ПРАКТИКУМ

Глава 1. Основы эконометрики

Тема 1.1 Предмет эконометрики

Развитие эконометрики в работах нобелевских лауреатов по экономике

Нобелевская премия по экономике была учреждена только в 1969 году, однако ей отмечена уже целая плеяда талантливых ученых, сделавших значительный вклад в развитие экономической теории. Так или иначе, работа этих экономистов была бы не возможна без создания и развития эконометрики, науки использующей статистико-математические методы и модели для количественного и качественного анализа экономических данных. Стоит отметить, что эконометрика обязана своим появлением как отдельной науки ученым, первыми получившим Нобелевскую премию по экономике в 1969 году. Рагнар Фриш и Ян Тинберген стали пионерами в создании эконометрической методологии и инструментария, они разработали и применили на практике первые динамические модели для анализа экономических процессов. Их работы стали толчком к дальнейшему развитию и совершенствованию эконометрики как самостоятельной науки, помогающей экономистам в их исследованиях. Достижения в области эконометрики были отмечены Нобелевскими премиями также и в 80-е годы. Лоуренс Клейн был награжден в 1980 году «за создание экономических моделей и их применение к анализу колебаний экономики и экономической политики». Ему удалось создать образец эконометрических макромоделей, используя которые можно было строить сравнительно реалистичные прогнозы дальнейшего экономического развития страны. Значительный вклад в развитие эконометрики сделал Трюгве Хаавельмо, получивший премию в 1989 году «за прояснение вероятностных основ эконометрики и анализ одновременных экономических структур». В своих трудах он стремился соединить методы статистики и теории вероятности для создания максимально адекватных моделей, также пытался решить проблему взаимозависимости переменных в экономических соотношениях. Работа Хаавельмо была продолжена Клайвом Грэнджером, разработавшим концепцию коинтегрированных временных рядов, позволяющую решить проблему кажущихся связей переменных. Вместе с Робертом Энглом он был удостоен Нобелевской премии в 2003 году за разработку новых методов анализа временных рядов. Тремя годами ранее премией были отмечены также достижения в области эконометрики. Джеймс Хекман и Дэниел Макфадден сделали значительный вклад в развитие теории и методов анализа дискретного выбора. Хекман исследовал методологическую проблему формирования статистической выборки, а Макфадден разработал способы моделирования дискретного выбора. Спустя десятилетие, в 2013 году, Нобелевскую премию «за эмпирический анализ цен на активы» получили сразу три широко известных ученых: Юджин Фама, Ларс Питер Хансен и 3 Роберта Шиллера. Обобщенный метод моментов Ларса Питера Хансена вывел анализ моделей ценообразования на качественно новый уровень, благодаря использованию спецификации рассматриваемых факторов. В следующем году Нобелевской премии по экономике был удостоен французский экономист Жан Тироль. Ученый детально проанализировал разного рода рыночные ситуации, уделив особое внимание «Регулированию доминирующих фирм», где представлен рынок монополии, на котором происходит процесс осуществления перевода средств со стороны государства в пользу фирмы-монополиста.

Глава 2. Основы статистического анализа.

Тема 2.1 Генеральная совокупность и выборка

Методические указания к решению задач

При анализе какого-то конкретного показателя X за фиксированный момент времени (либо без учета фактора времени) наблюдаемые значения x_1, $x_2, \ldots, x_n$ обычно упорядочивают по неубыванию: $x_1 \le x_2 \le \ldots \le x_n$. Разность между максимальным и минимальным значениями СВ X называется *размахом выборки*. Пусть количество различных значений в выборке равно k ($k \le n$). Для определённости положим $x_1 < x_2 < \ldots < x_k$.

Значения x_i, $i = 1, 2, \ldots, k$ называются *вариантами*.

Пусть значение x_i встретилось в выборке n_i раз, тогда число n_i называется *частотой* значения x_i, а $\omega_i = n_i/n$ – *относительной частотой* значения x_i. Тогда наблюдаемые значения можно сгруппировать в *статистический ряд*, показанный в табл.:

X	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k
$\omega_i = n_i/n$	n_1/n	n_2/n	…	n_k/n

$$\sum_{i=1}^{n} n_i = n$$

$$\sum_{i=1}^{n} \frac{n_i}{n} = 1$$

По статистическому ряду можно построить *эмпирическую функцию распределения* $F^*(x)$:

$$F^* = \frac{n_x}{n}$$

где n_x – число значений случайной величины X меньших, чем x; n – объем выборки.

Для любой СВ X кроме определения ее функции распределения желательно указать ее числовые характеристики, важнейшими из которых являются математическое ожидание, дисперсия, среднее квадратическое отклонение. Пусть объем генеральной совокупности равен N. Тогда математическим ожиданием СВ X является *генеральное среднее*:

$$\overline{x_{\Gamma}} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Дисперсией СВ X является *генеральная дисперсия*:

$$D_{\Gamma} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x_{\Gamma}})^2$$

Корень квадратный из генеральной дисперсии называется *генеральным средним квадратическим отклонением*:

$$\sigma_{\Gamma} = \sqrt{D_{\Gamma}}$$

Таким образом, для нахождения генеральных числовых характеристик необходим анализ всей генеральной совокупности. В силу того, что в реальности практически всегда имеют дело с выборками, приходится находить оценки указанных выше генеральных характеристик – выборочные числовые характеристики: выборочное среднее, выборочную дисперсию, выборочное среднее квадратическое отклонение.

Выборочное среднее – это среднее арифметическое наблюдаемых значений выборки.

$$\overline{x_{\text{В}}} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

При задании выборки в виде статистического ряда $\overline{x_{\text{В}}}$ рассчитывается по следующей формуле:

$$\overline{x_{\text{В}}} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i$$

Оценкой генеральной дисперсии является *выборочная дисперсия*:

$$D_{\text{В}} = \frac{1}{n}\sum_{i=1}^{n} n_i(x_i - \overline{x_{\text{В}}})^2$$

Зачастую для вычисления $D_{\text{В}}$ применяется следующая формула:

$$D_{\text{В}} = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2x_i\overline{x_{\text{В}}} + (\overline{x_{\text{В}}})^2) = \overline{x^2} - \bar{x}^2$$

При задании выборки в виде статистического ряда имеем:

$$D_{\text{В}} = \frac{1}{n}\sum_{i=1}^{n} n_i(x_i - \overline{x_{\text{В}}})^2$$

Корень квадратный из выборочной дисперсии называется *выборочным средним квадратическим отклонением*:

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} n_i(x_i - \overline{x_{\text{В}}})^2} = \sqrt{\overline{x^2} - \bar{x}^2}$$

При задании выборки в виде интервального статистического ряда вместо xi рассматривается среднее значение i-го подынтервала: $x_i = \frac{x_{i-1}+x_i}{2}$.

<u>Задача 1.</u>

Из большой группы предприятий одной из отраслей промышленности случайным образом отобрано 30, по которым получены показатели основных фондов в млн. руб.: 2; 3; 2; 4; 5; 2; 3; 3; 6; 4; 5; 4; 6; 5; 3; 4; 2; 4; 3; 3; 5; 4; 6; 4; 5; 3; 4; 3; 2; 4.

1) Составить дискретное статистическое распределение выборки.
2) Найти объем выборки.
3) Составить распределение относительных частот.
4) Построить полигон частот.
5) Составить эмпирическую функцию распределения и построить ее график.
6) Найти выборочную среднюю, выборочную дисперсию и среднее квадратическое отклонение.

<u>Решение:</u>

1. Расположим различные значения признака в порядке их возрастания и

под каждым из них запишем их частоты. Получим дискретное статистическое распределение выборки:

xi	2	3	4	5	6
ni	5	8	9	5	3

где xi - варианты, ni - частоты вариант xi.

2. Сумма частот всех вариант должна быть равной объему выборки. В данном примере объем выборки равен: n=5+8+9+5+3=30.

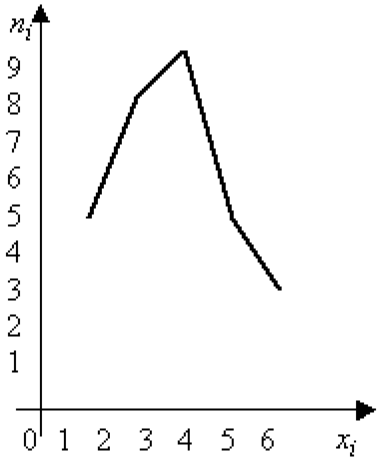
3. Найдем относительные частоты:

$$W = \frac{n_i}{\sum n_i}$$

Запишем искомое распределение относительных частот:

xi	2	3	4	5	6
Wi	1/6	4/15	3/10	1/6	1/10

4. Строим точки с координатами (xi, ni) и соединяем их последовательно отрезками. Полученная ломаная линия называется полигоном частот:



5. Согласно определению эмпирической функцией распределения называется функция вида $F(x) = \frac{n_x}{n}$, где n – объем выборки; n_x- сумма частот вариант, меньших x.

Эмпирическая функция является оценкой функции распределения генеральной совокупности.

Наименьшая варианта равна 2, поэтому при $x \le 2$, $n_x=0$ и $F(x)=0$.

Значение $X<3$, а именно, $X=x1=2$ наблюдалось 5 раз. Тогда для $2<x\le 3$ $n_x=5$ и $F(x)=5/30$.

Значение $X< 4$, а именно, $X=2$, $X=3$, наблюдалось $5 + 8 =13$ раз. Поэтому для $3<x\le 4$ $n_x=13$ и $F(x)=13/30$.

Аналогично рассуждая, получаем:

для $4<x\le 5$ $n_x=5+8+9=22$ и $F(x)=22/30$,

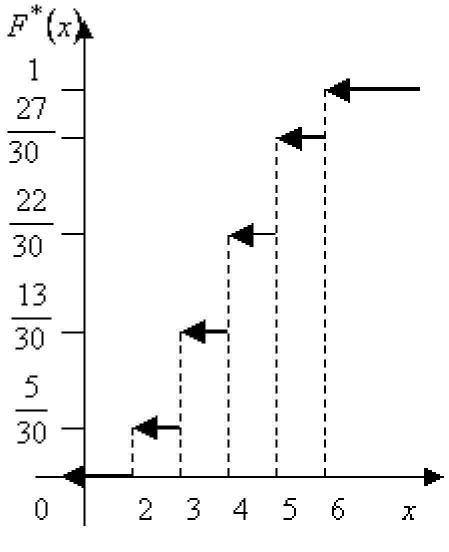
для $5<x\le 6$ $n_x=5+8+9+5=27$ и $F(x)=27/30$,

для $x>6$ $n_x=5+8+9+5+3=30$ и $F(x)=30/30=1$.

Таким образом,

$$F(x) = \begin{cases} 0 \text{ при } x \le 2 \\ \dfrac{5}{30} \text{ при } 2 < x \le 3 \\ \dfrac{13}{30} \text{ при } 3 < x \le 4 \\ \dfrac{22}{30} \text{ при } 4 < x \le 5 \\ \dfrac{27}{30} \text{ при } 5 < x \le 6 \\ 1 \text{ при } x > 6 \end{cases}$$

График эмпирической функции имеет вид:



6. Средняя выборочная:

$$\overline{x_в} = \frac{\sum x_i n_i}{n} = \frac{2*5+3*8+4*9+5*5+6*3}{30} = \frac{113}{30} = 3{,}77$$

Выборочная дисперсия:

$$D_{\text{В}} = \frac{1}{30}\sum_{i=1}^{5} n_i(x_i - \ \overline{x_{\text{В}}})^2 = M(X^2) - [M(X)]^2$$

$$= \sum x_i^2 p_i - x_{\text{В}}^2 = 2^2\frac{5}{30} + 3^2\frac{8}{30} + 4^2\frac{9}{30} + 5^2\frac{5}{30} + 6^2\frac{3}{30} - (3{,}77)^2$$

$$= 1{,}42$$

$$\sigma_{\text{В}} = \sqrt{D_{\text{В}}} = \sqrt{1{,}45} = 1{,}2$$

Ответ: 3,77; 1,42; 1,2.

Глава 3. Корреляционный анализ

Тема 3.1 Анализ количественной зависимости факторов

Методические указания у решению задач

Тесноту связи между двумя величинами можно определить при помощи ***коэффициента корреляции***, который определяется по формуле:

$$r = \frac{N\sum xy \ - \sum x \sum y}{\sqrt{N\sum x^2 - (\sum x)^2} * \sqrt{N\sum y^2 - (\sum y)^2}}$$

где x и y—текущие значения наблюдаемых величин; N — число наблюдений.

Существует несколько модификаций данной формулы, наиболее простая из которых имеет вид:

$$r = \frac{\overline{x\,y} - \overline{x} * \overline{y}}{\sigma_x * \sigma_y}$$

где $\overline{xy}$ —среднее значение произведения двух коррелируемых величин;

$\overline{x}$ и $\overline{y}$—средние значения этих величин;

σ_x, σ_y—среднеквадратичные (стандартные) отклонения соответствующих величин.

Отметим основные *свойства коэффициента корреляции:*

1) Коэффициент корреляции принимает значения на отрезке [-1,1], т.е

$$-1 \leq r \leq 1$$

В зависимости от того, насколько $|r|$ приближается к 1, различают связь слабую, умеренную, заметную, достаточно тесную, тесную и весьма тесную.

2) Если все значения переменных увеличить (уменьшить) на одно и то же число или в одно и то же число раз, то величина коэффициента корреляции не изменится.

3) При r = ± 1 корреляционная связь представляет линейную функциональную зависимость. При этом линии регрессии У по X и X по Y совпадают и все наблюдаемые значения переменных располагаются на общей прямой.

4) При r = 0 линейная корреляционная связь отсутствует. При этом

групповые средние переменных совпадают с их общими средними, а линии регрессии Y по X и X по Y параллельны осям координат.

Задача 1.

На основе имеющихся данных за 10 месяцев необходимо выявить есть ли зависимость между 2 экономическими показателями (x и y), если есть - то описать вид зависимости (линейная, криволинейная, положительная, отрицательная), определить коэффициент корреляции и изобразить зависимость графически.

	X	Y
1	152	413
2	135	615
3	140	536
4	164	330
5	133	466
6	154	513
7	155	643
8	164	616
9	138	420
10	172	585

Указание

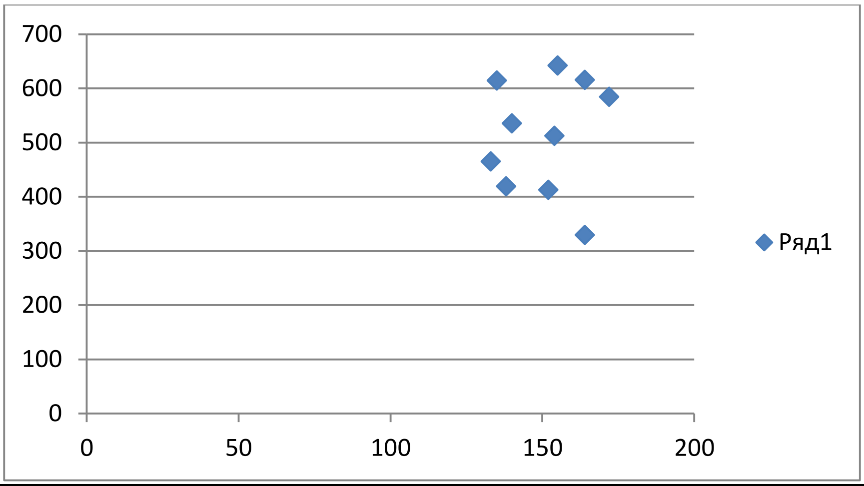
Расчет дисперсии можно провести по формулам:

$$D(x) = \sigma_x^2 = \overline{x^2} - \bar{x}^2$$

$$D(y) = \sigma_y^2 = \overline{y^2} - \bar{y}^2$$

Решение:

Построим исходные значения показателей на графике.



Визуальный анализ корреляционного поля показывает, что зависимости между x и y нет, и коэффициент корреляции скорее всего близок к нулю. Докажем это предположение с помощью расчетов.

Составим вспомогательную таблицу:

	x	**y**	**x*y**	**x*x**	**y*y**
1	152	413	62 776	23 104	170 569
2	135	615	83 025	18 225	378 225
3	140	536	75 040	19 600	287 296
4	164	330	54 120	26 896	108 900
5	133	466	61 978	17 689	217 156
6	154	513	79 002	23 716	263 169
7	155	643	99 665	24 025	413 449
8	164	616	101 024	26 896	379 456
9	138	420	57 960	19 044	176 400
10	172	585	100 620	29 584	342 225
S	**1 507**	**5 137**	**775 210**	**228 779**	**2 736 845**

$$r = \frac{N\sum xy - \sum x \sum y}{\sqrt{N\sum x^2 - (\sum x)^2} * \sqrt{N\sum y^2 - (\sum y)^2}}$$
$$= \frac{10 * 775210 - 1507 * 5137}{\sqrt{10 * 228779 - (1507)^2} * \sqrt{10 * 2736845 - (5137)^2}}$$
$$= \frac{10641}{\sqrt{16741} * \sqrt{979681}} = \frac{10641}{129{,}39 * 989{,}79} \approx 0{,}083$$

Проведем расчет по другой модификации данной формулы:

$$r = \frac{\overline{x\,y} - \overline{x} * \overline{y}}{\sigma_x * \sigma_y}$$

Для этого также составим вспомогательную таблицу:

	x	y	x*y	x*x	y*y
1	152	413	62 776	23 104	170 569
2	135	615	83 025	18 225	378 225
3	140	536	75 040	19 600	287 296
4	164	330	54 120	26 896	108 900
5	133	466	61 978	17 689	217 156
6	154	513	79 002	23 716	263 169
7	155	643	99 665	24 025	413 449
8	164	616	101 024	26 896	379 456
9	138	420	57 960	19 044	176 400
10	172	585	100 620	29 584	342 225
S	**1 507**	**5 137**	**775 210**	**228 779**	**2 736 845**
Среднее значение	**150,7**	**513,7**	**77 521**	**22 877,9**	**273 684,5**
σ^2	**167,41**	**9 796,8**			
σ	**12,94**	**98,98**			

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 = 22877{,}9 - 150{,}7^2 = 167{,}41$$

$$\sigma_x = \sqrt{167{,}41} = 12{,}94$$

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 = 273684{,}5 - 513{,}7^2 = 9796{,}8$$

$$\sigma_y = \sqrt{9796.8} = 98{,}98$$

$$r = \frac{\overline{x\,y} - \overline{x} * \overline{y}}{\sigma_x * \sigma_y} = \frac{77521 - 150{,}7 * 513{,}7}{12{,}94 * 98{,}98} = \frac{106{,}41}{1280{,}8} \approx 0{,}083$$

Ответ: линейной зависимости нет, r = 0,083

Задача 2

На основе имеющихся данных за 10 месяцев необходимо выявить есть ли зависимость между 2 экономическими показателями (x и y), если есть - то описать вид зависимости (линейная, криволинейная, положительная, отрицательная), определить коэффициент корреляции и изобразить зависимость графически.

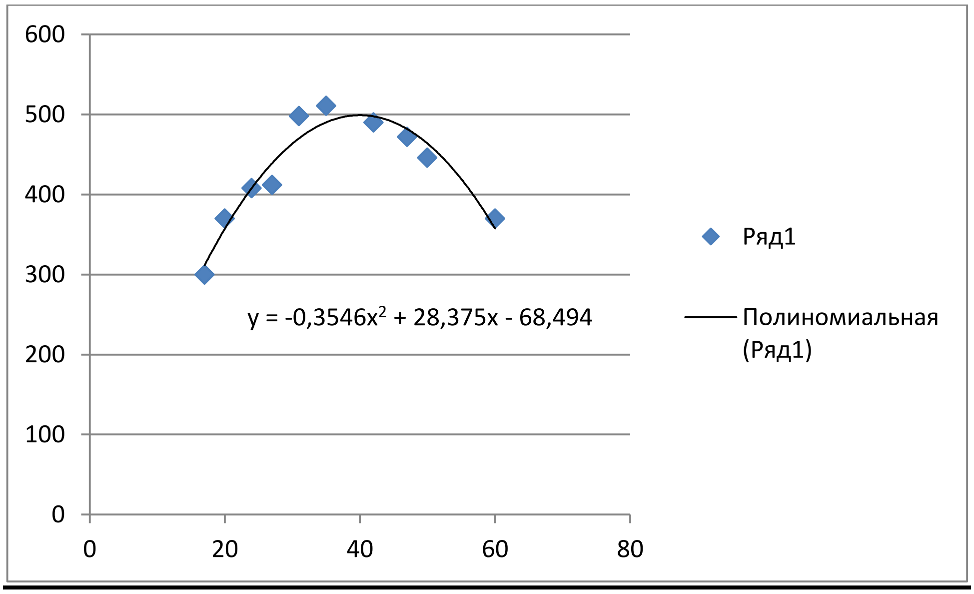
	x	y
1	35	511
2	17	300
3	50	446
4	31	498
5	42	490
6	47	472
7	60	370
8	24	408
9	27	412
10	20	370

Указание

Прежде чем приступить к расчетам, необходимо провести визуальный анализ графика корреляционного поля.

Решение:

Построим исходные значения показателей на графике.



Визуальный анализ корреляционного поля показывает, что зависимость между x и y носит нелинейный характер. Найдем коэффициент корреляции:

$$r = \frac{\overline{x\,y} - \overline{x} * \overline{y}}{\sigma_x * \sigma_y}$$

Для этого также составим вспомогательную таблицу:

	x	y	x*y	x*x	y*y
1	35	511	17 885	1 225	261 121
2	17	300	5 100	289	90 000
3	50	446	22 300	2 500	198 916
4	31	498	15 438	961	248 004
5	42	490	20 580	1 764	240 100
6	47	472	22 184	2 209	222 784
7	60	370	22 200	3 600	136 900
8	24	408	9 792	576	166 464
9	27	412	11 124	729	169 744
10	20	370	7 400	400	136 900
S	**353**	**4 277**	**154 003**	**14 253**	**1 870 933**
Среднее значение	**35,3**	**427,7**	**15 400,3**	**1 425,3**	**187 093,3**
σ^2	**179,21**	**4 166**			
σ	**13,39**	**64,54**			

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 = 1425{,}3 - 35{,}3^2 = 179{,}21$$

$$\sigma_x = \sqrt{179{,}21} = 13{,}39$$

$$\sigma_y^2 = \overline{y^2} - \bar{y}^2 = 187093{,}3 - 427{,}7^2 = 4\,166$$

$$\sigma_y = \sqrt{4166} = 64{,}54$$

$$r = \frac{\overline{x\,y} - \overline{x} * \overline{y}}{\sigma_x * \sigma_y} = \frac{15400{,}3 - 35{,}3 * 427{,}7}{13{,}39 * 64{,}54} = \frac{302{,}49}{864{,}2} \approx 0{,}35$$

Ответ: зависимость криволинейная, r = 0,35.

Глава 4. Регрессионный анализ данных

Тема 4.1 Линейная парная регрессия

Методические указания у решению задач

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессий, линейных по параметрам, используют метод наименьших квадратов (МНК), который позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических y_x минимальна:

$$\sum_{i=1}^{n}(y_i - \widehat{y_{xi}})^2 = \sum_{i=1}^{n}\varepsilon_i^2 \to min$$

Для линейных и нелинейных уравнений, приводимых к линейным, решается следующая система относительно a и b:

$$\begin{cases} na + b\sum x = \sum y \\ a\sum x + b\sum x^2 = \sum xy. \end{cases}$$

Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

$$a = \overline{y} - b\overline{x} \; ;$$

$$b = \frac{cov\,(x, y)}{\sigma_x^2} = \frac{\overline{xy} - \overline{x} * \overline{y}}{\overline{x^2} - \overline{x}^2}$$

Тесноту связи изучаемых явлений оценивает линейный коэффициент парной корреляции r_{xy} и индекс корреляции ρ_{xy}. Для линейной регрессии ($-1 \le r_{xy} \le 1$), причем, если коэффициент регрессии $b>0$, то $0 \le r_{xy} \le 1$, и, наоборот, при $b<0$ $\; -1 \le r_{xy} \le 0$.

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = \frac{cov(x, y)}{\sigma_x\,\sigma_y} = \frac{\overline{yx} - \overline{y} * \overline{x}}{\sigma_x\,\sigma_y}$$

Задача 1.

По семи территориям региона приводятся следующие данные:

Номер региона	Среднедушевой прожиточный минимум в день на одного трудоспособного, руб. x	Среднедневная заработная плата, у, руб.
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

Требуется:

– построить линейное уравнение парной регрессии у по х;

– рассчитать линейный коэффициент корреляции.

– выполнить прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x, составляющем 107% от среднего уровня.

– на одном графике отложить исходные данные и теоретическую прямую.

<u>Решение:</u>

Решение. Для расчета параметров a и b линейной регрессии y=a+bx построим таблицу:

№ п/п	**y**	**x**	**yx**	x^2	y^2
1	133	78	10 374	6 084	17 689
2	148	82	12 136	6 724	21 904
3	134	87	11 658	7 569	17 956
4	154	79	12 166	6 241	23 716
5	162	89	14 418	7 921	26 244
6	195	106	20 670	11 236	38 025
7	139	67	9 313	4 489	19 321
8	158	88	13 904	7 744	24 964
9	152	73	11 096	5 329	23 104
10	162	87	14 094	7 569	26 244
11	159	76	12 084	5 776	25 281
12	173	115	19 895	13 225	29 929
Итого	1 869	1 027	161 808	89 907	294 377

Среднее значение	**155,75**	**85,58**	**13 484**	**7 492,25**	**24 531,4**
σ^2	**16,53**	**12,97**	**-**	**-**	**-**
σ	**273,34**	**168,31**	**-**	**-**	**-**

$$b = \frac{cov\,(x,y)}{\sigma_x^2} = \frac{\overline{xy} - \overline{x} * \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{13484 - 155{,}75 * 85{,}58}{7492{,}25 - 85{,}58^2} = \frac{154{,}915}{168{,}31} \approx 0{,}92$$

$$a = \overline{y} - b\overline{x} = 155{,}75 - (0{,}92) * 85{,}58 \approx 77{,}02$$

Получено уравнение регрессии:

$$y = 77{,}02 + 0{,}92x$$

Параметр регрессии позволяет сделать вывод, что с увеличением среднедушевого прожиточного минимума на 1 руб. среднедневная заработная плата возрастает в среднем на 0,92 руб. (или 92 копейки).

Тесноту линейной связи оценит коэффициент корреляции:

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = 0{,}92 * \frac{12{,}97}{16{,}53} \approx 0{,}722$$

Значение коэффициента корреляции больше 0,7 говорит о наличии весьма тесной линейной связи между признаками.

Осуществим прогноз:

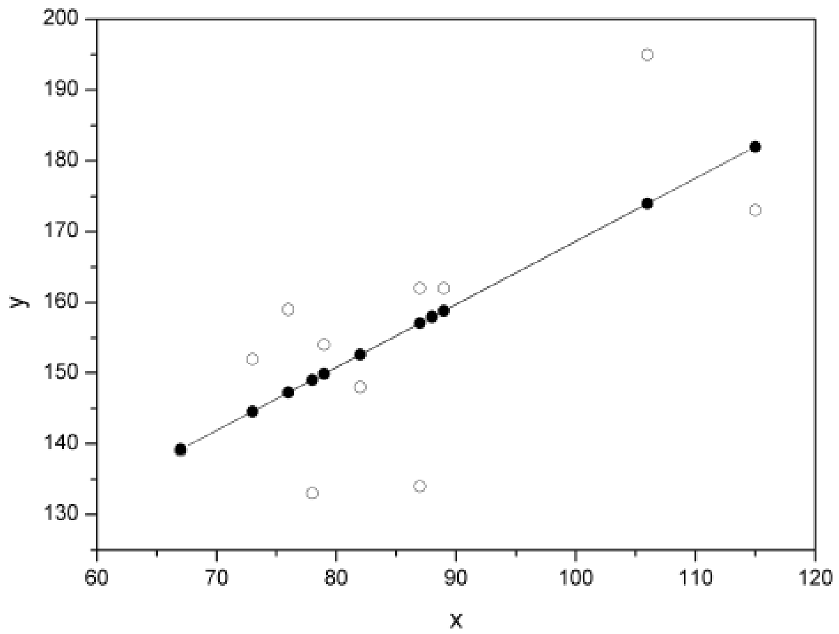
Если прогнозное значение прожиточного минимума составить

$$x_0 = \bar{x} * 1{,}07 = 85{,}6 * 1{,}07 = 91{,}6 \text{ руб.}$$

тогда индивидуальное прогнозное значение заработной платы составит:

$$y_0 = 77{,}02 + 0{,}92 * 91{,}6 = 161{,}29 \text{ руб.}$$

В заключении построим на одном графике исходные данные и теоретическую прямую:



<u>Ответ</u>: $y = 77{,}02 + 0{,}92x$; $r_{xy} = 0{,}722$

Тема 4.2 Проверка качества уравнения регрессии

Коэффициент детерминации (квадрат линейного коэффициента корреляции r_{xy}^2) характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака:

$$r_{xy}^2 = \frac{\sigma_{y\ \text{объсн}}^2}{\sigma_{y\ \text{общ}}^2} = \frac{\sum(\widehat{y_x} - \overline{y})^2}{\sum(y - \overline{y})^2}$$

Его можно интерпретировать как среднюю долю влияния x на y. Если, например, r_{xy}^2=0,13, то это значит, что вариация результат на 13 % объясняется вариацией фактора x.

Средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических.

Фактические значения результативного признака отличаются от теоретических, рассчитанных по уравнению регрессии, т. е. y и $\widehat{y_x}$. Чем меньше это отличие, тем ближе теоретические значения подходят к эмпирическим данным, лучше качество модели. Величина отклонений фактических и расчетных значений результативного признака (y – $\widehat{y_x}$) по каждому наблюдению представляет собой ошибку аппроксимации. Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации как среднюю арифметическую простую:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y - \hat{y}}{y}\right| * 100\%$$

Допустимый предел значений $\bar{A}$ – не более 8–10% (это свидетельствует о хорошем подборе модели к исходным данным).

Средний коэффициент эластичности $\overline{Э}$ показывает, на сколько процентов в среднем по совокупности изменится результат *y* от своей средней величины $\bar{y}$ при изменении фактора *x* на 1% от своего среднего значения:

$$\overline{Э} = f'(x)\frac{\overline{x}}{\overline{y}}$$

Оценка значимости уравнения регрессии в целом дается с помощью F-критерия Фишера. При этом выдвигается нулевая гипотеза, что коэффициент регрессии равен нулю, т. е. b=0, и, следовательно, фактор x не оказывает влияния на результат y.

$$F = \frac{D_{\text{факт}}}{D_{\text{ост}}}$$

Если нулевая гипотеза справедлива, то факторная и остаточная дисперсии не отличаются друг от друга. Фактическое значение F-критерия Фишера сравнивается с табличным значением $F_{\text{табл}}$ (α; k_1; k_2)при уровне значимости α и степенях свободы $k_1=m$ и $k_2= n–m –1$.

Табличное значение F-критерия – это максимальная величина отношения дисперсий, которая может иметь место при случайном их расхождении для данного уровня вероятности наличия нулевой гипотезы. Вычисленное значение F-критерия признается достоверным(отличным от единицы), если оно больше табличного. В этом случае нулевая гипотеза об отсутствии связи признаков отклоняется и делается вывод о существенности этой связи: $F_{\text{факт}}>F_{\text{табл}}$. H_0 отклоняется.

Если же величина окажется меньше табличной $F_{\text{факт}}<F_{\text{табл}}$, то вероятность нулевой гипотезы выше заданного уровня (например,0,05) и она не может быть отклонена без серьезного риска сделать неправильный вывод о наличии связи. В этом случае уравнение регрессии считается статистически незначимым, H_0 не отклоняется:

$$F = \frac{\sum(\hat{y} - \overline{y})^2 / m}{\sum(y - \hat{y})^2 / (n-m-1)} = \frac{r_{xy}^2}{1 - r_{xy}^2}\,(n-2)$$

где n – число единиц совокупности; m – число параметров при переменных x.

В линейной регрессии обычно оценивается значимость не только уравнения в целом, но и отдельных его параметров (по t-критерию Стьюдента). С этой целью по каждому из параметров определяется его стандартная ошибка m_b и m_a.

Стандартная ошибка коэффициента регрессии определяется последующей формуле:

$$m_b = \frac{\frac{\sqrt{\sum(y-\widehat{y_x})^2}}{(n-2)}}{\sum(x-\bar{x})^2} = \sqrt{\frac{S^2}{\sum(x-\bar{x})^2}}$$

где S^2 – остаточная дисперсия на одну степень свободы.

Величина стандартной ошибки совместно с t-распределением Стьюдента при n-2 степенях свободы применяется для проверки существенности коэффициента регрессии и для расчета его доверительных интервалов.

Для оценки существующего коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т. е. определяется фактическое значение t-критерия Стьюдента: $t_b = \frac{b}{m_b}$, которое затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы (n – 2). Если фактическое значение t-критерия превышает табличное, то гипотезу о несущественности коэффициента регрессии можно отклонить.

Стандартная ошибка параметра a определяется по следующей формуле:

$$m_a = \sqrt{\frac{\sum(y-\widehat{y_x})^2}{n-2}\frac{\sum x^2}{n\sum(x-\bar{x})^2}} = \sqrt{S^2 * \frac{\sum x^2}{\sum(x-\bar{x})^2}}$$

Процедура оценивания существенности данного параметра не отличается от рассмотренной выше для коэффициента регрессии, вычисляется t-критерий: $t_a = \frac{a}{m_a}$, его величина сравнивается с табличным значением при n-2 степенях свободы.

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции:

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

Фактическое значение t-критерия определяется как:

$$t_r = \frac{r}{m_r} = \frac{r}{\sqrt{1 - r^2}}\sqrt{n - 2}$$

Для расчета доверительных интервалов определяются предельные ошибки каждого показателя:

$$\Delta a = t_{табл} m_a$$

$$\Delta b = t_{табл} m_b$$

Формулы для расчета доверительных интервалов имеют следующий вид:

$$a - t_{табл} m_a \le a \le a + t_{табл} m_a$$

$$b - t_{табл} m_b \le b \le b + t_{табл} m_b$$

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное и отрицательное значения.

<u>Задача 1.</u>

По семи территориям Уральского района известны значения двух признаков:

Район	Расходы на покупку продовольственных товаров в общих расходах, *y*, %	Среднедневная заработная плата одного работающего, *x*, у. е.
Удмуртская респ.	68,8	45,1
Свердловская обл.	61,2	59,0
Башкортостан	59,9	57,2

Челябинская обл.	56,7	61,8
Пермская обл.	55,0	58,8
Курганская обл.	54,3	47,2
Оренбургская обл.	49,3	55,2

Требуется:

– для характеристики зависимости *y* от *x* рассчитать параметры a и b;

– оценить модель через ошибку аппроксимации $\overline{A}$ и *F*-критерий.

Указание

Чтобы сделать вывод о значимости регрессионного уравнения необходимо сравнить расчетные значения F-критерия с табличными.

Решение:

Решение. Для расчета параметров *a* и *b* линейной регрессии y=a+bx построим таблицу:

По исходным данным рассчитываем $\sum x, \sum y, \sum xy, \sum x^2, \sum y^2$

№ п/п	**y**	**x**	**yx**	$\boldsymbol{x^2}$	$\boldsymbol{y^2}$	$\boldsymbol{\hat{y}_x}$	**y-$\boldsymbol{\hat{y}_x}$**	$\boldsymbol{A_t}$
1	68,8	45,1	3102,88	2034,01	4733,44	61,3	7,5	10,9
2	61,2	59,0	3610,80	3481,00	3745,44	56,5	4,7	7,7
3	59,9	57,2	3426,28	3271,84	3588,01	57,1	2,8	4,7
4	56,7	61,8	3504,06	3819,24	3214,89	55,5	1,2	2,1
5	55,0	58,8	3234,00	3457,44	3025,00	56,5	–1,5	2,7
6	54,3	47,2	2562,06	2227,84	2948,49	60,5	–6,2	11,4

7	49,3	55,2	2721,36	3047,04	2430,49	57,8	–8,5	17,2
Итого	405,2	384,2	22162,34	21338,41	23685,76	405,2	0,0	56,7
Среднее значение	**57,89**	**54,9**	**3 166,05**	**3 048,34**	**3 383,68**	-	**-**	**8,1**
σ^2	**5,7**	**5,9**	**-**	**-**	**-**	**-**	**-**	**-**
σ	**32,43**	**34,33**	**-**	**-**	**-**	**-**	**-**	**-**

$$b = \frac{cov\,(x, y)}{\sigma_x^2} = \frac{\overline{xy} - \overline{x} * \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{3166{,}05 - 57{,}89 * 54{,}9}{3048{,}34 - 54{,}9^2} = \frac{-12{,}11}{34{,}33} \approx -0{,}35$$

$$a = \overline{y} - b\overline{x} = 57{,}89 - (-0{,}35) * 54{,}9 \approx 77$$

$$y = 77 - 0{,}35x$$

С увеличением среднедневной заработной платы на одну условную единицу (у. е.) доля расходов на покупку продовольственных товаров снижается в среднем на 0,35% пункта.

Рассчитаем линейный коэффициент парной корреляции:

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = -0{,}35 * \frac{5{,}9}{5{,}7} \approx -0{,}362$$

Связь умеренная, обратная.

Определим коэффициент детерминации:

$$r_{xy}{}^2 = (-0{,}362)^2 = 0{,}131.$$

Вариация результата на 13,1% объясняется вариацией фактора *x*.

Подставляя в уравнение регрессии фактические значения *x*, определим теоретические (расчетные) значения $\widehat{y}_x$.

Найдем величину средней ошибки аппроксимации $\bar{A}$:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y-\hat{y}}{y}\right| * 100\% = \frac{56{,}7}{7} * 100\% = 8{,}1\%$$

В среднем расчетные значения отклоняются от фактических на 8,1%.

Рассчитаем F-критерий:

$$F == \frac{r_{xy}^2}{1-r_{xy}^2}\ (n-2) = \frac{0{,}131}{1-0{,}131} * 5 = 0{,}75$$

Полученное значение указывает на необходимость применять гипотезу H_0 о случайной природе выявленной зависимости, т. к. $F_{факт.} < F_{табл.}$= 6,61 (табличные значения F-критерия приведены в приложении лекционного курса).

Ответ: a=0,77; b= – 0,35; A=8,1%; F=0,75

Задача 2.

По группе предприятий, производящих однородную продукцию, известно, как зависит себестоимость единицы продукции *y* от факторов, приведенных в табл.:

Признак-фактор	**Уравнение парной регрессии**	**Среднее значение фактора**
Объем производства, x_1, млн. р.	$\widehat{y_{x1}} = 0{,}65 + 58{,}85 * \frac{1}{x_1}$	$\overline{x_1} = 2{,}55$
Трудоемкость единицы продукции, x_2, чел.-ч.	$\widehat{y_{x2}} = 9{,}5 + 9{,}85 * x_2$	$\overline{x_2} = 1{,}58$
Оптовая цена за 1 т	$\widehat{y_{x3}} = 15{,}75 * x_3^{1{,}55}$	$\overline{x_3} = 1{,}53$

энергоносителя, x_3, млн. р.		
Доля прибыли, изымаемой государством, x_4, %	$\widehat{y_{x4}} = 14{,}07 * 1{,}01^{x_4}$	$\overline{x_4} = 28{,}35$

Требуется:

– определить с помощью коэффициентов эластичности силу влияния каждого фактора на результат;

– ранжировать факторы по силе влияния.

Указание

При нахождение коэффициентов эластичности можно руководствоваться следующей таблицей производных основных элементарных функций.

1. $c' = 0,\ c = \text{const}$
2. $\left(x^n\right)' = nx^{n-1}$
3. $\left(a^x\right)' = a^x \cdot \ln a$
4. $\left(e^x\right)' = e^x$
5. $(\log_a x)' = \dfrac{1}{x \ln a}$
6. $(\ln x)' = \dfrac{1}{x}$
7. $(\sin x)' = \cos x$
8. $(\cos x)' = -\sin x$
9. $\left(\sqrt{x}\right)' = \dfrac{1}{2\sqrt{x}}$
10. $(\mathrm{tg} x)' = \dfrac{1}{\cos^2 x}$
11. $(\mathrm{ctg} x)' = -\dfrac{1}{\sin^2 x}$
12. $(\arcsin x)' = \dfrac{1}{\sqrt{1-x^2}}$
13. $(\arccos x)' = -\dfrac{1}{\sqrt{1-x^2}}$
14. $(\mathrm{arctg} x)' = \dfrac{1}{1+x^2}$
15. $(\mathrm{arcctg} x)' = -\dfrac{1}{1+x^2}$
16. $(\mathrm{sh}\, x)' = \mathrm{ch}\, x$
17. $(\mathrm{ch}\, x)' = \mathrm{sh}\, x$
18. $(\mathrm{th}\, x)' = \dfrac{1}{\mathrm{ch}^2 x}$
19. $(\mathrm{th}\, x)' = -\dfrac{1}{\mathrm{sh}^2 x}$

Решение:

Для уравнения равносторонней гиперболы

$$\widehat{y_{x1}} = 0{,}65 + 58{,}85 * \frac{1}{x_1}$$

$$\overline{Э_{yx1}} = f'(x_1)\frac{\overline{x_1}}{\overline{y}} = -\frac{b}{\overline{x_1^2}} * \frac{\overline{x_1}}{a + {}^{b}/_{\overline{x_1}}} = -\frac{b}{a\overline{x_1} + b} = -\frac{58{,}85}{0{,}65 * 2{,}55 + 58{,}85}$$
$$= -0{,}973\%$$

Для уравнения прямой

$$\widehat{y_{x2}} = 9{,}5 + 9{,}85 * x_2$$

$$\overline{Э_{yx2}} = f'(x_2)\frac{\overline{x_2}}{\overline{y}} = \frac{b\overline{x_2}}{a + b\overline{x_2}} = \frac{9{,}85 * 1{,}58}{9{,}5 + 9{,}85 * 1{,}58} = 0{,}62\%$$

Для уравнения степенной зависимости

$$\widehat{y_{x3}} = 15{,}75 * x_3^{1{,}55}$$

$$\overline{Э_{yx3}} = f'(x_3)\frac{\overline{x_3}}{\overline{y}} = abx^b\frac{1}{x} * \frac{x}{ax^b} = b = 1{,}55\%$$

Для уравнения показательной зависимости

$$\widehat{y_{x4}} = 14{,}07 * 1{,}01^{x_4}$$

$$\overline{Э_{yx4}} = f'(x_4)\frac{\overline{x_4}}{\overline{y}} = ab^{\overline{x_4}}\ln b\,\frac{\overline{x_4}}{ab^{\overline{x_4}}} = \ln b\,\overline{x_4} = \ln 1{,}01 * 26{,}35 = 0{,}26\,\%$$

Сравнивая значения $\overline{Э_{yx\iota}}$, ранжируем x_j по силе их влияния на себестоимость единицы продукции:

а) $\overline{Э_{yx3}} = 1{,}55\%$, б) $\overline{Э_{yx1}} = -0{,}973\%$, в) $\overline{Э_{yx2}} = 0{,}62\%$,

г) $\overline{Э_{yx4}} = 0{,}26\,\%$

Для формирования уровня себестоимости продукции группы предприятий первоочередное значение имеют цены на энергоносители; в гораздо меньшей степени влияют трудоемкость продукции и отчисляемая часть прибыли. Фактором снижения себестоимости выступает размер производства: с ростом его на 1% себестоимость единицы продукции снижается на –0,973%.

Ответ: а) $\overline{Э_{yx3}} = 1{,}55\%$, б) $\overline{Э_{yx1}} = -0{,}973\%$, в) $\overline{Э_{yx2}} = 0{,}62\%$, г) $\overline{Э_{yx4}} = 0{,}26\ \%$

Тема 4.3 Нелинейная регрессия

Методические указания у решению задач

Метод наименьших квадратов не предназначен для оценки параметров нелинейных регрессий, кроме полиномиальной. Однако целый класс нелинейных зависимостей можно привести к зависимостям линейным преобразованием независимой и/или зависимой переменной. Такое преобразование называется *линеаризацией*.

Методы линеаризации:

1) Метод логарифмирования – применяется к степенным функциям;
2) Метод обратного преобразования – для дробных функций;
3) Комплексный метод – для дробных и степенных функций.

Способы преобразования данных и вычисления параметров a, b по оценкам a*, b*

Исходная спецификация	**Преобразование**	**Преобразование**	**Вычисление *b* по *b****	**Вычисление *a* по *a****
	$x \to x^*$	$y \to y^*$		

$y = a + \frac{b}{x} + \varepsilon$	$x^* = \frac{1}{x}$	$y^* = y$	$b = b^*$	$a = a^*$
$y = \frac{1}{a + bx + \varepsilon}$	$x^* = x$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
$y = \frac{x}{a + bx + x\varepsilon}$	$x^* = \frac{1}{x}$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
$y = ae^{bx+\varepsilon}$	$x^* = x$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$
$y = ae^{\frac{b}{x}+\varepsilon}$	$x^* = \frac{1}{x}$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$
$y = \frac{1}{a + be^{-x} + \varepsilon}$	$x^* = e^{-x}$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
$y = ax^b e^{\varepsilon}$	$x^* = \ln x$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$; $a^* = \ln(a)$

Задача 1.

Некоторая организация желает исследовать зависимость полученной прибыли Y (сотни тыс. руб.) от вложения средств в научные разработки выпускаемой продукции X (тыс. руб.). Для этого рассматриваются 4 регрессионных уравнения:

- линейное: $y = bx + a$,
- гиперболическое $y = \frac{b}{x} + a$,
- и степенное $y = ax^b$,
- экспоненциальное $y = ae^{bx}$.

В результате наблюдений, получены данные:

Прибыль Y	5	6	8	11	16	22	29	35	44	57	83
Вложения X	2	4	7	9	10	12	15	16	20	22	25

1) Линейная функция:

	x	y	x*y	x^2	y^2
1	2	5	10	4	25
2	4	6	24	16	36
3	7	8	56	49	64
4	9	11	99	81	121

5	10	16	160	100	256
6	12	22	264	144	484
7	15	29	435	225	841
8	16	35	560	256	1 225
9	20	44	880	400	1 936
10	22	57	1 254	484	3 249
11	25	83	2 075	625	6 889
сумма	**142**	**316**	**5 817**	**2 384**	**15 126**
среднее	**12,9**	**28,7**	**528,8**	**216,7**	**1 375,1**

$$r = \frac{\overline{x\,y} - \overline{x} * \overline{y}}{\sigma_x * \sigma_y} = \frac{528{,}8 - 12{,}9 * 28{,}7}{\sqrt{216{,}7 - 12{,}9^2} * \sqrt{1375{,}1 - 28{,}7^2}} = \frac{158{,}7}{\sqrt{50{,}29} * 551{,}41}$$

$$= \frac{158{,}7}{7{,}09 * 23{,}48} \approx 0{,}952$$

$$R = {r_{xy}}^2 = (0{,}952)^2 = 0{,}91.$$

$$b = \frac{cov\,(x, y)}{\sigma_x^2} = \frac{\overline{xy} - \overline{x} * \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{158{,}7}{50{,}29} \approx 3{,}15$$

$$a = \overline{y} - b\overline{x} = 28{,}7 - 3{,}15 * 12{,}9 \approx -11{,}94$$

$$y = 3{,}15x - 11{,}94$$

2) Гиперболическая функция:

	1/x	y	1/x *y	$(1/x)^2$	y*y
1	0,5	5	2,5	0,25	25
2	0,25	6	1,5	0,0625	36
3	0,142857	8	1,142857	0,020408	64
4	0,111111	11	1,222222	0,012346	121
5	0,1	16	1,6	0,01	256
6	0,083333	22	1,833333	0,006944	484
7	0,066667	29	1,933333	0,004444	841
8	0,0625	35	2,1875	0,003906	1 225
9	0,05	44	2,2	0,0025	1 936
10	0,045455	57	2,590909	0,002066	3 249
11	0,04	83	3,32	0,0016	6 889
сумма	**1,451923**	**316**	**22,03016**	**0,376715**	**15 126**
среднее	**0,13**	**28,7**	**2,00**	**0,034**	**1 375,1**

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

Сделаем замену: $x^* = \frac{1}{x}$; $y^* = y$; $b^* = b$; $a^* = a$;
Тогда получаем уравнение линейного вида: $y^* = b^*x^* + a^*$;

$$r = \frac{\overline{x\,y} - \overline{x} * \overline{y}}{\sigma_x * \sigma_y} = \frac{2 - 0{,}13 * 28{,}7}{\sqrt{0{,}034 - 0{,}13^2} * 23{,}48} = \frac{-1{,}73}{\sqrt{0{,}0171} * 23{,}48} = \frac{-1{,}73}{0{,}13 * 23{,}48} \approx -0{,}57$$

$$R = {r_{xy}}^2 = (-0{,}57)^2 = 0{,}32.$$

$$b^* = \frac{cov\,(x, y)}{\sigma_x^2} = \frac{\overline{xy} - \overline{x} * \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{-1{,}73}{0{,}0171} \approx -101 = b$$

$$a^* = \overline{y} - b\overline{x} = 28{,}7 + 101 * 0{,}13 \approx 41{,}86 = a$$

$$y = \frac{-101}{x} + 41{,}86$$

3) Степенная регрессия:

	ln x	ln y	ln x * ln y	$(\ln x)^2$	$(\ln y)^2$
1	0,69	1,61	1,12	0,48	2,59
2	1,39	1,79	2,48	1,92	3,21
3	1,95	2,08	4,05	3,79	4,32
4	2,20	2,40	5,27	4,83	5,75
5	2,30	2,77	6,38	5,30	7,69
6	2,48	3,09	7,68	6,17	9,55
7	2,71	3,37	9,12	7,33	11,34
8	2,77	3,56	9,86	7,69	12,64
9	3,00	3,78	11,34	8,97	14,32
10	3,09	4,04	12,50	9,55	16,35
11	3,22	4,42	14,22	10,36	19,53
сумма	**25,80**	**32,91**	**84,01**	**66,40**	**107,29**
среднее	**2,35**	**3**	**7,64**	**6,04**	**9,75**

$y = ax^b$,

Чтобы привести уравнение к линейному виду, прологарифмируем левую и правую часть:

$\ln y = \ln a + b \ln x$,

Сделаем замену: $y^* = \ln y$; $a^* = \ln a$; $x^* = \ln x$; $b^* = b$;
Тогда получаем уравнение линейного вида: $y^* = b^*x^* + a^*$;

$$r = \frac{\overline{x\,y} - \overline{x} * \overline{y}}{\sigma_x * \sigma_y} = \frac{7{,}64 - 2{,}35 * 3}{\sqrt{6{,}04 - 2{,}35^2} * \sqrt{9{,}75 - 9}} = \frac{0{,}59}{\sqrt{0{,}52} * \sqrt{0{,}75}} = \frac{0{,}59}{0{,}72 * 0{,}866} \approx 0{,}946$$

$$R = {r_{xy}}^2 = (0{,}946)^2 = 0{,}89.$$

$$b^* = \frac{cov\,(x,y)}{\sigma_x^2} = \frac{\overline{xy} - \overline{x} * \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{0{,}59}{0{,}52} \approx 1{,}13 = b$$

$$a^* = \overline{y} - b\overline{x} = 3 - 1{,}13 * 2{,}35 \approx 0{,}34$$

$$a = e^{0{,}34} = 2{,}71^{0{,}34} \approx 1{,}4$$

$$y = 1{,}4 * x^{1{,}13}$$

4) Экспоненциальная регрессия:

	x	ln y	x*ln y	x^2	$(\ln y)^2$
1	2	1,61	3,22	4,00	2,59
2	4	1,79	7,17	16,00	3,21
3	7	2,08	14,56	49,00	4,32
4	9	2,40	21,58	81,00	5,75
5	10	2,77	27,73	100,00	7,69
6	12	3,09	37,09	144,00	9,55
7	15	3,37	50,51	225,00	11,34
8	16	3,56	56,89	256,00	12,64
9	20	3,78	75,68	400,00	14,32
10	22	4,04	88,95	484,00	16,35
11	25	4,42	110,47	625,00	19,53
сумма	**142**	**32,91**	**493,84**	**2 384,00**	**107,29**
среднее	**12,9**	**3**	**44,89**	**216,7**	**9,75**

$y = ae^{bx}$

Чтобы привести уравнение к линейному виду, прологарифмируем левую и правую часть:

$\ln y = \ln a + \ln e^{bx} = \ln a + bx,$

Сделаем замену: $y^* = \ln y$; $a^* = \ln a$; $x^* = x$; $b^* = b$;

Тогда получаем уравнение линейного вида: $y^* = b^* x^* + a^*$;

$$r = \frac{\overline{x\,y} - \overline{x} * \overline{y}}{\sigma_x * \sigma_y} = \frac{44{,}89 - 12{,}9 * 3}{7{,}09 * 0{,}866} = \frac{6{,}19}{7{,}09 * 0{,}866} \approx 1$$

$$R = {r_{xy}}^2 = 1.$$

$$b^* = \frac{cov\,(x,y)}{\sigma_x^2} = \frac{\overline{xy} - \overline{x} * \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{6{,}19}{50{,}29} \approx 0{,}123 = b$$

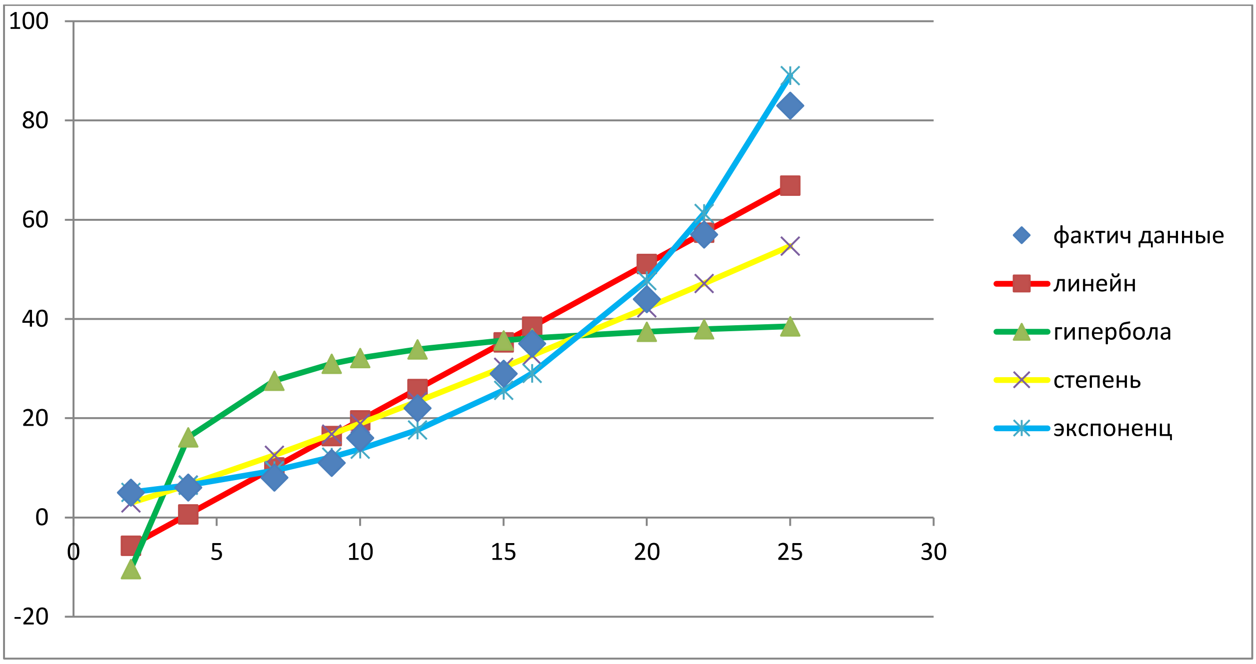
$$a^* = \overline{y} - b\overline{x} = 3 - 0{,}123 * 12{,}9 \approx 1{,}41$$

$$a = e^{1{,}41} \approx 4$$

$$y = 4 * e^{0{,}123x}$$

Для того, чтобы определить, какая модель наилучшим образом описывает данные, сравним индексы детерминации для каждой модели. Наибольшим является R_4= 1. Значит, опытные данные лучше описывать моделью $y = 4 * e^{0{,}123x}$.

На графике полученные зависимости выглядят следующим образом:



Тема 4.4 Множественная регрессия и корреляция

Методические указания у решению задач

Множественная регрессия – уравнение связи с несколькими независимыми переменными:

$$y = F(x_1, x_2, \dots, x_p),$$

где y – зависимая переменная (результативный признак); x_1, x_2, …, x_p – независимые переменные (факторы).

Для построения множественной регрессии используются линейная, степенная, экспоненциальная и гиперболическая функции, а также другие функции, приводимые к линейному виду. Для линейных и нелинейных уравнений, приводимых к линейным, строится следующая система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии:

$$\sum y = na + b_1 \sum x_1 + b_2 \sum x_2 + \dots + b_p \sum x_p,$$

$$\sum yx_1 = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + \dots + b_p \sum x_1 x_p,$$

……………………………………………………………………………

$$\sum yx_p = a \sum x_p + b_1 \sum x_1 x_p + b_2 \sum x_2 x_p + \dots + b_p \sum x_p^2.$$

Для ее решения может быть применен метод определителей:

$$a = \frac{\Delta a}{\Delta}, b_1 = \frac{\Delta b_1}{\Delta}, \dots, b_p = \frac{\Delta b_p}{\Delta},$$

где $\Delta = \begin{vmatrix} n & \sum x_1 & \sum x_2 & \dots & \sum x_p \\ \sum x_1 & \sum x_1^2 & \sum x_2 x_1 & \dots & \sum x_p x_1 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \dots & \sum x_p x_2 \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_p & \sum x_1 x_p & \sum x_2 x_p & \dots & \sum x_p^2 \end{vmatrix}$ – определитель системы.

Задача 1.

Имеются данные о ценах и дивидендах по обыкновенным акциям, а также о доходности компании.

№ п/п	Цена акции у.е	Доходность капитала, %	Уровень дивидендов, %
1	25	15,1	2,6
2	20	13,9	2,1
3	15	15,8	1,5
4	34	12,8	3,1
5	20	6,9	2,5
6	33	14,6	3,1
7	28	15,4	2,9
8	30	17,3	2,8
9	23	13,7	2,4
10	24	15,3	2,4
11	25	15,2	2,6
12	26	12,0	2,8
13	26	15,3	2,7
14	20	15,2	1,9
15	20	13,7	1,9
16	13	13,3	1,6
17	21	15,1	2,4

18	31	15,2	3,0
19	26	11,2	3,1
20	11	12,1	2,0

Задание: построить линейное уравнение множественной регрессии и пояснить экономический смысл его параметров.

Решение.

Необходимо построить расчетную табл.

№ п/п	**y**	x_1	x_2	$x_2 * x_2$	$x_1 * x_1$	**y** $*x_1$	$x_1 * x_2$
1	25	15.2	2.6	223.04	380.0	65.0	39.52
2	20	13.9	2.1	193.21	278.0	42.0	29.19
3	15	15.8	1.5	249.64	237.0	22.5	23.70
4	34	12.8	3.1	163.84	435.2	105.4	39.68
5	20	6.9	2.5	47.61	138.0	50.0	17.25
6	33	14.6	3.1	213.16	481.8	102.3	45.26
7	28	15.4	2.9	237.16	431.2	81.2	44.66
8	30	17.3	2.8	299.29	519.0	84.0	48.44
9	23	13.7	2.4	187.69	315.1	55.2	32.88
10	24	12.7	2.4	161.29	304.8	57.6	30.48
11	25	15.3	2.6	234.09	382.5	65.0	39.78
12	26	15.2	2.8	231.04	312.0	72.8	42.56
13	26	12.0	2.7	144.0	306.0	70.2	32.40

14	20	15.3	1.9	234.09	312.0	38.0	29.07
15	20	13.7	1.9	187.69	306.0	38.0	26.03
16	13	13.3	1.6	228.01	274.0	20.8	21.28
17	21	15.1	2.4	225.0	317.1	50.4	36.24
18	31	15.0	3.0	125.44	465.0	93.0	45.0
19	26	11.2	3.1	146.41	291.1	80.6	34.72
20	11	12.1	2.0	146.41	133.1	22.0	24.20
Итого	**471**	**276.5**	**49.4**	**3916.59**	**6569.1**	**1216**	**682.34**
Среднее Значение	**23.55**	**-**	**-**	**-**	**325.455**	**60.8**	**34.117**
σ	**6.07**	**2.168**	**0.484**				

По данным табл. строится система нормальных уравнений с тремя неизвестными:

$$\begin{cases} \sum y = na + b_1 \sum x_1 + b_2 \sum x_2 \,, \\ \sum y x_{1=} a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2, \\ \sum y x_2 = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_{2\,,}^2 \end{cases}$$

$$\begin{cases} 471 = 20a + 276{,}5b_1 + 49{,}4b_2, \\ 6569{,}1 = 276{,}5a + 3916{,}59b_1 + 682{,}34b_{2,} \\ 1216 = 49{,}4a + 682{,}34b_1 + 126{,}7b_2. \end{cases}$$

Из этой системы находятся коэффициенты a, b_1, b_2:

a= – 13,925; $b_1 = 0{,}686; b_2 = 11{,}331.$

Таким образом, уравнение множественной регрессии имеет следующий вид:

$$\hat{b} = -13{,}925 + 0{,}686x_1 + 11{,}331x_2.$$

Экономический смысл коэффициентов $b1$ и $b2$ в том, что это показатели силы связи, характеризующие изменение цены акции при изменении какого-либо факторного признака на единицу своего измерения при фиксированном влиянии другого фактора.

Глава 5. Использование различных функций Microsoft Excel в эконометрике

Тема 5.1 Расчет характеристик модели с помощью Microsoft Excel.

Задача 1.

Решим задачу п.4.1 с помощью MS Excel и сравним результаты.

По территориям региона приводятся следующие данные:

Номер региона	Средний прожиточный минимум, x, руб.	Среднедневная заработная плата, y, руб.
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

Задание:

1. Построить линейное уравнение парной регрессии y по x .

2. Рассчитать линейный коэффициент парной корреляции и коэффициент детерминации.

3. Оценить статистическую значимость уравнения регрессии в целом и отдельных параметров регрессии и корреляции с помощью F - критерия Фишера и t -критерия Стьюдента.

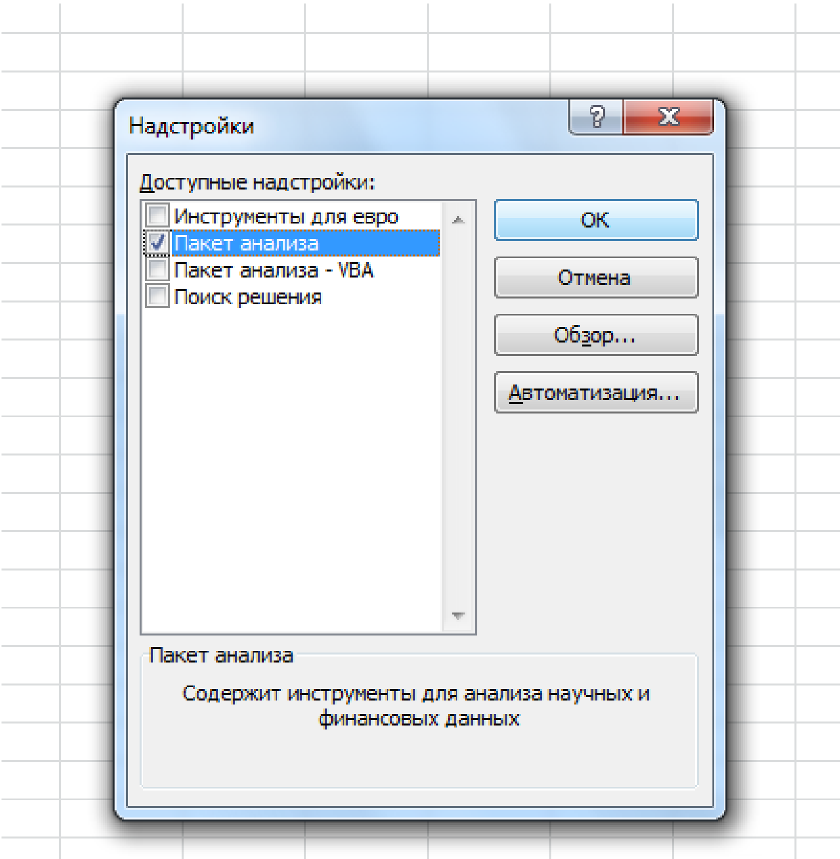
4. На одном графике отложить исходные данные и теоретическую прямую.

Решение.

С помощью инструмента анализа данных **Регрессия** можно получить результаты регрессионной статистики, дисперсионного анализа, доверительных интервалов, остатки и графики подбора линии регрессии.

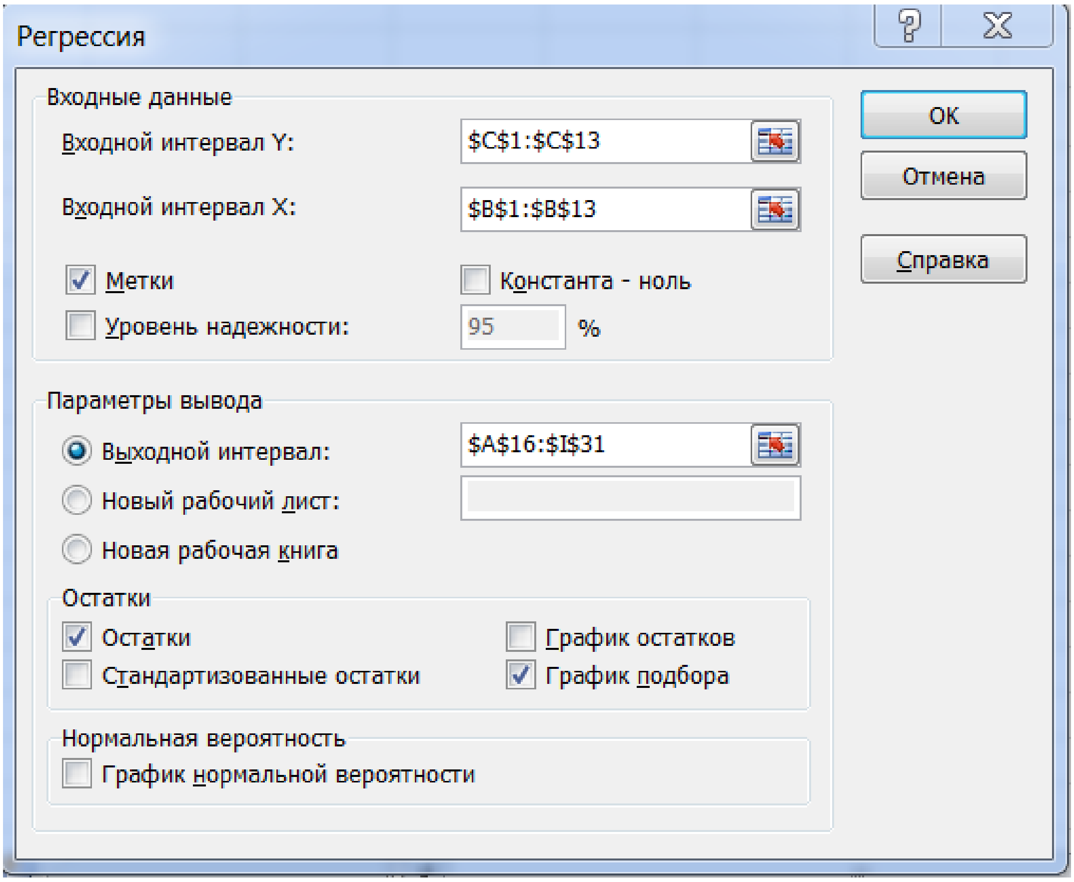
Если в меню «Данные» еще нет команды **Анализ данных**, то необходимо сделать следующее.

Вкладка Файл→Параметры→Надстройки→кнопка «перейти» →в окне «надстройки» установить флажок «Пакет анализа» →ОК.



Дальнейший алгоритм действий:

1. Вносим в таблицу исходные данные;
2. Выбираем Данные→Анализ данных→Регрессия.
3. Заполняем диалоговое окно ввода данных и параметров вывода:



Здесь:

Входной интервал Y – диапазон, содержащий данные результативного признака;

Входной интервал X – диапазон, содержащий данные признака-фактора;

Метки – «флажок», который указывает, содержит ли первая строка названия столбцов;

Константа – ноль – «флажок», указывающий на наличие или отсутствие свободного члена в уравнении;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно указать произвольное имя нового листа (или не указывать, тогда результаты выводятся на вновь созданный лист).

Получаем следующие результаты:

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественнı	0,721025214
R-квадрат	0,519877359
Нормированнı	0,471865095
Стандартная с	12,5495908
Наблюдения	12

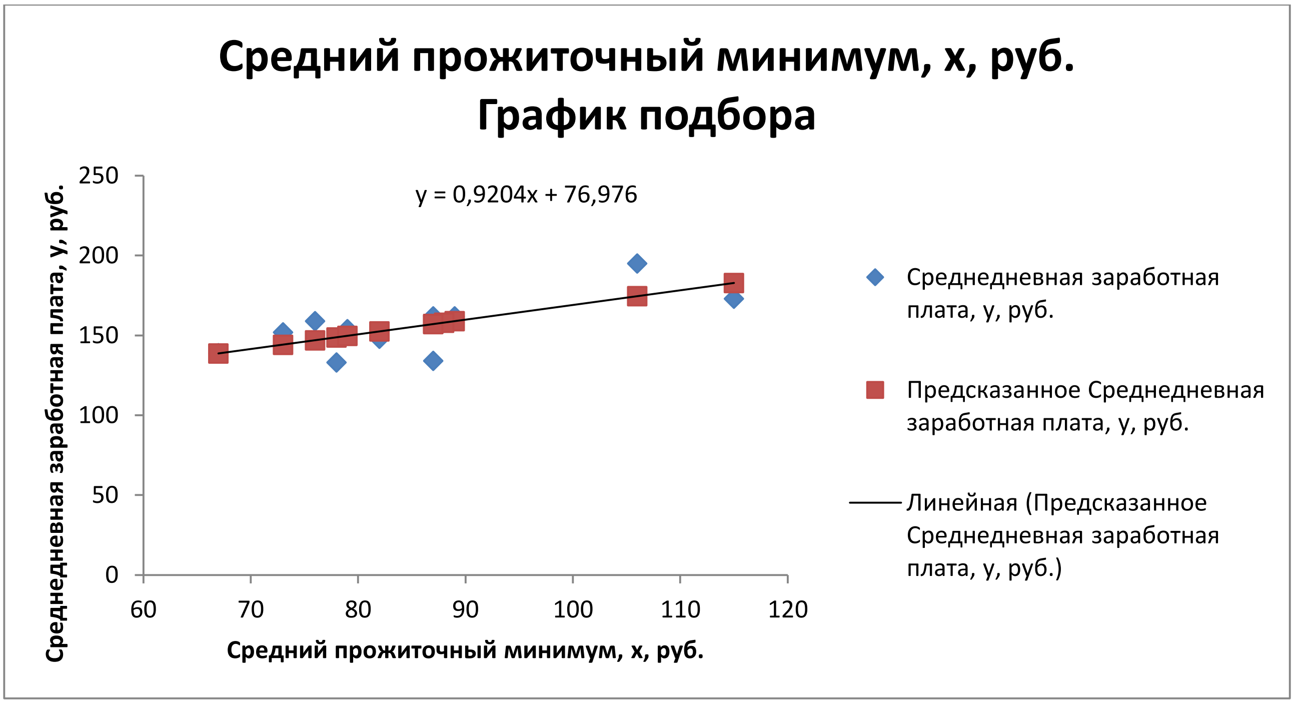
Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	1705,327706	1705,33	10,82801	0,008141843
Остаток	10	1574,922294	157,492		
Итого	11	3280,25			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечени	76,9764852	24,21156138	3,17933	0,009831	23,02976462	130,923	23,029765	130,923
Средний прож	0,920430553	0,279715587	3,29059	0,008142	0,297185386	1,54368	0,2971854	1,54368

ВЫВОД ОСТАТКА

Наблюдение	*Среднедневная заработна*	*Остатки*
1	148,7700683	-15,77006831
2	152,4517905	-4,45179052
3	157,0539433	-23,05394328
4	149,6904989	4,309501138
5	158,8948044	3,105195612
6	174,5421238	20,45787622
7	138,6453322	0,354667771
8	157,9743738	0,025626164
9	144,1679155	7,832084455
10	157,0539433	4,946056717
11	146,9292072	12,0707928
12	182,8259988	-9,825998758



Откуда выписываем, округляя до 4 знаков после запятой и переходя к нашим обозначениям:

Уравнение регрессии:

$$y = 76{,}9765 + 0{,}9204 * x$$

Коэффициент корреляции:

$$r_{xy} \approx 0{,}7210$$

Коэффициент детерминации:

$$r_{xy}^2 \approx 0{,}5199$$

Фактическое значение F -критерия Фишера:

$$F = 10{,}8280$$

Остаточная дисперсия на одну степень свободы:

$$S_{\text{ост}}^2 = 157{,}4922$$

Корень квадратный из остаточной дисперсии (стандартная ошибка):

$$S_{\text{ост}} = 12{,}5496$$

Стандартные ошибки для параметров регрессии:

$$m_a = 24{,}2116$$

$$m_b = 0{,}2797$$

Фактические значения t -критерия Стьюдента:

$$t_a = 3{,}1793$$

$$t_b = 3{,}2906$$

Доверительные интервалы:

$$23{,}0298 \leq a \leq 130{,}9232,$$

$$0{,}2972 \leq b \leq 1{,}5437 .$$

Как видим, результаты «ручного счета» от машинного отличаются незначительно (отличия связаны с ошибками округления).

Глава 6. Статистический анализ временных рядов

Тема 6.1 Основы анализа временных рядов

Временной ряд — это совокупность значений какого-либо показателя за несколько последовательных моментов (периодов) времени.

Метод скользящего среднего.

Скользящие средние позволяют сгладить как случайные, так и периодические колебания, выявить имеющуюся тенденцию в развитии процесса, и поэтому, являются важным инструментом при фильтрации компонент временного ряда.

Алгоритм сглаживания по простой скользящей средней может быть представлен в виде следующей последовательности шагов:

1. Определяют длину интервала сглаживания g, включающего в себя g последовательных уровней ряда (g<n). При этом надо иметь в виду, что чем шире интервал сглаживания, тем в большей степени взаимопогашаются колебания, и тенденция развития носит более плавный, сглаженный характер. Чем сильнее колебания, тем шире должен быть интервал сглаживания.

2. Разбивают весь период наблюдений на участки, при этом интервал сглаживания как бы скользит по ряду с шагом, равным 1.

3. Рассчитывают арифметические средние из уровней ряда, образующих каждый участок.

4. Заменяют фактические значения ряда, стоящие в центре каждого участка, на соответствующие средние значения.

При этом удобно брать длину интервала сглаживания g в виде нечетного числа: g=2p+1, т.к. в этом случае полученные значения скользящей средней приходятся на средний член интервала.

Наблюдения, которые берутся для расчета среднего значения, называются активным участком сглаживания. При нечетном значении g все уровни активного участка могут быть представлены в виде:

$$y_{t-p}, y_{t-p+1}, \cdots, y_{t-1}, y_t, y_{t+1}, \cdots, y_{t+p-1}, y_{t+p}$$

а скользящая средняя определена по формуле:

$$\widehat{y_t} = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \cdots + y_{t+p-1} + y_{t+p}}{2p+1}$$

где y_i – фактическое значение i-го уровня;

$\widehat{y_t}$ – значение скользящей средней в момент t;

2р+1– длина интервала сглаживания.

Процедура сглаживания приводит к полному устранению периодических колебаний во временном ряду, если длина интервала сглаживания берется равной или кратной циклу, периоду колебаний.

Тема 6.2 Фильтрация и сглаживание временного ряда

Метод экспоненциального сглаживания (Брауна) наиболее эффективен при разработке среднесрочных прогнозов. Он приемлем при прогнозировании только на один период вперед. Его основные достоинства простота процедуры вычислений и возможность учета весов исходной информации. Рабочая формула метода экспоненциального сглаживания:

$$U_{t+1} = \alpha * y_t + (1-\alpha) * U_t$$

где t – период, предшествующий прогнозному; t+1 – прогнозный период; U_{t+1} – прогнозируемый показатель; α – параметр сглаживания; $У_t$ – фактическое значение исследуемого показателя за период, предшествующий прогнозному; U_t – экспоненциально взвешенная средняя для периода, предшествующего прогнозному.

При прогнозировании данным методом возникает два затруднения:

- выбор значения параметра сглаживания α;
- определение начального значения Uo.

От величины α зависит, как быстро снижается вес влияния предшествующих наблюдений. Чем больше α, тем меньше сказывается влияние предшествующих лет. Если значение α близко к единице, то это приводит к учету при прогнозе в основном влияния лишь последних наблюдений. Если значение α близко к нулю, то веса, по которым взвешиваются уровни временного ряда, убывают медленно, т.е. при прогнозе учитываются все (или почти все) прошлые наблюдения.

Таким образом, если есть уверенность, что начальные условия, на основании которых разрабатывается прогноз, достоверны, следует использовать небольшую величину параметра сглаживания (α→0). Когда параметр сглаживания мал, то исследуемая функция ведет себя как средняя из большого числа прошлых уровней. Если нет достаточной уверенности в начальных условиях прогнозирования, то следует использовать большую величину α, что приведет к учету при прогнозе в основном влияния последних наблюдений.

Точного метода для выбора оптимальной величины параметра сглаживания α нет. В отдельных случаях автор данного метода профессор Браун предлагал определять величину α, исходя из длины интервала сглаживания. При этом α вычисляется по формуле:

$$\alpha = \frac{2}{n+1}$$

где n – число наблюдений, входящих в интервал сглаживания.

Задача выбора U_o (экспоненциально взвешенного среднего начального) решается следующими способами:

- если есть данные о развитии явления в прошлом, то можно воспользоваться средней арифметической и приравнять к ней U_o;
- если таких сведений нет, то в качестве U_o используют исходное первое значение базы прогноза $У_1$.

Задача 1

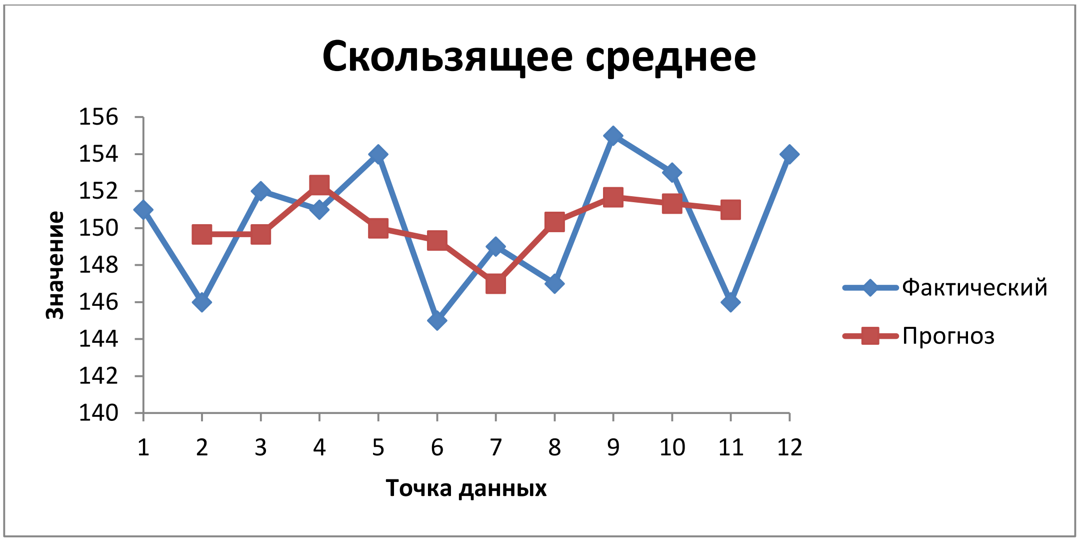
Методом простой (трехмесячной) скользящей средней сгладить имеющиеся данные по производству продукции, изобразить исходный и сглаженный временные ряды на графике.

Месяцы	Производство продукции (тыс. шт.)
январь	151
февраль	146
март	152
апрель	151
май	154
июнь	145
июль	149
август	147
сентябрь	155
октябрь	153
ноябрь	146
декабрь	154

Решение:

Построим сглаженный ряд:

Месяцы	Производство продукции (тыс. шт.)	Сглаженный ряд
январь	151	
февраль	146	149,6666667
март	152	149,6666667
апрель	151	152,3333333
май	154	150
июнь	145	149,3333333
июль	149	147
август	147	150,3333333
сентябрь	155	151,6666667
октябрь	153	151,3333333
ноябрь	146	151
декабрь	154	



<u>Задача 2.</u>

Имеются данные, характеризующие уровень безработицы в регионе, %

Январь	Февраль	Март	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь
2,99	2,66	2,63	2,56	2,40	2,22	1,97	1,72	1,56	1,42

- Постройте прогноз уровня безработицы в регионе на ноябрь, используя метод экспоненциального сглаживания.

- Рассчитайте ошибки полученного прогноза при использовании каждого метода.

Решение:

1) Определяем значение параметра сглаживания по формуле:

$$\alpha = \frac{2}{n+1}$$

где n – число наблюдений, входящих в интервал сглаживания.

$$\alpha = \frac{2}{n+1} = \frac{2}{10+1} \approx 0{,}2$$

2) Определяем начальное значение Uо двумя способами:

I способ (средняя арифметическая):

Uо=(2,99+2,66+2,63+2,56+2,40+2,22+1,97+1,72+1,56+1,42)/10=22,13/10=2,21;

II способ (принимаем первое значение базы прогноза): Uо = 2,99

3) Рассчитываем экспоненциально взвешенную среднюю для каждого периода, используя формулу:

$$U_{t+1} = \alpha * y_t + (1 - \alpha) * U_t$$

где t – период, предшествующий прогнозному; t+1 – прогнозный период; Ut+1 - прогнозируемый показатель; α - параметр сглаживания; Yt - фактическое значение исследуемого показателя за период, предшествующий прогнозному; Ut - экспоненциально взвешенная средняя для периода, предшествующего прогнозному.

Например:

Uфев=2,99*0,2+(1-0,2)*2,21=2,37 (I способ)

Uмарт=2,66*0,2+(1-0,2)*2,37=2,43 (I способ) и т.д.

Uфев.=2,99*0,2+(1-0,2)*2,99=2,99 (II способ)

Uмарт=2,66*0,2+(1-0,2)*2,99=2,92 (II способ)

Uапр = 2,63*0,2+(1-0,2) * 2,92 = 2,86 (II способ) и т.д.

4) По этой же формуле вычисляем прогнозное значение:

Uноябрь=1,42*0,2+(1-0,2)*2,08=1,95 (I способ)

Uноябрь=1,42*0,2+(1-0,2)*2,18=2,03 (II способ)

Результаты заносим в таблицу:

Месяцы	Уровень безработицы, Y_t, %	Экспоненциально взвешенная средняя U_t		Расчетный средней относительной ошибки, $\frac{\lvert Y_ф - Y_р \rvert}{Y_ф} * 100\%$	
		I способ	II способ	I способ	II способ
январь	2,99	2,21	2,99	26,09	0
февраль	2,66	2,37	2,99	10,90	12,41
март	2,63	2,43	2,92	7,60	11,03
апрель	2,56	2,47	2,86	3,52	11,72
май	2,40	2,49	2,80	3,75	16,67
июнь	2,22	2,47	2,72	11,26	22,52
июль	1,97	2,42	2,62	22,84	32,99
август	1,72	2,33	2,49	35,47	44,77

сентябрь	1,56	2,21	2,34	41,67	50
октябрь	1,42	2,08	2,18	46,48	53,52
			Итого:	209,58	255,63
Прогноз ноябрь		1,95	2,03		

5) Рассчитываем среднюю относительную ошибку по формуле:

$$\varepsilon = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\left|y_{ф} - y_{p}\right|}{y_{ф}} * 100\right]$$

$\varepsilon = \frac{209,58}{10} = 20,96\%$ (I способ)

$\varepsilon = \frac{255,63}{10} = 25,56\%$ (II способ)

В каждом случае точность прогноза является удовлетворительной поскольку средняя относительная ошибка попадает в пределы 20-50%.

Задача 3.

Построить тренды временного ряда – линейный, полиномиальный (6 степени), экспоненциальный (с отображением на диаграмме уравнения регрессии и коэффициента детерминации) для следующих данных, характеризующих ВВП на душу населения Швеции:

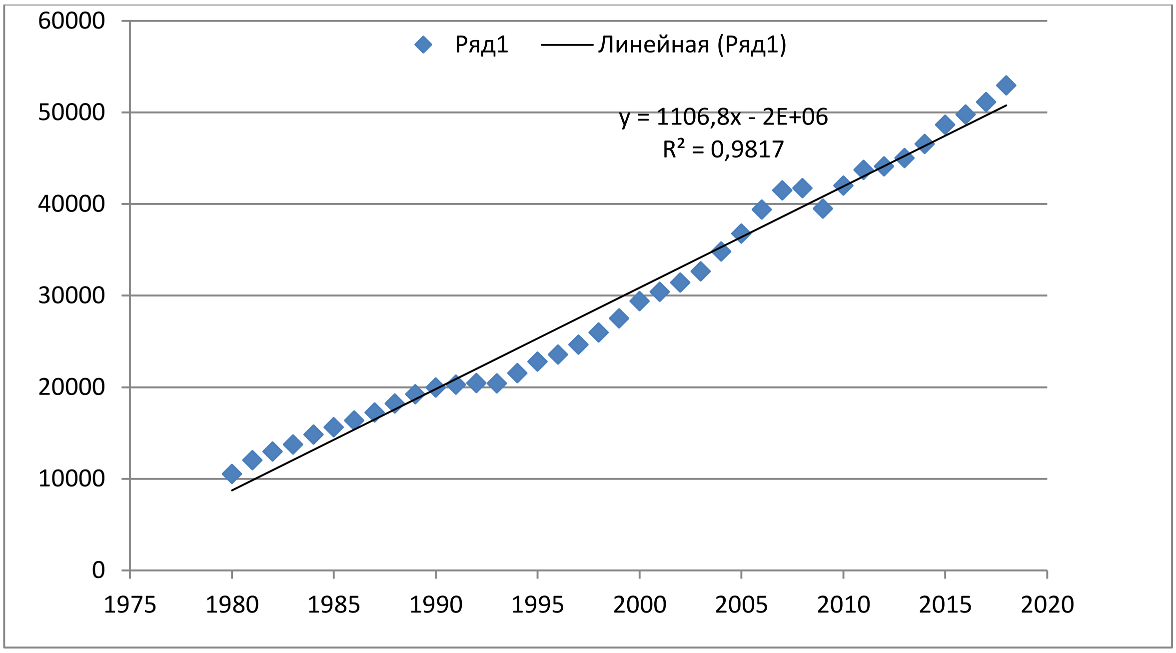
Год	**ВВП на душу населения, $**
1980	10550
1981	12065
1982	12987
1983	13768
1984	14856
1985	15645
1986	16381
1987	17267
1988	18219
1989	19244
1990	19975

1991	20312
1992	20459
1993	20433
1994	21546
1995	22828
1996	23580
1997	24677
1998	25993
1999	27530
2000	29410
2001	30437
2002	31449
2003	32661
2004	34848
2005	36797
2006	39407
2007	41527
2008	41765
2009	39536
2010	42057
2011	43773
2012	44139
2013	45050
2014	46602
2015	48681
2016	49806
2017	51180
2018	52984

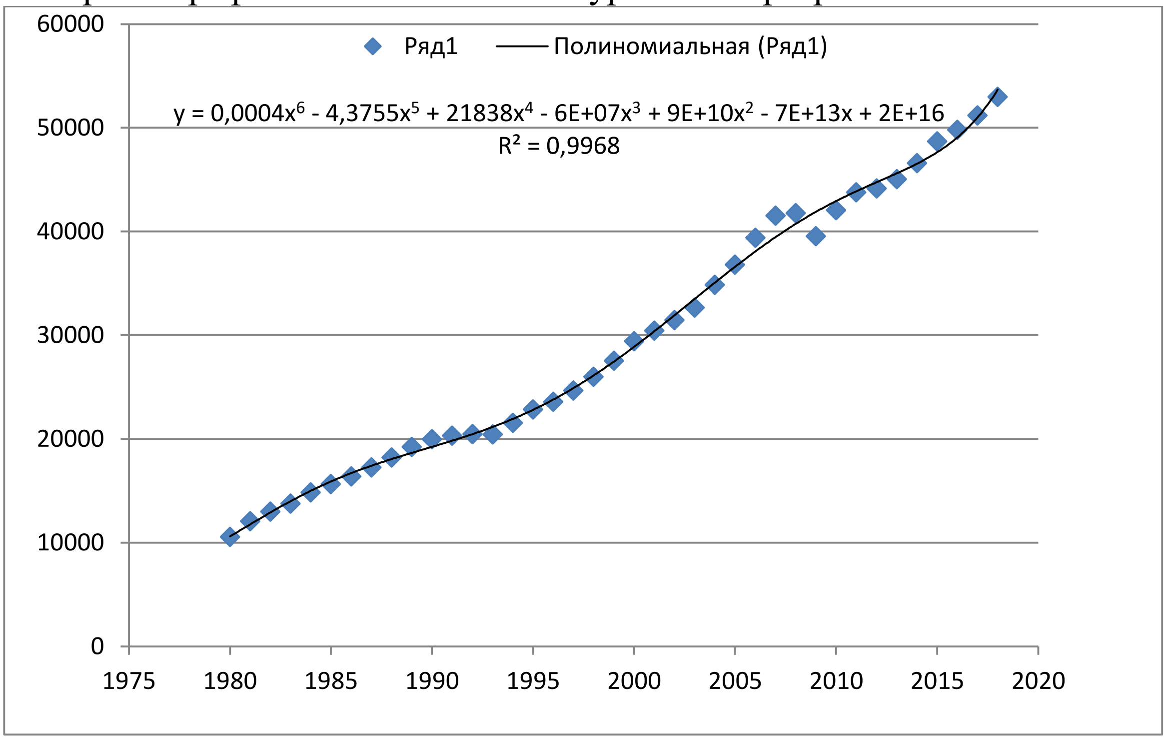
Выбрать уравнение, которое наилучшим образом описывает исходные данные.

Решение:

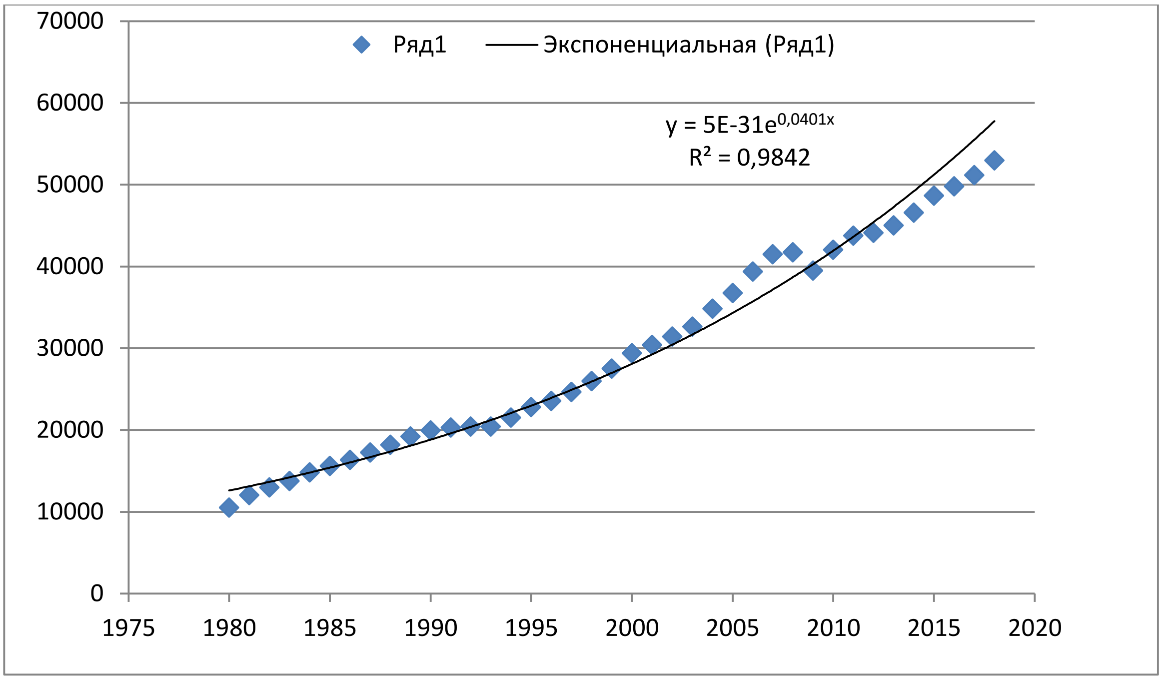
Построим график линейного уравнения регрессии:



Построим график полиномиального уравнения регрессии:



Построим график экспоненциального уравнения регрессии:



Исходя из анализа, делаем вывод, что полинома 6 степени (R^2 =0,9968) наилучшим способом описывает исходные данные ВВП на душу населения Швеции.

Глава 7. Системы эконометрических уравнений

Тема 7.1 Системы одновременных уравнений

Система совместных, одновременных уравнений (или структурная форма модели) обычно содержит эндогенные и экзогенные переменные.

Эндогенные переменные (y) – это зависимые переменные, которые число которых равно числу уравнений в системе.

Экзогенные переменные (x). Это предопределенные переменные, влияющие на эндогенные переменные, но не зависящие от них.

Простейшая структурная форма модели имеет вид:

$$\begin{pmatrix} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1 \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2 \end{pmatrix}$$

Приведенная форма модели представляет собой систему линейных функций эндогенных переменных от экзогенных:

$$\begin{pmatrix} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + \cdots + \delta_{1m}x_m \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + \cdots + \delta_{2m}x_m \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ y_n = \delta_{n1}x_1 + \delta_{n2}x_2 + \cdots + \delta_{nm}x_m \end{pmatrix}$$

где δ_i– коэффициенты приведенной формы модели.

С позиции идентифицируемости структурные модели можно подразделить на три вида:

- идентифицируемые;
- неидентифицируемые;
- сверхидентифицируемые.

Модель идентифицируема, если все ее структурные коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, т. е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные коэффициенты модели оцениваются через параметры приведенной формы модели, и модель идентифицируема.

Модель неидентифицируема, если число приведенных коэффициентов меньше числа структурных коэффициентов, и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Модель сверхидентифицируема, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы.

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых необходимо проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо.Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы. Для того, чтобы уравнение было идентифицируемо, нужно, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного.

Если обозначить число эндогенных переменных в j-м уравнении системы через H, а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, – через D, то условие идентифицируемости модели может быть записано в виде следующего счетного правила:

$D + 1 = H$ – уравнение идентифицируемо;

$D + 1 < H$ – уравнение неидентифицируемо;

$D + 1 > H$ – уравнение сверхидентифицируемо.

<u>Задача 1.</u>

Имеется следующая структурная модель:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2 \\ y_2 = b_{21}y_1 + b_{23}y_3 + a_{22}x_2 \\ y_3 = b_{32}y_2 + a_{31}x_1 + a_{33}x_3 \end{cases}$$

Соответствующая ей приведенная форма модели имеет вид:

$$\begin{cases} y_1 = 3x_1 - 4x_2 + 2x_3 \\ y_2 = 2x_1 + 4x_2 + 5x_3 \\ y_3 = -5x_1 + 6x_2 + 5x_3 \end{cases}$$

Определить, если это возможно, неизвестные параметры структурной модели.

<u>Решение.</u>

Сначала определим идентифицируемость структурной модели. Ограничимся для простоты применением счетного правила. Приведем кратко информацию об этом правиле.

Обозначим H – число эндогенных переменных в *i*- ом уравнении системы, D – число экзогенных переменных, которые содержатся в системе, но не входят в данное уравнение. Тогда условие идентифицируемости уравнения может быть записано в виде следующего счетного правила:

D+1 = H – уравнение идентифицируемо;

D+1 < H – уравнение неидентифицируемо;

D+1 > H – уравнение сверхидентифицируемо.

Первое и третье уравнения структурной модели имеют H = 2, D = 1. В первом уравнении две эндогенные переменные – y_1, y_2, в третьем тоже две –

y_2, y_3; в обоих уравнениях не хватает по одной экзогенной переменной: в первом отсутствует x_3, в третьем – x_2. В этих уравнениях выполняется равенство D + 1 = H, и они идентифицируемы. Во втором уравнении присутствуют все три эндогенные переменные (H=3), а отсутствуют две экзогенные – x_1 и x_3 (D=2). Здесь также выполняется равенство D + 1 = H, и второе уравнение также идентифицируемо. Поскольку все три уравнения структурной модели идентифицируемы, система также идентифицируема.

Для идентифицируемых систем методом оценки структурных параметров является косвенный МНК. Он заключается в том, что уравнения приведенной формы модели (ПФМ), полученные обычным МНК как уравнения множественной регрессии, с помощью алгебраических преобразований превращаются в уравнения структурной формы модели (СФМ). Здесь, как видим, МНК применяется только один раз – для оценки коэффициентов приведенной формы.

Начнем с построения первого уравнения СФМ. Из всех уравнений ПФМ к нему ближе всех по структуре первое уравнение: в обоих уравнениях слева стоит y_1, а справа стоят x_1 и x_2. Однако они отличаются тем, что в первом уравнении ПФМ стоит x_3, а в первом уравнении СФМ стоит y_2. Поэтому, чтобы получить первое уравнение СФМ из первого уравнения ПФМ, надо в последнем заменить x_3 на выражение, в котором появилась бы y_2. Эту замену делаем с помощью второго уравнения ПФМ:

$$x_3 = \frac{1}{5}(y_2 - 2x_1 - 4x_2).$$

Подставим в первое уравнение ПФМ, получаем после элементарных преобразований:

$$y_1 = 3x_1 - 4x_2 + 2\left[\frac{1}{5}(y_2 - 2x_1 - 4x_2)\right],$$

или

$$y_1 = 0{,}4y_2 + 2{,}2x_1 - 5{,}64x_2$$

.

Это и есть первое уравнение СФМ.

Для получения третьего уравнения СФМ действуем аналогично: в третьем уравнении ПФМ заменяем x_2 так, чтобы в результате замены появилась y_2. такую замену снова делаем через второе уравнение ПФМ:

$$x_2 = \frac{1}{4}(y_2 - 2x_1 - 5x_3).$$

.

Подставим в третье уравнение ПФМ, получаем:

$$y_3 = -5x_1 + 6\left[\frac{1}{4}(y_2 - 2x_1 - 5x_3)\right] + 5x_3,$$

или

$$y_3 = 1{,}5y_2 - 8x_1 - 2{,}5x_3.$$

Это и есть третье уравнение СФМ.

Для получения второго уравнения СФМ требуются более сложные преобразования. Это связано с тем, что из второго уравнения ПФМ, как наиболее похожего на второе уравнение СФМ, надо исключить сразу две переменные – x_1 и x_3, чтобы при этом появились y_1 и y_3.. Последовательное исключение здесь не годится, их надо исключать одновременно. Для этого запишем первое и третье уравнения ПФМ как систему относительно исключаемых переменных:

$$\begin{cases} 3x_1 + 2x_3 = y_1 + 4x_2 \\ -5x_1 + 5x_3 = y_3 - 6x_2 \end{cases}$$

Решаем эту систему любым способом, например, например, методом определителей:

$$\Delta = \begin{pmatrix} 3 & 2 \\ -5 & 5 \end{pmatrix} = 25$$

$$\Delta_1 = \begin{vmatrix} y_1 + 4x_2 & 2 \\ y_3 - 6x_2 & 5 \end{vmatrix} = 5(y_1 + 4x_2) - 2(y_3 - 6x_2) = 5y_1 - 2y_3 + 32x_2$$

$$\Delta_3 = \begin{vmatrix} 3 & y_1 + 4x_2 \\ -5 & y_3 - 6x_2 \end{vmatrix} = 3(y_3 + 6x_2) - (-5)(y_1 + 4x_2) = 5y_1 + 3y_3 + 2x_2$$

$$x_1 = \frac{\Delta_1}{\Delta} = 0{,}2y_1 - 0{,}08y_3 + 1{,}28x_2$$

$$x_3 = \frac{\Delta_3}{\Delta} = 0{,}2y_1 + 0{,}12y_3 + 0{,}08x_2$$

Подставим полученные решения во второе уравнение ПФМ, получаем второе уравнение СФМ:

$$y_2 = 2(0{,}2y_1 - 0{,}08y_3 + 1{,}28x_2) + 4x_2 + 5(0{,}2y_1 + 0{,}12y_3 + 0{,}08x_{2)}$$

или

$$y_2 = 1{,}4y_1 + 0{,}44y_3 + 6{,}96x_{2)}$$

Теперь можем полностью записать структурную модель:

$$\begin{cases} y_1 = 0{,}4y_2 + 2{,}2x_1 - 5{,}6x_2 \\ y_2 = 1{,}4y_1 + 0{,}44y_3 + 6{,}96x_2 \\ y_3 = 1{,}5y_2 - 8x_1 - 2{,}5x_3 \end{cases}$$

■