ЭКОНОМЕТРИКА

Лекции – 8 ч.

Практические занятия на ПЭВМ – 8 ч.

Контрольные работы 1,2 (зачет)

Экзамен

Кафедра Математики и информатики, к. 701

Консультации: пятница 17-00

Рекомендуемая литература

Эконометрика. Учебник (под ред. И.И.Елисеевой)

Практикум по эконометрике (под ред. И.И.Елисеевой)

Орлова И.В., Половников В.А. Экономико-математические методы и модели: компьютерное моделирование

Эконометрика. Методические указания по изучению курса и выполнению контрольной работы (**521600 «Экономика» бакалавр**): Москва – 2008

Кайгородова М.А., Поддубная М.Л. Эконометрика. Методические указания по решению задач и выполнению контрольной работы (**для студентов бакалавриата и 2-го высшего образования**): Барнаул – 2011

Поддубная М.Л., Кайгородова М.А. Основы эконометрики. Лекции в презентациях (для студентов бакалавриата и 2-го высшего образования): Барнаул – 2011

Учебно-методические материалы в электронном виде, размещенные на сервере филиала

Тема 1.
Введение. Эконометрика и эконометрическое моделирование: основные понятия и определения

План:

- Классификация эконометрических моделей;
- Основные этапы построения эконометрических моделей;
- Типы экономических данных.

Цель: дать обзор методов построения эконометрических моделей и их применения для анализа экономических и социальных систем

Задачи:

- рассмотреть общую постановку эконометрической задачи, этапы и методы ее решения;
- способствовать формированию научного подхода к решению экономических задач.

Литература:

Орлова И.В., Половников В.А. Экономико-математические методы и модели: компьютерное моделирование (глава 3): Учеб. пособие – М.: Вузовский учебник, 2011.

Эконометрика – даёт количественное выражение взаимосвязей экономических явлений и процессов – эмпирический вывод экономических законов.

Основные задачи эконометрики:

- построение количественно определённых экономико-математических моделей,
- разработка методов оценки их параметров по статистическим данным,
- анализ свойств построенных моделей и прогнозирование на их основе экономических процессов.

Выделяют **три класса моделей**, которые применяются для анализа и прогнозирования экономических процессов:

- модели временных рядов,
- регрессионные модели с одним уравнением,
- системы одновременных уравнений.

По способу вхождения в модель все переменные можно разбить на

- ***объясняемые -*** зависимые, исследуемые
- ***объясняющие -*** предопределённые, факторные.

Основные этапы эконометрического моделирования:

- ***Спецификация модели*** (этап подготовки)

- формулируются конечные цели моделирования, определяется наборы возможных исследуемых и факторных переменных;
- производится выбор общего вида модели– ***парной*** или ***множественной регрессии***; линейной или нелинейной;
- проводится сбор и предварительный анализ данных - проверка аномальных значений показателей, сглаживание, тестирование на наличие тенденции исследуемых показателей к изменению.

- ***Идентификация модели*** - построение модели.

- ***Верификация*** - проверка адекватности и качества модели, оценка точности модельных данных

- ***Интерпретация*** результатов - анализ, прогнозирование

В эконометрике выделяют **три типа данных**:

- ***Кросс секционные (перекрёстные)*** данные представляют ситуацию в группе переменных в отдельный момент времени: публикуемые в разделах газет списки цен на различные акции, процентные ставки по разным видам вкладов и обменные курсы разных валют.

- ***Пространственные данные*** характеризуют ситуацию по конкретной переменной (или набору переменных), относящейся к пространственно разделённым однотипным объектам в один момент времени: данные о курсах валют в один день по разным обменным пунктам города.

- ***Временные ряды*** отражают изменения (динамику) какой-либо переменной на промежутке времени: данные об обменном курсе валюты за каждый день в конкретном обменном пункте

Этап 1. Спецификация модели

Этап подготовки – ***выбор системы факторов*** для моделирования.

Цель исследования определяет **результирующую** (объясняемую) переменную: **Y**

Влияние на Y оказывают **факторные** (объясняющие) переменные: $\mathbf{X} = \{X_1, X_2, \ldots, X_p\}$.

Например:

- зависимость объемов продаж от цены на продукцию;
- зависимость цены от спроса;
- зависимость цены (на квартиру) от жилой площади, района города, типа дома;
- динамика изменения цен (зависимость цен от времени); …

Основные задачи спецификации:

1. количественная ***оценка тесноты*** взаимосвязи между результирующим признаком и рассматриваемыми факторами;
2. ***проверка достоверности*** наличия этой взаимосвязи;
3. ***выявление мультиколлинеарности*** факторов.

Эти задачи решаются с помощью **корреляционного анализа**.

Выделяют **два вида зависимостей**:

Функциональные связи характеризуются полным соответствием между изменением факторного признака (признаков) и исследуемого показателя.

В ***корреляционных связях*** между изменением факторного и результативного признаков нет однозначного соответствия, воздействие факторов проявляется лишь в среднем при многократном наблюдении фактических данных.

Спецификация модели: **оценка тесноты зависимости между переменными**

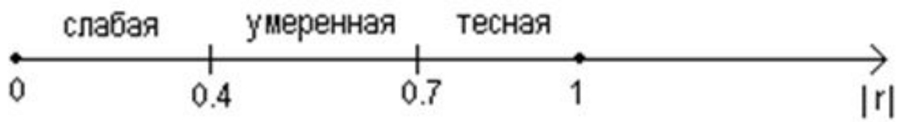
Показателем тесноты линейной взаимосвязи между переменными Y и X является выборочный ***коэффициент парной корреляции***

$$r = r(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

в Excel: данные / анализ данных / КОРРЕЛЯЦИЯ.

Свойства коэффициента корреляции:

1. $|r| \leq 1$;
2. если $\mathbf{r > 0}$, то говорят о *прямой* корреляционной зависимости между X и Y: при увеличении X величина Y имеет тенденцию в среднем возрастать;
3. если $\mathbf{r < 0}$, то говорят об *обратной* корреляционной зависимости между X и Y: при увеличении X величина Y имеет тенденцию в среднем уменьшаться;
4. если $\mathbf{r = 0}$, то между X и Y линейная корреляционная зависимость отсутствует;
5. если $|r| = 1$, то переменные X и Y связаны строгой функциональной зависимостью;
6. в зависимости от того, насколько $|r|$ приближается к $\mathbf{1}$, различают три уровня корреляционной зависимости:



Оценка тесноты зависимости между переменными

Пример В таблице представлены данные о средних доходах населения Y (тыс. $), уровне аграрности X_1 (% работников с/х) и уровень образованности населения X_2 (число лет, проведенных в учебных заведениях) для 15 развитых стран в 1983 г.

страна	Y	X1	X2
1	7	8	9
2	9	9	13
3	9	7	11
…	…	…	…
15	12	8	13

Требуется ***проанализировать направление и тесноту связи*** между средним доходом Y и факторами аграрности и образованности (X_1 и X_2).

Выбрать ***наиболее информативный фактор***.

Решение

С помощью программы «КОРРЕЛЯЦИЯ» получим матрицу коэффициентов парной корреляции между всеми имеющимися переменными:

Решение

С помощью программы «КОРРЕЛЯЦИЯ» получим матрицу коэффициентов парной корреляции между всеми имеющимися переменными:

	Y	X1	X2
Y	1		
X1	-0,47	1	
X2	0,69	-0,12	1

$r_1 = r(Y, X_1) = -0,47 < 0$, следовательно, между переменными Y и X_1 наблюдается обратная корреляционная зависимость: среднедушевой доход ниже для более «аграрных» стран.

$|r_1| = 0,47 \in (0,4;\ 0,7)$ – эта зависимость умеренная, ближе к слабой.

$r_2 = r(Y, X_2) = 0,69 > 0$, значит, между переменными Y и X_2 наблюдается прямая корреляционная зависимость: среднедушевой доход выше для более «образованных» стран.

$|r_2| = 0,69 \in (0,4;\ 0,7)$ – эта зависимость умеренная, ближе к тесной.

Более информативным является фактор образованности X_2, т.к. $|r_2| = 0,69 > |r_1| = 0,47$.

Наиболее информативным является фактор, для которого выборочный коэффициент парной корреляции r(Y,Xi) наибольший

Оценка значимости выборочного коэффициента корреляции

Это ответ на вопросы:

– ***закономерно ли отличие*** выборочного коэффициента корреляции $\boldsymbol{r}$ от нуля;

– ***достоверна ли зависимость*** между признаками.

Проверка выполняется с помощью ***t-критерия Стьюдента***.

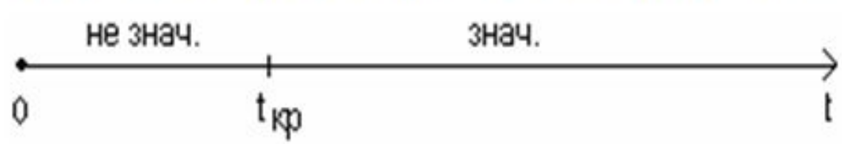
Алгоритм проверки:

1. для каждого проверяемого коэффициента корреляции $\boldsymbol{r}$ *вычисляют* ***t****-статистику*

$$t(r) = \sqrt{\frac{r^2(n-2)}{1-r^2}}$$, где n – количество уровней исходных данных;

2. по таблице критических точек распределения Стьюдента для случая двустороннего ограничения при заданном уровне значимости α и числе степеней свободы $k = n - 2$ *определяют* ***критическое значение*** $t_{кр}$ (в Excel функция СТЬЮДРАСПОБР);

3. *сопоставляют фактическое* значение $\boldsymbol{t}$ *с критическим* $\boldsymbol{t_{kp}}$, делают вывод:



Пример На 5%-ом уровне оценить значимость коэффициентов корреляции, характеризующих зависимость между средним доходом и факторами аграрности и образованности.

Решение

Используем критерий Стьюдента.

Рассчитаем *t-статистики* для каждого коэффициента корреляции:

$$t_1 = t(-0{,}47) = \sqrt{\frac{0{,}47^2 \cdot 13}{1-0{,}47^2}} = 1{,}90\ ; \qquad t_2 = t(0{,}69) = \sqrt{\frac{0{,}69^2 \cdot 13}{1-0{,}69^2}} = 3{,}41\ .$$

Критическое значение для уровня значимости $\alpha = 5\% = 0{,}05$ и числа степеней свободы $k = 15-2 = 13$ составляет $t_{кр} = 2{,}16$.

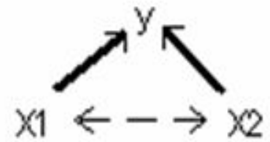
$t_1 = 1{,}90 < t_{кр} = 2{,}16$, следовательно, коэффициент $r(Y, X_1)$ не является значимым, т.е. на основании выборочных данных нельзя утверждать, что зависимость дохода Y от уровня аграрности X_1 является достоверной.

$t_2 = 3{,}41 > t_{кр} = 2{,}16$, следовательно, коэффициент $r(Y, X_2)$ значимо отличается от нуля, т.е. на уровне значимости 5% выборочные данные позволяют сделать вывод о наличии линейной корреляционной связи между признаками Y и X_2.

Таким образом, целесообразно рассмотреть модель зависимости среднедушевого дохода Y от фактора образованности X_2: $Y_T = f(X_2)$.

Мультиколлинеарность факторов

– это ***тесная*** корреляционная зависимость ***между факторными*** переменными, наличие которой ухудшает качество модели.



Для выявления м/к используют коэффициенты корреляции $r(X_i, X_j)$.

Если $|r(X_i, X_j)| > 0{,}8$, то м/к между факторами X_i и X_j считается *сильной*. В модель целесообразно включить только один из них, более информативный.

Если $|r(X_i, X_j)| < 0{,}8$, то сильной м/к между X_i и X_j нет; возможна м/к в слабой форме или ее отсутствие.

Если $|r(X_i, X_j)| < 0{,}8$ и, кроме того, выполняются два дополнительных условия:

$$\begin{cases} |r(X_i, X_j)| < |r(Y, X_i)| \\ |r(X_i, X_j)| < |r(Y, X_j)| \end{cases},$$

то м/к между факторами X_i и X_j отсутствует. В модель можно включать оба фактора.

Пример Оценить наличие мультиколлинеарности между факторами аграрности и образованности.

Решение

Из матрицы коэффициентов парной корреляции

	Y	*X1*	*X2*
Y	1		
X1	-0,47	1	
X2	0,69	-0,12	1

выпишем $\boldsymbol{r(X_1, X_2) = -0{,}12}$

Сравним:

$|r(X_1, X_2)| = 0{,}12 < 0{,}8 \Rightarrow$ сильной м/к между факторами X_1 и X_2 нет.

Кроме того, выполняются неравенства

$$|r(X_1, X_2)| = 0{,}12 < |r(Y, X_1)| = 0{,}47$$
$$|r(X_1, X_2)| = 0{,}12 < |r(Y, X_2)| = 0{,}69$$

Следовательно, между факторами аграрности и образованности м/к отсутствует, возможно построение двухфакторной модели $Y_T = f(X_1, X_2)$.

Этап 2. Идентификация эконометрической модели

Уравнение регрессии: $Y = Y_T + \varepsilon$

Y – фактическое значение результирующего показателя;

$Y_T = f(X_1, X_2, ..., X_p)$ – теоретическое (расчетное) значение этого показателя, выражающее зависимость результирующего признака Y от факторных переменных X;

ε – случайная величина, характеризующая отклонение фактического значения Y от расчетного Y_T из-за неучтенных в модели факторов и случайных ошибок.

Таким образом, ***эконометрическая модель*** имеет вид $Y = f(X_1, X_2, ..., X_p) + \varepsilon$, т.е.

Наблюдаемое значение зависимой переменной Y	=	Объясненная часть, зависящая от значений объясняющих переменных	+	Случайная составляющая

По типу функции f различают ***линейные*** и ***нелинейные*** модели.

По количеству включенных в модель факторов ***парные*** (*однофакторные*) и ***множественные*** (*многофакторные*) модели.

Наиболее часто используются линейные зависимости:

$Y_T = a + b \cdot X$ - парная линейная модель;

$Y_T = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + ... + b_p \cdot X_p$ - множественная линейная модель.

Идентификация модели состоит в *вычислении коэффициентов* выбранной зависимости

Оценка параметров линейной парной модели с помощью МНК

Основой *регрессионного анализа* является ***метод наименьших квадратов*** (МНК).

Расчетные формулы МНК следуют из условия минимизации функции ошибок:

$$\sum (y_i - y_{Ti})^2 = \sum E_i^2 \to \min$$

В случае *парной линейной модели* $Y_T = a + b \cdot X$ справедливы формулы

$$b = \frac{\sum (y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}; \qquad a = \bar{y} - b \cdot \bar{x}$$

В случае *множественной линейной модели* $Y_T = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \ldots + b_p \cdot X_p$

матричная формула $b = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$

Универсальным способом построения линейных моделей (как парных, так и множественных) является в Excel программа РЕГРЕССИЯ (данные / анализ данных / РЕГРЕССИЯ).

Опр. Коэффициенты b_j называются выборочными ***коэффициентами регрессии*** и показывают среднее изменение результата Y при увеличении соответствующего фактора X_j на 1 единицу.

Коэффициент a (или b_0) – ***свободный коэффициент***, который формально дает прогнозное значение Y при нулевых значениях факторных переменных (если это имеет смысл в данной задаче).

Парная модель

Пример Построить линейную модель зависимости среднедушевого дохода Y от уровня образованности X_2, объяснить смысл коэффициентов модели, исходные данные и результаты моделирования показать на чертеже.

Решение

Построим парную линейную модель вида $\boldsymbol{Y_T = a + b \cdot X}$.

Используем программу РЕГРЕССИЯ и найдем коэффициенты

	Коэффициенты
Y-пересечение	1,871194
X2	0,64637

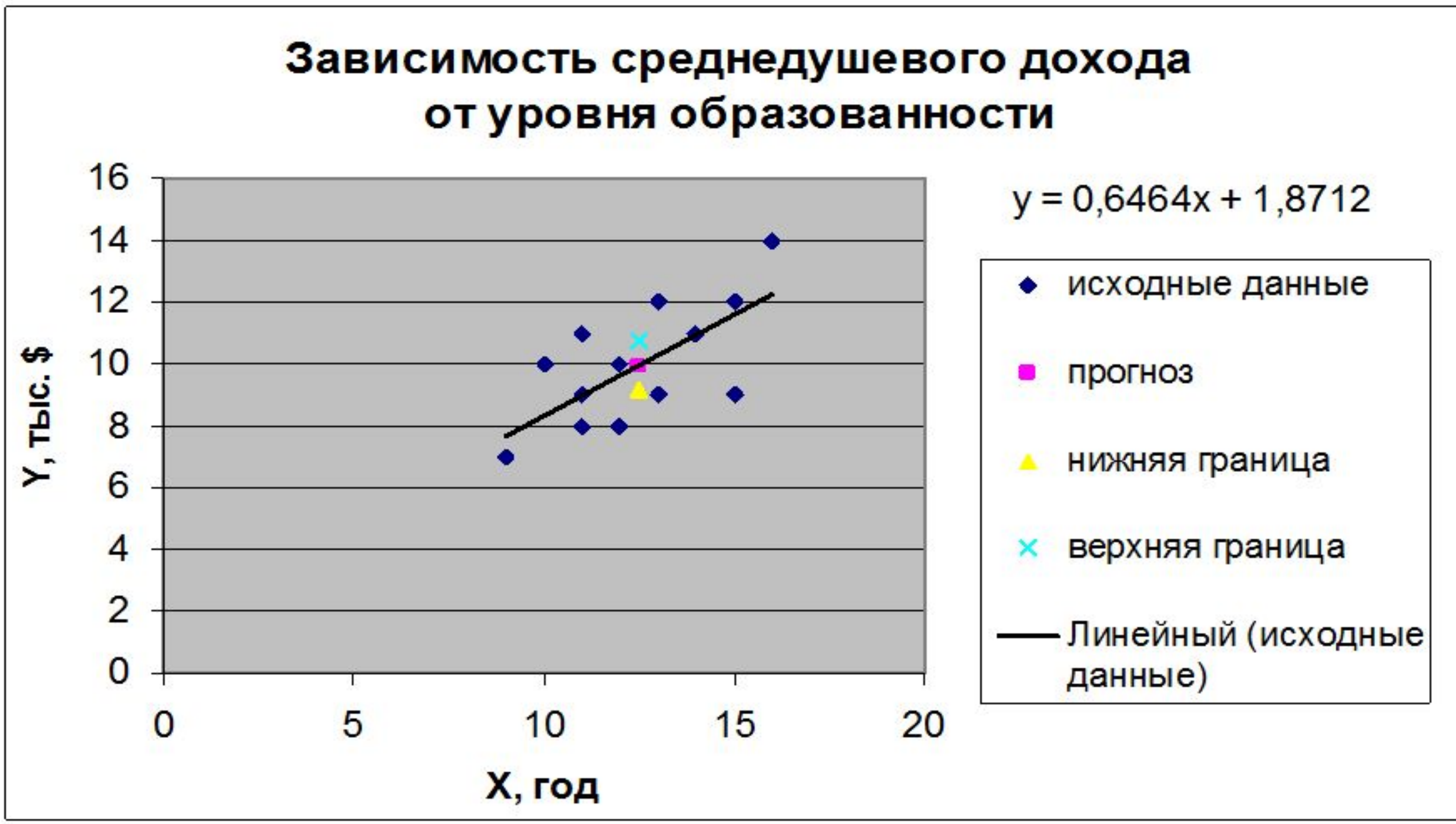
Следовательно, уравнение модели имеет вид $\boldsymbol{Y_T = 1{,}87 + 0{,}65 \cdot X_2}$.

Коэффициент регрессии $\boldsymbol{b = 0{,}65} \Rightarrow$ при увеличении уровня образованности X_2 на 1 год среднедушевой доход Y увеличивается в среднем на 0,65 тыс. $.

Свободный член $\boldsymbol{a = 1{,}87}$ в данном уравнении не имеет реального смысла (нет смысла рассматривать $X_2 = 0$).

Зам. 1 Для построения парных моделей можно использовать ***Мастер диаграмм***.

– На точечной диаграмме показать ***исходные данные*** (поле корреляции).
– ***Добавить линию тренда***: *тип* – линейная (или другой); *параметры* – показывать уравнение на диаграмме.



Линейная регрессия: **Многофакторная модель**

Пример По имеющимся данным о среднем доходе, уровнях аграрности и образованности построить двухфакторную модель. С помощью коэффициентов регрессии, эластичности, бета– и дельта–коэффициентов оценить влияние факторов на результирующий признак.

Решение

Построим двухфакторную модель $y_T = a_0 + a_1 x_1 + a_2 x_2$.

С помощью программы РЕГРЕССИЯ получим

	Коэффициенты
Y-пересечение	5,083965
X1	-0,40902
X2	0,602307

Следовательно, уравнение модели имеет вид $\boldsymbol{y_T = 5{,}08 - 0{,}41x_1 + 0{,}60x_2}$.

Коэффициенты регрессии: $a_1 = -0{,}41$, т.е. при увеличении уровня аграрности X_1 на 1% среднедушевой доход Y уменьшается в среднем на 0,41 тыс.$.

$a_2 = 0{,}60$, т.е. при увеличении уровня образованности X_2 на 1год среднедушевой доход Y увеличивается в среднем на 0,60 тыс.$.

Идентификация модели:

фиктивные переменные

Это специальные (двоичные, бинарные) переменные, которые используются ***для количественного описания качественных признаков***: пол, образование, фактор сезонности, принадлежность к определенному региону, …

$$z = \begin{cases} 1 \text{ - при наличии признака} \\ 0 \text{ - при отсутствии признака} \end{cases}$$

Все процедуры регрессионного анализа выполняются при наличии фиктивных переменных обычным образом.

Пример Исследовать зависимость цены за кв. метр (Y, тыс.$) от расположения квартиры (X, район города).

Решение Для проведения расчетов заменим в таблице исходных данных качественную переменную X фиктивной переменной Z, отражающей расположение квартиры в центре города (Октябрьский и Железнодорожный районы) или на его окраине (Индустриальный, Ленинский и Центральный районы)

$$z = \begin{cases} 1 \text{ - центр города} \\ 0 \text{ - окраина города} \end{cases}$$

С помощью программы РЕГРЕССИЯ построим модель $\boldsymbol{y_T = 1{,}13 + 0{,}52 \cdot z}$.

Коэффициент регрессии $\boldsymbol{b = 0{,}52}$ показывает, что цена за кв. метр для квартир в центре города в среднем на 0,52 тыс.$ выше, чем на окраине.

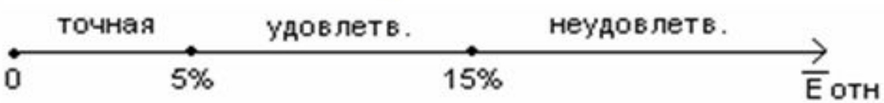
Этап 3. Верификация модели: оценка точности

Это ***оценка точности*** модели и ***дисперсионный анализ***: коэффициент детерминации,
критерий Фишера,
критерий Стьюдента.

Точность модели оценивают с помощью ***средней относительной ошибки*** *аппроксимации* $\overline{E}_{отн}$

Алгоритм расчета:

1. дополнить таблицу «Вывод остатка» итогов Регрессии столбцом «относительные погрешности», вычислив их по формуле $E_{отн\,i} = \left|\frac{E_i}{y_i}\right| \cdot 100$ (%)
2. определить среднее значение относительных погрешностей $\overline{E}_{отн}$ (функция СРЗНАЧ)
3. сопоставить найденную величину $\overline{E}_{отн}$ с уровнями 5% и 15%:



Пример Оценить точность построенной модели зависимости среднедушевого дохода от уровня «образованности» страны.

Решение

1. Используем исходные данные y_i и найденные программой «РЕГРЕССИЯ» остатки E_i.

Вычислим относительные погрешности $E_{отн\,i} = \left|\frac{E_i}{y_i}\right| \cdot 100$ (функция ABS).

2. По столбцу относительных погрешностей найдем среднее значение $\overline{E}_{отн} = 11{,}20\%$.

3. Сравним: $\overline{E}_{отн} = 11{,}20\% \in (5\%;\, 15\%)$ $\Rightarrow$ ***точность модели <u>удовлетворительная</u>***.

Проверка предпосылок регрессионного анализа (МНК)

Условия Гаусса-Маркова:

1. В линейном уравнении $\boldsymbol{Y} = \boldsymbol{Y_T}\left(X\right) + \boldsymbol{\varepsilon}$ слагаемое $\boldsymbol{Y_T}\left(X\right)$ объясняет закономерности изменения результирующего признака *Y*, слагаемое **ε** – ***случайная*** величина, выражает случайный характер результата *Y*.
2. Математическое ожидание случайной составляющей в любом наблюдении равно нулю, а ***дисперсия постоянна***: $\mathrm{M}\left(\varepsilon_i\right) = 0$, $D\left(\varepsilon_i\right) = const$.
3. Случайные составляющие для любых разных наблюдений ***независимы*** (некоррелированы).
4. Распределение случайной составляющей является ***нормальным***.

Теорема Гаусса-Маркова доказывает, что при выполнении перечисленных условий оценки коэффициентов регрессии, найденные методом наименьших квадратов, будут несмещенными, состоятельными и эффективными.

Несмещенность – это отсутствие систематических ошибок при оценивании.
Состоятельность – это увеличение точности оценок с увеличением объема выборки.
Эффективность – наименьшая дисперсия найденных оценок.

Таким образом, регрессионный анализ, основанный на методе наименьших квадратов, дает наилучшие из всех возможных результаты, ***МНК – оптимальный метод*** моделирования.

Схема анализа

МНК – оптимальный метод построения модели (оценки коэффициентов – несмещенные, состоятельные, эффективные)			
↑ **выполняются** (теорема Гаусса – Маркова) ↑			
Предпосылки МНК (условия Гаусса – Маркова)			
1. свойство случайности остатков	**2.** $M(\varepsilon)=0$; $D(\varepsilon)=const$	**3.** свойство независимости остатков	**4.** нормальное распределение остатков
Критерий **поворотных точек**	Критерий **стандартного отклонения;** критерий **Голдфельда - Квандта**	Критерий **Дарбина – Уотсона;** критерий **автокорреляции**	**R/S - критерий**
↓ **не выполняется** ↓	↓ **не выполняется** ↓	↓ **не выполняется** ↓	↓ **не выполняется** ↓
Рекомендуются нелинейные модели $y_p=\alpha\cdot x^{\beta}$ и др.	Рекомендуется обобщенный МНК	Рекомендуются модели авторегрессии $y_p(t)=a+b\cdot y(t-1)$	Невозможно оценить точность прогнозирования

Замечания

Проверка предпосылок МНК – всестороннее исследование остатков модели.

Зам. 1 Перед проверкой предпосылок регрессионного анализа, а именно, свойств случайности, гомоскедастичности и независимости, необходимо убедиться в том, что исходные данные для построенной модели ***упорядочены*** *по возрастанию факторной переменной X*.

!!! *Сортировка* исходных данных должна быть выполнена ***перед построением регрессии*** (сортировка и фильтр / настраиваемая сортировка).

Зам. 2 Для проверки предпосылок МНК программой РЕГРЕССИЯ найдены:

– остатки модели $E_i = y_i - y_{Ti}$ (столбец «*остатки*» итогов РЕГРЕССИИ);
– остаточная сумма квадратов $\mathbf{SS}_{\text{ост}}$; дисперсия остатков $\mathbf{MS}_{\text{ост}}$ (таблица «*Дисперсионный анализ*»);
– стандартная ошибка модели $S(E)$ (таблица «*Регрессионная статистика*»);
– график остатков (для удобства дальнейшей работы рекомендуется изменить тип диаграммы – соединить точки).

Коэффициент детерминации

Опр. *Коэффициент детерминации* определяется соотношением $R^2 = 1 - \dfrac{\sum E_i^2}{\sum (y_i - \bar{y})^2}$ и показывает, какая доля вариации (изменения) результирующего признака Y учтена в модели и объясняется изменением факторных переменных X.

Величина коэффициента детерминации определяется программой РЕГРЕССИЯ (строка «R-квадрат» вывода итогов), ее можно получить также с помощью *Мастера диаграмм*.

Пример Для задачи о зависимости среднедушевого дохода от уровня «образованности» найти коэффициент детерминации и дать интерпретацию полученному значению.

Решение

Выпишем из итогов РЕГРЕССИИ: $R^2 = 0{,}47 = 47\%$.

Таким образом, вариация среднедушевого дохода Y на 47% объясняется по полученному уравнению вариацией уровня «образованности» X_2.

Зам. Коэффициент детерминации R^2 является ***универсальной мерой качества*** эконометрических моделей:

- В случае ***парной линейной*** регрессии коэффициент парной корреляции r связан с коэффициентом детерминации R^2 соотношением $r^2 = R^2$.
- Для линейной модели ***множественной*** регрессии величина $R = \sqrt{R^2}$ называется *коэффициентом* ***множественной корреляции*** и характеризует тесноту зависимости между результирующей переменной и совокупностью учтенных в модели факторов.
- Для ***нелинейных*** регрессионных моделей величина $R = \sqrt{R^2}$ называется ***индексом корреляции*** и характеризует тесноту зависимости между переменными в соответствующей нелинейной форме.

Верификация модели:
оценка значимости уравнения регрессии

- хорошо ли модель соответствует исходным данным?
- достаточно ли включенных в нее факторных переменных?
- целесообразно ли использование этого уравнения в дальнейших расчетах?

Для проверки значимости уравнения регрессии используется ***F- тест Фишера***.

Алгоритм расчета:

1. рассмотреть статистику $F = \dfrac{R^2}{1-R^2} \cdot \dfrac{n-p-1}{p}$ (столбец «F» итогов программы РЕГРЕССИЯ);
2. определить критическое значение $F_{kp}(\alpha; p; n-p-1)$ – таблица критических точек распределения Фишера или функция FРАСПОБР.
3. Сравнить фактическое F с критическим F_{kp} :

Пример Оценить значимость полученного уравнения.

Решение

1. Выпишем из итогов РЕГРЕССИИ $F = 11{,}62$.

2. Критическое значение $F_{кр}= 4{,}67$ найдем с помощью функции «FРАСПОБР» при α=5%, k_1=1, k_2=13.

3. Сравним: $\boldsymbol{F = 11{,}62 > F_{кр}= 4{,}67} \Rightarrow$ ***уравнение модели <u>является значимым</u>***, его использование целесообразно, зависимая переменная Y достаточно хорошо описывается включенной в модель факторной переменной X_2.

Верификация модели:
оценка значимости коэффициентов модели

- закономерно или случайно эти коэффициенты отличаются от нуля?
- целесообразно ли сохранение в модели проверяемых коэффициентов и соответствующих им факторов?

Для проверки значимости коэффициентов модели используется ***критерий Стьюдента***:

1. для коэффициентов b_j модели рассмотреть t- статистики $t(b_j)=\dfrac{b_j}{S(b_j)}$ (столбец «*t-статистики*» итогов программы РЕГРЕССИЯ);
2. определить критическое значение $t_{kp}(\alpha;\ n-p-1)$ – таблица критических точек распределения Стьюдента или функция СТЬЮДРАСПОБР;
3. сравнить модуль $|t(b_j)|$ с критическим значением t_{kp}:

Пример

1. Выпишем из итогов РЕГРЕССИИ:

для свободного коэффициента $a=1{,}87$ статистика $t(a)=0{,}79$;

для коэффициента регрессии $b=0{,}65$ статистика $t(b)=3{,}41$.

2. Критическое значение $t_{kp}=2{,}16$ найдено для уровня значимости α=5% и числа степеней свободы $k=15-1-1=13$.

3. Сравним:

$|t(a)|=0{,}79 < t_{kp}=2{,}16 \Rightarrow$ ***свободный коэффициент а <u>не является значимым</u>***, его можно исключить из модели.

$|t(b)|=3{,}41 > t_{kp}=2{,}16 \Rightarrow$ ***коэффициент регрессии b <u>является значимым</u>***, его и фактор образованности нужно сохранить в модели.

Верификация модели: **замечания**

Зам.1 ***Для сравнения*** различных моделей по точности используют ***стандартную ошибку*** модели $\boldsymbol{S_E}$ (строка «стандартная ошибка» итогов программы РЕГРЕССИЯ).

Чем меньше величина S_E, тем лучше модель.

Зам.2 ***Для сравнения*** моделей с различным количеством учтенных в них факторов используют ***нормированный*** (исправленный, адаптированный) ***коэффициент детерминации*** $\boldsymbol{\hat{R}^2}$ (строка «нормированный R–квадрат» итогов программы РЕГРЕССИЯ).

Чем больше величина $\hat{R}^2$, тем лучше модель.

Зам.3 Для критерия Фишера программа РЕГРЕССИЯ определяет также величину «***Значимость F***» – предельный уровень значимости (вероятность ошибки вывода о целесообразности модели).

Зам.4 В третьей таблице итогов РЕГРЕССИИ представлена полная ***информация о коэффициентах модели***: их средние значения, стандартные ошибки, t–статистики, P–значения (предельные уровни их значимости), нижняя и верхняя границы доверительных интервалов для этих коэффициентов.

Для рассматриваемого примера:

	Коэфф-ты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	1,87	2,36	0,79	0,44	-3,22	6,96
X2	0,65	0,19	3,41	0,0047	0,24	1,06

Верификация модели:
итоги дисперсионного анализа

№	***F***- критерий Фишера	***t***- критерий Стьюдента	Вывод
1	Модель значима	***Все*** коэффициенты являются значимыми	***Модель адекватна***. Ее можно использовать для анализа и прогнозирования
2	Модель значима	***Некоторые*** коэффициенты значимыми не являются	Модель можно использовать для анализа. Для прогнозирования ее следует *улучшить*, исключив незначимые факторы.
3	Модель значима	***Все*** или ***большинство*** коэффициентов не являются значимыми	***Модель неадекватна***. Она непригодна ни для анализа, ни для прогнозирования.

Методы построения адекватной модели:

Метод пошагового включения – построение первоначально парной модели с наиболее информативным фактором и улучшение ее путем добавления новых факторов.

Метод пошагового исключения – построение первоначальной модели с полным набором факторов и улучшение ее путем постепенного исключения наименее информативных (корреляционный анализ) и наименее значимых (критерий Стьюдента) факторов.

«Улучшение» модели контролируют при помощи нормированного коэффициента детерминации $\hat{R}^2$ и стандартной ошибки S_E.

Этап 4. Анализ влияния факторов на результат по уравнению множественной модели

Для исследования влияния совокупности факторов X_j на результирующий признак Y используют множественную модель $\boldsymbol{Y_T = a_0 + a_1 X_1 + a_2 \cdot X_2 + \ldots}$.

Коэффициенты регрессии $a_1, a_2, \ldots$ показывают, на сколько единиц изменится результирующий признак Y при увеличении соответствующего фактора на 1 единицу и неизменных значениях других факторов, закрепленных на среднем уровне.

Коэффициенты a_j имеют разные степени колеблемости и единицы измерения, их нельзя использовать при сравнении влияния на Y различных факторов X_j.

Для этого рассчитывают специальные *безразмерные коэффициенты*.

Средние коэффициенты эластичности в случае линейной модели определяются формулами

$\boxed{Э_j = a_j \cdot \frac{\bar{x}_j}{\bar{y}}}$ где $j = 1, 2, \ldots, p$; $\bar{x}_j, \bar{y}$ – выборочные средние признаков X_j и Y.

Коэффициент $Э_j$ показывает, на сколько процентов изменится в среднем переменная Y при увеличении фактора X_j на 1% и неизменных значениях остальных факторов, закрепленных на их средних уровнях.

Средние коэффициенты эластичности в случае линейной модели определяются формулами

$$\boxed{\mathit{Э}_j = a_j \cdot \frac{\bar{x}_j}{\bar{y}}}$$

где $j = 1, 2, \ldots, p$; $\bar{x}_j$, $\bar{\mathrm{y}}$ – выборочные средние признаков X_j и Y.

Коэффициент $Э_j$ показывает, на сколько процентов изменится в среднем переменная Y при увеличении фактора X_j на 1% и неизменных значениях остальных факторов, закрепленных на их средних уровнях.

Бета–коэффициенты определяются формулами

$$\boxed{\beta_j = a_j \cdot \frac{S_{xj}}{S_y}}$$

где $S_{xj}, \mathrm{S}_{\mathrm{y}}$ – выборочные средние квадратичные отклонения признаков X_j и Y.

Коэффициент β_j показывает, на какую часть величины своего стандартного отклонения изменится в среднем результирующая переменная Y при увеличении только фактора X_j на одно его стандартное отклонение.

Дельта–коэффициенты определяются формулами

$$\boxed{\Delta_j = \beta_j \cdot \frac{r(Y, X_j)}{R^2}}$$

где $\mathrm{r}(\mathrm{Y}, \mathrm{X}_{\mathrm{j}})$ – соответствующие выборочные коэффициенты парной корреляции.

С помощью дельта–коэффициентов оценивается доля влияния каждого фактора в суммарном воздействии на результат всех факторов, учтенных в модели.

Анализ влияния факторов на результат

Пример По имеющимся данным о среднем доходе, уровнях аграрности и образованности построить двухфакторную модель. С помощью коэффициентов регрессии, эластичности, бета– и дельта–коэффициентов оценить влияние факторов на результирующий признак.

Решение

Построим двухфакторную модель $y_T = a_0 + a_1 x_1 + a_2 x_2$.

С помощью программы РЕГРЕССИЯ получим

	Коэффициенты
Y-пересечение	5,083965
X1	-0,40902
X2	0,602307

Следовательно, уравнение модели имеет вид $\boldsymbol{y_T = 5{,}08 - 0{,}41x_1 + 0{,}60x_2}$.

Коэффициенты регрессии: $a_1 = -0{,}41$, т.е. при увеличении уровня аграрности X_1 на 1% среднедушевой доход Y уменьшается в среднем на 0,41 тыс.$.

$a_2 = 0{,}60$, т.е. при увеличении уровня образованности X_2 на 1год среднедушевой доход Y увеличивается в среднем на 0,60 тыс.$.

Рассчитаем ***коэффициенты эластичности***.

Предварительно с помощью функции СРЗНАЧ найдем средние значения признаков X_1, X_2, Y: $\bar{x}_1 = 6{,}53$; $\bar{x}_2 = 12{,}27$; $\bar{y} = 9{,}80$.

$$\boldsymbol{Э_1 = a_1 \cdot \frac{\bar{x}_1}{\bar{y}} = -0{,}41 \cdot \frac{6{,}53}{9{,}80} = -0{,}27}$$ $\Rightarrow$ с ростом уровня аграрности X_1 на 1% и неизменном уровне образованности среднедушевой доход Y уменьшается в среднем на 0,27%.

$$\boldsymbol{Э_2 = a_2 \cdot \frac{\bar{x}_2}{\bar{y}} = 0{,}60 \cdot \frac{12{,}27}{9{,}80} = 0{,}75}$$ $\Rightarrow$ рост уровня образованности X_2 на 1% приводит к росту среднедушевого дохода Y в среднем на 0,75% (при неизменном уровне аграрности).

Для вычисления ***бета–коэффициентов*** предварительно подготовим стандартные отклонения $S_{X1} = 1{,}81;\quad S_{X2} = 2{,}02;\quad S_Y = 1{,}90$ (функция СТАНДОТКЛОН).

Тогда $\boldsymbol{\beta_1 = a_1 \cdot \frac{S_{X1}}{S_Y} = -0{,}41 \cdot \frac{1{,}81}{1{,}90} = -0{,}39}$; $\boldsymbol{\beta_2 = a_2 \cdot \frac{S_{X2}}{S_Y} = 0{,}60 \cdot \frac{2{,}02}{1{,}90} = 0{,}64}$

Таким образом, при увеличении только фактора X_1 на одно свое стандартное отклонение результат Y уменьшится в среднем на 0,39 своего стандартного отклонения S_Y, а при увеличении только фактора X_2 на одно его стандартное отклонение – увеличится на 0,64 S_Y.

Для вычисления ***дельта–коэффициентов*** используем коэффициенты парной корреляции $r(Y, X_j)$, найденные с помощью программы КОРРЕЛЯЦИЯ и коэффициент детерминации $R^2 = 0{,}62$, определенный для рассматриваемой двухфакторной модели программой РЕГРЕССИЯ.

$$\boldsymbol{\Delta_1 = \beta_1 \cdot \frac{r(Y, X_1)}{R^2} = -0{,}39 \cdot \frac{-0{,}47}{0{,}62} = 0{,}29;\quad \Delta_2 = \beta_2 \cdot \frac{r(Y, X_2)}{R^2} = 0{,}64 \cdot \frac{0{,}69}{0{,}62} = 0{,}71}$$

Значит, по уравнению полученной линейной двухфакторной модели изменение результирующего фактора Y (среднедушевого дохода) на 29% объясняется воздействием фактора X_1 (уровня аграрности) и на 71% влиянием фактора X_2 (уровня образованности).

Прогнозирование с помощью линейной модели

Это оценка значений результирующего показателя $\boldsymbol{Y}$ в некоторой, представляющей практический интерес прогнозной ситуации $\boldsymbol{x}^*$. Например:

– каким будет *объем реализации товара* (y^*), если *затраты на рекламу* составят x^* *ден. ед*.?
– какова ожидаемая *себестоимость продукции*(y^*), если *производительность труда* (X) увеличится на k% от достигнутого уровня?
– какой *курс доллара* (y^*) ожидается при сохранении сложившейся тенденции *в ближайший период*?
– …

Алгоритм прогнозирования:

1. зафиксировать $\boldsymbol{x^*}$ **–** *прогнозное значение факторной переменной* (задается в тексте задания или требует дополнительного расчета);
2. рассчитать по уравнению модели ***точечную прогнозную оценку*** $\boldsymbol{y_T^* = a + b \cdot x^*}$ – ожидаемое значение признака Y;
3. оценить ***точность прогнозирования*** – найти *границы доверительного прогнозного интервала*, который с заданной вероятностью γ устанавливает пределы возможного изменения признака Y (интервальное прогнозирование):

– *стандартная ошибка* прогнозирования $\boldsymbol{S(y_T^*) = S_E \cdot \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}}$,

– *размах* доверительного прогнозного интервала $\boldsymbol{U(y_T^*) = S(y_T^*) \cdot t_{кр}}$,

– *границы* доверительного прогнозного интервала для среднего значения $\boldsymbol{y_T^*}$

$$\boldsymbol{U_{нижн}(y_T^*) = y_T^* - U(y_T^*)}, \qquad \boldsymbol{U_{верх}(y_T^*) = y_T^* + U(y_T^*)}.$$

Пример Спрогнозировать величину среднего дохода на душу населения в стране, уровень образованности в которой составляет 12,5 лет. Результаты прогнозирования показать на диаграмме.

Решение

По условию задано $x^* = 12{,}5$.

По уравнению модели найдем *точечную прогнозную оценку* среднего дохода

$$\boldsymbol{y_T^* = 1{,}87 + 0{,}65 \cdot 12{,}5 = 9{,}951}.$$

Таким образом, в рассматриваемой стране ***ожидаемый среднедушевой доход будет составлять около 9,951 тыс.$.***

Для интервального прогнозирования предварительно подготовим:

– стандартную ошибку модели $\boldsymbol{S_E}$ =**1,43** («стандартная ошибка» итогов РЕГРЕССИИ);
– среднее значение $\boldsymbol{\overline{x_2} = 12{,}27}$ (СРЗНАЧ по столбцу исходных данных X_2);
– $\boldsymbol{\sum\left(x_{2i} - \overline{x_2}\right)^2 = 56{,}93}$ (КВАДРОТКЛ по столбцу исходных данных X_2).

Стандартная ошибка прогнозирования для среднего значения составляет

$$\boldsymbol{S\left(y_T^*\right) = S_E \cdot \sqrt{\frac{1}{n} + \frac{\left(x^* - \overline{x}\right)^2}{\sum\left(x_i - \overline{x}\right)^2}} = 1{,}43 \cdot \sqrt{\frac{1}{15} + \frac{\left(12{,}5 - 12{,}27\right)^2}{56{,}93}} = 0{,}372}.$$

При $\boldsymbol{t_{кр}(5\%,\, 13) = 2{,}16}$ размах доверительного интервала для среднего значения

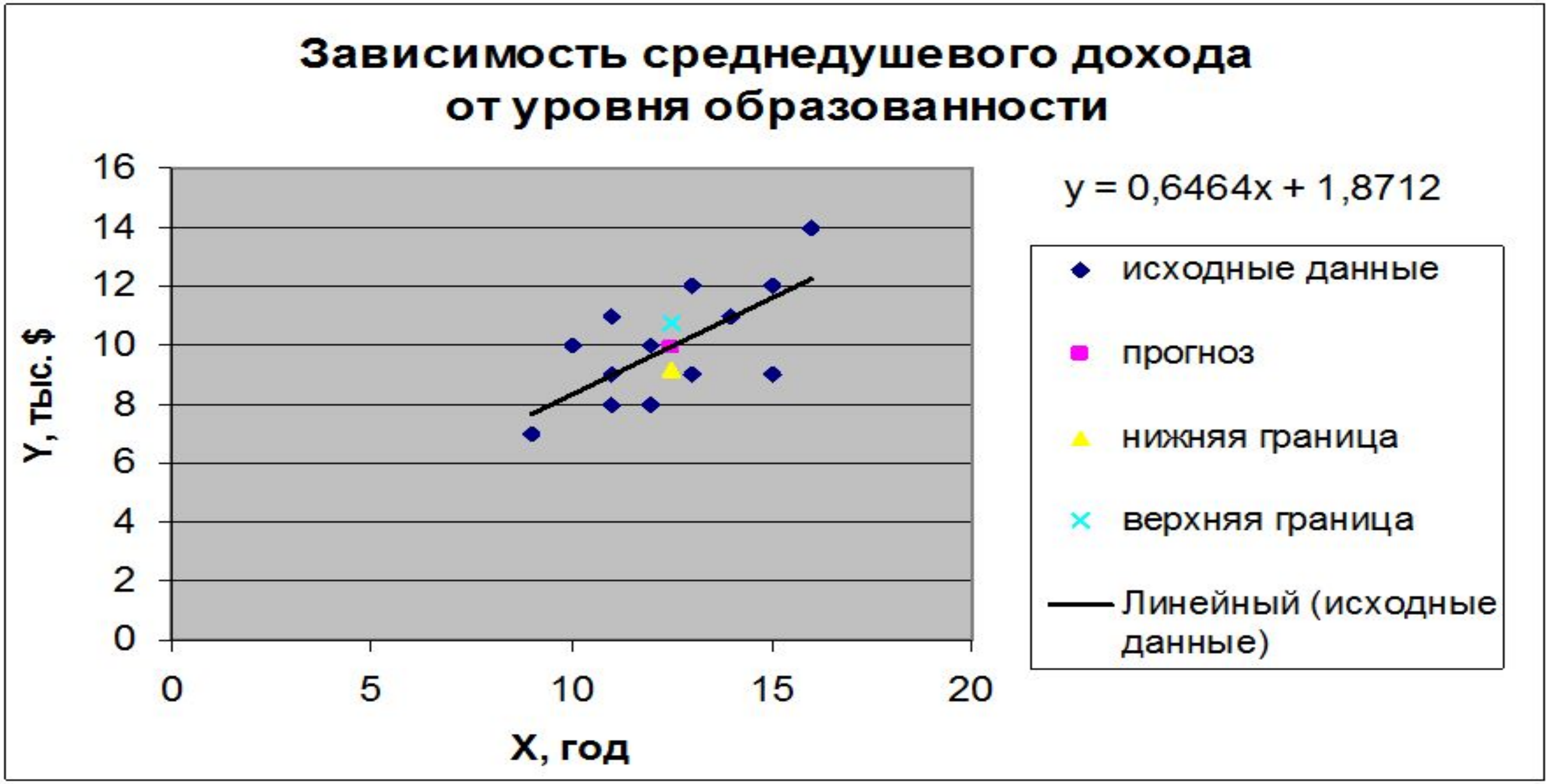
$$\boldsymbol{U\left(y_T^*\right) = S\left(y_T^*\right) \cdot t_{kp} = 0{,}372 \cdot 2{,}16 = 0{,}804}$$

Границами прогнозного интервала будут

$$\boldsymbol{U_{нижн} = y_T^* - U\left(y_T^*\right) = 9{,}951 - 0{,}804 = 9{,}147} \qquad \boldsymbol{U_{верх} = y_T^* + U\left(y_T^*\right) = 9{,}951 + 0{,}804 = 10{,}755}$$

Таким образом, с надежностью 95% можно утверждать, что в рассматриваемой стране ***средний доход на душу населения будет составлять от 9,147 тыс. $ до 10,755 тыс. $***.

Диаграмма результатов моделирования и прогнозирования



Замечания

Различают доверительные интервалы для средних и для индивидуальных значений результирующей переменной Y.

Зам.1 Формула $\boldsymbol{S}(y_T^*) = \boldsymbol{S}_E \cdot \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$ отражает ошибку, связанную с выборочным характером исходных данных, по которым определены коэффициенты модели и проведен расчет точечной прогнозной оценки y_T^*. Это *ошибка среднего прогнозного значения*.

Зам.2 Реальное (индивидуальное) значение результирующего признака может отклоняться от расчетного (теоретического) на случайную величину ε, что увеличивает стандартную *ошибку индивидуального прогнозного значения* $S(y^*)$ по сравнению с $S(y_T^*)$:

$$y^* = y_T^* + \varepsilon \quad \Rightarrow \quad S^2(y^*) = S^2(y_T^*) + S_E^2 \quad \Rightarrow \quad \boldsymbol{S}(y^*) = \boldsymbol{S}_E \cdot \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum (x_i - \bar{x})^2}}\,.$$

Соответственно увеличивается размах доверительного интервала $\boldsymbol{U}(y^*) = \boldsymbol{S}(y^*) \cdot t_{кр}$ и расширяются границы этого интервала $U_{нижн}(y^*) = y_T^* - U(y^*)$, $\quad U_{верх}(y^*) = y_T^* + U(y^*)$.

Зам.3 Формулы стандартных ошибок $S(y^*)$ и $S(y_T^*)$ показывают *пути улучшения качества прогнозирования*:

- уменьшать стандартную ошибку модели S_E;
- выбирать прогнозное значение x^* по возможности ближе к центру выборки $\bar{x}$;
- увеличивать объем исходных данных n.

Прогнозирование с помощью множественной модели

Для множественной линейной модели формулы стандартных ошибок прогнозирования сложнее, расчет проводится средствами ***матричного исчисления***.

Обозначим:

$\boldsymbol{Y}$ и $\boldsymbol{X}$ – матрицы исходных значений результирующей переменной Y и факторов $X_1, X_2, \ldots$;

$a = \left(X^T \cdot X\right)^{-1} \cdot X^T \cdot Y$ – столбец коэффициентов линейной модели $\boldsymbol{y_T = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \ldots}$;

$\boldsymbol{X_{пр}}$ – матрица-строка прогнозных значений факторных переменных $x_1^*, x_2^*, \ldots$

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}; \qquad \boldsymbol{X} = \begin{pmatrix} \boldsymbol{1} & \boldsymbol{x_{11}} & \boldsymbol{x_{21}} & \ldots \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{1} & \boldsymbol{x_{1n}} & \boldsymbol{x_{2n}} & \ldots \end{pmatrix}; \qquad a = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \end{pmatrix}; \qquad \boldsymbol{X_{пр}} = \left(\boldsymbol{1} \;\; \boldsymbol{x_1^*} \;\; \boldsymbol{x_2^*} \;\; \ldots\right).$$

Точечная прогнозная оценка (ожидаемое значение признака Y) определяется по уравнению модели $\boldsymbol{y_T^* = y_T\left(X^*\right) = a_0 + a_1 \cdot x_1^* + a_2 \cdot x_2^* + \ldots}$ (или в матричной форме $y_T^* = X_{пр} \cdot a$).

Для интервального прогнозирования

1. необходимо вычислить величину $\boldsymbol{v = X_{пр} \cdot \left(X^T \cdot X\right)^{-1} \cdot X_{пр}^T}$ и найти с ее помощью ***стандартные ошибки прогнозирования*** для среднего или для индивидуального значений результирующего признака Y: $\boldsymbol{S\left(y_T^*\right) = S_E \cdot \sqrt{v}}$ или $\boldsymbol{S\left(y^*\right) = S_E \cdot \sqrt{1+v}}$;
2. определить ***размах*** прогнозного интервала: $\boldsymbol{U\left(y_T^*\right) = S\left(y_T^*\right) \cdot t_{кр}}$ или $\boldsymbol{U\left(y^*\right) = S\left(y^*\right) \cdot t_{кр}}$;
3. найти ***границы*** прогнозного интервала: $\boldsymbol{U_{в/н}\left(y_T^*\right) = y_T^* \pm U\left(y_T^*\right)}$ или $\boldsymbol{U_{в/н}\left(y^*\right) = y_T^* \pm U\left(y^*\right)}$.

Пример Спрогнозировать величину среднего дохода на душу населения в стране, уровень аграрности в которой составляет 6,5%, а уровень образованности 12,5 лет.

Решение По условию задано $x_1^* = 6{,}5$; $x_2^* = 12{,}5$

По уравнению двухфакторной модели найдем **точечную прогнозную оценку** среднего дохода

$$y_T^* = b_0 + b_1 \cdot x_1^* + b_2 \cdot x_2^* = 5{,}08 - 0{,}41 \cdot 6{,}5 + 0{,}60 \cdot 12{,}5 = 9{,}954$$

⇨ **в рассматриваемой стране ожидаемый среднедушевой доход будет составлять около 9,954 тыс.$.**

Для интервального прогнозирования необходимо определить величину $v = X_{np} \cdot \left(X^T \cdot X\right)^{-1} \cdot X_{np}^T$

– подготовим матрицы $X = \begin{pmatrix} 1 & 8 & 9 \\ 1 & 9 & 13 \\ \dots & \dots & \dots \end{pmatrix}$ и $X_{np} = \begin{pmatrix} 1 & 6{,}5 & 12{,}5 \end{pmatrix}$;

– транспонируем их (Копировать; Специальная вставка… $\boxed{\text{v}}$ транспонировать);

– умножим $X^T \cdot X = \begin{pmatrix} 15 & 98 & 184 \\ 98 & 686 & 1196 \\ 184 & 1196 & 2314 \end{pmatrix}$ (МУМНОЖ); найдем обратную $\left(X^T \cdot X\right)^{-1} = \begin{pmatrix} 4{,}08 & -0{,}17 & -0{,}23 \\ -0{,}17 & 0{,}02 & 0{,}00 \\ -0{,}23 & 0{,}00 & 0{,}02 \end{pmatrix}$ (МОБР);

– умножим $X_{np} \cdot \left(X^T \cdot X\right)^{-1} = \begin{pmatrix} 0{,}018 & -0{,}000 & 0{,}004 \end{pmatrix}$; $v = 0{,}068$ (МУМНОЖ).

Стандартная ошибка прогнозирования для среднего значения составляет $S\left(y_T^*\right) = S_E \cdot \sqrt{v} = 1{,}26 \cdot \sqrt{0{,}068} = 0{,}328$.

При $t_{kp}(5\%, 12) = 2{,}18$ **размах** доверительного интервала для среднего значения $U\left(y_T^*\right) = S\left(y_T^*\right) \cdot t_{kp} = 0{,}328 \cdot 2{,}18 = 0{,}714$.

Границы прогнозного интервала

$$U_{нижн} = y_T^* - U\left(y_T^*\right) = 9{,}954 - 0{,}714 = 9{,}240 \quad \text{и} \quad U_{верх} = y_T^* + U\left(y_T^*\right) = 9{,}954 + 0{,}714 = 10{,}668.$$

Таким образом, с надежностью 95% можно утверждать, что **в рассматриваемой стране средний доход на душу населения будет составлять от 9,240 тыс. $ до 10,668 тыс. $.**

Нелинейная регрессия: Типы зависимости между экономическими показателями

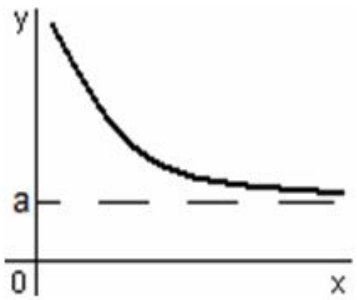
Линейная модель $y_T = a + bx$ отражает ***закон пропорциональности***, т.е. рост или убывание показателя Y с постоянной скоростью.

Например, зависимость затрат на производство (потребления электроэнергии, …) от объема выпускаемой продукции.

Однако многие экономические процессы описываются ***нелинейными функциями***, например:

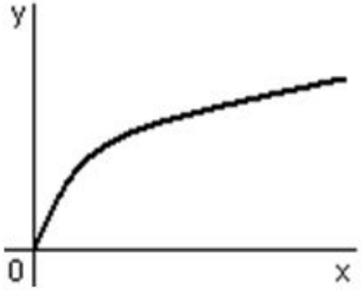
Квадратическая зависимость $y_T = a + bx + cx^2$ (парабола второй степени) целесообразна при наличии для рассматриваемого интервала значений фактора X ***точки максимума или минимума*** показателя Y.

Например, зависимость урожайности от количества внесенных удобрений.



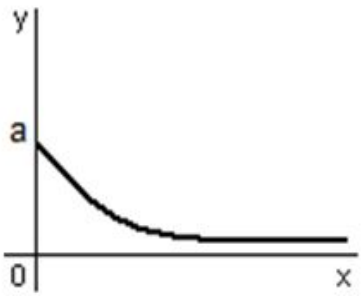
Гиперболическая зависимость $y_T = a + \frac{b}{x}$ используется для описания ***стабилизирующихся*** процессов.

Например, зависимость темпа роста зарплаты от уровня безработицы (кривая Филипса); зависимость себестоимости единицы продукции от объема выпуска.

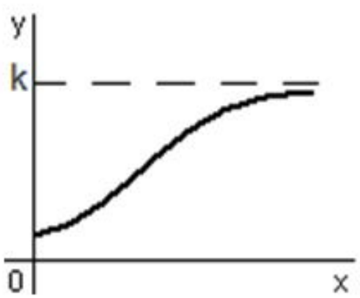


Для ***степенной*** зависимости $\boldsymbol{y_T = a \cdot x^b}$ характерно постепенное ***уменьшение скорости роста*** Y.

Она считается приемлемой для описания кривых Энгеля, характеризующих соотношение между спросом на определенный товар и общей суммой дохода. Степенными являются производственные функции $U_T = a_0 \cdot K^{a_1} \cdot L^{a_2}$, выражающие зависимость величины продукта от производственных факторов (затрат капитала, труда и др.).



Показательная функция $\boldsymbol{y_T = a \cdot b^x}$ применяется при ***стабильных темпах роста***, например, для описания кривых спроса (в зависимости от цены товара).



S – образные кривые $\boldsymbol{y_T = k \cdot a^{b^x}}$ (кривая Гомперца), $\boldsymbol{y_T = \dfrac{k}{1 + a \cdot b^{-x}}}$ (логистическая функция) отражают медленный рост показателя Y на 1-ом этапе развития, ускорение на 2-ом, замедление и стремление к некоторому предельному значению на 3-ем.

Такие зависимости целесообразны для описания производственного цикла, динамики показателей уровня жизни, спроса на новые товары.

Построение нелинейных моделей

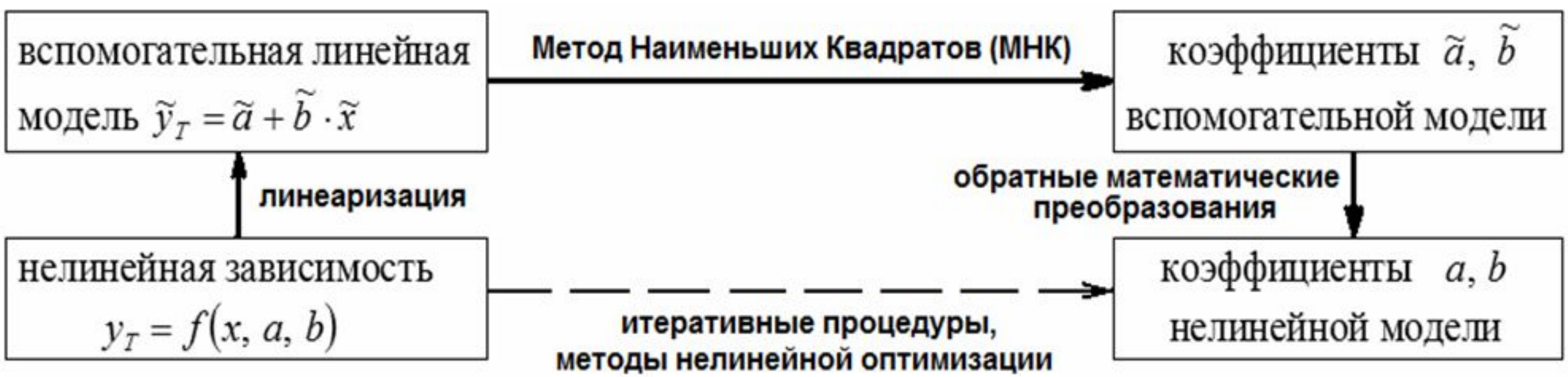
Для построения нелинейных моделей используют прием ***линеаризации*** – *преобразование нелинейной зависимости **к линейному виду***.

1. Модели, нелинейные относительно включенных в них переменных, но линейные по оцениваемым параметрам линеаризуют *с помощью введения* ***новых переменных***.

   Например: $\boldsymbol{y_T = a + \frac{b}{x}}$ (гиперболическая) $\rightarrow$ $\boldsymbol{\tilde{x} = \frac{1}{x}}$ $\rightarrow$ $\boldsymbol{y_T = a + b\tilde{x}}$ (линейная).

2. Модели, нелинейные по оцениваемым параметрам, линеаризуют *с помощью* ***математических преобразований***.

   Например: $\boldsymbol{y_T = a \cdot b^x}$ (показательная) $\rightarrow$ $\ln y_T = \ln a + \ln b \cdot x$ $\rightarrow$ $\boldsymbol{\tilde{y}_T = \tilde{a} + \tilde{b}x}$ (линейная).



Линейную и наиболее распространенные нелинейные зависимости можно построить с помощью *Мастера диаграмм* (опция *Добавить линию тренда…*).

Степенная модель

Пример По данным о рентабельности продукции Y (%) и ее трудоемкости X (чел./ед.), приведенным в таблице, построить нелинейные модели парной регрессии: степенную, показательную и гиперболическую. Исходные данные и результаты моделирования показать графически.

X_i	1,0	1,2	1,5	2,0	2,5	2,7	3,0
Y_i	32	28	22	20	16	15	10

Решение

Степенная модель $\boldsymbol{y_T = a \cdot x^b}$ является стандартной.

Для ее построения используем *Мастер диаграмм*: исходные данные покажем с помощью точечной диаграммы, затем добавим линию степенного тренда и выведем на диаграмму уравнение модели.

Таким образом, уравнение степенной модели $y_T = \mathbf{33{,}094} \cdot x^{-0{,}8924}$.

Показательная модель

Показательная модель $y_T = a \cdot b^x$ тоже стандартная (экспоненциальная). Построим ее с помощью *Мастера диаграмм*.

Таким образом, уравнение показательной модели $y_T = \mathbf{51{,}218} \cdot e^{-0{,}4953x}$

Можно вычислить $b = e^{-0{,}4953} = \mathbf{0{,}6094}$ (функция EXP) и записать уравнение модели в виде $y_T = \mathbf{51{,}218} \cdot (\mathbf{0{,}6094})^x$.

Гиперболическая модель

Гиперболическая модель $y_T = a + \frac{b}{x}$ не является стандартной.

Для ее построения выполним линеаризацию: $\tilde{x} = \frac{1}{x} \quad \rightarrow \quad y_T = a + b\tilde{x}$.

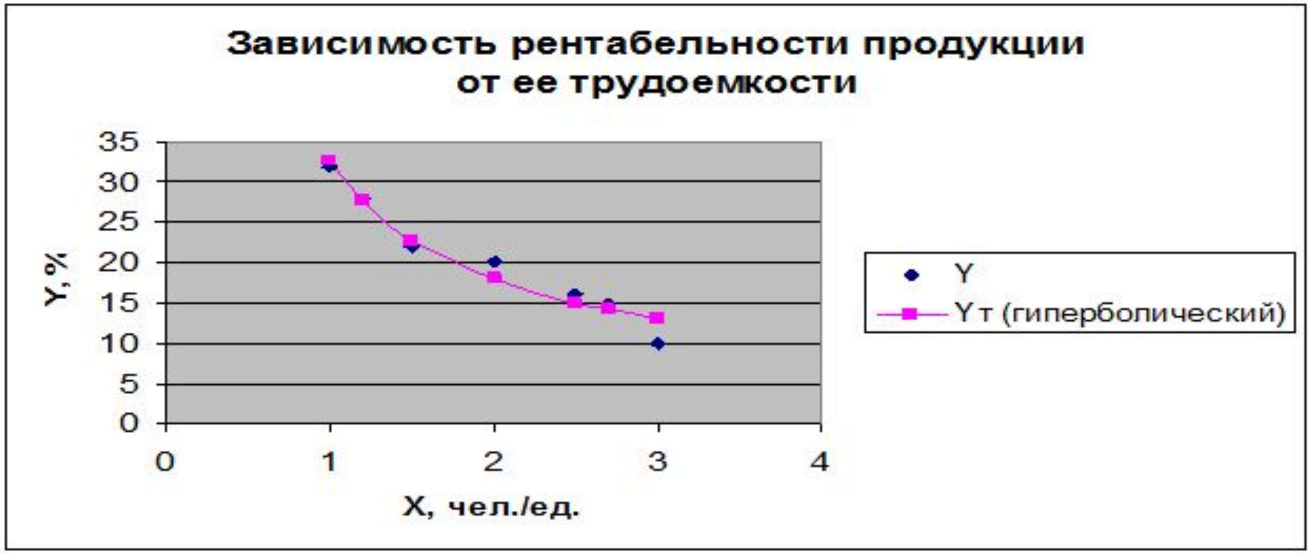
Коэффициенты линейной модели определим с помощью МНК, в качестве исходных данных используем столбец значений y_i (остается без изменений) и столбец преобразованных значений $\tilde{x}_i = \frac{1}{x_i}$.

С помощью программы РЕГРЕССИЯ получим

	Коэффициенты
Y-пересечение	3,269739
1/X	29,26913

Таким образом, найдены коэффициенты $a = 3{,}27$; $b = 29{,}27$ и получено уравнение гиперболической модели: $y_T = 3{,}27 + \frac{29{,}27}{x}$.

Для того, чтобы показать линию гиперболической модели на диаграмме, добавим к ряду исходных данных (x_i, y_i) ряд теоретических значений (x_i, y_{Ti}).



Временные ряды: основные понятия

Временным (динамическим) рядом называется последовательность наблюдений некоторого показателя (признака) Y во времени.

Отдельные наблюдения – ***уровни ряда*** – y_t или $y(t)$.

<u>Отличия временного ряда</u> от обычной статистической выборки:

– последовательные по времени наблюдения являются ***взаимозависимыми***;
– уровни временного ряда обладают ***разной информативностью***;
– ***точность статистических характеристик*** временного ряда может не повышаться, а даже понижаться с увеличением количества наблюдений.

<u>Свойства</u>, необходимые для математического моделирования временного ряда:

– ***Сопоставимость*** – формирование всех уровней по одной методике, использование одинаковых единиц измерения и шага наблюдений.

– ***Однородность*** данных – отсутствие нетипичных (аномальных) наблюдений.

– ***Устойчивость*** – преобладанием закономерности над случайностью в изменении уровней ряда, проявление тенденции.

– ***Полнота данных*** – для моделирования временного ряда число уровней должно быть не менее 7 – 10 (закономерность может обнаружиться лишь при достаточном объеме наблюдений).

Временные ряды: основные понятия

<u>Структурная модель</u> временного ряда:

$$Y(t) = f(t) + S(t) + C(t) + \varepsilon(t)$$

$f(t)$ – ***тренд*** – плавно меняющаяся компонента, длительная тенденцию изменения признака;

$S(t)$ – ***сезонная*** компонента – регулярные колебания, которые носят периодический или близкий к нему характер и заканчиваются в течение года;

$C(t)$ – ***циклическая*** компонента – длительные периоды относительного подъема и спада, циклы переменной длительности и амплитуды;

$f(t)$, $S(t)$, $C(t)$ – ***закономерные*** составляющие временного ряда;

$\varepsilon(t)$ – ***случайная*** компонента – влияние не поддающихся учету и регистрации случайных факторов.

Важнейшая **<u>задача</u>** исследования – ***выявление основной тенденции*** изучаемого процесса.

<u>Этапы</u> моделирования временного ряда:

Предварительный анализ → Построение кривой роста → Оценка качества модели → Прогнозирование

Предварительный анализ:
выявление аномальных наблюдений

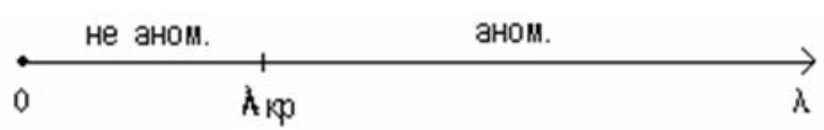
– выявление аномальных наблюдений;
– проверка наличия тренда;
– выявление формы тренда, наличия сезонной и циклической компонент.

Для выявления ***аномальных наблюдений*** используют **метод Ирвина**.

Алгоритм метода:

1. для $t \geq 2$ определить λ – статистики $\boxed{\lambda_t = \dfrac{|(y_t - y_{t-1})|}{S_y}}$

   S_y - выборочное среднее квадратичное (стандартное) отклонение признака Y;
2. по таблице критических значений найти $\lambda_{кр}$ (зависит от количества наблюдений n);
3. сравнить фактические значения λ_t с $\lambda_{кр}$:



Аномальные наблюдения, причинами возникновения которых являются ошибки технического характера, необходимо исключить из временного ряда и заменить их расчетными значениями (например, используя среднее по двум соседним значениям).

Если причины возникновения аномального наблюдения имеют объективный характер, то замену не производят.

Пример

На уровне значимости $\alpha = 1\%$ проверить наличие аномальных наблюдений в ряде данных по урожайности зерновых, собранных за период 10 лет.

год	1	2	3	4	5	6	7	8	9	10
урожайность	10,3	14,3	7,7	15,8	14,4	16,7	15,3	20,2	17,1	15,2

Решение

Используем метод Ирвина

1. Подготовим $\boldsymbol{S_y = 3{,}50}$ (СТАНДОТКЛОН) и рассчитаем статистики

$$\lambda_t = \left| \frac{y_t - y_{t-1}}{S_y} \right| \text{ (ABS).}$$

Результаты расчетов приведем в таблице:

t	1	2	3	4	5	6	7	8	9	10
Y	10,3	14,3	7,7	15,8	14,4	16,7	15,3	20,2	17,1	15,2
λt		1,14	1,88	2,31	0,40	0,66	0,40	1,40	0,88	0,54

2. Критическое значение при $\boldsymbol{n = 10}$ и уровне значимости $\boldsymbol{\alpha = 1\%}$ составляет $\boldsymbol{\lambda_{кр} = 2{,}0}$.
3. Сравним $\boldsymbol{\lambda_4 = 2{,}31 > \lambda_{кр} = 2{,}0} \Rightarrow$ соответствующее наблюдение $\boldsymbol{y_4 = 15{,}8}$ признается аномальным и требует выяснения причин его появления. Будем считать, что технические ошибки отсутствуют и сохраним значение $\boldsymbol{y_4}$ без изменений.

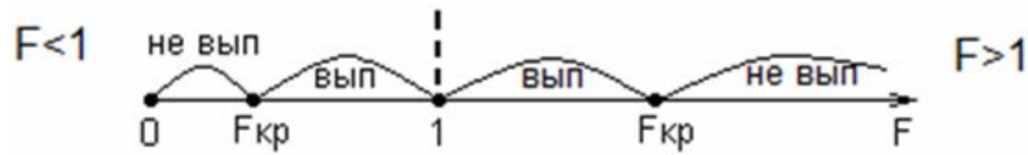
Предварительный анализ:
проверка наличия тренда

Метод сравнения средних уровней:

1. Исходный временной ряд разбить на две примерно равные по числу уровней части, каждую из которых рассмотреть как самостоятельную выборку со своими статистическими характеристиками (средними значениями $\bar{y}_1$ и $\bar{y}_2$, дисперсиями S_1^2 и S_2^2).

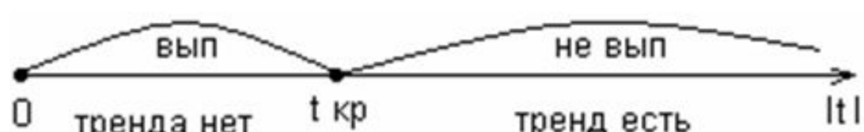
2. С помощью F-критерия Фишера проверить гипотезу о равенстве дисперсий обеих частей ряда. В Excel: сервис / анализ данных / ***двухвыборочный F-тест для дисперсии***.

   Результатом работы программы является статистика $F = \dfrac{S_1^2}{S_2^2}$ и критическое значение $F_{кр}$:



3. С помощью t-критерия Стьюдента проверить основную гипотезу о равенстве средних значений обеих частей ряда. В зависимости от результатов F-теста выбрать: сервис / анализ данных / ***двухвыборочный t-тест с одинаковыми дисперсиями*** или сервис / анализ данных / ***двухвыборочный t-тест с различными дисперсиями***.

   Программа определяет значение t-статистики $t = \dfrac{\bar{y}_1 - \bar{y}_2}{S_{общ}}$ и критический уровень $t_{кр}$:



Пример. Выполнить проверку наличия тренда урожайности.

Решение

1. Исходный временной ряд разделим **на две группы** по 5 наблюдений в каждой.

2. Для проверки гипотезы о равенстве дисперсий этих частей выполним **двухвыборочный F-тест для дисперсии**.

   Найдем $\boldsymbol{F} = \mathbf{2{,}778}$; $\boldsymbol{F_{kp}} = \mathbf{6{,}388}$.

   Сравнение $\boldsymbol{F} = \mathbf{2{,}778} < \boldsymbol{F_{kp}} = \mathbf{6{,}388}$ показывает, что ***различие выборочных дисперсий*** ***<u>незначительно</u>***, проверяемая гипотеза о равенстве дисперсий подтверждается.

3. Проверим гипотезу о равенстве средних с помощью **двухвыборочного t-теста с одинаковыми дисперсиями**.

   Найдем $\boldsymbol{t} = \mathbf{-2{,}498}$; $\boldsymbol{t_{kp}} = \mathbf{2{,}306}$.

   Анализ результатов выполнения *t*-теста показывает, что $|\boldsymbol{t}| = \mathbf{2{,}498} > \boldsymbol{t_{kp}} = \mathbf{2{,}306}$. Таким образом, ***различие средних значений следует считать <u>существенным</u>***, **тренд есть**.

Предварительный анализ:
сглаживание временного ряда

Метод простой скользящей средней:

1. Выберем интервал сглаживания $m\ (m < n)$, *например*: $m = 3$
2. Для m первых уровней ряда вычислим среднее арифметическое значение – сглаженное значение для уровня ряда, находящегося в середине интервала сглаживания:
$$\bar{y}_2 = \frac{1}{3} \cdot (y_1 + y_2 + y_3)$$
3. Сдвинем интервал сглаживания на один уровень и повторим вычисление среднего значения
$$\bar{y}_3 = \frac{1}{3} \cdot (y_2 + y_3 + y_4)$$
и т.д.

Процедура сглаживания устраняет случайные колебания результатов наблюдений и облегчает выявление закономерных компонент временного ряда.

Недостаток метода – сглаженный ряд укорачивается по сравнению с фактическим: теряет по $\frac{m-1}{2}$ крайних уровней.

Пример. Сгладить временной ряд урожайности с помощью простой скользящей средней (принять $m = 3$). Исходные данные и сглаженный ряд показать на диаграмме.

Построение кривой роста: виды тренда

Кривая роста – плавная кривая, аппроксимирующая временной ряд.

Аналитическое выравнивание временного ряда – построение такой кривой с целью установления основной тенденции развития и прогнозирования.

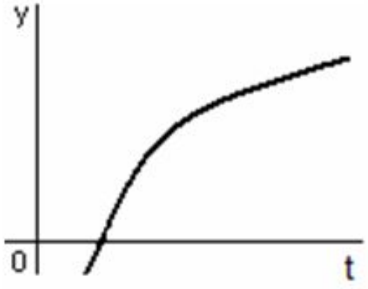
Эталонные типы развития социально-экономических явлений во времени.

$\boxed{\hat{y}_t = a + b \cdot t}$ – равномерное развитие;

отражает рост или убывание показателя Y с постоянной скоростью ($b > 0$ – рост, $b < 0$ – снижение).

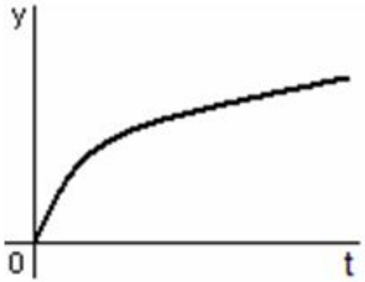
$\boxed{\hat{y}_t = a + b \cdot t + c \cdot t^2}$ – равноускоренное (равнозамедленное) развитие;

c – темп прироста ($c > 0$ – ускорение развития, $c < 0$ – замедление).



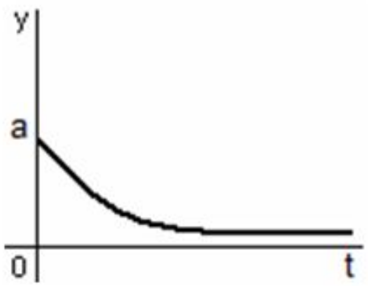
$\boxed{\hat{y}_t = a + b \cdot \ln t}$ – логарифмическая функция;

развитие с замедлением роста в конце периода.



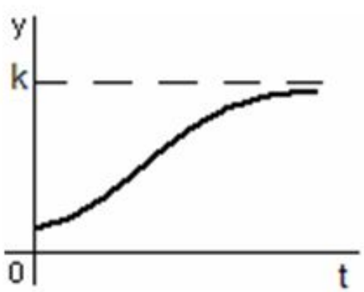
$\boxed{\hat{y}_t = a \cdot t^b}$ – степенная функция;

спрос на непродовольственные товары.



$\boxed{\hat{y}_t = a \cdot b^t}$ – показательная функция;

используют при стабильных темпах роста,
(b – темп роста или интенсивность развития).



$\boxed{y_T = k \cdot a^{b^t}}$ (кривая Гомперца), $\boxed{y_T = \dfrac{k}{1 + a \cdot b^{-t}}}$ (логистическая функция)

S – образные кривые – отражают медленный рост показателя Y на 1-ом этапе развития, ускорение на 2-ом, замедление и стремление к некоторому предельному значению на 3-ем.

Такие зависимости целесообразны для описания производственного цикла, динамики показателей уровня жизни, спроса на новые товары.

Для выбора типа модели используют интуитивные приёмы (по виду графика исходных данных), исследование статистических показателей (абсолютных приростов, темпов роста, …), а также содержательный анализ процесса. Предпочтение, как правило, отдаётся простым моделям.

Построение кривой роста:
линейная модель

Наиболее простой является **линейная модель** $\hat{y}_t = a + b \cdot t$.

Коэффициенты a и b линейной модели и дополнительную статистическую информацию о качестве этой модели можно получить с помощью программы РЕГРЕССИЯ

Коэффициент регрессии b показывает среднее изменение результата Y при увеличении фактора времени t на 1 единицу (с каждым годом, с каждым днем, …).

Смысл коэффициента a – значение Y в начальный момент времени – не представляет практического интереса.

Пример. Для временного ряда урожайности построить линейную регрессионную модель; объяснить смысл коэффициента регрессии; исходные данные и результаты моделирования показать на чертеже.

Решение С помощью программы «РЕГРЕССИЯ» получим

	Коэффициенты
Y-пересечение	10,467
t	0,770

Таким образом, $a = 10{,}467$; $b = 0{,}770$.

Уравнение модели имеет вид $\hat{y}_t = 10{,}467 + 0{,}770 \cdot t$.

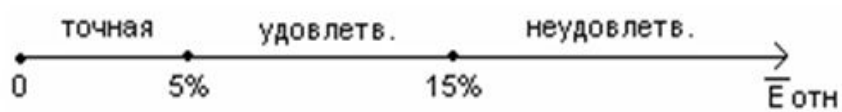
$b = 0{,}770 \Rightarrow$ с каждым годом урожайность Y увеличивается в среднем на 0,770 ц/га.

Качество модели временного ряда:
достоверность аппроксимации и точность модели

Зам.1 Для выбора типа модели временного ряда можно использовать «***Мастер диаграмм***»:

- построить точечную диаграмму исходных данных (t, y_t);
- добавить линию тренда, показать ее уравнение и величину достоверности аппроксимации R^2;
- сопоставить коэффициенты R^2, выбрать наибольший, соответствующий наиболее подходящему типу тренда.

Зам.2 ***Точность модели*** оценивают с помощью ***средней относительной ошибки*** $\bar{e}_{отн}$:



Пример Оценить точность линейной модели временного ряда урожайности.

Решение

1. Дополним таблицу «ВЫВОД ОСТАТКА» столбцом относительных погрешностей
$$e_{отн\,t} = \left|\frac{e_t}{y_t}\right| \cdot 100 \text{ (функция ABS).}$$
2. По столбцу относительных погрешностей найдем среднее значение $\bar{e}_{отн} = 15{,}81\%$.
3. Сравним: $\bar{e}_{отн} = 15{,}81\% > 15\%$. $\Rightarrow$ ***точность модели <u>неудовлетворительная</u>***.

Адекватность модели временного ряда: **теорема Гаусса-Маркова**

В структурной формуле временного ряда

$$\boxed{y_t = f_t + S_t + C_t + \varepsilon_t}$$

тренд, сезонная и циклическая компоненты являются ***регулярными*** или систематическими компонентами временного ряда.

Если при моделировании регулярные компоненты выделены правильно, то для остатков ε_t должны выполняться свойства:

1. случайности;
2. равенства нулю математического ожидания и постоянства дисперсии;
3. независимости;
4. нормального распределения.

Если все перечисленные свойства для остаточной компоненты выполняются, то модель правильно отражает систематические компоненты временного ряда. Такая модель называется ***адекватной***.

Теорема Гаусса-Маркова доказывает, что при выполнении перечисленных условий регрессионный анализ, основанный на методе наименьших квадратов, дает наилучшие из всех возможных результаты $\Rightarrow$ ***МНК – оптимальный метод*** моделирования.

Предпосылки МНК: **схема анализа**

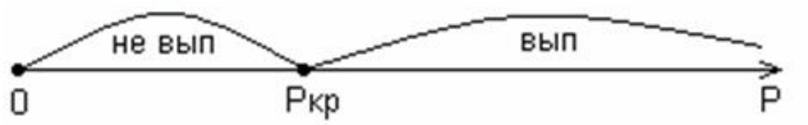
МНК – оптимальный метод построения модели (оценки коэффициентов – несмещенные, состоятельные, эффективные)			
↑ **выполняются** (теорема Гаусса – Маркова) ↑			
Предпосылки МНК (условия Гаусса – Маркова)			
1. свойство случайности остатков	**2.** $M(\varepsilon)=0$; $D(\varepsilon)=const$	**3.** свойство независимости остатков	**4.** нормальное распределение остатков
Критерий **поворотных точек**	Критерий **стандартного отклонения;** критерий **Голдфельда - Квандта**	Критерий **Дарбина – Уотсона;** критерий **автокорреляции**	**R/S - критерий**
↓ **не выполняется** ↓	↓ **не выполняется** ↓	↓ **не выполняется** ↓	↓ **не выполняется** ↓
Рекомендуются нелинейные модели $y_p=\alpha\cdot x^{\beta}$ и др.	Рекомендуется обобщенный МНК	Рекомендуются модели авторегрессии $y_p(t)=a+b\cdot y(t-1)$	Невозможно оценить точность прогнозирования

Проверка свойства случайности остатков:

критерий поворотных точек (пиков)

Алгоритм критерия:

1. определить количество поворотных точек (пиков) p;
2. вычислить критическую величину $p_{kp} = \left[\frac{2(n-2)}{3} - 1{,}96\sqrt{\frac{16n-29}{90}}\right]$;
3. сравнить фактическое p и p_{kp} :



Пример Проанализировать выполнение свойства случайности остатков для модели временного ряда урожайности.

Решение

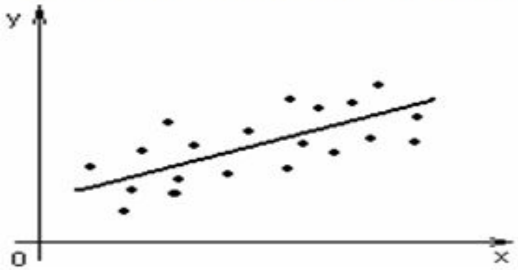
Используем критерий поворотных точек:

1. количество поворотных точек определим по графику остатков: $p = 7$;
2. при $n = 10$ критическое значение $p_{kp} = \left[\frac{2 \cdot 8}{3} - 1{,}96 \cdot \sqrt{\frac{16 \cdot 10 - 29}{90}}\right] = [2{,}97] = 2$;
3. сравним $p = 7 > p_{kp} = 2$, следовательно, ***свойство случайности для ряда остатков выполняется***.

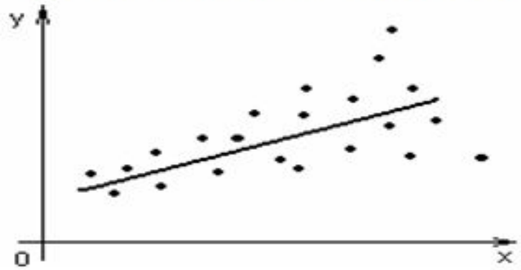
Свойства нулевого математического ожидания и постоянной дисперсии остатков

Свойство $M(\varepsilon_i) = 0$ для линейной модели, коэффициенты которой определены по методу наименьших квадратов, ***выполняется автоматически***.

Свойство $D(\varepsilon) = const$ характеризует степень однородности исходных данных:



дисперсия ***постоянная***
исходные данные ***однородные***
модель ***гомоскедастичная***



дисперсия ***непостоянная***
исходные данные ***неоднородные***
модель ***гетероскедастичная***

Для проверки свойства используют **критерий Голдфельда-Квандта**.

Алгоритм критерия:

1. Выбрать **m** первых и **m** последних наблюдений, по которым построить отдельные регрессии; средние $(n-2m)$ наблюдений не рассматривать; рассчитать статистику $F = \dfrac{SS_{max}}{SS_{min}}$.
2. Определить критическое значение $F_{kp}(\alpha;\, m-p-1;\, m-p-1)$ – таблица критических точек распределения Фишера или функция FРАСПОБР.
3. Сравнить фактическое значение F с критическим F_{kp}:

Пример Проанализировать выполнение свойства нулевого математического ожидания и постоянной дисперсии остатков модели временного ряда урожайности.

Решение

С помощью функции СРЗНАЧ для ряда остатков *убедимся*: $\bar{e} = 1{,}69E - 15 = 0$.

Для проверки *свойства гомоскедастичности* (постоянной дисперсии остаточной компоненты) используем ***критерий Голдфельда-Квандта***.

1. В исходных данных выделим ***первые*** **4** и ***последние*** **4** уровней, ***средние*** **2** уровня не рассматриваем.

С помощью программы РЕГРЕССИЯ построим модель по первым четырем наблюдениям (регрессия1). Для этой модели остаточная сумма квадратов $SS_1 = 36{,}21$.

С помощью программы РЕГРЕССИЯ построим модель по последним четырем наблюдениям (регрессия2). Для этой модели остаточная сумма квадратов $SS_2 = 15{,}79$.

Рассчитаем статистику критерия: $F = \frac{SS_{\max}}{SS_{\min}} = \frac{36{,}21}{15{,}79} = 2{,}29$.

2. Критическое значение при уровне значимости $\alpha = 5\%$ и числах степеней свободы $k_1 = k_2 = 4 - 1 - 1 = 2$ составляет $F_{kp} = 19$.

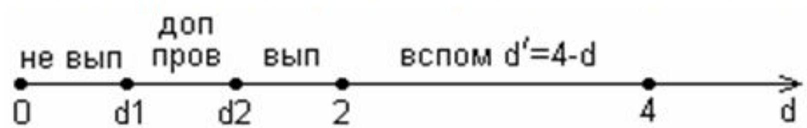
3. Сравним: $F = 2{,}29 < F_{kp} = 19 \Rightarrow$ ***свойство постоянной дисперсии остаточной компоненты выполняется; модель гомоскедастичная***.

Проверка свойства независимости остатков:

критерий Дарбина-Уотсона

Алгоритм критерия:

1. вычислить статистику $d = \dfrac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$;
2. определить критические уровни d_1 и d_2 (таблица d-статистик Дарбина-Уотсона);
3. сопоставить фактическое d с критическими в соответствии со схемой:



Пример Проанализировать выполнение свойства независимости остатков построенной модели временного ряда урожайности.

Решение Используем критерий Дарбина-Уотсона.

1. Для расчета d-статистики подготовим $\sum_{t=2}^{10}(e_t - e_{t-1})^2 = \mathbf{169{,}40}$ (СУММКВРАЗН);

$\sum_{t=1}^{10} e_t^2 = \mathbf{61{,}56}$ (остаточная сумма квадратов) $\Rightarrow$ $d = \dfrac{\mathbf{169{,}40}}{\mathbf{61{,}56}} = \mathbf{2{,}75}$.

2. Критические уровни $d_1 = \mathbf{0{,}88}$ и $d_2 = \mathbf{1{,}32}$ (таблица d-статистик при $n = \mathbf{10}$).

3. Согласно схеме проверки: $d = \mathbf{2{,}75} \in (2;\ 4) \Rightarrow d' = 4 - 2{,}75 = \mathbf{1{,}25}$.

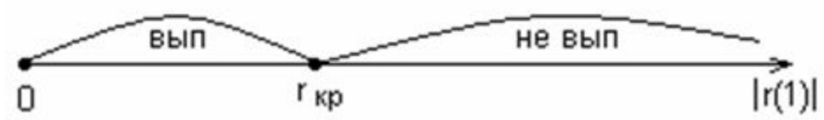
Сравним: $d' = \mathbf{1{,}25} \in (d_1 = \mathbf{0{,}88};\ d_2 = \mathbf{1{,}32}) \Rightarrow$ на основании критерия Дарбина-Уотсона нельзя сделать однозначного вывода о зависимости или независимости остаточной компоненты. Требуется выполнить проверку свойства с помощью другого критерия.

Проверка свойства независимости остатков:

критерий автокорреляции

Алгоритм критерия:

1. вычислить статистику $r(1)=\dfrac{\sum_{t=2}^{n} e_t \cdot e_{t-1}}{\sum_{t=1}^{n} e_t^2}$;
2. определить критическое значение $r_{kp}=\dfrac{1,96}{\sqrt{n}}$ (при уровне значимости $\alpha=5\%$);
3. сравнить модуль $|r(1)|$ с критическим r_{kp}:



Пример В учебных целях проанализировать выполнение свойства независимости остатков построенной модели с помощью критерия автокорреляции.

Решение

1. Для расчета коэффициента автокорреляции $r(1)$ подготовим

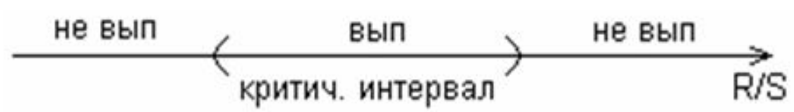
$$\sum_{t=2}^{10} e_t \cdot e_{t-1} = -27,97 \text{ (СУММПРОИЗВ)}; \quad \sum_{t=1}^{10} e_t^2 = 61,56 \quad \Rightarrow \quad r(1)=\frac{-27,97}{61,56}=-0,45.$$

2. Критическое значение коэффициента автокорреляции при $\alpha=5\%$ составляет $r_{kp}=\dfrac{1,96}{\sqrt{10}}=0,62$.
3. Сравним: $|r(1)|=0,45 < r_{kp}=0,62 \Rightarrow$ ***свойство независимости остаточной компоненты <u>выполняется</u>***.

Проверка свойства нормального распределения остатков:
R/S критерий

Алгоритм критерия:

1. вычислить статистику $\boldsymbol{R/S = \dfrac{e_{\max} - e_{\min}}{S(e)}}$;
2. определить критический интервал (таблица ***R/S*** – критерия);
3. сопоставить фактическую величину ***R/S*** с критическим интервалом:



Пример Проанализировать выполнение свойства нормального распределения остатков построенной модели временного ряда урожайности.

Решение Используем *R/S* критерий.

1. Подготовим $\boldsymbol{e_{\max} = 3{,}58}$ (МАКС); $\boldsymbol{e_{\min} = -5{,}08}$ (МИН);

$\boldsymbol{S(e) = 2{,}77}$ (стандартная ошибка модели) $\Rightarrow$ $\boldsymbol{R/S = \dfrac{3{,}58 - (-5{,}08)}{2{,}77} = 3{,}12}$.

2. Критический интервал при $\boldsymbol{n = 10}$ составляет $\boldsymbol{(2{,}67;\ 3{,}69)}$.

3. Сравним: $\boldsymbol{3{,}12 \in (2{,}67;\ 3{,}69)}$ $\Rightarrow$ для построенной модели ***свойство нормального распределения остаточной компоненты выполняется***.

Проведенная проверка предпосылок регрессионного анализа показала, что для модели выполняются ***все условия Гаусса-Маркова*** $\Rightarrow$ выбранный для ее построения метод (МНК) оптимален.

Прогнозирование
с помощью модели временного ряда

Прогнозирование основано на математическом ***методе экстраполяции*** $\boxed{\hat{y}^* = f(t^*)}$

$\hat{y}^*$ - прогнозируемое для момента $t^* = n + k$ значение признака Y (***точечный прогноз***),
n – номер последнего фактического наблюдения, k – период упреждения.

В случае линейной модели $\boxed{\hat{y}^* = a + b \cdot t^*}$ – ожидаемое в момент t^* значение признака Y.

Для учёта случайных колебаний найденное прогнозное значение $\hat{y}^*$ дополняют расчетом доверительного интервала $\left(\hat{y}^* - u;\ \hat{y}^* + u\right)$, т.е. выполняют ***интервальное прогнозирование***.

Прогнозный интервал показывает, в какие границы с требуемой вероятностью попадёт фактическое значение признака Y.

В случае линейной модели размах прогнозного интервала определяют по формуле

$$u = t_{kp} \cdot S(e) \cdot \sqrt{1 + \frac{1}{n} + \frac{\left(t^* - \bar{t}\right)^2}{\sum\limits_{t-1}^{n} \left(t - \bar{t}\right)^2}}\ .$$

Затем рассчитывают границы интервала вокруг точечного прогнозного значения

$$u_{нижн} = \hat{y}^* - u \quad \text{и} \quad u_{верх} = \hat{y}^* + u\ .$$

Размах u характеризует ошибку прогнозирования. Чем больше размах, тем менее точным является прогноз.

Пример С помощью линейной модели рассчитать точечную и интервальную прогнозную оценку урожайности на будущий год. Исходные данные, результаты моделирования и прогнозирования показать на чертеже.

Решение

«Будущий год» соответствует периоду упреждения $\boldsymbol{k=1}$, при этом $\boldsymbol{t^{*}=n+k=11}$.

Согласно уравнению модели получим точечную прогнозную оценку

$$\hat{y}_{11}^{*}=\mathbf{10{,}467+0{,}770\cdot 11=18{,}93}$$

$\Rightarrow$ ***ожидаемое значение урожайности в будущем году составит около 18,93 ц/га***.

Примем доверительную вероятность $\gamma=95\%$ и найдем границы прогнозного интервала.

Подготовим: $S(e)=2{,}77$ (строка «стандартная ошибка» итогов РЕГРЕССИИ);

$\bar{t}=5{,}5$ (СРЗНАЧ); $\sum_{t=1}^{10}\left(t-\bar{t}\right)^2=82{,}5$ (КВАДРОТКЛ); $t_{kp}=2{,}31$ (СТЬЮДРАСПОБР).

По формуле вычислим размах прогнозного интервала:

$$u=t_{kp}\cdot S(e)\cdot\sqrt{1+\frac{1}{n}+\frac{\left(t^{*}-\bar{t}\right)^2}{\sum_{t=1}^{n}\left(t-\bar{t}\right)^2}}=\mathbf{7{,}75}$$

Теперь можно определить границы: $u_{ниж}=\hat{y}_{11}^{*}-u=\mathbf{11{,}19}$ и $u_{верх}=\hat{y}_{11}^{*}+u=\mathbf{26{,}68}$

$\Rightarrow$ ***с вероятностью 95% можно утверждать, что урожайность в будущем году будет от 11,19 ц/га до 26,68 ц/га***.

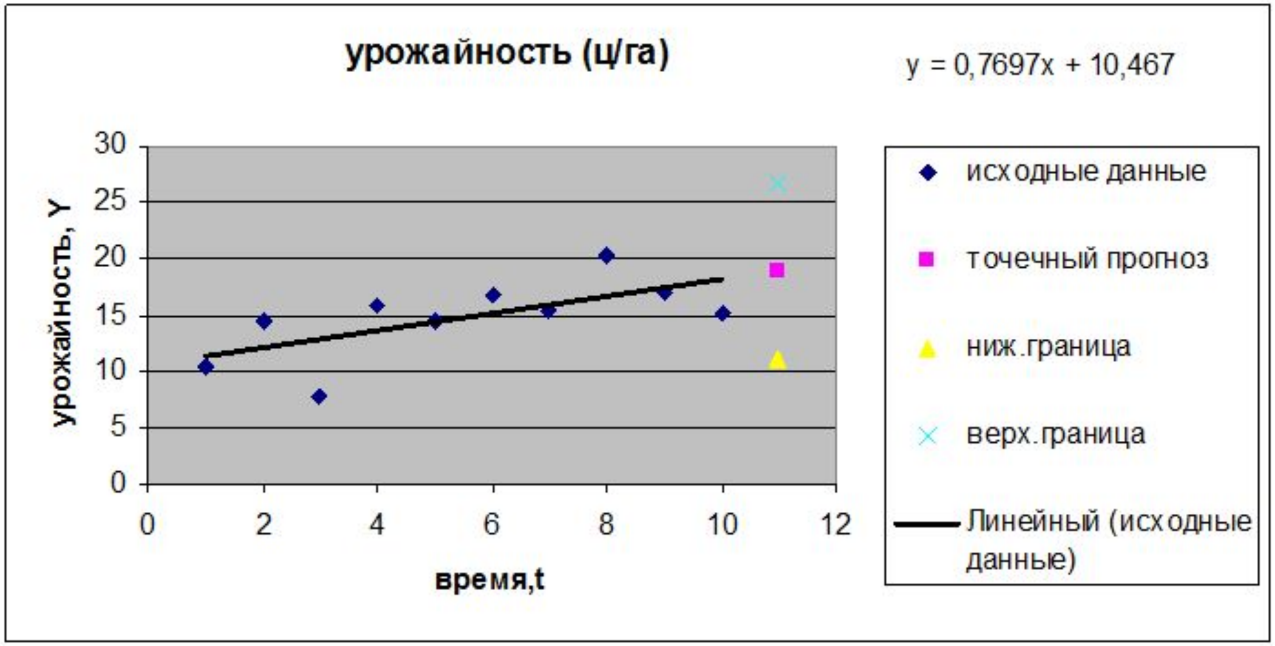
Диаграмма результатов моделирования и прогнозирования

1-ый шаг – построить диаграмму исходных данных $(t,\ y_t)$.

2-ой шаг – добавить линию тренда, показать уравнение тренда.

3-ий шаг – в опции *Исходные данные* добавить ряды:

- «точечный прогноз», $t^* = 11$, $\hat{y}^*_{11} = 18{,}93$;
- «нижняя граница», $t^* = 11$, $u_{ниж} = 11{,}19$;
- «верхняя граница», $t^* = 11$, $u_{верх} = 26{,}68$.



Большой размах прогнозного интервала говорит о низкой точности прогнозирования. Это объясняется низким качеством используемой модели.