5. Моделирование рядов динамики

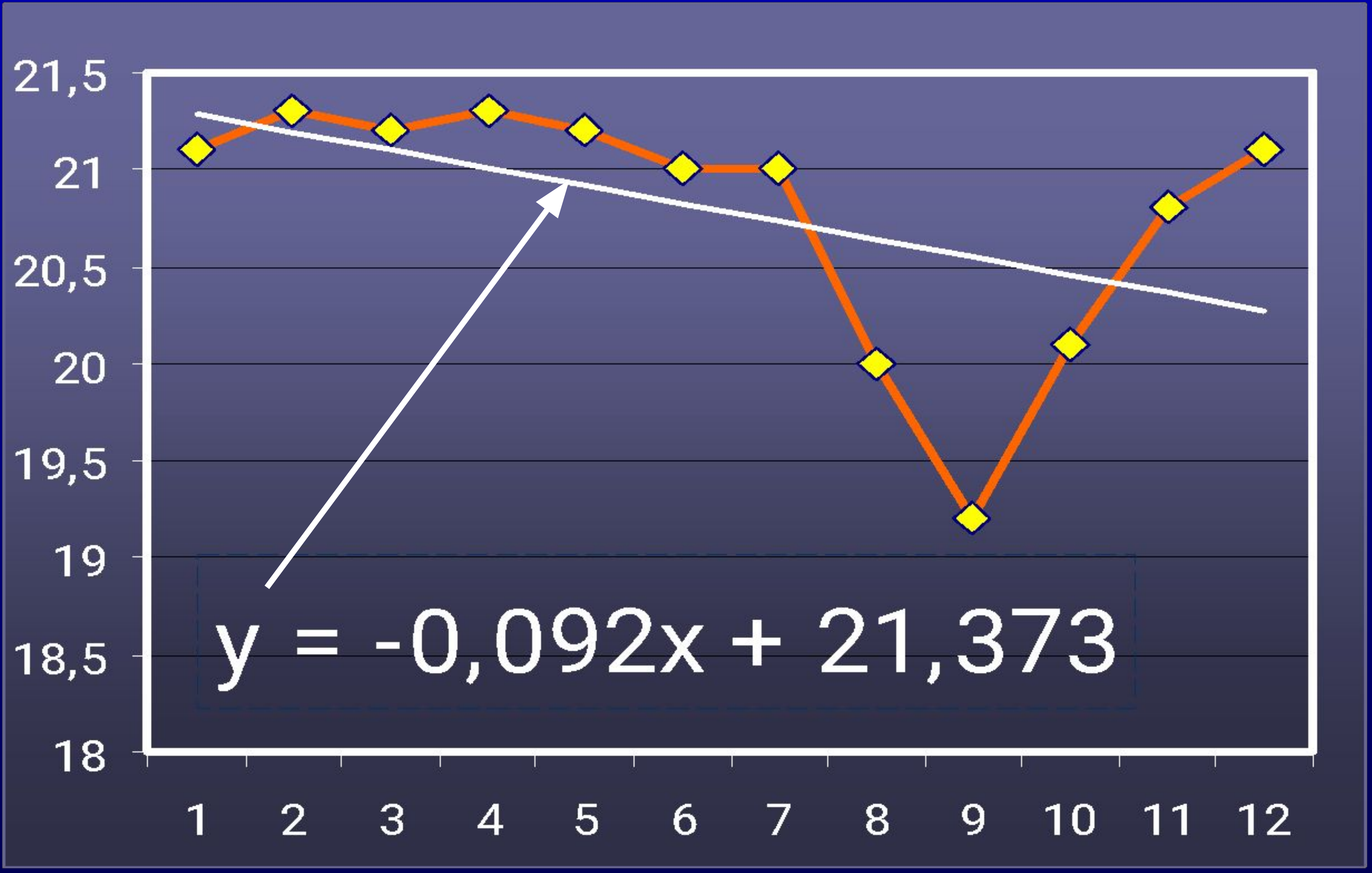
5. 1. Постановка задачи и общие сведения о временных рядах

Под временным рядом (рядом динамики) в экономике понимается совокупность наблюдений некоторого признака (случайной величины Y) в последовательные моменты времени.

Примером такого ряда могут быть данные о среднем размере товарных запасов в универмаге по месяцах в 1997 г., млн. руб.:

1	2	3	4	5	6
22,1	21,3	21,2	21,3	21,2	21,0
7	8	9	10	11	12
21,0	20,0	19,2	20,1	20,8	21,1

Графическое представление временного ряда. Показана линия и уравнение тренда



Каждый уровень (значение) временного ряда формируется под действием большого числа факторов, которые можно разделить на три группы:

1. факторы формирующие тенденции ряда (тренд);

2. Факторы формирующие сезонные колебания, отражающие повторяемость экономических процессов в течении не очень длительного периода;

3. Факторы отражающие повторяемость экономических процессов в течении длительных периодов;

4. Случайные факторы

Естественно предположить, что все четыре компоненты (трендовая, сезонная, циклическая и случайная) будут формировать наблюдаемое значение случайной величины Y.

Поэтому в общем случае временной ряд можно представить либо в виде

аддитивной модели

$$y_t = u_t + \nu_t + c_t + \varepsilon_t$$ либо

мультипликативной модели

$$y_t = u_t \cdot \nu_t \cdot c_t \cdot \varepsilon_t.$$

Важно подчеркнуть. что в отличие от

ε_t величины u_t, v_t, c_t

не являются случайными

Важнейшей классической задачей при исследовании экономических временных рядов является выявление и статистическое оценивание основной тенденции развития изучаемого явления.

Отметим основные этапы анализа временных рядов:

1. Графическое представление временного ряда;
2. Выделение и удаление закономерных (неслучайных составляющих временного ряда;
3. Исследование случайной составляющей временного ряда и проверка адекватности математической модели для ее описания;
4. прогнозирование развития изучаемого явления на основе имеющегося временного ряда.

На первый взгляд кажется. что набор величин

$$y_1, y_2, \ldots, y_n$$

можно рассматривать как элементы некоторой случайной выборки. В действительности это не так. В отличии от элементов случайной выборки значения временного ряда

$$y_t \ (t = 1,2,\ldots n)$$

не являются статистически независимыми. Например, потребление электроэнергии в городе подвержено сезонным колебаниям и поэтому будут сильно коррелировать данные, относящиеся к одному и тому же месяцу, взятые для разных лет.

5.2. Автокорреляция в рядах динамики

При наличии во временном ряде тенденции или циклических колебаний значения каждого последующего уровня ряда зависит от предыдущего.

Корреляционную зависимость между последовательными уровнями ряда называют автокорреляцией.

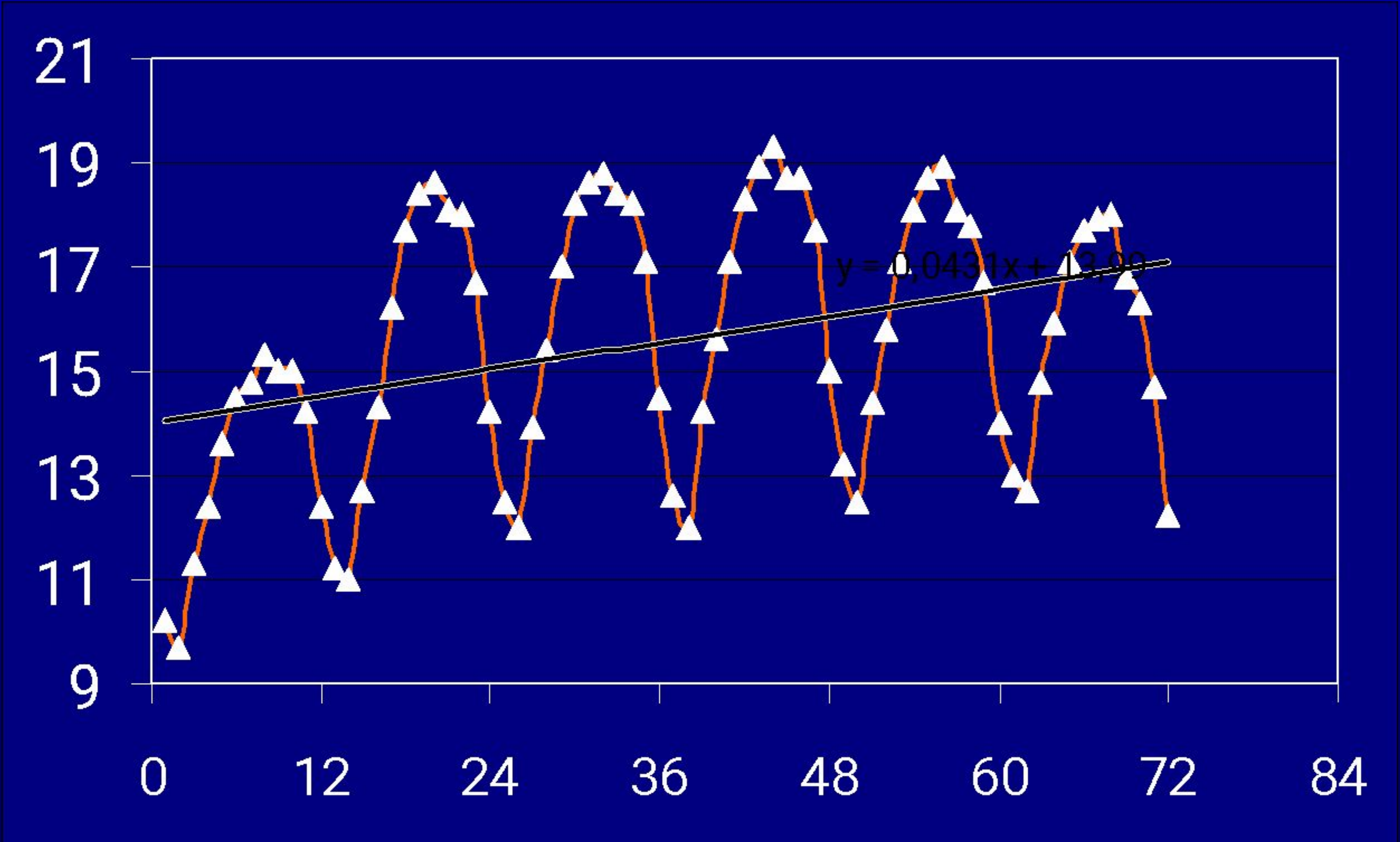
Количественно ее можно измерить при помощи вычисления коэффициента автокорреляции.

Ниже представлена таблица с данными о частных расходах на жилищное строительство в небольшом городке США за период с января 1988 по декабрь 1993 года.

Прогноз тенденций расходов на ближайшие 1-2 года мог бы заинтересовать не только строителей, но и, например, риэлторские организации.

Поскольку табличные данные воспринимаются очень плохо, представим исходные данные в графическом виде

Частные расходы на жилищное строительство в небольшом городке США за период с января 1988 по декабрь 1993



Очевидно, что если сдвинуть данные ровно на год, то картина повторяется и поэтому коэффициент корреляции данных с лагом (сдвигом) на 12 месяцев будет велик. Действительно расчет показывает, что в этом случае коэффициент корреляции равен 0,886, а при лаге 6 месяцев он отрицателен и равен – 0,535.

Можно построить график график зависимости коэффициента корреляции от номера лага. Этот график называется кореллограммой.

Пример 1. Аддитивная модель ряда.

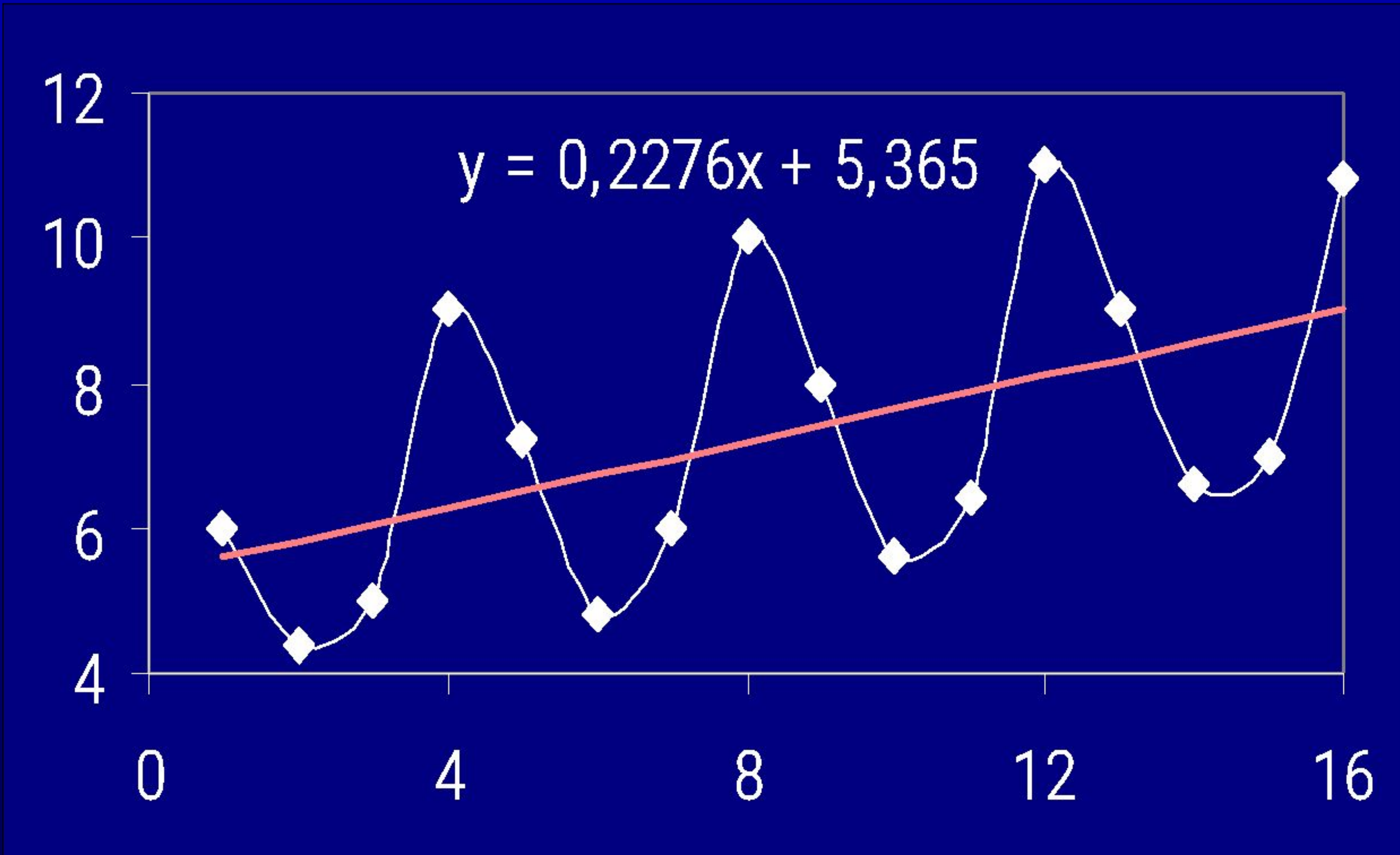
Рассмотрим более удобный для анализа пример зависимости поквартального потребления электроэнергии (млрд. Квт/час) жителями региона за 16 кварталов. Данные приведены на след. слайде.

Объем потребления электроэнергии (млрд. Квт/час) жителями региона за 16 кварталов

5

t(квартал)	y(t)	t(квартал)	y(t)
1	6	9	8
2	4,4	10	5,6
3	5	11	6,4
4	9	12	11
5	7,2	13	9
6	4,8	14	6,6
7	6	15	7
8	10	16	10,8

Данные об объемах потребления электроэнергии (млрд. Квт/час) жителями региона за 16 кварталов.



Вычислим коэффициенты корреляции исходных данных и данных сдвинутых на один кварта, два квартала три квартала 4 квартала (с лагами один, два, три, четыре).

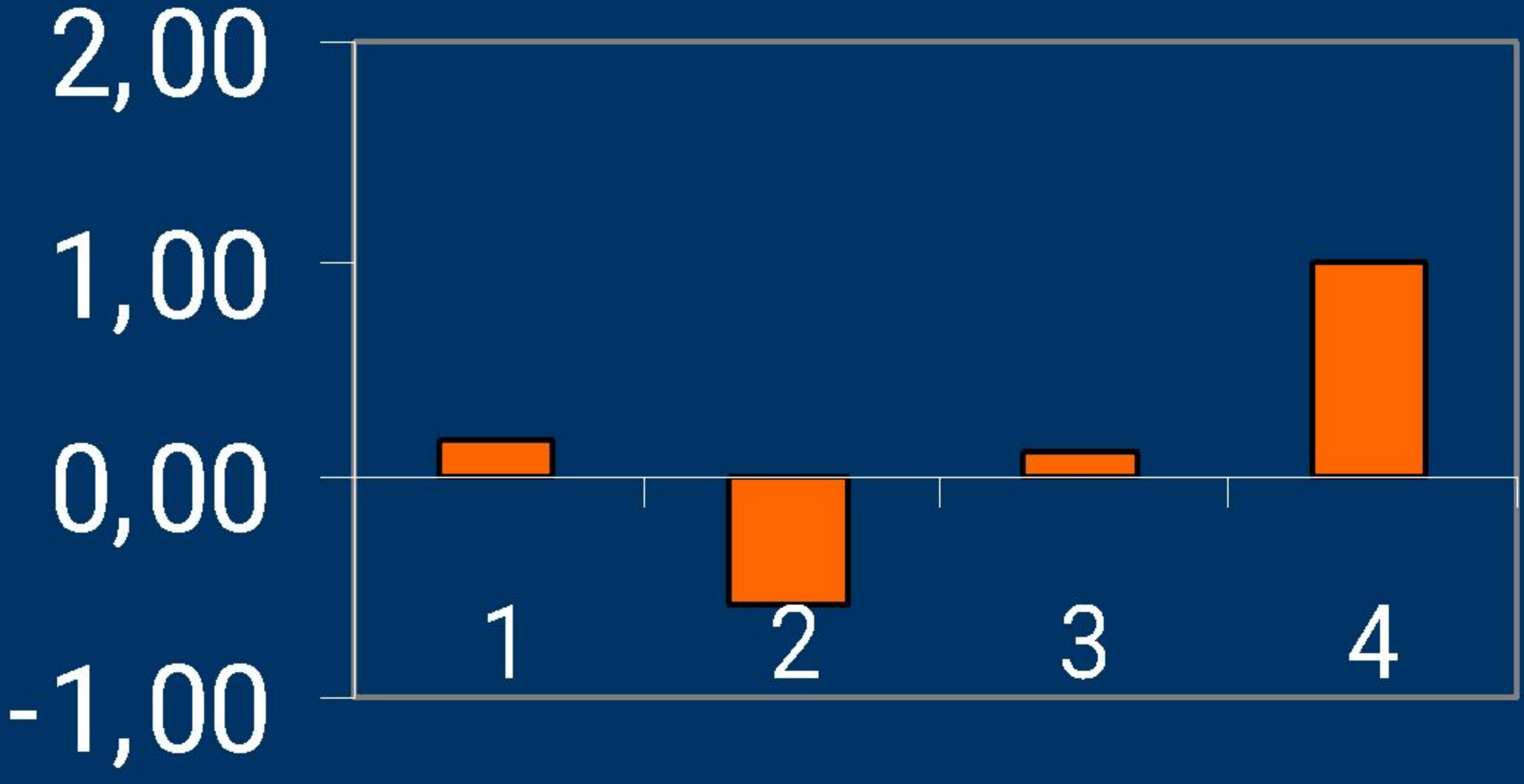
На следующем слайде показана структура данных при вычислении автокорреляционной функции с лагами 1,2,3,4 (Для наглядности часть данных опущена).

t(квартал)	Y(t)				7
1	6	лаг 1			
2	4,4	6	лаг 2		
3	5	4,4	6	лаг 3	
4	9	5	4,4	6	лаг 4
5	7,2	9	5	4,4	6
14	6,6	9	11	6,4	5,6
15	7	6,6	9	11	6,4
16	10,8	7	6,6	9	11
		10,8	7	6,6	9
			10,8	7	6,6
				10,8	7
					10,8

С помощью функции Корелл ( ) электронных таблиц Excel найдем значения коэффициентов автокорреляции и построим по этим данным кореллограмму

1	0,17
2	-0,57
3	0,11
4	0,98

Коррелограмма



Далеко не всегда автокорреляция столь заметна, как в рассмотренных выше примерах.

В то же время часто обнаруживается, что значения отклика в некоторой точке временного ряда сильно коррелировано с несколькими предшествующими и/или последующими значениями. Действительно, для многих явлений их современное состояние функционально определяется предшествующими состояниями системы, в большей степени недавними, в гораздо меньшей - далеко отстоящими от заданного по временному ряду. Подобные связи принято называть автокорреляцией - корреляцией ряда с самим собой.

Автокорреляция первого порядка характеризует тесноту связи между соседними значениями временного ряда, автокорреляция второго порядка - между отстоящими друг от друга на два периода etc. И вообще, автокорреляция n-го порядка относится к степени связанности откликов, разнесенных на n периодов. Предполагая, что возникшая связь между значениями сохранится некоторое время в будущем, мы получаем механизм прогнозирования, основанный на построении регрессии точек ряда на самих себя, то есть - авторегрессии.

Авторегрессионные модели разных порядков - первого, второго, в общем случае n-ого - можно описать уравнениями следующего вида:

$$y_{t_i} = a + b_1 \cdot y_{t_{i-1}} + \varepsilon;$$

$$y_{t_i} = a + b_1 \cdot y_{t_{i-1}} + b_2 \cdot y_{t_{i-2}} + \varepsilon;$$

$$y_{t_i} = a + b_1 \cdot y_{t_{i-1}} + \ldots + b_n \cdot y_{t_{i-n}} + \varepsilon.$$

5.3. Выделение тренда и сезонной составляющей для аддитивной и мультипликативной моделей временного ряда.

Как уже отмечалось, важнейшей задачей исследования временного ряда в экономике является выявление основной тенденции (тренда).

Для решения этой задачи необходимо выбрать вид функции, а затем с помощью метода наименьших квадратов получить коэффициенты теоретической линии регрессии.

Поскольку проблема построения регрессионной модели уже достаточно подробно обсуждалась, здесь мы не будем на ней останавливаться, а сразу перейдем к выявлению тренда и сезонной компоненты для аддитивной модели временного ряда (пример с потреблением электроэнергии).

В рассматриваемом примере 1 метод МНК для линейной модели регрессии приводит к уравнению

$$y_t = 5{,}365 + 0{,}228 \cdot t.$$

Значение параметров, которые возвращает функция ЛИНЕЙН ( ) приведены в табличке справа. Как следует из этих результатов, модель значима при уровне значимости 0,05.

0,228	5,365
0,102	0,986
0,263	1,88
4,983	14
17,62	49,5

Причина небольшого по величине фактора детерминации понятна, поскольку есть еще и сезонная составляющая, которую мы пока не учли.

Для выделения сезонной составляющей найдем разность

$$y_{\mathrm{i}} - y_{t_i}.$$

Эта разность представляет собой сезонную + случайную величину (мы исходим из аддитивной модели динамического ряда, предполагая, что случайная величина удовлетворяет всем требованиям регрессионной модели).

Найдем сезонную + случайную величину для рассматриваемого примера. Для нахождения сезонной компоненты за первый квартал найдем среднее отклонение за первый, пятый , девятый и тринадцатый кварталы.

$$c_1 = (0{,}4 + 0{,}7 + 0{,}6 + 0{,}7)/4 = 0{,}59.$$

t(квартал)	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0
Y	6,0	4,4	5,0	9,0	7,2	4,8	6,0	10,0
Y(t)	5,6	5,8	6,0	6,3	6,5	6,7	7,0	7,2
Y-Y(t)	0,4	-1,4	-1,0	2,7	0,7	-1,9	-1,0	2,8
t(квартал)	9,0	10,0	11,0	12,0	13,0	14,0	15,0	16,0
Y	8,0	5,6	6,4	11,0	9,0	6,6	7,0	10,8
Y(t)	7,4	7,6	7,9	8,1	8,3	8,6	8,8	9,0
Y-Y(t)	0,6	-2,0	-1,5	2,9	0,7	-2,0	-1,8	1,8

Аналогично найдем сезонную компоненту за второй, третий и четвертый кварталы. Соответствующие величины получились равными:

$$c_1 = 0{,}59;$$

$$c_2 = -1{,}84; \quad c_3 = -1{,}31; \quad c_4 = 2{,}559.$$

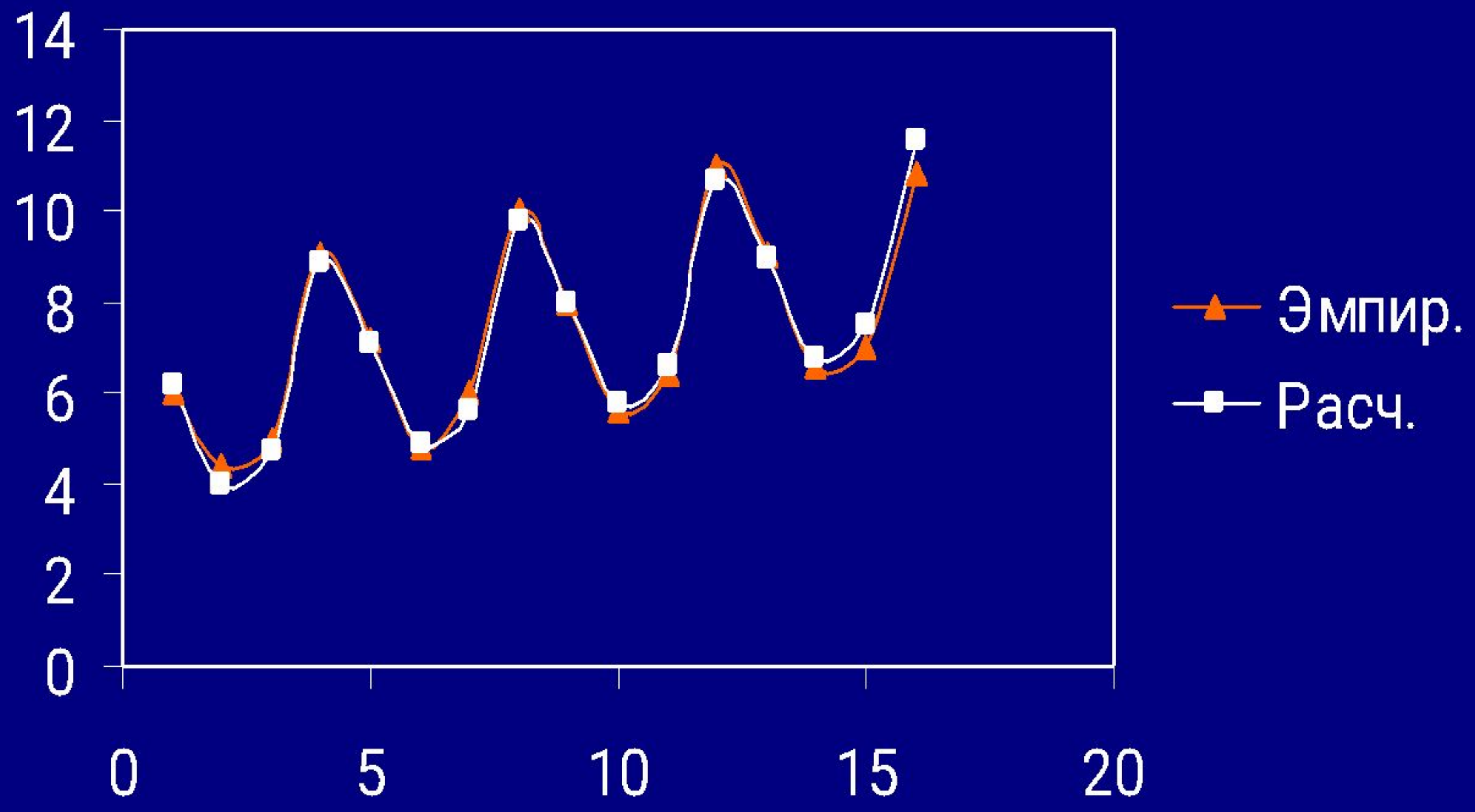
Легко проверить, что сумма сезонных составляющих с большой точностью равна нулю.

Сезонная составляющая + трендовая составляющая образуют детерминированную составляющую модели.

$$y_{\text{дет}}(t_i) = y_{t_i} + c_{t_i}.$$

Представляет интерес определить насколько хорошо детерминированная составляющая описывает эмпирический набор данных. Проведем это сравнение в графической форме, построив два графика зависимости энергопотребления от времени (эмпирический и расчетный). Результаты такого построения приведены на следующем слайде.

Сопоставление эмпирических и расчетных (с учетом сезонной составляющей) данных



Выделение трендовой, сезонной и случайной величин для примера с потреблением электроэнергии. Приведены первые 8 значений

t(квартал)	Y Эмп.	Y Расч.	Сез.+ случ.	Сезон.	Дет.	Случ.
1	6	5,59	0,41	0,59	6,18	-0,18
2	4,4	5,82	-1,42	-1,84	3,98	0,42
3	5	6,05	-1,05	-1,31	4,73	0,27
4	9	6,28	2,72	2,56	8,83	0,17
5	7,2	6,50	0,70	0,59	7,09	0,11
6	4,8	6,73	-1,93	-1,84	4,89	-0,09
7	6	6,96	-0,96	-1,31	5,64	0,36
8	10	7,19	2,81	2,56	9,74	0,26

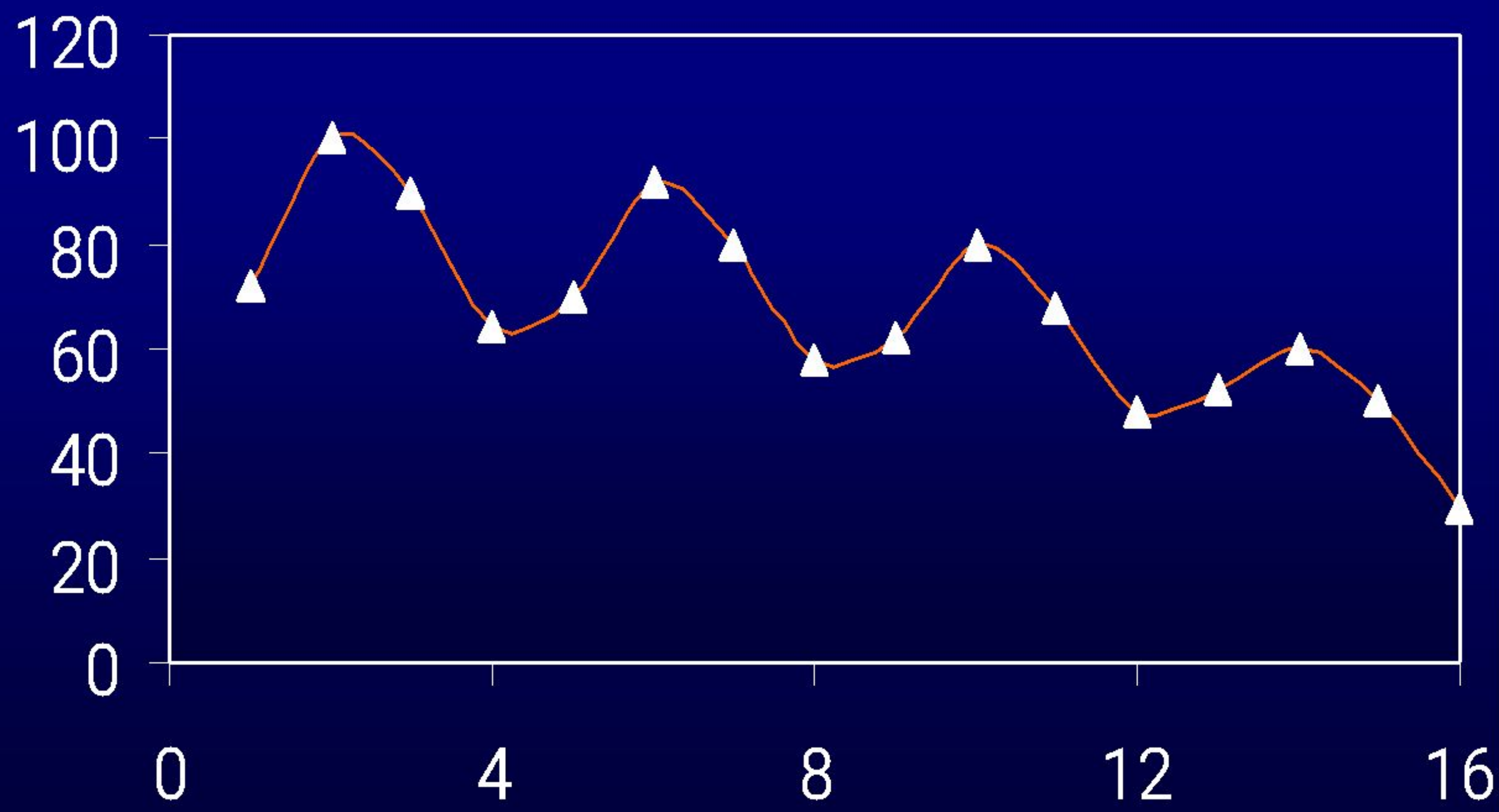
Перейдем теперь к рассмотрению

***Примера 2. мультипликативная модель*.**

Пусть имеются поквартальные данные о прибыли компании за последние годы

	Кварталы			
Год	I	II	III	IV
1	72	100	90	64
2	70	92	80	58
3	62	80	68	48
4	52	60	50	30

Поквартальные данные о прибыли компании за четыре года



Как видно из графика амплитуда осцилляций уменьшается, что и наводит на мысль использовать мультипликативную модель

Для выделения сезонной компоненты в мультипликативной модели временного ряда воспользуемся методом скользящей средней.

Метод скользящей средней в данном случае – это метод выравнивания ряда.

Напомним, что в предыдущем примере для выравнивания ряда мы воспользовались построением теоретической кривой тренда по методу наименьших квадратов, хотя могли использовать и метод скользящей средней.

Суть метода скользящей средней в том, что в данном случае наблюдается явная периодичность в четыре квартала. Чтобы устранить сезонные колебания будем использовать четырехзвенную скользящую среднюю. Для получения результатов по этому методу нужно найти среднее значение по первым четырем точкам, затем сдвинуться на одну точку и получить среднее значение 2,3,4,5 точек, и т. д.

В **Пакете анализа** Excel имеется функция, которая по входному набору данных рассчитывает скользящие средние.

t	y(t)	Скользящая средняя	Сезонная компон.
1,0	72,0		
2,0	100,0		
3,0	90,0	81,50	1,10
4,0	64,0	81,00	0,79
14,0	60,0	52,50	1,14
15,0	50,0	48,00	1,04
16,0	30,0		

После того, как рассчитана скользящая средняя, используя уравнение мультипликативной модели

$$y_i = \overline{y}_i \cdot c_i,$$

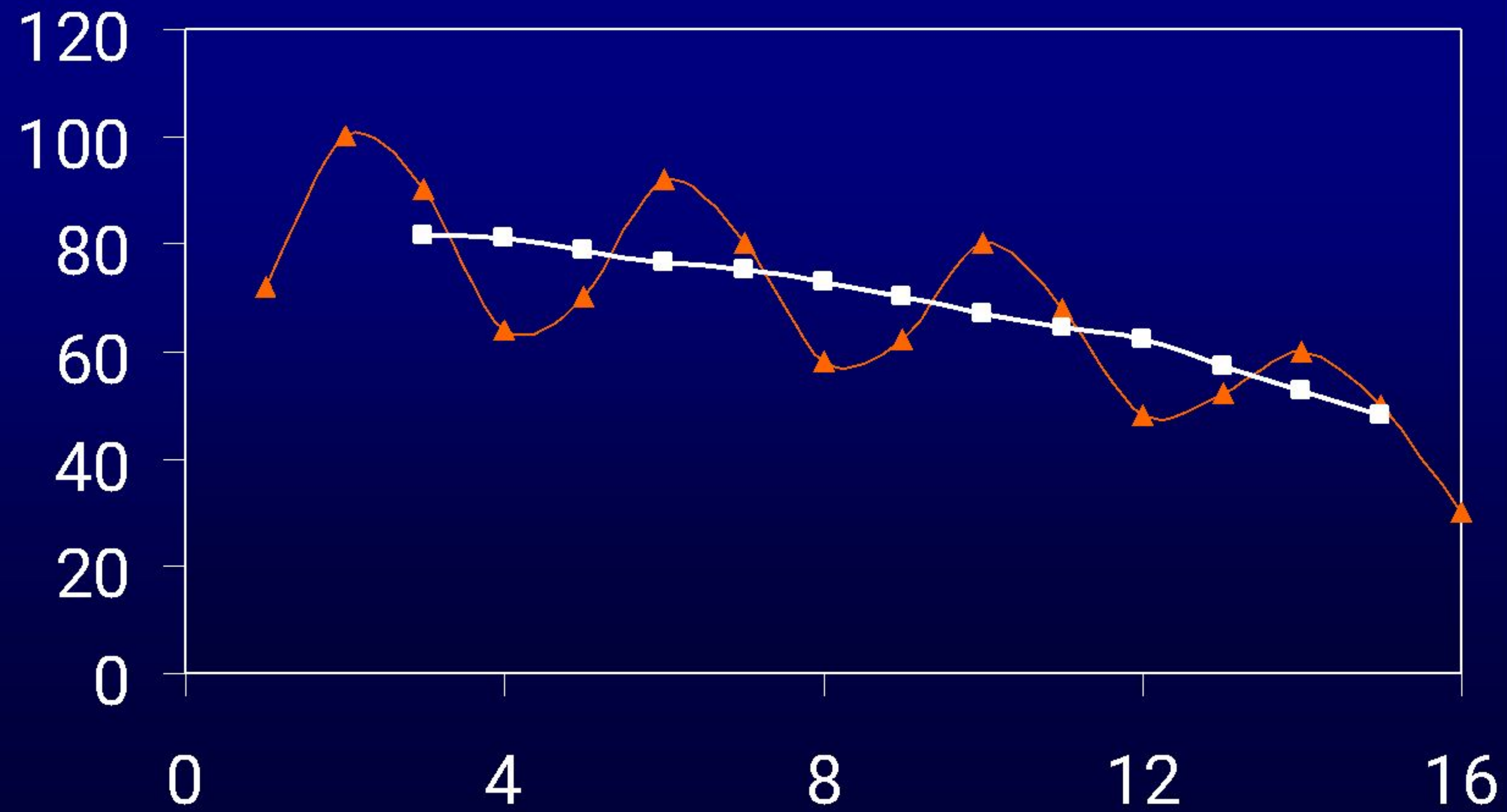
сезонную компоненту найдем как отношение эмпирических уровней ряда и тренда

$$c_i = \frac{y_i}{\overline{y}_i}.$$

Эти данные приведены в последнем столбце на предыдущем слайде.

На слайде отображена лишь часть данных. В графическом виде выровненные данные отображены на следующем слайде.

Выравнивание ряда с помощью четырехзвенной скользящей средней



Итоговые данные для сезонной составляющей

	Кварталы			
Год	I	II	III	IV
1	-	-	1,10	0,79
2	0,89	1,20	1,07	0,79
3	0,89	1,19	1,05	0,77
4	0,91	1,14	1,04	-
Средняя	0,89	1,18	1,07	**0,79**

Существует простой способ проверить правильность проведенных вычислений для сезонной составляющей. Если трендовая составляющая является постоянной и равной *A*, то при вычислении скользящей средней мы получаем

$$\bar{y} = \frac{A \cdot c_1 + A \cdot c_2 + A \cdot c_3 + A \cdot c_4}{4} =$$

$$A \cdot \frac{c_1 + c_2 + c_3 + c_4}{4} = A,$$

Иначе говоря сумма сезонных компонент должна быть равна числу точек, по которым вычисляется скользящее среднее. В рассматриваемом случае эта сумма равна 3, 93 т.е близка к четырем.

Рассчитаем теперь трендовую и случайную составляющие. Для этого эмпирические данные разделим на значение средней сезонной компоненты и построим линейное уравнение регрессии по полученным данным, используя функцию ЛИНЕЙН ( ).

-2,828	92,280
0,228	2,204
0,917	4,203
153,931	14,000
2719,072	247,3

Как следует из приведенных результатов модель и регрессионные коэффициенты являются значимыми при уровне значимости 0,05.

5.4. Прогнозирование по аддитивной и мультипликативной моделям

Предположим, что по данным рассмотренного ***Примера 1*** необходимо дать прогноз потребления электроэнергии жителями района в течении двух следующих кварталов ближайшего года.

Прогнозное значение F_t уровня временного ряда для аддитивной модели рассчитывается как сумма трендовой и сезонной компонент.

$$y_{17} = 5{,}365 + 0{,}228 \cdot 17 + 0{,}591 = 9{,}60;$$

$$y_{18} = 5{,}365 + 0{,}228 \cdot 18 - 1{,}84 = 7{,}17.$$

Прогнозное значение F_t уровня временного ряда в мультипликативной модели есть произведение трендовой и сезонной компонент. Предположим, что по данным ***Примера 2*** необходимо рассчитать прибыли компании за первое полугодие следующего года.

Для этой цели воспользуемся уравнением тренда и значениями сезонной составляющей:

$$y_{17} = (92{,}280 - 2{,}828 \cdot 17) \cdot 0{,}89 = 39{,}341;$$

$$y_{18} = (92{,}280 - 2{,}828 \cdot 18) \cdot 1{,}18 = 48{,}823.$$

Таким образом, прогноз ожидаемой прибыли составит 88,165 тыс. долл. США.

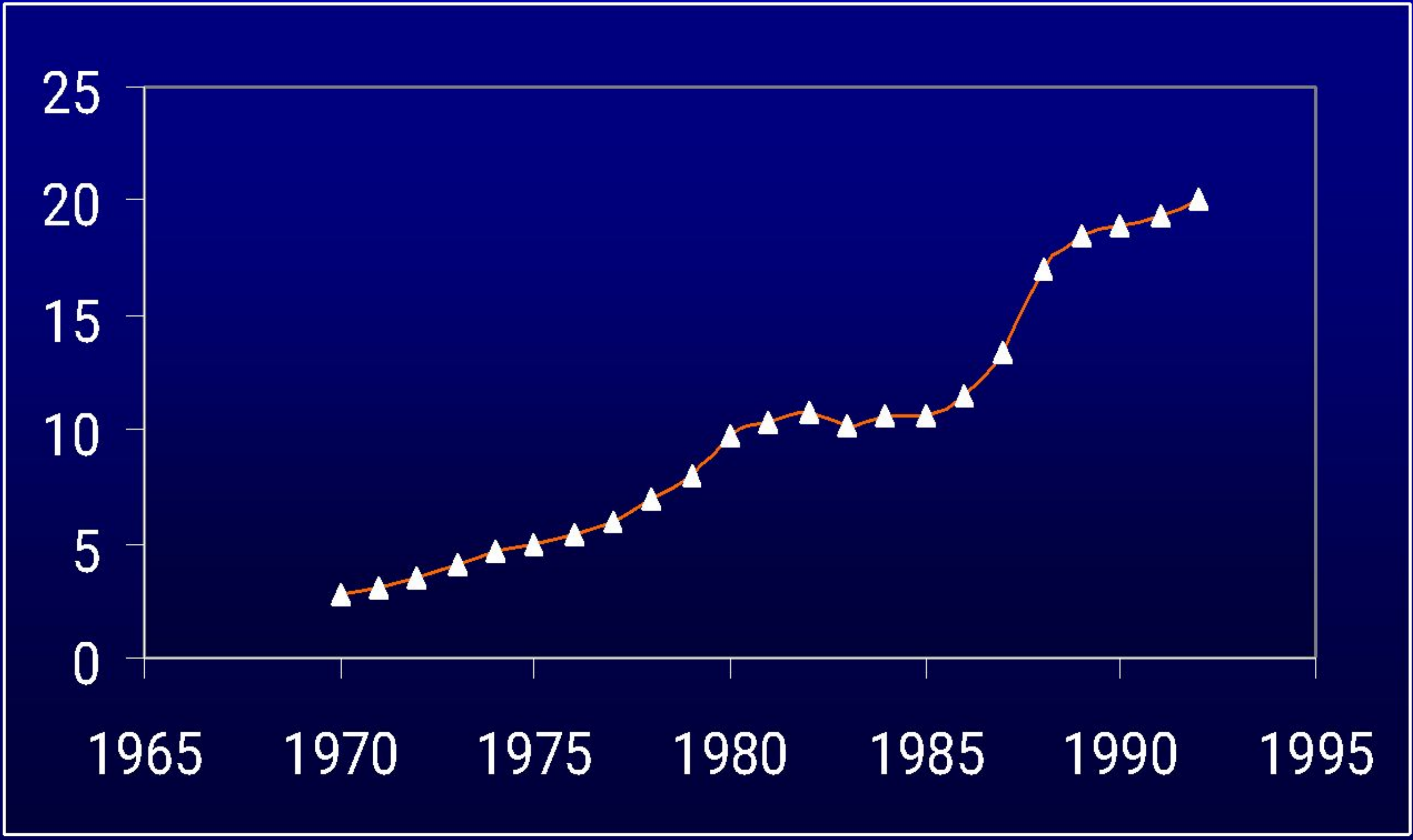
5.5. Обнаружение автокорреляции. Авторегрессионые модели первого порядка.

Рассмотрим прогнозирование для для авторегрессионной модели

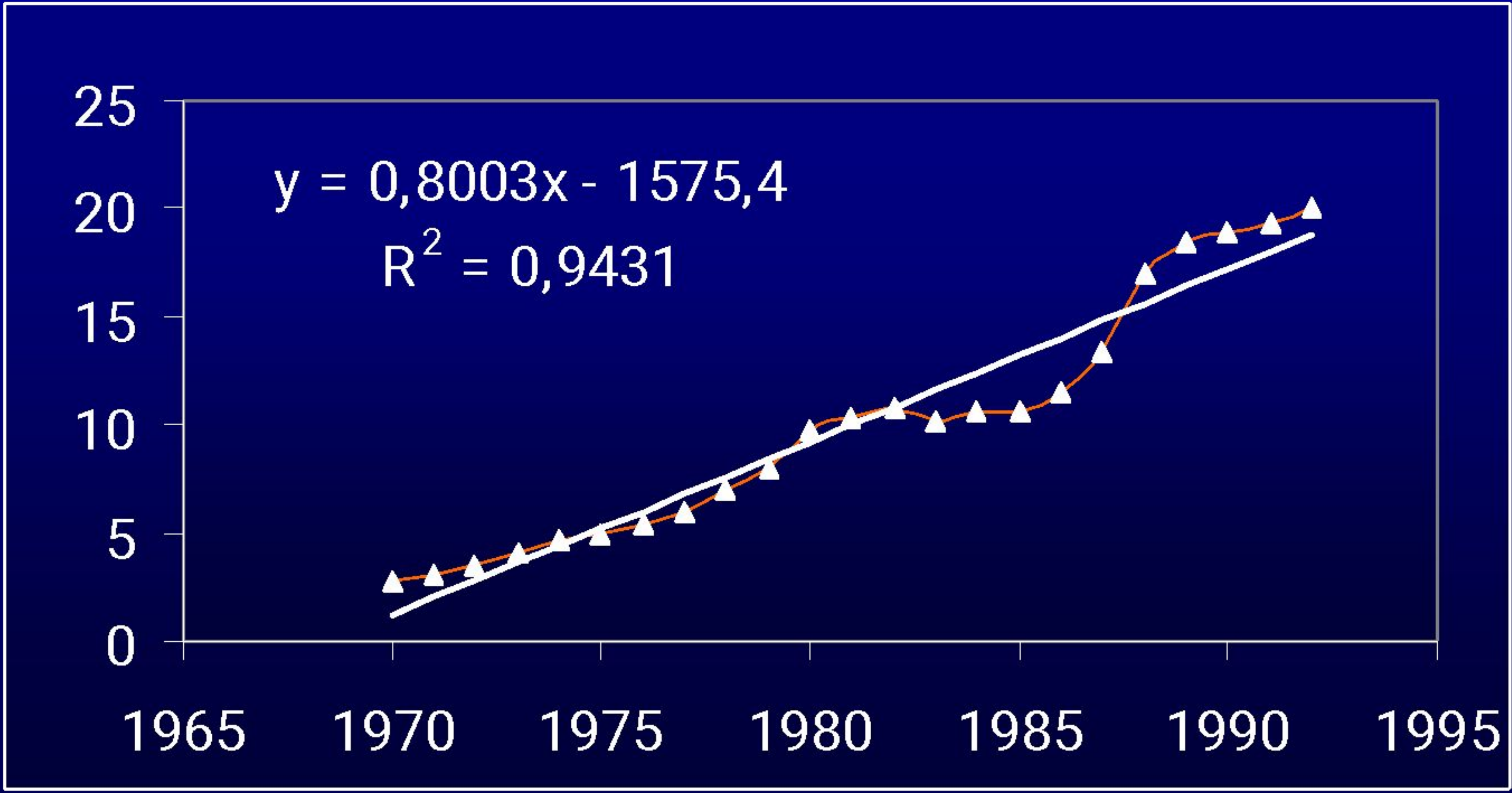
$$y_{t_i} = a + b_1 \cdot y_{t_{i-1}} + \varepsilon.$$

первого порядка на примере статистического материала об объеме выпуска продукции фирмой КОДАК в 1970 - 1992 гг. (млрд. долл.). Исходные статистические данные и промежуточные расчеты см. в файле Кодак.xls. Ниже приведены лишь итоговые расчеты и графические данные.

Объем производства фирмы Кодак в период с 1970 по 1992 г. (млрд. долл.)



Конечно, можно было бы не раздумывая применить линейную или экспоненциальную модель и получить достаточно хорошее согласие с эмпирическими данными.



Большая величина фактора детерминации и статистическая значимость коэффициентов регрессии еще не гарантируют правильность модели. поскольку нужно еще убедиться в выполнении критериев применимости самого метода МНК.

Одним из таких условий является статистическая независимость ошибок для разных наблюдений. Для рядов динамики это условие част нарушается.

Для обнаружения автокорреляции первого порядка используются различные тесты, но наиболее распространенным является тест Дарбина – Уотсона.

Тест Дарбина – Уотсона основан на простой идее: если корреляция ошибок уравнения регрессии не равна нулю, то она присутствует и в остатках обычного уравнения регрессии ε_{t_i}

$$y_{t_\mathrm{i}} = a + b_1 \cdot t_i + \varepsilon_{t_i}.$$

В тесте Дарбина – Уотсона используется величина

$$d = \frac{\sum_{\mathrm{i}=2}^{\mathrm{n}} \left(\varepsilon_{t_i} - \varepsilon_{t_{i-1}}\right)^2}{\sum_{i=1}^{n} \varepsilon^2{}_{t_i}}.$$

Несложные вычисления показывают, что статистика Дарбина – Уотсона просто связана с коэффициентом автокорреляции первого порядка

$$d \approx 2(1 - r),$$

где r коэффициент автокорреляции первого порядка для остатков регрессионной модели.

Применим тест Дарбина – Уотсона для рассматриваемой задачи.

Принципы вычисления показаны на след. слайде

Схема расчета статистики Дарбина – Уотсона

t	Y	Y Регресс	Ошибка	Ошибка Лаг 1	Числит.	Знаменат.
1970	2,8	1,201	1,599			2,557
1971	3	2,001	0,999	1,599	0,360	0,997
1972	3,5	2,802	0,698	0,999	0,090	0,488
1973	4	3,602	0,398	0,698	0,090	0,158
1974	4,6	4,402	0,198	0,398	0,040	0,039
1990	18,9	17,207	1,693	1,993	0,090	2,866
1991	19,4	18,007	1,393	1,693	0,090	1,940
1992	20,1	18,808	1,292	1,393	0,010	1,670
				1,292	**14,690**	**39,109**

Таким образом, $d = \dfrac{14,69}{39,10} = 0,37.$

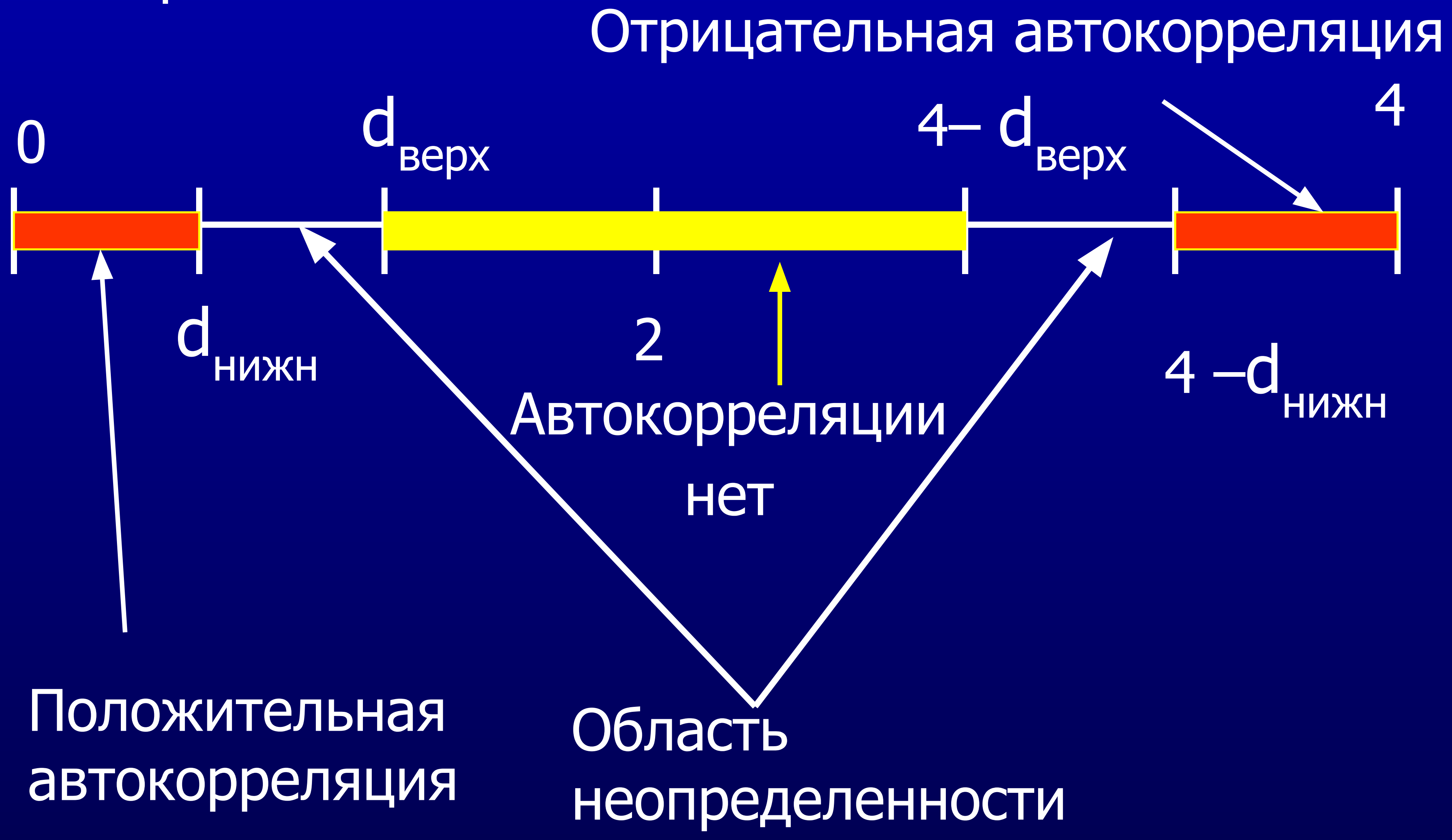
Теперь следует разобраться в каких пределах должна изменяться эта величина. Для этого нужно снова вернуться к оценке D через коэффициент корреляции

$$d \approx 2(1 - r).$$

Если корреляции нет, то $d = 2$. Если корреляция полная, то $d = 0$. Если корреляция полная и отрицательная, то $d = 4$. В нашем случае $d =$ 0,37 , что указывает на сильную положительную корреляцию.

Хотя тест Дарбина – Уотсона не является в полном смысле этого слова статистическим тестом, тем не менее для него разработаны специальные таблицы, которые для заданного уровня значимости α, числа наблюдений *n* и количества объясняющих переменных *m* дают два числа $d_{нижн}$ и $d_{верхн}$. В рассматриваемом нами примере примем уровень значимости равным 0,05, число точек наблюдения – 23, число объясняющих переменных – 1. По таблицам статистики Дарбина – Уотсона (см.. например, С. А. Бородич. Эконометрика, приложение 6) $d_{нижн}$ 1,257 и $d_{верхн}$=1,437.

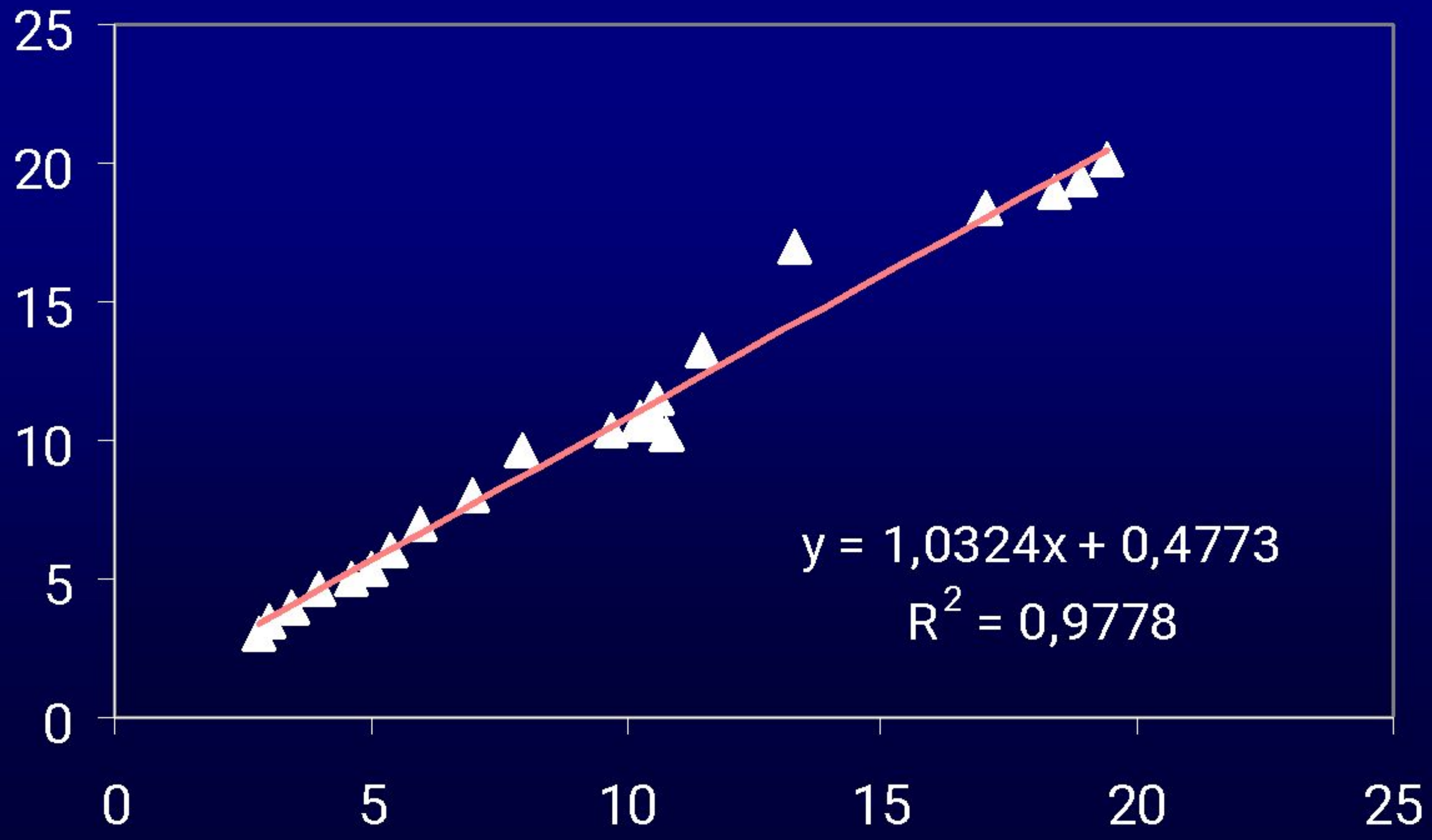
Теперь для проверки гипотезы об отсутствии автокорреляции следует обратиться к к диаграмме



В рассматриваемом случае эмпирическое значение d попадает в область сильной положительной корреляции. По этой причине использовать метод МНК не представляется возможным.

Используем авторегрессионную модель первого порядка, где в качестве регрессора выступают запаздывающие на один год значения объема производства.

Авторегрессионная модель первого порядка



На следующем слайде изображены часть исходных и расчетных данных. На основании расчетных данных можно строить и краткосрочные прогнозы динамики.

Год	Объем	Регр.	Расчет
1970	2,8	—	—
1971	3,00	2,80	3,37
1972	3,50	3,00	3,57
1973	4,00	3,50	4,09
1990	18,90	18,40	19,47
1991	19,40	18,90	19,99
1992	20,10	19,40	20,51

Эти данные позволяют построить два графика объема производства по годам эмпирический и расчетный. Обратите внимание на хорошее совпадение эмпирических и расчетных данных.

Сопоставление исходных и расчетных данных для авторегрессионной модели первого порядка