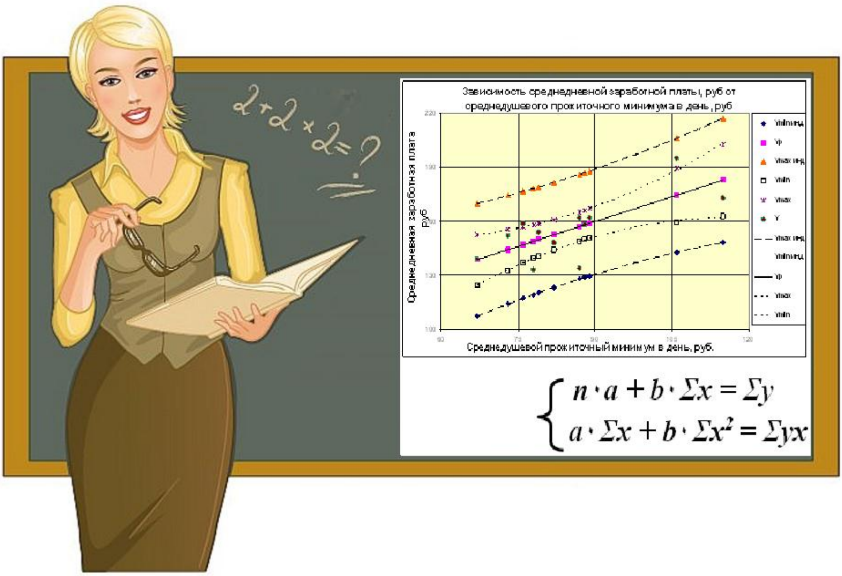
С.В. Амелин

ЭКОНОМЕТРИКА: ПРАКТИКУМ

Учебное пособие



Воронеж 2016

ФГБОУ ВО «Воронежский государственный
технический университет»

С.В. Амелин

ЭКОНОМЕТРИКА:
ПРАКТИКУМ

Утверждено Редакционно-издательским советом
университета в качестве учебного пособия

Воронеж 2016

УДК 330(075.8)

Амелин С.В. Эконометрика: практикум: учеб. пособие / С.В. Амелин. Воронеж: ФГБОУ ВО «Воронежский государственный технический университет», 2016. 125 с.

Представлены основные разделы эконометрического моделирования, необходимые экономистам (руководителям и специалистам) для анализа, прогнозирования и обоснования управленческих решений в сложных производственно-экономических ситуациях.

Издание соответствует требованиям Федерального государственного образовательного стандарта высшего образования по специальности 38.05.01 «Экономическая безопасность», по направлениям 38.03.01 «Экономика», 38.03.02 «Менеджмент», 27.03.05 «Инноватика», направленностям «Экономика предприятий и организаций», «Логистика и управление цепями поставок», «Производственный менеджмент», «Менеджмент высоких технологий», «Предпринимательство в инновационной деятельности», дисциплине «Эконометрика».

Табл. 8. Ил. 42. Библиогр.: 30 назв.

Рецензенты: кафедра информационных технологий и математических методов Воронежского государственного университета (зав. кафедрой д-р экон. наук, проф. В.В. Давнис);
канд. экон. наук, доц. И.А. Калашникова

ВВЕДЕНИЕ

Цель дисциплины «Эконометрика» состоит в подготовке будущих бакалавров менеджмента, умеющих применять современные эконометрические методы и модели для прогнозирования при обосновании управленческих решений; освоении различных способов выражения связей и закономерностей развития экономических процессов и явлений через эконометрические модели и методы проверки их адекватности, основанные на данных статистических наблюдений; освоении современного инструментария эконометрического моделирования.

Задачи дисциплины:

- обработка массивов экономических данных в соответствии с поставленной задачей, анализ, оценка, интерпретация полученных результатов и обоснование выводов;
- построение эконометрических моделей исследуемых процессов, явлений и объектов, относящихся к области профессиональной деятельности, анализ и интерпретация полученных результатов;
- прогнозирование на основе эконометрических моделей поведения экономических агентов, развития экономических процессов и явлений, на микро- и макроуровне.

В процессе проведения эконометрического моделирования важное значение имеет определение параметров моделей при анализе зависимостей между экономическими показателями. В математическом моделировании существуют два понятия, отражающие причинно-следственные связи: функциональная и корреляционная зависимость. Под функциональной понимается такая связь между величинами, когда значение зависимой величины – функции полностью и однозначно определяется значением других переменных величин – аргументов. Функциональная зависимость чаще всего встречается в естественных науках, например в физике.

Корреляционная зависимость имеет место, когда каждому значению одной величины соответствует множество случайных значений другой, возникающих с определенной вероятностью. Эту вероятность можно установить с помощью статистических исследований. При корреляционной связи изменение одной величины вызывает изменение среднего значения другой величины. При изучении экономических явлений мы имеем дело не с функциональными, а с корреляционными зависимостями. При парной корреляции наблюдается связь между двумя величинами. При множественной корреляции определенным значениям нескольких влияющих величин – факторов соответствует множество случайных значений зависимой результативной величины, распределенных по известному закону. Хотя такие зависимости не являются строго определенными, и выявляются лишь в виде некоторой тенденции, однако можно подобрать некоторую функцию, которая приближенно (в среднем) будет отражать зависимость результативной величины от нескольких факторов. Такая функция называется уравнением регрессии, а её график – линией регрессии. Уравнения регрессии могут быть линейными и нелинейными.

Корреляционный и регрессионный анализ позволяет решать такие задачи, которые пока другими методами выполнить нельзя, как, например, определение совместного и раздельного влияния многих взаимно связанных и одновременно действующих факторов на результативный признак.

С помощью корреляционного и регрессионного анализа можно рассчитать коэффициенты корреляции, которые оценивают силу связи между отдельными признаками (показателями), подобрать уравнение регрессии, которое определяет форму этой связи, и установить достоверность (реальность) существования связи.

Процесс корреляционного и регрессионного анализа экономических явлений состоит из следующих этапов:

а) предварительная обработка статистических данных и выбор факторов-аргументов (факторных признаков);

б) оценка тесноты связи между признаками и выявление форм связи;

в) разработка многофакторной и однофакторных моделей изучаемого явления и их анализ;

г) использование результатов анализа для совершенствования планирования и управления данным явлением.

Разработка регрессионной модели изучаемого экономического процесса или явления осуществляется на основе метода наименьших квадратов, согласно которому обеспечивается наилучшее приближение оценок результативного признака, рассчитанных по уравнению регрессии, к их фактическим значениям.

Важнейшими параметрами, характеризующими регрессионную модель, являются:

а) коэффициенты парной корреляции, которые определяют силу связи между двумя признаками ($\mathbf{r_{ij}}$);

б) коэффициент множественной корреляции, который определяет связь результативного признака с совокупностью факторных признаков (R);

в) коэффициент множественной детерминации, который определяет удельный вес совместного влияния всех включенных в модель факторных признаков на вариацию результативного признака (R^2).

Результаты корреляционно-регрессионного анализа позволяют выявить факторы, оказывающие существенное влияние на исследуемый экономический процесс или явление. Они могут также быть использованы для разработки норм и нормативов.

Разработка прогнозов развития экономических процессов осуществляется с помощью метода экстраполяции на основе данных рядов динамики. Для прогнозирования наиболее часто применяется метод аналитического выравнивания с помощью регрессионных моделей. При этом подбираются аналитические функции (уравнения регрессии), наиболее точно характеризующие развитие конкретных явлений или процессов во времени. Найденная функция позволяет получить выравненные значения уровней ряда динамики, т.е. его теоретические оценки. Определение параметров функции, по которой выравнивается ряд динамики, осуществляется методом наименьших квадратов, который требует, чтобы сумма квадратов отклонений значений, лежащих на линии регрессии (теоретических оценок уровней), от фактических значений уровней была минимальной, т.е. чтобы соблюдалось условие

$$\boldsymbol{S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \rightarrow min.}$$

где $\boldsymbol{y_i}$ – фактические значения зависимой переменной;
$\hat{y}_i$ – значения переменной, вычисленные по уравнению регрессии.

ПРАКТИЧЕСКИЕ ЗАНЯТИЯ

Практическое занятие № 1. Математическое ожидание, ковариация, дисперсия, корреляция

Случайная переменная – это любая переменная, значение которой не может быть точно предсказано. *Дискретной* называется случайная величина, имеющая определенный набор возможных значений (например, число выпавшее при бросании игральной кости).

Математическое ожидание дискретной случайной величины – это взвешенное среднее всех ее возможных значений по генеральной совокупности, причем в качестве весового коэффициента берется вероятность соответствующего исхода.

Предположим, что случайная величина X может принимать n конкретных значений (x_1, x_2, …, x_n) и вероятность получения x_i равна p_i . Тогда математическое ожидание M(x) или μ_x

$$M(x) = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n .$$

Свойства математического ожидания:

1) M(k) = k, где k – константа;
2) M(k•X) = k·M(X);
3) M(X ± Y) = M(X) ± M(Y);
4) M(X•Y) = M(X) • M(Y), где X и Y – независимые случайные величины;
5) M(X ± k) = M(X) ± k;
6) M(X – μ_x) = 0, где μ_x = M(X).

Задание 1. 1) Определить M(y), где y = a + bx – x^2, a и b – константы.

Оценкой математического ожидания для выборки размера n является выборочное среднее значение

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i .$$

Ковариация является мерой взаимосвязи между двумя случайными величинами. Ковариация случайных величин X и Y определяется как математическое ожидание произведения отклонения этих величин от своих математических ожиданий

$$Cov(X,Y) = \mathrm{M}[(X - \mu_x)(Y - \mu_y)].$$

Свойства ковариации двух случайных величин:

1) $Cov(X,Y) = 0$, если X и Y независимы;

2) $Cov(X,Y) = \mathrm{M}(X \bullet Y) - \mu_x \bullet \mu_y$;

3) $|Cov(X,Y)| \le \sigma_x \sigma_y$, где $\sigma_x \sigma_y$ – средние квадратические отклонения величин X и Y.

Выборочная ковариация (корреляционный момент):

$$\hat{Cov}(X,Y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \cdot \frac{1}{n}\sum_{i=1}^{n} y_i =$$

$$= \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}).$$

Правила расчета ковариации:

1) если $Y=V+W$, то $\hat{Cov}(X,Y) = \hat{Cov}(X,V) + \hat{Cov}(X,W)$;

2) если $Y=kZ$, где k – константа, то $\hat{Cov}(X,Y) = k \cdot \hat{Cov}(X,Z)$;

3) если $Y= k$, где k – константа, то $\hat{Cov}(X,Y) = 0$.

Задание 2. Допустим $y = a + bz$, где a и b – константы. Найти ковариацию между переменными величинами x и y.

Средним квадратическим отклонением (СКО, стандартным отклонением) σ_x случайной величины X называется арифметическое значение корня квадратного из ее дисперсии:

$$\sigma_x = \sqrt{D(x)}.$$

Теоретическая дисперсия случайной величины X – это мера разброса для ее вероятностного распределения. Она определяется как математическое ожидание квадрата разности между величиной X и ее средним значением по генеральной совокупности

$$D(X) = \sigma_x^2 = M(X - \mu_x)^2 = \sum_{i=1}^{n} (x_i - \mu_x)^2 p_i = M(X^2) - \mu_x^2.$$

Свойства дисперсии случайной величины:

1) $\boldsymbol{D}(\boldsymbol{k}) = 0$, где $\boldsymbol{k}$ – константа;

2) $\boldsymbol{D}(\boldsymbol{kX}) = \boldsymbol{k}^2\boldsymbol{D}(\boldsymbol{X})$;

3) $\boldsymbol{D}(\boldsymbol{X}) = \mathrm{M}(\boldsymbol{X}^2) - \mu_x^2$;

4) $\boldsymbol{D}(\boldsymbol{X} \pm \boldsymbol{Y}) = \boldsymbol{D}(\boldsymbol{X}) + \boldsymbol{D}(\boldsymbol{Y}) \pm 2\boldsymbol{Cov}(\boldsymbol{X},\boldsymbol{Y})$

Оценкой дисперсии по выборке размера $\boldsymbol{n}$ является выборочная дисперсия

$$s_x^2 = \overline{(x - \overline{x})^2} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \overline{x})^2 = \overline{x^2} - \overline{x}^2;$$

несмещенная состоятельная оценка генеральной дисперсии

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2.$$

Задание 3. Имеются данные по шести семьям (домохозяйствам):

Семья	Доход семьи (x)	Расходы на питание и одежду (y)	Расходы на питание (v)	Расходы на одежду (w)	Расходы на питание и одежду (вторая выборка) (z)
1	3000	1100	850	250	2200
2	2500	850	700	150	1700
3	4000	1200	950	250	2400
4	6000	1600	1150	450	3200
5	3300	1000	800	200	2000
6	4500	1300	950	350	2600

Определить ковариацию между величинами $\boldsymbol{x}$ и $\boldsymbol{y}$, $\boldsymbol{x}$ и $\boldsymbol{v}$, $\boldsymbol{x}$ и $\boldsymbol{w}$, $\boldsymbol{x}$ и **z**. Если каждая семья в выборке имеет по два взрослых человека, чему будет равна ковариация между x и числом взрослых в семье?

Задание 4. Допустим годовой доход индивида (y) определяется выражением:

$$\boldsymbol{y = 10000 + 500\, s + 200\, t.}$$

Для выборки из пяти индивидов определить: 1) ковариацию между величинами ***x*** и ***y***, ***x*** и **s**, ***x*** и ***t***; 2) дисперсии величин ***y***, **s** и ***t***:

Индивид	Возраст (x), годы	Годы (s) обучения	Трудовой стаж (t)	Доход (y), ден. ед.
1	18	11	1	15700
2	29	14	6	18200
3	33	12	8	17600
4	35	16	10	20000
5	45	12	5	17000

Домашнее задание

Поскольку ковариация – это величина размерная, то для облегчения оценки степени зависимости случайных величин также используется, лишенный этого недостатка коэффициент корреляции.

Теоретический коэффициент корреляции двух случайных величин определяется отношением их ковариации к произведению средних квадратических отклонений этих величин:

$$\rho_{xy} = \frac{Cov(x,y)}{\sigma_x \sigma_y}.$$

Коэффициент корреляции – это безразмерная величина, характеризующая тесноту линейной зависимости между случайными величинами.

Свойства коэффициента корреляции:

1) $-1 \le \rho \le 1$;

2) $\rho = 0$, если случайные величины Y и X независимы;

3) если $|\rho| = 1$, то между случайными величинами существует линейная функциональная зависимость.

Из независимости двух случайных величин следует их некоррелированность, т.е. равенство $\rho = 0$. Однако некоррелированность двух случайных величин еще не означает их независимость. (Поскольку между ними может быть нелинейная зависимость, а коэффициент корреляции определяется для линейной зависимости).

Выборочный коэффициент парной корреляции определяется по следующим выражениям:

$$r=\frac{\hat{Cov}(x,y)}{s_x s_y}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n\cdot s_x s_y}=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\sqrt{\overline{x^2}-\bar{x}^2}\sqrt{\overline{y^2}-\bar{y}^2}};$$

$$r=\frac{\frac{\sum_{i=1}^{n}x_i y_i}{n}-\frac{\sum_{i=1}^{n}x_i}{n}\cdot\frac{\sum_{i=1}^{n}y_i}{n}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2-\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2}\cdot\sqrt{\frac{1}{n}\sum_{i=1}^{n}y_i^2-\left(\frac{1}{n}\sum_{i=1}^{n}y_i\right)^2}}=$$

$$=\frac{n\sum_{i=1}^{n}x_i y_i-\sum_{i=1}^{n}x_i\cdot\sum_{i=1}^{n}y_i}{\sqrt{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}\cdot\sqrt{n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2}}.$$

Задание 5. По данным о годовых темпах прироста продукции рассчитать коэффициент парной корреляции:

Страна	Занятость (прирост) (х)	Производительность (прирост) (у)
Австрия	2,0	4,2
Бельгия	1,5	3,9
Канада	2,3	1,3
Дания	2,5	3,2
Франция	1,9	3,8
Италия	4,4	4,2
Япония	5,8	7,8
Нидерланды	1,9	4,1
Норвегия	0,5	4,4
ФРГ	2,7	4,5
Англия	0,6	2,8
США	0,8	2,6

Задание 6. Определить коэффициент корреляции по двум признакам (расходы на покупку продовольственных товаров в общих расходах, % - *у* и среднедневная заработная плата одного работающего, руб. - ***х***) для семи территорий:

Район	(y)	(x)
Удмуртия	68,8	45,1
Свердл. обл	61,2	59,0
Башкирия	59,9	57,2
Челябинск. обл	56,7	61,8
Пермск. обл	55,0	58,8
Курганск. обл	54,3	47,2
Оренбургск. обл	49,3	55,2

Практическое занятие № 2.
Парный регрессионный анализ

Линейное уравнение регрессии $\hat{y} = b_0 + b_1 x$, ($\hat{y}$ - расчётное значение зависимой переменной, x – факторная переменная).

Параметры уравнения регрессии b_0 и b_1 находят из системы нормальных уравнений

$$\begin{cases} b_0 \cdot n + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i . \end{cases}$$

или $b_1 = \frac{C\hat{o}v(X,Y)}{s_x^2}$, $\hat{y} - \bar{y} = b_1(x - \bar{x})$; $b_0 = \bar{y} - b_1 \bar{x}$.

Выборочная остаточная дисперсии:

$$s^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2}.$$

Оценка дисперсии групповой средней при заданном значении x_k:

$$s_{\hat{y}}^2 = s_e^2 \left(\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right).$$

Доверительный интервал для условного математического ожидания зависимой переменной для принятого значения x_k

$$\hat{y}_k - t_{1-\alpha;n-m-1} s_{\hat{y}} \le M_x(Y) \le \hat{y}_k + t_{1-\alpha;n-m-1} s_{\hat{y}},$$

где $t_{\gamma,df}$ – критерий Стьюдента, γ – уровень надежности, $\gamma = 1-\alpha$, α – уровень значимости, df – количество степеней свободы, $df = n - m - 1$, n – количество наблюдений, m – количество независимых переменных.

Дисперсия оценки индивидуального значения

$$s_{\hat{y}^*}^2 = s_e^2 \left(1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right).$$

Доверительный интервал для индивидуального значения

$$\hat{y}_{np} - t_{1-\alpha;n-m-1} s_{\hat{y}^*} \le y_{np}^* \le \hat{y}_{np} + t_{1-\alpha;n-m-1} s_{\hat{y}^*}.$$

Интервальная оценка коэффициента регрессии

$$b_1 - t_{1-\alpha;n-m-1} \frac{s_e}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \le \beta_1 \le b_1 + t_{1-\alpha;n-m-1} \frac{s_e}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2}}.$$

Доверительный интервал для дисперсии остатков

$$\frac{n s_e^2}{\chi^2_{\alpha/2;n-m-1}} \le \sigma^2 \le \frac{n s_e^2}{\chi^2_{1-\alpha/2;n-m-1}},$$

где χ^2 - критерий Пирсона хи-квадрат.

Задание. Проанализировать зависимость между сменной добычей угля на одного рабочего Y(т) и мощностью пласта X(м) по данным для n = 10 шахт:

i шахта	1	2	3	4	5	6	7	8	9	10
x_i	8	9	8	12	8	12	9	8	9	11
y_i	6	6	5	10	5	8	5	6	7	10

1. Изобразить имеющуюся зависимость в виде поля корреляции – точками на координатной плоскости;

2. Определить параметры уравнения регрессии;

3. Рассчитать парный коэффициент корреляции.

4. Оценить сменную добычу угля на одного рабочего для шахт с мощностью пласта 8 м;

5. Найти 95 % - ные доверительные интервалы для среднего и индивидуального значений сменной добычи угля на 1 рабочего;

6. Найти с надежностью 0,95 интервальные оценки коэффициента регрессии β_1 и дисперсии σ^2.

Практическое занятие № 3.
Оценка значимости уравнения регрессии и его параметров

Дисперсионный анализ. Сумма квадратов отклонений:

1) общая $\boldsymbol{Q=\sum_{i=1}^{n}(y_i-\bar{y})^2}$; $\mathbf{df}_{общ}=(\boldsymbol{n}-1)$ - число степеней свободы;

2) обусловленная регрессией $\boldsymbol{Q_R=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2}$; $\mathrm{df}_R=\boldsymbol{m}$;

3) остаточная $\boldsymbol{Q_e}=\boldsymbol{Q}-\boldsymbol{Q_R}$; $\boldsymbol{Q_e=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}$; $\mathrm{df}_e=(\boldsymbol{n}-\boldsymbol{m}-1)$.

Значимость уравнения по критерию Фишера-Снедекора определяется условием:

значение $\boldsymbol{F}$-статистики $\boldsymbol{F=\frac{Q_R(n-m-1)}{Q_e\cdot m}>F_{\alpha;m;n-m-1}}$.

Значимость коэффициентов уравнения регрессии по критерию Стьюдента определяется условием:

$$|\boldsymbol{t}|=\frac{|\boldsymbol{b}_1|}{\boldsymbol{s_e}}\sqrt{\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}}(\boldsymbol{x_i}-\bar{\boldsymbol{x}})^2}>\boldsymbol{t}_{1-\alpha;n-m-1}.$$

Дисперсия случайных факторов $s_e^2 = Q_e /(n - m - 1)$.

Условие значимости коэффициента корреляции

$$|t| = \frac{|r|\sqrt{n-m-1}}{\sqrt{1-r^2}} > t_{1-\alpha;n-m-1}$$

Задание. Торговое предприятие имеет сеть, состоящую из 10 магазинов, информация о деятельности которых представлена в таблице (торговая площадь, тыс. м^2 – х и годовой товарооборот, млн. руб. – у):

i - магазин	1	2	3	4	5	6	7	8	9	10
x_i	0,31	0,48	0,98	1,12	1,12	1,49	1,29	1,29	0,94	0,78
y_i	38,09	41,08	68,51	89,05	91,26	108,55	99,84	91,13	75,01	56,29

1. Построить диаграмму рассеяния годового товарооборота в зависимость от торговой площади и определить форму связи;

2. Оценить уравнение регрессии;

3. Определить коэффициент парной корреляции;

4. Найти 95 % - ные доверительные интервалы для среднего и индивидуального значений торговой площади;

5. Оценить на уровне α = 0,05 значимость уравнения регрессии Y по X и его параметров.

Домашнее задание

По данным 30 нефтяных компаний получено следующее уравнение регрессии между оценкой Y (ден.ед.) и фактической стоимостью X (ден.ед.) этих компаний:

$$y_x = 296 + 0.875\, x.$$

Найти 95 %-ные доверительные интервалы для среднего и индивидуального значений оценки предприятий, фактическая стоимость которых составила 1300 ден.ед., если коэффициент корреляции между переменными равен 0,76, а среднее квадратическое отклонение переменной X равно 270 ден.ед.

Практическое занятие № 4.
Множественный регрессионный анализ

Вектор коэффициентов множественного уравнения регрессии $\boldsymbol{b} = (\boldsymbol{X'X})^{-1}X'Y$.

Стандартизованные коэффициенты регрессии

$$b_j' = b_j \frac{s_{xj}}{s_y}.$$

Коэффициенты эластичности $E_j = b_j \dfrac{\bar{x}_j}{\bar{y}}$.

Оценка дисперсии коэффициента регрессии $\boldsymbol{b_j}$

$$s_{b_j}^2 = s^2 [(X'X)^{-1}]_{jj}.$$

Выборочная остаточная дисперсия

$$s^2 = \frac{e'e}{n-m-1} = \frac{\sum_{i=1}^{n} e_i^2}{n-m-1}.$$

Оценка значимости коэффициента регрессии

$$|t| = \frac{|b_j|}{s_{b_j}} > t_{1-\alpha;n-m-1}.$$

Доверительный интервал для параметра $\boldsymbol{\beta_j}$:

$$b_j - t_{1-\alpha;n-m-1}s_{bj} \le \beta_j \le b_j + t_{1-\alpha;n-m-1}s_{bj}.$$

Доверительный интервал для функции регрессии

$$\hat{y} - t_{1-\alpha;n-m-1}s_{\hat{y}} \le M_x(Y) \le \hat{y} + t_{1-\alpha;n-m-1}s_{\hat{y}},$$

где $s_{\hat{y}} = s\sqrt{X_o'(X'X)^{-1}X_o}$ - стандартная ошибка групповой средней $\hat{y}$, определяемой по уравнению регрессии.

Доверительный интервал для индивидуальных значений зависимой переменной:

$$\hat{y}_o - t_{1-\alpha;n-m-1}s_{\hat{y}o} \le y_o^* \le \hat{y}_o + t_{1-\alpha;n-m-1}s_{\hat{y}o}$$

где $s_{\hat{y}o} = s\sqrt{1 + X_o'(X'X)^{-1}X_o}$.

Доверительный интервал для дисперсии остатков

$$\frac{ns^2}{\chi^2_{\alpha/2;n-m-1}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{1-\alpha/2;n-m-1}} .$$

Задание. Проанализировать зависимость между сменной добычей угля на одного рабочего ***Y***(т) и мощностью пласта ***X***(м) по данным для n = 10 шахт:

i шахта	1	2	3	4	5	6	7	8	9	10
x_{i1}	8	9	8	12	8	12	9	8	9	11
X_{i2}	5	6	7	8	5	7	4	8	5	8
y_i	6	6	5	10	5	8	5	6	7	10

1. Изобразить имеющуюся зависимость в виде поля корреляции – точками на координатной плоскости;

2. Определить параметры уравнения регрессии;

3. Рассчитать парный коэффициент корреляции.

4. Оценить сменную добычу угля на одного рабочего для шахт с мощностью пласта 8 м;

5. Найти 95 % - ные доверительные интервалы для среднего и индивидуального значений сменной добычи угля на 1 рабочего;

6. Найти с надежностью 0,95 интервальные оценки коэффициента регрессии $\boldsymbol{\beta}_1$ и дисперсии $\boldsymbol{\sigma}^2$.

Домашнее задание

Имеются данные о выработке литья на одного работающего X_1 (т), браке литья X_2 (%) и себестоимости 1т литья Y (руб) по литейным цехам 20 заводов:

i - завод	X_{1i}	X_{2i}	y_i	i - завод	X_{1i}	X_{2i}	y_i
1	44,8	1,2	158	11	33,4	2,3	195
2	17,4	7,7	251	12	16,6	4,1	251
3	21,5	5,5	262	13	27,6	1,1	201
4	13,5	6,7	254	14	75,8	3,4	172
5	14,6	4,2	239	15	40,6	3,3	197
6	19,7	2,7	2,5	16	56,0	1,3	170
7	17,0	9,3	282	17	22,3	4,2	204
8	33,4	2,3	196	18	28,1	1,4	186
9	22,6	5,2	238	19	20,1	8,4	259
10	30,1	3,5	186	20	11,9	2,2	101

Определить величины, предусмотренные заданием на практическом занятии.

Практическое занятие № 5.
Оценка значимости множественного уравнения регрессии. Определение доверительных интервалов для коэффициентов и функции регрессии

Сумма квадратов отклонений:

1) общая $\boldsymbol{Q = Y'Y - n\,\bar{y}^2}$;

2) остаточная $\boldsymbol{Q_e = Y'Y - b'X'Y}$;

3) обусловленная регрессией $\boldsymbol{Q_R = b'X'Y - n\,\bar{y}^2}$.

Значение $\boldsymbol{F}$-статистики: $\boldsymbol{F = \dfrac{Q_R(n-m-1)}{Q_e \cdot m} > F_{\alpha;m;n-m-1}}$.

Множественный коэффициент детерминации:

$$\boldsymbol{R^2 = \frac{Q_R}{Q} = \frac{b'X'Y - n\,\bar{y}^2}{Y'Y - n\,\bar{y}^2}}.$$

Скорректированный коэффициент детерминации

$$\hat{R}^2 = 1 - \frac{n-1}{n-m-1}(1-R^2) = 1 - \frac{(n-1)e'e}{(n-m-1)y'y}.$$

Значимость коэффициента детерминации

$$F = \frac{R^2(n-m-1)}{(1-R^2)m} > F_{\alpha;m;n-m-1}.$$

Задание. Торговое предприятие имеет сеть, состоящую из 10 магазинов, информация о деятельности которых представлена в таблице (i – номер магазина, x_1 – торговая площадь, тыс. м^2, x_2 – среднее число посетителей в день, тыс. чел., y – годовой товарооборот, млн. руб.):

i	1	2	3	4	5	6	7	8	9	10
x_{i1}	0,31	0,48	0,98	1,12	1,12	1,49	1,29	1,29	0,94	0,78
x_{i2}	10,24	11,01	7,51	10,81	13,72	13,92	12,27	9,89	12,36	8,54
y_i	38,09	41,08	68,51	89,05	91,26	108,55	99,84	91,13	75,01	56,29

1. Построить диаграмму рассеяния годового товарооборота в зависимость от торговой площади и определить форму связи;

2. Оценить уравнение регрессии;

3. Определить коэффициент парной корреляции;

4. Найти 95 % - ные доверительные интервалы для среднего и индивидуального значений торговой площади;

5. Оценить на уровне $\alpha = 0,05$ значимость уравнения регрессии Y по X и его параметров.

Практическое занятие № 6.
Линейные регрессионные модели с переменной структурой. Фиктивные переменные. Тест Чоу

Регрессионная модель с фиктивной переменной имеет вид:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_m x_{im} + \alpha z_{i1} + \varepsilon_i, \; i = 1, \ldots, n,$$

где $z_{i1} = \begin{cases} 1, \textit{если } \mathbf{i} - \textit{й работник мужского пола}; \\ 0, \textit{если } \mathbf{i} - \textit{й работник женского пола}. \end{cases}$

Согласно критерию Г.Чоу нулевая гипотеза H_0 о равенстве параметров двух моделей отвергается на уровне значимости α, если статистика

$$F = \frac{\left(\sum_{i=1}^{n} e_i^2 - \sum_{i=1}^{n_1} e_i^2 - \sum_{i=n_1+1}^{n} e_i^2\right)(n-2m-2)}{\left(\sum_{i=1}^{n_1} e_i^2 - \sum_{i=n_1+1}^{n} e_i^2\right)(m+1)} > F_{\alpha;m+1;n-2m-2},$$

где $\sum_{i=1}^{n} e_i^2$, $\sum_{i=1}^{n_1} e_i^2$, $\sum_{i=n_1+1}^{n} e_i^2$ - остаточные суммы квадратов соответственно для объединенной, первой и второй выборок размером: $n = n_1 + n_2$; m – количество факторных переменных.

Задание. Исследовать зависимость между результатами письменных вступительных и семестровых (на 1 курсе) экзаменов по математике. Имеются данные о числе решаемых задач на вступительных экзаменах X (из 10 заданий) и экзаменов по курсу Y (из 7 заданий) для 10 студентов, а также распределение этих студентов по фактору «пол»:

i - студент	Число решаемых задач		Пол студента
	x_i	y_i	
1	10	6	муж.
2	6	4	жен.
3	8	4	муж.
4	8	5	жен.
5	6	4	жен.
6	7	7	муж.
7	5	9	жен.
8	7	4	муж.
9	9	7	муж.
10	6	3	жен.

Построить линейную регрессионную модель Y по X с использованием фиктивной переменной по фактору «пол». Можно ли считать, что эта модель одна и та же для юношей и девушек? Определить коэффициент детерминации. Оценить значимость уравнения и его коэффициентов регрессии.

Домашнее задание

1) По данным о двадцати рабочих цеха оценить регрессию заработной платы рабочего за месяц Y ден.ед от количественного фактора X – возраст рабочего (лет) и качественного фактора Z – пол:

i	y_i	x_i	z_i	i	y_i	x_i	z_i
1	250	28	Ж	11	300	29	Ж
2	350	30	М	12	400	40	М
3	200	25	М	13	300	36	Ж
4	400	48	М	14	200	32	Ж
5	220	30	Ж	15	350	23	М
6	320	40	М	16	400	45	М
7	390	40	М	17	380	38	Ж
8	360	38	М	18	400	40	М
9	260	29	Ж	19	380	50	М
10	250	25	М	20	400	47	М

2) Критерий (тест) Чоу. В результате исследования зависимости заработной платы двадцати рабочих от их возраста были построены отдельные модели для рабочих-мужчин и рабочих-женщин, а также объединенная модель. Определить следовало ли разбивать выборку на части и существует ли на исследуемом предприятии дискриминация в оплате труда по фактору «пол»:

Выборка	Оцененное уравнение	R^2	Сумма квадратов остатков
Мужчины	$\boldsymbol{y_i}=55{,}00+7{,}39\boldsymbol{x_1}$ $t_{набл}$ (1,39) (6,88)	0,735	24027
Женщины	$\boldsymbol{y_i}=59{,}43+7{,}3\boldsymbol{x_1}$ $t_{набл}$ (1,43) (6,48)	0,712	24831
Объединенная	$\boldsymbol{y_i}=62{,}27+7{,}23\boldsymbol{x_1}$ $t_{набл}$ (4,29) (4,104)	0,728	24888

Практическое занятие № 7. Мультиколлинеарность. Частные коэффициенты корреляции

Частный коэффициент корреляции между переменными x_i и x_j при элиминировании влияния переменной x_k:

$$r_{ij.k} = \frac{r_{ij} - r_{ik} r_{jk}}{\sqrt{1 - r_{ik}^2} \cdot \sqrt{1 - r_{jk}^2}},$$

где выборочный коэффициент парной корреляции r_{ij} или r_{xy} определяется выражением:

$$r = \frac{\hat{Cov}(x, y)}{s_x s_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x s_y} =$$

$$= \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2} \cdot \sqrt{n \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i \right)^2}}.$$

Значимость коэффициента корреляции определяется с помощью t-критерия Стьюдента

$$|t| = \frac{|r| \sqrt{n-2}}{\sqrt{1 - r^2}} > t_{1-\alpha; n-2}\ .$$

Задание. Для исследования зависимости между производительностью труда X_1, возрастом X_2 и производственным стажем X_3 произведена выборка из 100 рабочих одной специальности. Определены парные коэффициенты корреляции, которые оказались значимы: r_{12}=0,2; r_{13}=0,41; r_{23}=0,82. Определить частные коэффициенты корреляции и оценить их значимость на уровне α = 0,05.

Домашнее задание

С целью исследования влияния факторов X_1 - среднемесячного количества профилактических наладок автоматической линии и X_2 – среднемесячного количества сбоев на показатель Y – среднемесячную характеристику качества продукции (в баллах) по данным 37 предприятий были вычислены парные коэффициенты корреляции: $r_{y,x1}$=0,105, r_{yx2}=0,024 и r_{x1x2}=0,996. Определить частные коэффициенты корреляции $r_{yx1/x2}$ и $r_{yx2/x1}$ и определить их значимость на 5 %-ном уровне.

Практическое занятие № 8.
Характеристики временных рядов

Выборочный коэффициент автокорреляции временного ряда $\boldsymbol{y_t}$ для лага τ:

$$\boldsymbol{r(\tau)=\frac{(n-\tau)\sum\limits_{t=1}^{n-\tau} y_t y_{t-\tau}-\sum\limits_{t=1}^{n-\tau} y_y \sum\limits_{t=1}^{n-\tau} y_{y-\tau}}{\sqrt{(n-\tau)\sum\limits_{t=1}^{n-\tau} y_t^2-(\sum\limits_{t=1}^{n-\tau} y_y)^2}\cdot\sqrt{(n-\tau)\sum\limits_{t=1}^{n-\tau} y_{t-\tau}^2-(\sum\limits_{t=1}^{n-\tau} y_{t+\tau})^2}}}\,.$$

Выборочный частный коэффициент автокорреляции 1-го порядка между членами временного ряда y_t и y_{t+2} при элиминировании влияния y_{t+1} определяется из выражения:

$$\boldsymbol{r_{частн}(2)=r_{02.1}=\frac{r(2)-r(1)r(1,2)}{\sqrt{1-r^2(1)}\cdot\sqrt{1-r^2(1,2)}}},$$

где r(1), r(1,2), r(2) – выборочные коэффициенты автокорреляции между $\boldsymbol{y_t}$ и $\boldsymbol{y_{t+1}}$, $\boldsymbol{y_{t+1}}$ и $\boldsymbol{y_{t+2}}$, $\boldsymbol{y_t}$ и $\boldsymbol{y_{t+2}}$, $\boldsymbol{t} = 1, \ldots, \boldsymbol{n}$.

Задание 1. Имеются данные, отражающие спрос на некоторый товар за восьмилетний период (усл.ед.), т.е. временной ряд спроса $\boldsymbol{y_t}$:

Год, t	1	2	3	4	5	6	7	8
Спрос, y_t	213	171	291	309	317	362	351	361

Найти среднее значение, среднее квадратическое отклонение, коэффициенты автокорреляции для лагов τ = 1; 2 и частный коэффициент автокорреляции первого порядка.

Согласно методу наименьших квадратов параметры линейного тренда $\hat{y}_t = b_0 + b_1 t$ находятся из системы нормальных уравнений

$$\begin{cases} b_0 n + b_1 \sum\limits_{t=1}^{n} t = \sum\limits_{t=1}^{n} y_t, \\ b_0 \sum\limits_{t=1}^{n} t + b_1 \sum\limits_{t=1}^{n} t^2 = \sum\limits_{t=1}^{n} y_t t. \end{cases}$$

Сумма квадратов отклонений:

1) обусловленная регрессией $Q_R = SS_{регр} = \sum\limits_{t=1}^{n} (\hat{y}_t - \bar{y}_t)^2$;

2) общая $Q = SS_{общ} = \sum\limits_{t=1}^{n} (y_t - \bar{y}_t)^2$;

3) остаточная $Q_e = Q - Q_R = SS_{ост} = \sum\limits_{t=1}^{n} (\hat{y}_t - y_t)^2$.

Значение F-статистики: $F = \dfrac{Q_R(n-m-1)}{Q_e \cdot m}$.

Оценка $\mathbf{s^2}$ остаточной дисперсии: $s^2 = \dfrac{\sum\limits_{t=1}^{n} e_t^2}{n-m-1} = \dfrac{\sum\limits_{t=1}^{n} e_t^2}{n-2}$.

Оценка дисперсии групповой средней

$$s_{\hat{y}_t}^2 = s^2 \left(\frac{1}{n} + \frac{(t_o - \bar{t})^2}{\sum\limits_{t=1}^{n} (t - \bar{t})^2} \right).$$

Интервальная оценка прогноза среднего значения

$$\hat{y}_t - t_{1-\alpha;n-m-1} s_{\hat{y}_t} \le M_t(Y) \le \hat{y}_t + t_{1-\alpha;n-m-1} s_{\hat{y}_t}.$$

Дисперсия оценки прогноза индивидуального значения

$$s^2_{\hat{y}_{t_0}} = s^2\left(1+\frac{1}{n}+\frac{(t_o-\bar{t})^2}{\sum_{t=1}^{n}(t-\bar{t})^2}\right).$$

Интервальная оценка прогноза индивидуального значения

$$\hat{y}_t - t_{1-\alpha;n-m-1}s_{\hat{y}(t_0)} \le y(t_o) \le \hat{y}_t + t_{1-\alpha;n-m-1}s_{\hat{y}(t_0)}$$

Задание 2. По данным из предыдущего задания найти уравнение неслучайной составляющей – линейного тренда для временного ряда $\boldsymbol{y_t}$. Определить значимость уравнения тренда. Найти интервальные оценки прогноза для модели временного ряда.

Домашнее задание

Имеются данные, отражающие динамику курса акций некоторой компании (ден.ед):

t	y_t	t	y_t	t	y_t
1	971	6	727	11	823
2	1166	7	752	12	1112
3	1644	8	1019	13	1386
4	907	9	972	14	1428
5	957	10	815	15	1364

Определить параметры линейного тренда, оценить значимость уравнения, дать точечный и интервальный прогноз среднего и индивидуального значений курса акций в момент $\boldsymbol{t}$=16, т.е. на глубину в один интервал времени.

Рассчитать параметры авторегрессионной модели

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \ (t = 1, 2, \ldots, n)$$

ЛАБОРАТОРНЫЕ РАБОТЫ

Лабораторная работа № 1.
Расчет параметров линейной модели при аппроксимации опытных данных с помощью функции Excel ЛИНЕЙН

Цель работы: изучение методики использования функции рабочего листа Excel ЛИНЕЙН для нахождения параметров линейного уравнения регрессии.

Исходные положения. При вызове Excel на экране появится окно электронной таблицы. В главном меню выбрать указателем мыши значок вызова справочных данных - [**?**] или [Справка] и нажать левую кнопку мыши. В появившемся меню выбрать строку *Вызов справки* или *Справка: Microsoft Excel*. В диалоговом окне справочной системы выбрать вкладку *Предметный указатель* и произвести поиск нужной темы по ключевому слову. В текстовое поле ввести с клавиатуры: *функция листа* ЛИНЕЙН. Выбрать соответствующую строчку с помощью мыши и нажать кнопку *Вывести*. Другой способ – выбрать вкладку *Содержание*, затем раскрыть строку *Справка по функциям*, выбрать *Статистические функции* и среди этих функций найти *функцию* ЛИНЕЙН. На экране появится описание функции ЛИНЕЙН. Внимательно изучите и законспектируйте основные положения описания функции: назначение функции, математическую модель, синтаксис, порядок ввода, форму представления результатов и описание получаемых результатов решения.

Порядок выполнения работы. Получив у преподавателя исходные данные для расчета, введите их в поле электронной таблицы. Отформатируйте исходные данные с помощью меню *Формат* и команды *Ячейки*. Толщину линий рамки и цвет граф введенной таблицы исходных данных подобрать в соответствии со вкусом и чувством меры.

С помощью функции ЛИНЕЙН определить параметры уравнения регрессии и коэффициент детерминации для модели множественной регрессии и для всех $\boldsymbol{m}$ однофакторных моделей в рамках рассматриваемой задачи:

$$\boldsymbol{y = b_0+b_1x_1+b_2x_2+\ldots+b_mx_m};$$

$$y = b_0+b_1x_1;\; y = b_0+b_2x_2;\; y = b_0+b_mx_m.$$

Используя найденные параметры уравнения регрессии, определить расчетные значения зависимой переменной ($y_{расч}$) и прогнозное значение функции при известных значениях влияющих переменных и найденных параметрах уравнения регрессии.

Определить значение коэффициента корреляции (**R**). Проанализируйте найденные величины коэффициентов детерминации и корреляции.

Пример выполнения работы. Консалтинговая фирма провела анализ влияния издержек реализации, включающих закупку, транспортировку, складирование на величину прибыли, получаемой фирмами, реализующими свой товар. Специалисты-консультанты, используя полученные данные как средне-нормативные, смогли определить ориентировочный размер прибыли для новой фирмы, появившейся на рынке:

	A	B	C	D	E
1		ИЗДЕРЖКИ РЕАЛИЗАЦИИ			
2	Товар	Закупка	Транспорт	Склады	Прибыль
3	Продукт 1	24782.60	8522.85	2346.19	15048.37
4	Продукт 2	333.10	239.61	69.16	148.13
5	Продукт 3	7674.20	4639.26	1886.79	2799.74
6	Продукт 4	61.80	30.69	9.49	43.02
7	Продукт 5	4308.00	1889.78	307.93	2594.29
8	Продукт 6	1025.25	1025.25	399.32	1013.83
9	Продукт 7	864.70	345.09	53.27	76.94
10	Продукт 8	1363.60	765.14	210.69	540.57
11	Продукт 9	95.40	37.65	14.55	49.41
12	Продукт 10	7381.80	3797.41	811.70	2409.09
13	Данные для расчета прибыли				
14	Продукт 11	3786.60	2980.75	363.59	
15	Продукт 12	3547.20	2024.51	291.42	
16	Продукт 13	201.70	122.70	27.15	

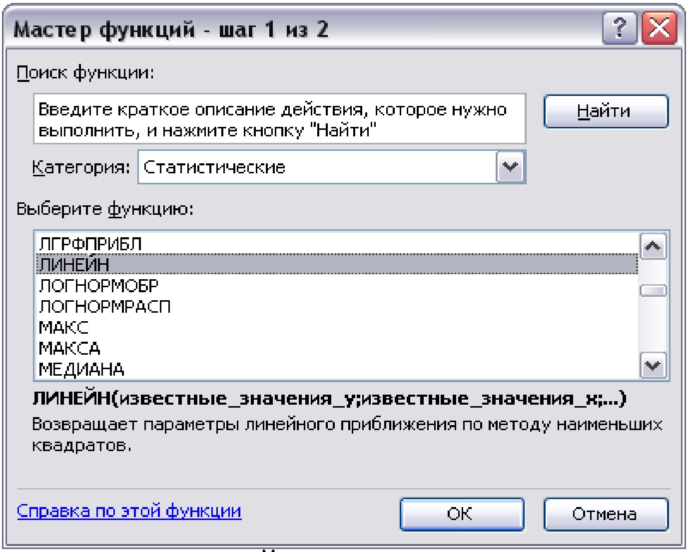


Рис. 1. Функция ЛИНЕЙН из категории *Статистические*

Ввести в таблицу исходные данные. Вызвать меню *Формат* команда *Ячейки*, вкладка *Границы* (внешние и внутренние).

Выделите на свободном месте таблицы с помощью левой кнопки мыши массив ячеек в пять строк и с количеством столбцов, равным количеству столбцов исходных данных (количеству неизвестных величин в модели $\boldsymbol{b}_0$, $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$), например, B18:E22. Вызвать функцию ЛИНЕЙН. Для этого следует нажать левой кнопкой мыши на кнопку [f_x]. Или из меню *Вставка* выбрать команду *Функция*. На экране появится диалоговое окно *Мастера функций*. Выберите в списке Категория – Статистические, а в списке Функция – ЛИНЕЙН. Щелкните по кнопке [ОК]. В появившемся диалоговом окне функции ЛИНЕЙН в соответствующие текстовые поля введите аргументы функции: *известные_значения_y* – адреса E3:E12; *известные_значения_x* – адреса B3:D12; *константа* (для получения свободного члена уравнения регрессии) – ввести число 1 или слово ИСТИНА; *статистика* (для получения коэффициента детерминации $\boldsymbol{R}^2$) – ввести число 1 или слово ИСТИНА. Для ввода функции (как формулы массива) щелкните по формуле в строке формул (или нажмите кнопку F2) и нажмите сочетание клавишей Ctrl+Shift+Enter.

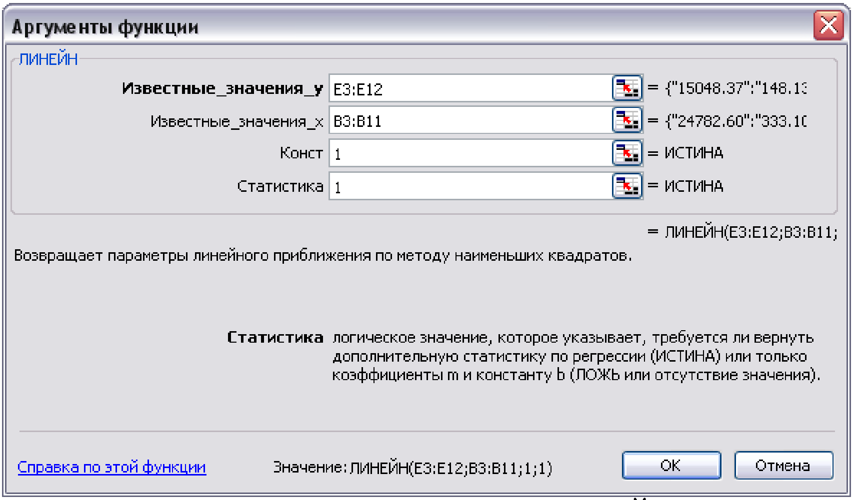


Рис. 2. Ввод аргументов функции ЛИНЕЙН

В выделенном массиве ячеек появляется следующая информация:

	B	C	D	E
18	1,0723562	-1,52565	1,021341	174,507899
19	0,9372989	0,519529	0,116526	243,260813
20	0,9909402	531,455	#Н/Д	#Н/Д
21	218,75572	6	#Н/Д	#Н/Д
22	185358993	1694666	#Н/Д	#Н/Д

Из таблицы следует: свободный член уравнения регрессии $\mathbf{b}_0$ = 174.507 коэффициенты уравнения $\mathbf{b}_1$ = 1.021, $\mathbf{b}_2$ = -1.525, $\mathbf{b}_3$= 1.072. Коэффициент детерминации $\mathbf{R}^2$=0.99. Для расчета коэффициента корреляции выделить свободную ячейку, вызвать в категории *Математические* функцию с именем КОРЕНЬ. В качестве аргумента ввести адрес ячейки, содержащей значение $\mathbf{R}^2$. Получим **R**=0.995.

Для определения расчетных значений зависимой величины $y_{расч}$. в ячейку, расположенную рядом с первым значением $y_{факт.}$ ввести формулу:

$$y_{расч} = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3.$$

Для того, чтобы не вводить похожие формулы и во все последующие ячейки, их можно скопировать. Поскольку при копировании формул адреса ячеек в относительной форме (например, E3 или D15) автоматически изменяются на величину смещения по строкам и столбцам, а адреса ячеек в абсолютной форме (например, $D$15) остаются неизменными, используем это свойство. Введем в ячейку F3 знак =, щелкнем левой кнопкой мыши по ячейке E18 (это адрес свободного члена уравнения b_0) и нажмем на клавиатуре кнопку [F4] (для быстрой установки знаков $), введем с клавиатуры знак +, щелкнем по ячейке D18 и нажмем [F4] (это коэффициент b_1), введем знак * (умножить), щелкнем по ячейке B3 (это значение x_1), введем знак +, щелкнем по ячейке C18 и нажмем [F4] (это коэффициент b_2), введем знак * , щелкнем по ячейке C3 (значение x_2) и так далее. В результате в ячейке F3 получим формулу: =$E$18+$D$18*B3+$C$18*C3+$B$18*D3 .

Для ввода формулы нажать клавишу Enter или мышью кнопку [V] в строке формул. Для копирования выделенной формулы установить курсор мыши на *маркер заполнения* – небольшой черный квадрат в нижнем правом углу выделенной ячейки или диапазона. Попав на маркер заполнения, указатель примет вид черного креста. Чтобы скопировать содержимое выделенного диапазона в соседние ячейки, нажмите левую кнопку мыши и перемещайте мышь в нужном направлении (до ячейки F12).

Во всех ячейках диапазона F3:F12 появятся расчетные значения для зависимой величины ($y_{расч.}$). Если выделить ячейку F12 в строке формул можно увидеть формулу:

=$E$18+$D$18*B12+$C$18*C12+$B$18*D12 .

Как видно, при копировании относительные адреса изменились, а абсолютные остались неизменными.

Определим коэффициент детерминации с помощью функции КВПИРСОН. В разделе Статистические выбрать функцию КВПИРСОН и в качестве аргумента известные_значения_y ввести адреса фактических значений зависимой переменной (E3:E12), а для аргумента известные_значения_x ввести адреса расчетных значений зависимой переменной (F3:F12). В выделенной ячейке после нажатия кнопки [ОК] появится значение $\mathbf{R}^2$, совпадающее с тем, которое возвращает функция ЛИНЕЙН.

Для определения прогнозного значения функции для новых данных выделим ячейку с формулой (например, F12), подведем указатель мыши к границе рамки, нажмем клавишу Ctrl (рядом со стрелкой появится черный крестик) и не отпуская ее, при нажатой левой кнопке мыши перетащим копию формулы в ячейку F14. Отпустив клавиши, в ячейке F14 получим искомую прогнозную величину. Для завершения прогноза скопируйте формулу в ячейки F15 и F16.

Для получения параметров уравнения регрессии для однофакторных моделей, вызвать функцию ЛИНЕЙН и в строке известные_значения_x указать в первом случае адреса первого столбца данных (B3:B12), для другой модели вызвать снова функцию ЛИНЕЙН и выделить второй столбец (C3:C12), а для третьей модели – соответственно (D3:D12).

Порядок выполнения работы

1. Получить у преподавателя данные для расчета.

2. Ввести исходные данные в таблицу Excel.

3. Провести на ЭВМ серию расчетов по определению параметров множественной корреляционной зависимости и парных зависимостей для каждой из независимых переменных в отдельности.

4. Определить расчетные значения зависимой переменной, прогнозные значения и коэффициенты корреляции и детерминации.

5. Зафиксировать результаты расчетов в тетради.

6. Сделать выводы по результатам моделирования и записать в тетради.

Отчет по работе должен содержать

1. Название и цель работы.
2. Основные теоретические и методические положения.
3. Исходные данные для расчета.
4. Результаты расчета.
5. Выводы по результатам моделирования.

Лабораторная работа № 1а. Использование функций листа Excel в определении параметров моделей линейных и нелинейных зависимостей

Цель работы: изучение методики использования функций рабочего листа Excel ТЕНДЕНЦИЯ, ПРЕДСКАЗ, КОРРЕЛ, ПИРСОН, ЛГРФПРИБЛ, РОСТ для прогнозирования линейных и нелинейных процессов.

Исходные положения. В главном меню Excel вызвать справочную систему с помощью кнопки [Справка] из главного меню. Вывести на экран описание функций листа ТЕНДЕНЦИЯ, ПРЕДСКАЗ, КОРРЕЛ, ПИРСОН, ЛГРФПРИБЛ, РОСТ. Внимательно изучите и законспектируйте основные положения из описания этих функций: назначение функций, математические модели, синтаксис, порядок ввода, форму представления результатов и описание расшифровки получаемых результатов.

Порядок выполнения работы.

Получив у преподавателя исходные данные для расчета, введите их в поле электронной таблицы. С помощью функции ТЕНДЕНЦИЯ определить прогнозные значения зависимой переменной $\boldsymbol{y}$. С помощью функции ПРЕДСКАЗ определить прогнозные значения функции $\boldsymbol{y}$ отдельно для каждой из независимых переменных. С помощью функции КОРРЕЛ и ПИРСОН определить парные коэффициенты корреляции между всеми независимыми переменными, а также между независимыми и зависимой переменной. С помощью функции ЛГРФПРИБЛ определить параметры множественного уравнения регрессии для нелинейной зависимости. С помощью функции РОСТ определить прогнозные значения зависимой величины для нелинейной зависимости.

Модели, используемые в функциях:

ТЕНДЕНЦИЯ $y = b_0 + b_1 * x_1 + b_2 * x_2 + \ldots + b_m * x_m$

ПРЕДСКАЗ $y = b_0 + b_1 * x$

ЛГРФПРИБЛ и РОСТ $y = b_0 * b_1^{x_1} * b_2^{x_2} * \ldots * b_m^{x_m}$

КОРРЕЛ

$$r = \frac{n(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

ПИРСОН

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}; \qquad \sigma_X^2 = \frac{1}{n}\sum(x_j - \mu_X)^2$$

$$Cov(X,Y) = \frac{1}{n}\sum_{j=1}^{n}(x_j - \mu_x)(y_j - \mu_y); \qquad \sigma_Y^2 = \frac{1}{n}\sum(y_j - \mu_y)^2$$

Пример выполнения работы.

Используя исходные данные из предыдущей работы, найдем расчетные значения прогнозируемой величины прибыли. Для этого выделим массив ячеек, соответствующий неизвестным значениям величины прибыли для Продуктов 11, 12 и 13 – это ячейки E14:E16. Вызвать функцию ТЕНДЕНЦИЯ из категории *Статистические*. В появившемся диалоговом окне в соответствующие текстовые поля ввести аргументы функции: *известные_значения_у* – адреса E3:E12; *известные_значения_х* – адреса B3:D12; *новые_значения_х* – адреса B14:D16; *константа* – ввести число 1 (или слово ИСТИНА). Для ввода функции (как формулы массива) щелкните по формуле в строке формул и нажмите сочетание клавишей Ctrl+Shift+Enter. В выделенном массиве ячеек появятся прогнозируемые значения зависимой переменной. Для определения расчетных значений зависимой переменной, используем функцию ТЕНДЕНЦИЯ. Для этого выделим ряд свободных ячеек, аналогичный по величине ряду фактических значений зависимой величины (например, ячейки F3:F12). Вызвать функцию ТЕНДЕНЦИЯ, в текстовое поле *известные_значения_у* ввести адреса E3:E12; в поле *известные_значения_х* – адреса B3:D12, остальные поля можно оставить незаполненными. В результате ввода функции как формулы массива, в выделенных ячейках получим искомые расчетные значения функции.

Найдем расчетные значения прогнозируемой величины прибыли в отдельности от каждой из независимых величин. Для этого в свободную ячейку введем функцию ПРЕДСКАЗ из категории *Статистические*. В появившемся диалоговом окне в текстовые поля ввести аргументы функции: *значение независимой величины х*, для которого предсказывается величина зависимой переменной – ввести адрес ячейки B14; *известные_значения_у* – адреса E3:E12; *известные_значения_х* – адреса B3:B12. После ввода функции в выделенной ячейке появится значение прогнозируемой величины.

Аналогичные действия повторяются и для определения прогноза при рассмотрении зависимости функции от других независимых переменных, расположенных в столбцах C и D. Эти действия можно упростить, если использовать операцию копирования функции. Для этого при вводе функции ПРЕДСКАЗ (например, в ячейку G14), необходимо для значений зависимой величины указать абсолютные (неизменяемые при копировании) адреса, то есть при вводе адресов *известные_значения_у*, следует нажать клавишу F4. В результате адреса примут вид $E$3:$E$12. Скопируем функцию вправо на две ячейки, потянув за *маркер заполнения*. В ячейках H14 и I 14 окажутся прогнозные значения от величины второго и третьего факторов.

Найдем расчетные значения парных коэффициентов корреляции. В свободную ячейку введем функцию КОРРЕЛ из категории *Статистические*. В появившемся диалоговом окне в текстовые поля введем в качестве *массива1* – адреса значений переменной $\boldsymbol{x}_1$ (B3:B12), в качестве *массива2* – адреса значений переменной $\boldsymbol{x}_2$ (C3:C12). После ввода функции в выделенной ячейке появится значение коэффициента корреляции $\boldsymbol{r}_{12}$. Аналогичные действия повторить для определения коэффициентов корреляции для всех сочетаний независимых и зависимой переменной и представить их в виде корреляционной матрицы:

	x_1	x_2	x_3	y
x_1	r_{11}	r_{12}	r_{13}	r_{1y}
x_2			r_{23}	r_{2y}
x_3			r_{33}	r_{3y}
y				

Оставшиеся свободными ячейки заполнить значениями парных коэффициентов корреляции, рассчитанными с помощью функции ПИРСОН из категории *Статистические*. В соответствующую ячейку введите функцию и в качестве аргументов укажите адреса двух рядов данных, например для расчета коэффициента $\boldsymbol{r}_{21}$ в качестве *массива1* ввести адреса переменной $\boldsymbol{x}_2$ (С3:С12), а в качестве *массива2* – адреса переменной $\boldsymbol{x}_1$ (В3:В12).

Найдем расчетные значения параметров нелинейного уравнения регрессии и коэффициент детерминации. Выделим массив ячеек в пять строк и с количеством столбцов, соответствующим количеству искомых параметров уравнения регрессии (в нашем примере – четыре столбца). Вызвать функцию ЛГРФПРИБЛ из категории *Статистические*. В появившемся диалоговом окне в соответствующие текстовые поля ввести аргументы функции: *известные_значения_у* – адреса Е3:Е12, *известные_знечения_х* – адреса В3:D12, *константа* – ввести число 1 (или слово ИСТИНА), *статистика* – ввести число 1 (или слово ИСТИНА). Для ввода функции (как формулы массива) щелкните по формуле в строке формул и нажмите сочетание клавишей Ctrl+Shift+Enter. В выделенном массиве ячеек появятся параметры нелинейного уравнения регрессии и коэффициент детерминации. Расположение результатов расчета аналогичное функции ЛИНЕЙН.

Найдем расчетные значения прогнозируемой величины прибыли для нелинейной зависимости. Выделить массив ячеек, соответствующий неизвестным значениям прибыли для продуктов 11, 12 и 13 – это ячейки H14:H16. Вызвать функцию РОСТ из категории *Статистические*. В появившемся диалоговом окне в соответствующие тестовые поля ввести аргументы функции: *известные_значения_y* – адреса E3:E12, *известные_значения_x* – адреса B3:D12, *новые_значения_x* – адреса B14:D16, *константа* – можно ввести число 1 (или слово ИСТИНА). Для ввода функции (как формулы массива) щелкните по формуле в строке формул и нажмите сочетание клавишей Ctrl+Shift+Enter. В выделенном массиве ячеек появятся прогнозируемые значения зависимой величины – прибыли.

Порядок выполнения работы

1. Получить у преподавателя данные для расчета.
2. Ввести исходные данные в таблицу Excel.
3. Провести на ЭВМ серию расчетов по определению прогнозных значений зависимой переменной с использованием модели множественной корреляционной линейной и нелинейной зависимости, определению парных коэффициентов корреляции.
4. Определить параметры нелинейного множественного уравнения регрессии, расчетные значения зависимой переменной, прогнозные значения и коэффициенты корреляции и детерминации.
5. Зафиксировать результаты расчетов в тетради.
6. Сделать выводы по результатам моделирования и записать в тетради.

Отчет по работе должен содержать

1. Название и цель работы.
2. Основные теоретические и методические положения.
3. Исходные данные для расчета.
4. Результаты расчета.
5. Выводы по результатам моделирования.

Лабораторная работа № 2. Применение инструмента электронной таблицы Excel *Анализ данных* при определении параметров линейных моделей

Цель работы: изучение методики использования программы *Пакет анализа* при определении параметров уравнения регрессии, парных и множественных коэффициентов корреляции и коэффициента детерминации.

Исходные положения. Решение задач аппроксимации с помощью *Пакета анализа* становится возможным после загрузки соответствующего приложения, которое представлено в Excel как дополнительная программа (надстройка) – *Анализ данных*. Для загрузки этой надстройки, ее имя следует отметить в диалоговом окне *Надстройки*, которое открывается с помощью команды *Надстройки* из меню *Сервис*. В Excel 2010 сначала выбрать закладку *Меню*, в которой находится меню *Сервис*. После установки в списке надстроек опции программы *Пакет анализа*, соответствующая команда *Анализ данных* будет доступна при открытии меню *Сервис* (в Excel 2010 команда *Анализ данных* находится в меню *Данные*).

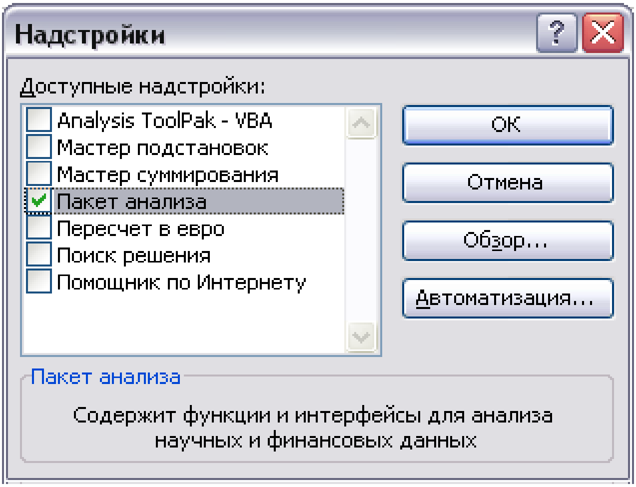


Рис. 3. Диалоговое окно *Надстройки* при установке *Пакета анализа*

Теперь в меню *Сервис* при вызове команды *Анализ данных* можно выбрать из списка инструменты *Регрессия* и *Корреляция*, которые возвращают информацию о параметрах уравнения регрессии, коэффициенте детерминации и коэффициентах парной корреляции.

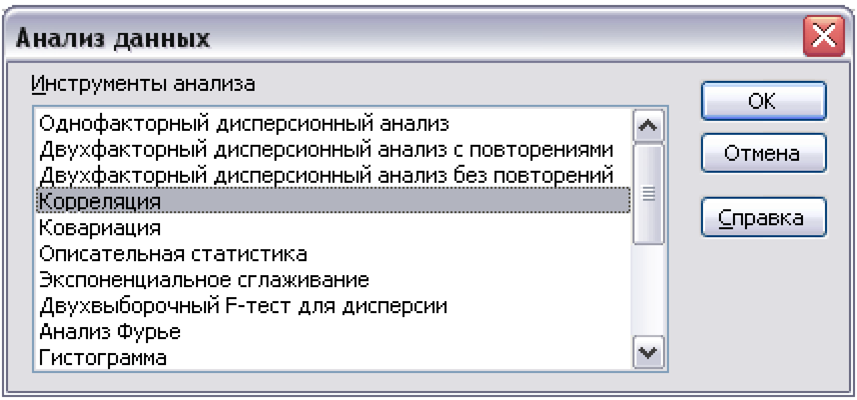


Рис. 4. Выбор инструмента анализа

Порядок выполнения работы. Получив у преподавателя данные для расчета, введите их в поле электронной таблицы. С помощью программы *Анализ данных* определите параметры уравнения регрессии и коэффициент детерминации. Для этого в списке инструментов выбрать строку *Регрессия*. В появившемся диалоговом окне в качестве входного массива укажите адреса исходных данных – зависимой и независимых переменных. Результаты расчетов представить на новом листе.

При определении параметров уравнения регрессии используется линейная модель

$$\boldsymbol{y = b_0 + b_1 * x_1 + b_2 * x_2 + \ldots + b_m * x_m} \,.$$

С помощью программы *Пакет анализа* определите коэффициенты парной корреляции. Для этого вызовите из меню *Сервис* программу *Анализ данных*. В появившемся меню выберите инструмент *Корреляция*. В открывшемся диалоговом окне в качестве входного массива указать адреса исходных данных (включая зависимую и независимые переменные). Результаты расчетов представить на новом листе.

С помощью полученных параметров уравнения регрессии определить расчетные значения функции и прогнозные значения.

С помощью программы *Анализ данных* определить параметры однофакторных уравнений регрессии для каждой из независимых переменных в отдельности: $\boldsymbol{y=b_0+b_1*x_1}$; $\boldsymbol{y=b_0+b_2*x_2}$; … $\boldsymbol{y=b_0+b_m*x_m}$.

По результатам расчетов сделать выводы.

Пример выполнения работы.

Используем исходные данные предыдущих работ. Из меню *Сервис* вызовем программу *Анализ данных*. В появившемся окне выберем инструмент Регрессия. В открывшемся диалоговом окне в соответствующих текстовых полях зададим ссылки на ячейки - адреса значений результативного признака - зависимой величины $\boldsymbol{y}$ – *Входной интервал Y*: E3**:**E12 и адреса значений независимых величин - факторов $\boldsymbol{x_1}, \ldots, \boldsymbol{x_i}, \ldots, \boldsymbol{x_m}$ ($\boldsymbol{m} \leq 16$) – *Входной интервал X:* B3**:**D12. *Уровень надежности* (доверительная вероятность) по умолчанию предполагается равным 95%. Если исследователя это значение не устраивает, то рядом со словами *Уровень надежности* нужно поставить «галочку» и указать требуемое значение. Поставив «галочку» рядом со словом Константа-ноль, исследователь получит $\boldsymbol{b_0}$=0 по умолчанию. Если нужны значения остатков $\mathbf{e}_i$ и их график, то нужно поставить «галочки» рядом со словами ***Остатки*** и ***График остатков***. ОК.

Вывод результатов произведем на новый рабочий лист.

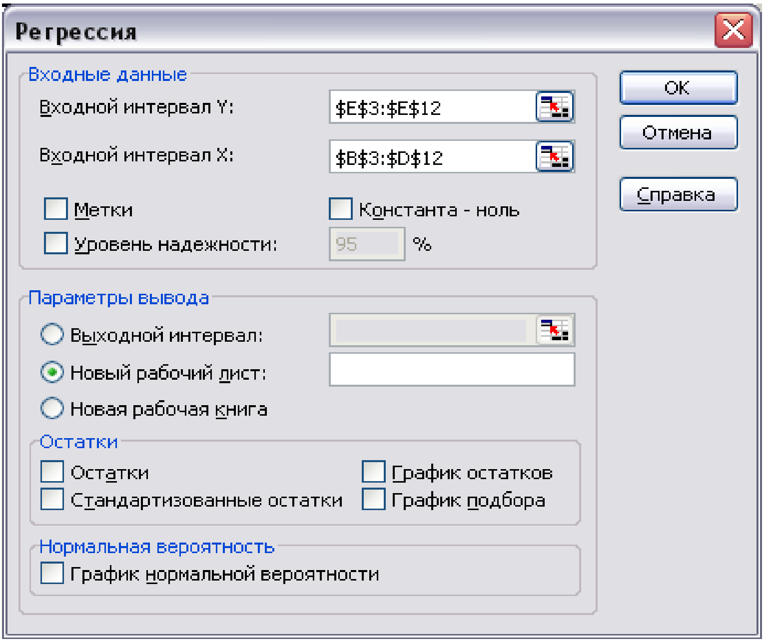


Рис. 5. Диалоговое окно инструмента *Регрессия*

	A	B	C	D	E	F	G
1	ВЫВОД ИТОГОВ						
2							
3	*Регрессионная статистика*						
4	Множественный R	0.990950856					
5	R-квадрат	0.981983598					
6	Нормированный R-кв	0.972975397					
7	Стандартная ошибка	749.4480065					
8	Наблюдения	10					
9							
10	Дисперсионный анализ						
11		*df*	*SS*	*MS*	*F*	*Значимость F*	
12	Регрессия	3	183683625.3	61227875.09	109.0099574	1.27057E-05	
13	Остаток	6	3370033.887	561672.3144			
14	Итого	9	187053659.2				
15							
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
17	Y-пересечение	-305.10886	319.9863232	-0.95350594	0.377151605	-1088.087759	477.8700392
18	Переменная X 1	0.71707091	0.074878267	9.576489163	7.40903E-05	0.533850257	0.900291562
19	Переменная X 2	0.094195122	0.080959208	1.163488682	0.28880417	-0.103905068	0.292295311
20	Переменная X 3	-1.557350187	0.719314026	-2.165049104	0.073562625	-3.317449489	0.202749115

Рис. 6. Результаты расчета с помощью инструмента Регрессия

Результаты расчета: коэффициент множественной корреляции $\boldsymbol{R}$ = 0,99; коэффициент множественной детерминации $\boldsymbol{R}^2$=0,98 ; коэффициенты уравнения регрессии $\boldsymbol{b}_0$= -305,1 , $\boldsymbol{b}_1$= 0,717 , $\boldsymbol{b}_2$= 0,09 , $\boldsymbol{b}_3$= -1,557 .

Расчётную величину F-статистики (адрес E12) $F_{расч}$ = 109 необходимо сравнить с критическим значением $F_{кр}$, получаемым с помощью функции листа из категории *Статистические* FРАСПОБР() (в Excel 2010 =F.ОБР.ПХ(*вероятность*; *степени свободы* 1; *степени свободы* 2)). Вероятность $\boldsymbol{\alpha}$ = 0,05 – уровень значимости (обычно принимается равным 5% и вводится в соответствующее окно с клавиатуры); *степени свободы* 1 – $\boldsymbol{df}_1 = \boldsymbol{m}$ = 3 – регрессионное число степеней свободы, которое располагается в данной задаче по адресу B12 на листе с результатами расчёта; *степени свободы* 2 – $\boldsymbol{df}_2 = \boldsymbol{n} - \boldsymbol{m}$ – 1 = 6 – остаточное число степеней свободы, которое располагается в данной задаче по адресу B13 на листе с результатами расчёта ($\boldsymbol{n}$ = 10 – число наблюдений – адрес B8, $\boldsymbol{m}$ = 3 – число независимых переменных – $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, $\boldsymbol{x}_3$). Если $F_{расч} > F_{кр}$, то нулевая гипотеза о незначимости уравнения регрессии в целом отвергается и модель по критерию Фишера считается пригодной для анализа и прогнозирования. Величина *Значимость F* (адрес F12) означает вероятность ошибки $\boldsymbol{p}$ = 1,27E-05 = 0,0000127. Если число в графе *Значимость F* превышает 5% - уровень надёжности, то принимается гипотеза $\boldsymbol{R^2=0}$. Иначе принимается гипотеза $\boldsymbol{R^2 \neq 0}$.

Для проверки статистической значимости коэффициентов уравнения регрессии (B17:B20) применяется критерий Стьюдента $\boldsymbol{t}$-статистика. Расчётные значения $\boldsymbol{t}_{расч}$ (D17:D20) необходимо сравнить с критическим значением $\boldsymbol{t}_{кр}$, получаемым с помощью функции листа из категории статистические =СТЬЮДРАСПОБР(*вероятность*; *степени свободы; хвосты*) (в Excel 2010 =СТЬЮДЕНТ.ОБР.2Х(*вероятность*; *степени свободы*)). Вероятность $\boldsymbol{\alpha}$ = 0,05 – уровень значимости вводится в соответствующее окно с клавиатуры; *степени свободы* – $\boldsymbol{df} = \boldsymbol{n} - \boldsymbol{m} - 1$ = 6 – остаточное число степеней свободы

(B13), *хвосты* = 2. Если $|t_{расч}| > t_{кр}$, то нулевая гипотеза о незначимости соответствующего коэффициента уравнения регрессии отвергается и данный параметр считается статистически значимым. В противном случае прогноз будет недостоверным. ***P***-значение - это величины уровней значимости, соответствующие вычисленным ***t***-статистикам. Если ***P***-значение превышает 5% уровень надежности, то соответствующая переменная статистически незначима и её можно исключить из модели.

Нижние 95% и Верхние 95% - это нижние и верхние границы 95-процентных доверительных интервалов для коэффициентов теоретического уравнения линейной регрессии. Если исследователь согласился с принятым по умолчанию значением доверительной вероятности - *уровень надёжности* 95%, то последние два столбца будут дублировать два предыдущих столбца. Если исследователь вводил свое значение доверительной вероятности, например, 99%, то последние два столбца будут содержать иные значения доверительных интервалов.

Расширить колонки можно установив указатель мыши на границе между заголовками столбцов на сером фоне, указатель примет форму креста со стрелками, тогда нужно дважды щелкнуть левой кнопкой мыши или при нажатой левой кнопке потянуть границу столбца в нужную сторону.

Для определения расчетных значений зависимой величины $\boldsymbol{y}_{расч}$ в ячейку (F3), расположенную рядом с первым значением $\boldsymbol{y}_{фактич}$ введем формулу:

$$\boldsymbol{y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_3}.$$

Скопировать формулу для определения всех расчетных значений ***y***.

Параметры уравнения регрессии отдельно для каждой из независимых переменных (однофакторные модели) рассчитываются аналогично, как и для многофакторной модели, но в текстовом поле инструмента *Регрессия* указываются адреса только одного ряда данных, соответствующего переменной $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ и затем $\boldsymbol{x}_3$.

Найдем парные коэффициенты корреляции с помощью программы *Пакет анализа*. Для этого в меню *Сервис* выберите строку *Анализ данных*. В появившемся окне выберите инструмент *Корреляция*. В текстовом поле *Входной интервал* укажите адреса всех известных значений исходных данных (включая независимые и зависимую переменную).

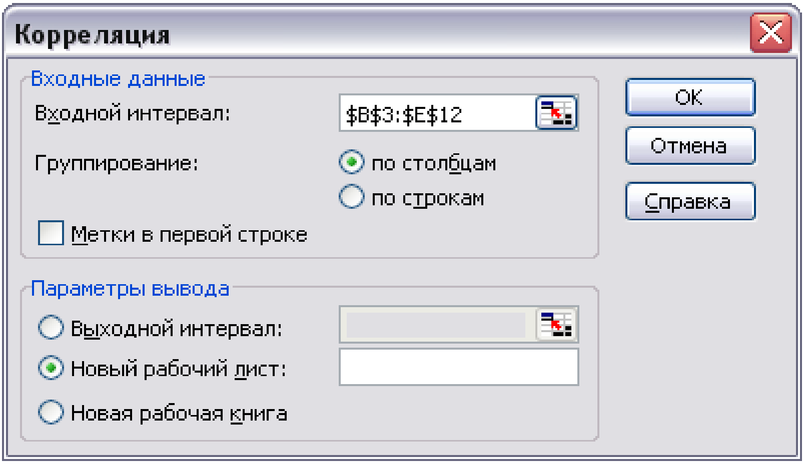


Рис. 7. Диалоговое окно инструмента Корреляция

На новом листе получим матрицу парных коэффициентов корреляции (табл. 1).

Таблица 1

Коэффициенты корреляции

	Столбец 1	*Столбец 2*	*Столбец 3*	*Столбец 4*
Столбец 1	1			
Столбец 2	0.54709691	1		
Столбец 3	0.898639153	0.610267956	1	
Столбец 4	0.983456355	0.560539112	0.83832548	1

Порядок выполнения работы

1. Получить у преподавателя данные для расчета

2. Ввести исходные данные в таблицу Excel.

3. Провести на ЭВМ серию расчетов по определению прогнозных значений зависимой переменной с использованием модели множественной корреляционной линейной зависимости, определению парных коэффициентов корреляции.

4. Определить параметры множественного уравнения регрессии, расчетные значения зависимой переменной, прогнозные значения и коэффициенты корреляции и детерминации.

5. Зафиксировать результаты расчетов в тетради.

6. Сделать выводы по результатам моделирования и записать в тетради.

Отчет по работе должен содержать

1. Название и цель работы.
2. Основные теоретические и методические положения.
3. Исходные данные для расчета.
4. Результаты расчета.
5. Выводы по результатам моделирования.

Лабораторная работа № 3. Графическое моделирование линейных и нелинейных зависимостей

Цель работы: изучение использования возможностей графических моделей для определения формы связи, параметров уравнения регрессии, аппроксимации с помощью линейных и нелинейных функций и использования экстраполяции найденного тренда для прогнозирования.

Исходные положения. Методика позволяет производить выбор графических моделей для парных регрессионных зависимостей. Парная регрессия показывает зависимость в виде $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$. Excel позволяет построить график зависимости и находить уравнение парной регрессии для построения графика зависимости. Для этого ввести исходные данные в рабочий лист. Вызвать инструмент Мастер диаграмм (или из меню *Вставка* команду *Диаграмма*). 

В появившемся диалоговом окне выбрать тип диаграммы – Точечная). Чтобы перейти к следующему окну, нажать кнопку Далее>.

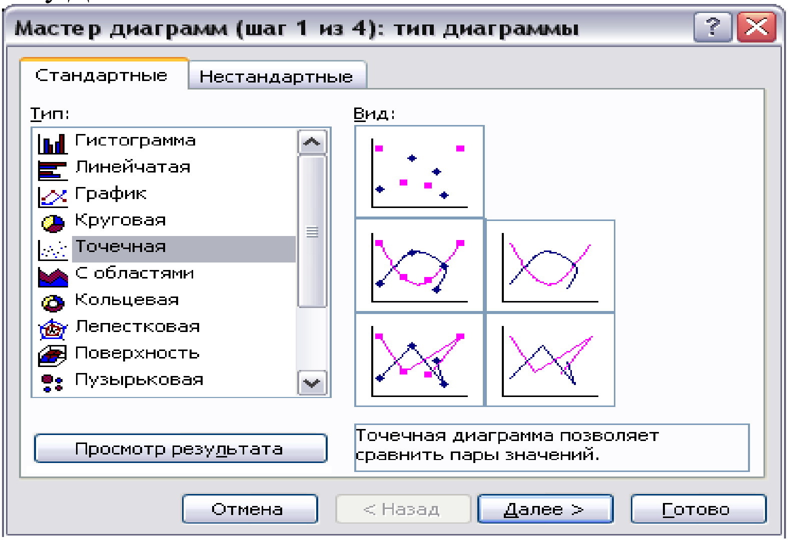


Рис. 8. Диалоговое окно *Мастера диаграмм* (шаг 1 – выбор типа диаграммы)

В текстовое поле Диапазон диалогового окна *Мастер диаграмм* (шаг 2) ввести адреса массива исходных данных (B3:B12;E3:E12). Для задания диапазона несмежных данных необходимо ввести его с клавиатуры или с помощью мыши. Для этого выделить мышью один ряд данных и при нажатой клавише Ctrl выделить другой ряд данных. С помощью следующего диалогового окна *Мастер диаграмм* (шаг 3) можно ввести название диаграммы и названия осей графика.

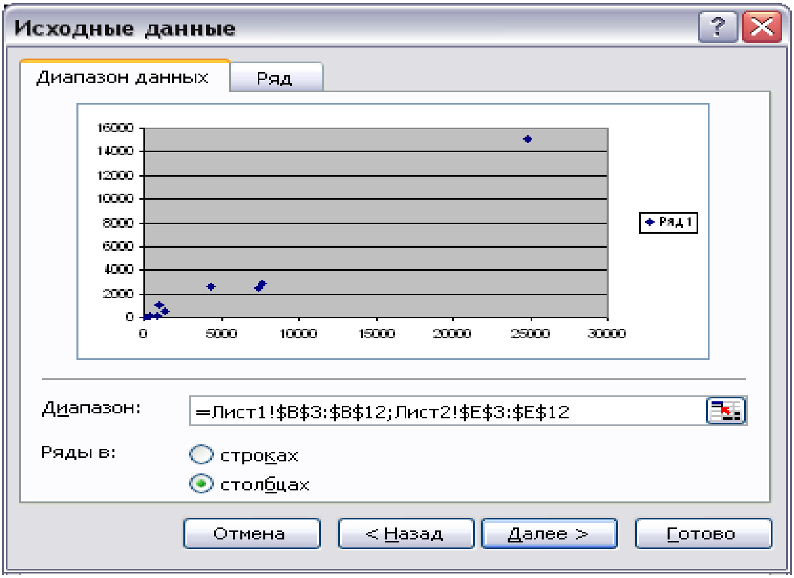


Рис. 9. Диалоговое окно *Мастера диаграмм* (шаг 2 – задать диапазон)

После нажатия кнопки *Готово* на экране появится график. Поместите указатель мыши в область диаграммы (рядом с внешней границей), нажмите левую кнопку мыши и переместите график на свободное место в рабочем листе. Щелкнув левой кнопкой мыши по элементу графика, например, по легенде, выделим этот элемент. Выделение сопровождается появлением на элементе графика меленьких черных квадратиков. Если установить указатель мыши на один из этих квадратиков, указатель примет вид разнонаправленной стрелки и при нажатой левой кнопке, перемещение мыши позволяет изменить размеры, положение выделенного элемента. Если щелкнуть по выделенному элементу правой кнопкой мыши, появится контекстное меню. Выбрав в меню команду *Удалить*, можно убрать из рабочего листа выделенный элемент диаграммы.

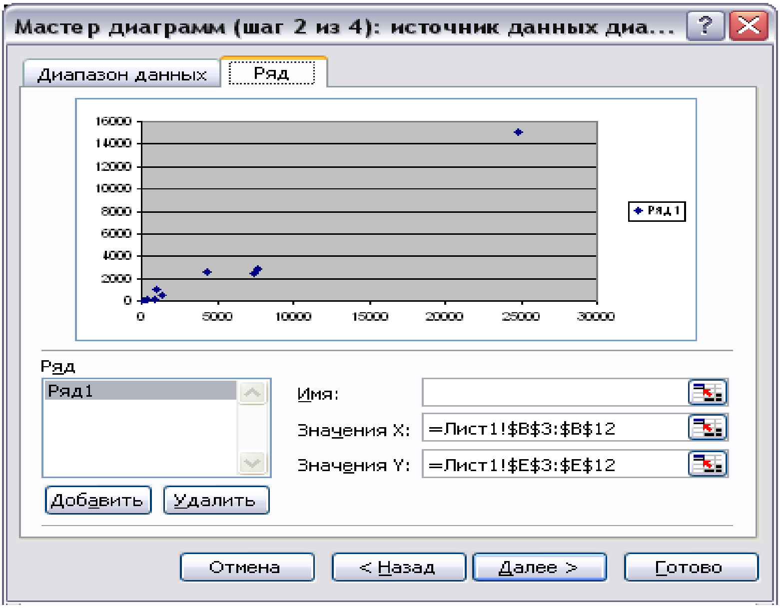


Рис. 10. Диалоговое окно *Мастера диаграмм* (шаг 2 – исходные данные)

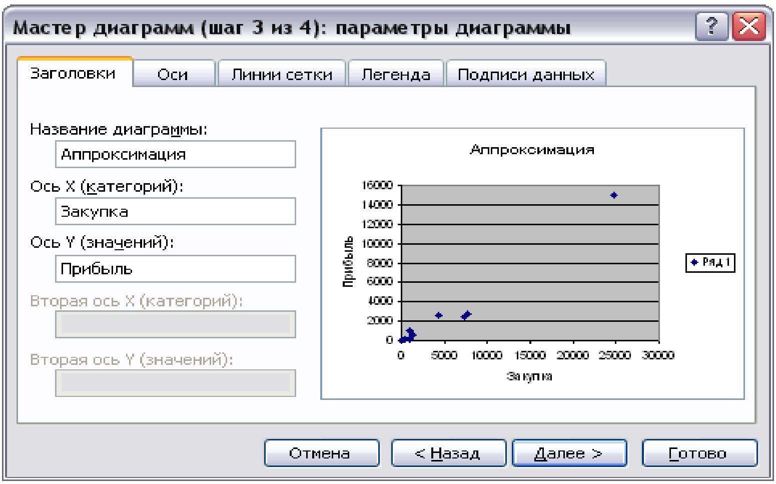


Рис. 11. Диалоговое окно *Мастера диаграмм* (шаг 3 – названия осей)

Если график выделен метками, то в главном меню появляется раздел *Диаграмма*. Вызвать меню *Диаграмма* и выбрать команду *Добавить линию тренда*. На экране появится диалоговое окно Линия тренда. В окне с ярлычком Тип выбрать форму линии для аппроксимации.

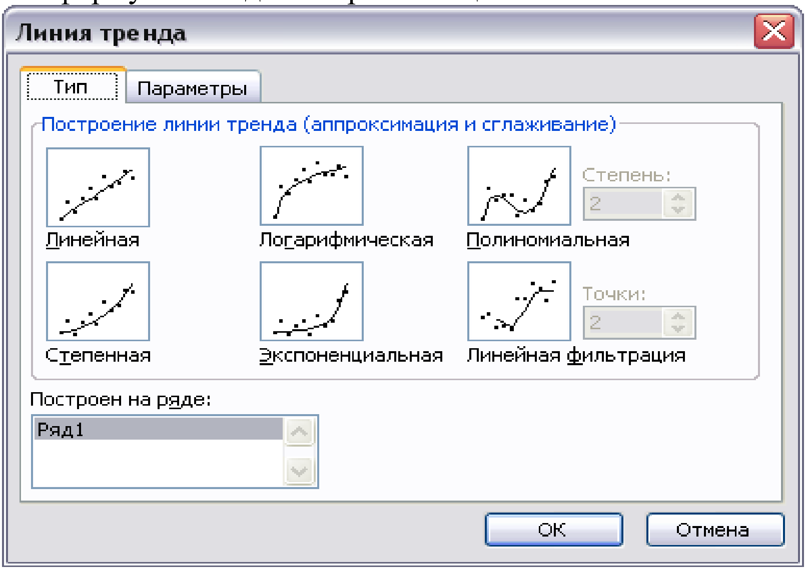


Рис. 12. Диалоговое окно команды *Добавить линию тренда*

Таблица 2

Модели регрессии

Характеристики различного типа регрессий	
Тип парной регрессии	Уравнение
Линейная	$y=b_0+b_1x$
Логарифмическая	$y=b_0+b_1 * ln(x)$
Полиномиальная	$y=b_0+b_1x+b_2x^2+\ldots+b_mx^m$
Степенная	$y=b_1x^{b_0}$
Экспоненциальная	$y=b_0e^{b_1x}$

Выбрать тип регрессии – *Линейная*. Щелкнуть левой кнопкой мыши ярлычок *Параметры*. Отметить в соответствующих окошках: *Показывать уравнение на диаграмме* и *Поместить на диаграмму* $\boldsymbol{R}^2$. Нажать кнопку [ОК]. Если тенденция имеет возрастающий или убывающий характер, то возможно на основе экстраполяции данных достроить линию тренда вперед на несколько периодов для осуществления прогноза.

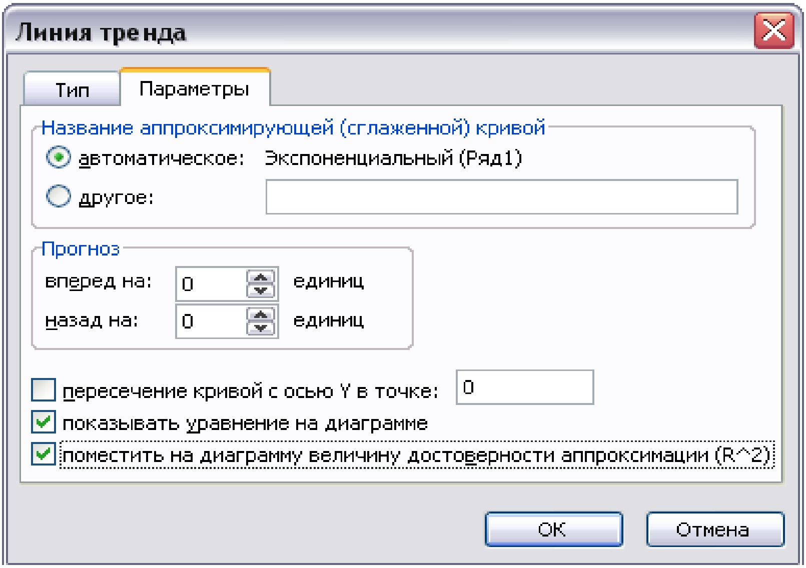


Рис. 13. Диалоговое окно для показа *Параметров* тренда на диаграмме

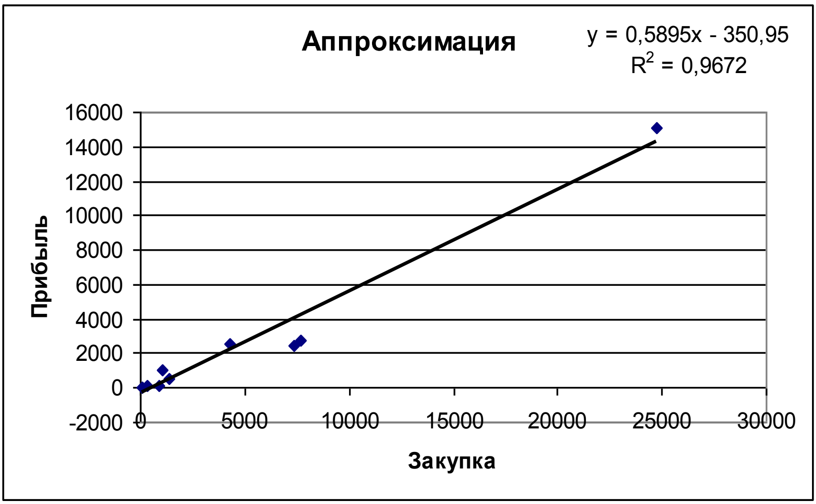
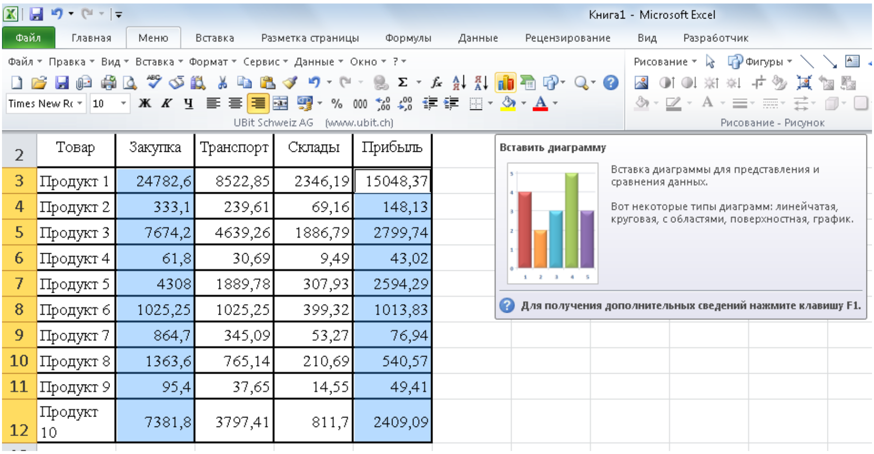


Рис. 14. Результат добавления линейного тренда к диаграмме

На графике появляется сглаженная кривая – линия тренда, уравнение парной регрессии и величина $\boldsymbol{R}^2$. Уравнение и коэффициент детерминации следует для наглядности выделить и переместить на свободное место. Для этого щелкнем по уравнению левой кнопкой мыши, появится рамка выделения. Подвести указатель мыши к границе рамки и при нажатой левой кнопке мыши протянуть рамку в нужную сторону.

Повторите все построения для всех сочетаний независимых и зависимой переменных и для всех форм зависимостей. Например, для зависимости $\boldsymbol{y}$ от $\boldsymbol{x}_1$ построить линейную, степенную, логарифмическую, экспоненциальную и полиномиальную зависимости, тоже для зависимостей $\boldsymbol{y}$ от $\boldsymbol{x}_2$ и $\boldsymbol{y}$ от $\boldsymbol{x}_3$). Результаты расчетов и построений отобразить в тетради.

В Excel 2010 выделяем исходные данные – если диапазон включает несмежные столбцы, то действие производится при нажатой клавише Ctrl. В закладке *Меню* выбираем пиктограмму *Вставить диаграмму* (рис. 15).



2	Товар	Закупка	Транспорт	Склады	Прибыль
3	Продукт 1	24782,6	8522,85	2346,19	15048,37
4	Продукт 2	333,1	239,61	69,16	148,13
5	Продукт 3	7674,2	4639,26	1886,79	2799,74
6	Продукт 4	61,8	30,69	9,49	43,02
7	Продукт 5	4308	1889,78	307,93	2594,29
8	Продукт 6	1025,25	1025,25	399,32	1013,83
9	Продукт 7	864,7	345,09	53,27	76,94
10	Продукт 8	1363,6	765,14	210,69	540,57
11	Продукт 9	95,4	37,65	14,55	49,41
12	Продукт 10	7381,8	3797,41	811,7	2409,09

Рис. 15. Вставка диаграммы в рабочий лист Excel

В появившемся диалоговом окне Вставка диаграммы выбираем *Точечную* с несвязанными линиями точками (рис. 16).

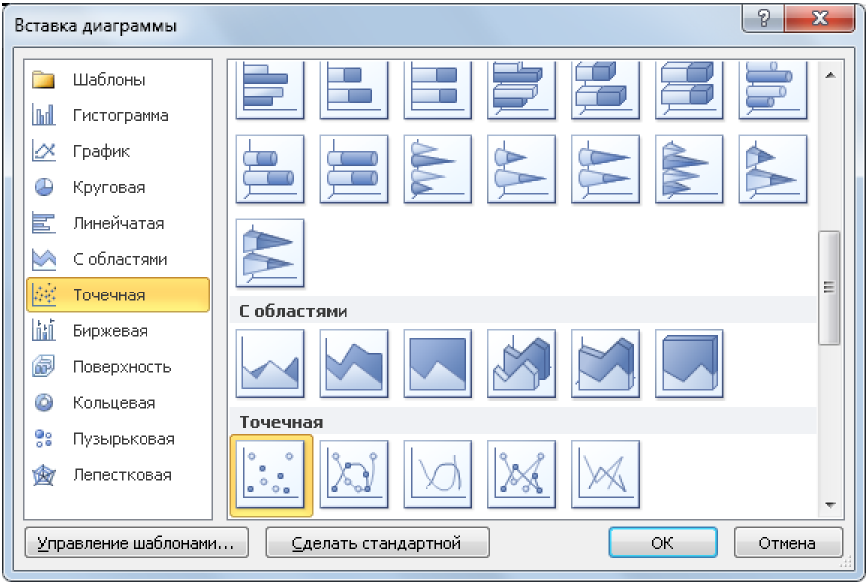


Рис. 16. Выбор вида диаграммы - *Точечной* диаграммы

На полученной диаграмме вызываем контекстное меню нажатием правой кнопки мыши на любой из полученных точек. Выбираем в меню опцию *Добавить линию тренда* (рис. 17).

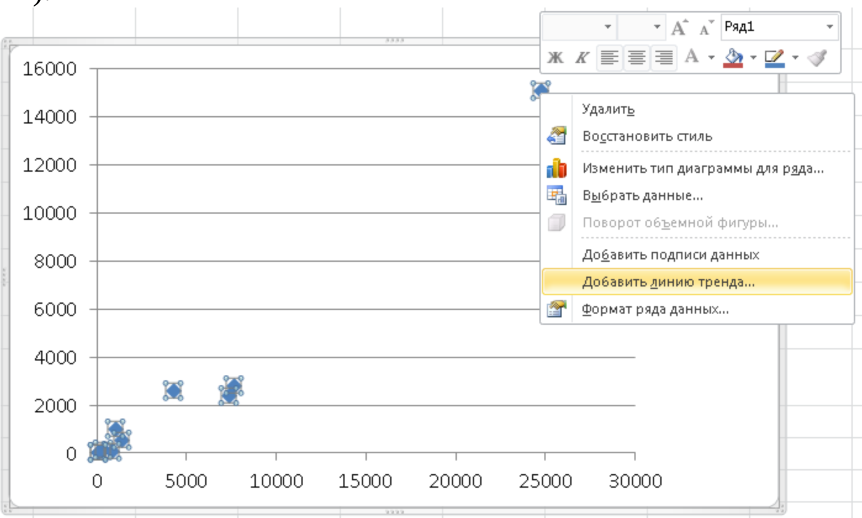


Рис. 17. Добавление линии тренда на диаграмме

В появившемся диалоговом окне *Формат линии тренда* выбираем вид зависимости – *Линейная* и устанавливаем галочки в квадратных окошках строк *Показывать уравнение на диаграмме* и *Поместить на диаграмму величину достоверности аппроксимации* R^2 – коэффициент детерминации $\boldsymbol{R}^2$ (рис. 18).

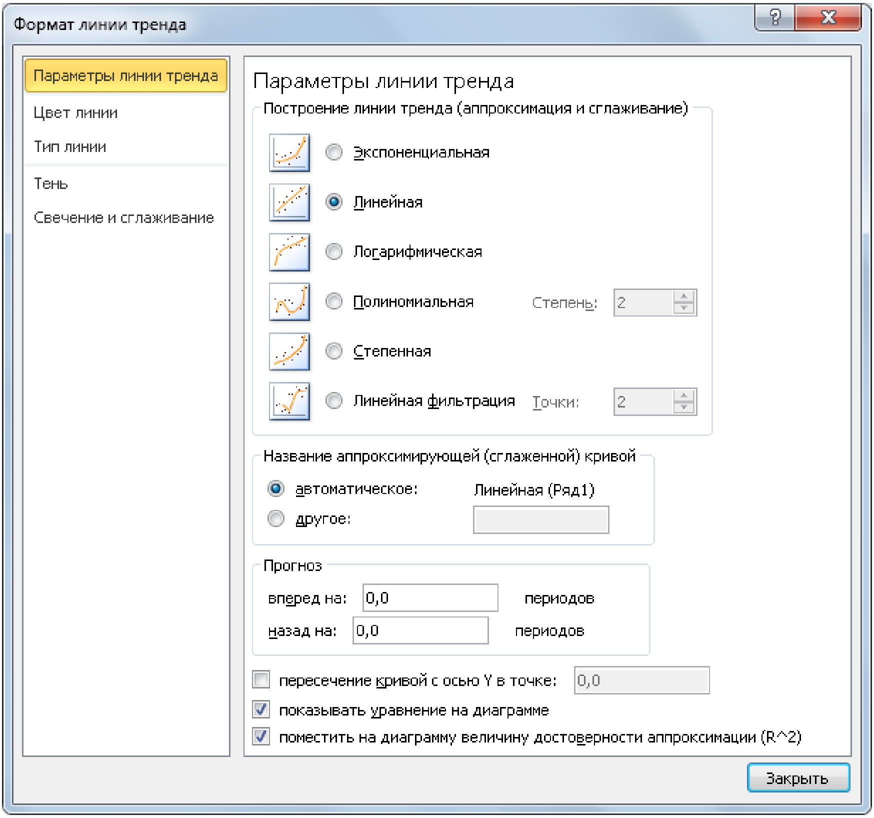


Рис. 18. Выбор вида зависимости и её параметров

Результат построения диаграммы представлен на рис. 19.

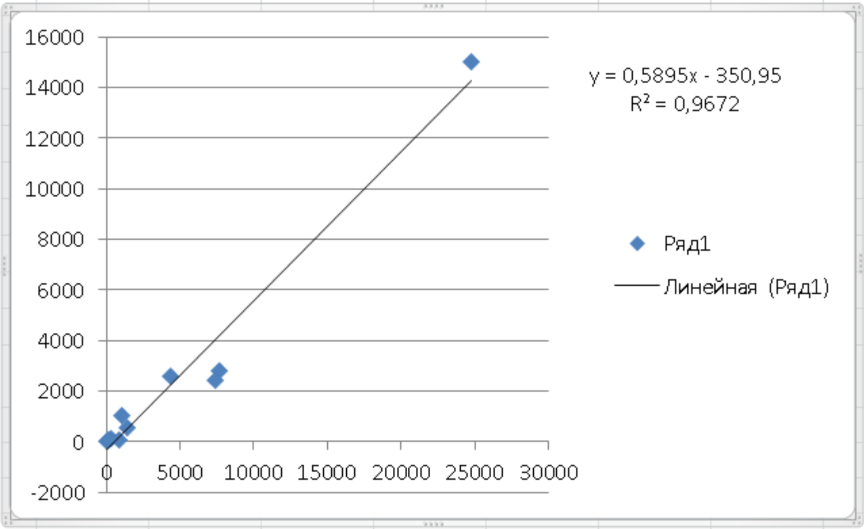


Рис. 19. Линейная регрессия и её параметры

Порядок выполнения работы

1. Получить у преподавателя данные для расчета.
2. Ввести исходные данные в таблицу Excel.
3. Провести на ЭВМ серию экспериментов с различными типами моделей регрессии с определением прогнозных значений зависимой переменной, параметров регрессионных моделей и коэффициентов детерминации.
4. Зафиксировать результаты расчетов в тетради.
5. Сделать выводы по результатам моделирования и записать в тетради.

Отчет по работе должен содержать

1. Название и цель работы.
2. Основные теоретические и методические положения.
3. Исходные данные для расчета.
4. Результаты расчета.
5. Выводы по результатам моделирования.

Лабораторная работа № 4.
Метод наименьших квадратов

Цель работы: изучение использования возможностей инструмента Поиск решений для определения параметров уравнения регрессии методом наименьших квадратов, закрепление навыков аппроксимации с помощью линейных и нелинейных функций и использования экстраполяции найденного тренда для прогнозирования.

Исходные положения. Рассмотрим задачу построения регрессионной модели. С помощью средства поиска решений решим задачу нахождения уравнения регрессии для одной зависимой и одной независимой переменных. Данный подход позволяет исследовать любое уравнение регрессии. Используем функции рабочего листа, непосредственно вычисляющие различные характеристики линейного уравнения регрессии и экспоненциального уравнения регрессии, которые позволяют значительно упростить процедуру регрессионного анализа для наиболее часто встречающихся на практике моделей.

1. Общий подход к построению уравнения регрессии на примере линейной модели

Рассмотрим, как решается задача нелинейной оптимизации с помощью средства поиска решений на примере построения линейного уравнения регрессии. Имеются две наблюдаемые величины ***x*** и ***y***, например, объем реализации фирмы, торгующей подержанными автомобилями, за шесть недель ее работы. Значения этих наблюдаемых величин приведены на рис. 20, где ***x*** - отчетная неделя, а ***y*** - объем реализации за эту неделю.

	A	B	C	D	E	F
1	x	y	Теор.знач y			
2	1	7	7,29	b	m	Z
3	2	9	9,17	5,400003	1,886	1,771
4	3	12	11,06			
5	4	13	12,94			
6	5	14	14,83			
7	6	17	16,71			

Рис. 20. Исходные данные (x, y) для построения линейной модели

Необходимо построить линейную модель $y = mx + b$, наилучшим образом описывающую наблюдаемые значения. Обычно m и b подбираются так, что бы минимизировать сумму квадратов разностей между наблюдаемыми и теоретическими значениями зависимой переменной y, т.е. минимизировать

$$Z = \sum_{i=1}^{n} (y_i - (b + m x_i))^2 \to \min,$$

где n - число наблюдений (в данном случае n = 6).

Для решения этой задачи отведем под переменные b и m ячейки D3 и E3, соответственно, а в ячейку F3 введем минимизируемую функцию

{ = СУММКВРАЗН (B2 : B7**;** D3 + E3*****A2 **:** A7) }

Функция СУММКВРАЗН (SUMSQ) вычисляет сумму квадратов разностей для элементов указанных массивов.

Теперь выберем команду *Сервис, Поиск решения* (Too1s, So1ver) и заполним открывшееся диалоговое окно *Поиск решения* (So1ver) как показано на рис. 21.

Отметим, что на переменные b и m ограничения не налагаются. В результате вычислений средство Поиска решений найдет: b = 5,400003 и m = 1,886.

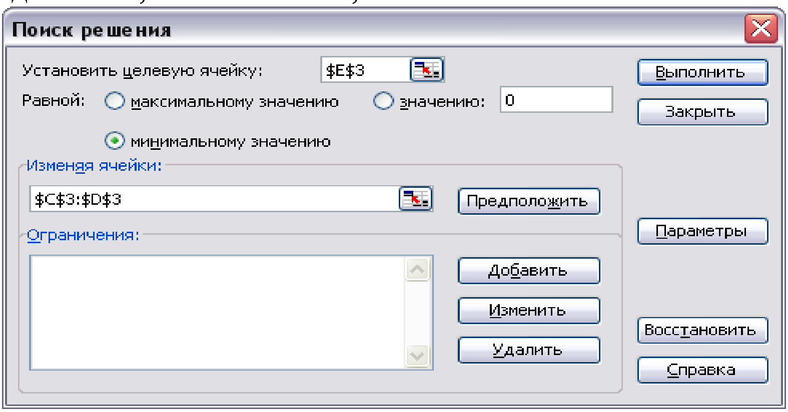


Рис. 21. Диалоговое окно *Поиск решения* для расчета уравнения регрессии

2. Функции рабочего листа для уравнения линейной регрессии

Параметры ***m*** и ***b*** линейной модели ***y* = *mx* + *b*** из предыдущего раздела можно определить с помощью функций НАКЛОН (SLOPE) и ОТРЕЗОК (INTERCEPT).

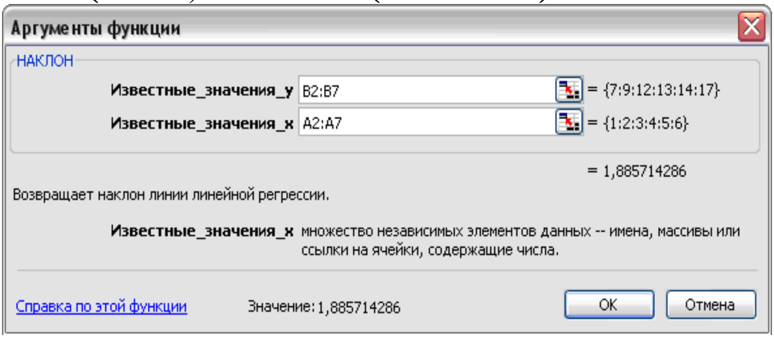


Рис. 22. Диалоговое окно функции НАКЛОН

Функция НАКЛОН (SLOPE) определяет коэффициент наклона линейного тренда.

Синтаксис

НАКЛОН (известные_значения_ y; известные_значения_x)

Функция ОТРЕЗОК (INTERCEPT) определяет точку пересечения линии линейного тренда с осью ординат.

Синтаксис:

ОТРЕЗОК (известные_значения_x; известные_значения_ y)

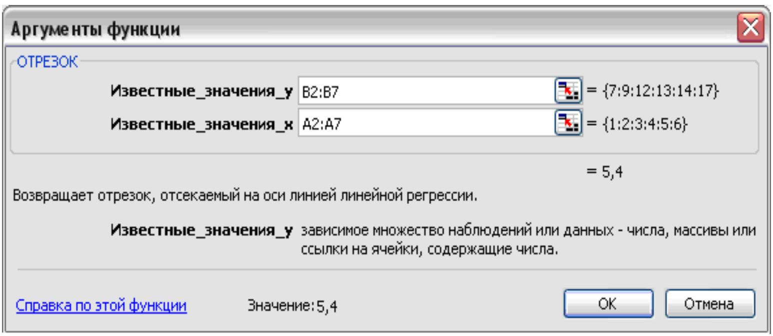


Рис. 23. Диалоговое окно функции ОТРЕЗОК

Аргументы функций НАКЛОН (SLOPE) и ОТРЕЗОК (INTERCEPT):

известные_значения_y - Массив известных значений зависимой наблюдаемой величины

известные_значения_x - Массив известных значений независимой наблюдаемой величины. Если аргумент на *известные_значения_x* опущен, то предполагается, что это массив {1 2; 3;...} такого же размера, как и аргумент *известные_значения_y*

Функции НАКЛОН и ОТРЕЗОК вычисляются по следующим формулам:

$$\boldsymbol{m} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \qquad \boldsymbol{b} = \bar{y} - m\bar{x},$$

где

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}.$$

В ячейках D2 и E2 (рис. 1) найдены ***m*** и ***b***, соответственно, по формулам:

= НАКЛОН (B2 **:** B7 **;** A2 **:** A7)

= ОТРЕЗОК (B2 **:** B7 **;** A2 **:** A7)

Коэффициенты ***m*** и ***b*** можно найти и другим способом. Постройте точечный график по диапазону ячеек A2**:**B7, выделите точки графика двойным щелчком, а затем щелкните их правой кнопкой мыши. В раскрывшемся контекстном меню выберите команду *Линии тренда* (Trendline) (рис. 24).

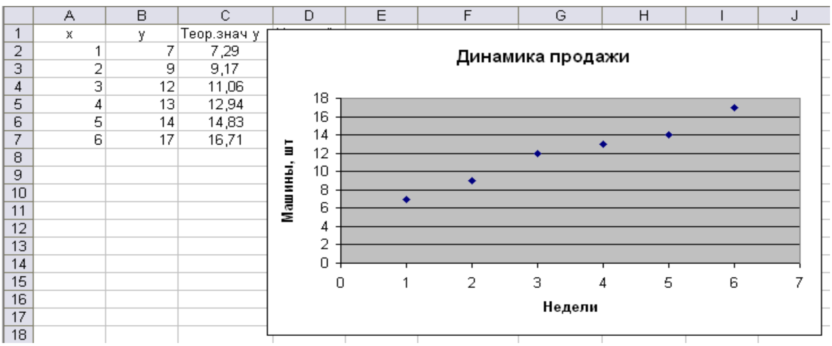


Рис. 24. Начало построения линии тренда

В диалоговом окне *Линия тренда* (Trendline) на вкладке *Тип* (Type) в группе *Построение линии тренда* (аппроксимация и сглаживание) (Trend/Regression type) выберите параметр *Линейная* (Linear) (рис. 25), а на вкладке параметры (Options) установите флажки *Показывать уравнение на диаграмме* (Display Equation on Chart) и *Поместить на диаграмму величину достоверности аппроксимации* (R^2) (Display R-squared) (т. е. на диаграмму необходимо поместить значение квадрата коэффициента корреляции) (рис. 26).

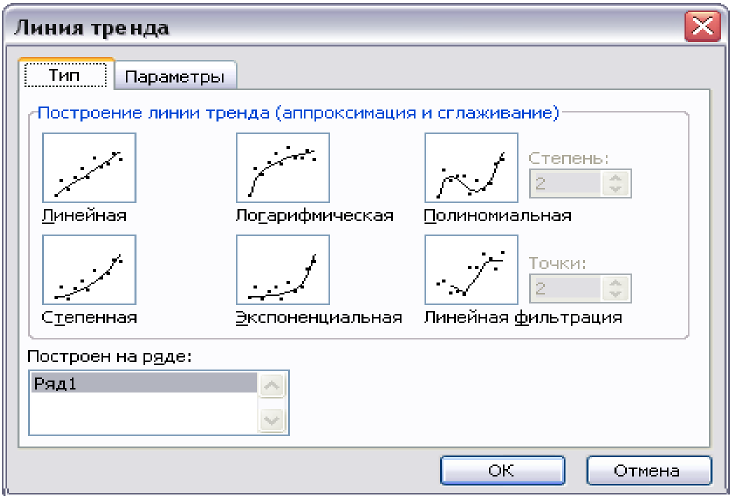


Рис. 25. Вкладка *Тип* диалогового окна *Линия тренда*

По коэффициенту корреляции можно судить о правомерности использования линейного уравнения регрессии. Если он лежит в диапазоне от 0 до 1, то данную зависимость можно использовать для предсказания результата. Чем ближе к единице коэффициент корреляции, тем более обоснованно это указывает на линейную зависимость между наблюдаемыми величинами. Если коэффициент корреляции близок к -1, то это говорит об обратной зависимости между наблюдаемыми величинами.

Флажок *Пересечение кривой с осью Y в точке* (Set Intercept) (рис. 26) устанавливается только в случае, если эта точка известна. Например, если этот флажок установлен и в его поле введен 0, это означает, что ищется модель ***y*** = ***mx***.

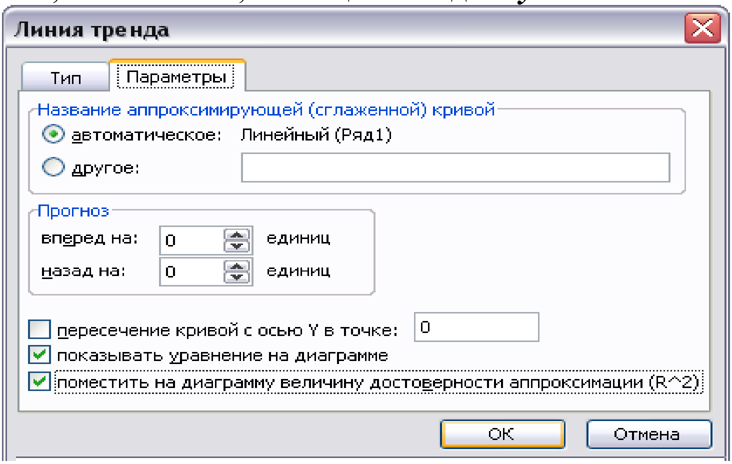


Рис. 26. Вкладка *Параметры* диалогового окна *Линия тренда*

Результат выполнения команды *Линии тренда* (Trendline) приведен на рис. 27.

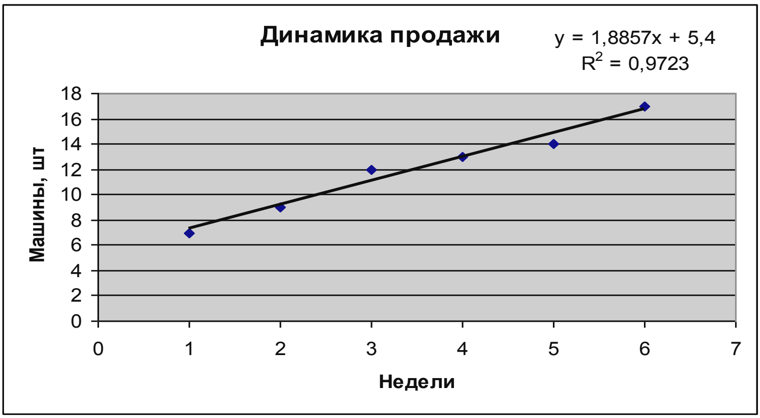


Рис. 27. График линии тренда

Как видно из рисунка, квадрат коэффициента корреляции равен 0,9023, следовательно, линейная модель может быть использована для предсказания результатов.

На основе найденных коэффициентов уравнения регрессии можно определить теоретическое значение наблюдаемой величины ***y***. Вычислим теоретическое значение ***y*** в ячейке С2 (рис. 20) при ***x*** из А2 по формуле `= $D$2*A2+$E$2` .

Однако теоретическое значение ***y*** в фиксированной точке можно вычислить и без предварительного определения коэффициентов линейной модели с помощью функции

ПРЕДСКАЗ (FORECAST).

Синтаксис:

= ПРЕДСКАЗ (t; известные_значения_ y; известные_значения_x)

Аргументы:

t - Точка данных, для которой предсказывается значение

*известные_значения_**y*** - Массив известных значений зависимой наблюдаемой величины

*известные_значения_**x*** - Массив известных значений независимой наблюдаемой величины. Если аргумент *известные_значения_**x*** опущен, то предполагается, что это массив (1 2; 3; ...} такого же размера, как и массив *известные_значения_**y***.

Например, теоретическое значение в ячейке С2 (рис. 20) можно также определить по формуле

= ПРЕДСКАЗ (А2; $B$2 : $B$7; $A$2 : $A$7)

Функция ТЕНДЕНЦИЯ (TREND) вычисляет значения уравнения линейной регрессии для целого диапазона значений независимой переменной как для одномерного, так и для многомерного уравнения регрессии. Многомерная линейная модель регрессии имеет вид:

$$\boldsymbol{y = m_1x_1 + \ldots + m_kx_k + b.}$$

Свитаксис:

= ТЕНДЕНЦИЯ (известные_значения_ y; известные_значения_x; новые_значения_x; конст)

Аргументы:
*известные_значения_**y*** - Массив известных значений зависимой наблюдаемой величины
*известные_значения_**x*** - Массив известных значений независимой наблюдаемой величины. Если аргумент *известные_значения_**x*** опущен, то предполагается, что это массив {1; 2; 3;..} такого же размера, как и массив *известные_значения_**y***
*новые_значения_**x*** - Новые значения ***x***, для которых функция ТЕНДЕНЦИЯ возвращает соответствующие значения **y**
конст - Логическое значение, которое указывает, требуется ли, чтобы константа ***b*** была равна 0. Если аргумент *конст* имеет значение ИСТИНА или опущен, то ***b*** вычисляется обычным образом. Если *конст* имеет значение ложь, то ***b*** полагается равным 0.

Если строится многомерная линейная модель, то аргументы *известные_значения_**x*** и *новые_значения_**x*** должны содержать столбец (или строку) для каждой независимой переменной. Если аргумент *новые_значения_**x*** опущен, ***b*** то предполагается, что он совпадает с аргументом *известные_значения_**x.***

Функция ЛИНЕЙН (LINEST) возвращает массив $\{m_n, ..., m_1, b\}$ значений параметров уравнения многомерной линейной регрессии.

Синтаксис:
= ЛИНЕЙН (известные_значения_y; известные_значения_x; конст; статистика)
Аргументы:
*известные_значения_**y*** - Массив известных значений зависимой наблюдаемой величины
*известные_значения_**x*** - Массив известных значений независимой наблюдаемой величины. Если аргумент *известные_значения_x* опущен, то предполагается, что это массив {1; 2: 3;...} такого же размера, как и *известные_значения_**y***

конст - Логическое значение, которое указывает, требуется ли, чтобы константа ***b*** была равна 0. Если аргумент *конст* имеет значение ИСТИНА или опущен, то ***b*** вычисляется обычным образом. Если *конст* имеет значение ложь, то ***b*** полагается равным 0

статистика - Логическое значение, которое указывает, требуется ли вывести дополнительную статистику по регрессии, на пример, коэффициент корреляции. Если *статистика* имеет значение ИСТИНА, то функция ЛИНЕЙН возвращает дополнительную регрессионную статистику. Если аргумент *статистика* имеет значение ЛОЖЬ или опущен, то функция ЛИНЕЙН возвращает только значения коэффициентов

3. Экспоненциальная модель

Другой часто встречающейся на практике регрессионной моделью является экспоненциальная модель, которая описывается уравнением

$$y = bm^x$$

Значения экспоненциального тренда можно предсказывать с помощью функции РОСТ (GROWTH).

Синтаксис:

= РОСТ (известные_значения_ у; известные_значения_x; новые_значения_x; конст)

Аргументы:

*известные_значения_**y*** - Массив известных значений зависимой наблюдаемой величины

*известные_значения_**x*** - Массив известных значений независимой наблюдаемой величины. Если аргумент *известные_значения_**x*** опущен, то предполагается, что это массив {1; 2; 3...} такого же размера, как и *известные_значения_**y*** *новые_значения_**x***, для которых ТЕНДЕНЦИЯ возвращает соответствующие значения ***y***

конст - Логическое значение, которое указывает, требуется ли, чтобы константа ***b*** была равна 0. Если аргумент *конст* имеет значение ИСТИНА или опущен, то ***b*** вычисляется обычным образом. Если *конст* имеет значение ЛОЖЬ то ***b*** полагается равным 0

Значения параметров зкспоненциальной модели определяются с помощью функции ЛГРФПРИБЛ (LOGRST)

Синтаксис:

=ЛГРФПРИБЛ (известные_значения_ у; известные_значения_х; конст; статистика)

Аргументы:

*известные_значения_**y*** - Массив известных значений зависимой наблюдаемой величины

*известные_значения_**x*** - Массив известных значений независимой наблюдаемой величины. Если аргумент *известные_значения_**x*** опущен, то предполагается, что это массив {1; 2; 3;...} такого же размера, как и *известные_значения_**y***

конст - Логическое значение, которое указывает, требуется ли, чтобы константа ***b*** была равна 0. Если аргумент *конст* имеет значение ИСТИНА или опущен, то ***b*** вычисляется обычным образом. Если *конст* имеет значение ЛОЖЬ то ***b*** полагается равным 0

статистика - Логическое значение, которое указывает, требуется ли вывести дополнительную статистику по регрессии, на пример, коэффициент корреляции. Если статистика имеет значение ИСТИНА, то функция ЛГРФПРИБЛ возвращает дополнительную регрессионную статистику. Если статистика имеет значение ЛОЖЬ или опущена, то функция ЛГРФПРИБЛ возвращает только значения коэффициентов

Кроме того, одномерную экспоненциальную модель можно построить графически (рис. 25). На рис. 28 приведены результаты построения экспоненциального уравнения тренда продажи подержанных автомобилей за 7, 8 и 9-ю недели торговли.

	A	B	C	D	E	F	G	H	I	J
1	x	y	Теор.знач y	Нелинейн						
2	1	7	7,29	7,591		b	m			
3	2	9	9,17	8,970		5,400003	1,886	1,771		
4	3	12	11,06	10,599						
5	4	13	12,94	12,524		5,4	1,8857143			
6	5	14	14,83	14,799			m	b		
7	6	17	16,71	17,488		ЛИНЕЙН	1,8857143	5,4		
8	7	18,60	18,60	20,664		ЛГРФПРИБЛ	1,1816549	6,4237659		
9	8	20,49	20,49	24,418						
10	9	22,37	22,37	28,854						

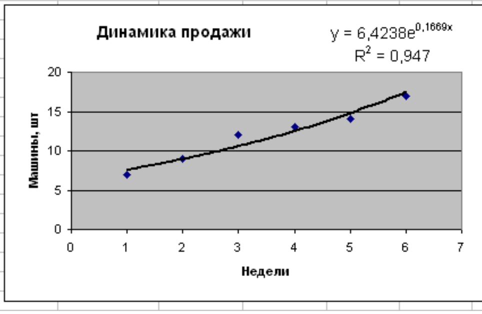


Рис. 28. Экспоненциальная линия тренда

В диапазоне ячеек B8:B10 введена формула построения линейного тренда

{= ТЕНДЕЦИЯ (B2 : B7; A2 : A7; A8 : A10)}

В диапазоне ячеек D2:D10 введена формула построения экспоненциального тренда

{= РОСТ (B2 : B7; A2 : A7; A2 : A10)}

Линейный и экспоненциальный тренды тесно связаны между собой. В диапазоне ячеек D2:D10 можно было бы получить такой же результат, введя формулу

{= EXP (ТЕНДЕНЦИЯ (LN (B2 : B7); A2 : A7; A2 : A10))}

В диапазоны ячеек G7: H7 и G8: H8 введены формулы

{= ЛИНЕЙН (B2 : B2; A2 : A7)}

{= ЛГРФПРИБЛ (B2 : B2; A2 : A7)}

для определения параметров линейной и экспоненциальной моделей.

Квадрат коэффициента корреляции экспоненциальной модели равен 0,947 (рис. 28) и меньше квадрата коэффициента корреляции линейной модели (0,9923) (рис. 27). Таким образом, в данном примере линейная модель более достоверно описывает зависимость между наблюдаемыми величинами.

Порядок выполнения работы

1. Получить у преподавателя данные для расчета.
2. Ввести исходные данные в таблицу Excel.
3. Провести на ЭВМ серию экспериментов с различными типами моделей регрессии с определением прогнозных значений зависимой переменной, параметров регрессионных моделей и коэффициентов детерминации.
4. Зафиксировать результаты расчетов в тетради.
5. Сделать выводы по результатам моделирования и записать в тетради.

Отчет по работе должен содержать

1. Название и цель работы.
2. Основные теоретические и методические положения.
3. Исходные данные для расчета.
4. Результаты расчета.
5. Выводы по результатам моделирования.

Лабораторная работа № 5.
Определение параметров моделей нелинейных зависимостей в форме, определенной пользователем

Цель работы: изучение порядка нахождения параметров уравнения нелинейной регрессии в форме, определенной пользователем и методики оценки достоверности уравнения и коэффициентов регрессии.

Исходные положения. Для нахождения параметров линейного уравнения регрессии используется функция ЛИНЕЙН, для нелинейного уравнения – функция ЛГРФПРИБЛ. Обычно анализ данных производят имея представления о форме зависимости. Если не подходит линейная форма, используют нелинейную функцию. Если же обе функции – ЛИНЕЙН и ЛГРФПРИБЛ не дают удовлетворительного результата, то в этом случае уравнение регрессии можно находить в виде функции, вид которой назначает пользователь. Например, уравнение нелинейной регрессии можно искать в виде полинома второй степени:

$$\boldsymbol{y = b_0 + b_1x_1 + b_2x_2 + b_3x_1^2 + b_4x_2^2 + b_5x_1x_2.}$$

Такую зависимость можно представить в линейной форме (линеаризовать) путем замены квадратов и произведений переменных другими переменными. Например, введем замену $x_3=x_1^2$, $x_4=x_2^2$, $x_5=x_1x_2$; тогда уравнение примет вид:

$$\boldsymbol{y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5.}$$

Оценка достоверности (α) величины $\boldsymbol{R}^2$ уравнения регрессии производится с помощью *Статистической* функции FРАСП. В диалоговое окно ввести адрес ячейки, содержащей $\boldsymbol{F}_{\text{расч}}$.; число степеней свободы, равное количеству регрессоров уравнения (x_1, x_2 …, x_m); адрес ячейки, содержащей число степеней свободы (df) – число измерений минус количество регрессоров минус единица ($\boldsymbol{n-m-1}$).

$$\alpha = \text{FРАСП}(F_{\text{расч}};\ \text{df}_{\text{регр}};\ \text{df}_{\text{остат}})$$

в Excel 2010: α = ***F.РАСП.ПХ(F; m; n-m-1).***

Величина $\boldsymbol{\alpha}$ - это вероятность того, что зависимость $\boldsymbol{y}$ от $\boldsymbol{x_i}$ отсутствует. Для оценки достоверности наличия зависимости $\boldsymbol{y}$ от $\boldsymbol{x_i}$ нужно из единицы вычесть значение, полученное с помощью функции FРАСП.

Достоверность значения определяемых величин $\boldsymbol{b_0}$ и $\boldsymbol{b_i}$ оценивается с помощью вероятности, найденной из распределения Стьюдента. Для этого нужно вычислить величины $\boldsymbol{t_i = b_i / Sb_i}$ ($\boldsymbol{i=\overline{0,n}}$). Далее определим **δ**-вероятность того, что значения $\boldsymbol{b_i}$ и $\boldsymbol{Sb_i}$ не достоверны: вызвать статистическую функцию СТЬЮДРАСП и в диалоговое окно ввести адрес найденного значения $\boldsymbol{t_i}$ (для исключения отрицательных величин используем функцию абсолютной величины ABS); адрес ячейки, содержащей число степеней свободы (df) – число измерений $\boldsymbol{n}$ минус число параметров модели $\boldsymbol{m+1}$ ($\boldsymbol{b_0}$, …, $\boldsymbol{b}_{\text{m}}$); хвосты – 2 (это признак используемого двухстороннего распределения Стьюдента):

δ =СТЬЮДРАСП(ABS(t_i); df; хвосты)
в Excel 2010: δ ***=СТЬЮДЕНТ.РАСП.2Х(ABS(t); df).***

Определим (1 - δ) – вероятность того, что значения b_i достоверны.

Пример выполнения работы. Введем исходные данные в таблицу:

	A	B	C
1	x1	x2	y
2	17	2	50
3	15	17	250
4	13	12	400
5	12	14	500
6	10	10	550
7	9	20	600
8	8	11	550
9	7	15	400
10	6	13	375
11	4	18	150
12	1	5	50

Используя функцию листа ЛИНЕЙН определим параметры уравнения регрессии, коэффициент детерминации и среднеквадратические отклонения:

15	18.73282693	6.21542	61.32999
16	11.36621921	12.82976	203.4766
17	0.259628468	192.5387	#Н/Д
18	1.402692872	8	#Н/Д
19	103998.9033	296569.3	#Н/Д

15	b_2	b_1	b_0
16	Sb_2	Sb_1	Sb_0
17	R^2	$S_{ост}$	#Н/Д
18	F	df = n – m - 1	#Н/Д
19	$SS_{регр}$	$SS_{ост}$	#Н/Д

Параметры уравнения регрессии b_i в матричной форме определяются из выражения $\mathbf{B = (X'X)^{-1}X'Y}$

Дисперсии и средние квадратические отклонения для параметров уравнения регрессии определяются, используя диагональные элементы обратной матрицы $(\mathbf{X'X})^{-1}$ и остаточную дисперсию S^2

$$S_{b_j}^2 = S^2 [(X'X)^{-1}]_{jj} \qquad S^2 = \frac{e'e}{n-(m+1)} = \frac{\sum_{i=1}^{n} e_i^2}{n-m-1}.$$

Выборочный коэффициент множественной детерминации R^2 рассчитывается через суммы квадратов отклонений (SSобщей, SSрегрессионной и SSостаточной) по формуле

$$R^2 = \frac{SS_{регр}}{SS_{общ}} = 1 - \frac{SS_{ост}}{SS_{общ}},$$

где $SS_{общ} = \sum_{i=1}^{n}(y_i - \bar{y})^2$, $SS_{регр} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$, $SS_{ост} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$.

Выборочная стандартная ошибка остатков

$$S = S_e = S_{ост} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n-m-1}} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-m-1}} = \sqrt{\frac{SS_{ост}}{n-m-1}}.$$

Статистическая значимость полученного уравнения регрессии в целом оценивается с помощью F-критерия Фишера. Уравнение множественной регрессии значимо, если

$$F = \frac{R^2 \cdot (n-m-1)}{(1-R^2)\cdot m} > F_{\alpha;m;n-m-1}.$$

Критическое значение F-критерия Фишера определяется с помощью функции =FРАСПОБР(***α; m; n-m-1***) в Excel 2010: =***F.ОБР.ПХ(α; m; n-m-1)*** .

Определим оценки достоверности коэффициента детерминации и коэффициентов уравнения регрессии:

	A	B	C	D
15	18.73282693	6.21542	61.32999	
16	11.36621921	12.82976	203.4766	α
17	0.259628468	192.5387	#Н/Д	=FРАСП(A18;2;8)
18	1.402692872	8	#Н/Д	1 - α
19	103998.9033	296569.3	#Н/Д	=1-D17
20	t_2	t_1	t_0	
21	=A15/A16	=B15/B16	=C15/C16	
22	δ_2	δ_1	δ_0	
23	=СТЬЮДРАСП(ABS(A21);$B$18;2)	=СТЬЮДРАСП(ABS(B21);$B$18;2)	=СТЬЮДРАСП(ABS(C21);$B$18;2)	
24	1- δ_2	1- δ_1	1- δ_0	
25	=1-A23	=1-B23	=1-C23	

В Excel 2010 используются следующие функции

=F.РАСП.ПХ(F; m; n-m-1)   =СТЬЮДЕНТ.РАСП.2Х(ABS(t); df)

В результате получим: коэффициент детерминации R^2=0.259 , оценка достоверности по большинству параметров малозначима.

15	18.73283	6.2154196	61.32999	
16	11.36622	12.829756	203.4766	α
17	0.259628	192.53872	#Н/Д	0.300468
18	1.402693	8	#Н/Д	1 - α
19	103998.9	296569.28	#Н/Д	0,699532
20	t_2	t_1	t_0	
21	1.648114	0.4844535	0.301411	
22	δ_2	δ_1	δ_0	
23	0.137942	0.641049	0.770787	
24	1- δ_2	1- δ_1	1- δ_0	
25	0.862058	0.358951	0.229213	

Воспользуемся функцией листа ЛГРФПРИБЛ для определения параметров нелинейной модели и оценим значимость параметров:

	A	B	C	D
27	1.119235523	1.033221055	48.65104965	
28	0.045577723	0.0514464	0.815926393	α
29	0.438003529	0.77206645	#Н/Д	**=FРАСП(A30;2;8)**
30	3.117482412	8	#Н/Д	1 - α
31	3.716579003	4.768692826	#Н/Д	=1-D29
32	t_2	t_1	t_0	
33	**=(LN(A27)/A28)**	**=(LN(B27)/B28)**	**=(LN(C27)/C28)**	
34	δ_2	δ_1	δ_0	
35	**=СТЬЮДРАСП(ABS(A33); $B$30;2)**	**=СТЬЮДРАСП(ABS(B33); $B$30;2)**	**=СТЬЮДРАСП(ABS(C33); $B$30;2)**	
36	1- δ_2	1- δ_1	1- δ_0	
37	=1-A35	=1-B35	=1-C35	

При расчете параметра $\boldsymbol{t}$, применяя формулу $\boldsymbol{t_i = (\ln(b_i) / \ln(\sigma_i))}$, необходимо использовать функцию натурального логарифма в числителе, а в знаменателе использовать логарифм не следует, поскольку функция ЛГРФПРИБЛ итак возвращает натуральные логарифмы среднеквадратических отклонений.

Результат расчета представлен в следующем виде:

	A	B	C	D
27	1.119235523	1.033221055	48.65104965	
28	0.045577723	0.0514464	0.815926393	α
29	0.438003529	0.77206645	#Н/Д	0,099755
30	3.117482412	8	#Н/Д	1 - α
31	3.716579003	4.768692826	#Н/Д	0,900245
32	t_2	t_1	t_0	
33	2.471511891	0.635246794	4.761058615	
34	δ_2	δ_1	δ_0	
35	0.038619312	0.543005729	0.001424864	
36	1- δ_2	1- δ_1	1- δ_0	
37	0.961380688	0.456994271	0.998575136	

Как видно, коэффициент детерминации невелик, а значимость параметров выше, чем при использовании линейной функции.

Используем модель зависимости в форме, определенной пользователем. Введем исходные данные и преобразуем ряды данных, как это показано в таблице. Новые ряды данных получаются путем возведения рядов $\boldsymbol{x}_1$ и $\boldsymbol{x}_2$ в квадрат и перемножения значений в этих рядах.

В результате получим:

	E	F	G	H	I	J
1	x1	x2	x1^2	x2^2	x1x2	y
2	17	2	289	4	34	50
3	15	17	225	289	255	250
4	13	12	169	144	156	400
5	12	14	144	196	168	500
6	10	10	100	100	100	550
7	9	20	81	400	180	600
8	8	11	64	121	88	550
9	7	15	49	225	105	400
10	6	13	36	169	78	375
11	4	18	16	324	72	150
12	1	5	1	25	5	50

Анализ результатов полученных с помощью последней модели, показывает, что коэффициент детерминации в данном случае значительно выше, чем для линейной и нелинейной зависимостей. Оценки достоверности также более значимы для модели, определенной пользователем.

	E	F	G	H	I	J	K
15	2.437183	2.4131	-11.504	-91.586	188.916	255.877	
16	0.354402	0.4496	0.6870	13.8513	11.1155	66.9332	α
17	0.990664	27.349	#Н/Д	#Н/Д	#Н/Д	#Н/Д	4,5E-05
18	106.11	5	#Н/Д	#Н/Д	#Н/Д	#Н/Д	1- α
19	396828.4	3739.8	#Н/Д	#Н/Д	#Н/Д	#Н/Д	0,9999
20	t_5	t_4	t_3	t_2	t_1	t_0	
21	6.876886	5.3671	-16.744	-6.6121	16.9958	3.82288	
22	δ_5	δ_4	δ_3	δ_2	δ_1	δ_0	
23	0.000995	0.0030	1.4E-05	0.00119	1.3E-05	0.01234	
24	1- δ_5	1- δ_4	1- δ_3	1- δ_2	1- δ_1	1- δ_0	
25	0.999005	0.9969	0.9999	0.99881	0.9999	0.98766	

Порядок выполнения работы

1. Получить у преподавателя данные для расчета.
2. Ввести исходные данные в таблицу Excel.

3. Провести на ЭВМ серию расчетов по определению параметров множественной корреляционной зависимости для линейной, нелинейной модели и модели, определяемой пользователем.

4. Определить оценку достоверности коэффициентов корреляции и детерминации.

5. Зафиксировать результаты расчетов в тетради.

6. Сделать выводы по результатам моделирования и записать в тетради.

Отчет по работе должен содержать

1. Название и цель работы.
2. Основные теоретические и методические положения.
3. Исходные данные для расчета.
4. Результаты расчета.
5. Выводы по результатам моделирования.

Лабораторная работа № 6.
Точечные и интервальные оценки линейной модели

Цель работы: освоение методов оценки уравнения линейной регрессии, определения значимости её параметров и уравнения в целом и построения доверительных интервалов для параметров модели, для линии регрессии и для индивидуальных значений зависимой переменной.

Исходные положения. Для осуществления надежного прогнозирования изменения производственно-экономических процессов необходимо с определить доверительные интервалы, в которые с заданной вероятностью попадают истинные значения анализируемой величины. Порядок проведения расчетов рассмотрим на следующей ситуации.

Таблица 3

Данные по среднедневной заработной плате y_i, руб. и среднедушевому прожиточному минимуму в день одного трудоспособного x_i, руб.

i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	133	148	134	154	162	195	139	158	152	162	159	173
x_i	78	82	87	79	89	106	67	88	73	87	76	115

Требуется:

1. Построить выборочное уравнение линейной парной регрессии (найти значения коэффициентов b_1, b_0).

2. Рассчитать значение выборочного коэффициента корреляции r_{xy}, общую сумму квадратов Q, сумму квадратов, объясненную регрессией Q_r, остаточную сумму квадратов Q_e, несмещенные оценки соответствующих дисперсий S^2, S^2_R, S^2_e, средних квадратических отклонений S, S_R, S_e, выборочный коэффициент детерминации R^2_{yx} и стандартные отклонения коэффициентов регрессии S_{b1}, S_{b0}.

3. На уровне значимости $\alpha = 0{,}05$ оценить значимость коэффициентов и уравнения регрессии. Найти доверительные интервалы для значимых коэффициентов регрессии и значений y_i.

4. Построить графики зависимостей y_i, и $\hat{y}_i$ от x_i, а также доверительные интервалы для значений y_i, и $\hat{y}_i$.

5. Проверить полученные результаты с помощью стандартных статистических функций ТЕНДЕНЦИЯ, ЛИНЕЙН и программы РЕГРЕССИЯ из пакета анализа Microsoft Excel.

Решение

1. Для определения параметров уравнения линейной регрессии строим расчетную таблицу (рис. 29).

2. Строим выборочное уравнение регрессии. Находим выборочные средние:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 85{,}58; \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = 155{,}75; \quad \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i = 13484;$$

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 = 7492{,}25; \quad \overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 = 24531{,}42$$

и, используя *Мастер функций* Microsoft Excel, проверяем полученные значения с помощью стандартной функции СРЗНАЧ (-) из категории *Статистические*, подставляя в нее в качестве аргументов-столбцов векторы соответствующих переменных (например, $\boldsymbol{x} = (\boldsymbol{x_1}\ \boldsymbol{x_2}\ \ldots\ \boldsymbol{x_n})^{\mathrm{T}}$, $\boldsymbol{y} = (\boldsymbol{y_1}\ \boldsymbol{y_2}\ \ldots\ \boldsymbol{y_n})^{\mathrm{T}}$ и т.д.).

	A	B	C	D	E	F	G	H	I	J
1	x	y	(x-x ср)2	Se i	Se i инд	y расч	y мин	y макс	y мин инд	y макс инд
2	78	133	57,507	4,198	13,233	148,770	139,416	158,124	119,285	178,255
3	82	148	12,840	3,759	13,100	152,452	144,077	160,827	123,262	181,641
4	87	134	2,007	3,644	13,068	157,054	148,934	165,174	127,937	186,171
5	79	154	43,340	4,064	13,191	149,690	140,636	158,745	120,299	179,082
6	89	162	11,674	3,747	13,097	158,895	150,547	167,243	129,713	188,077
7	106	195	416,840	6,763	14,256	174,542	159,473	189,611	142,778	206,306
8	67	139	345,340	6,336	14,058	138,645	124,528	152,763	107,321	169,969
9	88	158	5,840	3,685	13,080	157,974	149,763	166,186	128,831	187,117
10	73	152	158,340	5,051	13,528	144,168	132,913	155,422	114,026	174,310
11	87	162	2,007	3,644	13,068	157,054	148,934	165,174	127,937	186,171
12	76	159	91,840	4,507	13,334	146,929	136,888	156,971	117,219	176,640
13	115	173	865,340	8,991	15,438	182,826	162,794	202,858	148,429	217,223
14	85,58333	x ср								

Рис. 29. Исходные данные и доверительные интервалы

Находим значения выборочных дисперсий и средних квадратических отклонений

$$S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \overline{x^2} - (\bar{x})^2 = 167{,}74;\ S_x = 12{,}95;$$

$$S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \overline{y^2} - (\bar{y})^2 = 273{,}35;\ S_y = 16{,}53$$

и проверяем полученные значения с помощью стандартных статистических функций ДИСПР (-) и СТАНДОТКЛОНП (-) соответственно. В Excel 2010: ДИСП.Г(-) и СТАНДОТКЛОН.Г (-).

Находим выборочный коэффициент ковариации

$$C\hat{o}v(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \overline{xy} - \bar{x}\,\bar{y} = 154{,}396$$

и проверяем полученное значение с помощью стандартной статистической функции =КОВАР(***x***; ***y***) или в Excel 2010: =КОВАРИАЦИЯ.Г(***x***; ***y***).

Рассчитываем значения выборочных коэффициентов регрессии

$$\boldsymbol{b}_1 = \frac{\hat{Cov}(\boldsymbol{x},\boldsymbol{y})}{\boldsymbol{S}_x^2} = 0{,}920431;\ \boldsymbol{b}_0 = \bar{\boldsymbol{y}} - \boldsymbol{b}_1\bar{\boldsymbol{x}} = 76{,}98$$

и проверяем полученные значения с помощью стандартных статистических функций НАКЛОН ($\boldsymbol{y}$; $\boldsymbol{x}$) и ОТРЕЗОК ($\boldsymbol{y}$; $\boldsymbol{x}$) соответственно.

Величина коэффициента $\boldsymbol{b}_1$ показывает, что с увеличением прожиточного минимума на 1 руб. среднедневная заработная плата возрастает в среднем на 0,92 руб. Параметр $\boldsymbol{b}_0$ мы не интерпретируем, поскольку в выборке отсутствуют значения $\boldsymbol{x_i}$ факторного признака, близкие к нулю.

3. Рассчитываем значение выборочного коэффициента корреляции по формуле

$$\boldsymbol{r}_{xy} = \boldsymbol{b}_1 \frac{\boldsymbol{S}_x}{\boldsymbol{S}_y} = \frac{\hat{Cov}(\boldsymbol{x},\boldsymbol{y})}{\boldsymbol{S}_x \boldsymbol{S}_y} = 0{,}721$$

и проверяем полученное значение с помощью стандартной статистической функции =КОРРЕЛ (x; y).

Подставляя рассчитанные значения $\boldsymbol{b}_0$ и $\boldsymbol{b}_1$ в формулу $\hat{\boldsymbol{y}}_i = \boldsymbol{b}_0 + \boldsymbol{b}_1\boldsymbol{x}_i$, находим величины $\hat{\boldsymbol{y}}_i$, ($\boldsymbol{i}$ = 1, 2, ..., $\boldsymbol{n}$). Для одного, произвольно выбранного $\boldsymbol{k}$-го значения $\hat{\boldsymbol{y}}_k$, отвечающего аргументу $\boldsymbol{x_k}$, проверяем полученный результат с помощью стандартной статистической функции ПРЕДСКАЗ ($\boldsymbol{x_k}$, $\boldsymbol{y}$; $\boldsymbol{x}$).

Вычисляем значения $(y_i - \bar{y})^2, (\hat{y}_i - \bar{y})^2, (y_i - \hat{y}_i)^2$ и рассчитываем соответствующие суммы квадратов (общую, обусловленную регрессией и остаточную), дисперсия на степень свободы и средние квадратические отклонения (если $\boldsymbol{n}$ – число наблюдений, $\boldsymbol{m}$ – число объясняющих переменных):

$$\boldsymbol{Q} = \sum_{i=1}^{n}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^2 = 3280{,}25;\ \boldsymbol{s}^2 = \frac{\boldsymbol{Q}}{\boldsymbol{n}-1} = 298{,}20;\ \boldsymbol{s} = 17{,}27;$$

$$Q_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = 1705{,}33;\ s_R^2 = \frac{Q_R}{m} = 1705{,}33;\ s_R = 41{,}296;$$

$$Q_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 1574{,}92;\ s_e^2 = \frac{Q_e}{n-m-1} = 157{,}49;\ s_e = 12{,}55.$$

Полученное значение стандартной ошибки S_e проверяем с помощью статистической функции =СТОШYX (***y***; ***x***).

Рассчитываем величину выборочного коэффициента детерминации

$$R^2_{yx} = \frac{Q_R}{Q} = 1 - \frac{Q_e}{Q} = 0{,}52$$

и проверяем полученное значение с использованием эквивалентного выражения $R^2_{xy} = r^2_{xy}$ а также с помощью стандартной статистической функции =КВПИРСОН (***y***; ***x***).

Величина коэффициента R^2_{xy} показывает, что 52% вариации зависимой переменной объясняется вариацией предикторной переменной, а остальные 48% - влиянием неучтенных и случайных факторов.

Находим стандартные отклонения оценок коэффициентов регрессии по формулам

$$s_{b_1} = \frac{s_e}{s_x\sqrt{n}} = 0{,}28; \; s_{b_0} = \frac{s_e}{s_x\sqrt{n}}\sqrt{\overline{x^2}} = s_{b_1}\sqrt{\overline{x^2}} = 24{,}21.$$

Теперь выборочное уравнение регрессии можно записать в общепринятом виде (под коэффициентами в скобках указаны их стандартные отклонения):

$$\hat{y}_i = \underset{(24,21)}{76{,}976} + \underset{(0,2797)}{0{,}92}x_i.$$

4. Вычисляем статистики критерия значимости коэффициентов регрессии:

$$t_{b_1} = \frac{b_1}{s_{b_1}} = 3{,}29; \; t_{b_0} = \frac{b_0}{s_{b_0}} = 3{,}179.$$

Находим значение критической точки с помощью стандартной статистической функции =СТЬЮДРАСПОБР(α; ***n***–***m***–1) для заданного уровня значимости α = 0,05:

$t_{кр}(\alpha; k = n - m - 1) = 2{,}228$,   ($n - m - 1 = 12 - 2 = 10$).

Поскольку $|t_{b_j}| > t_{кр}$, с уровнем значимости 0,05 (с доверительным уровнем 95%) делаем вывод о том, что коэффициенты β_0 и β_1 значимы.

Вычисляем $\boldsymbol{P}$-значения для коэффициентов с помощью статистической функции =СТЬЮДРАСП ($|t_{b_j}|$, $n - m - 1$; хвосты):

P_{b0} = 0,0098 - для коэффициента β_0;

P_{b1} = 0,0081 - для коэффициента β_1;

хвосты = 2 – двустороннее t-распределение.

В Excel 2010 используется ***=СТЬЮДЕНТ.РАСП.2Х(ABS(t); df).***

В силу того, что $P_{b_j} < \alpha$, вывод о значимости коэффициентов регрессии подтверждается.

Определяем значение F-статистики по формуле

$$F = \frac{S_r^2}{S_e^2} = \frac{Q_r(n-m-1)}{Q_e \cdot m} = 10{,}828$$

и проверяем полученное значение с использованием эквивалентных формул – формулы $F = \dfrac{R_{yx}^2(n-m-1)}{(1-R_{yx}^2)\cdot m}$ и соотношения $F = t_{b_i}^2$.

Критическое значение статистики Фишера - Снедекора для заданного уровня значимости α = 0,05 находим с помощью стандартной статистической функции =FРАСПОБР (α; m; $n - m - 1$)

$$F_{кр} = (\alpha;\ k_1 = m = 1;\ k_2 = n - m - 1 = 10) = 4{,}96,$$

проверяя полученное значение по формуле $F_{кр} = t_{кр}^2$.

В Excel 2010 используется =***F.ОБР.ПХ(α; m; n-m-1).***

В силу того, что $F > F_{кр}$, с доверительным уровнем 0,95 делаем вывод о том, что уравнение регрессии значимо.

Вычисляем величину P-значения с помощью статистической функции =FРАСП (F; $k_1 = m = 1$; $k_2 = n - m - 1 = 10$)

$$p = 0{,}008.$$

В Excel 2010 используется ***=F.РАСП.ПХ(F; m; n-m-1) .***

Поскольку $p < \alpha$, вывод о значимости уравнения регрессии подтверждается.

	A	B	C	D	E	F	G	H	I
16	b1min	0,297185	23,02976	b0min	Qx	2012,917			
17	b1max	1,543676	130,9232	b0max	S2x	167,743	167,743	273,354	S2y
18					Sx	12,952	12,952	16,533	Sy
19	b1	0,920431	76,97649	b0					
20	Sb1	0,279716	24,21156	Sb0			Cov	154,3958	
21	R2xy	0,519877	12,54959	Se	Se2	157,49223			
22	F	10,82801	10	df	Fкр	4,9646027	b0	76,97649	
23	Qr	1705,328	1574,922	Qe	P	0,0081418	b1	0,920431	
24									
25		tb1	tb0	tкр			Se	12,54959	
26		3,290594	3,179328	2,228139					
27		Pb1	Pb0				R2xy	0,519877	
28		0,008142	0,009831						

Рис. 30. Проверка значимости модели

Нижние и верхние границы доверительного интервала коэффициентов регрессии β_j (нижние и верхние $\gamma \cdot 100\%$) с доверительной вероятностью $\gamma = 1 - \alpha$ найдем по формуле

$$b_j - t_{кр}Sb_j < \beta_j < b_j + t_{кр}Sb_j$$

$$\beta_{1min} = b_1 - t_{кр}Sb_1 = 0{,}297;\ \beta_{1max} = b_1 + t_{кр}Sb_1 = 1{,}544;$$

$$\beta_{0min} = b_0 - t_{кр}Sb_0 = 23{,}03;\ \beta_{0max} = b_0 + t_{кр}Sb_0 = 130{,}92.$$

Для получения доверительного интервала для линии регрессии находим несмещенную оценку дисперсии $s_{\hat{y}}^2$ прогноза величин y, соответствующих значениям по формуле

$$s_{\hat{y}}^2 = s_e^2\left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{ns_x^2}\right],$$

и, вычисляя корень, определим значения Se_i или $s_{\hat{y}}^2$.

Нижние $y_{i\,min}$ и верхние $y_{i\,max}$ границы доверительного интервала для математического ожидания зависимой величины определяем по формуле

$$y_{min} < M_x(Y) < y_{max},$$

где $y_{i\,min} = \hat{y}_i - t_{\alpha;n-m-1}s_{\hat{y}}$, $y_{i\,max} = \hat{y}_i + t_{\alpha;n-m-1}s_{\hat{y}}$

и приводим их величины в таблице на рис. 29.

Для получения доверительного интервала для индивидуальных значений зависимой переменной находим несмещенную оценку дисперсии $s_{\hat{y}_o}^2$ прогноза величин y, соответствующих значениям по формуле

$$s_{\hat{y}_o}^2 = s_e^2\left[1+\frac{1}{n}+\frac{(x_i-\bar{x})^2}{ns_x^2}\right]$$

и, вычисляя корень, определим значения $S_{ei\text{инд}}$ или $s_{\hat{y}_o}^2$

Нижние $y_{i\,min}$ и верхние $y_{i\,max}$ границы доверительного интервала для значений y_i определяем по формуле

$$y_{i\,min} < y_i < y_{i\,max},$$

где $y_{i\,min} = \hat{y}_i - t_{\alpha;n-m-1}s_{\hat{y}_o}$, $y_{i\,max} = \hat{y}_i + t_{\alpha;n-m-1}s_{\hat{y}_o}$

и приводим их величины в таблице на рис. 29.

5. При построении графиков используем *Мастер диаграмм* Microsoft Excel в следующем порядке.

Шаг 1 - тип диаграммы. На вкладке *Стандартные* выбираем *Точечную диаграмму*, позволяющую сравнить пары значений. Нажимаем кнопку *Далее*.

Шаг 2 - источник данных диаграммы. На вкладке диапазон данных выделяем диапазон (x; y) и указываем, что ряды находятся в столбцах. Переходим на вкладку *Ряд*. В поле *Графика* просматриваем полученный результат. В поле *Имя* указываем название «y».

Последовательно нажимая кнопку *Добавить*, добавляем ряды значений $\hat{y}_i$, $y_{i\,min}$, $y_{i\,max}$ аналогичным образом и задаем названия «y_p», «y_{min}», «y_{max}». После просмотра результатов нажимаем кнопку *Далее*.

Шаг 3 - параметры диаграммы. На вкладке *Заголовки* в полях *Название диаграммы*, *Ось X* (*категорий*) и *Ось Y* (*значений*) задаем соответствующие названия «Зависимость среднедневной заработной платы, руб. от среднедушевого прожиточного минимума в день, руб.», «Среднедушевой прожиточный минимум в день, руб.» и «Среднедневная заработная плата, руб.». На вкладке *Линии сетки* добавляем *основные линии на оси X* (*категорий*). Остальные вкладки оставляем без изменения. После просмотра результатов нажимаем кнопку *Далее*.

Шаг 4 - размещение диаграммы. Помещаем диаграмму на имеющемся листе и нажимаем кнопку *Готово*.

В результате получаем диаграмму, показанную на рис. 31.

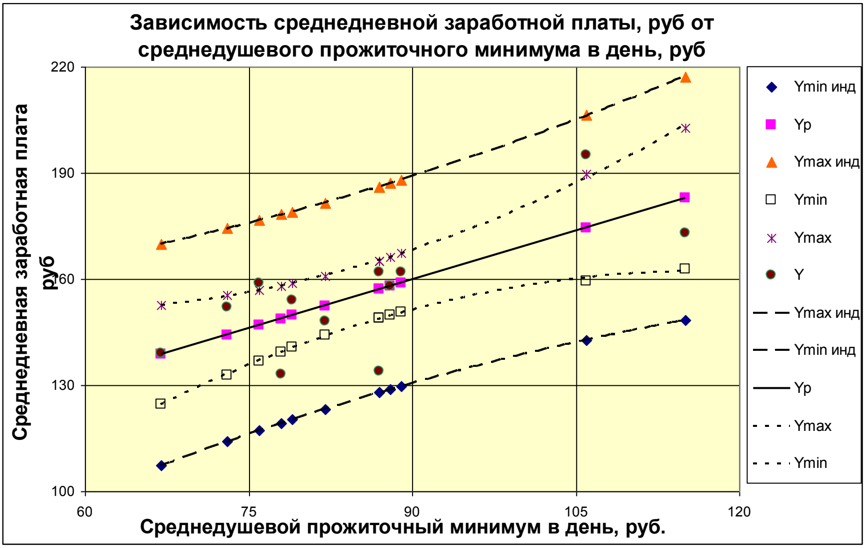


Рис. 31. Линия тренда и доверительные интервалы

Y - исходные данные (y_i); Y_p - линейная регрессия ($\hat{y}_i$); Y_{min} - нижняя граница доверительного интервала для линии регрессии ($y_{i\,min}$); Y_{max} - верхняя граница доверительного интервала для линии регрессии ($y_{i\,max}$); $Y_{min\ инд}$ - нижняя граница доверительного интервала для индивидуальных значений; $Y_{max\ инд}$ - верхняя граница доверительного интервала для индивидуальных значений.

Параметры полученной диаграммы можно изменять, используя меню *Диаграмма* или контекстное меню, вызываемое щелчком правой кнопки мыши.

В частности, целесообразно задать новые значения шкалы осей, чтобы расположить графики наилучшим образом. Для этого необходимо выбрать команду *Формат оси* и на вкладке

Шкала задать требуемые величины в полях *Минимальное значение, Максимальное значение* и *Цена основных делений,* убрав флажки из соответствующих полей *Авто*.

Для более наглядного представления результатов необходимо выбрать ряд $\boldsymbol{y_p}$, с помощью контекстного меню выбрать команду *Формат ряда данных* и на вкладке *Вид* задать параметры линии н маркера (можно также вызвать команду *Добавить линию тренда* и в поле *Линия тренда* на вкладке *Тип* выбрать поле *Линейная*). Для рядов $\boldsymbol{y_{i\,min}}$ и $\boldsymbol{y_{i\,max}}$ аналогичным образом добавляется линия тренда *Полиномиальная* при значении степени, равном 2 (по умолчанию).

С помощью команды *Формат линии тренда* при необходимости на вкладке *Вид* выбирается тип, цвет и толщина линии, а на вкладке *Параметры* - название аппроксимирующей кривой и величина интервала прогноза вперед или назад на заданное число единиц. Здесь также задается возможность показать уравнение регрессии и коэффициент детерминации в поле диаграммы.

6. Для определения значений результативного признака по линейному уравнению регрессии с помощью стандартной статистической функции ТЕНДЕНЦИЯ выполняем следующие операции:

в расчетной таблице (рис. 29) озаглавливаем столбец (например, символом «$\boldsymbol{y_p}$» или словом «тенденция» и выделяем 12 значащих позиций этого столбца ($\boldsymbol{i}$ = 1, 2, ..., $\boldsymbol{n}$);

• с помощью *Мастера функций* выбираем статистическую функцию ТЕНДЕНЦИЯ;

• в поля *Изв_знач_у* и *Изв_знач_х* вводим значения векторов $\boldsymbol{y}$ и $\boldsymbol{x}$ соответственно;

• поле *Нов_знач_х* оставляем пустым (при этом предполагается, что *Нов_ знач_х* совпадают с *Изв_знач_х*);

• поле *Константа* оставляем пустым (если Константа имеет значение ИСТИНА, 1 или опущена, то коэффициент $\boldsymbol{b_0}$ вычисляется обычным образом, если *Константа* имеет значение ЛОЖЬ или 0, то коэффициент $\boldsymbol{b_0}$ полагается равным нулю);

• контролируем результат решения в окне функции (первый элемент массива) $\hat{y}_1$ = 148,770;

• для получения массива результатов (вывода формулы массива) нажимаем комбинацию клавиш *Ctrl+Shift+Enter* (в выделенном столбце появятся результаты вычислений).

Для определения параметров линейного уравнения регрессии с помощью стандартной статистической функции ЛИНЕЙН выполняем следующие операции:

• с помощью Мастера функций выбираем статистическую функцию ЛИНЕЙН;

• в поля *Изв_знач_y* и *Изв_знач_x* вводим значения векторов $\boldsymbol{y}$ и $\boldsymbol{x}$ соответственно;

• поле *Константа* оставляем пустым (если Константа имеет значение ИСТИНА, 1 или опущена, то коэффициент $\boldsymbol{b_0}$ вычисляется обычным образом, если Константа имеет значение ЛОЖЬ или 0, то коэффициент $\boldsymbol{b_0}$ полагается равным нулю);

• в поле *Стат* вводим значение ИСТИНА или 1 (если *Стат* имеет значение ИСТИНА или 1, то вычисляется дополнительная статистика - строки 3-6 в табл. 4, если *Стат* имеет значение ЛОЖЬ, 0 или опущена, то вычисляются только значения коэффициентов и - вторая строка в табл. 4);

• контролируем результат решения в окне функции (первый элемент массива) $\boldsymbol{b_1}$ = 0,920431;

• для получения массива результатов (вывода формулы массива) нажимаем комбинацию клавиш *Ctrl+Shift+Enter*;

• в выделенных ячейках появятся результаты вычислений, представленные в табл. 4 ($\boldsymbol{b_1}$, $\boldsymbol{b_0}$ - выборочные оценки коэффициентов регрессии; $\boldsymbol{Sb_1}$, $\boldsymbol{Sb_0}$ - стандартные отклонения коэффициентов регрессии; $\boldsymbol{R^2_{xy}}$ - выборочный коэффициент детерминации; $\boldsymbol{S_e}$ - значение стандартной ошибки; $\boldsymbol{F}$ - значение $\boldsymbol{F}$-статистики для уравнения регрессии; $\boldsymbol{df}$ = $\boldsymbol{n}$-2 - число степеней свободы; $\boldsymbol{Q_r}$ и $\boldsymbol{Q_e}$ - факторная и остаточная суммы квадратов соответственно).

Таблица 4

Результаты расчета

b_1	0,920431	76,97649	b_0
Sb_1	0,279716	24,21156	Sb_0
R^2_{xy}	0,519877	12,54959	S_e
F	10,82801	10	df
Q_r	1705,328	1574,922	Q_e

Для получения решения с помощью подпрограммы РЕГРЕССИЯ из пакета анализа выполняем следующие операции:

• выбираем команду *Анализ данных* в меню *Сервис* (если она отсутствует, необходимо в меню *Сервис* выбрать команду *Надстройки* и в появившемся окне диалога выбрать пункт *Пакет анализа*);

• в окне *Анализ данных* выбираем инструмент *Регрессия* (при использовании этого инструмента данные обязательно должны быть расположены по столбцам);

• в категории *Входные данные* в поля *Входной интервал Y* и *Входной интервал X* вводим значения векторов $\boldsymbol{y}$ и $\boldsymbol{x}$ соответственно, а остальные поля оставляем пустыми (флажок в поле *Метки* ставится, если в соответствующие входные интервалы включены названия столбцов; флажок в поле *Константа* - ноль ставится, когда коэффициент $\boldsymbol{b_0}$ полагается равным нулю; флажок в поле *Уровень надежности* ставится в случаях, когда необходимо задать величину доверительного уровня $\boldsymbol{\gamma}\cdot 100\%$, отличную от 95%);

• в категории *Параметры вывода* оставляем переключатель в положении *Новый рабочий лист*, при необходимости задавая имя листа в поле ввода рядом с параметром (этот параметр вставляет новый лист в рабочую книгу и располагает результаты, начиная с ячейки A1 нового листа; параметр *Выходной интервал* позволяет ввести ссылку для левой верхней ячейки интервала, в который выводятся результаты на текущем рабочем листе; параметр *Новая рабочая книга* создает новую рабочую книгу, добавляя в нее новый лист и вставляя результаты в ячейку A1 этого листа);

• в категории *Остатки* ставим флажки в полях *Остатки*, *Стандартизированные остатки*, *График остатков*, *График подбора*. Последний позволяет вывести точечные графики зависимости наблюдаемых y и теоретических $\hat{y}$ результативных значений от факторных признаков x_i;

• в категории *Нормальная вероятность* ставим флажок в поле *График нормальной вероятности*. Это позволяет вывести точечный график зависимости наблюдаемых значений у от автоматически формируемых интервалов персентилей.

Результаты расчетов выводятся в виде пяти таблиц и трех диаграмм. Содержание таблиц под общим названием *Вывод итогов* показано на рис. 32.

Таблица *Регрессионная статистика*. В таблице представлены:

• *Множественный R* - множественный выборочный коэффициент корреляции R_{xy}, равный квадратному корню из коэффициента детерминации и, для парной регрессии, совпадающий с выборочным коэффициентом корреляции r_{xy};

• *R-квадрат* - коэффициент детерминации R^2_{xy};

• *Нормированный R-квадрат* - для парной регрессии определяется выражением $R^2_{adj} = 1 - \dfrac{Q_e/(n-m-1)}{Q/(n-1)} = 1 - \dfrac{(n-1)}{(n-m-1)}(1 - R^2_{xy})$ (вычисление этого коэффициента целесообразно только для множественной регрессии); m – число факторных признаков;

• *Стандартная ошибка* - корень из несмещенной оценки остаточной дисперсии

$$S_e = \sqrt{\frac{Q_e}{n-(m+1)}} = \sqrt{\frac{1}{n-(m+1)} \cdot \sum_{i=1}^{n} (\hat{y}_i - y_i)^2};$$

• наблюдения - число наблюдений в выборке n.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,721025
R-квадрат	0,519877
Нормированный R-кв	0,471865
Стандартная ошибка	12,54959
Наблюдения	12

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	1705,327706	1705,3277	10,82801	0,00814184
Остаток	10	1574,922294	157,49223		
Итого	11	3280,25			

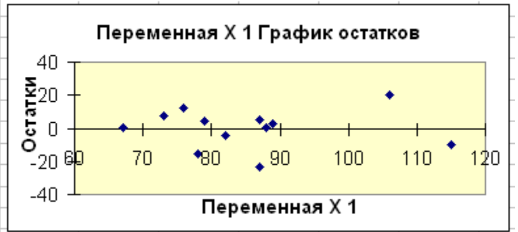
	Коэффиц	*Стандартная ош*	*t-статист*	*P-Значен*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	76,97649	24,21156138	3,1793276	0,009831	23,0297648	130,923206
Переменная X 1	0,920431	0,279715587	3,2905944	0,008142	0,29718539	1,54367572

ВЫВОД ОСТАТКА

Наблюдение	*Предсказ*	*Остатки*	*Стандартные ост*
1	148,7701	-15,77006831	-1,317954
2	152,4518	-4,45179052	-0,37205
3	157,0539	-23,05394328	-1,926691
4	149,6905	4,309501138	0,3601586
5	158,8948	3,105195612	0,259511
6	174,5421	20,45787622	1,7097292
7	138,6453	0,354667771	0,0296407
8	157,9744	0,025626164	0,0021417
9	144,1679	7,832084455	0,654552
10	157,0539	4,946056717	0,4133575
11	146,9292	12,0707928	1,0087942
12	182,826	-9,825998758	-0,82119

ВЫВОД ВЕРОЯТНОСТИ

Персентиль	*Y*
4,16666667	133
12,5	134
20,8333333	139
29,1666667	148
37,5	152
45,8333333	154
54,1666667	158
62,5	159
70,8333333	162
79,1666667	162
87,5	173
95,8333333	195



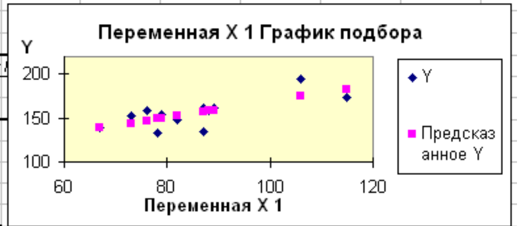


Рис. 32. Вывод итогов расчета параметров линейной регрессионной модели

Таблица *Дисперсионный анализ*. В таблице представлены (по столбцам соответственно для строк *Регрессия*, *Остаток*, *Итого*):

• $\boldsymbol{df}$ - число степеней свободы ($\boldsymbol{df} = \boldsymbol{m}$ - для объясненной дисперсии, $\boldsymbol{df} = \boldsymbol{n}\text{-}\boldsymbol{m}\text{-}1$ - для остаточной дисперсии, $\boldsymbol{df}=\boldsymbol{n}-1$ - для общей дисперсии $\boldsymbol{df} = \boldsymbol{m} + \boldsymbol{n}\text{-}\boldsymbol{m}\text{-}1$);

• $\boldsymbol{SS}$ – сумма квадратов ($Q_R = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$, объясненная регрессией, $Q_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ - остаточная, $Q = Q_R + Q_e = \sum_{i=1}^{n}(y_i - \bar{y})^2$ - общая);

• $\boldsymbol{MS}$ – несмещенные оценки дисперсий ($s_R^2 = \frac{Q_R}{m}$ - объясненная регрессией, $s_e^2 = \frac{Q_e}{n-m-1}$ - остаточная);

• $\boldsymbol{F}$ - вычисленное значение статистики Фишера - Снедекора $F = \frac{s_R^2}{s_e^2} = \frac{Q_R(n-m-1)}{Q_e \cdot m}$;

• Значимость $\boldsymbol{F}$ - величина $\boldsymbol{P}$-значения для выборочного уравнения регрессии,

$$P = \int_F^\infty f_F(\xi; k_1, k_2) d\xi).$$

Таблица с информацией о параметрах выборочного уравнения регрессии. В ней по столбцам соответственно для строк *Y-пересечение* (коэффициент $\boldsymbol{b_0}$) и *Переменная X1* (коэффициент $\boldsymbol{b_1}$) представлены:

• *Коэффициенты* - значения коэффициентов $\boldsymbol{b_0}$ и $\boldsymbol{b_1}$;

• *Стандартная ошибка* - стандартные отклонения коэффициентов регрессии $\boldsymbol{Sb}_0$ и $\boldsymbol{Sb}_1$;

• $\boldsymbol{t}$-статистика - статистики критерия значимости $\boldsymbol{tb}_0$ и $\boldsymbol{tb}_1$ коэффициентов регрессии $\boldsymbol{\beta}_0$ и $\boldsymbol{\beta}_1$;

• $\boldsymbol{P}$-значения - величины $\boldsymbol{P}$-значений $\boldsymbol{Pb}_0$ и $\boldsymbol{Pb}_1$ для коэффициентов $\boldsymbol{\beta}_0$ и $\boldsymbol{\beta}_1$;

• Нижние 95% и Верхние 95% - значения соответствующих интервальных оценок $\boldsymbol{\beta}_{\mathbf{j}\,min} = \boldsymbol{b}_\mathbf{j} - \boldsymbol{t}_{\text{кр}}\boldsymbol{Sb}_\mathbf{j}$ и $\boldsymbol{\beta}_{\mathbf{j}max} = \boldsymbol{b}_\mathbf{j} +$

$t_{кр}S_{b_j}$ для коэффициентов β_j при уровне значимости $\alpha = 0{,}05$, $\gamma \cdot 100\% = 95\%$ (в случае задания другого доверительного уровня, например $\gamma = 1 - \alpha = 0{,}9$, в этих столбцах все равно будут указаны 95% границы, а в следующих двух столбцах - 90%).

Таблица *Вывод остатка*. В таблице представлены:

• *Наблюдение* - порядковые номера i выборочных значений y_i и x_i, ($i = 1, 2, \ldots, n$);

• *Предсказанное Y* – значения $\hat{y}_i$ рассчитанные по выборочному уравнению регрессии $\hat{y}_i = b_0 + b_1 x_i$;

• *Остатки* - значения остатков регрессии e_i (выборочная оценка возмущений ε_i);

• *Стандартные остатки* - значения нормированных остатков регрессии $e_i^0 = \dfrac{e_i}{s^0}$, где $s^0 = \sqrt{\dfrac{1}{n-1} \cdot \sum_{i=1}^{n} (e_i - \bar{e})^2}$; $\bar{e} = \dfrac{1}{n} \sum_{i=1}^{n} e_i$.

Таблица *Вывод вероятности*. В таблице представлены:

• *Персентиль* - рассчитывается для каждого значения y_i как сумма предшествующего вычисленного значения персентиля и шага $h = 100\% / n$ (при этом начальное и конечное значения равны $h / 2$ и $100\% - h / 2$ соответственно);

• *Y* - значения y_i, расположенные в неубывающем порядке.

Три диаграммы (которые показаны на рис. 27) включают в себя:

• диаграмму *Переменная X1 График остатков* - график зависимости e_i от x_i;

• диаграмму *Переменная X1 График подбора* - графики зависимостей y_i и $\hat{y}_i$ от x_i ;

• диаграмму *График нормального распределения*, строящуюся по данным таблицы *Вывод вероятности*.

Порядок выполнения работы

1. Получить у преподавателя данные для расчета.
2. Ввести исходные данные в таблицу Excel.
3. Провести на ЭВМ серию расчетов по определению параметров регрессионной зависимости, точечных и интервальных оценок.
4. Построить графическую интерпретацию доверительных интервалов для линии регрессии и индивидуальных значений зависимой переменной.
5. Зафиксировать результаты расчетов в тетради.
6. Сделать выводы по результатам моделирования и записать в тетради.

Отчет по работе должен содержать

1. Название и цель работы.
2. Основные теоретические и методические положения.
3. Исходные данные для расчета.
4. Результаты расчета.
5. Выводы по результатам моделирования.

Лабораторная работа № 7. Регрессионная модель с гетероскедастичностью. Метод взвешенных наименьших квадратов

Цель работы: изучение частного случая регрессионной модели - модели с гетероскедастичностью. Это означает, что ошибки (отклонения ε) не коррелированы, но имеют непостоянные дисперсии.

Исходные положения. Классическая модель с постоянными дисперсиями ошибок называется гомоскедастичной. Возникает гетероскедастичность чаще при анализе неоднородных объектов.

Наиболее распространенным в практике статистического оценивания параметров уравнений регрессии является метод наименьших квадратов (МНК). Этот метод основан на ряде предпосылок относительно природы данных и результатов построения модели. Основные из них - это некоррелированность факторов, исходящих в уравнение, линейность связи, отсутствие автокорреляции остатков, равенство их математических ожиданий нулю и постоянная дисперсия. Эмпирические данные не всегда обладают такими характеристиками, т.е. предпосылки МНК нарушаются. Применение этого метода в чистом виде может привести к таким нежелательным результатам, как смещение оцениваемых параметров, снижение их состоятельности, устойчивости, а в некоторых случаях может совсем не дать решения. Для смягчения нежелательных эффектов при построении регрессионных уравнений, повышения адекватности моделей существует ряд специальных методов, которые применяются в случае нарушения предпосылок применения МНК.

Тесты на гетероскедастичность

Существует несколько тестов на гетероскедастичность. Во всех этих тестах проверяется основная нулевая гипотеза о равенстве дисперсий H_o: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_n^2$ (наличие гомоскедастичности, отсутствие гетероскедастичности) против альтернативной гипотезы H_1: не H_o.

Рассмотрим один из самых распространенных тестов - *тест Голдфелда-Квандта*. При проведении проверки по этому критерию предполагается, что стандартное отклонение распределения вероятностей ε_i, пропорционально значению x в этом наблюдении. Предполагается также, что случайный член распределён нормально и не подвержен автокорреляции.

Все n наблюдений в выборке упорядочиваются по величине x, после чего оцениваются отдельные регрессии для первых n^1 и для последних n^2 наблюдений; средние $(n - n^1 - n^2)$ наблюдений отбрасываются. Если предположение относительно природы гетероскедастичности верно, то дисперсия ε в последних n^2 наблюдениях будет больше, чем в первых и это будет отражено в сумме квадратов остатков в двух указанных «частных» регрессиях. Обозначая суммы квадратов остатков в регрессиях для первых n^1 и последних n^2 наблюдений соответственно через $\mathbf{RSS_1}$ и $\mathbf{RSS_2}$ рассчитаем отношение $\mathbf{RSS_2}$/ $\mathbf{RSS_1}$, которое имеет F-распределение с $(n^2 - m - 1)$ и $(n^1 - m - 1)$ степенями свободы, где m - число объясняющих переменных в регрессионном уравнении. (В числителе должна быть наибольшая из сумм квадратов отклонений.) Мощность критерия зависит от выбора n' по отношению к n. Основываясь на результатах некоторых проведенных ими экспериментов, С.Голфелд и Р.Квандт утверждают, что n' должно составлять порядка 11, когда n = 30, и порядка 22, когда n = 60, Если в модели имеется более одной объясняющей переменной, то наблюдения должны упорядочиваться по той из них, которая, как предполагается, связана с σ_i.

Метод Голфелда - Квандта может также использоваться для проверки на гетероскедастичность при предположении, что σ_i, обратно пропорционально x_i. При этом используется та же процедура, что и описанная выше, но тестовой статистикой теперь является показатель $\mathbf{RSS_1/RSS_2}$, который вновь имеет F-распределение с $(n^1 – m – 1)$ и $(n^2 – m – 1)$ степенями свободы.

Рассмотрим пример выполнения теста в Excel. По 34 странам оценивалась регрессия расходов на образование от валового национального продукта (ВНП). На основе данных, приведенных на рис.33, с помощью функции ЛИНЕЙН были оценены регрессии сначала по наблюдениям 12 стран с наименьшим ВНП а затем для 12 стран с наибольшим ВНП. Сумма квадратов отклонений в первой регрессии равна 2,68, а во второй - 389,83.

Соотношение $\mathbf{RSS_2/RSS_1}$, следовательно, составило 145,27. Критическое значение F(10,10) равно 4,849 при однопроцентном уровне значимости, и нулевая гипотеза об отсутствии гетероскедастичности отклоняется.

G11 fx =H19/H7

	A	B	C	D	E	F	G	H
1								
2	Страна	Расход	ВНП	Страна	Расход	ВНП		
3	1	0,34	5,67	16	1,6	66,97	0,041351	0,082066
4	2	0,22	10,13	17	4,26	76,88	0,012292	0,327561
5	3	0,32	11,34	18	5,31	101,65	0,530905	0,518027
6	4	1,23	18,88	19	6,4	115,97	11,31764	10
7	5	1,81	20,94	20	7,15	119,49	3,037108	2,683517
8	6	1,02	22,16	21	11,22	124,15		
9	7	1,27	23,83	22	8,66	140,98		
10	8	1,07	24,67	23	5,56	153,85	F 2/1	Fкр 0,01
11	9	0,67	27,56	24	13,41	169,38	145,2667	4,849142
12	10	1,25	27,57	25	5,46	186,33		
13	11	0,75	40,15	26	4,79	211,78		
14	12	2,8	51,62	27	8,92	249,72		
15	13	4,9	57,71	28	18,9	261,41	0,071104	-8,19012
16	14	3,5	63,03	29	15,95	395,52	0,002744	2,450414
17	15	4,45	66,32	30	29,9	534,97	0,98533	6,243602
18	16	1,6	66,97	31	33,59	655,29	671,6619	10
19	17	4,26	76,88	32	38,62	816	26183,11	389,8256
20	18	5,31	101,65	33	61,61	1040,45		
21	19	6,4	115,97	34	181,3	2586,4		

Рис. 33. Тест на гетероскедастичность

Метод взвешенных наименьших квадратов

На первом этапе данного метода оценивается линейная регрессионная модель

$$\mathbf{Y} = \mathbf{BX} + \varepsilon$$

с помощью обычного МНК. Предполагается, что остатки e_i независимы между собой, но имеют разные дисперсии. Поскольку теоретические отклонения нельзя рассчитать, их обычно заменяют фактическими отклонениями зависимой переменной от линии регрессии ε_i. Предполагается, что ковариационная матрица вектора ошибок ε диагональна,

$$V(e_i) = \sigma_i^2,\ t = 1,\ldots,n.$$

Если величины σ_i^2 известны, то делением исходного регрессионного уравнения на σ_i, получаем (выписав каждое уравнение):

$$\frac{Y_i}{\sigma_i} = \sum_{J=1}^{k} b_i \frac{X_{ij}}{\sigma_i} + u_i, \qquad i = 1,\ldots,n,$$

где $u_i = \varepsilon_i / \sigma_i$, причем $V(u_i) = 1$, $Cov(u_i, u_s) = 0$ при $i \neq s$. Применяя к полученному уравнению стандартный метод наименьших квадратов, оценку получаем минимизацией по $b = (b_1, \ldots, b_k)$ суммы:

$$f(b) = \sum_{i=1}^{n} \left[\frac{1}{\sigma_i} \left(Y_i - \sum_{j=1}^{k} b_j X_{ij} \right) \right]^2 .$$

Содержательный смысл этого преобразования заключается в следующем. Используя обычный метод наименьших квадратов, мы минимизируем сумму квадратов отклонений, в которую, говоря не строго, разные слагаемые вносят разный статистический вклад из-за различных дисперсий, что в конечном итоге и приводит к неэффективности МНК-оценки. «Взвешивая» каждое наблюдение с помощью коэффициента $1/\sigma_i$, мы устраняем такую неоднородность. Применение метода взвешенных наименьших квадратов приводит к уменьшению дисперсий оценок по сравнению с обычным методом наименьших квадратов.

Таким образом, для учета гетероскедастичности в случае пропорциональности дисперсии одному или нескольким регрессорам можно использовать двухшаговую процедуру оценки. Такая двухшаговая процедура дает асимптотически несмещенные оценки стандартных ошибок коэффициентов регрессии.

Предполагается, что дисперсия ошибки есть линейная функция от нескольких регрессоров. Допустим, например:

$$\sigma_i^2 = a_0 + b_1 x_1 + b_2 x_2.$$

На первом шаге процедуры оценивается регрессионное уравнение модели:

$$y = a_0 + \Sigma b_i x_i + \varepsilon.$$

Устанавливаются остатки e_i, которые подставляются в первое уравнение вместо σ_i. По полученному уравнению находится состоятельная оценка вектора дисперсий σ_i^2.

На втором шаге полученные оценки $\boldsymbol{\sigma_i}^2$ используются в качестве весовых коэффициентов для взвешенного метода наименьших квадратов ($\boldsymbol{i}$-е уравнение делится на $\boldsymbol{\sigma_i}^2$).

Препятствием для этой процедуры является то, что на практике, как правило, неизвестны фактические значения $\boldsymbol{\sigma_i}$. Однако процедура будет применимой, если мы сможем подобрать некоторую величину, пропорциональную, по нашему мнению, $\boldsymbol{\sigma}$ в каждом наблюдении, и разделим на нее обе части уравнения.

Например, можно предположить, что $\boldsymbol{\sigma}$ приблизительно пропорциональна $\boldsymbol{x}$, как в критерии Голдфелда-Квандта. Если после этого мы разделим каждое наблюдение на соответствующее ему значение $\boldsymbol{x}$, то исходное уравнение примет вид:

$$\frac{y}{x} = a\frac{1}{x} + b + \frac{u}{x},$$

и при этом, возможно, новый случайный член $\boldsymbol{u/x}$ будет иметь постоянную дисперсию. Затем необходимо оценить регрессионную зависимость $\boldsymbol{y/x}$ от $\boldsymbol{1/x}$, включив в уравнение постоянный член. Коэффициент при $\boldsymbol{1/x}$ будет эффективной оценкой $\boldsymbol{a}$, постоянный член - эффективной оценкой $\boldsymbol{b}$. В предыдущем примере зависимой переменной будет доля расходов на образование в ВНП, а объясняющей переменной - обратная к ВНП величина. На рис. 34 приведены результаты расчетов для этого случая.

	A	B	C	D	E	F	G	H	I	J	K
1											
2	Страна	Расход	ВНП			Расх/ВНП	1/ВНП	-0,06567	0,053202		
3	1	0,34	5,67			0,0599647	0,176367	0,094413	0,004112	Вся регрессия	
4	2	0,22	10,13			0,0217177	0,098717	0,014893	0,019453		
5	3	0,32	11,34			0,0282187	0,088183	0,483778	32		
6	4	1,23	18,88			0,0651483	0,052966	0,000183	0,012109		
7	5	1,81	20,94			0,0864374	0,047755	-0,69554	12,93933	t	
8	6	1,02	22,16			0,0460289	0,045126				
9	7	1,27	23,83			0,0532942	0,041964	0,02968	0,043811		
10	8	1,07	24,67			0,0433725	0,040535	0,145349	0,010513		
11	9	0,67	27,56			0,0243106	0,036284	0,004153	0,021041	Первые 12	
12	10	1,25	27,57			0,0453391	0,036271	0,041698	10	наблюдений	
13	11	0,75	40,15			0,01868	0,024907	1,85E-05	0,004427		
14	12	2,8	51,62			0,0542425	0,019372				
15	13	4,9	57,71			0,0849073	0,017328				
16	14	3,5	63,03			0,0555291	0,015865	-2,62554	0,058441		
17	15	4,45	66,32			0,0670989	0,015078	2,586447	0,009839		
18	16	1,6	66,97			0,0238913	0,014932	0,093419	0,017935	Последние 12	
19	17	4,26	76,88			0,055411	0,013007	1,03046	10	наблюдений	
20	18	5,31	101,65			0,0522381	0,009838	0,000331	0,003217		
21	19	6,4	115,97			0,0551867	0,008623				
22	20	7,15	119,49			0,0598376	0,008369				
23	21	11,22	124,15			0,0903745	0,008055	1,376294		F-статистика	
24	22	8,66	140,98			0,0614272	0,007093		2,97824	Fкр	0,05
25	23	5,56	153,85			0,0361391	0,0065		4,849142	Fкр	0,01
26	24	13,41	169,38			0,0791711	0,005904				
27	25	5,46	186,33			0,0293028	0,005367				
28	26	4,79	211,78			0,0226178	0,004722				
29	27	8,92	249,72			0,03572	0,004004				
30	28	18,9	261,41			0,0723002	0,003825				
31	29	15,95	395,52			0,0403267	0,002528				
32	30	29,9	534,97			0,055891	0,001869				
33	31	33,59	655,29			0,0512597	0,001526				
34	32	38,62	816			0,0473284	0,001225				
35	33	61,61	1040,45			0,0592148	0,000961				
36	34	181,3	2586,4			0,0700974	0,000387				

Рис. 34. Коррекция на гетероскедастичность при ошибке пропорциональной независимой переменной

Как видно из рис. 34, $\boldsymbol{RSS_1}$ = 0,0044 больше, чем $\boldsymbol{RSS_2}$ = 0,0032; это показывает, что пересчет более чем компенсировал гетероскедастичность. Тестовая статистика в этом случае $\boldsymbol{RSS_1/RSS_2}$ = 1,376 невысокая и указывает на статистическую незначимость гетероскедастичности.

Иногда в нашем распоряжении может оказаться несколько переменных, каждую из которых можно использовать для масштабирования уравнения. В рассмотренном примере альтернативной переменной может быть численность населения страны ($\boldsymbol{H}$). Разделив обе части исходного уравнения на эту величину, получаем:

$$\frac{y}{H} = a\frac{1}{H} + b\frac{x}{H} + \frac{u}{H},$$

и надеемся на то, что случайный член $\boldsymbol{u/H}$, будет иметь постоянную дисперсию для всех наблюдений. Таким образом, теперь оценивается регрессионная зависимость государственных расходов на образование на душу населения от ВНП на душу населения и обратной величины от численности населения, причем на этот раз без постоянного члена. Статистика численности населения приведена на рис. 35.

	A	B	C	D
1				
2	Страна	Расход	ВНП	Н
3	1	0,34	5,67	0,36
4	2	0,22	10,13	2,9
5	3	0,32	11,34	2,39
6	4	1,23	18,88	3,44
7	5	1,81	20,94	3,87
8	6	1,02	22,16	10,71
9	7	1,27	23,83	3,1
10	8	1,07	24,67	9,93
11	9	0,67	27,56	5,07
12	10	1,25	27,57	11,1
13	11	0,75	40,15	9,6
14	12	2,8	51,62	4,78
15	13	4,9	57,71	4,09
16	14	3,5	63,03	22,34
17	15	4,45	66,32	5,12
18	16	1,6	66,97	44,92
19	17	4,26	76,88	7,51
20	18	5,31	101,65	6,37
21	19	6,4	115,97	8,37
22	20	7,15	119,49	9,86
23	21	11,22	124,15	8,31
24	22	8,66	140,98	14,62
25	23	5,56	153,85	27,06
26	24	13,41	169,38	14,14
27	25	5,46	186,33	67,4
28	26	4,79	211,78	37,43
29	27	8,92	249,72	123,03
30	28	18,9	261,41	23,94
31	29	15,95	395,52	57,04
32	30	29,9	534,97	55,95
33	31	33,59	655,29	53,71
34	32	38,62	816	61,56
35	33	61,61	1040,45	116,78
36	34	181,3	2586,4	227,64

Рис. 35. Данные по численности населения (Н)

Данные для новой регрессии необходимо упорядочить по переменной ***ВНП / Н***. Сортировку всего диапазона, начиная с номера страны, можно произвести с помощью функции *Данные, Сортировка*. Диалоговое окно этой опции приведено на рис. 36.

К отсортированным данным применена указанная схема расчетов, а результаты представлены на рис. 37. В функции ЛИНЕЙН в данном случае аргументу «константа» присвоить значение 0.

В данном случае $\boldsymbol{RSS_2 / RSS_1}$ равняется 4,6, что указывает на то, что нулевая гипотеза о гомоскедастичности должна быть отклонена при уровне значимости в 5% (критическое значение $\boldsymbol{F}$ составляет 2,978).

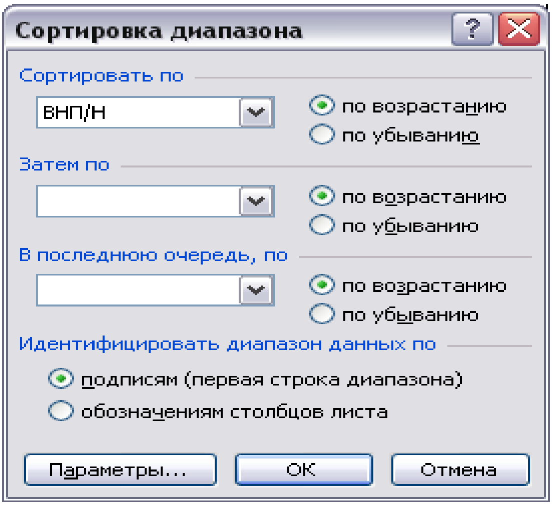


Рис. 36. Диалоговое окно Сортировка диапазона

Порядок выполнения работы

1. Получить у преподавателя данные для расчета.
2. Ввести исходные данные в таблицу Excel.
3. Провести на ЭВМ тесты на гетеростичность и расчеты с помощью взвешенного метода наименьших квадратов.
4. Провести анализ полученных результатов.
5. Зафиксировать результаты расчетов в тетради.

6. Сделать выводы по результатам моделирования и записать в тетради.

I22 f_x {=ЛИНЕЙН(E25:E36;F25:G36;0;1)}

	A	B	C	D	E	F	G	H	I	J
1										
2	Страна	Расход	ВНП	Н	Расх/Н	1/Н	ВНП/Н			
3	16	1,6	66,97	44,92	0,035619	0,022262	1,490873	0,062057	-0,02207	
4	27	8,92	249,72	123,03	0,072503	0,008128	2,029749	0,003083	0,057483	
5	6	1,02	22,16	10,71	0,095238	0,093371	2,069094	0,83404	0,148891	Вся регрессия
6	10	1,25	27,57	11,1	0,112613	0,09009	2,483784	80,40897	32	
7	8	1,07	24,67	9,93	0,107754	0,100705	2,484391	3,565102	0,709394	
8	25	5,46	186,33	67,4	0,081009	0,014837	2,76454	20,13113	-0,384	
9	14	3,5	63,03	22,34	0,15667	0,044763	2,821397			
10	2	0,22	10,13	2,9	0,075862	0,344828	3,493103			
11	11	0,75	40,15	9,6	0,078125	0,104167	4,182292	0,043645	-0,0634	
12	3	0,32	11,34	2,39	0,133891	0,41841	4,74477	0,015108	0,278492	
13	5	1,81	20,94	3,87	0,4677	0,258398	5,410853	0,351266	0,094276	Первые 12
14	9	0,67	27,56	5,07	0,13215	0,197239	5,435897	2,707324	10	наблюдений
15	4	1,23	18,88	3,44	0,357558	0,290698	5,488372	0,048125	0,088879	
16	26	4,79	211,78	37,43	0,127972	0,026717	5,658028			
17	23	5,56	153,85	27,06	0,205469	0,036955	5,685514			
18	29	15,95	395,52	57,04	0,279628	0,017532	6,934081			
19	7	1,27	23,83	3,1	0,409677	0,322581	7,687097	0,066369	-0,02878	
20	33	61,61	1040,45	116,78	0,527573	0,008563	8,909488	0,00489	0,080708	
21	30	29,9	534,97	55,95	0,534406	0,017873	9,561573	0,203807	0,202258	Последние 12
22	22	8,66	140,98	14,62	0,592339	0,068399	9,642955	1,279887	10	наблюдений
23	17	4,26	76,88	7,51	0,567244	0,133156	10,23702	0,104716	0,409081	
24	12	2,8	51,62	4,78	0,585774	0,209205	10,79916			
25	28	18,9	261,41	23,94	0,789474	0,041771	10,91938			
26	34	181,3	2586,4	227,64	0,796433	0,004393	11,3618		4,602671	F-стат
27	24	13,41	169,38	14,14	0,948373	0,070721	11,97878		2,97824	Fкр 0,05
28	20	7,15	119,49	9,86	0,725152	0,10142	12,11866		4,849142	Fкр 0,01
29	31	33,59	655,29	53,71	0,625396	0,018619	12,20052			
30	15	4,45	66,32	5,12	0,869141	0,195313	12,95313			
31	32	38,62	816	61,56	0,627355	0,016244	13,25536			
32	19	6,4	115,97	8,37	0,764636	0,119474	13,85544			
33	13	4,9	57,71	4,09	1,198044	0,244499	14,11002			
34	21	11,22	124,15	8,31	1,350181	0,120337	14,93983			
35	1	0,34	5,67	0,36	0,944444	2,777778	15,75			
36	18	5,31	101,65	6,37	0,833595	0,156986	15,95761			

Рис. 37. Коррекция на гетероскедастичность
с использованием новой переменной

Отчет по работе должен содержать

1. Название и цель работы.
2. Основные теоретические и методические положения.
3. Исходные данные для расчета.
4. Результаты расчета.
5. Выводы по результатам моделирования.

Лабораторная работа № 8.
Регрессионный анализ временных рядов

Цель работы: изучение автокорреляции во временных рядах и применения статистики Дарбина-Уотсона для ее обнаружения.

Исходные положения. Периодическая и сезонная зависимость (сезонность) представляет собой другой общий тип компонент временного ряда. Периодическая зависимость может быть формально определена как корреляционная зависимость порядка k между каждым i-м элементом ряда и i_k-м элементом. Её можно измерить с помощью автокорреляции (т.е. корреляции между самими членами ряда). Величину k обычно называют шагом (иногда используют другие термины: сдвиг, запаздывание). Если ошибка измерения не слишком большая, то сезонность можно определить визуально, рассматривая поведение членов ряда через каждые k временных единиц.

Автокорреляция данных временного ряда

Автокорреляция - это корреляционная связь между значениями одного и того же случайного процесса $X(t)$ в моменты времени t_1 и t_2. Функция, характеризующая эту связь, называется автокорреляционной. Корреляционная связь в этом случае измеряется с помощью коэффициента автокорреляции.

При анализе временных рядов автокорреляционная функция характеризует внутреннюю зависимость между значениями временного ряда и значениями того же самого ряда, но сдвинутыми на некоторый промежуток (лаг) времени. Иначе говоря, это корреляция членов ряда и передвинутых на L единиц времени членов того же ряда: x_1, x_2, x_3 ... и $x_{1+L}, x_{2+L}, x_{3+L}$. Запаздывание L называется лагом и представляет собой положительное целое число. Поскольку самое широкое распространение получили модели с лагом, равным одному году, то иногда автокорреляция определяется как корреляционная зависимость между значениями для одноименных месяцев года временного ряда.

Понятие автокорреляции иллюстрируется данными, представленными в табл. 5

Таблица 5

Данные об объемах продаж автомобилей

Время, t	Месяц	Исходные данные, Y_t	Y с запаздыванием на один период Y_{t-1}	Y с запаздыванием на два периода Y_{t-2}
1	Январь	7100		
2	Февраль	7200	7100	
3	Март	7350	7200	7100
4	Апрель	7400	7350	7200
5	Май	7455	7400	7350
6	Июнь	7470	7455	7400
7	Июль	7500	7470	7455
8	Август	7532	7500	7470
9	Сентябрь	7640	7532	7500
10	Октябрь	7667	7640	7532
11	Ноябрь	7700	7667	7640
12	Декабрь	7800	7700	7667

Автокорреляция бывает нескольких видов. Различные виды автокорреляции и взаимосвязь между ними представлены на рис. 38.

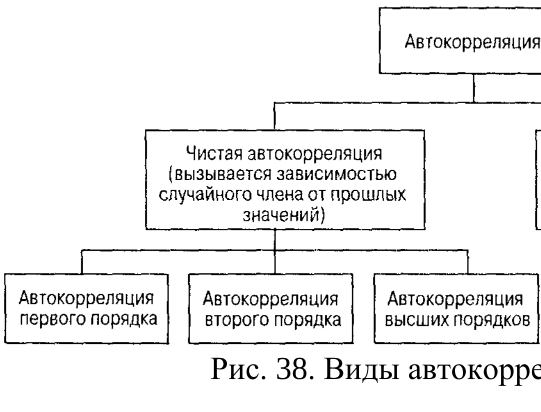


Рис. 38. Виды автокорреляции

Обычный тип автокорреляции, который можно еще назвать серийной корреляцией первого порядка, характеризуется тем, что слагаемое ошибки в текущий момент времени прямо связано со слагаемым ошибки в предыдущий момент времени. В этом случае, используя для обозначения времени индекс t, модель простой линейной регрессии можно записать в виде:

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \tag{1}$$

с условием

$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t \tag{2}$$

где ε_t – величина ошибки в момент t;
ρ – коэффициент автокорреляции с запаздыванием на один период, измеряющий корреляцию между последовательными слагаемыми ошибки;
ν_t – нормально распределенные независимые ошибки с математическим ожиданием **0** и дисперсией σ^2_ν.

Уравнение (2) показывает, что величина одного слагаемого ошибки ε_{t-1} непосредственно влияет на величину следующего ε_t. Значение коэффициента автокорреляции ρ, где -1 < ρ < 1, указывает на степень серийной корреляции. Если ρ равно 0, то серийной корреляции нет и значения ошибок независимы ($\varepsilon_t = \nu_t$).

Коэффициент автокорреляции может использоваться для того, чтобы определить, являются ли данные случайными, имеется ли тренд (нестационарность), являются ли данные стационарными, есть ли в них сезонные колебания.

Коэффициент автокорреляции r_k с запаздыванием на k моментов наблюдения, т.е. между наблюдениями Y_t и Y_{t-k} определяется по формуле:

$$r_k = \frac{\sum_{i=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{i=1}^{n}(Y_t - \bar{Y})^2},$$

где r_k – коэффициент автокорреляции для запаздывания на k периодов;
$\bar{Y}$ – среднее значение ряда;
Y_t – наблюдение в момент времени t;

Y_{t-k} – наблюдение в момент времени $t-1$.

По аналогии, коэффициент автокорреляции с запаздыванием на один период r_1 или корреляция между Y_t и Y_{t-1}, (корреляция первого порядка), вычисляется по формуле:

$$r_1 = \frac{\sum_{i=2}^{n}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{\sum_{i=1}^{n}(Y_t - \bar{Y})^2}.$$

Коэффициент автокорреляции с запаздыванием на два периода r_2 или корреляция между Y_t и Y_{t-2}, (корреляция второго порядка), вычисляется по формуле:

$$r_2 = \frac{\sum_{i=3}^{n}(Y_t - \bar{Y})(Y_{t-2} - \bar{Y})}{\sum_{i=1}^{n}(Y_t - \bar{Y})^2}.$$

В том случае, если ряд данных случаен, коэффициенты автокорреляции между Y_t и Y_{t-k} для любого запаздывания близки к нулю. Последовательные значения временного ряда не связаны друг с другом.

Если у ряда имеется тренд, то значения Y_t и Y_{t-1} имеют сильную корреляцию, причем коэффициенты автокорреляции существенно отличны от нуля для первых нескольких периодов запаздывания, а с увеличением периода постепенно убывают до нуля.

В том случае, если ряд имеет сезонную компоненту, значительный коэффициент автокорреляции будет наблюдаться для периодов запаздывания, равных сезонному периоду или кратных ему. Сезонный период запаздывания равен 4 для ежеквартальных данных и 12 для ежемесячных данных.

Коэффициент автокорреляции имеет выборочное распределение, которое может быть аппроксимировано нормальной кривой со средним значением, равным нулю, и среднеквадратическим отклонением $1/\sqrt{n}$. Однако некоторые пакеты прикладных программ используют несколько отличную формулу для вычисления стандартных ошибок корреляционной функции. В этой формуле предполагается, что любая автокорреляция для запаздывания, меньшего $\boldsymbol{k}$ (где $\boldsymbol{k} > \boldsymbol{1}$), отлична от нуля, а любая автокорреляция для запаздывания, большего или равного $\boldsymbol{k}$, равна нулю. Для автокорреляции, соответствующей запаздыванию в один период, используется стандартная ошибка $\mathbf{1}/\sqrt{\boldsymbol{n}}$.

$$\boldsymbol{SE(r_k)=\sqrt{\frac{1+2\sum_{i=1}^{k-1} r_i^2}{n}}},$$

где $SE(r_k)$ – стандартная ошибка автокорреляции с запаздыванием k; r_i – автокорреляция с запаздыванием i; k –время запаздывания; n – количество наблюдений во временном ряду.

Для того чтобы графически показать автокорреляционную функцию, используют коррелограмму (автокоррелограмму) представляющую значения коэффициентов автокорреляции и (и их стандартных ошибок) для последовательности лагов из определённого диапазона (например, от 1 до 30). На коррелограмме обычно отмечается диапазон в размере двух стандартных ошибок на каждом лаге, однако обычно величина автокорреляции более интересна, чем ее надежность, потому что интерес в основном представляют очень сильные (а следовательно, значимые) автокорреляции.

При изучении коррелограмм следует помнить, что автокорреляции последовательных лагов формально зависимы между собой. Рассмотрим следующий пример. Если первый член ряда тесно связан со вторым, а второй с третьим, то первый элемент должен также каким-то образом зависеть от третьего и т.д. Это приводит к тому, что периодическая зависимость может существенно измениться после удаления автокорреляций первого порядка, т.е. после взятия разности с лагом 1. Взятие разности также удаляет тренд, который обычно подавляет другие автокорреляции. Например, если имеется устойчивый линейный тренд, то каждое наблюдение в значительной степени является линейной функцией предыдущего наблюдения.

Тест Дарбина-Уотсона

Автокорреляция затрудняет применение ряда классических методов анализа временных рядов. В моделях регрессии, описывающих зависимости между случайными значениями взаимосвязанных величин, она снижает эффективность применения метода наименьших квадратов. Поэтому были выработаны и широко применяются специальные статистические приемы для ее выявления (например, критерий Дарбииа-Уотсона) и исключения (например, преобразование временного ряда в ряд значений разностей между его соседними членами), а также для модификации самого метода наименьших квадратов.

Статистика Дарбина-Уотсона предназначена для обнаружения автокорреляции первого порядка. Она основана на изучении остатков уравнения регрессии и определяет, можно ли считать равным нулю параметр $\boldsymbol{\rho}$, имеющийся в уравнении $\boldsymbol{\varepsilon_t = \rho\ \varepsilon_{t-1} + v_t}$. Для этого выбирается одна из двух гипотез:

$$\boldsymbol{H_0 : \rho = 0}$$
$$\boldsymbol{H_1 : \rho > 0.}$$

В качестве альтернативной гипотезы выбирается $\boldsymbol{\rho > 0}$, поскольку временные ряды, которые обычно рассматривают в экономике, чаще всего имеют положительную автокорреляцию.

В том случае, если регрессионная модель не свободна, остатки будут автокоррелирующими. Поэтому в критерии Дарбина-Уотсона выводы строятся на основании величин остатков, полученных при регрессионном анализе. Статистика Дарбина-Уотсона определяется следующим равенством:

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

где $\boldsymbol{e_i = Y_t - \widehat{Y}_t}$ – остаток для периода времени $\boldsymbol{t}$;

$\boldsymbol{e_{i-1} = Y_{t-1} - \widehat{Y}_{t-1}}$ – остаток для периода времени $\boldsymbol{t-1}$.

Критерий Дарбина-Уотсона изменяется в диапазоне $\mathbf{0<\mathit{DW}<4}$ (рис. 39). При отсутствии автокорреляции $\boldsymbol{DW = 2}$.

Поэтому, если выполняется указанное условие, можно сделать следующий вывод:

• $\boldsymbol{0 < DW < d_1}$, - в ряду есть положительная автокорреляция; ($\boldsymbol{d_1}$ - нижняя граница);

• $\boldsymbol{(4 - d_2) < DW < 4}$, - в ряду есть отрицательная автокорреляция; (d_2 - нижняя граница);

• $\boldsymbol{d_2 < DW < (4 - d_1)}$, - автокорреляция в ряду отсутствует;

• $\boldsymbol{d_1 < DW < d_2}$ или $\boldsymbol{(4 - d_2) < DW < (4 - d_1)}$, нужны дополнительные исследования.

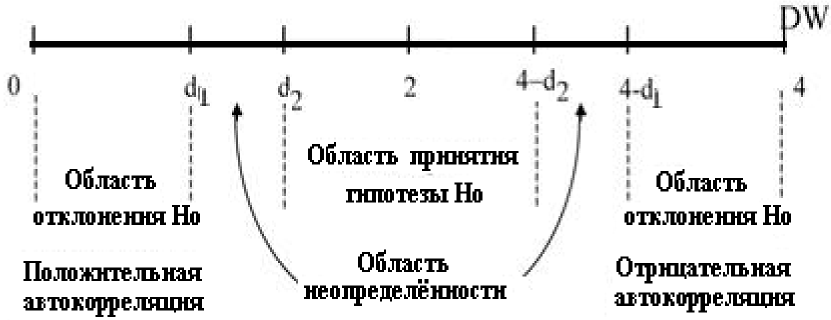


Рис. 39. Критические области статистики Дарбина-Уотсона

Упоминаемые выше граничные значения обычно приводятся в специальных таблицах, называемых “Граничные значения для статистик Дарбина-Уотсона”.

Пусть перед аналитиками стоит задача прогнозирования будущих объемов продаж для небольшой компании, которая реализует продукцию в различных регионах. Доходы этой компании в каждом регионе связаны с объемами продаж. В исходных данных содержатся объемы продаж и доходы компании за период с 2001 по 2015 год, как показано на рис.40, столбцы А – С.

	A	B	C	D	E	F	G
1	Данные объемов продаж и вычисление статистики Дарбина-Уотсона						
2	Год	Продажи `	Доход X	Остатки е	е i - е i-1	(е i - е i-1	(е i)^2
3	2001	1805	23400	48,41			2343,63
4	2002	1890	25600	-15,52	-63,93	4087,47	240,94
5	2003	1940	26700	-39,99	-24,47	598,62	1599,11
6	2004	1970	27300	-50,61	-10,62	112,75	2561,08
7	2005	2058	28400	-37,07	13,53	183,15	1374,47
8	2006	2178	29700	-5,08	31,99	1023,61	25,80
9	2007	2215	30560	-26,30	-21,22	450,26	691,65
10	2008	2376	31700	57,53	83,83	7026,71	3309,26
11	2009	2390	32157	40,59	-16,94	286,88	1647,44
12	2010	2410	32789	17,80	-22,78	519,13	316,99
13	2011	2457	33123	42,19	24,39	594,83	1780,28
14	2012	2498	33987	24,70	-17,49	305,91	610,25
15	2013	2675	36875	6,19	-18,51	342,58	38,37
16	2014	2690	37134	3,66	-2,53	6,42	13,40
17	2015	2745	38983	-66,51	-70,17	4924,07	4423,69
18					Сумма	20462,38	20976,38

Рис. 40. Доходы компании за период с 2001 по 2015 год

На рис. 40 также приведены расчетные данные, необходимые для вычисления статистики Дарбина-Уотсона. Уравнение регрессии и остатки получены с помощью функции *Регрессия* из пакета *Анализ данных*. Прежде чем использовать для прогноза прямую, полученную с помощью функции *Регрессия*, которая выглядит следующим образом:

$$Y = 172.48 + 0.068X,$$

следует применить к исходным данным критерий Дарбина-Уотсона для проверки наличия серийной корреляции. На рис. 40 представлены результаты следующих вычислений

$$e_t - e_{t-1} = -15,5 - 48,4 = -63,9$$

$$(e_t - e_{t-1})^2 = (-63,9)^2 = 4087,47$$

$$(e_t)^2 = (-15,5)^2 = 240,94$$

Статистика Дарбина-Уотсона вычисляется следующим образом:

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} = \frac{20462{,}38}{20976{,}38} = 0{,}975.$$

При уровне значимости 0,05 для выборки n = 15 и количества независимых переменных k = 1 найденные по таблице "Граничные значения для статистик Дарбина-Уотсона" значения нижней и верхней границ равны соответственно 1,08 и 1,36.

Если значение статистики DW находится между значениями верхней и нижней границ, необходимы дополнительные исследования, так как в этом случае критерий не дает ответа о наличии или отсутствии автокорреляции в исходных данных.

Поскольку статистика DW = 0,975 имеет значение меньшее нижней границы, имеется положительная автокорреляция остатков.

Таблица 6

Статистические таблицы критических уровней при уровне значимости 0,05

Количество наблюдений	Модель	
	Однопараметрическая (m = 1)	
	d_1	d_2
10	--	--
15	1,08	1,36
20	1,20	1,41
25	1,28	1,45
30	1,35	1,49

Решение проблемы автокорреляции

Если во временном ряду обнаружена автокорреляция данных, ее необходимо устранить или каким-либо образом учесть, прежде чем полученное уравнение регрессии можно будет использовать для прогноза. В этом случае начинать следует с оценки самого уравнения регрессии, чтобы получить ответы на следующие вопросы: правильна ли выбранная форма уравнения, не пропущена ли важная независимая переменная, имеются ли повторяющиеся явления, которые накладывают свой отпечаток на значения данных?

Есть несколько методов устранения автокорреляции. Один из них заключается в добавлении в уравнение регрессии дополнительной переменной, которая влияет на значение зависимой переменной в разные периоды времени.

Для того чтобы устранить серийную корреляцию сильно автокоррелирующих данных, можно также использовать в расчетах не сами значения ряда, а их разности. Иначе говоря, вместо определения уравнения регрессии относительно исходных переменных $\boldsymbol{Y}, \boldsymbol{X_1}, \boldsymbol{X_2},\ldots, \boldsymbol{X_k}$, это уравнение отыскивается для разностей $\boldsymbol{Y'_t = Y_t - Y_{t-1}}$ и $\boldsymbol{X'_{t1} = X_{t1} - X_{t-1,1}}$, $\boldsymbol{X'_{t2} = X_{t2} - X_{t-2,2}}$. Разности следует использовать, когда значение статистики $\boldsymbol{DW}$, вычисленное для исходных переменных, близко к $\mathbf{0}$.

Использование регрессионных моделей, построенных для обобщенных разностей в виде $\boldsymbol{Y'_t = Y_t - \rho Y_{t-1}}$ и $\boldsymbol{X'_t = X_t - \rho X_{t-1}}$, также позволяет устранить серийную корреляцию. Однако если серийная корреляция очень велика, целесообразно использовать обычные разности.

Для устранения влияния автокорреляции также может использоваться модель авторегрессии. Модель авторегрессии первого порядка записывается в виде уравнения $\boldsymbol{Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t}$, где предполагается, что ошибки $\boldsymbol{\varepsilon_t}$ удовлетворяют обычным предположениям регрессионной модели. Вычисляя параметры этой модели методом наименьших квадратов, получаем уравнение для прогнозирования: $\boldsymbol{\widehat{Y}_t = b_0 + b_1 Y_{t-1}}$. В модели авторегрессии прогнозируемые значения вычисляются как функция предыдущих значений временного ряда.

При другом методе устранения этого влияния используется логарифмирование и нахождение разностей. Исходные значения переменных логарифмируются, и используются разности прологарифмированных значений. В данном случае для прогнозирования используется следующее уравнение:

$$\boldsymbol{Ln\widehat{Y}_t = Ln\widehat{Y}_t + 1{,}01(LnX_t - LnX_{t\text{-}1})}.$$

Порядок выполнения работы

1. Получить у преподавателя данные для расчета.
2. Ввести исходные данные в таблицу Excel.
3. Провести на ЭВМ серию расчетов по определению параметров статистики Дарбина-Уотсона.
4. Провести анализ полученных результатов.
5. Зафиксировать результаты расчетов в тетради.
6. Сделать выводы по результатам моделирования и записать в тетради.

Отчет по работе должен содержать

1. Название и цель работы.
2. Основные теоретические и методические положения.
3. Исходные данные для расчета.
4. Результаты расчета.
5. Выводы по результатам моделирования.

Лабораторная работа № 9. Системы одновременных уравнений

Цель работы. Освоение двухшагового метода наименьших квадратов для поиска параметров системы эконометрических уравнений.

Основные понятия. При моделировании достаточно сложных экономических объектов часто приходится вводить не одно, а несколько связанных между собой уравнений. А значит, при проведении регрессионного анализа модели может возникнуть необходимость оценивать систему уравнений. Оценка систем уравнений требует введения новых понятий и методов.

Рассмотрим классический пример системы одновременных уравнений, который демонстрирует основные проблемы, возникающие при попытке оценить неизвестные параметры. Таким примером является исследование зависимости спроса и предложения некоторого товара от его цены и дохода:

$$\mathbf{S_t} = \boldsymbol{a}_1 + \boldsymbol{a}_2\mathbf{P_t} + \boldsymbol{\varepsilon}_t \text{ (предложение);}$$

$$\mathbf{D_t} = \boldsymbol{b}_1 + \boldsymbol{b}_2\mathbf{P_t} + \boldsymbol{b}_3\boldsymbol{Y}_t + \boldsymbol{u}_t \text{ (спрос);}$$

$$\mathbf{S_t} = \mathbf{D_t} \text{ (равновесие),}$$

где P_t - цена товара; Y_t - доход в момент времени t.

Записывая каждое уравнение для упрощения в отклонениях от средних значений получаем следующую систему:

$$s_t = a_2 p_t + \varepsilon_t\,;$$

$$d_t = b_2 p_t + b_3 y_t + u_t.$$

(*Уравнение в отклонениях*. Обозначим через $x_t = X_t - \bar{X},\ y_t = Y_t - \bar{Y}$ х, = Х, — , у, = - отклонения от средних по выборке значений $\bar{X} = \frac{1}{n}\sum X_t,\ \bar{Y} = \frac{1}{n}\sum Y_t.$ Х, и У,, Х = У = Решением задачи оценки параметров линейной функции $\boldsymbol{y = a + bx}$ будет та же прямая на плоскости (х, у), что и для исходных данных X_t, Y_t. Переход от X, Y к отклонениям x,y означает лишь перенос начала координат в точку $(\bar{X}, \bar{Y})$. Учитывая, что $\bar{x} = \bar{y} = \Sigma x_t = \Sigma y_t = 0$, получим:

$$\hat{a} = 0, \quad \hat{b} = \frac{\sum x_t y_t}{\sum x_t^2}.$$

Данная система уравнений называется структурной формой модели, соответственно, коэффициенты этих уравнений называются структурными коэффициентами. В соответствии с этой моделью цена и величина спроса-предложения определяются одновременно и поэтому эти переменные должны считаться *эндогенными*, а доход y_t, в отличие от них, - *экзогенной* переменной. Деление переменных на экзогенные и эндогенные определяется содержательной стороной модели. Предполагается, что в каждом уравнении экзогенные переменные не коррелированы с ошибкой. В то же время эндогенные переменные, стоящие в правых частях уравнений, как правило, имеют ненулевую корреляцию с ошибкой в соответствующем уравнении. В модели со стохастическими регрессорами наличие корреляции между регрессорами и ошибками приводит к смещённости и несостоятельности МНК-оценок.

Двухшаговый метод наименьших квадратов

Методы оценивания систем одновременных уравнений можно разделить на методы, позволяющие оценивать каждое из уравнений поочередно, и методы, предназначенные для оценивания всех уравнений сразу, т. е. всей модели в целом. Примерами первой группы методов служат двухшаговый МНК и метод ограниченной информации для одного уравнения, а примерами методов второй группы - трехшаговый метод наименьших квадратов и метод максимального правдоподобия полной информации.

В Excel нет встроенного двухшагового метода наименьших квадратов. Поэтому основные возможности следующие: последовательные вычисления с использованием функции ЛИНЕЙН, учитывая, что она выводит вектор-строку коэффициентов регрессии в обратном порядке, поэтому вектор-столбец коэффициентов при транспонировании также, к сожалению, выходит в обратном порядке. Следующая возможность - использовать матричные функции или комбинацию матричных функций и функции ЛИНЕЙН.

Последовательность действий рассмотрим на конкретном примере. Возьмем систему уравнений «спрос-предложение». Обозначим y_1 - спрос, y_2 - предложение, Y_1, Y_2 - цена, X_1, X_2 - доход, a_1, a_2, β_1, β_2, γ_1, γ_2 - искомые коэффициенты регрессий:

$$y_1 = a_1 + Y_1\,\beta_1 + X_1\,\gamma_1 + \varepsilon_1,$$
$$y_2 = a_2 + Y_2\,\beta_2 + X_2\,\gamma_2 + \varepsilon_2.$$

Введем исходные данные: y_1 - в A2:A11, Y_1 - в B2:B11, X_1 - в C2:C11, y_2 – в D2:D11, Y_2 - в E2:E11, X_2 - в F2:F11 (табл. 7).

Таблица 7

Данные для расчета

y_1	Y_1	X_1	y_2	Y_2	X_2
79	1	5	15	2	10
74	3	7	26	4	15
77	2	9	36	6	17
74	5	13	46	8	20
70	6	8	60	11	23
56	4	15	70	15	18
48	8	20	67	13	28
49	6	18	86	17	30
44	9	22	96	19	34
34	11	24	125	25	40

Определим для сравнения коэффициенты уравнений обычным МНК, т. е. применим функцию ЛИНЕЙН к обоим уравнениям (табл. 8).

Таблица 8

Результаты расчета

-1,8902	-0,954	92,398		0,66145	3,8559	0,885
0,5699	1,21135	4,028		0,13181	0,1703	1,4745
0,9189	5,22005	#Н/Д		0,99875	1,3525	#Н/Д
39,667	7	#Н/Д		2791,52	7	#Н/Д
2161,8	190,743	#Н/Д		10213,3	12,805	#Н/Д
t_X	t_Y	t_0		t_X	t_Y	t_0
-3,3167	-0,7875	22,93902		5,0183	22,643	0,60019

Таким образом, уравнения выглядят:

$$\boldsymbol{y_1} = 92{,}398 - 0{,}954\,\boldsymbol{Y_1} - 1{,}98\,\boldsymbol{X_1},$$

$$\boldsymbol{y_2} = 0{,}885 + 3{,}856\,\boldsymbol{Y_2} + 0{,}661\,\boldsymbol{X_2}.$$

Теперь перейдем к двухшаговому методу наименьших квадратов. Для удобства образуем матрицу $\boldsymbol{X}$: запишем в столбцы G и H векторы $\boldsymbol{X_1}$ и $\boldsymbol{X_2}$. Теперь определим коэффициенты приведенной формы для первого уравнения. Для этого применим функцию ЛИНЕЙН к $\boldsymbol{Y_1}$ и матрице $\boldsymbol{X} = [\boldsymbol{X_1}\boldsymbol{X_2}]$ и выведем результат в ячейки A13:B13 (рис. 41).

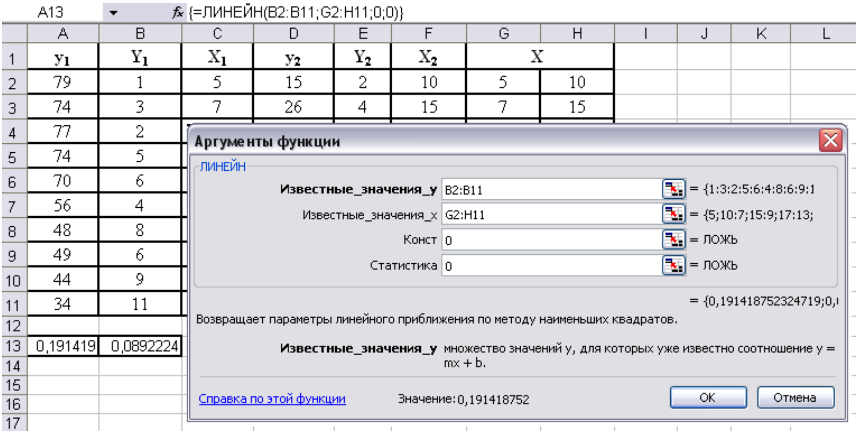


Рис. 41. Первое применение МНК в двухшаговой процедуре

Сформируем вектор коэффициентов приведенной формы, скопировав значение ячейки A13 в ячейку B14. Рассчитаем вектор прогнозных значений в столбец J2:J11, перемножив матрицу ***X*** на вектор коэффициентов приведенной формы:

= МУМНОЖ(G2:H11;B13:B14).

Теперь определим коэффициенты регрессии для первого уравнения между y_1 и матрицей $[\hat{\boldsymbol{Y}}_{11}\ \ \boldsymbol{X}_1]$, сформированной в ячейках J2:K11, функцией ЛИНЕЙН, введя ответ в ячейки F13:H17:

= ЛИНЕЙН(A2:A11;J2:K11,1,1).

Теперь проделаем аналогичные вычисления для второго уравнения. В ячейки D13:E13 введем коэффициенты приведенной формы:

= ЛИНЕЙН(E2:E11);G2:H11;0;0).

Сформируем вектор коэффициентов приведенной формы, вставив значение ячейки D13 в ячейку E14. Рассчитаем в L2:L11 вектор прогнозных значений $\hat{\boldsymbol{Y}}_{12}$:

=МУМНОЖ(G2:H11;E13:E14).

Сформируем матрицу $[\hat{\boldsymbol{Y}}_{12}\boldsymbol{X}_2]$ в ячейках L2:M11 и определим коэффициенты регрессии для второго уравнения между $\boldsymbol{y}_2$

и матрицей $[\hat{Y}_{12} X_2]$ функцией ЛИНЕЙН, введя ответ в ячейки I13:K17:

ЛИНЕЙН(D2:D11;L2:M11,1,1).

Решение приведено на рис. 42.

	G	H	I	J	K	L	M
1	X			$\hat{Y}_{11}$	X_1	$\hat{Y}_{12}$	X_2
2	5	10		2,3603	5	4,80324	10
3	7	15		3,4958	7	6,94204	15
4	9	17		4,0571	9	8,42833	17
5	13	20		4,9883	13	11,1834	20
6	8	23		5,1164	8	9,20772	23
7	15	18		4,7839	15	11,7997	18
8	20	28		7,1442	20	16,6029	28
9	18	30		7,3486	18	15,9866	30
10	22	34		8,4711	22	18,9592	34
11	24	40		9,7981	24	21,3155	40
12							
13	-1,536658	-2,276754	95,27273		2,0008	2,551019	-16,265
14	0,808824	2,321323	4,656124		1,45712	2,492113	9,909965
15	0,9224	5,106788	#Н/Д		0,91913	10,86915	#Н/Д
16	41,60285	7	#Н/Д		39,7802	7	#Н/Д
17	2169,945	182,555	#Н/Д		9399,13	826,9682	#Н/Д
18	t_X	t_Y	t_0		t_X	t_Y	t_0
19	-1,899867	-0,9808	20,46181		1,37312	1,023637	-1,64127

Рис. 42. Окончательный результат применения двухшагового МНК

Таким образом, система уравнений выглядит так:

$$y_1 = 95{,}272 - 2{,}277\, Y_1 - 1{,}537\, X_1,$$

$$y_2 = -16{,}265 + 2{,}551\, Y_2 + 2{,}001\, X_2.$$

Порядок выполнения работы

1. Получить у преподавателя данные для расчета.
2. Ввести исходные данные в таблицу Excel.
3. Провести на ЭВМ серию расчетов по определению параметров системы эконометрических уравнений.

4. Провести анализ полученных результатов.
5. Зафиксировать результаты расчетов в тетради.
6. Сделать выводы по результатам моделирования и записать в тетради.

Отчет по работе должен содержать

1. Название и цель работы.
2. Основные теоретические и методические положения.
3. Исходные данные для расчета.
4. Результаты расчета.
5. Выводы по результатам моделирования.

ЗАКЛЮЧЕНИЕ

В пособии даны методические материалы, предназначенные для освоения основ эконометрического моделирования и анализа. Большое внимание уделено классической парной и множественной регрессии, классическому, взвешенному, двухшаговому методам наименьших квадратов. Подробно рассмотрены проблемы, возникающие при построении многомерной регрессии: мультиколлинеарность, фиктивные переменные, гетероскедастичность модели. Приведен анализ временных рядов. Описаны методы одновременных уравнений. Изучение представленного материала предполагает широкое использование техники расчетов в среде электронной таблицы Excel.

ПРИЛОЖЕНИЕ

СТАТИСТИКО-МАТЕМАТИЧЕСКИЕ ТАБЛИЦЫ

1. Таблица значений F-критерия Фишера при уровне значимости α = 0,05

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48

k_2 \ k_1	1	2	3	4	5	6	8	12	24	∞
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

2. Критические значения t-критерия Стьюдента при уровне значимости 0,10, 0,05, 0,01(двухсторонний)

Число степеней свободы d.f.	α			Число степеней свободы d.f.	α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

3. Критические значения корреляции для уровневой значимости 0,05 и 0,01

d.f.	$\alpha = 0,05$	$\alpha = 0,01$	d.f.	$\alpha = 0,05$	$\alpha = 0,01$
1	0,996917	0,9998766	17	0,4555	0,5751
2	0,95000	0,99000	18	0,4438	0,5614
3	0,8783	0,95873	19	0,4329	0,5487
4	0,8114	0,91720	20	0,4227	0,5368
5	0,7545	0,8745	25	0,3809	0,4869
6	0,7067	0,8343	30	0,3494	0,4487
7	0,6664	0,7977	35	0,3246	0,4182
8	0,6319	0,7646	40	0,3044	0,3932
9	0,6021	0,7348	45	0,2875	0,3721
10	0,5760	0,7079	50	0,2732	0,3541
11	0,5529	0,6835	60	0,2500	0,3248
12	0,5324	0,6614	70	0,2319	0,3017
13	0,5139	0,6411	80	0,2172	0,2830
14	0,4973	0,6226	90	0,2050	0,2673
15	0,4821	0,6055	100	0,1946	0,2540
16	0,4683	0,5897			

для простой корреляции iГ. на 2 меньше, чем число пар вариантов; в случае частной корреляции необходимо также вычесть число исключаемых переменных.

4. Значения статистик Дарбина - Уотсона d_L d_U при 5%-ном уровне значимости

n	$k^1=1$		$k^1=2$		$k^1=3$		$k^1=4$		$k^1=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40	–	–	–	–				
7	0,70	1,36	0,47	1,90	–	–				
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Айвазян С.А. Эконометрика [Текст]: учеб. пособие / С.А. Айвазян, С.С. Иванова.. – М.: Маркет ДС, 2010. – 104 с.

2. Артамонов Н.В. Введение в эконометрику [Текст]: учеб. пособие / Н.В. Артамонов. – М.: Изд-во МЦНМО, 2011. – 204 с.

3. Амелин С.В. Эконометрика: [Текст]: учеб. пособие / С.В. Амелин. - Воронеж: ГОУВПО «Воронежский государственный технический университет», 2009. – 96 с.

4. Амелин С.В. Эконометрика: практикум: [Текст]: учеб. пособие / С.В. Амелин. - Воронеж: ГОУВПО «Воронежский государственный технический университет», 2007. – 120 с.

5. Методические указания по выполнению самостоятельной работы и индивидуальных заданий по дисциплине "Эконометрика" для студентов, обучающихся по направлениям 080100 «Экономика», 080200 «Менеджмент», 222000 «Инноватика» очной формы обучения [Текст] / ФГБОУ ВПО "Воронежский государственный технический университет"; сост. С.В. Амелин. - Воронеж, 2011. - 46 с.

6. Баклушина О.А. Эконометрика. Краткий курс [Текст]: учеб. пособие / О.А. Баклушина. – М.: Окей-книга, 2010. – 128 с

7. Белько И.В. Эконометрика. Практикум [Текст]: учеб. пособие / И.В. Белько, Е.А. Криштапович. – М.: Издательство Гревцова, 2011. – 224 с.

8. Бывшев В.А. Эконометрика [Текст]: учеб. пособие / В.А. Бывшев. – М.: Финансы и статистика, 2008. – 480 с.

9. Гладилин А.В. Эконометрика [Текст]: учеб. пособие / А.В. Гладилин, А.Н. Герасимов, Е.И. Громов. – Ростов-на-Дону: Феникс, 2011. – 304 с.

10. Гладилин А.В. Практикум по эконометрике [Текст]: учеб. пособие / А.В. Гладилин, А.Н. Герасимов, Е.И. Громов. – Ростов-на-Дону: Феникс, 2011. – 336 с.

11. Дуброва Т.А. Прогнозирование социально–экономических процессов. Статистические методы и модели: [Текст]: учеб. пособие / Т.А. Дуброва. – М.: Маркет ДС, 2007. – 192 с

12. Елисеева И.И. Практикум по эконометрике: [Текст]: учеб. пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко; под ред. И.И. Елисеевой. – 2–е изд., перераб. и доп. – М.: Финансы и статистика, 2007. – 344 с

13. Эконометрика: [Текст]: учебник / под ред. И.И. Елисеевой. – М.: Проспект, 2009. – 288 с.

14. Каплан В.Е. Статистическая обработка и анализ экономических данных [Текст]: учеб. пособие / В.Е. Каплан, А.В. Каплан, Е.В. Овечкина, М.В. Мащенко. – Ростов-на-Дону: Феникс, 2007. – 336 с.

15. Кочетыгов А.А. Основы эконометрики [Текст]: учеб. пособие / А.А. Кочетыгов, Л.А. Толоконникова. – М.: ИКЦ "МарТ", 2007. – 352 с.

16. Кремер Н.Ш. Эконометрика [Текст]: учеб. пособие / Н.Ш. Кремер, Б.А. Путко.. – М.: Юнити-Дана, 2010. – 328 с.

17. Куфель Т. Эконометрика. Решение задач с применением пакета программ GRETL [Текст] / Т. Куфель. – М.: Горячая Линия - Телеком, 2007. – 200 с.

18. Магнус Я.Р. Эконометрика. Начальный курс [Текст]: учебник / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий.. – М.: Дело, 2007. – 504 с.

19. Мхитарян В.С. Эконометрика [Текст]: учеб. пособие / В.С. Мхитарян, М.Ю. Архипова, В.А. Балаш, О.С. Балаш, Т.А. Дуброва, В.С. Сиротин. – М.: Проспект, 2010. – 384 с.

20. В.П. Носко. Эконометрика [Текст]: Книга 1. Части 1 и 2. / – М.: Издательский дом "Дело" РАНХиГС, 2011. – 672 с.

21. Носко В.П. Эконометрика. [Текст]: Книга 2. Части 3 и 4./ В.П. Носко. – М.: Издательский дом "Дело" РАНХиГС, 2011. – 576 с.

22. Орлов А.И. Эконометрика [Текст]: учебник / А.И. Орлов. – Ростов-на-Дону: Феникс, 2009. – 576 с.

23. Просветов Г.И. Эконометрика. Задачи и решения [Текст]: учеб. пособие / Г.И. Просветов. – М.: Альфа-Пресс, 2008. – 192 с.

24. Палий И.А. Прикладная статистика [Текст]: учеб. пособие / И.А. Палий. – М.: Издательско–торговая корпорация "Дашков и К", 2008. – 224 с

25. Симчера В.М. Методы многомерного анализа статистических данных [Текст]: учеб. пособие / В.М. Симчера. – М.: Финансы и статистика, 2008. – 400 с.

26. Тихомиров Н.П. Методы эконометрики и многомерного статистического анализа [Текст]: учебник / Н.П. Тихомиров, Т.М. Тихомирова, О.С. Ушмаев. – М.: Экономика, 2011. – 640 с.

27. Эконометрика [Текст]: учебник / Под редакцией В.Б. Уткина. – М.: Дашков и Ко, 2011. – 564 с.

28. Цымбаленко Т.Т. Методы математической статистики в обработке экономической информации [Текст]: учеб. пособие / Т.Т. Цымбаленко, А.Н. Баудаков, О.С. Цымбаленко; под ред. проф. Т.Т. Цымбаленко. – М.: Финансы и статистика; Ставрополь: АРГУС, 2007. – 200 с

29. Чураков Е.П. Прогнозирование эконометрических временных рядов [Текст]: учеб. пособие / Е.П. Чураков. – М.: Финансы и статистика, 2008. – 208 с

30. Яновский Л.П. Введение в эконометрику [Текст]: учеб. пособие / Л.П. Яновский, А.Г. Буховец. – М.: КноРус, 2010. – 256 с.

ОГЛАВЛЕНИЕ

Учебное издание

Амелин Станислав Витальевич

ЭКОНОМЕТРИКА: ПРАКТИКУМ

В авторской редакции

Подписано в печать 29.11.2016.
Формат 60х84/16. Бумага для множительных аппаратов.
Усл. печ. л. 7,0. Уч.-изд. л. 5,1. Тираж 200 экз.
Заказ № .

ФГБОУ ВО «Воронежский государственный технический университет»

394026 Воронеж, Московский просп., 14