$$F = \frac{S_{\text{факт}}}{S_{\text{ост}}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}$$

Оценка значимости уравнения множественной регрессии (F-критерий): Если Fфакт > Fтабл, то признается статистическая значимость и надежность уравнения. Если Fфакт < Fтабл, то уравнение регрессии незначимо.

Дисперсионный анализ для модели с 2 факторами

	df	SS	MS	F	Значимость F
Регрессия	2	88,425	44,213	77,87	1,65E-05
Остаток	7	3,974	0,5677		
Итого	9	92,4			

частный F-критерий

- Во множественной регрессии оценивается значимость не только уравнения в целом, но и фактора, дополнительно включенного в регрессионную модель. Это связано с тем, что не каждый фактор, вошедший в модель, может существенно увеличивать долю объясненной вариации результативного признака. Ввиду корреляции между факторами значимость одного и того же фактора может быть разной в зависимости от последовательности его введения в модель. **Мерой для оценки включения фактора в модель служит частный F-критерий, т.е. F_{xi}.**

Частный F-критерий

- Частные критерии F_{x1} оценивает статистическую значимость включения фактора x_1 в уравнение множественной регрессии после другого фактора , т.е. F_{x1} оценивает целесообразность включения в уравнение x_1 после включения в него, например, фактора x_2.

Методика построения Fxj

- Частный F-критерий построен на сравнении **прироста факторной дисперсии**, обусловленного влиянием дополнительно включенного фактора, **с остаточной дисперсией на одну степень свободы по** регрессионной **модели в целом**. Предположим, что для регрессии с двумя факторами оцениваем значимость влияния X1 как дополнительно включенного в модель фактора. Используем следующую формулу:

$$\hat{y} = a + b_1 x_1 + b_2 x_2$$

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot (n - 3)$$

$$F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot (n - 3)$$

пример

- **N=23   Y=20+5X1+ 2X2   SST=1000**
  SSR=700
- **R^2=0,7**

	df	SS	MS	F
Регрес сия	2	700	350	23,3
x1	1	600	600	40
Доп. X2	1	100	100	6,67
оста ток	20	300	15	
ИТОГО	22	1000		

Fx2

- Fx2 =(0,7-0,6) x 20/0,3

=6,67

F(a=0,05; 1 и 20)=4,35

$$t_{b_i} = \sqrt{F_{x_i}}$$

Зная величину F_{xi} можно определить и t – критерий для коэффициента регрессии при i-том факторе.

связь частного Fкритерия

- Частный F-критерий связан с **частной корреляцией**. Частный F-критерий в числителе формулы содержит прирост факторной дисперсии, т.е сокращение остаточной дисперсии, которая учитывается в частной корреляции. Поэтому отсев факторов при построении модели множественной регрессии возможен при использовании как частной корреляции, так и частного F-критерия, а также **t-критерия** Стьюдента и **стандартизованных коэффициентов регрессии (β).**

Прогноз по множественной регрессии

$$\hat{y}_p = f(x_{1_p}, x_{2_p}, ..., x_{m_p})$$

$$\hat{y}_p - \Delta_{\hat{y}_p} \le \hat{y}_p \le \hat{y}_p + \Delta_{\hat{y}_p}$$

$$\hat{y}_p - \Delta_{\hat{y}_p} \le \hat{y}_p \le \hat{y}_p + \Delta_{\hat{y}_p}$$

$$\Delta_{\hat{y}_p} = t_{табл} m_{\hat{y}_p}$$

Модели на основе рядов динамики

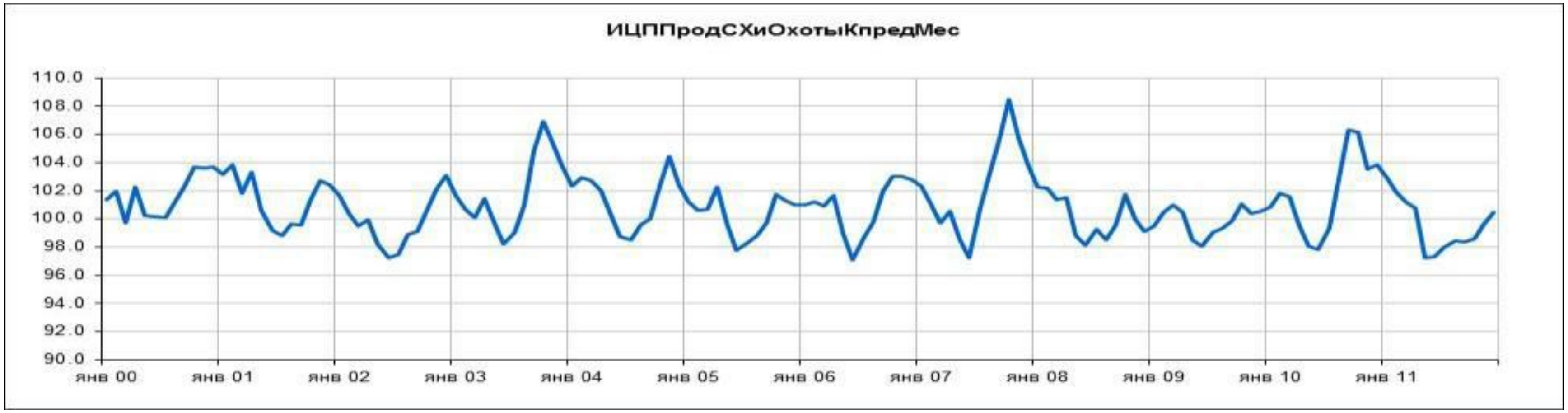
- Модели изолированного динамического ряда.
- Модели системы взаимосвязанных рядов динамики.
- Модели автрегрессии.
- Модели с распределенным лагом.

Компоненты временного ряда

- Тенденция *(T)*
- Периодические колебания *(P)*
- Случайные колебания *(E)*

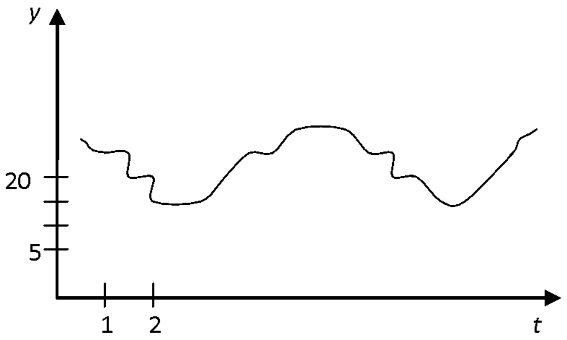
$$y_t = f(T, P, E)$$

ИЦ производителей с/х,2000-2012



Динамика инвестиций в основной капитал по РФ

- 2000-2012гг



- Ряд с периодическими и случайными колебаниями

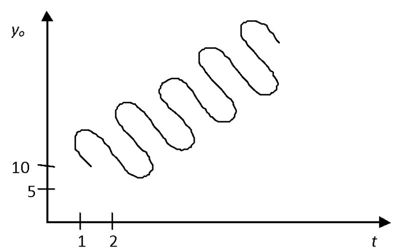
- Ряд с тенденцией, периодическими и случайными колебаниями

$$y_t = f(P, E)$$

$$y_t = f(T, P, E)$$

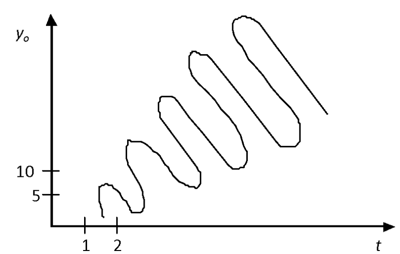
Аддитивная модель

$$y_t = T + P + E$$



Мультипликативная модель

$$y_t = T \times P \times E$$

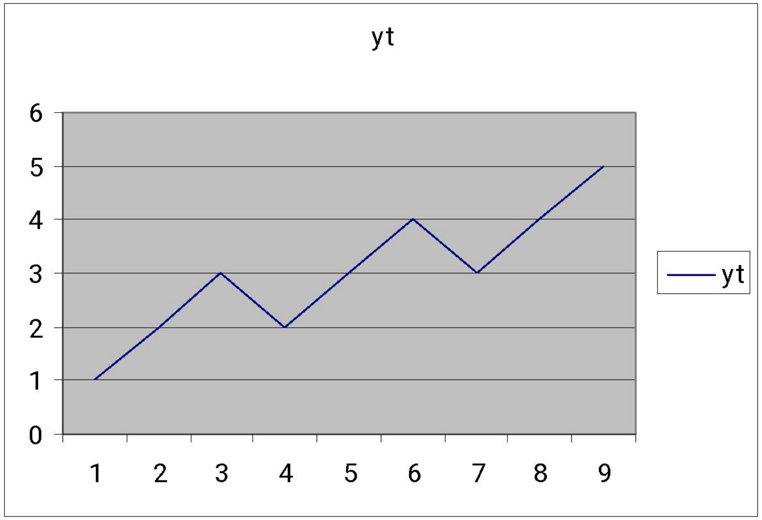


Автокорреляция уровней ряда и ее последствия

- Корреляционная зависимость между последовательными значениями уровней временного ряда называется *автокорреляцией* уровней ряда

$$r_{y_t y_{t-1}} = \frac{\overline{y_t y_{t-1}} - \overline{y_t} \cdot \overline{y_{t-1}}}{\sigma_{y_t} \sigma_{y_{t-1}}}$$

t	y_t	y_{t-1}	y_{t-2}	y_{t-3}
1	1	-	-	-
2	2	1	-	-
3	3	2	1	-
4	2	3	2	1
5	3	2	3	2
6	4	3	2	3
7	3	4	3	2
8	4	3	4	3
9	5	4	3	4



$$r_1 = 0,6; r_2 = 0,4; r_3 = 1$$

Уравнения трендов

- линейная: $y = a + bt$
- параболическая: $y = a + bt + ct^2$
- степенная: $y = at^b$
- гипербола: $y = a + \frac{b}{t}$
- показательная: $y = a \cdot b^t$
- экспонента: $y = e^{a+bt}$

Линейный тренд

$$y = a + bt$$

t	$y = a + bt$	$\Delta y = y_t - y_{t-1}$
1	$a + b$	-
2	$a + 2b$	b
3	$a + 3b$	b
4	$a + 4b$	b

Линейный тренд :Y=a+bt

- равным абсолютным приростом (параметр *b*)
- индекс потребительских цен за 12 месяцев
- $= 100{,}5 + 2t$, где $t = 1, 2, \ldots, 12$ y_{1}**=102,5;** y_{12}**=124,5**

Парабола 2-го порядка

t	$y_t = a + bt + ct^2$	Скорость $\Delta y_t = y_t - y_{t-1}$	Ускорение $\Delta_t = \Delta_t - \Delta_{t-1}$
1	$a + b +$	-	-
2	$a + 2b + 4$	$b + 3$	-
3	$a + 3b + 9c$	$b + 5$	$2c$
4	$a + 4b + 16c$	$b + 7$	$2c$

Парабола :$Y=a+bt+ct^2$

- постоянное абсолютное ускорение(Δ_2)
- параметр «а»−У при t=0
- Параметр «с»=0,5(Δ_2)

численность детей в возрасте 7 лет за 15 лет

- **$Y=323.7+10.8t-1.6t^2$,** где *y* – тыс. чел., *t* = 1, 2,…, 15.
- ежегодно численность детей сокращалась в среднем с ускорением в 3,2 тыс. чел.

Показательная функция

$$y = a \cdot b^t$$

t	$y = a \cdot b^t$	$K = y_t / y_{t-1}$
1	ab	-
2	ab^2	b
3	ab^3	b
4	ab^4	b

Показательная функция

- $Y=ab^t$
- стабильный коэффициент роста (b)
- $Y=13.5*1.5^t$
- $Y=13.5e^{0.405t}$ – экспонента

Степенной тренд

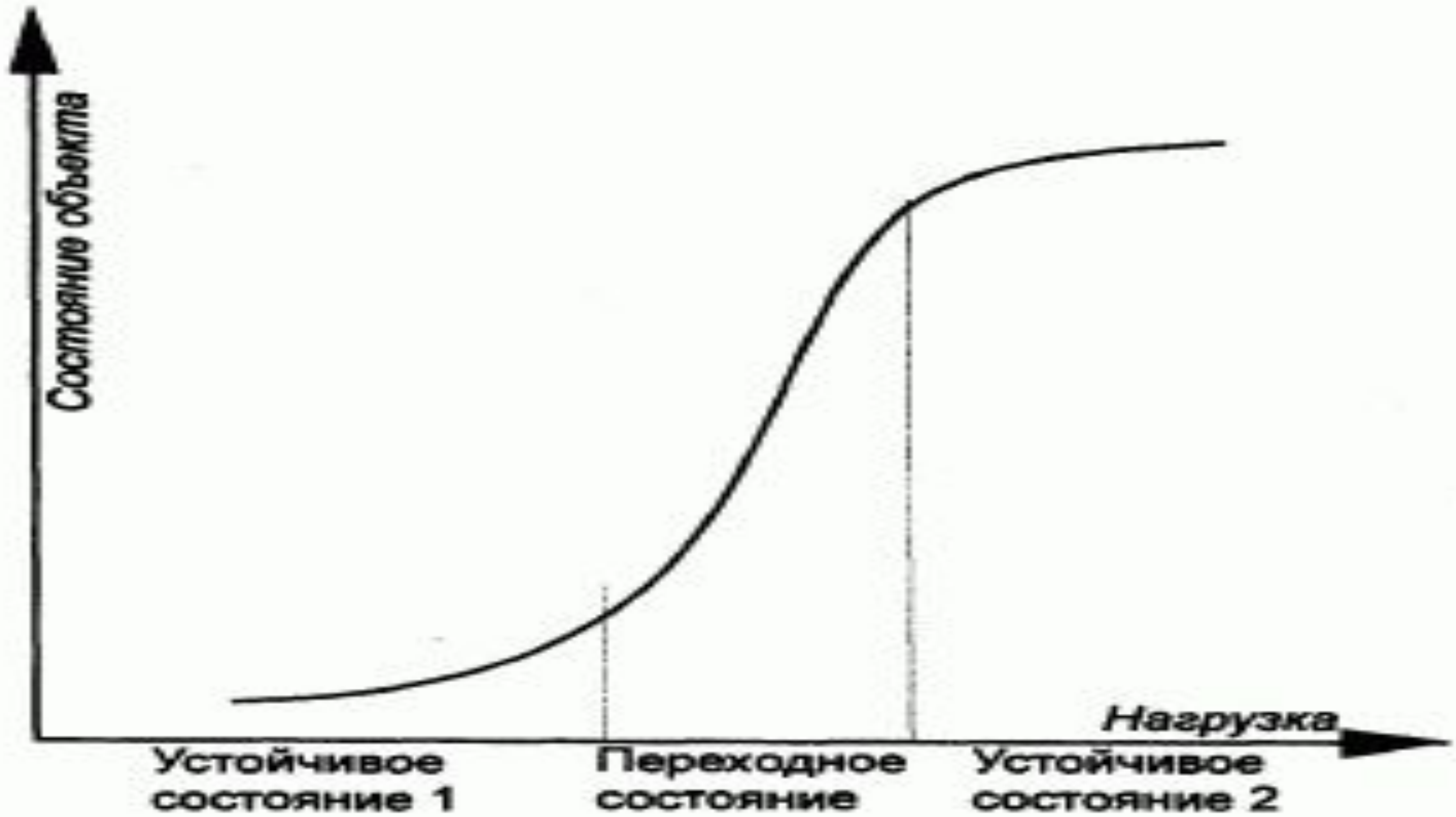
$$y = at^b$$

- При $b > 0$ она характеризует непрерывный рост уровней с падающими темпами роста, а при $b < 0$ – их ускоренное снижение. Величина t^b означает базисный коэффициент роста

t	$y = at^b$	Базисный коэффициент роста
1	a	1
2	$a2^b$	2^b
3	$a3^b$	3^b
4	$a4^b$	4^b
5	$a5^b$	5^b

Равносторонняя гипербола

- при $b > 0$ означает, что уровни ряда снижаются во времени и асимптотически приближаются к параметру a.Так, выручка предприятия за 7 месяев
- У=400+85/t ,т.е. падающая тенденция, при которой У не может быть меньше 400. Если $b < 0$, то уравнение тренда характеризует тенденцию к росту уровней ряда с асимптотической границей равной параметру "a". Так, У=500-20/ t



Состояние объекта
Нагрузка
Устойчивое состояние 1
Переходное состояние
Устойчивое состояние 2

Оценка параметров уравнения тренда

- При использовании полиномов разных степеней оценка параметров уравнения тренда производится методом наименьших квадратов (МНК) точно также, как оценки параметров уравнения регрессии на основе пространственных данных. В качестве зависимой переменной -уровни динамического ряда, а в качестве независимой переменной – фактор времени t, который обычно выражается рядом натуральных чисел: 1, 2,…, n.
-

нелинейные функции тренла

- Оценка параметров нелинейных функций проводится МНК после линеаризации, т. е. приведения их к линейному виду.

Показательная функция

- Для оценки параметров показательной кривой Y=ab^t или экспоненты Y=ae^bt путем логарифмирования функции приводятся к линейному виду и применяется МНК к ln Y и t
- Число зарегистрированных ДТП (на 100000 человек населения) по области за 2005-2013 годы характеризуется данными:105,7; 105,3; 156; 158,1; 160,1; 178; 191,5; 274,6; 287,3.

- Для построения системы нормальных уравнений были рассчитаны вспомогательные величины:ln Y
- получим: ln Y= 4,517598 + 0,123523t, где 4,517598= lna   0,123523=lnb $a = e^{4,5176}$ = 91,61524 $b = e^{0,12352}$ = 1,131476 Соответственно, имеем экспоненту $y=91,615e^{0,1235t}$
- или показательную кривую: $Y=91,615*1,1315^{t}$. Число ДТПвозрастало в среднем ежегодно на 13,5%.

Использование трендовых моделей для прогнозирования

$$S_{e(\hat{y}_p)} = \sqrt{MS_{\text{ост}}\left[1 + \frac{1}{n} + \frac{\left(t_p - \bar{t}\right)^2}{\sum\left(t - \bar{t}\right)^2}\right]}$$

$$MS_{\text{ост}} = \frac{\sum\left(y - \hat{y}\right)^2}{n - m - 1}$$

$$\hat{y}_p - t_{\text{òàáë}} \cdot Se_{(\hat{y}_p)} \le \hat{Y}_p \le \hat{y}_p + t_{\text{òàáë}} \cdot Se_{(\hat{y}_p)}$$

±

$Y=13.028+3.0167t$

- $t=1.2\ldots.9$мес. t_p =10 $У_p$=43,19
- $\sqrt{MS_{ост}}=(14,87/7)^{0,5}=1,4576$-стандартная ошибка регрессии;
- $Q=(1+(1/9)+(10-5)^2/60)^{0,5}=1.236$
- ±
- S_p =1,4576* 1.236=1.801- ошибка прогноза
- $a=0.05$; df=7 ; t_a = 2,365 ;
- Δ_p=2,365*1,801=4,26-предельная ошибка прогноза
- 43,19 ± 4,26 , т .е интервал от 38,9 до 47,4.

Оценка адекватности модели тенденции

- Модель тенденции считается адекватной реальному процессу, если теоретические (найденные по уравнению тренда) уровни ряда достаточно близко подходят к фактическим их значениям. Для оценки адекватности модели проводится анализ остатков . Модели тенденции можно сравнивать по величине остаточной суммы квадратов:$S^2=\sum(Y - Yteor)^2$. Чем меньше эта величина, тем в большей мере уравнение тренда подходит для описания тенденции временного ряда.

Предположим, что было рассчитано

- уравнение линейного тренда и экспоненциального тренда. Для линейного тренда остаточная сумма квадратов составила 3874,62, а для экспоненты 2617,701. Следовательно, экспонента лучше описывает тенденцию ряда.
- Другим показателем при выборе функции тренда является коэффициент детерминации R^2. Чем выше R^2, тем соответственно выше вероятность того, что данная модель тенденции описывает исходные данные. В примере R^2 для экспоненты составил 0,9202, а для линейного тренда 0,8832, подтверждая еще раз, что экспонента в большей мере подходит для описания тенденции.

Автокорреляция в остатках

$$r_{a_e} = \frac{\overline{e_t e_{t-1}} - \overline{e_t} \cdot \overline{e_{t-1}}}{\sigma_{e_t} \cdot \sigma_{e_{t-1}}}$$

$$r_{a_e} = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

- автокорреляция в остатках оценивается также, как и автокорреляция уровней ряда с тем лишь отличием, что в расчетах используются остаточные величины , а не уровни динамического ряда .Пусть коэффициент автокорреляции остатков оказался равным 0,627. Его величина не столь мала, чтобы утверждать об отсутствии автокорреляции остатков. Очевидно уравнение тренда не является наилучшим, ибо нарушена предпосылка МНК об отсутствии автокорреляции остатков.

- Уравнение тренда хорошо описывает тенденцию, если остатки текущего периода не коррелируют с остатками предыдущего периода.
- Проверка модели на автокорреляцию остатков обычно проводится с помощью критерия Дарбина-Уотсона.

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

$$d \cong 2(1 - r_{a_e})$$

$$0 \le d \le 4$$

Границы критерия Дарбина-Уотсона

- При полной положительной автокорреляции остатков ($\rho=1$) критерий d=0, а при полной отрицательной автокорреляции ($\rho=-1$) критерий d=4. Если же автокорреляция в остатках отсутствует, т. е. $\rho=0$, то d=2. Иными словами критерий Дарбина-Уотсона изменяется в пределах:
- $0 \le d \le 4$.

- Дарбин и Уотсон разработали пороговые значения показателя d, позволяющие принять или отвергнуть гипотезу об отсутствии автокорреляции в остатках.
- При заданном числе наблюдений n (длина динамического ряда) и m параметров при t в уравнении тренда (или m объясняющих переменных в уравнении регрессии) установлены при 5%-ом уровне значимости верхняя (u – upper) и нижняя (ℓ - low) границы критерия.

сравнение с табличными значениями

- **Если d<2**, то возможны следующие варианты:
- 1) при d< нижней границы делается вывод о наличии положительной автокорреляции в остатках;
- 2)при d›верхней границы делается вывод об отсутствии корреляционной связи последующих остатков с предыдущими;
- 3) при d между нижней и верхней границами нельзя ни отвергнуть, ни принять нулевую гипотезу об отсутствии автокорреляции в остатках т. е. значение d попало в область неопределенности и необходимы дальнейшие исследования, например, по большему числу наблюдений.

фактическое значение d › 2

- означает отрицательную автокорреляцию, то с пороговыми табличными значениями сравнивается величина 4-d. При этом возможны следующие варианты:
- 1) 4-d ‹ нижней границы: делается вывод о наличии отрицательной автокорреляции в остатках;
- 2)4-d › верхней границы: отсутствует автокорреляция в остатках;
- 3) 4-d между нижней и верхней границами: нельзя сделать определенного вывода о наличии или отсутствии автокорреляции в остатках по имеющимся данным

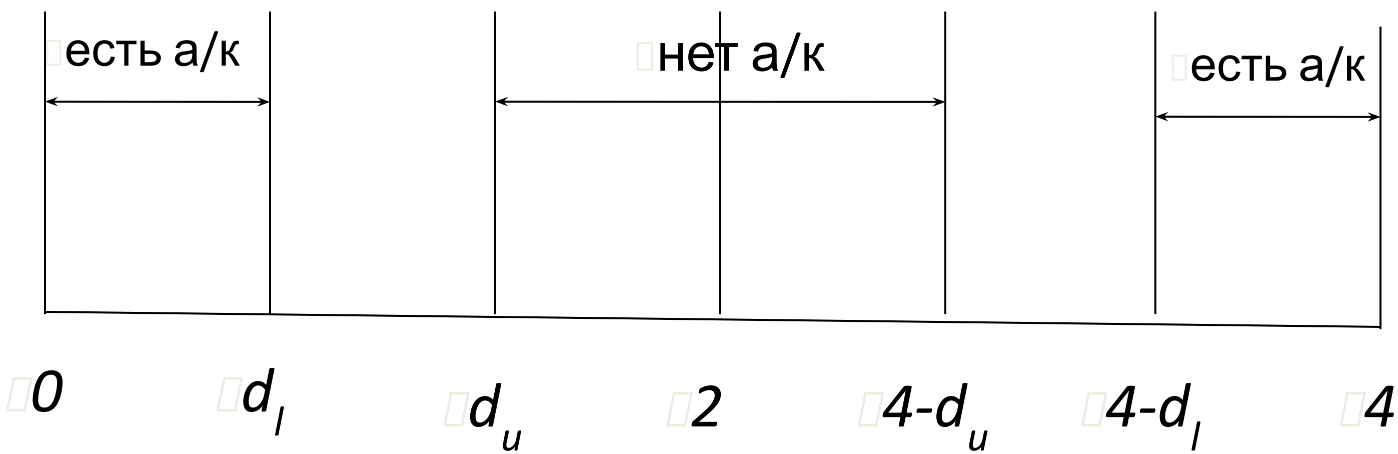
- По величине критерия Дарбина-Уотсона можно определить величину коэффициента автокорреляции остатков, исходя из соотношения: $d \approx 2(1-\rho)$. Отсюда $0{,}5d \approx 1-\rho$ и соответственно $\rho \approx 1-0{,}5d$.
- Поэтому, если $d > 2$, то $\rho < 0$, а при $d < 2$ $\rho > 0$. Таким образом, если фактическое значение критерия Дарбина-Уотсона не слишком отличается от 2, то можно сделать вывод об отсутствии автокорреляции в остатках.

Пример $\hat{y}_t = 5,857 + 1,07t$

t	y_t	$\hat{y}_t$	e_t	e_{t-1}	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$	$e_t^{\ 2}$
1	7	6,928571	0,071429	-	-	-	0,005102
2	8	8	0	0,071429	-0,07143	0,005102	0
3	10	9,071429	0,928571	0	0,928571	0,862245	0,862245
4	9	10,14286	-1,14286	0,928571	-2,07143	4,290816	1,306122
5	11	11,21429	-0,21429	-1,14286	0,928571	0,862245	0,045918
6	12	12,28571	-0,28571	-0,21429	-0,07143	0,005102	0,081633
7	14	13,35714	0,642857	-0,28571	0,928571	0,862245	0,413265
Итого	X	X	X	X	X	6,887755	2,714286

$$d = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^{\ 2}} = \frac{6,888}{2,714} = 2,538$$

$$d = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2} = \frac{6,888}{2,714} = 2,538$$



табличные значения 0,7 и 1,36 4-d=1.462

Аддитивная модель с тенденцией с фиктивными переменными

- Аддитивная модель уровней динамического ряда при наличии тенденции и сезонности может быть построена как модель регрессии с включением в нее фактора времени (*t*) и фиктивных переменных (*z*).
- При квартальном разрезе информации модель примет вид:

$$y_t = a + bt + c_1 z_1 + c_2 z_2 + c_3 z_3 + \varepsilon_t$$

Аддитивная модель при наличии тенденции

данные за 3 года о численности безработных

$$\hat{y}_t = 12{,}417 - 0{,}344\,t - 2{,}031\,z1 - 3{,}688\,z2 - 5{,}010\,z3$$

(t) *38,5* *-11* *-6,7* *-12,5* *-17,3*

$R^2 = 0{,}984$ $F = 108{,}25$

- Параметр "b" = -0,344 указывает на тенденцию снижения уровней ряда при элиминировании сезонности. Его величина по содержанию и численно практически совпадает с величиной параметра "b" в уравнении тренда по данным с устранением сезонности, найденным ранее.
- Иными словами, ежеквартально независимо от сезона уровни ряда снижаются в среднем на 0,34 тыс. чел.
- Параметры c_1, c_2, c_3 показывают, что в I, II и III кварталах уровни ряда независимо от влияния тенденции были в среднем ниже, чем в четвертом квартале на соответствующие величины. Параметр "a" = 12,417 характеризует уровень IV квартала 2012 г. вместе с сезонной компонентой.

прогноз

- Прогноз по данной модели на I квартал 2015 г. составит 5,914 тыс. чел.:
- Ур= 12,417-0,344 х13 -2,031 х1=5,914 тыс. чел.

МОДЕЛИ РЕГРЕССИИ ПО ВРЕМЕННЫМ РЯДАМ

- **Специфика изучения взаимосвязей по рядам динамики**
- Временные ряды как источник информации накладывают отпечаток на методологию построения регрессионных моделей .Это связано с возможной ***ложной корреляцией*** *и ложной регрессией.* Высокая корреляция между уровнями временных рядов может иметь место и при отсутствии реальной связи между явлениями.

- Если ряды динамики характеризуются наличием тренда, то при построении модели регрессии надо учесть тренд ,

например, исключить его. В противном случае корреляция уровней рядов динамики будет преувеличена (коэффициент корреляции будет близок к +1 при одинаковой тенденции в рядах и к -1 - при противоположной тенденции).

- Если ряды динамики характеризуются не только тенденцией, но и периодическими колебаниями, то при построении модели регрессии следует учесть обе компоненты динамических рядов. В этом случае можно из первоначальных данных исключить как тенденцию, так и периодическую составляющую. Модель регрессии может быть построена либо по остаточным величинам, либо с включением в нее обоих компонент динамического ряда наряду с экономическими переменными.

- Однако можно строить регрессию и по уровням рядов динамики, если удается при этом устранить автокорреляцию в остатках, применяя, например, обобщенный метод наименьших квадратов. Устранение автокорреляции в остатках возможно также путем изменения спецификации модели, включая, например, в правую часть модели регрессии лагированные (запаздывающие переменные, например, прибыль не только текущего года, но и предыдущих лет).

Методы учёта тенденции при моделировании взаимосвязей по временным рядам

- *Метод отклонений от тренда*
- *Метод последовательных разностей*
- *Включение в модель регрессии по временным рядам фактора времени*

Метод отклонений от тренда

$$e_{y_t} = y_t - \hat{y}_t$$

$$e_{x_t} = x_t - \hat{x}_t$$

$$e_{y_t} = a + b \cdot e_{x_t}$$

Метод последовательных разностей

$$\Delta_{y_t} = y_t - y_{t-1}$$

$$\Delta_{x_t} = x_t - x_{t-1}$$

$$\Delta_{y_t} = a + b \cdot \Delta_{x_t}$$

Включение в модель регрессии по временным рядам фактора времени

$$y_t = a + bx_t + ct$$

$$y_t = a + b_1x_1 + b_2x_2 + b_3x_3 + ct$$

Учет сезонности при построении модели регрессии

$$y_t = a + bx_t + c_1 z_1 + c_2 z_2 + c_3 z_3$$

- $z_1 = 1$ – для первого квартала,
- *0* – для остальных;
- $z_2 = 1$ – для второго квартала,
- *0* – для остальных;
- $z_3 = 1$ – для третьего квартала,
- *0* – для остальных.

- Пример. По промышленному предприятию имеются данные за 3 года в поквартальном разрезе об уровне производительности труда (*y*, в тыс.руб. на одного работника) и доле активной части основных фондов (*x*, в %):

t	1	2	3	4	5	6	7	8	9	10	11	12
y	5	6	6	7	8	10	11	11	13	12	13	15
x	9,9	18,9	19,8	27,9	22,2	29,7	38,7	36	46	37,8	45	54

- Модель регрессии с включением в нее фактора времени *t*
- оказалась следующей:

$$y_t = 2{,}943 + 0{,}104x + 0{,}533t$$

t -критерий 5,47 *2,43* 3,44

- В модели параметр *b*=0,104 показывает, что рост доли активной части основных фондов на 1 процентный пункт в условиях неизменной тенденции способствует росту уровня производительности труда на 0,104 тыс.руб. Параметр *с* характеризует среднеквартальный прирост производительности труда независимо от изменения доли активной части основных фондов, т. е. обусловленный влиянием других факторов, не учитываемых в регрессии.

год	квартал	t	y_t	x_t	z_1	z_2	z_3
2006	1	1	11	9	1	0	0
	2	2	15	10	0	1	0
	3	3	6	8	0	0	1
	4	4	12	9	0	0	0
2007	1	5	11	10	1	0	0
	2	6	16	9	0	1	0
	3	7	4	3	0	0	1
	4	8	13	11	0	0	0
2008	1	9	10	7	1	0	0
	2	10	14	10	0	1	0
	3	11	7	8	0	0	1
	4	12	12	11	0	0	0
2009	1	13	10	12	1	0	0
	2	14	16	9	0	1	0
	3	15	8	11	0	0	1
	4	16	13	12	0	0	0
2010	1	17	11	8	1	0	0
	2	18	18	16	0	1	0
	3	19	7	6	0	0	1
	4	20	12	12	0	0	0

ВЫВОД ИТОГОВ

$$\hat{y}_t = 8{,}896 + 0{,}319x_t - 1{,}227z_1 + 3{,}464z_2 - 4{,}789z_3$$

рессионная ст ат ист ика	
Множестве	0,977311
R-квадрат	0,955137
Нормиров:	0,943173
Стандартн	0,865054
Наблюден	20

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*ıачимост ь F*
Регрессия	4	238,9752	59,74381	79,83739	6,32E-10
Остаток	15	11,22478	0,748319		
Итого	19	250,2			

	Коэффициент	*дарт ная ош*	*т ат ист ик*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересеч	8,895575	1,072979	8,290537	5,53E-07	6,608574	11,18258
xt	0,318584	0,090983	3,501588	0,003213	0,124659	0,512509
z1	-1,22655	0,571093	-2,14772	0,048484	-2,44381	-0,00929
z2	3,463717	0,547411	6,327455	1,36E-05	2,296938	4,630495
z3	-4,78938	0,647194	-7,40023	2,22E-06	-6,16884	-3,40992