75 - 3,18 * 2,453 < y_p < 13,75 + 3,18 * 2,453$$

$$5,948 < y_p < 21,55$$

**Вывод:** с вероятностью 0,95 истинное значение прибыли попадет в полученный интервал.

**Тема 3:**  
**Нелинейная регрессия.**

*Вопросы:*

5. Регрессии, нелинейные по переменным.
6. Регрессии, нелинейные по параметрам.
7. Индекс корреляции и индекс детерминации.
8. Эластичность функции.

Многие экономические зависимости не являются линейными по своей сути и их моделирование линейными регрессиями не дает положительного результата. Так для описания зависимости спроса на некоторый товар от его цены наиболее целесообразно использовать логарифмическую модель. При анализе зависимостей издержек от объема выпуска наиболее обоснованной является полиномиальная модель. Широко используемая функция Кобба-Дугласа, является степенной функцией

$$Y = A * K^\alpha * L^\beta$$

Y – объем выпуска.

K-затраты капитала.

L - затраты труда.

A,  $\alpha$ ,  $\beta$  – параметры.

В современной экономике применяются также достаточно часто обратные и экспоненциальные модели. Различают регрессии нелинейные по переменным и нелинейные по параметрам.

**1.**

К регрессиям, нелинейным по переменным относят полиномы различных степеней.:

$$y = \alpha + \beta x + \gamma x^2 + \varepsilon, \quad (1)$$

$$y = \alpha + \beta x + \gamma x^2 + \theta x^3 + \varepsilon, \quad (2)$$

$$\text{равносторонняя гипербола } y = \varepsilon + \frac{\beta}{x} + \varepsilon, \quad (3)$$

$$\text{функции вида } y = \alpha + \beta \sqrt{x} + \varepsilon \quad (4)$$

Нелинейность по переменным устраняется путем замены переменной. Так в регрессии (1) сделаем замену  $x = x_1$ ,  $x^2 = x_2$  и получим двухфакторную линейную регрессию.

$$y = \alpha + \beta x_1 + \lambda x_2 + \varepsilon$$

В уравнении (3) замена переменной имеет вид:  $z = \frac{1}{x}$ , а в (4) -  $z = \sqrt{x}$ .

Применение метода МНК для оценки коэффициентов соответствующих выборочной регрессии приводит к следующим системам уравнений. Для регрессии (!):

$$\begin{cases} \sum y_i = na + b \sum x_i + c \sum x_i^2 \\ \sum y_i x_i = a \sum x_i + b \sum x_i^2 + c \sum x_i^3 \\ \sum y_i x_i^2 = a \sum x_i^2 + b \sum x_i^3 + c \sum x_i^4 \end{cases} \quad (5).$$

Для равносторонней гиперболы система уравнений имеет вид:

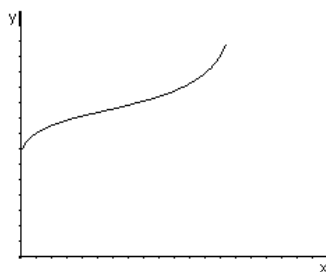
$$\begin{cases} \sum y_i = na + b \sum \frac{1}{x_i} \\ \sum \frac{y_i}{x_i} = a \sum \frac{1}{x_i} + b \sum \frac{1}{x_i^2} \end{cases} \quad (6)$$

Для уравнения (4):

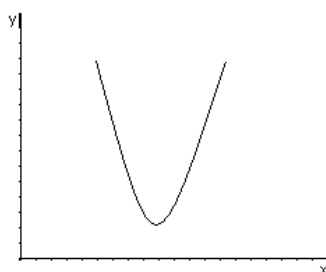
$$\begin{cases} \sum y_i = na + b \sum \sqrt{x_i} \\ \sum y_i \sqrt{x_i} = a \sum \sqrt{x_i} + b \sum x_i \end{cases} \quad (7)$$

Приведем некоторые примеры использования уравнений (1-4) в экономике:

1. Полином третьей степени уравнения (2) часто моделирует зависимость общих издержек  $Y$  от объема выпуска  $X$ . график имеет вид:



2. Полином второй степени (уравнение (1)) парабола может описать зависимость между объемом выпуска  $X$  и средними (либо предельными) издержками  $Y$



3. Гипербола (3) (обратная модель) применяется в тех случаях, когда неограниченное увеличение объясняющей переменной  $X$  асимптотически приближает зависимую переменную  $Y$  к некоторому пределу. Если  $a$  и  $b$  - оценки параметров гиперболы соответственно, то в зависимости знаков  $a$  и  $b$  возможны следующие ситуации:

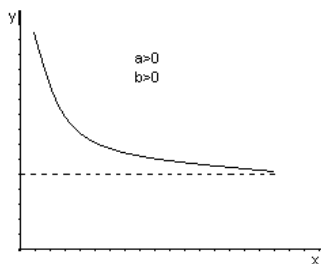


рис.1

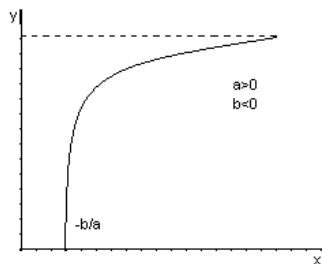


рис.2

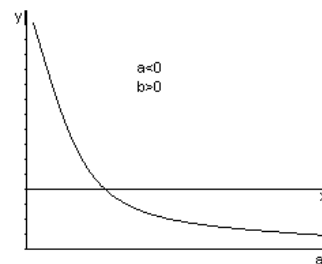


рис.3

График на рисунке 1 может отражать зависимость между объемом выпуска  $X$  и средними фиксированными издержками  $Y$ . график на рисунке 2

может описывать зависимость между доходом  $X$  и спросом на блага  $Y$ . Такие функции называются функциями Тронквиста. Важным приложением графика на рисунке 3 является кривая Филипса, отражающая зависимость между уровнем безработицы  $X$  (%) и процентным изменением заработной платы  $Y$ .

4. Уравнения с квадратными корнями (4) использовались в исследовании урожайности и трудоемкости  $c/x$  производства.

*Пример 1:*

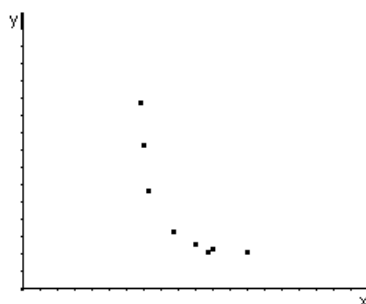
На основании информации о норме безработицы и темпах инфляции (таблица 1) построить :

1. диаграмму рассеяния.
2. уравнение регрессии, описывающее зависимость темпов инфляции от нормы безработицы.

Таблица 1

№ наблюдения, $i$	Темпы инфляции, $y_i$	Норма безработицы, $x_i$	$z_i$
1	1,1	6,5	0,154
2	1,1	5,4	0,185
3	1,2	5,5	0,182
4	1,3	5,0	0,2
5	1,7	4,4	0,227
6	2,9	3,7	0,270
7	4,2	3,5	0,286
8	5,4	3,4	0,294

Строим диаграмму рассеяния:



Из диаграммы рассеяния видно, что зависимость можно описать гиперболой  $\hat{y} = a + \frac{b}{x}$ . Сделаем замену переменных  $z = \frac{1}{x}$  и уравнение регрессии примет вид:

$$\hat{y} = a + bz,$$

$$z_i = \frac{1}{x_i}, i = \overline{1,8}$$

Обратимся в Excel к программе регрессия и введем данные  $z_i, y_i$ , получим:

$$\hat{y} = -3,9 + 27,47z$$

$$\hat{y} = -3,9 + 27,47 * \frac{1}{x}$$

*Замечание:*

Для оценки коэффициентов гиперболы можно построить систему уравнений (6) и решить ее.

## 2.

К нелинейным по параметрам регрессиям относятся:

$$\text{степенная: } y = \alpha * x^\beta * \varepsilon, \quad (8)$$

$$\text{показательная } y = \alpha * \beta^x * \varepsilon, \quad (9)$$

$$\text{экспоненциальную } y = \alpha * e^{\beta x} * \varepsilon. \quad (10)$$

Нелинейные по параметрам регрессии сводятся к линейным путем логарифмирования.

$$(8') \quad \ln y = \ln(\alpha x^\beta \varepsilon) = \ln \alpha + \beta \ln x + \ln \varepsilon,$$

$$(9') \quad \ln y = \ln(\alpha \beta^x \varepsilon) = \ln \alpha + x \ln \beta + \ln \varepsilon,$$

$$(10') \quad \ln y = \ln(\alpha * e^{\beta x} * \varepsilon) = \ln \alpha + \beta x + \ln \varepsilon$$

Для нахождения оценок соответствующих коэффициентов выборочных регрессии для (8'), (9'), (10') используется МНК при условии, что  $\ln \varepsilon$  распределен нормально.

*Пример 2:*

В таблице 2 приведены данные о расходах на питание и доходах 5 групп населения. Построить степенную регрессию, описывающую зависимость расходов на питание  $Y$  от доходов населения  $X$ .

Таблица 2

№	Доходы, $x$	Расходы, $y$	$V = \ln x$	$Z = \ln y$
1	2	1	1,69	0
2	6	2	1,79	0,69
3	10	4	2,3	1,39
4	14	11	2,6	2,4
5	18	12	2,9	2,48

Степенная регрессия имеет вид:

$$y = \alpha * x^\beta * \varepsilon;$$

$$\ln y = \ln \alpha + \beta \ln x + \ln \varepsilon;$$

$$z = \ln y;$$

$$V = \ln x; \Rightarrow$$

$$z = \ln \alpha + \beta V + \ln \varepsilon$$



Получим линейное уравнение. Обратимся к программе «Регрессия», введем данные столбцов v и z, получим:

$$\begin{aligned}\hat{z} &= -1,049 + 1,183V; \\ \ln \hat{y} &= -1,049 + 1,183 \ln x; \\ e^{\ln \hat{y}} &= e^{-1,049 + 1,183 \ln x}\end{aligned}\quad (11)$$

Выполним обратные преобразования (пропотенцируем полученное уравнение):

$$\hat{y} = e^{-1,049} * x^{1,183} = 0,35 * x^{1,183}$$

*Замечание:*

Для построения регрессии (11) можно воспользоваться формулами (8) темы 2.

### 3.

Уравнение нелинейной регрессии также как и линейной дополняются показателями корреляции и детерминации. Для оценки тесноты связи между переменными рассчитывается индекс корреляции:

$$R = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}} = \sqrt{\frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}} \quad (12)$$

Индекс корреляции (R) меняется от 0 до 1. чем ближе R к 1, тем сильнее нелинейная связь между переменными. Величина

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \quad (13)$$

используется для оценки качества уравнения регрессии. Для проверки значимости индекса детерминации используется F-статистика

$$F = \frac{R^2}{1 - R^2} * \frac{n - m - 1}{m}$$

n – объем выборки

m – число параметров при независимых переменных.

Так для параболы m=2, а для степенной функции m=1

### 4.

В экономическом анализе часто используется эластичность функции. Эластичность функции  $y = f(x)$  рассчитывается как относительное изменение y к относительному изменению x:

$$\mathcal{E} = \left(\frac{dy}{y}\right) / \left(\frac{dx}{x}\right) = \frac{x}{y} f'(x) \quad (15)$$

Эластичность показывает, насколько процентов изменяется функция  $y = f(x)$  при изменении независимой переменной на 1 %.

Для степенной функции  $y = ax^e$  эластичность представляет собой постоянную величину, равную e, действительно :

$$\begin{aligned}f'(x) &= abx^{e-1} \\ \mathcal{E} &= \frac{x}{ax^e} * aex^{e-1} = e\end{aligned}$$

В примере 2 степенная регрессия  $\hat{y} = 0,35x^{1,183}$  описывает зависимость расходов на питание от доходов. Коэффициент 0,35 экономического смысла не имеет, а коэффициент  $\epsilon$ , равный 1,183, показывает, что увеличение личного дохода на 1% приведет к увеличению расходов на питание в среднем на 1,183%.

Для остальных функций эластичность не является постоянной величиной. Так для линейной функции  $y = ax + \epsilon$  эластичность  $\mathcal{E} = \epsilon \frac{x}{y}$ , т.е. эластичность зависит от  $x$ , поэтому для остальных функций вычисляется средний показатель эластичности, в частности для линейной функции по формуле:

$$\mathcal{E} = \epsilon \frac{\bar{x}}{\bar{y}} \quad (16)$$

Рассчитаем коэффициент эластичности в примере 1 темы 2:

$$\epsilon = 0,775; \bar{x} = 10; \bar{y} = 6;$$

$$\mathcal{E} = 0,775 \frac{10}{6} = 1,29$$

Можно утверждать, что с увеличением расходов на новое оборудование на 1%, прибыль предприятия возрастет на 1,29%.

**Тема 4:**  
**Множественная регрессия.**

*Вопросы:*

4. Оценка параметров линейной модели множественной регрессии.
5. Оценка качества множественной линейной регрессии.
6. Анализ и прогнозирование на основе многофакторных моделей.

Множественная регрессия является обобщением парной регрессии. Она используется для описания зависимости между объясняемой (зависимой) переменной  $Y$  и объясняющими (независимыми) переменными  $X_1, X_2, \dots, X_k$ . Множественная регрессия может быть как линейная, так и нелинейная, но наибольшее распространение в экономике получила линейная множественная регрессия.

**1.**

Теоретическая линейная модель множественной регрессии имеет вид:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + \varepsilon \quad (1)$$

соответствующую выборочную регрессию обозначим:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_k x_k + e \quad (2)$$

Как и в парной регрессии случайный член  $\varepsilon$  должен удовлетворять основным предположениям регрессионного анализа. Тогда с помощью МНК получают наилучшие несмещенные и эффективные оценки параметров теоретической регрессии. Кроме того переменные  $X_1, X_2, \dots, X_k$  должны быть некоррелированы (линейно независимы) друг с другом. Для того, чтобы записать формулы для оценки коэффициентов регрессии (2), полученные на основе МНК, введем следующие обозначения:

$$\vec{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix};$$
$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix};$$
$$\vec{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \dots \\ \alpha_k \end{pmatrix}; \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}; \vec{e} = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}; \vec{a} = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_k \end{pmatrix}$$

Тогда можно записать в векторно-матричной форме теоретическую модель:

$$\vec{Y} = X\vec{\alpha} + \vec{\varepsilon};$$

и выборочную регрессию

$$\bar{Y} = X\bar{a} + \bar{e}.$$

МНК приводит к следующей формуле для оценки вектора  $\vec{a}$  коэффициентов выборочной регрессии:

$$\vec{a} = (X'X)^{-1}X'Y \quad (3)$$

Для оценки коэффициентов множественной линейной регрессии с двумя независимыми переменными  $\hat{y} = a_0 + a_1x_1 + a_2x_2$ , можно решить систему уравнений:

$$\begin{cases} \sum y_i = na_0 + a_1 \sum x_{i1} + a_2 \sum x_{i2}; \\ \sum x_{i1}y_i = a_0 \sum x_{i1} + a_1 \sum x_{i1}^2 + a_2 \sum x_{i1} * x_{i2}; \\ \sum x_{i2}y_i = a_0 \sum x_{i2} + a_1 \sum x_{i1} * x_{i2} + a_2 \sum x_{i2}^2 \end{cases} \quad (4)$$

Как и в парной линейной регрессии для множественной регрессии рассчитывается стандартная ошибка регрессии S:

$$S = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - k - 1}} \quad (5)$$

и стандартные ошибки коэффициентов регрессии:

$$Sa_j = S\sqrt{(X'X)^{-1}_{jj}} \quad (6)$$

значимость коэффициентов проверяется с помощью t-критерия.

$$t = \frac{a_j}{Sa_j} \quad (7)$$

имеющего распространение Стьюдента с числом степеней свободы  $\nu = n - k - 1$ .

## 2.

Для оценки качества регрессии используется коэффициент (индекс) детерминации:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}, \quad (8)$$

$$0 \leq R^2 \leq 1$$

чем ближе  $R^2$  к 1, тем выше качество регрессии.

Для проверки значимости коэффициента детерминации используется критерий Фишера или F- статистика.

$$F = \frac{R^2(n - k - 1)}{(1 - R^2)k} \quad (9)$$

с  $\nu_1 = k$ ,  $\nu_2 = n - k - 1$  степенями свободы.

В многофакторной регрессии добавление дополнительных объясняющих переменных увеличивает коэффициент детерминации. Для компенсации такого увеличения вводится скорректированный (или нормированный) коэффициент детерминации:

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1} \quad (10)$$

Если увеличение доли объясняемой регрессии при добавлении новой переменной мало, то  $\overline{R^2}$  может уменьшиться. Значит, добавлять новую переменную нецелесообразно.

*Пример 4:*

Пусть рассматривается зависимость прибыли предприятия от затрат на новое оборудование и технику и от затрат на повышение квалификации работников. Собраны статистические данные по 6 однотипным предприятиям. Данные в млн. ден. ед. приводятся в таблице 1.

Таблица 1

Номер предприятия, $i$	Прибыль $i$ -го предприятия, $y_i$	Затраты на новое оборудование $i$ -го предприятия, $x_{i1}$	Затраты на повышение квалификации на $i$ -м предприятии, $x_{i2}$
1	2	3	1
2	3	3	4
3	5	5	5
4	6	7	6
5	8	9	8
6	8	10	11

Построить двухфакторную линейную регрессию  $\hat{y} = a_0 + a_1x_1 + a_2x_2$  и оценить ее значимость. Введем обозначения:

$$\vec{y} = \begin{pmatrix} 2 \\ 3 \\ 5 \\ 6 \\ 8 \\ 8 \end{pmatrix};$$

$$X = \begin{pmatrix} 1 & 3 & 1 \\ 1 & 3 & 4 \\ 1 & 5 & 5 \\ 1 & 7 & 7 \\ 1 & 9 & 8 \\ 1 & 10 & 11 \end{pmatrix};$$

$$\vec{a} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix}$$

Транспонируем матрицу  $X$ :

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 3 & 5 & 7 & 9 & 10 \\ 1 & 4 & 5 & 7 & 8 & 11 \end{pmatrix};$$

$$XX' = \begin{pmatrix} 6 & 37 & 36 \\ 37 & 273 & 271 \\ 36 & 271 & 276 \end{pmatrix}$$

Обращение этой матрицы:

$$(XX')^{-1} = \begin{pmatrix} 1,099 & -0,263 & 0,115 \\ -0,263 & 0,208 & -0,170 \\ 0,115 & -0,170 & 0,155 \end{pmatrix}$$

$$\vec{a} = (XX')^{-1} X' \vec{Y} = \begin{pmatrix} 0,396 \\ 0,661 \\ 0,144 \end{pmatrix}$$

таким образом зависимость прибыли от затрат на новое оборудование и технику и от затрат на повышение квалификации работников можно описать следующей регрессией:

$$\hat{y} = 0,396 + 0,661x_1 + 0,144x_2$$

Используя формулу (5), где  $k=2$  рассчитаем стандартную ошибку регрессии  $S=0,636$ .

Стандартные ошибки коэффициентов регрессии рассчитаем, используя формулу (6):

$$Sa_0 = S\sqrt{1,099} = 0,636\sqrt{1,099} = 0,667$$

Аналогично:

$$Sa_1 = 0,636\sqrt{0,208} = 0,29$$

$$Sa_2 = 0,636\sqrt{0,155} = 0,25$$

Проверим значимость коэффициентов регрессии  $a_1, a_2$ . посчитаем  $t_{расч.}$

$$a_1 : t_{расч.} = \frac{a_1}{Sa_1} = \frac{0,661}{0,29} = 2,274$$

Выберем уровень значимости  $\alpha = 0,1$ , число степеней свободы

$$\nu = n - 2 - 1 = 6 - 3 = 3; \Rightarrow t_{кр.} = 1,638$$

$$|t_{расч.}| = 2,274 > t_{кр.} = 1,638$$

значит коэффициент  $a_1$  значим.

Оценим значимость коэффициента  $a_2$ :

$$a_2 : t_{расч.} = \frac{a_2}{Sa_2} = \frac{0,144}{0,25} = 0,576 < t_{кр.} = 1,638$$

Коэффициент  $a_2$  незначим.

Рассчитаем коэффициент детерминации по формуле (7)  $R^2 = 0,96$ . Прибыль предприятия на 96% зависит от затрат на новое оборудование и технику и повышение квалификации на 4% от прочих и случайных факторов. Проверим значимость коэффициента детерминации. Рассчитаем  $F_{расч.}$ :

$$F_{расч} = \frac{0,96(6-2-1)}{(1-0,96)*2} = 37,3$$

$$\alpha = 0,1; \nu_1 = 2; \nu_2 = 6 - 2 - 1 = 3; \Rightarrow F_{кр} = 5,46$$

$$F_{расч} = 37,3 > F_{кр} = 5,46;$$

т.о. коэффициент детерминации значим, уравнение регрессии значимо.

### 3.

Большое значение в анализе на основе многофакторной регрессии имеет сравнение влияния факторов на зависимый показатель  $y$ . Коэффициенты регрессии для этой цели не используется, из-за различий единиц измерения и различной степени колеблемости. От этих недостатков свободные коэффициенты эластичности:

$$\mathcal{E}_j = a_j \frac{\overline{x_j}}{y_j} \quad (11)$$

Эластичность показывает, на сколько процентов в среднем изменяется зависимый показатель  $y$  при изменении переменной  $x_j$  на 1% при условии неизменности значений остальных переменных. Чем больше  $\mathcal{E}_j$ , тем больше влияние соответствующей переменной. Как и в парной регрессии для множественной регрессии различают точечный прогноз и интервальный прогноз. Точечный прогноз (число) получают при подстановке прогнозных значений независимых переменных в уравнение множественной регрессии. Обозначим через:

$$\overline{x_p}' = (1, x_{1p}, x_{2p}, \dots, x_{kp}) \quad (12)$$

вектор прогнозных значений независимых переменных, тогда точечный прогноз

$$\hat{y}_p = \overline{x_p}' * \overline{a} \quad (13)$$

или

$$\hat{y}_p = a_0 + a_1 x_{1p} + a_2 x_{2p} + \dots + a_k x_{kp} \quad (14)$$

Стандартная ошибка предсказания в случае множественной регрессии определяется следующим образом:

$$S_p = S \sqrt{1 + \overline{x_p}' (X'X)^{-1} \overline{x_p}} \quad (15)$$

Выберем уровень значимости  $\alpha$  по таблице распределения Стьюдента. Для уровня значимости  $\alpha$  и числа степеней свободы  $\nu = n-k-1$  найдем  $t_{кр}$ . Тогда истинное значение  $y_p$  с вероятностью  $1 - \alpha$  попадает в интервал:

$$\hat{y}_p - t_{кр} * S_p < y_p < \hat{y}_p + t_{кр} * S_p \quad (16)$$

## **Тема 5: Временные ряды.**

*Вопросы:*

4. Основные понятия временных рядов.
5. Основная тенденция развития – тренд.
6. Построение аддитивной модели.

### **1.**

**Временные ряды** представляют собой совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени.

Момент (или период) времени обозначают  $t$ , а значение показателя в момент времени обозначают  $y(t)$  и называют **уровнем ряда**.

Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые можно разделить на 3 группы:

Длительные, постоянно действующие факторы, оказывающие на изучаемое явление определяющее влияние и формирующие основную тенденцию ряда – тренд  $T(t)$ .

Кратковременные периодические факторы, формирующие сезонные колебания ряда  $S(t)$ .

Случайны факторы, которые формируют случайные изменения уровней ряда  $\varepsilon(t)$ .

**Аддитивной моделью** временного ряда называется модель, в которой каждый уровень ряда представлен суммой тренда, сезонной и случайной компоненты:

$$y(t) = T(t) + S(t) + \varepsilon(t). \quad (1)$$

**Мультипликативная модель** – это модель, в которой каждый уровень ряда представляет собой произведение перечисленных компонент:

$$y(t) = T(t) * S(t) * \varepsilon(t). \quad (2)$$

Выбор одной из моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда колебаний примерно постоянна, то строят аддитивную модель. Если амплитуда возрастает, то мультипликативную модель.

Основная задача эконометрического анализа заключается в выявлении каждой из перечисленных компонент.

### **2.**

**Основной тенденцией развития (трендом)** называют плавное и устойчивое изменение уровней ряда во времени свободное от случайных и сезонных колебаний.

Задача выявления основных тенденций развития называется **выравниванием временного ряда**.

К методам выравнивания временного ряда относят:

- 1) метод укрупнения интервалов,



- 2) метод скользящей средней,
- 3) аналитическое выравнивание.

1) Укрупняются периоды времени, к которым относятся уровни ряда. Затем по укрупненным интервалам суммируются уровни ряда. Колебания в уровнях, обусловленные случайными причинами, взаимно погашаются. Более четко обнаружится общая тенденция.

2) Для определения числа первых уровней ряда рассчитывается средняя величина. Затем рассчитывается средняя из такого же количества уровней ряда, начиная со второго уровня и т.д. средняя величина скользит по ряду динамики, продвигаясь на 1 срок (момент времени). Число уровней ряда, по которому рассчитывается средняя, может быть четным и нечетным. Для нечетного скользящую среднюю относят к середине периода скольжения. Для четного периода нахождение среднего значения не сопоставляют с определением  $t$ , а применяют процедуру центрирования, т.е. вычисляют среднее из двух последовательных скользящих средних.

3) Построение аналитической функции, характеризующей зависимость уровня ряда от времени. Для построения трендов применяют следующие функции:

$$\hat{y}_t = a + bt;$$

$$\hat{y}_t = a + \frac{b}{t};$$

$$\hat{y}_t = e^{a+bt};$$

$$\hat{y}_t = a * t^b;$$

$$\hat{y}_t = a + b_1t + b_2t^2.$$

Параметры трендов определяются с помощью МНК. Выбор наилучшей функции осуществляется на основе коэффициента  $R^2$ .

### 3.

Построение аддитивной модели проведем на примере.

*Пример 7:*

Имеются поквартальные данные об объеме потребления электроэнергии в некотором районе за 4 года. Данные в млн. кВт в таблице 1.

Таблица 1

Год \ Квартал	1	2	3	4
1	6,0	7,2	8,0	9,0
2	4,4	4,8	5,6	6,6
3	5,0	6,0	6,4	7,0
4	9,0	10,0	11,0	10,8

Построить модель временного ряда.

В этом примере в качестве независимой переменной рассматриваем номер квартала  $t = \overline{1,16}$ , а в качестве зависимой переменной  $y(t)$  потребление электроэнергии за квартал.

Из диаграммы рассеяния можно увидеть, что тенденция (тренд) носит линейный характер. Видно также наличие сезонных колебаний (период = 4) одинаковой амплитуды, поэтому будем строить аддитивную модель.

$$y(t) = T(t) + S(t) + \varepsilon(t)$$

Построение модели включает следующие шаги:

1. Проведем выравнивание исходного ряда методом скользящей средней за 4 квартала и проведем центрирование:
  - 1.1. Просуммируем уровни ряда последовательно за каждые 4 квартала со сдвигом на 1 момент времени.
  - 1.2. Разделив полученные суммы на, 4 найдем скользящие средние.
  - 1.3. Приводим эти значения в соответствие с фактическими моментами времени, для чего найдем среднее значение из двух последовательных скользящих средних – центрированные скользящие средние.
2. Рассчитаем сезонную вариацию. Сезонная вариация  $(t) = y(t) -$  центрированная скользящая средняя. Построим таблицу 2 .

Таблица 2

Сквозной № квартала t	Потребление электроэнергии Y(t)	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной вариации
1	6,0	-	-	-
2	4,4	6,1	-	-
3	5,0	6,4	6,25	-1,25
4	9,0	6,5	6,45	2,55
5	7,2	6,75	6,625	0,575
:	:	:	:	:
14	6,6	8,35	8,375	-1,775
15	7,0	-	-	-
16	10,8	-	-	-

3. На основе сезонной вариации в таблице 3 рассчитывается сезонная компонента.

Таблица 3

Показатели	Год	Номер квартала в году				
		I	II	III	IV	
	1	-	-	-1,250	2,550	
	2	0,575	-2,075	-1,100	2,700	
	3	0,550	-2,025	-1,475	2,875	
	4	0,675	-1,775	-	-	
Итого		1,8	-5,875	-3,825	8,125	Сумма
Среднее		0,6	-1,958	-1,275	2,708	0,075
Сезонная компонента		0,581	-1,977	-1,294	2,690	

4. Устраняем сезонную компоненту из исходных уровней ряда:

$$T(t) + \varepsilon(t) = y(t) - S(t),$$

$$T(1) + \varepsilon(1) = y(1) - S(1) = 6 - 0,581 = 5,419.$$

5. Аналитически выравниваем ряд  $T + \varepsilon$ . Строим регрессию:

$$T(t) = 5,706 + 0,187t.$$

6. Рассчитываем ошибку:

$$\varepsilon(t) = y(t) - T(t) - S(t).$$

Результаты вычислений приведем в таблице 4.

Таблица 4

t	Y(t)	S(t)	T(t)+ $\varepsilon$ (t)=Y(t)-S(t)	T(t)	e
1	6,0	0,581	5,419	5,893	-0,474
2	4,4	-1,977	6,337	6,088	0,256
3	5,0	-1,294	6,294	6,268	0,025
4	9,0	2,690	6,310	6,455	-0,145
5	7,2	0,581	6,619	6,642	-0,023
:	:	:	:	:	:
16	10,8	2,690	8,11	8,701	-5,91

7. Рассчитываем коэффициент детерминации  $R^2$ :

$$R^2 = 1 - \frac{\text{var}(e)}{\text{var}(y)} = 1 - \frac{0,0684}{4,196} = 0,984.$$

*Вывод:*

Аддитивная модель объясняет 98,4% общей вариации уровней исходного временного ряда.

## **Тема 6.**

### ***Эконометрический анализ при нарушении предпосылок метода наименьших квадратов***

4. Понятие мультиколлинеарности.
5. Понятие автокорреляции.
6. Понятие гетероскедастичности

#### **1.**

Одним из условий регрессионного анализа является предположение о линейной независимости объясняющих переменных. Однако, может оказаться, что несколько или все объясняющие переменные могут иметь общий временной тренд, относительно которого они совершают малые

колебания. Тогда условие линейной независимости объясняющих переменных нарушается.

**Мультиколлинеарность** – высокая взаимная коррелированность (линейная зависимость) объясняющих переменных. Различают функциональную и стохастическую мультиколлинеарность.

При функциональной мультиколлинеарности определитель матрицы  $X'X$  равен 0. В этом случае невозможно решить матричное уравнение (3).

При стохастической мультиколлинеарности определитель матрицы  $X'X$  очень мал. Она имеет место, когда хотя бы между двумя объясняющими переменными существует тесная корреляционная связь).

Для определения наличия мультиколлинеарности существуют 2 способа:

1. Рассчитывается матрица коэффициентов парной корреляции. Если между какими-либо независимыми переменными коэффициент парной корреляции больше 0,8, то считают, что мультиколлинеарность имеет место.

2. Рассчитывают определитель матрицы  $X'X$  и его близость к 0 также свидетельствует о наличии мультиколлинеарности.

Методы устранения мультиколлинеарности:

1. Из двух объясняющих переменных, имеющих высокий коэффициент корреляции, исключают из рассмотрения ту, которая имеет меньший коэффициент корреляции с зависимой переменной.

2. Метод включения: независимая переменная включается в уравнение регрессии в том случае, если включение существенно увеличивает значение коэффициента множественной корреляции.

3. Метод исключения: после построения уравнения регрессии проверяется значимость всех коэффициентов. Из уравнения исключаются независимые переменные с незначимым коэффициентом. Затем получают новое уравнение регрессии и опять проводят оценку значимости коэффициентов.

*Пример 5:*

Пусть по данным бюджетного обследования 7 случайно выбранных семей изучалась зависимость накоплений  $y$  от дохода  $x_1$ , расходов на питание  $x_2$  и стоимости имущества  $x_3$ . После применения к исходным данным программы «Корреляция» была получена следующая корреляционная матрица:

	$y$	$x_1$	$x_2$	$x_3$
$y$	1			
$x_1$	0,85	1		
$x_2$	0,8	0,93	1	
$x_3$	-	-	-	1
	0,75	0,38	0,28	

$r_{x_1x_2} = 0,93 > 0,8 \Rightarrow$  тесная корреляционная связь. значит, одну из этих переменных следует исключить из уравнения регрессии. Исключаем  $x_2$ , т.к.  $r_{yx_1} = 0,85 > 0,8$ . Таким образом в уравнение регрессии включаются такие факторы как доход  $x_1$  и стоимость имущества  $x_3$ .

*Замечание:*

При построении регрессии и подборе независимых переменных необходимо помнить, что объем выборки  $n$  (число независимых наблюдений) должно быть в 6-7 раз больше чем число независимых переменных.

## 2.

Третья предпосылка регрессионного анализа гласит: случайные члены  $\varepsilon_i$  теоретической регрессии должны быть независимы друг от друга.

**Автокорреляция** – зависимость текущего значения случайного члена от непосредственно предшествующего значения. Т.о. автокорреляция случайного члена теоретической регрессии нарушает третью предпосылку регрессионного анализа. Причинами автокорреляции могут быть:

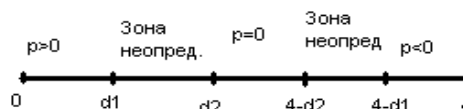
- ошибки спецификации, т.е. неправильно подобранная математическая функция,
- необходимость введения в модель новой переменной,
- ошибки наблюдения.

Наличие или отсутствие автокорреляции проверяют с помощью критерия (статистики) Дарбина Уотсона.

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (11)$$

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

Значение статистики DW распределено в интервале (0,4). По таблице распределения статистики DW на основании уровня значимости  $\alpha$ , объема выборки  $n$  и числа объясняющих переменных  $k$  находят критические точки  $d_1, d_2$ . Эти точки разбивают отрезок (0,4) на 5 зон:



Проверка автокорреляции:

1. Формируется гипотеза  $H_0$  об отсутствии автокорреляции:

$$H_0 : \rho = 0,$$

и альтернативные гипотезы  $H_1 : \rho > 0$  о наличии положительной автокорреляции и  $H_2 : \rho < 0$  о наличии отрицательной автокорреляции.

2. Выбирается уровень значимости  $\alpha = 0,05 \vee 0,01$ .

3. По таблице распределения DW на основании  $\alpha$ ,  $n$  и  $k$  находят критические точки  $d_1$  и  $d_2$

4. На основании выборочных данных для построенной регрессии по формуле (11) рассчитывается значение статистики DW:

- если  $0 < DW < d_1$ , то с вероятностью  $1 - \alpha$  принимается гипотеза  $H_1$ ,
- если  $d_1 < DW < d_2$ , то нет оснований для принятия или непринятия всех гипотез,

- если  $d_2 < DW < 4 - d_2$ , то с вероятностью  $1 - \alpha$  принимается гипотеза  $H_0$ ,

- если  $4 - d_2 < DW < 4 - d_1$ , то неопределенность,

- если  $4 - d_1 < DW < 4$ , то с вероятностью  $1 - \alpha$  принимается гипотеза  $H_2$ .

Если DW попадает в зону неопределенности, то для обнаружения автокорреляции используются другие методы. Если утверждается наличие автокорреляции, то тогда пытаются ее устранить.

### 3.

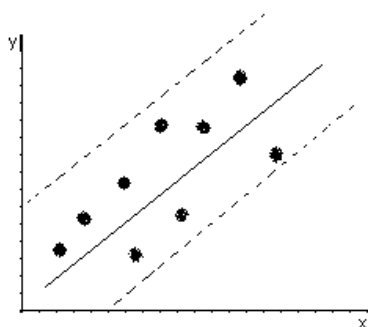
Вторая предпосылка регрессионного анализа гласит, что дисперсия случайного члена регрессионной модели может быть постоянной для любого наблюдения, т.е.:

$$D(\varepsilon_i) = \sigma^2, i = \overline{1, n}$$

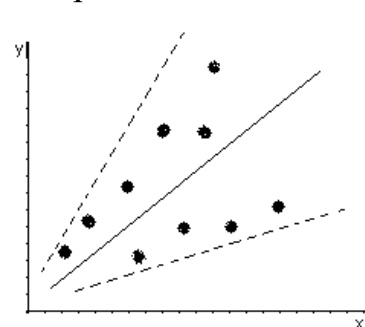
Это условие называется **гомоскедастичностью** (одинаковой разбросанностью).

Зависимость дисперсии случайного члена от номера наблюдения называется гетероскедастичностью,  $D(\varepsilon_i) = \sigma_i^2, i = \overline{1, n}$

А) гомоскедастичность



Б) гетероскедастичность



Для обнаружения гетероскедастичности используются различные тесты. Например, тест ранговой корреляции Спирмена. При наличии гетероскедастичности для оценки параметров регрессионной модели используют обобщенные (или взвешенный МНК(ОМНК)).

Пример 6:

Пусть исследуется зависимость денежных сбережений  $y$  от среднедушевых доходов  $x$  в 12 семьях. Данные в млн. руб. приводятся в таблице1:

Таблица1

i	$y_i$	$x_i$	i	$y_i$	$x_i$
1	0,3	1	7	5,8	7
2	0,1	2	8	2,5	8
3	2,2	3	9	7,5	9
4	0,9	4	10	3,0	10
5	4,0	5	11	9,0	11
6	1,7	6	12	3,4	12

Из диаграммы рассеяния видно, что зависимость носит линейный характер, и что с ростом доходов  $x$  вариация (разброс) отклонений сбережений  $y$  от линии регрессии растет пропорционально  $x$ , что позволяет сделать вывод о наличии гетероскедастичности.

Построим регрессию  $\hat{y} = a_0 + a_1x$  используя МНК и ОМНК. Обратимся к программе «Регрессия» и введем данные таблицы. Получим уравнение регрессии на основе МНК:

$$\hat{y} = -0,097 + 0,533x$$

(1,316) (0,179)

Предположим, что имеет место гетероскедастичность, тогда регрессионная модель имеет вид:

$$y_i = a_0 + a_1x_i + \varepsilon_i x_i$$

Разделим обе части этой модели на  $x_i$ :

$$\frac{y_i}{x_i} = a_0 \frac{1}{x_i} + a_1 + \varepsilon_i$$

В последней модели случайная ошибка  $\varepsilon$  не зависит от объясняющей переменной  $x_i$ . Сделаем замену переменных:

$$\frac{y_i}{x_i} = y_i'$$

$$\frac{1}{x_i} = x_i'$$

Получим регрессионную модель:

$$y_i' = a_0 x_i' + a_1 + \varepsilon_i,$$

для которой применим классический МНК. В таблице2 подготовим данные для построения последней регрессионной модели:

Таблица2

i	$y_i' = \frac{y_i}{x_i}$	$x_i' = \frac{1}{x_i}$	i	$y_i' = \frac{y_i}{x_i}$	$x_i' = \frac{1}{x_i}$
1	0,3	1	7	0,828	0,142
2	0,05	0,5	8	0,312	0,125
3	0,733	0,333	9	0,833	0,111
4	0,225	0,25	10	0,3	0,1
5	0,8	0,2	11	0,818	0,09
6	0,283	0,166	12	0,283	0,083

Для преобразованных данных применим МНК. Получим регрессию

$$\hat{y}' = 0,576 - 0,386x' \\ (0,12) \quad (0,333)$$

Делаем обратную замену переменных:

$$\frac{\hat{y}}{x} = 0,576 - 0,386 \frac{1}{x}$$

ОМНК:

$$\hat{y} = -0,363 + 0,576x \\ (0,333) \quad (0,12)$$

МНК:

$$\hat{y} = -0,097 + 0,533x \\ (1,316) \quad (0,179)$$

Стандартные ошибки коэффициентов регрессии, построенной на основе ОМНК, меньше соответствующих ошибок регрессии, построенной на основе МНК. Поэтому в данном примере уравнение регрессии, построенное на основе ОМНК предпочтительнее.



## Литература

1. Магнус Л.Р., Катышев П.К., Пересецкий А.А. Эконометрика. М., Дело, 2004.
2. Бородич С.А. Эконометрика. Минск: Новое знание, 2001.
3. Эконометрика. Под редакцией Елисеевой И.И. М.: Финансы и статистика, 2007.
4. Доугерти К. Введение в эконометрику. М.: ИНФРА-М, 2004.
5. Экономико-математические методы и модели. Под ред. Миксюк С.Ф. Мн.: БГЭУ, 2006.
6. Экономико-математические методы и модели; практикум. Под ред. Миксюк С.Ф. Мн.: БГЭУ, 2006.