МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

НАБЕРЕЖНОЧЕЛНИНСКИЙ ИНСТИТУТ (ФИЛИАЛ) ФЕДЕРАЛЬНОГО
ГОСУДАРСТВЕННОГО АВТОНОМНОГО ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

КАФЕДРА «БИЗНЕС-ИНФОРМАТИКА И МАТЕМАТИЧЕСКИЕ МЕТОДЫ
В ЭКОНОМИКЕ»

ЭКОНОМЕТРИКА

Электронный образовательный ресурс

2018

УДК 330
ББК 65
П 12

Печатается по решению учебно-методической комиссии
Набережночелнинского филиала
федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ»

протокол № 9 от « 19 » мая 2014 г.

Рецензент:
Макаров А.Н., филиал федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет" в г. Набережные Челны, докт. экон. наук, профессор

С.В. Павликов, А.Г. Исавнин. Эконометрика. Электронный образовательный ресурс. – Набережные Челны, 2018. – 94 с.: ил.

Электронный образовательный ресурс по эконометрике содержит контрольные задания из 20 вариантов по следующим пяти разделам эконометрики: парная регрессия; нелинейные модели; множественная регрессия; системы эконометрических уравнений; временные ряды. Для каждого раздела приводятся методические указания к решению задач, разобраны примеры решения задач. Ресурс соответствует требованиям государственного стандарта по эконометрике, содержит программу курса, вопросы для самопроверки, таблицы критических точек распределения Стьюдента и Фишера.

Электронный образовательный ресурс адресован студентам экономических направлений вузов различных форм обучения. Контрольные задания пособия могут использоваться для составления контрольных работ и типовых расчетов по эконометрике.

УДК 330
ББК 65

ОДЕРЖАНИЕ

ВВЕДЕНИЕ

Моделирование экономических процессов сопряжено с рядом трудностей. Это и многообразие экономической жизни и конфликт интересов различных социальных групп и внешний фактор в силу открытости современной экономики. Возникает определенный пессимизм по отношению к возможностям и полезности количественного моделирования, стремление к качественному описанию взаимосвязей экономических величин. Тем не менее, конкретные решения, влекущие материальную ответственность, не могут опираться на качественные рассуждения и требуют точных вычислений. Востребованные практикой средства анализа данных, на которые можно опираться в процессе принятия решений, предоставляет эконометрика. В этой науке соединились возможности экономической теории и математики.

В предлагаемом пособии представлены контрольные задания по пяти основным темам, соответствующим стандарту по эконометрике. В первом задании предлагается построить линейную регрессионную модель с одним фактором, влияющим на результат (парная регрессия). В рамках построенной модели требуется получить оценки параметров, оценить качество модели. Во втором задании необходимо построить нелинейную модель с теми же исходными данными. В третьем задании число регрессоров (факторов) увеличивается до двух (множественная регрессия). Необходимо оценить параметры, проверить нулевые гипотезы относительно значений параметров, оценить качество множественной регрессии, сделать прогноз. В четвертом задании предлагается идентифицировать параметры системы одновременных уравнений. В пятом задании изучается автокорреляционная функция временного ряда, и прогнозирование в условиях авторегрессии.

По каждому из пяти заданий предлагаются методические указания, которые включают теоретические выкладки и примеры решения эконометрических задач. Уровень сложности предлагаемых заданий и относительно небольшое количество наблюдений позволяют выполнить предлагаемую работу с помощью обычного калькулятора. Допускается использование специализированных пакетов программ, например оболочки EXCEL.

Выполнение работы следует начинать с проработки методических указаний, параллельно изучая теорию в соответствии со стандартом и рабочей программой курса. Затем выполняются задания своего варианта. При подготовке к экзамену, рекомендуется письменно ответить на вопросы для самопроверки.

1. ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ КОНТРОЛЬНОЙ РАБОТЫ. ВЫБОР ВАРИАНТА

Контрольную работу следует выполнять в ученических тетрадях или на отдельных листах. Титульный лист включает: название института, название кафедры, название контрольной работы, фамилию, имя, отчество и личный шифр студента. Задания могут выполняться с применением компьютера. Вычисления производятся с точностью до трех знаков после запятой.

Номер варианта N определяется по двум последним цифрам nn Вашей зачетной книжки по правилу:

$N=20$, если $nn=00$, $N=nn$, если $00<nn\leq 20$; $N=nn-20$, если $20<nn\leq 40$; $N=nn-40$, если $40<nn\leq 60$, и т.д.

2. КОНТРОЛЬНЫЕ ЗАДАНИЯ

Задание 1

Линейная регрессионная модель

В табл.1 даны наблюдения $\mathbf{x_t}$ и $\mathbf{y_t}$. Предполагается, что зависимую переменную **y** и независимую **x** связывает линейное регрессионное уравнение $\mathbf{y_t=a+b\cdot x_t+\varepsilon_t}$,
где **a** и **b** неизвестные параметры уравнения, $\boldsymbol{\varepsilon_t}$ – случайные отклонения.

1. Постройте диаграмму рассеяния наблюдений и визуально проверьте гипотезу о возможной линейной зависимости между **x** и **y**;
2. По методу наименьших квадратов (МНК) определите оценки параметров **a** и **b** линейной регрессионной модели;
3. На диаграмме рассеяния постройте график прогнозных значений $\hat{y}_t = \hat{a} + \hat{b} \cdot x_t$, где $\hat{a}$ - оценка параметра **a**, а $\hat{b}$ - оценка параметра **b**.
4. Вычислите оценку дисперсии остатков. Оцените дисперсию $\hat{a}$ и $\hat{b}$;
5. С уровнем значимости 0,05 проверьте гипотезу **a**=100 и гипотезу **b**=0.
6. Постройте 95% доверительные интервалы для параметров **a** и **b**.
7. Определите коэффициент детерминации $\mathbf{R^2}$, качественно оценить тесноту связи между **x** и **y**;
8. Вычислите дисперсионное отношение **F**, с уровнем значимости 0,05 проверьте гипотезу о наличии связи между **x** и **y**;
9. Определите прогнозное значение $\hat{y}_{11}$ при $\mathbf{x_{11}=N}$, где **N** – номер Вашего варианта. Постройте 95% доверительный интервал для найденного прогнозного значения.
10. Оцените с помощью эластичности силу влияния фактора на результат в точке $\mathbf{x_{11}}$.

Задание 2

Нелинейная модель. Линеаризация

Для тех же наблюдений $\mathbf{x_t}$ и $\mathbf{y_t}$, предполагается, что зависимую переменную **y** и независимую **x** связывает нелинейное регрессионное уравнение:

$y_t = a + b \cdot \dfrac{1}{x_t} + \varepsilon_t$	для вариантов	1,6,11,16;
$y_t = a \cdot b^{x_t} \cdot \varepsilon_t$	для вариантов	2,7,12,17;
$\ln y_t = a + b \cdot \dfrac{1}{x_t} + \varepsilon_t$	для вариантов	3,8,13,18;

$y_t = a \cdot e^{b \cdot x_t} \cdot \varepsilon_t$ для вариантов 4,9,14,19;

$y_t = a \cdot x_t^{\ b} \cdot \varepsilon_t$ для вариантов 5,10,15,20.

1. Проведите линеаризацию модели, определите оценки параметров нелинейной модели.
2. Оцените качество модели с помощью коэффициента детерминации и дисперсионного отношения **F**.
3. Определите прогнозное значение $\hat{y}_{11}$ при $\mathbf{x_{11}=N}$, где **N** – номер Вашего варианта. Постройте 95% доверительный интервал для прогноза.
4. Оцените с помощью эластичности силу влияния фактора на результат в точке $\mathbf{x_{11}}$.
5. На диаграмме рассеяния постройте график прогнозных значений. Определите сумму квадратов отклонений наблюдений от нелинейного прогноза.

Таблица 1

t	Вариант 1		Вариант 2		Вариант 3		Вариант 4		Вариант 5	
	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t
1	9	31	15	60	13	25	14	110	19	26
2	11	31	17	84	9	22	18	136	11	33
3	15	35	7	19	15	25	16	125	6	42
4	6	28	17	75	15	25	8	84	9	36
5	17	33	15	59	18	27	19	140	15	28
6	13	31	16	65	6	18	6	77	17	28
7	14	34	14	50	11	24	16	120	18	28
8	16	34	13	55	14	25	13	100	11	34
9	7	28	6	32	18	27	8	84	19	25
10	11	31	15	53	8	21	6	75	11	33
t	Вариант 6		Вариант 7		Вариант 8		Вариант 9		Вариант 10	
	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t
1	9	131	7	16	9	25	12	97	10	33
2	7	130	11	26	18	31	6	74	19	25
3	16	144	14	38	20	31	14	107	14	31
4	11	140	9	25	17	29	13	108	17	28
5	17	157	7	21	6	17	18	128	19	27
6	8	123	15	40	7	23	14	113	14	29
7	19	155	10	24	7	21	20	141	6	43

8	18	159	12	32	8	21	14	109	16	28
9	12	145	6	18	20	30	10	88	9	39
10	14	153	7	19	18	30	14	112	10	36

t	Вариант 11		Вариант 12		Вариант 13		Вариант 14		Вариант 15	
	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t
1	16	32	13	52	9	22	5	72	19	26
2	19	34	15	66	8	20	18	136	19	25
3	12	31	15	62	20	27	10	88	13	30
4	16	32	14	61	9	22	7	75	19	27
5	14	34	6	26	15	26	15	114	7	41
6	6	29	19	89	15	26	10	97	17	26
7	13	32	11	41	14	26	5	73	9	37
8	10	32	14	54	19	26	8	84	18	27
9	12	34	7	25	18	27	19	143	8	38
10	20	35	16	70	19	27	7	79	9	37

t	Вариант 16		Вариант 17		Вариант 18		Вариант 19		Вариант 20	
	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t	x_t	y_t
1	17	158	7	23	9	24	8	84	19	27
2	15	138	6	18	11	28	19	140	16	27
3	8	129	14	36	17	32	6	77	19	26
4	5	114	7	21	19	27	16	120	14	29
5	16	155	16	39	14	32	13	100	17	29
6	14	151	19	52	9	27	8	84	14	32
7	20	163	14	37	19	33	6	75	12	31
8	16	145	9	20	16	31	12	97	6	45
9	12	144	16	41	12	27	6	74	14	30
10	12	148	6	17	6	21	14	107	19	26

Задание 3

Множественная регрессия

К тем же наблюдениям $\mathbf{x_t}$, и $\mathbf{y_t}$. добавляются значения $\mathbf{z_t} = \sqrt{x_t}$. Предполагается, что зависимую переменную **y** и факторы связывает уравнение множественной линейной регрессии

$\mathbf{y_t = a + b \cdot x_t + c \cdot z_t + \varepsilon_t}$,

где **a**, **b** и **c** неизвестные параметры уравнения, $\boldsymbol{\varepsilon_t}$ – случайные отклонения.

1. Определите МНК оценки параметров уравнения.
2. С уровнем значимости 0,05 проверьте гипотезу **b**=0 (о влиянии фактора **x** на результат) и **c**=0 (о влиянии фактора **z** на результат).
3. Определите коэффициент детерминации и скорректированный коэффициент детерминации.
4. По критерию Фишера F с уровнем значимости 0,05 оцените качество модели в целом.
5. Составьте корреляционную таблицу наблюдений и вычислите частные коэффициенты корреляции.
6. Сравните по качеству модели заданий 1, 2 и 3.

Задание 4

Системы регрессионных уравнений

В каждом из заданий (табл.2) предлагается структурная система эконометрических уравнений, приведенная система уравнений и данные наблюдений.

1. Определите к какому типу относится каждое из уравнений структурной системы эконометрических уравнений (идентифицируемо, неиндентифицируемо или сверхидентифицируемо).
2. Опираясь на данные наблюдений и построенную на их основе приведенную систему эконометрических уравнений, проведите идентификацию параметров структурной системы.

Таблица 2

Вариант 1.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{23}x_{t3} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 2 - x_{t1} + 3x_{t2} - x_{t3}$$
$$\hat{y}_{t2} = -1 + 2x_{t1} - 2x_{t2} + x_{t3}$$
$$\hat{y}_{t3} = 5 + 2x_{t1} - 4x_{t2} + 3x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	14	-8	-5	1	5	3
2	6	0	3	2	3	2
3	20	-7	-8	4	8	3
4	-5	14	23	6	2	5
5	6	2	14	2	4	7

Вариант 2.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + b_{13}y_{t3} + a_{13}x_{t3} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 5 + 3x_{t1} + 2x_{t2} + 5x_{t3}$$
$$\hat{y}_{t2} = 1 + 4x_{t1} - 2x_{t2} - 2x_{t3}$$
$$\hat{y}_{t3} = -1 - 3x_{t1} + x_{t2} + x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	26	-5	-1	1	1	3
2	19	4	-4	2	2	1
3	47	-4	-1	4	7	3
4	32	19	-16	6	3	1
5	48	-5	-1	3	4	5

Вариант 3.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + a_{12}x_{t2} + a_{13}x_{t3} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{21}x_{t1} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{32}y_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 2 - 3x_{t1} + x_{t2} + 4x_{t3}$$
$$\hat{y}_{t2} - 4 + 2x_{t1} - x_{t2} + 2x_{t3}$$
$$\hat{y}_{t3} = -3 - 2x_{t1} + 2x_{t2} + 5x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	-6	3	-1	1	5	3
2	15	0	20	2	3	2
3	7	-1	17	4	8	3
4	4	18	13	6	2	5
5	29	20	38	2	4	7

Вариант 4.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + a_{21}x_{t1} + a_{23}x_{t3} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + b_{32}y_{t2} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = -3 + 4x_{t1} + x_{t2} - 5x_{t3}$$
$$\hat{y}_{t2} = 1 - 2x_{t1} - 2x_{t2} + 3x_{t3}$$
$$\hat{y}_{t3} = 5 - 1x_{t1} - x_{t2} - x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	11	-12	-9	1	1	3
2	-1	-6	-4	2	2	1
3	3	-13	-7	4	7	3
4	-5	4	-8	6	3	1
5	-26	12	-7	3	4	5

Вариант 5.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{23}x_{t3} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + x_{t1} + 2x_{t2} + 2x_{t3}$$
$$\hat{y}_{t2} = -1 + 3x_{t1} + x_{t2} + 3x_{t3}$$
$$\hat{y}_{t3} = 2 + 3x_{t1} + 2x_{t2} + 3x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	26	35	44	1	5	3
2	13	17	21	2	3	2
3	14	16	25	4	8	3
4	13	23	24	6	2	5
5	20	17	26	2	4	7

Вариант 6.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + b_{13}y_{t3} + a_{13}x_{t3} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + 3x_{t1} + x_{t2} - 2x_{t3}$$
$$\hat{y}_{t2} = -2 + 2x_{t1} + 3x_{t2} + 2x_{t3}$$
$$\hat{y}_{t3} = 1 + 3x_{t1} + 4x_{t2} + 5x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	13	25	47	1	1	3
2	15	18	33	2	2	1
3	16	35	62	4	7	3
4	11	27	50	6	3	1
5	10	16	28	3	4	5

Вариант 7.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + a_{12}x_{t2} + a_{13}x_{t3} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{21}x_{t1} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{32}y_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + 2x_{t1} + 2x_{t2} + 2x_{t3}$$
$$\hat{y}_{t2} = 2 - 1x_{t1} - 2x_{t2} + x_{t3}$$
$$\hat{y}_{t3} = -1 - 4x_{t1} + 4x_{t2} + 3x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	26	-11	11	1	5	3
2	30	-3	-18	2	3	2
3	14	-4	15	4	8	3
4	29	-9	37	6	2	5
5	16	-4	9	2	4	7

Вариант 8.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + a_{21}x_{t1} + a_{23}x_{t3} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + b_{32}y_{t2} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + 1x_{t1} + 4x_{t2} + 2x_{t3}$$
$$\hat{y}_{t2} = -2 - 1x_{t1} + 1x_{t2} - 3x_{t3}$$
$$\hat{y}_{t3} = 2 + 2x_{t1} + x_{t2} + 2x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	30	-10	16	1	1	3
2	24	-12	25	2	2	1
3	44	-15	29	4	7	3
4	23	-19	25	6	3	1
5	16	-15	15	3	4	5

Вариант 9.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{23}x_{t3} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Вариант 10.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + b_{13}y_{t3} + a_{13}x_{t3} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 2 + 2x_{t1} + x_{t2} + 2x_{t3}$$
$$\hat{y}_{t2} = 3 - 3x_{t1} - 2x_{t2} + x_{t3}$$
$$\hat{y}_{t3} = 1 + x_{t1} + x_{t2} + 3x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	16	-9	15	1	5	3
2	12	-6	12	2	3	2
3	25	-23	24	4	8	3
4	24	-12	23	6	2	5
5	25	-4	28	2	4	7

Вариант 11.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + a_{12}x_{t2} + a_{13}x_{t3} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{21}x_{t1} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{32}y_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 6 - 3x_{t1} + x_{t2} + 4x_{t3}$$
$$\hat{y}_{t2} = 2 + x_{t1} - x_{t2} + 2x_{t3}$$
$$\hat{y}_{t3} = -3 - 2x_{t1} + 2x_{t2} + 3x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	-2	3	-3	1	5	3
2	19	5	14	2	3	2
3	11	1	13	4	8	3
4	8	18	3	6	2	5
5	33	22	20	2	4	7

Вариант 13.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{23}x_{t3} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 2 + x_{t1} + 3x_{t2} - 4x_{t3}$$
$$\hat{y}_{t2} = -4 + 5x_{t1} - 3x_{t2} + 3x_{t3}$$
$$\hat{y}_{t3} = 3 + x_{t1} + 4x_{t2} + 3x_{t3}$$

Приведенная система:

$$\hat{y}_{t1} = 2 + x_{t1} + 2x_{t2} + 2x_{t3}$$
$$\hat{y}_{t2} = 6 - 4x_{t1} - 5x_{t2} + 2x_{t3}$$
$$\hat{y}_{t3} = -2 + 3x_{t1} + 4x_{t2} + x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	12	1	7	1	1	3
2	9	-9	13	2	2	1
3	27	-40	43	4	7	3
4	14	-29	28	6	3	1
5	24	-16	28	3	4	5

Вариант 12.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + a_{21}x_{t1} + a_{23}x_{t3} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + b_{32}y_{t2} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 2 + 4x_{t1} + x_{t2} - 3x_{t3}$$
$$\hat{y}_{t2} = -3 + 2x_{t1} - 2x_{t2} - 2x_{t3}$$
$$\hat{y}_{t3} = 5 - 1x_{t1} + 4x_{t2} - x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	22	-7	11	1	1	3
2	8	-12	21	2	2	1
3	14	-20	33	4	7	3
4	10	-1	-3	6	3	1
5	-7	-19	8	3	4	5

Вариант 14.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + b_{13}y_{t3} + a_{13}x_{t3} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + 3x_{t1} + 5x_{t2} - 5x_{t3}$$
$$\hat{y}_{t2} = 2 + 3x_{t1} + 2x_{t2} + 2x_{t3}$$
$$\hat{y}_{t3} = 1 - 5x_{t1} - 4x_{t2} + 5x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	14	28	39	1	5	3
2	-8	8	24	2	3	2
3	12	9	24	4	8	3
4	-15	20	23	6	2	5
5	15	-6	35	2	4	7

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	24	28	-25	1	1	3
2	13	27	-23	2	2	1
3	39	35	-34	4	7	3
4	0	37	-14	6	3	1
5	23	18	-20	3	4	5

Вариант 15.

Структурная система:

$$y_{t1} = b_{10} + b_{12} y_{t2} + a_{12} x_{t2} + a_{13} x_{t3} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21} y_{t1} + b_{23} y_{t3} + a_{21} x_{t1} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{32} y_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + 2x_{t1} - 3x_{t2} - 4x_{t3}$$
$$\hat{y}_{t2} = 1 - 2x_{t1} - 2x_{t2} + x_{t3}$$
$$\hat{y}_{t3} = -2 - 4x_{t1} + 2x_{t2} + 3x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	-17	-16	0	1	5	3
2	-5	-13	-21	2	3	2
3	-13	-6	8	4	8	3
4	-41	-12	20	6	2	5
5	-11	-7	2	2	4	7

Вариант 16.

Структурная система:

$$y_{t1} = b_{10} + b_{12} y_{t2} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21} y_{t1} + a_{21} x_{t1} + a_{23} x_{t3} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31} y_{t1} + b_{32} y_{t2} + a_{32} x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 3 + x_{t1} + 3x_{t2} + x_{t3}$$
$$\hat{y}_{t2} = -3 - x_{t1} - x_{t2} + 2x_{t3}$$
$$\hat{y}_{t3} = 1 - x_{t1} + 4x_{t2} + 2x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	24	-6	24	1	1	3
2	21	-9	9	2	2	1
3	34	-5	37	4	7	3
4	18	1	12	6	3	1
5	13	2	11	3	4	5

Вариант 17.

Структурная система:

$$y_{t1} = b_{10} + b_{12} y_{t2} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21} y_{t1} + b_{23} y_{t3} + a_{23} x_{t3} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31} y_{t1} + a_{31} x_{t1} + a_{32} x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = -1 + 2x_{t1} + 4x_{t2} + 2x_{t3}$$
$$\hat{y}_{t2} = 3 - 3x_{t1} - 2x_{t2} + 4x_{t3}$$
$$\hat{y}_{t3} = -2 + x_{t1} + 3x_{t2} + 3x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	28	0	22	1	5	3
2	18	0	15	2	3	2
3	46	-14	37	4	8	3
4	27	3	24	6	2	5
5	34	17	33	2	4	7

Вариант 18.

Структурная система:

$$y_{t1} = b_{10} + b_{12} y_{t2} + b_{13} y_{t3} + a_{13} x_{t3} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21} y_{t1} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31} y_{t1} + a_{31} x_{t1} + a_{32} x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + x_{t1} + 2x_{t2} + 4x_{t3}$$
$$\hat{y}_{t2} = 5 + 4x_{t1} + 6x_{t2} + 2x_{t3}$$
$$\hat{y}_{t3} = -4 - 3x_{t1} - 5x_{t2} + x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	17	19	-10	1	1	3
2	10	28	-19	2	2	1
3	32	68	-46	4	7	3
4	15	51	-37	6	3	1
5	33	51	-28	3	4	5

Вариант 19.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + a_{12}x_{t2} + a_{13}x_{t3} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{21}x_{t1} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{32}y_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 - 3x_{t1} + 3x_{t2} + x_{t3}$$
$$\hat{y}_{t2} = -3 - 2x_{t1} - x_{t2} + 2x_{t3}$$
$$\hat{y}_{t3} = 1 - 2x_{t1} + 2x_{t2} + x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	0	-20	-1	1	5	3
2	15	-3	12	2	3	2
3	16	-16	13	4	8	3
4	-8	-5	-3	6	2	5
5	5	5	6	2	4	7

Вариант 20.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$
$$y_{t2} = b_{20} + b_{21}y_{t1} + a_{21}x_{t1} + a_{23}x_{t3} + \varepsilon_{t2}$$
$$y_{t3} = b_{30} + b_{31}y_{t1} + b_{32}y_{t2} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 3 + 4x_{t1} - 2x_{t2} - 5x_{t3}$$
$$\hat{y}_{t2} = -2 - 2x_{t1} + 3x_{t2} + 3x_{t3}$$
$$\hat{y}_{t3} = 5 - x_{t1} + 4x_{t2} - x_{t3}$$

Данные наблюдений:

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	5	5	11	1	1	3
2	-10	16	21	2	2	1
3	-15	24	33	4	7	3
4	-2	6	-3	6	3	1
5	-29	24	8	3	4	5

Задание 5

Временные ряды. Авторегрессия

В табл.3 представлены наблюдения временного ряда.

Таблица 3

Варианты 1 - 20																				
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t	y_t
1	50	52	49	49	52	49	52	50	54	52	100	100	98	100	103	90	98	101	100	98
2	50	50	49	51	55	47	51	51	55	46	100	99	101	101	101	89	98	99	101	101
3	52	52	49	48	47	48	52	52	55	44	101	100	103	103	106	90	99	97	99	102
4	51	48	51	50	48	49	51	51	52	51	104	102	98	99	104	92	98	100	97	99
5	51	51	49	50	55	49	51	50	53	48	104	100	102	101	101	90	100	96	97	101
6	51	49	50	52	53	50	52	49	52	50	103	102	102	102	104	89	99	95	98	95
7	52	52	51	50	56	49	51	51	53	44	103	101	99	104	103	88	100	97	98	100
8	51	51	52	53	50	49	51	49	49	41	103	103	101	102	109	89	98	96	95	101
9	52	52	51	51	50	49	49	49	52	47	103	101	101	103	107	89	100	94	96	98
10	51	52	50	55	58	50	50	49	52	46	103	104	100	104	102	88	98	96	96	96
11	51	53	51	51	56	49	46	46	51	47	104	101	102	106	107	88	100	97	97	93
12	52	54	52	53	59	51	48	49	48	42	103	104	102	105	106	86	98	92	95	99
13	52	53	51	53	53	50	48	49	52	40	102	103	102	104	111	88	99	94	95	97
14	54	54	53	58	54	49	49	46	50	45	102	104	103	106	110	88	96	96	94	96
15	52	53	52	54	60	48	48	48	52	43	101	105	103	108	105	89	98	92	95	94
16	53	54	51	56	58	50	48	48	47	45	102	103	104	107	108	88	97	95	93	91
17	53	54	54	55	62	48	48	46	50	41	102	106	105	107	106	87	96	94	94	96
18	51	54	54	59	56	47	46	46	47	38	104	104	104	108	113	86	96	92	93	95
19	51	55	52	56	57	48	46	47	49	41	104	104	105	110	111	87	96	95	93	94
20	50	55	55	57	62	48	45	46	45	43	103	103	105	108	108	87	95	93	92	92
21	50	55	54	58	60	48	46	45	47	40	103	103	105	108	112	87	97	92	93	89
22	51	56	53	60	65	49	45	47	46	37	102	102	106	110	110	86	94	93	91	90
23	54	55	55	57	59	47	46	45	46	36	102	104	108	111	114	87	96	93	91	91
24	54	57	57	59	60	48	46	45	44	38	103	106	107	111	113	88	94	92	89	93
25	53	55	57	61	64	47	47	46	45	38	104	105	107	110	111	88	95	94	91	90
26	53	56	55	62	62	47	46	45	44	36	104	106	108	111	115	87	93	94	90	87
27	53	56	57	59	67	45	46	45	44	33	107	105	108	112	113	85	93	91	88	87
28	53	55	57	59	63	46	46	44	41	34	105	106	108	111	115	87	94	93	87	89
29	53	55	56	63	62	45	46	43	42	34	106	106	107	111	116	88	93	91	89	90
30	54	55	59	62	68	46	45	44	42	35	104	106	108	112	114	87	95	90	90	87

1. Постройте диаграмму наблюдений временного ряда. Определите для него линейный тренд. Вычислите отклонения наблюдений от тренда (остатки регрессии). Установите, является ли данный тренд значимым.

2. Определите и постройте выборочную автокорреляционную функцию остатков (r_i для i=1,2,..,5). Установите пиковое значение автокорреляционной функции. Постройте соответствующую найденному пиковому значению модель временного ряда с корреляцией остатков. Оцените качество построенной модели.
3. С помощью построенной модели сделайте прогноз для следующих за тридцатым пяти наблюдений временного ряда.

3. МЕТОДИЧЕСКИЕ УКАЗАНИЯ

3.1. Линейная регрессионная модель

Для анализа работы торгового предприятия произведено 10 наблюдений числа покупателей $\mathbf{x_t}$ и выручки $\mathbf{y_t}$ (табл.4):

Таблица 4

t	1	2	3	4	5	6	7	8	9	10
$\mathbf{x_t}$	31	75	89	26	35	73	91	21	56	21
$\mathbf{y_t}$	64	100	103	50	63	95	109	43	93	37

Предполагается, что зависимую переменную (выручку) и независимую (число покупателей в магазине) связывает линейное регрессионное уравнение $y_t = a + bx_t + \varepsilon_t$.

1. Построим диаграмму рассеяния наблюдений (рис.1), откладывая на координатной плоскости 10 точек с координатами (31; 64), (75; 100),…, (21; 37):

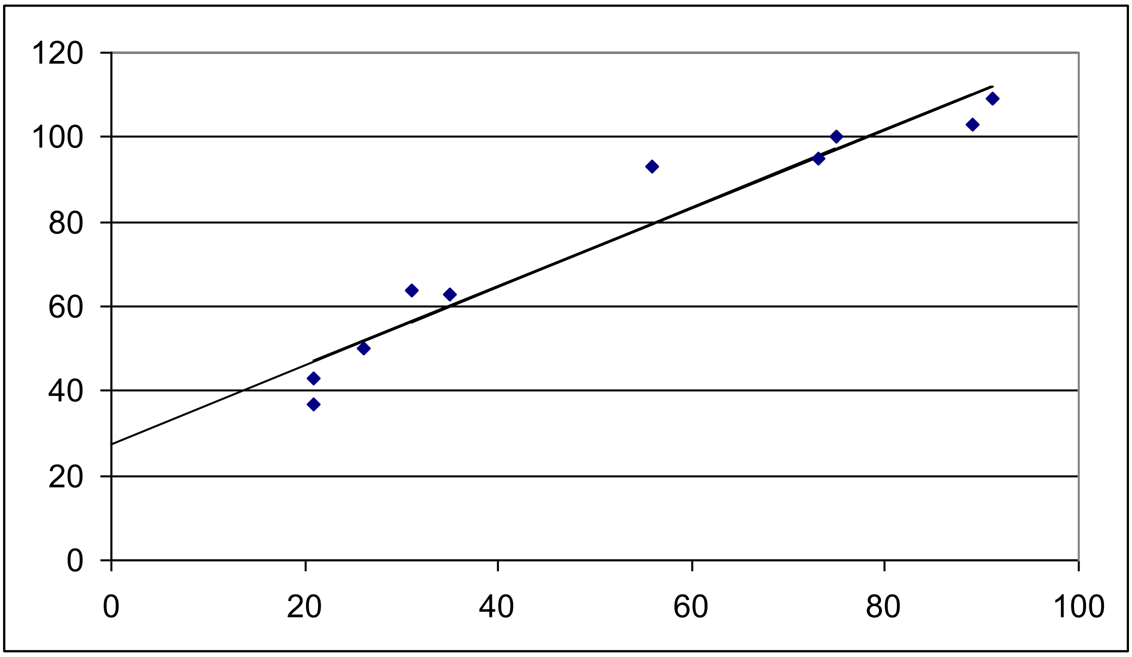


Рис.1 Диаграмма рассеяния наблюдений (точки), линейный тренд (сплошная прямая).

По типу диаграммы рассеяния можно предположить, что между наблюдениями **x** и **y** существует линейная зависимость.

2. Применяя метод наименьших квадратов, получим оценки параметров **a** и **b** линейной регрессионной модели:

Оценка параметра **b** вычисляется по формуле

$$\hat{b} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{4589{,}8 - 51{,}8 \cdot 75{,}7}{3403{,}6 - 51{,}8^2} = 0{,}928,$$

где $\overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t = 51{,}8$; $\overline{y} = \frac{1}{n}\sum_{t=1}^{n} y_t = 75{,}7$; $\overline{x^2} = \frac{1}{n}\sum_{t=1}^{n} x_t^2 = 3403{,}6$; $\overline{xy} = \frac{1}{n}\sum_{t=1}^{n} x_t y_t = 4589{,}8$;

n – число наблюдений. В представленном примере n=10.

Оценка параметра **a** вычисляется по формуле

$$\hat{a} = \overline{y} - \hat{b} \cdot \overline{x} = 75{,}7 - 0{,}928 \cdot 51{,}8 = 27{,}626.$$

Для получения оценок параметров модели удобно использовать табл.5:

Таблица 5

t	x_t	y_t	x_t^2	y_t^2	$x_t y_t$	$\hat{y}_t$	e_t	e_t^2
1	31	64	961	4096	1984	56,396	7,604	57,817
2	75	100	5625	10000	7500	97,231	2,769	7,667
3	89	103	7921	10609	9167	110,224	-7,224	52,186

4	26	50	676	2500	1300	51,756	-1,756	3,083
5	35	63	1225	3969	2205	60,109	2,891	8,361
6	73	95	5329	9025	6935	95,375	-0,375	0,141
7	91	109	8281	11881	9919	112,080	-3,080	9,487
8	21	43	441	1849	903	47,116	-4,116	16,938
9	56	93	3136	8649	5208	79,598	13,402	179,617
10	21	37	441	1369	777	47,116	-10,116	102,326
Σ	518	757	34036	63947	45898	757,000	0,000	437,623
Σ/n	51,8	75,7	3403,6	6395	4589,8	75,700	0,000	

Можно сделать следующие выводы:

- среднее число покупателей 51,8;
- средняя выручка 75,7 ед.;
- каждый покупатель приносит в среднем 0,928 ед. выручки.

3. Уравнение прогнозных значений имеет вид:

$$\widehat{y}_t = \widehat{a} + \widehat{b} \cdot x_t = 27,626 + 0,928 \cdot x_t .$$

Заполним соответствующий столбец в таблице и построим график прогнозных значений на диаграмме рассеяния.

4. Остатки линейной регрессионной модели определим по формуле

$$e_t = y_t - \widehat{y}_t$$

Оценка дисперсии остатков равна

$$\widehat{\sigma}^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-2} = \frac{437,623}{10-2} = 54,703 .$$

Оценка дисперсии $\widehat{a}$ равна

$$\widehat{\sigma}_{\widehat{a}}^2 = \frac{\widehat{\sigma}^2\, \overline{\overline{x^2}}}{n\left(\overline{\overline{x^2}} - \overline{x}^2\right)} = 25,846 .$$

Оценка дисперсии $\widehat{b}$ равна

$$\hat{\sigma}_{\hat{b}}^2 = \frac{\hat{\sigma}^2}{n\left(\overline{x^2} - \bar{x}^2\right)} = 0{,}00759 \ .$$

5. Гипотеза $a = a_0$ будет проверяться исходя из того, что случайная величина

$$t = \frac{\hat{a} - a_0}{\hat{\sigma}_{\hat{a}}}$$

в нормальной классической линейной регрессионной модели подчиняется распределению Стьюдента с (n-2) степенями свободы. Если $|t|$ окажется меньше некоторого критического значения $\mathbf{t_\alpha}$, которое находится по таблицам критических точек распределения Стьюдента, то гипотеза принимается. Если больше, то гипотеза отвергается. Таблица критических точек распределения Стьюдента приводится в Приложении 1. По таблице находим, что для уровня значимости 0,05 и восьми степеней свободы

t_α=2,306.

Проверим нулевую гипотезу H_0: **a**=0, при конкурирующей H_1: $a \neq 0$. Вычислим

$$t = \frac{27{,}626}{\sqrt{25{,}846}} = 5{,}434.$$

Поскольку 5,434>2,306, то нулевая гипотеза отвергается.

Выше сказанное справедливо и для параметра **b**. Проверим гипотезу **b**=1, которая означает, что один покупатель в среднем приносит торговой точке единицу выручки. Вычислим

$$\left|\frac{0{,}928 - 1}{\sqrt{0{,}00759}}\right| = 0{,}826.$$

Поскольку 0,826<2,306, то гипотеза принимается.

6. Из неравенств

$$\left|\frac{\hat{a} - a}{\hat{\sigma}_{\hat{a}}}\right| < t_\alpha \ \text{и} \ \left|\frac{\hat{b} - b}{\hat{\sigma}_{\hat{b}}}\right| < t_\alpha,$$

находим

$$\hat{a} - t_\alpha \hat{\sigma}_{\hat{a}} < a < \hat{a} + t_a \hat{\sigma}_{\hat{a}}$$

и

$$\hat{b} - t_\alpha \hat{\sigma}_{\hat{b}} < b < \hat{b} + t_\alpha \hat{\sigma}_{\hat{b}}.$$

Подставляя найденные ранее значения, находим 95% доверительные интервалы

$15{,}903 < a < 39{,}350$ и $0{,}727 < b < 1{,}129$.

Последнее неравенство означает, что с вероятностью 0,95 средняя выручка, которую приносит один покупатель, принадлежит интервалу (0,727; 1,129).

7. Коэффициент детерминации равен отношению суммы квадратов отклонений регрессии к общей сумме квадратов отклонений:

$$R^2 = \frac{\sum_{t=1}^{n}(\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}.$$

Можно доказать, что для парной регрессии данное отношение равно

$$R^2 = \hat{b}^2 \frac{S_x^2}{S_y^2} = 0{,}934,$$

где $S_x^2 = \overline{x^2} - \bar{x}^2 = 720{,}36$ и $S_y^2 = \overline{y^2} - \bar{y}^2 = 664{,}21$ выборочные дисперсии.

В построенной модели дисперсия результата на 93,4% объясняется линейной зависимостью выручки от числа покупателей и только на 6,6% дисперсией неучтенных факторов. Полученное значение коэффициента детерминации близко к единице. Поэтому связь между **x** и **y** сильная (число покупателей заметно влияет на выручку данного предприятия).

8. Если фактор (в нашей задаче число покупателей) не влияет на результат (выручку), тогда дисперсионное отношение

$$F = (n-2)\frac{R^2}{1-R^2} = 113{,}422$$

в классической нормальной линейной регрессионной модели, подчиняется распределению Фишера с (1; n-2) числом степеней свободы. По таблице критических точек распределения Фишера (Приложение 2) находим, что для уровня значимости α=0,05, величина F_α=10,128 (число степеней свободы: 1; 8). Найденное значение F>> F_α, что указывает на сильное влияние фактора на результат.

9. Допустим, что планируется расширение предприятия, при этом среднее количество покупателей должно вырасти на 20% и составит $x_{11} = 51{,}8 \cdot 1{,}2 = 67$ чел. Необходимо определить среднюю выручку, которую в этом случае получит предприятие. Выручку найдем из прогнозного уравнения

$$\hat{y}_{11} = \hat{a} + \hat{b} \cdot x_{11} = 27{,}626 + 0{,}928 \cdot 67 = 89{,}807.$$

Определим, насколько точным является данный прогноз. Для этого построим 95% доверительный интервал для прогнозируемой выручки:

$$\hat{y}_{11} - \hat{\sigma}_{e_{11}} \cdot t_\alpha < y_{11} < \hat{y}_{11} + \hat{\sigma}_{e_{11}} \cdot t_\alpha$$

$$71{,}66 < y_{11} < 107{,}953,$$

где $\hat{\sigma}_{e_{11}} = \sigma\sqrt{1 + \frac{1}{n} + \frac{(x_{11} - \bar{x})^2}{nS_x^2}} = 7{,}869$ - стандартное отклонение e_{11}.

10. Эластичность в произвольной точке определяется по формуле

$$\varepsilon_{x_t} = \hat{y}_t{}' \frac{x_t}{\hat{y}_t},$$

где $\hat{y}_t{}' = (\hat{a} + \hat{b}x_t)' = \hat{b}$ - производная по фактору **x**. В точке $\mathbf{x_{11}}$ эластичность равна

$$\varepsilon_{x_{11}} = \hat{b}\frac{x_{11}}{\hat{a} + \hat{b}x_{11}} = 0{,}692.$$

Найденная величина означает, что при увеличении числа покупателей на 1% выручка возрастает в среднем на 0,692%. Справедливо это в окрестности точки x_{11}.

3.2. Нелинейная модель. Линеаризация

Построим нелинейную регрессионную модель в виде

$$y_t = a \cdot x_t^b \cdot \varepsilon_t.$$

Тип нелинейной зависимости определяется формой диаграммы рассеяния, содержанием и теоретической моделью соответствующего экономического процесса. Иногда приходится применять различные нелинейные модели, а затем выбирать из них лучшую.

1. Предложенная модель становится линейной после логарифмирования:

$$\ln y_t = \ln a + b \ln x_t + \ln \varepsilon_t.$$

Обозначая: $Y_t = \ln y_t$; $A = \ln a$; $X_t = \ln x_t$; $u_t = \ln \varepsilon_t$, получим

$$Y_t = A + bX_t + u_t.$$

Преобразуем таблицу исходных данных и, опираясь на результаты вычислений в табл.6, вычислим оценки

$$\hat{A} = 1{,}779;\quad \hat{a} = e^{\hat{A}} = 5{,}923;\ \hat{b} = 0{,}654.$$

Уравнение прогнозных значений имеет вид

$$\hat{Y}_t = \underset{(0{,}199)}{1{,}779} + \underset{(0{,}052)}{0{,}654} \cdot X_t.$$

В уравнении в скобках показаны стандартные отклонения оценок параметров.

Прогнозирование в нелинейной модели может осуществляться по формуле

$$\hat{y}_t = 5{,}923 \cdot x_t^{0{,}654}$$

или с помощью экспоненты

$$\hat{y}_t = e^{\hat{Y}_t}.$$

Таблица 6

t	$X_t = \ln x_t$	$Y_t = \ln y_t$	X_t^2	Y_t^2	X_tY_t	$\hat{Y}_t$	$\hat{y}_t = e^{\hat{Y}_t}$	$(y_t - \hat{y}_t)^2$
1	3,434	4,159	11,792	17,296	14,282	4,023	55,878	65,961
2	4,317	4,605	18,641	21,208	19,883	4,601	99,546	0,206
3	4,489	4,635	20,148	21,481	20,804	4,712	111,328	69,347
4	3,258	3,912	10,615	15,304	12,746	3,908	49,810	0,036
5	3,555	4,143	12,640	17,166	14,730	4,102	60,491	6,294
6	4,290	4,554	18,408	20,738	19,538	4,583	97,803	7,855
7	4,511	4,691	20,348	22,009	21,162	4,727	112,956	15,652
8	3,045	3,761	9,269	14,147	11,451	3,769	43,321	0,103
9	4,025	4,533	16,203	20,544	18,245	4,410	82,243	115,706
10	3,045	3,611	9,269	13,039	10,994	3,769	43,321	39,951
Σ	38	42,604	147,334	182,930	163,834	42,604	756,697	321,112
Σ/n	3,797	4,260	14,733	18,293	16,383	4,260	75,670	

2. Вычислим коэффициент детерминации

$$R^2 = \hat{b}^2 \frac{S_X^2}{S_Y^2} = 0{,}654 \frac{0{,}317}{0{,}142} = 0{,}952,$$

где $S_X^2 = \overline{X^2} - \overline{X}^2$ и $S_Y^2 = \overline{Y^2} - \overline{Y}^2$.

Дисперсионное отношение

$$F = (n-2)\frac{R^2}{1-R^2} = 158{,}484.$$

Найденное значение F>> F_α=10,128, что указывает на сильное влияние фактора на результат в логарифмической модели.

Чтобы сравнить по качеству нелинейную и линейную модель, вычислим сумму квадратов отклонений нелинейного прогноза от наблюдений

$$\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 = 321{,}112,$$

и определим отношение, которое называют псевдодетерминацией:

$$\hat{R}^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2} = 1 - \frac{321{,}112}{6642{,}1} = 0{,}952 .$$

Сравнивая значение 0,952 с коэффициентом детерминации линейной модели 0,934, делаем вывод о том, что нелинейная модель лучше согласуется с данными наблюдений.

3. Как и в предыдущем задании, считаем, что планируется расширение предприятия, при этом среднее количество покупателей должно вырасти на 20% и составит $x_{11} = 51{,}8 \cdot 1{,}2 = 67$ чел. Прогноз средней выручки в нелинейной модели

$$\hat{y}_{11} = 5{,}923 \cdot 67^{0{,}654} = 92{,}471 .$$

Определим, насколько точным является данный прогноз. Для этого построим 95% доверительный интервал для логарифма прогнозируемой выручки:

$$\hat{Y}_{11} - \hat{\sigma}_{e_{11}} \cdot t_\alpha < Y_{11} < \hat{Y}_{11} + \hat{\sigma}_{e_{11}} \cdot t_\alpha$$
$$4{,}298 < Y_{11} < 4{,}756,$$

где $\hat{\sigma}_{e_{11}} = \sigma\sqrt{1 + \frac{1}{n} + \frac{(X_{11} - \bar{X})^2}{nS_X^2}} = 0{,}099$ - стандартное отклонение e_{11};

X_{11}=ln x_{11}=ln 67=4,205; $\hat{Y}_{11} = 1{,}779 + 0{,}654 \cdot 4{,}205 = 4{,}527$.

Для прогнозируемой выручки в рамках нелинейной модели доверительный интервал имеет вид

$$e^{4{,}298} < y_{11} < e^{4{,}756}$$
$$73{,}564 < y_{11} < 116{,}238.$$

4. Для степенной нелинейной модели эластичность в произвольной точке определяется по формуле

$$\varepsilon_{x_t} = \hat{y}_t{}' \frac{x_t}{y_t} = \hat{a} \cdot \hat{b} \cdot x_t^{\hat{b}-1} \frac{x_t}{\hat{a} \cdot x_t^{\hat{b}}} = \hat{b} = 0{,}654.$$

Найденная величина означает, что при увеличении среднего числа покупателей на 1% выручка возрастает в среднем на 0,654%.

5.

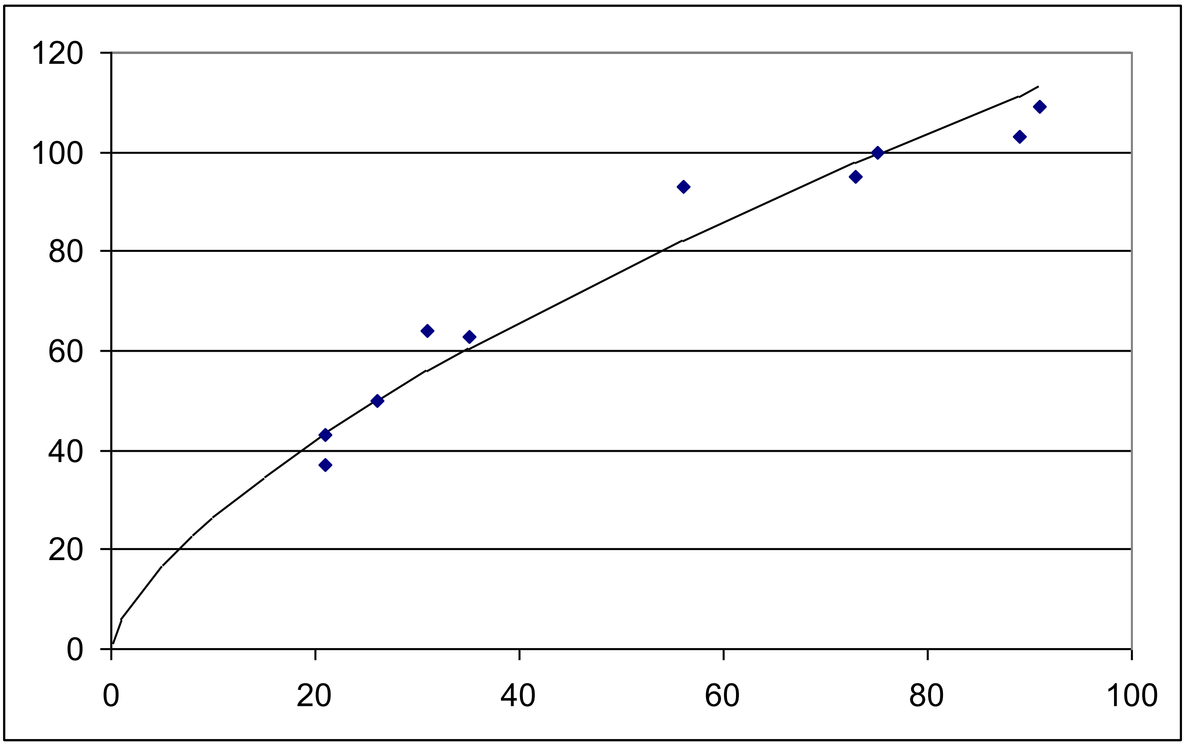


Рис.2. Диаграмма рассеяния наблюдений (точки), нелинейный тренд (сплошная линия).

3.3. Множественная регрессия

На выручку в торговом предприятии могло повлиять наличие или отсутствие рекламы. Чтобы определить эффективность рекламных мероприятий введем фиктивную переменную $\mathbf{z_t}$, которая принимает значение 1, когда реклама применялась, и 0, когда рекламы не было. Данные наблюдений представлены в табл.7.

Таблица 7

t	1	2	3	4	5	6	7	8	9	10
$\mathbf{x_t}$	31	75	89	26	35	73	91	21	56	21
$\mathbf{z_t}$	0	1	1	0	0	0	1	1	1	0
$\mathbf{y_t}$	64	100	103	50	63	95	109	43	93	37

1. Линейное уравнение множественной регрессии при наличии двух факторов, влияющих на результат, имеет вид

$$y_t = a + b \cdot x_t + c \cdot z_t + \varepsilon_t .$$

В соответствии с МНК, оценки параметров множественной регрессии являются решением системы нормальных уравнений:

$$\begin{cases} \hat{a} + \hat{b} \cdot \overline{x} + \hat{c} \cdot \overline{z} = \overline{y} \\ \hat{a} \cdot \overline{x} + \hat{b} \cdot \overline{x^2} + \hat{c} \cdot \overline{xz} = \overline{xy} \\ \hat{a} \cdot \overline{z} + \hat{b} \cdot \overline{xz} + \hat{c} \cdot \overline{z^2} = \overline{zy} \end{cases}$$

Коэффициенты при неизвестных в данной системе уравнений найдем с помощью табл.8:

Таблица 8

t	x_t	z_t	y_t	x_t^2	z_t^2	y_t^2	$x_t z_t$	$x_t y_t$	$y_t z_t$	$\hat{y}_t$	e_t	e_t^2
1	31	0	64	961	0	4096	0	1984	0	56,109	7,891	62,275
2	75	1	100	5625	1	10000	75	7500	100	97,495	2,505	6,277
3	89	1	103	7921	1	10609	89	9167	103	110,346	-7,346	53,969
4	26	0	50	676	0	2500	0	1300	0	51,519	-1,519	2,306
5	35	0	63	1225	0	3969	0	2205	0	59,780	3,220	10,366
6	73	0	95	5329	0	9025	0	6935	0	94,664	0,336	0,113
7	91	1	109	8281	1	11881	91	9919	109	112,182	-3,182	10,127
8	21	1	43	441	1	1849	21	903	43	47,924	-4,924	24,242
9	56	1	93	3136	1	8649	56	5208	93	80,053	12,947	167,625
10	21	0	37	441	0	1369	0	777	0	46,929	-9,929	98,579
Σ	518	5	757	34036	5	63947	332	45898	448	757	0,0	435,880
Σ/n	51,8	0,5	75,7	3403,6	0,5	6394,7	33,2	4589,8	44,8	75,7	0,0	

С найденными коэффициентами система нормальных уравнений приобретает вид

$$\begin{cases} \hat{a} + \hat{b} \cdot 51{,}8 + \hat{c} \cdot 0{,}5 = 75{,}7 \\ \hat{a} \cdot 51{,}8 + \hat{b} \cdot 3403{,}6 + \hat{c} \cdot 33{,}2 = 4589{,}8 \\ \hat{a} \cdot 0{,}5 + \hat{b} \cdot 33{,}2 + \hat{c} \cdot 0{,}5 = 44{,}8. \end{cases}$$

Полученная система уравнений может быть решена методом Крамера.

Найдем определитель матрицы коэффициентов при неизвестных

$$\Delta = \begin{vmatrix} 1 & 51{,}8 & 0{,}5 \\ 51{,}8 & 3403{,}6 & 33{,}2 \\ 0{,}5 & 33{,}2 & 0{,}5 \end{vmatrix} = 126{,}8.$$

Заменим первый столбец свободными коэффициентами и вычислим определитель

$$\Delta_1 = \begin{vmatrix} 75,7 & 1,8 & 0,5 \\ 4589,8 & 3403,6 & 33,2 \\ 44,8 & 33,2 & 0,5 \end{vmatrix} = 3506,160.$$

Аналогично заменим второй и третий столбцы свободными коэффициентами и вычислим определители

$$\Delta_2 = \begin{vmatrix} 1 & 75,7 & 0,5 \\ 51,8 & 4589,8 & 33,2 \\ 0,5 & 44,8 & 0,5 \end{vmatrix} = 116,4$$

и

$$\Delta_3 = \begin{vmatrix} 1 & 51,8 & 75,7 \\ 51,8 & 3403,6 & 4589,8 \\ 0,5 & 33,2 & 44,8 \end{vmatrix} = 126,16.$$

В соответствии с методом Крамера оценки параметров равны:

$$\hat{a} = \frac{\Delta_1}{\Delta} = \frac{3506,16}{126,8} = 27,651;$$

$$\hat{b} = \frac{\Delta_2}{\Delta} = \frac{116,4}{126,8} = 0,918;$$

$$\hat{c} = \frac{\Delta_3}{\Delta} = \frac{126,16}{126,8} = 0,995.$$

Дисперсия остатков оценивается по формуле

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-k-1} = \frac{435,880}{10-2-1} = 62,269.$$

Оценки дисперсий параметров и их стандартные отклонения равны:

$$\hat{\sigma}_{\hat{a}}^2 = \hat{\sigma}^2 \cdot \frac{A_{11}}{n \cdot \Delta} = 62,269 \cdot \frac{599,56}{10 \cdot 126,8} = 29,443; \qquad \hat{\sigma}_{\hat{a}} = 5,426;$$

$$\sigma_{\hat{b}}^2 = \sigma^2 \cdot \frac{A_{22}}{n \cdot \Delta} = 0,0123; \qquad \hat{\sigma}_{\hat{b}} = 0,111;$$

$$\sigma_{\hat{c}}^2 = \hat{\sigma}^2 \cdot \frac{A_{33}}{n \cdot \Delta} = 35,375; \qquad \sigma_{\hat{c}}^2 = 5,948,$$

где

$$A_{11} = \begin{vmatrix} \overline{x^2} & \overline{xz} \\ \overline{xz} & \overline{z^2} \end{vmatrix} = \begin{vmatrix} 3403,6 & 33,2 \\ 33,2 & 0,5 \end{vmatrix} = 599,56; \quad A_{22} = \begin{vmatrix} 1 & \bar{z} \\ \bar{z} & \overline{z^2} \end{vmatrix} = \begin{vmatrix} 1 & 0,5 \\ 0,5 & 0,5 \end{vmatrix} = 0,25;$$

$A_{33} = \begin{vmatrix} 1 & \bar{x} \\ \bar{x} & \overline{x^2} \end{vmatrix} = \begin{vmatrix} 1 & 51{,}8 \\ 51{,}8 & 3403{,}6 \end{vmatrix} = 720{,}36$ - соответствующие алгебраические дополнения.

.

Прогнозное уравнение имеет вид

$$\hat{y}_t = \hat{a} + \hat{b} \cdot x_t + \hat{c} \cdot z_t = \underset{(5,426)}{27{,}651} + \underset{(0,111)}{0{,}918} \cdot x_t + \underset{(5,948)}{0{,}995} \cdot z_t.$$

С помощью данного уравнения заполним в таблице столбцы $\hat{y}_t$, $e_t = y_t - \hat{y}_t$ и e_t^2.

2. Если $b = b_0$, то в нормальной классической линейной регрессионной модели случайная величина

$$t = \frac{\hat{b} - b_0}{\hat{\sigma}_{\hat{b}}}$$

подчиняется распределению Стьюдента с n-k-1=10-2-1=7 степенями свободы. Как и в парной регрессии, если $|t| < t_\alpha$, то принимается гипотеза $b = b_0$, в противном случае эта гипотеза отвергается.

Проверяем гипотезу **b**=0, т.е. b_0=0. Находим отношение

$$|t| = \left| \frac{\hat{b}}{\hat{\sigma}_{\hat{b}}} \right| = \frac{0{,}918}{0{,}111} = 8{,}270.$$

По таблице Приложения 1 находим $t_\alpha = 2{,}365$. Поскольку 8,270>2,365, то гипотеза **b**=0 отвергается, фактор **x** значимо влияет на результат.

Проверяем гипотезу **c**=0. Находим отношение

$$|t| = \left| \frac{\hat{c}}{\hat{\sigma}_{\hat{c}}} \right| = \frac{0{,}995}{5{,}948} = 0{,}167.$$

Поскольку 0,167<2,365, то гипотеза **c**=0 принимается, фактор **z** не влияет на результат. В нашем случае нет оснований считать, что рекламная компания повлияла на выручку. Следует учитывать, что десяти наблюдений для получения значимых результатов в множественной регрессии чаще всего оказывается недостаточно. Как правило, их требуется на порядок больше.

3. Коэффициент детерминации определяется по формуле

$$R^2 = \frac{RSS}{TSS},$$

где $RSS = \sum_{t=1}^{n}(\hat{y}_t - \bar{y})^2$ - сумма квадратов регрессии;

$$TSS = \sum_{t=1}^{n}(y_t - \bar{y})^2 = n(\overline{y^2} - \bar{y}^2) = 10 \cdot (6394{,}7 - 75{,}7^2) = 6642{,}10$$ - общая сумма квадратов.

Для МНК оценок параметров линейной регрессии справедливо

TSS=RSS+ESS,

где $ESS = \sum_{t=1}^{n}(y_t - \hat{y}_t)^2 = \sum_{t=1}^{n} e_t^2 = 435{,}88$ - сумма квадратов ошибок.

Найдем RSS=TSS-ESS=6642,10-435,88=6206,22. Коэффициент детерминации равен

$$R^2 = \frac{6206{,}22}{6642{,}10} = 0{,}934.$$

Скорректированный коэффициент детерминации равен

$$\tilde{R}^2 = 1 - \frac{n-1}{n-3}(1 - R^2) = 1 - \frac{9}{7}(1 - 0{,}934) = 0{,}916.$$

Сравнивая значение скорректированного коэффициента детерминации с коэффициентом детерминации, полученным в задании 1, делаем вывод, что введение дополнительного регрессора **z** не улучшило качества регрессионной модели.

4. Дисперсионное отношение Фишера равно

$$F = \frac{n-k-1}{k} \cdot \frac{RSS}{ESS} = \frac{10-2-1}{2} \cdot \frac{6206{,}22}{435{,}88} = 49{,}834,$$

где k – число регрессоров в модели (в нашем случае **x** и **z** всего два регрессора). С помощью таблицы Приложения 2 находим $F_\alpha = 4{,}737$, число степеней свободы 2 и 7. Поскольку 49,834>4,737 модель является значимой.

5. Выборочный коэффициент корреляции между **x** и **y** равен

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x S_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2}\sqrt{\overline{y^2} - \bar{y}^2}} = \frac{4589{,}8 - 51{,}8 \cdot 75{,}7}{\sqrt{3403{,}6 - 51{,}8^2}\sqrt{6394{,}7 - 75{,}7^2}} = 0{,}966.$$

Аналогично определяются коэффициенты корреляции **x** и **z**, а также между **z** и **y**. Результаты расчетов представлены в табл.9.

Таблица 9

Коэффициенты корреляции

r	**y**	**x**	**z**
y	1	0,966	0,539
x	0,966	1	0,544
z	0,539	0,544	1

В общем случае корреляционная матрица наблюдений имеет вид $r = \begin{pmatrix} 1 & r_{yx} & r_{yz} \\ r_{yx} & 1 & r_{xz} \\ r_{yz} & r_{xz} & 1 \end{pmatrix}$.

Алгебраические дополнения ее элементов определяются стандартно. Например

$$A_{12} = -\begin{vmatrix} r_{yx} & r_{xz} \\ r_{yz} & 1 \end{vmatrix} = -(r_{yx} - r_{xz} r_{yz}).$$

Частный коэффициент корреляции между **x** и **y** при фиксированном значении **z** равен

$$r_{yx.z} = -\frac{A_{12}}{\sqrt{A_{11}}\sqrt{A_{22}}} = \frac{r_{yx} - r_{xz} r_{yz}}{\sqrt{1-r_{xz}^2}\sqrt{1-r_{yz}^2}} = \frac{0{,}966 - 0{,}544 \cdot 0{,}539}{\sqrt{1-0{,}544^2}\sqrt{1-0{,}539^2}} = 0{,}953.$$

Частный коэффициент корреляции между **x** и **z** при фиксированном значении **y** равен

$$r_{xz.y} = -\frac{A_{23}}{\sqrt{A_{22}}\sqrt{A_{33}}} = \frac{r_{xz} - r_{yx} r_{yz}}{\sqrt{1-r_{yx}^2}\sqrt{1-r_{yz}^2}} = \frac{0{,}544 - 0{,}966 \cdot 0{,}539}{\sqrt{1-0{,}966^2}\sqrt{1-0{,}539^2}} = 0{,}105.$$

Частный коэффициент корреляции между **y** и **z** при фиксированном значении **x** равен

$$r_{yz.x} = -\frac{A_{13}}{\sqrt{A_{22}}\sqrt{A_{33}}} = \frac{r_{yz} - r_{xz} r_{yx}}{\sqrt{1-r_{xz}^2}\sqrt{1-r_{yx}^2}} = \frac{0{,}539 - 0{,}544 \cdot 0{,}966}{\sqrt{1-0{,}544^2}\sqrt{1-0{,}966^2}} = 0{,}063.$$

Результаты расчетов частных коэффициентов корреляции представлены в табл.10.

Таблица 10

Коэффициенты частной корреляции

r	**y**	**x**	**z**
y	1	0,953	0,063
x	0,953	1	0,105
z	0,063	0,105	1

3.4. Системы регрессионных уравнений

Задана структурная система эконометрических уравнений:

$$\begin{cases} y_{t1} = b_{10} + b_{12} y_{t2} + a_{11} x_{t1} + a_{12} x_{t2} + \varepsilon_{t1} \\ y_{t2} = b_{20} + b_{21} y_{t1} + b_{23} y_{t3} + a_{23} x_{t3} + \varepsilon_{t2} \\ y_{t3} = b_{30} + b_{32} y_{t1} + \varepsilon_{t3}. \end{cases}$$

В табл.11 приведены данные наблюдений эндогенных y_{ti} и экзогенных x_{ti} переменных модели:

Таблица 11

t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	19	23	13	5	5	4
2	15	31	15	7	8	3
3	22	13	17	4	8	2
4	6	32	9	6	2	5
5	8	66	16	12	2	9

Необходимо, опираясь на данные наблюдений, оценить параметры структурной системы.

Определим, являются ли уравнения системы идентифицируемыми. Для этого применим необходимое условие идентификации или счетное правило:

D+1=H – уравнение идентифицируемо;

D+1<H – уравнение неидентифицируемо;

D+1>H – уравнение сверхидентифицируемо.

Здесь - H – число эндогенных переменных в уравнении; D – число отсутствующих экзогенных и лаговых переменных в уравнении. В приведенной системе лаговые или запаздывающие переменные отсутствуют.

В первом уравнении эндогенных переменных две: y_{t1}, y_{t2}, следовательно, H=2. Всего у нас три экзогенных переменных, из них в первом уравнении отсутствует одна x_{t3}. Поэтому D=1. Для первого уравнения выполняется D+1=H=2. Первое уравнение идентифицируемо. Аналогично определяем, что для второго уравнения выполняется D+1=H=3. Для третьего уравнения находим, что D+1=4>H=2. Третье уравнение сверхидентифицируемо.

Прямая оценка параметров структурной системы уравнений недопустима. Из-за корреляции регрессоров с остатками получаются смещенные оценки. Для оценки параметров используют косвенный МНК (в случае идентификации) или двухшаговый МНК (в случае идентификации и сверхидентификации). Оба метода требуют оценки параметров приведенной формы модели:

$$\begin{cases} \hat{y}_{t1} = \hat{c}_{10} + \hat{c}_{11}x_{t1} + \hat{c}_{12}x_{t2} + \hat{c}_{13}x_{t3} \\ \hat{y}_{t2} = \hat{c}_{20} + \hat{c}_{21}x_{t1} + \hat{c}_{22}x_{t2} + \hat{c}_{23}x_{t3} \\ \hat{y}_{t1} = \hat{c}_{30} + \hat{c}_{31}x_{t1} + \hat{c}_{32}x_{t2} + \hat{c}_{33}x_{t3}. \end{cases}$$

Опираясь на данные наблюдений, выполняем регрессию y_{t1} на x_{t1}, x_{t2}, x_{t3}. Получаем уравнение прогноза

$$\underset{}{\hat{y}_{t1}} = \underset{(2,726)}{-14,982} - \underset{(2,726)}{5,421}\, x_{t1} + \underset{(0,367)}{5,228}\, x_{t2} + \underset{(0,809)}{8,632}\, x_{t3} \qquad R^2 = 0,998$$

В скобках приводятся стандартные ошибки оценок параметров.

Затем выполняем регрессию y_{t2} на x_{t1}, x_{t2}, x_{t3}

$$\hat{y}_{t2} = \underset{(2,726)}{1,129} + \underset{(0,469)}{5,912}\, x_{t1} - \underset{(0,367)}{1,327}\, x_{t2} - \underset{(0,809)}{0,368}\, x_{t3} \qquad R^2 = 0,9997$$

и y_{t3} на x_{t1}, x_{t2}, x_{t3}

$$\hat{y}_{t3} = \underset{(7,496)}{-7,298} - \underset{(1,291)}{1,342}\, x_{t1} + \underset{(1,008)}{2,623}\, x_{t2} + \underset{(2,223)}{3,763}\, x_{t3} \qquad R^2 = 0,920$$

Приведенная система эконометрических уравнений имеет вид

$$\begin{cases} \hat{y}_{t1} = -14,982 - 5,421x_{t1} + 5,228x_{t2} + 8,632x_{t3} \\ \hat{y}_{t2} = 1,129 + 5,912x_{t1} - 1,327x_{t2} - 0,368x_{t3} \\ \hat{y}_{t3} = -7,298 - 1,342x_{t1} + 2,623x_{t2} + 3,763x_{t3}. \end{cases}$$

Параметры первого уравнения структурной системы можно идентифицировать методом, который называется косвенный МНК. Для этого второе уравнение приведенной системы решается относительно x_{t3}:

$$x_{t3} = (\hat{y}_{t2} - 1,129 - 5,912x_{t1} + 1,327x_{t2})/(-0,368) = 3,068 - 2,717\hat{y}_{t2} + 16,065x_{t1} - 3,606x_{t2}$$

Затем найденное значение подставляется в первое уравнение:

$$\begin{aligned} \hat{y}_{t1} &= -14,982 - 5,421x_{t1} + 5,228x_{t2} + 8,632 \cdot (3,068 - 2,717\hat{y}_{t2} + 16,065x_{t1} - 3,606x_{t2}) \\ &= 11,501 - 23,453\hat{y}_{t2} + 133,252x_{t1} - 25,899x_{t2}. \end{aligned}$$

Для идентификации второго уравнения структурной системы решаем первое и третье уравнения приведенной системы относительно x_{t1} и x_{t2} и подставляем найденные значения во второе уравнение:

$$\begin{cases} -5,421x_{t1} + 5,228x_{t2} = \hat{y}_{t1} + 14,982 - 8,632x_{t3} \\ -1,342x_{t1} + 2,623x_{t2} = \hat{y}_{t3} + 7,298 - 3,763x_{t3} \end{cases}$$

$$x_{t1} = \frac{\Delta_1}{\Delta} = \frac{\begin{vmatrix} \hat{y}_{t1} + 14,982 - 8,632x_{t3} & 5,228 \\ \hat{y}_{t3} + 7,298 - 3,763x_{t3} & 2,623 \end{vmatrix}}{\begin{vmatrix} -5,421 & 5,228 \\ -1,342 & 2,623 \end{vmatrix}} = -0,159 - 0,364\hat{y}_{t1} + 0,726\hat{y}_{t3} + 0,412x_{t3}$$

$$x_{t2} = \frac{\Delta_2}{\Delta} = \frac{\begin{vmatrix} -5,421 & \hat{y}_{t1} + 14,982 - 8,632x_{t3} \\ -1,342 & \hat{y}_{t3} + 7,298 - 3,763x_{t3} \end{vmatrix}}{\begin{vmatrix} -5,421 & 5,228 \\ -1,342 & 2,623 \end{vmatrix}} = 2,701 - 0,186\hat{y}_{t1} + 0,753\hat{y}_{t3} - 1,224x_{t3}.$$

$$\hat{y}_{t2} = 1,129 + 5,912 \cdot (-0,159 - 0,364\hat{y}_{t1} + 0,726\hat{y}_{t3} + 0,412x_{t3}) -$$
$$-1,327 \cdot (2,701 - 0,186\hat{y}_{t1} + 0,753\hat{y}_{t3} - 1,224x_{t3}) - 0,368x_{t3} =$$
$$= -3,395 - 1,905\hat{y}_{t1} + 3,293\hat{y}_{t3} + 3,692x_{t3}.$$

Третье уравнение структурной системы сверхидентифицируемо. Для идентификации его параметров необходимо применить двухшаговый МНК. На первом шаге определяем значения инструментальной переменной

$$\hat{y}_{t1} = -14,982 - 5,421x_{t1} + 5,228x_{t2} + 8,632x_{t3}.$$

Значения $\hat{y}_{t1}$ рассчитываются в табл.12.

Таблица 12

t	y_{t3}	x_{t1}	x_{t2}	x_{t3}	$\hat{y}_{t1}$
1	13	5	5	4	18,579
2	15	7	8	3	14,789
3	17	4	8	2	22,421
4	9	6	2	5	6,105
5	16	12	2	9	8,105

На втором шаге осуществляем регрессию y_{t3} на $\hat{y}_{t1}$. Получаем уравнение

$$\underset{(3,348)}{\hat{y}_{t3} = 10,370} + \underset{(0,219)}{0,259\,\hat{y}_{t1}} \qquad R^2 = 0,319$$

Идентификация параметров произведена. Структурная система эконометрических уравнений имеет вид:

$$\begin{cases} \hat{y}_{t1} = 11,501 - 23,453\hat{y}_{t2} + 133,252x_{t1} - 25,899x_{t2} \\ \hat{y}_{t2} = -3,395 - 1,905\hat{y}_{t1} + 3,293\hat{y}_{t3} + 3,692x_{t3} \\ \hat{y}_{t3} = 10,370 + 0,259\hat{y}_{t1} + \varepsilon_{t3}. \end{cases}$$

3.5. Временные ряды. Авторегрессия

В табл.13 приводятся наблюдения продаж некоторого товара. Переменная **t** - номер наблюдения. Переменная $\mathbf{y_t}$ – выручка за один день.

Таблица 13

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\mathbf{y_t}$	19	15	20	21	21	19	16	20	20	22	19	16	22	22	23

t	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
$\mathbf{y_t}$	21	19	23	21	23	22	19	21	21	23	21	19	22	21	24

1. Данные таблицы представляют временной ряд. Оценим характер временного ряда визуально. Для этого построим его диаграмму, представленную на рис.3.

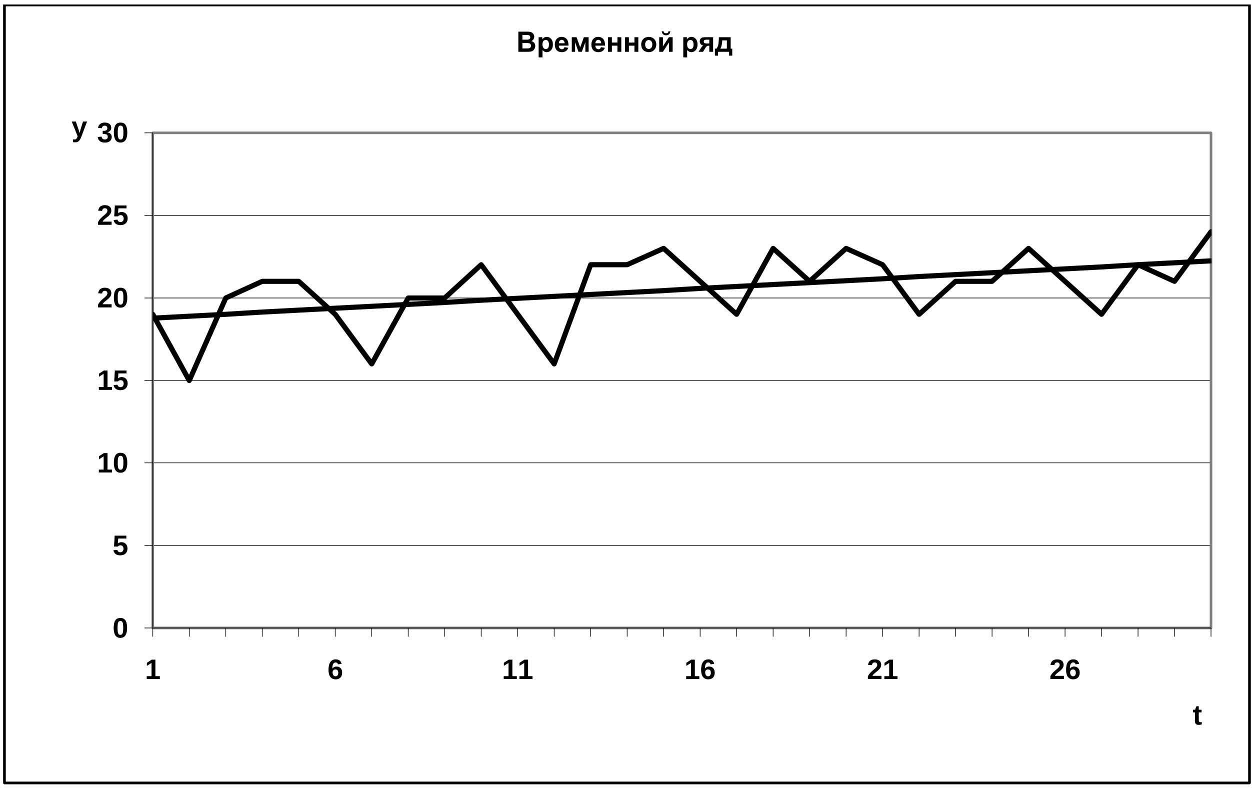


Рис.3. Временной ряд и тренд (прямая линия) продаж товара.

По графику временного ряда можно увидеть, что значения временного ряда возрастают с увеличением номера наблюдения. Это означает, что у временного ряда может быть тренд. Определим параметры тренда. Для этого построим линейную регрессионную модель с регрессором **t**:

$$y_t = a + b \cdot t + \varepsilon_t.$$

Определим оценки параметров: $\widehat{a} = 18{,}641\ (t = 25{,}932)$; $\widehat{b} = 0{,}120\ (t = 2{,}961)$ (в скобках представлено отношение Стьюдента, критическое значение t_a=2,048, число степеней свободы 30-2=28). Поскольку 2,961>2,048, оценка параметра **b** является значимой и тренд у временного ряда существует.

Таблица 14

t		y_t	e_t	t-5	y_{t-5}	t^*	y^*_t	$\hat{y}_t$	$\hat{y}_t^{пр}$	$e_t^{пр}$
	1	19	0,239					19,676		
	2	15	-3,881					19,749		
	3	20	0,999					19,822		
	4	21	1,879					19,895		
	5	21	1,759					19,968		
	6	19	-0,361	1	19	5,323	6,137	20,041	19,583	-0,583
	7	16	-3,481	2	15	5,646	5,845	20,114	16,899	-0,899
	8	20	0,399	3	20	5,969	6,46	20,187	20,307	-0,307
	9	20	0,279	4	21	6,292	5,783	20,260	21,008	-1,008
	10	22	2,160	5	21	6,615	7,783	20,333	21,032	0,968
	11	19	-0,960	6	19	6,938	6,137	20,406	19,701	-0,701
	12	16	-4,080	7	16	7,261	5,168	20,479	17,694	-1,694
	13	22	1,800	8	20	7,584	8,46	20,552	20,425	1,575
	14	22	1,680	9	20	7,907	8,46	20,625	20,449	1,551
	15	23	2,560	10	22	8,23	8,106	20,698	21,827	1,173
	16	21	0,440	11	19	8,553	8,137	20,771	19,819	1,181
	17	19	-1,680	12	16	8,876	8,168	20,844	17,812	1,188
	18	23	2,200	13	22	9,199	8,106	20,917	21,897	1,103
	19	21	0,080	14	22	9,522	6,106	20,990	21,921	-0,921
	20	23	1,960	15	23	9,845	7,429	21,063	22,622	0,378
	21	22	0,840	16	21	10,168	7,783	21,136	21,291	0,709
	22	19	-2,279	17	19	10,491	6,137	21,209	19,961	-0,961
	23	21	-0,399	18	23	10,814	5,429	21,282	22,692	-1,692
	24	21	-0,519	19	21	11,137	6,783	21,355	21,362	-0,362
	25	23	1,361	20	23	11,46	7,429	21,428	22,739	0,261
Наблюдения	26	21	-0,759	21	22	11,783	6,106	21,501	22,086	-1,086
	27	19	-2,879	22	19	12,106	6,137	21,574	20,079	-1,079
	28	22	0,001	23	21	12,429	7,783	21,647	21,456	0,544
	29	21	-1,119	24	21	12,752	6,783	21,720	21,480	-0,480
	30	24	1,761	25	23	13,075	8,429	21,793	22,857	1,143
Σ	**465**	**615**	**0,0**	**325,0**	**508,0**	**230,0**	**175,084**	**622,0**	**519,0**	**0,0**
Σ/n	**15,5**	**20,5**	**0,0**	**10,833**	**16,933**	**7,666**	**5,836**	**20,734**	**17,300**	**0,0**

Прогноз	31	21,866	22,105
	32	21,939	21,452
	33	22,012	23,506
	34	22,085	17,324
	35	22,158	22,000

Вычислим прогнозные значения

$$\hat{y}_t = \hat{a} + \hat{b} \cdot t = 18{,}641 + 0{,}120 \cdot t$$

и отклонения от прогноза

$$e_t = y_t - \hat{y}_t .$$

Результаты расчетов представлены в табл.14. Линия тренда показана на рис.3.

Сумма квадратов отклонений остатков построенной модели ESS=103,2, общая сумма квадратов TSS=135,5. Коэффициент детерминации R^2=0,239, отношение Фишера F=8,771 (критическое значение F_a=4,196). Модель тренда является значимой по критерию Фишера.

2. Исследуем возможности увеличения значимости данной модели. Для этого определим выборочную автокорреляционную функцию остатков построенной модели. Ее значения рассчитываются по формуле

$$r_i \approx \frac{\sum_{t=i+1}^{n} e_t e_{t-i}}{\sum_{t=1}^{n} e_t^2}$$

и представлены в табл.15. Гистограмма автокорреляционной функции показана на рис.4.

Таблица 15

Автокорреляционная функция

i	0	1	2	3	4	5	6
r_i	1	-0,015	-0,273	-0,336	-0,04	0,677	-0,094

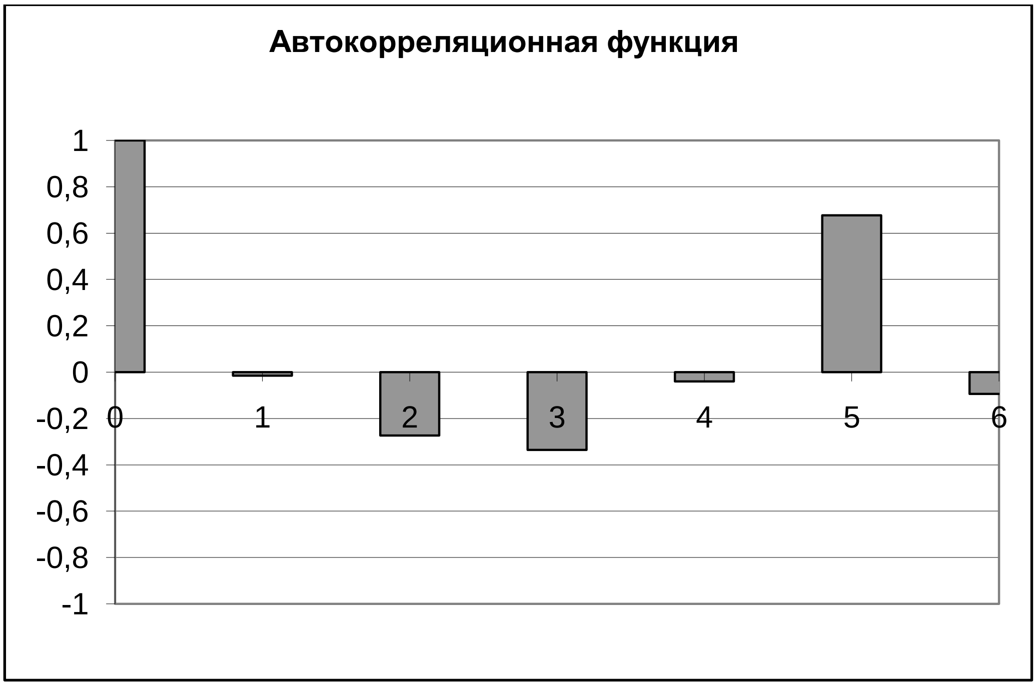


Рис.4. Гистограмма автокорреляционной функции.

При i=5 автокорреляция (0,677) значительно превосходит по модулю соседние значения, и указывает на корреляцию остатков через 5 наблюдений. Для продаж, которые осуществляются по рабочим дням (5 дней в неделю), характерна так называемая недельная сезонная компонента. Если фиксируются поквартальные данные, то может наблюдаться корреляция остатков при i=4 (квартальная сезонная компонента). Если рассматриваются месячные данные, то может наблюдаться значимая корреляция при i=12 и т.д.

В связи с тем, что отклонения продаж товара от тренда, коррелируют через 5 дней, предлагается следующая модель временного ряда:

$$y_t = a + b \cdot t + \varepsilon_t$$
$$\varepsilon_t = \rho \cdot \varepsilon_{t-5} + u_t,$$

где u_t – случайные отклонения, удовлетворяющие требованиям классической ЛРМ.

Вычислим

$$y_t^* = y_t - \rho y_{t-5} = a + bt + \varepsilon_t - \rho(a + b(t-5) + \varepsilon_{t-5}) =$$
$$= a(1-\rho) + b(t - \rho(t-5)) + \varepsilon_t - \rho\varepsilon_{t-5} = a^* + bt^* + u_t$$

Получена классическая ЛРМ, параметры которой могут оцениваться с помощью МНК при этом оценки являются состоятельными, несмещенными и эффективными. Оценка параметра $\hat{\rho} = r_5 = 0{,}677$. Вычислим $y_t^* = y_t - \hat{\rho} y_{t-5}$ и $t^* = t - \hat{\rho}(t-5)$ и включим найденные значения в таблицу. МНК оценки параметров $\hat{a}^* = 6{,}332;\ \hat{a} = \frac{\hat{a}^*}{1-\hat{\rho}} = \frac{6{,}332}{1-0{,}677} = 19{,}603;\ \hat{b} = 0{,}073$. Получена ЛРМ с учетом сезонных колебаний

$$\hat{y}_t = \hat{a} + \hat{b}t = 19{,}603 + 0{,}073 \cdot t.$$

Прогнозирование в этой модели осуществляется с учетом последних наблюдений временного ряда:

$$\hat{y}_t^{пр} = \hat{y}_t^* + \hat{\rho} y_{t-5} = \hat{y}_t + \hat{\rho}(y_{t-5} - \hat{y}_{t-5}).$$

Рис.5. Временной ряд, линейный тренд и прогноз с учетом корреляции остатков.

Сумма квадратов остатков построенной модели ESS=26,5, общая сумма квадратов TSS=135,5. Коэффициент детерминации R^2=0,804, отношение Фишера F=94,6 (критическое значение F_a=4,279, число степеней свободы 30-2-5=23). Значимость модели заметно улучшилась. Прогноз продаж товара на следующую рабочую неделю представлен в таблице и на рис.5.

4. ЛЕКЦИИ

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

Эконометрика – это наука, которая даёт количественное выражение взаимосвязей экономических явлений и процессов.

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

Можно выделить три основных класса моделей, которые применяются для анализа и прогнозирования экономических процессов:

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

При этом все переменные любой эконометрической модели по способу их вхождения в эту модель можно разбить на ***объясняемые (зависимые, исследуемые)*** переменные и ***объясняющие (предопределённые, факторные)*** переменные.

Например, если мы будем решать задачу прогнозирования продаж мороженого в определённый день каким-либо торговым предприятием, то объясняемой переменной будет объём продаж, а объясняющими переменными могут выступать: температура воздуха, торговая наценка, среднедушевой доход населения и другие.

Необходимым условием использования той или иной переменной при построении модели является наличие ряда данных наблюдений (измерений) величины этой переменной, либо получение ряда значений с использованием дополнительных вычислений на основе наблюдений о показателях, объясняющих интересующую нас переменную.

Например, определение достоверных значений среднедушевого дохода непосредственно по результатам опросов и бухгалтерской отчётности может оказаться сложнее оценки изменения дохода на основе информации об изменении розничного оборота товаров и услуг, а также изменении общей суммы банковских вкладов населения.

В эконометрике выделяют три типа данных:

I. ***Кросс секционные (перекрёстные)*** данные представляют ситуацию в группе переменных в отдельный момент времени. Таковыми, например, являются публикуемые в деловых разделах газет списки цен на различные акции, процентные ставки по разным видам вкладов и обменные курсы разных валют. Другим примером может служить информация о продажах торговым предприятием в определённый день товаров различных групп (пищевых, хозяйственных и т.д.)

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

III. ***Временные ряды*** отражают изменения (динамику) какой-либо переменной на промежутке времени. Например, данные об обменном курсе валюты за каждый день в конкретном обменном пункте или данные о продажах мороженого в одном киоске за каждый день будут являться ежедневным временным рядом.

Эконометрическое моделирование состоит из следующих этапов:

1. *На постановочном этапе* формулируются конечные цели моделирования, определяется наборы возможных исследуемых (объясняемых) переменных $\overline{Y}=(y_1, y_2, \dots y_k)$ и факторных (объясняющих) переменных $\overline{X}=(x_1, x_2, \dots x_m)$.
2. *На предварительном этапе* осуществляется предварительный анализ экономической сути изучаемого явления, возможностей сбора и обработки статистических данных.
3. *На этапе параметризации* производится *выбор общего вида модели*, в том числе состава и формы входящих в неё связей. Например, может быть выбрана модель с одной объясняющей и одной объясняемой переменными – ***модель парной регрессии***. Если объясняющих (факторных) переменных используется две или более, то говорят об использовании ***модели множественной регрессии***. При этом, в качестве вариантов могут быть выбраны линейная, экспоненциальная, гиперболическая, показательная и другие виды функций, связывающие эти переменные.
4. *Информационный этап* заключается в сборе информации (проведение наблюдений, использование материалов отчётности и т.д.) и предварительном анализе данных (проверка аномальных значений показателей, сглаживание, тестирование на наличие тенденции исследуемых показателей к изменению).
5. *Идентификация модели* посвящена определению неизвестных параметров (коэффициентов) модели с использованием имеющегося набора данных. Наибольшее распространение для оценки параметров получил метод наименьших квадратов.
6. *Проверка (верификация) модели и прогнозирование* предполагает сопоставление реальных и модельных данных, проверку адекватности модели, оценку точности модель-

ных данных. Если модель адекватна и имеет приемлемую точность, то на её основе строится прогноз – точечный и интервальный.

Тема 2. Парная корреляция и регрессия

Изучение действительности показывает, что изменение каждого исследуемого (объясняемого) показателя находится в связи и взаимодействии с изменением объясняющих (факторных) показателей. Например, изменение производительности труда работников предприятия зависит от степени совершенства применяемого оборудования, технологии, организации труда, управления и других факторов.

Рассматривая зависимости между признаками, необходимо выделить два вида зависимостей: функциональные и корреляционные.

Функциональные связи характеризуются полным соответствием между изменением факторного признака (признаков) и исследуемого показателя. Так, величина начисленной зарплаты при повременной оплате труда однозначно определяется количеством отработанных часов.

В корреляционных связях между изменением факторного и результативного признаков нет однозначного соответствия, воздействие факторов проявляется лишь в среднем при многократном наблюдении фактических данных. Например, чем больше у человека заработная плата, тем больше он тратит денег на покупку одежды. Однако, точную величину таких расходов при определенной величине заработной платы назвать нельзя. Можно только определить среднюю величину расходов на одежду у людей с определённым размером заработной платы.

В отличие от жёсткости функциональной связи корреляционные связи характеризуются множеством причин и следствий и устанавливаются лишь тенденции изменения исследуемого признака при изменении факторного признака (признаков).

2.1. Ковариация. Выборочный коэффициент парной корреляции

Основная задача ***корреляционного анализа*** заключается в выявлении взаимосвязи между случайными переменными путём оценки коэффициентов корреляции и детерминации, а также проверки значимости полученных значений.

В *эконометрике* корреляционный анализ применяется для *отбора факторов*, оказывающих наибольшее влияние на исследуемый показатель и *оценки качества* построенных эконометрических *моделей*.

Мерой взаимосвязи между двумя переменными v и w является выборочная ковариации, вычисляемая по правилу:

$$Cov(v,w) = \frac{1}{n-1}\sum_{i=1}^{n}\left[(v_i - v_{cp})(w_i - w_{cp})\right],$$

где $v_i, w_i, i = 1,\ldots n$ - результаты наблюдений, n – число наблюдений, $v_{cp} = \frac{1}{n}\sum_{i=1}^{n} v_i$; $w_{cp} = \frac{1}{n}\sum_{i=1}^{n} w_i$. Обозначения переменных специально выбраны отличные от x и y, чтобы подчеркнуть возможность наличия связи между двумя любыми переменными, не обязательно являющимися объясняющей и объясняемой переменными.

Существенным недостатком ковариации является зависимость от единиц, в которых измеряются переменные v и w. Если мы одни и те же данные запишем с использованием различных единиц измерения, то получим различные значения ковариации. То есть любое ненулевое значение ковариации само по себе не позволяет сделать вывод о тесноте связи между переменными.

Поэтому для измерения силы связи между двумя переменными используется парный коэффициент корреляции. Парный коэффициент корреляции является показателем тесноты связи между переменными v и w лишь в случае линейной зависимости между этими переменными.

В практических расчётах обычно используется выборочный парный коэффициент парной корреляции, определяемый по имеющемуся набору фактических данных:

$$r(v,w) = \frac{\sum_{i=1}^{n}\left[(v_i - v_{cp}) * (w_i - w_{cp})\right]}{\sqrt{\sum_{i=1}^{n}(v_i - v_{cp})^2 * \sum_{i=1}^{n}(w_i - w_{cp})^2}} = \frac{Cov(v,w)}{S_v * S_w} \quad , \qquad (2.1)$$

где $S_v^2 = \frac{1}{n-1}\sum_{i=1}^{n}(v_i - v_{cp})^2, S_w^2 = \frac{1}{n-1}\sum_{i=1}^{n}(w_i - w_{cp})^2$ - выборочные значения дисперсии переменных v и w.

Парный коэффициент корреляции обладает следующими свойствами:

1) принимает значение в интервале [-1;1], то есть $\left| r(v,w) \right| \le 1$;

2) не зависит от выбора начала отсчёта и единицы измерения $r(av+b; cw+d) = r(v, w)$, где *a, b, c, d* – постоянные величины, причём *a* и *c* – положительны;
3) если $r(v, w) > 0$, то между переменными имеется *прямая* связь, то есть при возрастании (убывании) одной из них другая также возрастает (убывает); если $r(v, w) < 0$, то связь является *обратной*, то есть при возрастании одной переменной другая убывает;
4) если $r(v, w) = \pm 1$, то между переменными имеется функциональная линейная зависимость, а если $r(v, w) = 0$, то линейная связь между переменными отсутствует; соответственно, чем ближе модуль коэффициента парной корреляции к единице, тем теснее связь между переменными.

Заметим, что при отсутствии линейной связи между двумя переменными, между ними может существовать тесная связь другого вида.

Пример 1. В таблице приведены данные об объёмах продаж мороженого в магазине за день *y*, в зависимости от температуры воздуха в городе *x2* и процента торговой надбавки *x3*. Видно, что спрос быстро растёт при повышении температуры воздуха. При наступлении очень высоких температур, предприятие резко увеличивает наценку, поскольку оказывается не в состоянии физически удовлетворить резко возрастающий спрос и сдерживает его повышением цен.

Требуется определить наличие между переменными линейных корреляционных связей, сделать выводы об их тесноте и охарактеризовать как прямые или обратные.

Таблица 1.

y	x1	x2
2	5	20
3,5	10	20
5	15	20
12	20	20
22	25	20
40	30	25
42	35	50

Решение. В первую очередь вычислим средние значения переменных в предложенной выборке данных: $y_{ср} = (2+3{,}5+5+12+22+40+42)/7 \approx 18{,}0714$. Аналогично, $x1_{ср} = 20{,}0; \quad x2_{ср} = 25{,}0$. Тогда выборочные коэффициенты парной корреляции:

$$r(y,x1) = \frac{(2-18{,}0714)*(5-20)+(3{,}5-18{,}0714)*(10-20)+\ldots+(42-18{,}0714)*(35-20)}{\sqrt{(2-18{,}0714)^2+\ldots+(42-18{,}0714)^2} * \sqrt{(5-20)^2+\ldots+(35-20)^2}} \approx$$

$\approx 0{,}9494; \quad r(y,x2) \approx 0{,}7229; \quad r(x1,x2) \approx 0{,}6901.$

Следовательно, мы можем сказать, что между переменными y (объём продаж) и x1 (температура воздуха) имеется тесная прямая линейная связь. Между переменными x1 (температура воздуха) и x2 (торговая наценка) также наблюдается тесная прямая линейная зависимость. То же самое можно сказать о взаимосвязи между переменными y и x2.

Для того, чтобы проверить, можем ли мы делать вывод о наличии линейной корреляционной связи между переменными по полученному значению коэффициента парной корреляции производится оценка его значимости, то есть определяется действительно ли полученное значение отражает наличие линейной связи, или же ненулевое значение коэффициента получено в результате случайных колебаний показателей или является следствием погрешности в вычислениях.

2.2. Оценка значимости выборочного коэффициента парной корреляции

Для оценки значимости выборочного коэффициента парной корреляции применяется t-критерий Стьюдента. При этом фактическое значение этого критерия определяется по формуле:

$$t_{набл} = \sqrt{\frac{r^2}{1-r^2}(n-2)}, \qquad (2.2)$$

где n – число наблюдений. Полученное значение сравнивается с табличным критическим значением $t_{кр}(\alpha,\nu)$, зависящим от уровня значимости α и числа степеней свободы $\nu = n-2$. Критическое значение может быть найдено по соответствующим таблицам, а при использовании табличного процессора Excel – с помощью функции СТЬЮДРАСПОБР (α ; γ).

При $t_{набл} > t_{кр}$ полученное значение коэффициента корреляции r признается значимым, то есть между переменными имеется линейная корреляционная зависимость.

Для рассмотренного **Примера 1** при $\alpha = 0{,}1$, с учётом количества степеней свободы $\nu = 7 - 2 = 5$ критическое значение $t_{кр}(0{,}1;5) = 2{,}02$. Вычислим $t_{набл}$ для каждой пары переменных и сделаем вывод о значимости соответствующих коэффициентов корреляции.

Для пары переменных *y, x1* :

$$t_{набл} = \sqrt{\frac{0{,}9494 * 0{,}9494}{1 - 0{,}9494 * 0{,}9494} * 5} \approx 6{,}76 \quad > \quad t_{кр} = 2{,}02.$$

Следовательно, значение коэффициента $r(y, x1) = 0{,}9494$ является значимым.

Для пары переменных *y, x2* :

$$t_{набл} = \sqrt{\frac{0{,}7229 * 0{,}7229}{1 - 0{,}7229 * 0{,}7229} * 5} \approx 2{,}34 \quad > \quad t_{кр} = 2{,}02.$$

Следовательно, мы можем утверждать, что значение коэффициента $r(y, x2) = 0{,}7229$ является значимым.

Для пары переменных *x1, x2* :

$$t_{набл} = \sqrt{\frac{0{,}6901 * 0{,}6901}{1 - 0{,}6901 * 0{,}6901} * 5} \approx 2{,}13 \quad > \quad t_{кр} = 2{,}02.$$

Следовательно, значение коэффициента $r(x1, x2) = 0{,}6901$ является значимым.

Поскольку мы выбрали уровень значимости $\alpha = 0{,}1$, то с вероятностью 10% мы сделали ошибочные выводы, а с вероятностью $P = 1 - \alpha = 0{,}9$ наши выводы верны.

2.3. Модель парной регрессии. Основные понятия. Линейная парная регрессия

Регрессионное уравнение, разрешённое относительно исследуемой переменной *y* при наличии одной факторной переменной *x*, в общем виде записывается как:

$$y_p = f(x),$$

и показывает, каково будет *<u>в среднем</u>* значение переменной *y*, если переменная *x* примет конкретное значение. Индекс *p* указывает на то, что мы получаем расчётное значение переменной *y*. Мы говорим *в среднем*, поскольку под влиянием неучтённых в модели факторов и в результате погрешностей измерения фактическое значение переменной *y* может принимать различные значения для одного значения *x*.

Если *f(x)* является линейной функцией, то мы имеем общий вид модели парной линейной регрессии:

$$y_p = a + b * x, \tag{2.3}$$

где *a* – постоянная величина (или свободный член уравнения), *b* – коэффициент регрессии, определяющий наклон линии, вдоль которой рассеяны наблюдения. Коэффициент регрессии характеризует изменение переменной *y* при изменении значения *x* на единицу. Если $b > 0$, то переменные положительно коррелированны, если $b < 0$ - отрицательно коррелированны. Фактическое значение исследуемой переменной у тогда может быть представлено в виде:

$$y = a + b * x + \varepsilon, \tag{2.4}$$

где *ε* – разность между фактическим значением (результатом наблюдения) и значением, рассчитанным по уравнению модели. Если модель адекватно описывает исследуемый процесс, то *ε* – независимая нормально распределённая случайная величина с нулевым математическим ожиданием ($M\varepsilon = 0$) и постоянной дисперсией ($D\varepsilon = \sigma^2$). Наличие случайной компоненты *ε* отражает тот факт, что присутствуют другие факторы, влияющие на исследуемую переменную и не учтённые в модели.

2.4. Определение параметров линейной парной модели методом МНК

Для оценки параметров *a* и *b* линейной парной регрессии с использованием имеющегося набора результатов наблюдений наиболее часто используют метод наименьших квадратов (МНК), который минимизирует сумму квадратов ε_i - отклонения результатов наблюдений y_i от рассчитанных по линейной модели (2.3) значений y_{pi}:

$$b = \frac{\sum_{i=1}^{n}\left[(x_i - x_{cp}) * (y_i - y_{cp})\right]}{\sum_{i=1}^{n}(x_i - x_{cp})^2},$$

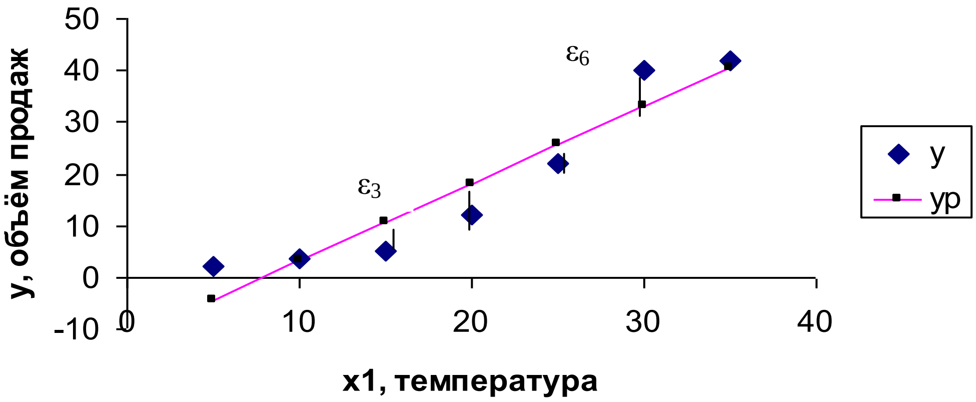
$$a = y_{cp} - b * x_{cp} \tag{2.5}$$

Такое решение может существовать только при выполнении условия $\sum_{i=1}^{n}(x_i - x_{cp})^2 \neq 0$, то есть когда не все наблюдения проводились при одном и том же значении факторной переменной (сумма квадратов равна нулю, если каждое слагаемое равно нулю). Это условие называется ***условием идентифицируемости*** модели.

По данным, приведённым в Примере 1, построим линейную модель для объёма продаж мороженного *y* в зависимости от температуры воздуха *x1*. Промежуточные данные вычислений и модельные значения y_p приведены в Таблице 2.

Таблица 2.

	x1	y	$x1_i$-$x1_{cp}$	y_i-y_{cp}	$(x1_i$-$x1_{cp})^2$	$(x1_i$-$x1_{cp})$*$(y_i$-$y_{cp})$	y_p	ε
	5,0	2	-15,0	-16,07	225,00	241,07	-4,43	6,43
	10,0	3,5	-10,0	-14,57	100,00	145,71	3,07	0,43
	15,0	5	-5,0	-13,07	25,00	65,36	10,57	-5,57
	20,0	12	0,0	-6,07	0,00	0,00	18,07	-6,07
	25,0	22	5,0	3,93	25,00	19,64	25,57	-3,57
	30,0	40,0	10,0	21,93	100,00	219,29	33,07	6,93
	35,0	42,0	15,0	23,93	225,00	358,93	40,57	1,43
Сумма	140,0	126,5	0,0	0,00	700,00	1050,00	126,50	0,00
Сред-нее	20,0	18,1	b=	1,5	a=	-11,93		



Исходные данные наблюдений и результаты расчётов приведены на следующем рисунке

Рис 1. Модель парной линейной регрессии

Таблица и график построены средствами табличного процессора Excel.

Таким образом уравнение парной линейной модели имеет вид:

$$y_p = -11{,}93 + 1{,}5 * x1.$$

2.5. Проверка значимости параметров парной линейной модели

Поскольку в результате наблюдений мы имеем случайные значения y_i, то и вычисленные с их помощью параметры парной линейной модели a и b также являются случайными величинами. Для оценки надёжности полученных значений a и b производится проверка их значимости с использованием ***стандартной ошибки оценки,*** которая, в свою очередь, определяется по значениям ряда остатков ε_i:

$$S_{cm} = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n - m - 1}}, \tag{2.6}$$

где n – количество наблюдений, m – количество факторных переменных в модели. Выражение (2.6) для определения стандартной ошибки оценки будет использоваться нами в дальнейшем неоднократно, поскольку применимо в случае нелинейных моделей, а также при наличии в модели двух и более факторных переменных, то есть является универсальным.

Собственно проверка значимости параметров линейной модели производится в три этапа, аналогично тому, как это делалось для проверки значимости выборочного коэффициента корреляции.

На первом этапе вычисляются t –статистики:

$$t_a = \frac{|a|}{S_a}, \quad t_b = \frac{|b|}{S_b}, \tag{2.7}$$

где

$$S_a = S_{cm} * \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - x_{cp})^2}}, \qquad S_b = S_{cm} * \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - x_{cp})^2}}. \tag{2.8}$$

На втором этапе определяется критическое значение $t_{кр}(\alpha; n\text{-}m\text{-}1)$ по таблицам или с помощью функции СТЮДРАСПОБР в Excel. Уровень значимости α задаётся, а число степеней свободы вычисляется по числу наблюдений n и числу факторов m (в парной модели фактор x единственный).

Наконец, на третьем этапе вычисленные значения *t*-статистик сравниваются с критическими значениями $t_{кр}$. Если расчётное значение больше табличного, то соответствующий параметр (коэффициент уравнения) считается значимым. В противном случае коэффициент значимым не является, то есть его можно положить равным нулю.

Произведём проверку значимости линейной модели парной регрессии, которую мы построили по данным **Примера 1.** Стандартная ошибка оценки вычисляется по значениям ряда остатков линейной модели ε_i (приведён в последней колонке Таблицы 2):

$$S_{ст} = \sqrt{\frac{6{,}43^2 + 0{,}43^2 + \ldots + 1{,}43^2}{7-1-1}} \approx 5{,}869 .$$

Тогда, с использованием результатов вычислений из Таблицы 2, получаем:

$$S_a = 5{,}869\sqrt{\frac{5^2 + \ldots + 35^2}{7 * 700}} \approx 4{,}96; \; S_b = 5{,}869\sqrt{\frac{1}{700}} \approx 0{,}22; \quad t_a = \frac{11{,}93}{4{,}96} \approx 2{,}4; \; t_b = \frac{1{,}5}{0{,}22} \approx 6{,}76;$$

При уровне значимости 10% и числе степеней свободы 7-1-1=5 имеем $t_{кр}$=2,02. Поскольку расчётные значения *t*-статистик для обоих параметров больше критического значения, то с вероятностью 90% можно утверждать, что оба параметра линейного уравнения - *a* и *b* являются значимыми.

2.6. Проверка выполнения предпосылок МНК.

Проверка выполнения предпосылок МНК выполняется на основе анализа остаточной компоненты ε. Ряд остатковε должен удовлетворять ряду требований, а именно: равенство нулю математического ожидания, случайный характер отклонений от математического ожидания, отсутствие автокорреляции и неизменность дисперсии остатков при изменении факторной переменной, нормальный закон распределения. Рассмотрим способы проверки этих условий:

1. Проверка равенства математического ожидания уровней ряда остатков нулю осуществляется в ходе проверки соответствующей $H_0 : |\varepsilon| = 0$. С этой целью строится *t*-статистика

$$\hat{t} = \frac{|\bar{\varepsilon}|}{S_\Sigma}\sqrt{n}, \qquad S_\Sigma = \sqrt{\frac{\sum_{t=1}^{n}(\varepsilon_t - \bar{\varepsilon})^2}{n-1}}, ,$$

где $\bar{\varepsilon}$ - среднее арифметическое значение уровней ряда остатков ε_t, S_Σ - среднеквадратическое отклонение для этой последовательности, рассчитанное по формуле для малой выборки. На уровне значимости α гипотеза отклоняется, если $\hat{t} > t_{\alpha,\nu}$, где $t_{\alpha,\nu}$ - критерий распределения Стьюдента с доверительной вероятностью (1-α) и $\gamma = n-1$ степенями свободы.

2. Для проверки условия случайности возникновения отдельных отклонений от тренда часто используется *критерий поворотных точек*. Значение случайной переменной считается поворотной точкой, если оно одновременно больше (или одновременно меньше) значений предыдущего и последующего члена. Если остатки случайны, то поворотная точка приходится в среднем примерно на каждые 1,5 наблюдения.

Существует определённая зависимость между средней арифметической $\overline{p}$, дисперсией σ_p^2 количества поворотных точек в ряде остатков p и числом членов исходного ряда наблюдений n. С использованием этих зависимостей критерий случайности отклонений от тренда при с доверительной вероятностью 0,95 можно представить в виде:

$$p > \left[\frac{2}{3}(n-2) - 1{,}96\sqrt{\frac{16n-29}{90}} \right],$$

где квадратные скобки означают, что от результата вычисления в правой части необходимо взять целую часть (не путать с процедурой округления!).

Если неравенство (5.6) не выполняется, то ряд остатков нельзя назвать случайным (то есть он содержит регулярную компоненту) и, следовательно, модель не является адекватной.

3. Наличие (отсутствие) автокорреляции в отклонениях ε_t фактических значений от модели роста проще всего проверить с помощью критерия Дарбина-Уотсона. С этой целью строится статистика Дарбина-Уотсона (*d* – статистика), в основе которой лежит расчётная формула

$$d = \frac{\sum_{t=1}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n} \varepsilon_t^2}.$$

Для формулирования вывода о наличии (отсутствии) автокорреляции полученное значение необходимо сравнить с критическими значениями d_1 (нижнее) и d_2 (верхнее), которые определяются по специальным таблицам для трёх уровней значимости (α =0,01; α =0,025; α =0,05). При сравнении могут возникнуть следующие ситуации: $d < d_1$ - остатки содержат автокорреляцию; $d_1 < d < d_2$ - область неопределённости, когда нет оснований принять или отвергнуть гипотезу о существовании автокорреляции; $d_2 < d < 2$ - ряд остатков некоррелирован. Если d превышает 2, то это свидетельствует о наличии отрицательной корреляции. Перед входом в таблицу такие значения следует преобразовать по формуле $d' = 4 - d$.

Если установлено наличие автокорреляции остатков, нужно улучшить модель (изменить кривые роста, попытаться выделить дополнительные регулярные компоненты и т.п.). Если же ситуация оказалась неопределённой, применяют другие критерии. В частности можно воспользоваться первым коэффициентом автокорреляции:

$$r_1 = \frac{\left(\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1}\right)}{\sum_{t=1}^{n} \varepsilon_t^2}.$$

Для суждения о наличии или отсутствии автокорреляции с исследуемом ряду фактическое значение коэффициента автокорреляции (5.8) сопоставляется с табличным (критическим) $r_{1кр}$ для 5%-го или 1%-го уровня значимости (вероятность допустить ошибку при принятии гипотезы о независимости уровней ряда). Если $r_1 < r_{1кр}$, то гипотеза об отсутствии автокорреляции в ряду может быть принята. Когда же фактическое значение больше табличного, делают вывод о наличии автокорреляции во временном ряду.

4. Неизменность дисперсии остатков при изменении факторной переменной (исследование на гетероскедастичность) обычно проверяется с помощью трёх тестов, в которых делаются различные предположения о зависимости между дисперсией случайной компоненты и факторной переменной: тест ранговой корреляции Спирмена, тест Голдфельда-Квандта и тест Глейзера.

При малом объёме выборки для оценки гетероскедастичности может использоваться метод Голдфельда-Квандта. Для проведения такого теста необходимо выполнить следующие шаги:

- упорядочить n наблюдений по мере возрастания переменной x;

- разделить совокупность наблюдений на две группы (соответственно с малыми и большими значениями фактора x) и построить по каждой из групп уравнение регрессии

- определить остаточную сумму квадратов для первой регрессии $S_1 = \sum_{i=1}^{n_1} (y_i - y_{p1i})^2$ и второй регрессии $S_2 = \sum_{i=n_1+1}^{n} (y_i - y_{p2i})^2$.

- вычислить отношения Fнабл = S2/S1 (или S1/S2). В числителе должна быть большая сумма квадратов.

- полученное отношение следует сравнить с Fкр(α, k1, k2), где k1 = n1 – m, k2 = n2 – m. Здесь n1 и n2 – количество наблюдений попавших в 1-ю и 2-ю группы. Если Fнабл > Fкр , то гетероскедастичность имеет место, то есть условие о неизменности дисперсии при изменении факторной переменной не выполняется.

5. Соответствие ряда остатков нормальному закону распределения проверим с помощью *R/S* – критерия:

$$R/S = \frac{(\varepsilon_{\max} - \varepsilon_{\min})}{S_{\Sigma}}.$$

Полученное значение проверяется на предмет попадания в интервал, границы которого являются табличными значениями, и зависят от уровня доверия α и количества наблюдений n.

Если все четыре пункта проверки 1-5 дают положительный результат, делается вывод о том, что выбранная трендовая модель является адекватной реальному ряду наблюдений. Только в этом случае её можно использовать для построения прогнозных оценок. В противном случае модель нужно улучшать.

2.7. Оценка качества уравнения регрессии

Для общей оценки качества построенной эконометрической определяются такие характеристики как коэффициент детерминации, индекс корреляции, средняя относительная ошибка аппроксимации, а также проверяется значимость уравнения регрессии с помощью *F*-критерия Фишера. Перечисленные характеристики являются достаточно универсальными и могут применяться как для линейных, так и для нелинейных моделей, а также моделей с двумя и более факторными переменными. Определяющее значение при вычислении всех перечисленных характеристик качества играет ряд остатков ε_i, который вычисляется путем вычитания из фактических (полученных по наблюдениям) значений исследуемого признака y_i значений, рассчитанных по уравнению модели y_{pi}.

Коэффициент детерминации

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \varepsilon_i^2}{\sum_{i=1}^{n} (y_i - y_{cp})^2} = \frac{\sum_{i=1}^{n} (y_{pi} - y_{cp})^2}{\sum_{i=1}^{n} (y_i - y_{cp})^2} \qquad (2.9)$$

показывает, какая доля изменения исследуемого признака учтена в модели. Другими словами коэффициент детерминации показывает, какая часть изменения исследуемой переменной может быть вычислена, исходя из изменений включённых в модель факторных переменных с помощью выбранного типа функции, связывающей факторные переменные и исследуемый признак в уравнении модели.

Коэффициент детерминации R^2 может принимать значения от 0 до 1. Чем ближе коэффициент детерминации R^2 к единице, тем лучше качество модели.

Индекс корреляции можно легко вычислить, зная коэффициент детерминации:

$$R = \sqrt{R^2}\ . \tag{2.10}$$

Индекс корреляции R характеризует тесноту выбранного при построении модели типа связи между учтёнными в модели факторами и исследуемой переменной. В случае линейной парной регрессии его значение по абсолютной величине совпадает с коэффициентом парной корреляции $r(x, y)$, который мы рассмотрели ранее, и характеризует тесноту линейной связи между x и y. Значения индекса корреляции, очевидно, также лежат в интервале от 0 до 1. Чем ближе величина R к единице, тем теснее выбранный вид функции связывает между собой факторные переменные и исследуемый признак, тем лучше качество модели.

Средняя относительная ошибка аппроксимации

$$E_{отн.ср} = 100 * \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\varepsilon_i}{y_i} \right| \tag{2.11}$$

выражается в процентах и характеризует точность модели. Приемлимая точность модели при решении практических задач может определяться, исходя из соображений экономической целесообразности с учётом конкретной ситуации. Широко применяется критерий, в соответствии с которым точность считается удовлетворительной, если средняя относительная погрешность меньше 15%. Если $E_{отн.ср.}$ меньше 5%, то говорят, что модель имеет высокую точность. Не рекомендуется применять для анализа и прогноза модели с неудовлетворительной точностью, то есть, когда $E_{отн.ср.}$ больше 15%.

F-критерий Фишера используется для оценки значимости уравнения регрессии. Расчётное значение F-критерия определяется из соотношения:

$$F = \frac{(n - m - 1) * R^2}{m * (1 - R^2)}. \qquad (2.12)$$

Критическое значение *F*-критерия определяется по таблицам при заданном уровне значимости α и степенях свободы $\nu_1 = m$, $\nu_2 = n - m - 1$(можно использовать функцию FРАСПОБР в Excel). Здесь, по-прежнему, *m* – число факторов, учтённых в модели, *n* – количество наблюдений. Если расчётное значение больше критического, то уравнение модели признаётся значимым. Чем больше расчётное значение *F*-критерия, тем лучше качество модели.

Определим характеристики качества построенной нами линейной модели для **Примера 1**. Воспользуемся данными Таблицы 2. *Коэффициент детерминации*:

$$R^2 = 1 - \frac{6{,}43^2 + 0{,}43^2 + \ldots + 1{,}43^2}{(-16{,}07)^2 + (-14{,}57)^2 + \ldots + 23{,}93^2} \approx 0{,}901.$$

Следовательно, в рамках линейной модели изменение объёма продаж на 90,1% объясняется изменением температуры воздуха.

Индекс корреляции

$$R = \sqrt{0{,}85} \approx 0{,}95 .$$

Значение индекса корреляции в случае парной линейной модели как мы видим, действительно по модулю равно коэффициенту корреляции между соответствующими переменными (объём продаж и температура). Поскольку полученное значение достаточно близко к единице, то можно сделать вывод о наличии тесной линейной связи между исследуемой переменной (объём продаж) и факторной переменноё (температура).

F-критерий Фишера

$$F = \frac{(7 - 1 - 1) * 0{,}901}{(1 - 0{,}901)} \approx 45{,}73$$

Критическое значение Fкр при α = 0,1; ν1=1; ν2=7-1-1=5 равно 4,06. Расчётное значение F-критерия больше табличного, следовательно, уравнение модели является значимым.

Средняя относительная ошибка аппроксимации

$$E_{отн.ср.} = \frac{100}{7}\left(\frac{6,43}{2} + \frac{0,43}{3,5} + \ldots \frac{1,43}{42}\right) \approx 76,09$$

Построенная линейная модель парной регрессии имеет неудовлетворительную точность (>15%), и её не рекомендуется использовать для анализа и прогнозирования.

В итоге, несмотря на то, что большинство статистических характеристик удовлетворяют предъявляемым к ним критериям, линейная модель парной регрессии непригодна для прогнозирования объёма продаж в зависимости от температуры воздуха. Нелинейный характер зависимости между указанными переменными по данным наблюдений достаточно хорошо виден на Рис.1. Проведённый анализ это подтвердил.

2.8. Нелинейные модели парной регрессии

Если между исследуемыми и факторными переменными связь имеет нелинейный характер, то для построения модели необходимо использовать нелинейные функции.

Рассмотрим наиболее распространённые парные нелинейные модели.

Парабола второй степени определяет следующий вид модели:

$$y_p = a + bx + cx^2. \tag{2.13}$$

Параболическую модель целесообразно использовать, если связь меняет свой характер: прямая связь меняется на обратную или, наоборот, обратная связь меняется на прямую. Например, размер заработной платы работников физического труда в среднем растёт до некоторого возраста, а затем начинает убывать. Для определения параметров модели *a, b, c* модель (2.13) сводится путём замены переменных $x_1 = x$; $x_2 = x^2$ к линейной модели двухфакторной модели

$$y_p = a + bx_1 + cx_2 \tag{2.14}$$

Для оценки параметров модели вида (2.14), как будет показано далее, используется метод наименьших квадратов (МНК).

В основе ***гиперболической модели*** лежит уравнение гиперболы:

$$y_p = a + \frac{b}{x} \tag{2.15}$$

Классическим примером гиперболической модели является *кривая Филипса,* характеризующая нелинейное соотношение между нормой безработицы *x* и процентом прироста заработной платы *y*: при росте *x* до некоторого уровня *y* также растёт, а при дальнейшем росте x рост y приостанавливается. Этот же характер связи проявляется при изучении зависимости расходов на единицу продукции сырья, материалов, топлива (то есть переменных затрат) от объёма выпускаемой продукции. Другим примером гиперболической зависимости является зависимость времени оборота товаров в зависимости от величины товарооборота. Кривые Энгеля, описывающие долю доходов, расходуемых на непродовольственные товары, в зависимости от размера доходов также описываются гиперболическими функциями.

Сделав замену $x_1 = 1/x$, мы сведём уравнение (2.15) к линейному виду:

$$y_p = a + bx_1, \qquad (2.16)$$

для оценки параметров которого используется МНК.

Степенная модель

$$y_p = a * x^b \qquad (2.17)$$

применяется для описания изменения спроса при изменении цены на товар. Параметр b в ней показывает, на сколько процентов уменьшится в среднем спрос, если цена увеличится на 1% (то есть b – отрицательная величина) и называется *коэффициентом эластичности*. Логарифмирование соотношения (2.17) приводит его к линейному виду:

$$\ln y_p = \ln a + b * \ln x \qquad (2.18)$$

Применение метода наименьших квадратов (с использованием прологарифмированных данных рядов наблюдений x и y) позволит нам найти коэффициенты уравнения (2.18) $\ln a$ и b, тем самым позволит найти параметры исходной степенной модели a и b.

В эконометрических исследованиях применяется также ***показательная модель***:

$$y_p = a * b^x. \qquad (2.19)$$

Она также сводится к линейному виду путём логарифмирования:

$$\ln y_p = \ln a + \ln b * x. \qquad (2.20)$$

После логарифмирования ряда фактических значений y и применения МНК получим значения $\ln a$ и $\ln b$. Возводя основание логарифма (в данном случае число e) в степень с использованием полученных значений, мы получим оценки параметров a и b исходной показательной модели.

Необходимо отметить, что не все нелинейные модели можно свести к линейной. Если модель не сводится к линейной, то она называется ***внутренне нелинейной***.

Построим ***показательную модель*** по данным **Примера 1**. Для этого построим таблицу, аналогичную Таблице 2, в качестве исходных данных которой будут выступать *x1 и,* $z = \ln y$.

Таблица 3

	x1	z	$x1_i - x1_{ср}$	$z_i - z_{ср}$	$(x1_i - x1_{ср})^2$	$(x1_i - x1_{ср})*(z_i - z_{ср})$
	5,0	0,69	-15,00	-1,67	225,000	25,084
	10,0	1,25	-10,00	-1,11	100,000	11,126
	15,0	1,61	-5,00	-0,76	25,000	3,780
	20,0	2,48	0,00	0,12	0,000	0,000
	25,0	3,09	5,00	0,73	25,000	3,628
	30,0	3,69	10,00	1,32	100,000	13,235
	35,0	3,74	15,00	1,37	225,000	20,584
Сумма					700,000	77,437
Среднее	20,00	2,37	**ln b**	**0,111**	**ln a**	**0,153**

Тогда $a = e^{\ln a} = e^{0,153} \approx 1,165; \quad b = e^{\ln b} = e^{0,111} \approx 1,117$. Зная параметры степенной модели a и b, мы можем вычислить расчётные значения исследуемого признака по формуле (2.17) и составить ряд остатков.

Таблица 4

x1	5,0	10,0	15,0	20,0	25,0	30,0	35,0
y	2,0	3,5	5,0	12,0	22,0	40,0	42,0
y_p	2,03	3,52	6,12	10,65	18,51	32,19	55,97
ε	-0,03	-0,02	-1,12	1,35	3,49	7,81	-13,97

Вычислим характеристики качества полученной показательной модели:

$$R^2 = 1 - \frac{0{,}03^2 + 0{,}02^2 + \ldots + 13{,}97^2}{16{,}07^2 + 14{,}57^2 + \ldots + 23{,}93^2} \approx 0{,}845; \qquad R = \sqrt{0{,}845} \approx 0{,}92;$$

$$F = \frac{5 * 0{,}845}{(1 - 0{,}845)} \approx 27{,}2; \qquad E_{отн.ср.} = \frac{100}{7}\left(\frac{0{,}03}{2} + \frac{0{,}02}{3{,}5} + \ldots + \frac{13{,}97}{42}\right) \approx 14{,}9.$$

Характеристики качества показательной модели оказались лучше соответствующих характеристик линейной модели. Точность модели можно считать удовлетворительной.

Построив несколько моделей, выбрав из них лучшую, удовлетворяющую необходимым требованиям к качеству и точности модели, мы можем использовать эту модель для прогнозирования.

2.9. Прогнозирование с применением парного уравнения регрессии

Регрессионные модели могут использоваться для прогнозирования возможных ожидаемых значений исследуемой переменной при заданных (или определённых за рамками модели) значениях факторной переменной. При этом различают точечный и интервальный прогнозы.

Рассмотрим прогнозирование на основе ***парной линейной модели регрессии***

$y_p = a_0 + a_1 * x$,

Точечный прогноз вычисляем путём подстановки в уравнение прогнозного значения факторной переменной:

$$y_{прогн}^{точ} = a_0 + a_1 * x_{прогн}. \qquad (2.21)$$

Вероятность реализации точечного прогноза практически равна нулю. Поэтому в дополнение к точечному прогнозу рассчитывается средняя ошибка прогноза или доверительный интервал прогноза с достаточно большой надёжностью. Размах прогнозного интервала L зависит от стандартной ошибки (3.8), удаления $x_{прогн}$ от своего среднего значения в ряде наблюдений $x_{ср}$, количества наблюдений n и уровня значимости прогноза α :

$$L = S_{ст} * t_{\alpha, n-m-1} * \sqrt{1 + \frac{1}{n} + \frac{(x_{прогн} - x_{ср})^2}{\sum_{i=1}^{n}(x_i - x_{ср})^2}}\ . \qquad (2.22)$$

Тогда фактические значения исследуемого признака с вероятностью (1-α) попадут в интервал

$$y_{прогн} \in \left[y_{прогн}^{точеч} - L;\ y_{прогн}^{точеч} + L \right] \qquad (2.23)$$

Чем больше количество наблюдений *n* и чем ближе прогнозное значение факторной переменной $x_{прогн}$ к среднему в ряду наблюдений значению $x_{ср}$, тем меньше прогнозный интервал, то есть лучше качество прогнозирования. Качество самой эконометрической модели влияет на величину прогнозного интервала через стандартную ошибку, которая зависит от величин элементов ряда остатков ε_i. Чем хуже качество модели, тем больше величины остатков ε, тем больше размах доверительного интервала. Наконец, на величину прогнозного интервала влияет задаваемый уровень значимости (вероятность ошибки). Чем меньше мы задаём уровень значимости, тем больше будет надёжность прогноза. Однако размах доверительного интервала при этом будет расти, поскольку величина *t*-статистики будет увеличиваться.

При определённых значениях размаха доверительного интервала прогноз теряет актуальность. Например, прогноз температуры воздуха на завтра с размером прогнозного интервала в 20-30 градусов никого не интересует.

Рассчитаем точечный и интервальный прогноз для объёма продаж в Примере 1 с использованием построенной нами в п. 2.4 линейной модели парной регрессии. Прогнозное значение факторной переменной $x1_{прогн}$ мы можем взять по данным Гидрометеоцентра, который, в свою очередь, делает прогноз на основе соответствующих математических моделей. Допустим прогнозное значение температуры воздуха $x1_{прогн}$ = 28 градусов. Тогда точечный прогноз по линейной модели:

$$y_{прогн}^{точ} = -11{,}93 + 1{,}5 * 28 \approx 30{,}07.$$

Для построения доверительного интервала используем стандартную ошибку, вычисленную нами в п. 2.5 и данные Таблицы 2. С учётом $t_{0.1;5} \approx 2{,}02$ получим размах доверительного интервала:

$$L = 5{,}869 * 2{,}02 * \sqrt{1 + \frac{1}{7} + \frac{(28-20)^2}{700}} \approx 13{,}14.$$

Следовательно, ожидаемое значение объёма продаж с вероятностью 90% будет находиться в интервале:

$$y_{прогн} \in [16{,}93;\ 43{,}21].$$

Прогнозный интервал получился достаточно большой, что и следовало ожидать исходя из неудовлетворительной точности линейной модели в данной задаче.

Прогнозирование на основе парных нелинейных моделей, которые заменой переменных сводятся к линейной модели, можно произвести, применив формулы (2.21)-(2.23) к линеаризованному виду нелинейной модели. Если исследуемая переменная не участвовала в заменах переменных, то полученный прогнозный интервал является конечным результатом прогнозирования. Если же мы произвели замену исследуемой переменной, то с помощью обратной замены мы должны будем вычислить прогнозный интервал для исходной исследуемой переменной.

Построим прогноз по данным нашего Примера 1 на основе построенной в п.2.7 парной показательной модели, у которой характеристики точности были выше, чем у линейной. В линеаризованном виде показательную модель можно записать в виде:

$$z_p = 0{,}153 + 0{,}111 * x1, \qquad где \quad z_p = \ln y_p .$$

Построим дополнительную вспомогательную таблицу:

Таблица 5

x1	5,0	10,0	15,0	20,0	25,0	30,0	35,0
z	0,69	1,25	1,61	2,48	3,09	3,69	3,74
z_p	0,71	1,26	1,81	2,37	2,92	3,47	4,02
ε	-0,013	-0,006	-1,203	0,120	0,173	0,217	-0,287

Значение точечного прогноза для переменной $z = ln\,y$ будет равно:

$$z^{точ}_{прогн} = 1{,}153 + 1{,}111 * 28 \approx 3{,}25 .$$

Для построения прогнозного интервала вычислим стандартную ошибку линеаризованной модели:

$$S_{ст} = \sqrt{\frac{0{,}013^2 + 0{,}006^2 + \ldots + 0{,}287^2}{7-1-1}} \approx 0{,}207 ,$$

а с её использованием размах прогнозного интервала для z:

$$L = 0{,}207 * 2{,}02 * \sqrt{1 + \frac{1}{7} + \frac{(28-20)^2}{700}} \approx 0{,}464.$$

Таким образом, мы получаем прогнозный интервал:

$$z_{прогн} \in [2{,}79;\ 3{,}71].$$

Для определения прогнозного интервала исходной исследуемой переменной применим обратную замену:

$y_{прогн}^{нижн} = e^{2.79} \approx 16{,}22; \quad y_{прогн}^{верх} = e^{3.71} \approx 41{,}04.$

В итоге получим прогнозный интервал для исходной исследуемой переменной с использованием показательной модели:

$y_{прогн} \in [16{,}22;\ 41{,}04].$

Длина интервала получилась меньше, чем длина прогнозного интервала, построенного с использованием линейной модели, чего и следовало ожидать, учитывая лучшие характеристики качества показательной модели по сравнению с линейной.

Однако, величина прогнозного интервала осталась достаточно большой, то есть прогноз остался достаточно грубым. Одним из способов улучшения качества модели, а значит, качества прогнозирования является введение в рассмотрение дополнительных факторных переменных, влияющих на исследуемый признак.

Тема 3. Модель множественной регрессии

3.1. Общий вид линейной модели множественной регрессии

Линейная модель множественной регрессии имеет вид:

$$y_p = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_m x_m, \qquad (3.1)$$

где y_p - расчётные значения исследуемой переменной, $x_1, x_2, \ldots, x_m$ - факторные переменные. Каждый из коэффициентов уравнения $a_1, a_2 \ldots, a_m$ имеет следующую экономическую интерпретацию: он показывает, насколько изменится значение исследуемого признака при изменении соответствующего фактора на 1 при неизменных прочих факторных переменных.

Фактическое значение исследуемой переменной тогда представимо в виде:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_m x_m + \varepsilon \qquad (3.2)$$

Для адекватности модели необходимо, чтобы случайная величина ε, являющаяся разностью между фактическими и расчётными значениями, имела нормальный закон распределения с математическим ожиданием равным нулю и постоянной дисперсией σ^2.

Имея *n* наборов данных наблюдений, с использованием представления (2.2), мы можем записать *n* уравнений вида:

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \ldots + a_m x_{mi} + \varepsilon_i, \qquad (3.3)$$

где $y_i, x_{1i}, x_{2i}, \ldots, x_{mi}$ - значения исследуемой и факторных переменных в *i*-м наблюдении, а ε_i – отклонение фактического значения y_i от расчётного значения y_{pi}, которое может быть рассчитано с помощью (2.1) по значениям факторных переменных $x_{1i}, x_{2i}, \ldots, x_{mi}$ в *i*-м наблюдении.

Систему уравнений (2.3) удобно исследовать в матричном виде:

$$Y_в = X_в A + E\ , \qquad (3.4)$$

где $Y_в$ – вектор выборочных данных наблюдений исследуемой переменной (*n* элементов), $X_в$ – матрица выборочных данных наблюдений факторных переменных ($n \times (m+1)$ элементов), А – вектор параметров уравнения (*m+1* элементов), а Е – вектор случайных отклонений (*n* элементов):

$$Y_в = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, \quad X_в = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix}, \quad A = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{pmatrix}, \quad E = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_m \end{pmatrix}. \tag{3.5}$$

3.2. Оценка параметров модели с помощью МНК. Отбор факторов

При построении модели множественной регрессии возникает необходимость оценки (вычисления) коэффициентов линейной функции, которые в матричной форме записи обозначены вектором *А*. Формулу для вычисления параметров регрессионного уравнения методом наименьших квадратов (МНК) по данным наблюдений приведём без вывода:

$$A = (X_в^T X_в)^{-1} X_в^T Y_в\ . \tag{3.6}$$

При $m = 1$ соотношение (3.6) принимает вид (2.5). Нахождение параметров с помощью соотношения (3.6) возможно лишь тогда, когда между различными столбцами и различными строками матрицы исходных данных X отсутствует строгая линейная зависимость (иначе не существует обратная матрица). Это условие не выполняется, если существует линейная или близкая к ней связь между результатами двух различных наблюдений, или же если такая связь существует между двумя различными факторными переменными. Линейная или близкая к ней связь между факторами называется ***мультиколлениарностью***. Чтобы избавиться от мультиколлениарности, в модель включают один из линейно связанных между собой факторов, причём тот, который в большей степени связан с исследуемой переменной.

На практике чтобы избавиться от мультиколлениарности мы будем проверять *для каждой пары факторных переменных* выполнение следующих условий:

$$\begin{cases} \left| r_{x_i x_j} \right| < 0{,}8 \\ \left| r_{x_i x_j} \right| < \left| r_{y x_i} \right| \\ \left| r_{x_i x_j} \right| < \left| r_{y x_j} \right| \end{cases} . \tag{3.7}$$

То есть коэффициент корреляции между двумя факторными переменными должен быть меньше 0,8 и, одновременно, меньше коэффициентов корреляции между исследуемой переменной и каждой из этих двух факторных переменных. *Если хотя бы одно из условий*

(3.7) не выполняется, то в модель включают только один из этих двух факторов, а именно, тот, у которого модуль коэффициента корреляции с Y больше.

Пример. Будем считать, что торговое предприятие из Примера 1 находится в г. Барнауле, *x1* – температура воздуха в г. Барнауле. Дополним данные наблюдений значениями факторной переменной *x3* – значениями температуры воздуха в г. Новосибирске в период наблюдений:

Таблица 6

y	x1	x2	x3
2	5,0	20	4
3,5	10,0	20	8
5	15,0	20	14
12	20,0	20	21
22	25,0	20	23
40	30,0	25	30
42	35,0	50	32

Проверим наличие мультиколлениарности между факторными переменными, произведём отбор факторов и найдём параметры линейной модели множественной регрессии. Для нахождения коэффициентов парной корреляции можно воспользоваться формулой (2.1). Поскольку вычисления будут достаточно громоздкими, эффективнее использовать средства табличного процессора Microsoft Excel. Применив к данным из Таблицы 6 обработку *Сервис/ Анализ данных/ Корреляция*, получим набор коэффициентов парной корреляции:

	y	x1	x2	x3
y	1			
x1	0,949	1		
x2	0,723	0,690	1	
x3	0,938	0,992	0,630	1

Проверим выполнение условий (3.7) для каждой пары факторных переменных.

Для *x1, x2*:

$|r(x1, x2)| = 0{,}690 \quad < \quad 0{,}8$ - выполняется,

$|r(x1, x2)| = 0{,}690 \quad < \quad |r(y, x1)| = 0{,}949$ - выполняется,

$|r(x1, x2)| = 0{,}690 \quad < \quad |r(y, x2)| = 0{,}723$ - выполняется.

Все три условия (3.7) выполняются, значит мультиколлениарность между факторными переменными *x1* (температура воздуха в г. Барнауле) и *x2* (размер торговой наценки) отсутствует, то есть они могут использоваться в модели одновременно.

Для *x1, x3*:

$|r(x1, x3)| = 0{,}992 \quad < \quad 0{,}8$ - не выполняется,

$|r(x1, x3)| = 0{,}992 \quad < \quad |r(y, x1)| = 0{,}949$ - не выполняется,

$|r(x1, x3)| = 0{,}992 \quad < \quad |r(y, x3)| = 0{,}938$ - не выполняется.

Ни одно из условий не выполняется, следовательно, факторы *x1* (температура воздуха в г. Барнауле) и *x3* (температура воздуха в г. Новосибирске) мультиколлениарны, то есть не рекомендуется использовать их в модели одновременно. Поскольку $|r(y, x1)| = 0{,}949 \quad > \quad |r(y, x3)| = 0{,}938$, то фактор *x1* теснее связан с исследуемой переменной у (объём продаж), чем фактор *x3*. Поэтому исключить из рассмотрения следует фактор *x3*.

Для *x2, x3*:

$|r(x2, x3)| = 0{,}630 \quad < \quad 0{,}8$ - выполняется,

$|r(x2, x3)| = 0{,}630 \quad < \quad |r(y, x2)| = 0{,}723$ - выполняется,

$|r(x2, x3)| = 0{,}630 \quad < \quad |r(y, x3)| = 0{,}938$ - выполняется.

Все три условия выполняются, значит мультиколлениарность между факторными переменными *x2* и *x3* отсутствует, и они могут использоваться в модели одновременно.

Можно резюмировать, что в модели можно оставить либо пару факторов *x1, x2,* либо пару *x3, x2*. То есть выбор необходимо сделать между факторами *x1* и *x3*. Как уже отмечалось выше, фактор *x1* имеет преимущество, поскольку теснее, чем *x3*, связан с *y*. Поэтому модель для объёма продаж *y* мы будем строить с учётом влияния факторов *x1* и *x2*:

$$y_p = a_0 + a_1 * x1 + a_2 * x2 .$$

Для вычисления параметров модели по данным наблюдений выпишем вектор $Y_в$ и матрицу $X_в$:

$$Y_в = \begin{pmatrix} 2 \\ 3,5 \\ 5 \\ 12 \\ 22 \\ 40 \\ 42 \end{pmatrix}, \qquad X_в = \begin{pmatrix} 1 & 5 & 20 \\ 1 & 10 & 20 \\ 1 & 15 & 20 \\ 1 & 20 & 20 \\ 1 & 25 & 20 \\ 1 & 30 & 25 \\ 1 & 35 & 50 \end{pmatrix}.$$

Опуская операции транспонирования матрицы, перемножения матриц и нахождения обратной матрицы (можно воспользоваться в Excel функциями ТРАНСП, МУМНОЖ, МОБР), запишем промежуточный результат вычислений, необходимых для нахождения вектора параметров модели *A* по формуле (3.6):

$$(X_в^T X_в)^{-1} = \begin{pmatrix} 1,0065 & -0,0091 & -0,0273 \\ -0,0091 & 0,0027 & -0,0018 \\ -0,0273 & -0,0018 & 0,0026 \end{pmatrix}.$$

Продолжая операции с матрицами в соответствии с (3.6), получим искомый вектор параметров модели:

$$A = \begin{pmatrix} -14,04 \\ 1,36 \\ 0,20 \end{pmatrix}.$$

То есть мы получили уравнение линейной регрессии следующего вида:

$$y_p = -14,04 + 1,36 * x1 + 0,20 * x2 . \qquad (3.8)$$

Значения параметров модели указывают, что в среднем при увеличении температуры воздуха в г. Барнауле на 1 градус объём продаж на изучаемом предприятии увеличивается на 1,36 единицы, а при увеличении торговой наценки на 1% объём продаж увеличивается на 0,20 единицы. Последний вывод выглядит некорректно, поскольку в реальном процессе, наоборот, увеличение наценки сдерживает рост объёма продаж.

Определим по (3.8) расчётные значения исследуемой переменной для набора значений факторов, полученных в наблюдениях (Таблица 6), и составим ряд отклонений ε_i фактических значений объёма продаж от расчётных значений.

Таблица 7

y	2	3,5	5	12	22	40	42
y_p	-3,30	3,49	10,29	17,09	23,88	31,66	43,39
ε	5,30	0,01	-5,29	-5,09	-1,88	8,34	-1,39

3.3. Анализ статистической значимости параметров модели

Значимость параметров модели множественной регрессии a_j проверяется с помощью t-критерия Стьюдента аналогично тому, как мы проверяли значимость коэффициентов модели парной регрессии. Для каждого параметра уравнения вычисляется t-статистика:

$$t_{a_j} = \frac{a_j}{S_{a_j}}, \qquad где \qquad S_{a_j} = S_{ст} * \sqrt{b_{jj}}\ . \tag{3.9}$$

Здесь $S_{ст}$ – стандартная ошибка оценки, задаваемая соотношением (2.6), b_{jj} – диагональный элемент матрицы $(X_в^T X_в)^{-1}$.

Далее по таблицам (или в Excel с помощью функции СТЬЮДРАСПОБР) определяется значение $t_{кр}$ в зависимости от уровня значимости α и параметра n-m-1. Наконец, каждая из t-статистик (3.9) сравнивается с табличным значением. Если $|t_{aj}| > t_{кр}$, то коэффициент a_j считается значимым. В противном случае коэффициент не является значимым и его можно положить равным нулю, тем самым исключить из модели фактор x_j (качество модели при этом не ухудшится).

Проверим значимость коэффициентов полученного нами уравнения регрессии (3.8). Вычислим стандартную ошибку оценки:

$$S_{ст} = \sqrt{\frac{5{,}3^2 + 0{,}01^2 + \ldots + 1{,}39^2}{7-2-1}} \approx 6{,}26\,.$$

Тогда

$$t_{a_0} = \frac{-14{,}04}{\sqrt{1{,}0065} * 6{,}26} \approx -2{,}23; \quad t_{a_1} = \frac{1{,}36}{\sqrt{0{,}0027} * 6{,}26} \approx 4{,}15; \quad t_{a_2} = \frac{0{,}20}{\sqrt{0{,}0026} * 6{,}26} \approx 0{,}62.$$

Находим табличное значение $t_{кр}(0{,}1;4) = 2{,}13$. Для коэффициентов a_0, a_1 вычисленные *t*-статистики по модулю больше критического значения. Следовательно, с вероятностью 90% мы можем утверждать, что коэффициенты a_0, a_1 уравнения регрессии (3.8) являются значимыми.

$$\left|t_{a_2}\right| = 0{,}62 \quad < \quad t_{кр} = 2{,}13,$$

следовательно, коэффициент a_2 не является значимым, то есть его можно положить равным нулю, тем самым, исключив фактор *x2* из рассмотрения.

3.4. Оценка качества линейной модели множественной регрессии

Качество модели оценивается стандартным способом для уравнений регрессии: по адекватности и точности на основе анализа остатков регрессии *ε*.

Как и в случае парной линейной регрессии, ***коэффициент детерминации*** R^2 можно вычислить по формуле (2.9), ***индекс корреляции R*** (в случае линейной множественной регрессии он называется ***коэффициентом множественной регрессии***) по формуле (2.10), ***среднюю относительную ошибку*** $E_{отн.ср}$ по формуле (2.11). Процедура проверка значимости уравнения регрессии в целом также производится аналогично случаю парной регрессии. Вычисляется ***F-критерий Фишера*** по формуле (2.12), затем определяется критическое значение и сравнивается с расчётным значением.

Произведём оценку качества модели (3.8) с использованием ряда остатков, приведённого в Таблице 6 и промежуточных результатов расчётов из Таблицы 2.

$$R^2 = 1 - \frac{5{,}3^2 + 0{,}01^2 + \ldots 1{,}39^2}{16{,}07^2 + 14{,}57^2 + \ldots 23{,}93^2} \approx 0{,}91.$$

Исходя из полученного значения коэффициента детерминации, можно сказать, что в рамках линейной модели множественной регрессии изменение объёма продаж на 91% объясняется изменением температуры воздуха и торговой наценки.

$$R = \sqrt{0{,}91} \approx 0{,}954.$$

Следовательно, связь между исследуемой переменной и используемым набором факторов тесная.

$$F = \frac{(7-2-1)*0{,}91}{2*(1-0{,}91)} \approx 20{,}27 .$$

Критическое значение $F_{кр}(0{,}1;2;4) \approx 4{,}32$. Расчётное значение F-критерия больше критического, поэтому мы можем утверждать, что уравнение регрессии (3.8) является значимым.

$$E_{отн.ср.} = \frac{100}{7}\left(\frac{5{,}3}{2} + \frac{0{,}01}{3{,}5} + \ldots + \frac{1{,}39}{42}\right) \approx 63{,}73$$

Средняя относительная ошибка аппроксимации составила 63,73%, то есть точность модели следует признать неудовлетворительной и дальнейшее использованием модели признать нецелесообразным.

3.5.Оценка влияния отдельных факторов на исследуемую переменную

Важную роль при оценке влияния отдельных факторов играют коэффициенты регрессионной модели a_j. Однако непосредственно с их помощью нельзя сопоставить факторы по степени их влияния на зависимую переменную из-за различия единиц измерения и разного масштаба колебаний (степени колеблемости) при использовании разных наборов результатов наблюдений.

Для устранения таких различий применяются частные ***коэффициенты эластичности***:

$$Э_j = a_j * \frac{x_{jср}}{y_{ср}}, \tag{3.10}$$

где $x_{jср}, y_{ср}$ - средние значения переменных в рядах наблюдений и ***бета - коэффициенты***

$$\beta_j = a_j * \frac{S_{x_j}}{S_y}, \tag{3.11}$$

где S_{x_j}, S_y - среднеквадратические отклонения переменных:

$$S_{x_j} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - x_{jср})^2}, \quad S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - y_{ср})^2}. \tag{3.12}$$

Коэффициент эластичности показывает, на сколько процентов изменяется исследуемая переменная при изменении факторной переменной на 1 процент. Если коэффициент эластичности меньше 0, то при увеличении значения фактора исследуемая переменная уменьшается. Таким образом, коэффициенты эластичности можно сравнивать между собой по модулю для выяснения того, изменения какого фактора больше влияют на изменение исследуемой переменной. Однако коэффициент эластичности не учитывает степень колеблемости факторов.

Бета – коэффициент показывает, на какую часть величины среднеквадратического отклонения S_y изменится переменная y с изменением соответствующей независимой переменной x_j на величину своего среднеквадратического отклонения при фиксированном уровне значений остальных факторных переменных.

Указанные коэффициенты позволяют упорядочить факторы по степени их влияния на исследуемую переменную.

Долю влияния фактора в суммарном влиянии всех факторов можно оценить по величине ***дельта – коэффициентов***:

$$\Delta_j = r(x_j, y) * \frac{\beta_j}{R^2}, \qquad (3.13)$$

где $r(x_j, y)$ - коэффициент парной корреляции между фактором x_j и исследуемой переменной y.

Рассчитаем значения коэффициентов эластичности, бета - и дельта – коэффициентов для уравнения регрессии (3.8), частично используя уже произведённые вычисления.

$$Э_1 = 1{,}36 * \frac{20}{18{,}07} \approx 1{,}5; \quad Э_2 = 0{,}2 * \frac{25}{18{,}07} \approx 0{,}27;$$

Таким образом, при увеличении температуры на 1% следует ожидать увеличения объёма продаж на 1,5%, а при увеличении торговой наценки на 1% ожидаемое увеличение объёма продаж составляет 0,27%.

Для определения бета – коэффициентов рассчитаем среднеквадратические отклонения:

$$S_y = \sqrt{\frac{16{,}07^2 + 14{,}57^2 + \ldots 23{,}93^2}{7-1}} \approx 18{,}07;$$

$$S_{x_1} = \sqrt{\frac{700}{7-1}} \approx 10{,}8; \qquad S_{x_1} = \sqrt{\frac{(20-25)^2 + (20-25)^2 + \ldots + (50-25)^2}{7-1}} \approx 11{,}18.$$

Тогда

$$\beta_1 = 1{,}36 * \frac{10{,}8}{18{,}07} \approx 0{,}86; \qquad \beta_2 = 0{,}20 * \frac{11{,}18}{18{,}07} \approx 0{,}13.$$

Значения бета – коэффициентов показывают, что при изменении *x1* на одно своё среднеквадратическое отклонение значение *y* в среднем будет изменяться на 0,86 от своего среднеквадратического отклонения, а при изменении *x2* на величину S_{x_2} значение *y* в среднем изменится на $0{,}13 * S_y$.

Наконец, дельта – коэффициенты:

$$\Delta_1 = 0{,}949 * \frac{0{,}86}{0{,}91} \approx 0{,}9; \qquad \Delta_2 = 0{,}723 * \frac{0{,}13}{0{,}91} \approx 0{,}1.$$

Значения дельта – коэффициентов показывают, что доля влияния первого фактора составила 90%, а второго – 10%. Заметим, что сумма дельта – коэффициентов всегда равна 1. Этот факт можно использовать для проверки правильности произведённых вычислений.

3.6. Построение прогнозов на основе модели множественной линейной регрессии

Одной из важнейших целей построения эконометрической модели является прогнозирование поведения исследуемого процесса или объекта. Если в модели присутствует фактор времени, то прогнозирование подразумевает предсказание состояния системы в будущем. Если фактор времени в модели отсутствует, то прогнозирование величины исследуемой переменной (вычисление $y_{прогн}$) производится при некотором наборе (наборах) значений факторных переменных. Эти значения факторов ($x_{прогн1}$, $x_{прогн2}$, … , $x_{прогн\,m}$) должны быть заданы исследователем или вычислены с помощью других моделей.

Как и в случае парной регрессии вычисляются точечное и интервальное прогнозные значения исследуемой переменной.

Точечный прогноз осуществляется подстановкой прогнозного набора факторных переменных в уравнение регрессии:

$$y_{прогн}^{точ} = a_0 + a_1 x_{прогн1} + a_2 x_{прогн2} + \ldots + a_m x_{прогнm}. \qquad (3.14)$$

Если прогноз осуществляется не для одного набора факторных переменных, а для некоторого ряда наборов, то ряд точечных прогнозов исследуемой переменной можно представить в виде вектора, и вычислять его удобнее с использованием операций с матрицами:

$$Y_{прогн}^{точ} = X_{прогн} A, \tag{3.15}$$

где

$$Y_{прогн}^{точ} = \begin{pmatrix} y_{1прогн}^{точ} \\ y_{2прогн}^{точ} \\ \dots \\ y_{l\,прогн}^{точ} \end{pmatrix}; \qquad X_{прогн} = \begin{pmatrix} 1 & x_{1прогн1} & x_{1прогн2} & \dots & x_{1прогнm} \\ 1 & x_{2прогн1} & x_{2прогн2} & \dots & x_{2прогнm} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{l\,прогн1} & x_{l\,прогн2} & \dots & x_{l\,прогнm} \end{pmatrix}; \quad A = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{pmatrix}. \tag{3.16}$$

Интервальный прогноз в рамках модели множественной регрессии строится с использованием соотношений, являющихся обобщением формул (2.22), (2.23), позволяющих строить прогноз на основе парной регрессионной модели.

Для нахождения размаха доверительного интервала необходимо вычислить матрицу *V*:

$$V = X_{прогн} \times (X_{в}^{T} \times X_{в})^{-1} \times X_{прогн}^{T}. \tag{3.17}$$

В выражении (3.17) участвуют матрица $X_в$, составленная из значений факторных переменных, имевших место в рядах наблюдений по правилу (3.5), и матрица $X_{прогн}$, составленная из прогнозируемых значений факторных переменных по правилу (3.16). Размерность матрицы *V* равна $(l \times l)$, то есть зависит от числа прогнозируемых наборов факторных переменных. Если мы хотим рассчитать прогноз для одного набора факторных переменных, то получим матрицу *V* размером (1×1), то есть число. Размах прогнозного интервала для *i*-го набора факторных переменных равен:

$$L_i = S_{cm} * t_{\alpha, n-m-1} * \sqrt{1 + v_{ii}}\ . \tag{3.18}$$

Величины $S_{cn}, t_{\alpha, n-m-1}$ вычисляются тем же образом, что и в (2.22), а v_{ii} является диагональным элементом матрицы (3.17). Тогда фактические значения исследуемой величины у для *i*-го набора значений факторных переменных с вероятностью *(1-α)* попадают в интервал:

$$y_{i\,прогн} \in \left[y_{i\,прогн}^{точеч} - L_i;\ y_{i\,прогн}^{точеч} + L_i \right]. \tag{3.19}$$

Несмотря на то, что в ходе исследования качества построенной нами модели (3.8) мы сделали вывод о нецелесообразности её использования для анализа и прогнозирования, рассчитаем прогноз для прогнозного значения температуры $x1_{прогн} = 28$ и величины торговой наценки $x2_{прогн} = 25$, то есть матрица $X_{прогн}$ примет у нас вид вектора:

$$X_{прогн} = (1;\ 28;\ 25).$$

Точечный прогноз будет тогда равен:

$$y_{прогн}^{точ} = -14{,}04 + 1{,}36 * 28 + 0{,}2 * 25 = 28{,}94.$$

Вычислим матрицу *V* по правилу (3.17), имея в виду, что матрицу $\left(X_в^T \times X_в\right)^{-1}$ мы уже вычислили в п.3.2, получим число (поскольку один прогнозируемый набор факторов)*: V=0,32*. Далее, с учётом приведённых в п.3.3 стандартной ошибки $S_{ст} = 6{,}26$ и значения $t_{\alpha, n-m-1} = 2{,}13$, получим по формуле (3.18) размах интервала: *L = 15,33*. В итоге получим прогнозный интервал для фактического значения объёма продаж:

$$y_{прогн} \in [13{,}62;\ 44{,}27].$$

Если мы сравним прогноз, полученный по двухфакторной линейной модели, с прогнозом, который мы сделали в п.2.8. на основе парной показательной модели, то увидим, что прогнозный интервал у двухфакторной модели больше, чем у однофакторной, то есть качество прогнозирования, несмотря на введение нового фактора, ухудшилось. Рекомендации о нецелесообразности использования, сделанные нами при исследовании качества линейной двухфакторной модели, оправдались.

Этот результат обусловлен, в первую очередь явно нелинейным характером связи между исследуемым объёмом продаж и основным фактором – температурой воздуха. Для улучшения парной показательной модели достаточно логично было бы ввести в модель дополнительную факторную переменную, не меняя показательной связи между объёмом продаж *y* и температурой воздуха *x1*. Это оказывается возможным с использованием техники вычислений, применявшейся нами при построении множественной линейной модели регрессии.

Построим по данным Примера 1 нелинейную модель вида:

$$y_p = a_0 * a_1^{x1} * x2^{a_2}. \tag{3.20}$$

Применив операцию логарифмирования к уравнению (3.20) и сделав замены переменных, получим уравнение линейной модели множественной регрессии:

$$z_p = A_0 + A_1 * x1 + a_2 * v2, \tag{3.21}$$

где $z_p = \ln y_p$, $v2 = \ln x2$, $A_0 = \ln a_0$, $A_1 = \ln a_1$. Соответственно, для нахождения коэффициентов линейной модели (3.21) A_0, A_1, a_2, исследования свойств полученной модели и прогнозирования, будем использовать данные наблюдений из Таблицы 6, при этом каждое из значений в первом и третьем столбцах (данные для *y* и *x2*) необходимо предварительно прологарифмировать.

Применив процедуру МНК, получим модель:

$$z_p = 1{,}86 + 0{,}13 * x1 - 0{,}63 * v2 . \qquad (3.22)$$

В соответствии с уравнением (3.22), в отличие от уравнения линейной модели (3.8), при увеличении торговой наценки объём продаж будет уменьшаться, что соответствует реальному процессу.

Произведя все операции для построения прогнозного интервала на основе линейной модели множественной регрессии, аналогично тому, как это описано выше, получим:

$$z_{прогн} \in [2{,}99; 3{,}67].$$

Тогда с учётом соотношений $e^{2,99} \approx 19{,}85$; $e^{3,67} \approx 39{,}14$ прогнозный интервал для исходной исследуемой переменной с уровнем значимости *α = 0,1*:

$$y_{прогн} \in [19{,}85; 39{,}14].$$

С помощью построения нелинейной двухфакторной модели нам удалось уменьшить длину прогнозного интервала, полученного с помощью однофакторной показательной модели. Однако, интервал остаётся достаточно большим.

Если выбрать уровень значимости *α = 0,3*, то прогнозный интервал значительно уменьшится:

$y_{прогн} \in [23{,}06; 33{,}69]$.

При этом, однако, вероятность выполнения прогноза уменьшится с 90% до 70%.

В итоге наилучшей из построенных нами по данным Примера 1 моделей оказалась нелинейная двухфакторная модель вида:

$$y_p = 6{,}436 * 1{,}332^{x1} * x2^{-0,632}.$$

Здесь использовано обратное преобразование коэффициентов: $e^{1,86} \approx 6{,}436$; $e^{0,13} \approx 1{,}332$.

3.7. Применение обработки РЕГРЕССИЯ для определения параметров модели множественной линейной регрессии и её исследования

Построение и исследование модели множественной линейной регрессии является достаточно трудоёмкой процедурой. Трудоёмкость вычислений можно существенно снизить с помощью применения в MS Excel обработки Сервис/Анализ данных/РЕГРЕССИЯ.

Рассмотрим возможности использования обработки РЕГРЕССИЯ на данных примера из п.3.2. Данные для факторной переменной x3 мы использовать не будем, поскольку x3 была удалена из рассмотрения в результате проверки факторных переменных на мультиколлениарность. После вызова обработки РЕГРЕССИЯ зададим в соответствующих окнах диапазон ячеек, в которых находятся данные для Y вместе с заголовком столбца, диапазон ячеек, в которых находятся данные для факторных переменных x1, x2 также с заголовками столбцов, поставим флажок *Метки* (указывает, что в первой строке диапазонов стоят названия столбцов), зададим начальную ячейку для выходного интервала, поставим флажок *Остатки*. После выполнения обработки в ячейках, расположенных ниже и правее ячейки, указанной нами как начальная ячейка выходного интервала будут расположены результаты. Результаты обработки группируются в 4 таблицы. Если при вызове обработки мы дополнительно поставим флажок *График остатков*, то будут выданы графики остатков, по горизонтальной оси которых будут отложены значения одной из факторных переменных, а по вертикальной – значения ряда остатков ε_i. Число графиков будет совпадать с числом факторных переменных. Рассмотрим полученные результаты.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,9540
R-квадрат	0,9102
Нормированный R-квадра	0,8653
Стандартная ошибка	6,2635
Наблюдения	7

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	2	1590,289	795,144	20,268	0,008
Остаток	4	156,926	39,231		
Итого	6	1747,214			

	Коэффициен	*Стандартная оши*	*t-статистик*	*P-Значение*	*Нижние 95%*	*Верхние 95*
Y-пересечение	-14,042	6,284	-2,235	0,089	-31,489	3,404
X1	1,359	0,327	4,155	0,014	0,451	2,267
X2	0,197	0,316	0,624	0,566	-0,680	1,075

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное Y*	*Остатки*	Е отн
1	-3,301	5,301	265,06
2	3,494	0,006	0,17
3	10,290	-5,290	105,79
4	17,085	-5,085	42,38
5	23,881	-1,881	8,55
6	31,662	8,338	20,84
7	43,390	-1,390	3,31

Во-первых, в колонке *Коэффициенты* третьей таблицы возьмём значения параметров множественной модели линейной регрессии. Уравнение модели имеет вид:

$$y_p = -14{,}04 + 1{,}36 * x1 + 0{,}20 * x2\,.$$

В колонке *t–статистика* этой же таблицы находятся t-статистики для коэффициентов уравнения регрессии. Если возьмём при α=0,1 критическое значение $t_{кр}(0{,}1;\ 7\text{-}2\text{-}1)=2{,}13$, то получим, что модули первых двух параметров превышают критической значение, а модуль третьего параметра нет. Таким образом значения a_0*=-14,04* и a_1*=1,36* следует признать значимыми, а значение a_2*=0,2* – незначимым. Следует отметить, что для определения значимости коэффициентов не обязательно определять критическое значение t-статистики. Достаточно сравнить соответствующие значение колонки *Р-Значение* с выбранным уровнем значимости α и, если оно меньше чем α, то соответствующий параметр можно признать значимым. У нас получилось 0,089 < 0,1 и 0,014 < 0,1, то есть первые два параметра можно признать значимыми с вероятностью 90%, а 0,566 > 0,1, то есть третий параметр значимым не является, то есть наценку можно исключить из рассмотрения в рамках данной модели.

В первой таблице приведено значение коэффициента детерминации *R-квадрат = 0,9102.* Следовательно, можно сделать вывод, что в рамках линейной модели множественной регрессии изменение объёма продаж на 91% объясняется изменением температуры воздуха и торговой наценки.

В колонке *F* третьей таблицы приведено значение F-статистики Фишера равное 20,268. Для оценки значимости уравнения регрессии в целом сравним его с критическим значением $F_{кр}(0,1;\ 2;\ 7\text{-}2\text{-}1) = 4,32$. Поскольку F-статистика больше критического значения можно сделать вывод о значимости уравнения в целом. Этот же вывод можно сделать без определения критического значения $F_{кр}$ путём сравнения значения из следующей колонки третьей таблицы *Значимость F*, равное 0,008, с выбранным уровнем значимости $\alpha = 0,1$ (для возможности сделать вывод о значимости уравнения в целом это значение не должно превышать выбранный уровень значимости).

Для определения средней ошибки аппроксимации можно воспользоваться имеющимся в четвёртой таблице рядом остатков ε_i (колонка *Остатки*). Однако, потребуются дополнительные вычисления. Указанную таблицу следует дополнить колонкой $E_{отн} = \left|\frac{\varepsilon_i}{Y_i}\right| * 100$, где Y_i – ряд наблюдений переменной Y (в учебных задачах задан в условии) и вычислить среднее значение для этой колонке. В результате получим:

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное Y*	*Остатки*	E отн	
1	-3,301	5,301	265,06	
2	3,494	0,006	0,17	
3	10,290	-5,290	105,79	
4	17,085	-5,085	42,38	
5	23,881	-1,881	8,55	
6	31,662	8,338	20,84	
7	43,390	-1,390	3,31	
			63,73	**E отн.ср.**

Модуль вычисляется с помощью функции ABS. Мы получили E отн.ср. = 63,73%, что значительно превышает 15%, следовательно, точность модели неудовлетворительная, и её не рекомендуется использовать для прогнозирования.

Заметим, что в первой таблице итоговых результатов имеется значение стандартной ошибки оценки, которое необходимо при построении интервального прогноза, а в последней четвёртой таблице имеется ряд расчётных значений исследуемого признака Y_{pi} (колонка *Предсказанное Y*).

Тема 4. Системы линейных одновременных уравнений

Использование одного регрессионного уравнения в экономических исследованиях часто оказывается недостаточным. На практике ряд факторных переменных чаще всего влияет на целый набор взаимозависимых результирующих переменных. Так, при оценке эффективности производства нельзя руководствоваться только моделью рентабельности. Она должна быть дополнена моделью производительности труда, а также моделью себестоимости единицы продукции. В качестве факторных переменных, при этом, могут выступать показатели квалификации сотрудников, обеспечения необходимыми средствами производства, удалённости от рынков сбыта и другие.

В том же Примере 1, помимо объёма продаж нас будут интересовать сумма затрат и прибыль. При этом сумма затрат будет зависеть от объёма продаж, а прибыль от обеих этих исследуемых переменных.

Таким образом, возникает потребность рассмотрения систем эконометрических уравнений. Выделяются три основных вида систем эконометрических уравнений: ***система независимых уравнений, система рекурсивных уравнений и система одновременых уравнений.***

В общем случае уравнения могут быть нелинейными, однако здесь мы ограничимся рассмотрением систем ***линейных*** уравнений.

Система линейных независимых уравнений имеет следующий общий вид:

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1 \\ y_2 = a_{20} + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2 \\ \ldots\ldots\ldots\ldots \\ y_k = a_{k0} + a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{km}x_m + \varepsilon_k \end{cases} \tag{4.1}$$

Уравнения системы независимых уравнений могут рассматриваться самостоятельно в произвольном порядке, то есть к каждому их них применимы все операции, которые мы рассматривали выше для линейных уравнений.

Если зависимая (исследуемая переменная) одного уравнения выступает в качестве факторных переменной в последующих уравнениях, то может быть построена модель в виде ***системы линейных рекурсивных уравнений***:

$$
\begin{cases}
y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1 \\
y_2 = b_{21}y_1 + a_{20} + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2 \\
y_3 = b_{31}y_1 + b_{32}y_2 + a_{30} + a_{31}x_1 + a_{32}x_2 + \ldots + a_{3m}x_m + \varepsilon_3 \\
\ldots\ldots\ldots\ldots \\
y_k = b_{k1}y_1 + b_{k2}y_2 + \ldots + b_{kk-1}y_{k-1} + a_{k0} + a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{km}x_m + \varepsilon_k
\end{cases}
. \qquad (4.2)
$$

Уравнения системы рекурсивных уравнений также могут рассматриваться по отдельности. В случае системы линейных уравнений параметры модели могут определяться с помощью МНК. При выполнении прогнозных значений необходимо будет производить вычисления последовательно, начиная с первого уравнения.

Наибольшее распространение в эконометрических исследованиях получила ***система одновременных (взаимозависимых) уравнений***. В ней одни и те же зависимые (исследуемые) переменные в одних уравнениях входят в левую часть, а других – в правую часть системы. Даже в простейшем случае ***системы одновременных <u>линейных</u> уравнений (***её также называют ***структурной формой модели – СФМ) :***

$$
\begin{cases}
y_1 = a_{10} + b_{12}y_2 + b_{13}y_3 + \ldots + b_{1k}y_k + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1 \\
y_2 = a_{20} + b_{21}y_1 + b_{23}y_3 + \ldots + b_{2k}y_k + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2 \\
y_3 = a_{30} + b_{31}y_1 + b_{32}y_2 + \ldots + b_{3k}y_k + a_{31}x_1 + a_{32}x_2 + \ldots + a_{3m}x_m + \varepsilon_3 \\
\ldots\ldots\ldots\ldots \\
y_k = a_{k0} + b_{k1}y_1 + b_{k2}y_2 + \ldots + b_{kk-1}y_{k-1} + a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{km}x_m + \varepsilon_k
\end{cases}
. \qquad (4.3)
$$

определение параметров модели сталкивается с большими трудностями и не всегда возможно в принципе. Для нахождения параметров модели исходная система одновременных линейных уравнений сводится к ***приведённой форме модели (ПФМ)***, которая имеет вид системы независимых переменных:

$$
\begin{cases}
y_1 = \delta_{10} + \delta_{11}x_1 + \delta_{12}x_2 + \ldots + \delta_{1m}x_m + \varepsilon_1 \\
y_2 = \delta_{20} + \delta_{21}x_1 + \delta_{22}x_2 + \ldots + \delta_{2m}x_m + \varepsilon_2 \\
\ldots\ldots\ldots\ldots \\
y_k = \delta_{k0} + \delta_{k1}x_1 + \delta_{k2}x_2 + \ldots + \delta_{km}x_m + \varepsilon_k
\end{cases}
\qquad (4.1)
$$

Такое сведение всегда возможно произвести с помощью алгебраических преобразований исходной системы уравнений. Параметры приведённой системы δ_{ij} можно находить с помощью МНК. Основная трудность заключается в том, что не всегда возможно по коэффициентам приведённой системы восстановить коэффициенты исходной системы уравнений, то есть осуществить обратный переход (подобно тому, как мы это делали, сводя нелинейное уравнение к линейному, находя параметры линейной модели, а затем производя обратный пересчёт параметров нелинейной модели).

Проблема перехода от приведённой формы (ПФМ) системы уравнений к исходной СФМ называется ***проблемой идентификации***. Различаются *идентифицируемые, неидентифицируемые и сверхидентифицируемые* модели.

1. Модель идентифицируема, если все коэффициенты исходной модели определяются однозначно, единственным образом по коэффициентам приведённой модели. Это возможно когда число параметров исходной модели равно числу параметров приведённой формы (здесь и далее не учитывается число свободных коэффициентов в уравнениях). Процедура нахождения коэффициентов идентифицируемой модели носит название косвенного метода наименьших квадратов (КМНК) и содержит следующие этапы:

а) исходная модель преобразуется в приведённую форму модели;

б) для каждого уравнения приведённой формы модели применяется обычный МНК;

в) коэффициенты приведённой модели трансформируются в коэффициенты исходной модели.

2. Модель неидентифицируема, если число параметров приведённой системы меньше чем, число параметров исходной модели, и в результате коэффициенты исходной модели не могут быть оценены через коэффициенты приведённой формы.

3. Модель сверхидентифицируема, если число приведённых коэффициентов больше числа коэффициентов в исходной модели. В этом случае на основе коэффициентов приведённой формы можно получить два и более значений одного коэффициента исходной модели. Сверхидентифицируемая модель в отличие от неидентифицируемой модели практически разрешима, но требует специальных методов исчисления параметров. Наиболее распространённым является двух шаговый метод наименьших квадратов (ДНМК). Основная идея ДНМК – на основе приведённой формы модели получить для сверхидентифицируемого уравнения (имеются критерии для определения идентифицируемости каждого уравнения исходной системы) теоретические значения исследуемых переменных, содержащегося в правой части уравнения. Далее, подставив эти значения вместо фактических значений (результатов наблюдений), применяется МНК к сверхидентифицируемому уравнению исходной системы.

Для того, чтобы модель была идентифицируема, необходимо, чтобы каждое уравнение модели было идентифицируемо. *Если хотя бы одно уравнение СФМ неидентифицируемо, то вся модель считается неидентифицируемой.*

Рассмотрим необходимые и достаточные условия идентифицируемости отдельного уравнения модели.

Необходимым условием идентифицируемости отдельного уравнения модели является ***счетное правило***. Если обозначить через H число исследуемых переменных y_l, *присутствующих* в i-м уравнении, а через D обозначить число факторных переменных x_j, *отсутствующих* в i-м уравнении, то счётное правило формулируется следующим образом:

- если $D + 1 < H$, то уравнение неидентифицируемо;
- если $D + 1 = H$, то уравнение идентифицируемо;
- если $D + 1 > H$, то уравнение сверхидентифицируемо.

Достаточное условие идентифицируемости отдельного уравнения модели выполняется, если определитель матрицы, составленной из коэффициентов *в других уравнениях* при переменных (как исследуемых y, так и факторных x), отсутствующих в данном i-м уравнении не равен нулю, а ранг этой матрицы, одновременно, не меньше, чем количество всех исследуемых переменных в системе уравнениё за вычетом 1.

Пример 4.1. Дана структурная модель:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2 \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 \end{cases}$$

Необходимо проверить каждое уравнение системы на выполнение необходимого и достаточного условия идентифицируемости и сделать вывод об идентифицируемости системы уравнений в целом.

Всего в системе присутствуют три исследуемые переменные y_1, y_2, y_3 и четыре факторные переменные x_1, x_2, x_3 и x_4.

В первом уравнении три исследуемые переменные: y_1, y_2, y_3 $(H=3)$. В нём отсутствуют две факторные переменные: x_3 и x_4 $(D=2)$. Необходимое условие идентифицируемости $D + 1 = H$ выполняется. Для проверки достаточного условия составим матрицу из коэффициентов при отсутствующих в первом уравнении x_3 и x_4, взятых во втором и третьем уравнениях:

Уравнения, из которых взяты эффициенты при переменнь	Переменные	
	x_3	x_4
2	a_{23}	a_{24}
3	0	0

Во второй строке матрицы стоят нули, поскольку x_3 и x_4 отсутствуют в третьем уравнении. Определитель такой матрицы равен нулю. Значит, достаточное условие не выполнено, и первое уравнение нельзя считать идентифицируемым. Следовательно, и вся система не является идентифицируемой. Тем не менее проверим, являются ли другие уравнения системы идентифицируемыми.

Во втором уравнении присутствуют две исследуемые переменные: y_1, y_2 $(H=2)$. В нём же отсутствует одна факторная переменная x_1 $(D=1)$. Необходимое условие идентифицируемости $D + 1 = H$ выполняется. Для проверки достаточного условия составим матрицу из коэффициентов при отсутствующих во втором уравнении y_3 и x_1 , взятых в первом и третьем уравнениях:

Уравнения, из которых взяты эффициенты при переменнь	Переменные	
	y_3	x_1
1	b_{13}	a_{11}
3	-1	a_{31}

В третьем уравнении (вторая строка таблицы) при y_3 коэффициент равен -1, так как эта пе-

$$0 = b_{31}y_1 + b_{32}y_2 - y_3 + a_{31}x_1 + a_{32}x_2$$

ременная стоит в левой части уравнения. Третье уравнение можно записать в виде

и тогда равенство $b_{33} = -1$ становится очевидным. Определитель матрицы не равен нулю. Ранг матрицы равен 2, что совпадает с числом исследуемых переменных минус один. Значит, достаточное условие выполняется, и второе уравнение является идентифицируемым.

В третьем уравнении присутствуют три исследуемые переменные: y_1, y_2, y_3 $(H=3)$. В нём отсутствует две факторные переменные x_3 и x_4 $(D=2)$. Необходимое условие идентифицируемости $D + 1 = H$ выполняется. Для проверки достаточного условия составим матрицу из коэффициентов при отсутствующих в третьем уравнении x_3 и x_4 , взятых во первом и втором уравнениях:

Уравнения, из которых взяты коэффициенты при переменных	Переменные	
	x_3	x_4
1	0	0
2	a_{23}	a_{24}

Определитель такой матрицы равен нулю. Следовательно, достаточное условие не выполняется, и третье уравнение нельзя считать идентифицируемым.

В итоге мы получили что идентифицируемым является только второе уравнение, а первое и третье уравнения не являются идентифицируемыми, поэтому система в целом не является идентифицируемой.

Рассмотрим на примере применение ***косвенного метода наименьших квадратов (косвенного МНК)***.

Пример 4.2. Пусть дана идентифицируемая модель из двух уравнений, содержащая две исследуемые и две факторные переменные:

$$\begin{cases} y_1 = a_{10} + b_{12}y_2 + a_{11}x_1 + \varepsilon_1 \\ y_2 = a_{20} + b_{21}y_1 + a_{22}x_2 + \varepsilon_2 \end{cases}.$$

Задан набор фактических данных:

№ наблюдения	y_1	y_2	x_1	x_2
1	33,0	37,1	3	11
2	45,9	49,3	7	16
3	42,2	41,6	7	9
4	51,4	45,9	10	9
5	49	37,4	10	1
6	49,3	52,3	8	16

Решение: Исходную модель можно преобразовать в приведённую форму модели вида:

$$\begin{cases} y_1 = d_{10} + d_{11}x_1 + d_{12}x_2 + \varepsilon_1 \\ y_2 = d_{20} + d_{21}x_1 + d_{22}x_2 + \varepsilon_2 \end{cases}.$$

Приведённая форма модели является системой независимых уравнений, к каждому из которых для нахождения коэффициентов можно применить МНК, подобно тому, как это делается для построения линейной модели множественной регрессии, состоящей из одного уравнения. Для нахождения коэффициентов первого уравнения мы применим в MS Excel обработку Сервис/ Анализ данных/ РЕГРЕССИЯ выбрав в качестве диапазона данных для исследуемой переменной колонку данных для y_1, а в качестве диапазона данных для факторных переменных – колонки данных для x_1 и x_2. Аналогично для определения коэффициентов второго уравнения применим обработку РЕГРЕССИЯ, взяв данные для y_1 , x_1 и x_2. В итоге получим следующую систему уравнений (ПФМ):

$$\begin{cases} y_1 = 19{,}9046 + 2{,}8214x_1 + 0{,}3937x_2 + \varepsilon_1 \\ y_2 = 19{,}0661 + 1{,}6844x_1 + 1{,}1855x_2 + \varepsilon_2 \end{cases}$$

Для перехода от приведённой формы к структурной форме модели найдём x_2 из второго уравнения:

$$x_2 = \frac{y_2 - 19{,}0661 - 1{,}6844x_1}{1{,}1855}.$$

Подставим это выражение в первое уравнение вместо x_2 , и после необходимых арифмети-

$$y_1 = 13{,}573 + 0{,}332y_2 + 2{,}262x_1 + \varepsilon_1$$

ческих преобразований, получим первое уравнение структурной формы:

Далее выразим x_1 из первого уравнения ПФМ

$$x_1 = \frac{y_1 - 19{,}9046 - 0{,}3937x_1}{2{,}8214}$$

и подставим это выражение во второе уравнение ПФМ вместо x_1. После очевидных преоб-

$$y_2 = 7{,}183 + 0{,}597y_1 + 0{,}951x_2 + \varepsilon_2$$

разований получим второе уравнение структурной формы:

Окончательный вид структурной модели:

$$\begin{cases} y_1 = 13{,}573 + 0{,}332y_2 + 2{,}262x_1 + \varepsilon_1 \\ y_2 = 7{,}183 + 0{,}597y_1 + 0{,}951x_2 + \varepsilon_2 \end{cases}$$

Тема 5. Многомерный статистический анализ

Компонентный анализ является методом определения структурной зависимости между случайными переменными. В результате его использования получается сжатое описание малого объёма, несущее почти всю информацию, содержащуюся в исходных данных. Основой компонентного анализа является построение таких линейных комбинаций исходных переменных (главных компонент), которые бы имели максимальную дисперсию и минимальную зависимость друг от друга.

Более общим методом преобразования исходных переменных по сравнению с компонентным анализом является ***факторный анализ***. Центральной проблемой, которую приходится решать при обработке экспериментальных данных, является задача её “сжатия”, выделения существенной информации, которая затемнена разного рода данными, не имеющими отношения к сути изучаемого явления. Поэтому задача уменьшения размеров исходного массива данных тесно связана с задачей выявления закономерностей изучаемого явления. Наблюдаемые параметры зачастую являются лишь косвенными характеристиками изучаемого объекта. На самом деле существуют внутренние (не наблюдаемые непосредственно) параметры или свойства, число которых мало и которые определяют значения наблюдаемых параметров. Эти внутренние параметры принято называть факторами. Задача факторного анализа – представить наблюдаемые параметры в виде линейных комбинаций факторов.

Кластерный анализ – это совокупность методов, позволяющих классифицировать многомерные наблюдения, каждое из которых описывается набором признаков (параметров). Целью кластерного анализа является образование групп схожих между собой объектов, которые принято называть кластерами (классами). Особое место кластерный анализ занимает в тех отраслях науки, которая связана с изучением массовых явлений и процессов. Необходимость развития кластерного анализа и их использования продиктована тем, что они помогают построить научно обоснованные классификации, выявить взаимосвязи между единицами наблюдаемой совокупности. Кроме того, методы кластерного анализа могут использоваться в целях сжатия информации, что является важным фактором в условиях постоянного увеличения и усложнения потоков статистических данных.

Дискриминантный анализ является разделом многомерного статистического анализа, который включает в себя методы классификации многомерных (по ряду показателей) наблюдений по принципу максимального сходства при наличии обучающих факторов (то есть используется алгоритм, автоматически учитывающий изменения в данных).

Если в кластерном анализе рассматриваются методы многомерной классификации без обучения, то в дискриминантном анализе новые кластеры не образуются, а формулируется правило, по которому на основании данных наблюдений за новым объектом производится отнесение его к одному из уже существующих классов (кластеров, обучающих подмножеств). Такое правило базируется на сравнении определённых статистических характеристик изучаемого объекта со значениями дискриминантной функции, которая строится, чаще всего, в виде линейной статистических характеристик имеющихся классов.

Предположим, что существуют две или более совокупности (группы) и что мы располагаем множеством выборочных наблюдений над ними. Основная задача дискриминантного анализа состоит в построении с помощью этих выборочных наблюдений правила, позволяющего отнести новое наблюдение к одной из совокупностей.

Дискриминантный анализ может использоваться и для прогнозирования поведения наблюдаемого объекта путем сопоставления изменения его показателей с поведением аналогичных показателей объектов обучающих подмножеств.

Например, можно по ряду показателей выделить группы развитых и развивающихся стран. При этом мы должны уже иметь некоторые группы стран, явно относящиеся к одной из этих групп, а также иметь наборы значений некоторых показателей (среднедушевой доход, продолжительность жизни, уровень образования, производительность труда и т.д.). При отнесении других стран к одному из этих классов, мы должны построить дискриминантную функцию, зависящую от статистических характеристик имеющихся наборов данных, и сравнивать значения этой функции для каждой изучаемой страны со значениями этой же функции для каждой из двух групп. Та группа, которая будет иметь более близкое значение дискриминантной функции и примет в свои ряды новую страну. Далее зная динамику изменений показателей в этой группе, мы можем делать некоторые прогнозы изменения показателей изучаемой страны. В простейшем случае одного показателя, например, среднедушевого дохода, мы можем просто вычислить среднее значение этого показателя для каждой из групп и сравнить среднедушевой доход изучаемой страны с полученными средними значениями. Если у изучаемой страны этот показатель будет ближе к доходу осреднённому для развитых стран, то мы и отнесём её к группе развитых стран.

Аналогичный подход можно применить к предприятиям, разбив их на группы: крупные, средние, мелкие. Проделав соответствующий анализ, мы можем отнести новое предприятие к одной из групп, а далее постараться сделать прогноз развития предприятия на основании сравнения с изменением показателей предприятий этой группы. Такой подход может быть достаточно продуктивным, особенно если все предприятия относятся к какой-то одной отрасли.

5. ПРОГРАММА КУРСА

1. Линейная регрессионная модель (ЛРМ). Оценка параметров модели методом наименьших квадратов (МНК).
2. Векторное, матричное представление ЛРМ. Геометрическая интерпретация.
3. Классическая ЛРМ (КЛРМ). Несмещенность и состоятельность оценок МНК.
4. Эффективность оценок МНК. Теорема Гаусса-Маркова для парной регрессии.
5. Статистические свойства оценок параметров в нормальной КЛРМ.
6. Проверка гипотез и доверительные интервалы в нормальной КЛРМ.
7. Дисперсионный анализ в КЛРМ. F-отношение Фишера и коэффициент детерминации. Связь между ними.
8. Нелинейные модели. Эластичность.
9. Прогнозирование в КЛРМ. Доверительные интервалы для прогнозных значений.
10. Множественная ЛРМ. Оценка параметров МНК.
11. Классическая множественная линейная регрессионная модель (КМЛРМ). Состоятельность, несмещенность, эффективность оценок МНК.
12. Оценка качества множественной регрессии. Проверка гипотез. Доверительные интервалы. Коэффициент детерминации и отношение Фишера.
13. Стандартизованная множественная регрессия.
14. Частные коэффициенты корреляции. Геометрическая интерпретация частной корреляции.
15. Скорректированный коэффициент детерминации.
16. Ошибки спецификации модели. «Лишние» и «пропущенные» регрессоры.
17. Мультиколлинеарность.
18. Фиктивные переменные. Моделирование «излома».
19. Обобщенный метод наименьших квадратов. Теорема Айткена.
20. Гетероскедастичность. Взвешенный метод наименьших квадратов.
21. Моделирование временных рядов. Авторегрессия первого порядка.
22. Оценка параметров в модели с авторегрессией. Процедуры Кохрейна-Оркатта, Хилдрета-Лу и Дарбина.
23. Автокорреляция остатков. Отношение Дарбина-Уотсона и его статистические свойства.
24. Инструментальные переменные. Двухшаговый МНК.
25. Системы внешне не связанных регрессионных уравнений.
26. Системы одновременных регрессионных уравнений.

6. ВОПРОСЫ ДЛЯ САМОПРОВЕРКИ

1. Парная линейная регрессия

1. Что называется линейной регрессионной моделью (ЛРМ)? Какие практические задачи могут решаться с помощью парной регрессии?
2. Как оценить параметры ЛРМ методом наименьших квадратов (МНК)?
3. Каким требованиям должна удовлетворять классическая ЛРМ? Сформулируйте условия Гаусса-Маркова.
4. Докажите несмещенность и состоятельность оценок параметров в классической ЛРМ.
5. Какая оценка параметра называется эффективной. Сформулируйте теорему Гаусса-Маркова для парной регрессии.
6. Какие статистические свойства у оценок параметров в нормальной классической ЛРМ?
7. Как проверяются гипотезы для значений параметров и строятся доверительные интервалы в нормальной классической ЛРМ?
8. Как вычисляется коэффициент детерминации и дисперсионное отношение Фишера? Как проверяется гипотеза о значимости модели?
9. Установите связь между F-отношением Фишера и коэффициентом детерминации R^2.
10. Представьте основные нелинейные модели. Опишите процесс линеаризации.
11. Как осуществляется прогнозирование в ЛРМ? Как строятся доверительные интервалы для прогноза.
12. Как определяется эластичность в линейной и нелинейных моделях.

2. Множественная линейная регрессия

1. Определите множественную ЛРМ. Какие практические задачи могут решаться с помощью множественной регрессии?
2. Как производится оценка параметров множественной ЛРМ? Как выглядит система нормальных уравнений? Опишите способ ее решения.
3. Какая множественная ЛРМ называется классической? Перечислите условия Гаусса-Маркова для множественной регрессии.
4. Как оценивается качество множественной регрессии с помощью коэффициента детерминации и отношения Фишера? Как проверить гипотезу о значении параметра модели?

5. Опишите процедуру вычисления скорректированного коэффициента детерминации. Для каких целей он используется?
6. Когда наблюдается полная, а когда частичная мультиколлинеарность? Опишите признаки частичной мультиколлинеарности и методы ее устранения.
7. Какие переменные называются фиктивными? Какие возможности появляются у исследователя, использующего фиктивные переменные?
8. Как выглядит диаграмма рассеяния наблюдений в случае гетероскедастичности остатков? Опишите процедуру «взвешенного» МНК.

4. Системы регрессионных уравнений

1. Приведите пример системы регрессионных уравнений. Какие переменные в системе называются экзогенными? Какие переменные в модели называются эндогенными?
2. Какая форма модели называется структурной?
3. Какая форма модели называется приведенной? Как преобразовать приведенную форму в структурную?
4. Сформулируйте необходимое условие идентифицируемости параметров уравнения в системе регрессионных уравнений.

5. Временные ряды. Авторегрессия

1. Приведите примеры временных рядов. Как выделятся тренд временного ряда?
2. Как вычисляется ряд разностей временного ряда? Как вычисляется автокорреляционная функция временного ряда?
3. Как по виду автокорреляционной функции определить статистические свойства временного ряда?
4. Что называется авторегрессией? Как осуществляется оценка параметров и прогнозирование в авторегрессии?
5. Как вычисляется отношение Дарбина-Ватсона? Как с помощью этого отношения определить наличие или отсутствие корреляции остатков у временного ряда?

ЗАКЛЮЧЕНИЕ

Приведенные в пособии примеры являются простейшими и рассчитаны на неискушенного слушателя, впервые столкнувшегося с методологическим аппаратом эконометрики. Малое число наблюдений в заданиях позволяет практически «вручную» (на калькуляторе) выполнить необходимые расчеты. Поставленные задания носят скорее технический характер, хотя приведенные в пособии примеры связаны с реальной экономикой. На начальном этапе изучения эконометрики эта работа необходима. Слушатель учится правильно читать листинги эконометрических отчетов, понимает, как получена та или иная характеристика модели. Эта работа делает более доступным лекционный материал. Нечто схожее наблюдается в линейном программировании, когда графически, с помощью симплекс-таблиц, методом «потенциалов», «венгерским» методом и т.д. решают определенные задачи, хотя давно уже доступны специализированные пакеты для решения задач линейного программирования.

Использование эконометрического пакета позволило бы существенно поднять планку заданий и максимально приблизить их к практике современных эконометрических исследований. Планируется вторая часть пособия, рассчитанная на решения задач повышенного уровня сложности.

ЛИТЕРАТУРА

1. Эконометрика: Учебник / Под ред. И.И. Елисеевой. – М: Финансы и статистика, 2002.
2. Практикум по эконометрике: Учебное пособие / Под ред. И.И. Елисеевой. – М: Финансы и статистика, 2002.
3. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика: Начальный курс. – М.: Дело, 2001.
4. Катышев П.К., Пересецкий А.А. Сборник задач к начальному курсу эконометрики. – М; Дело, 1999.
5. Кошмак В.К. Эконометрика: Методические указания и контрольные задания по курсу/ Кошмак В.К. – Псков: Издательство ППИ, 2007. – 33 с.
6. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики: Учебник. – М.: ЮНИТИ, 1998.
7. Доугерти К. Введение в эконометрику. – М.: ИНФРА-М, 2001.
8. Колемаев В.А. Эконометрика: Учебник. – М.: ИНФРА-М, 2004.
9. Мардас А.Н. Эконометрика. – СПб: Питер, 2001.
10. Бородич С.А. Эконометрика. – Минск: ООО «Новое знание», 2001.
11. Тихомиров Н.П., Дорохина Е.Ю. Эконометрика. М.: «Экзамен», 2003.

Приложение 1

Критические точки распределения Стьюдента

Число степеней свободы k	Уровень значимости α (двусторонний)			Число степеней свободы k	Уровень значимости α (двусторонний)		
	0,1	0,05	0,01		0,1	0,05	0,01
1	6,314	12,706	63,656	26	1,706	2,056	2,779
2	2,920	4,303	9,925	27	1,703	2,052	2,771
3	2,353	3,182	5,841	28	1,701	2,048	2,763
4	2,132	2,776	4,604	29	1,699	2,045	2,756
5	2,015	2,571	4,032	30	1,697	2,042	2,750
6	1,943	2,447	3,707	31	1,696	2,040	2,744
7	1,895	2,365	3,499	32	1,694	2,037	2,738
8	1,860	2,306	3,355	33	1,692	2,035	2,733
9	1,833	2,262	3,250	34	1,691	2,032	2,728
10	1,812	2,228	3,169	35	1,690	2,030	2,724
11	1,796	2,201	3,106	36	1,688	2,028	2,719
12	1,782	2,179	3,055	37	1,687	2,026	2,715
13	1,771	2,160	3,012	38	1,686	2,024	2,712
14	1,761	2,145	2,977	39	1,685	2,023	2,708
15	1,753	2,131	2,947	40	1,684	2,021	2,704
16	1,746	2,120	2,921	41	1,683	2,020	2,701
17	1,740	2,110	2,898	42	1,682	2,018	2,698
18	1,734	2,101	2,878	43	1,681	2,017	2,695
19	1,729	2,093	2,861	44	1,680	2,015	2,692
20	1,725	2,086	2,845	45	1,679	2,014	2,690
21	1,721	2,080	2,831	46	1,679	2,013	2,687
22	1,717	2,074	2,819	47	1,678	2,012	2,685
23	1,714	2,069	2,807	48	1,677	2,011	2,682
24	1,711	2,064	2,797	49	1,677	2,010	2,680
25	1,708	2,060	2,787	50	1,676	2,009	2,678

Приложение 2

Критические точки распределения Фишера (уровень значимости 0,05)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10
1	161,446	199,499	215,707	224,583	230,160	233,988	236,767	238,884	240,543	241,882
2	18,513	19,000	19,164	19,247	19,296	19,329	19,353	19,371	19,385	19,396
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,785
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637
8	5,318	4,459	4,066	3,838	3,688	3,581	3,500	3,438	3,388	3,347
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,896	2,854
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,796	2,753
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,714	2,671
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,646	2,602
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,588	2,544
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,538	2,494
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,494	2,450
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,456	2,412
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,423	2,378
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,393	2,348
21	4,325	3,467	3,072	2,840	2,685	2,573	2,488	2,420	2,366	2,321
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,342	2,297
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,320	2,275
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,300	2,255
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,282	2,236
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,265	2,220
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,250	2,204
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,236	2,190
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,223	2,177
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,211	2,165
31	4,160	3,305	2,911	2,679	2,523	2,409	2,323	2,255	2,199	2,153
32	4,149	3,295	2,901	2,668	2,512	2,399	2,313	2,244	2,189	2,142
33	4,139	3,285	2,892	2,659	2,503	2,389	2,303	2,235	2,179	2,133
34	4,130	3,276	2,883	2,650	2,494	2,380	2,294	2,225	2,170	2,123
35	4,121	3,267	2,874	2,641	2,485	2,372	2,285	2,217	2,161	2,114
36	4,113	3,259	2,866	2,634	2,477	2,364	2,277	2,209	2,153	2,106
37	4,105	3,252	2,859	2,626	2,470	2,356	2,270	2,201	2,145	2,098
38	4,098	3,245	2,852	2,619	2,463	2,349	2,262	2,194	2,138	2,091
39	4,091	3,238	2,845	2,612	2,456	2,342	2,255	2,187	2,131	2,084
40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,124	2,077
45	4,057	3,204	2,812	2,579	2,422	2,308	2,221	2,152	2,096	2,049
50	4,034	3,183	2,790	2,557	2,400	2,286	2,199	2,130	2,073	2,026
55	4,016	3,165	2,773	2,540	2,383	2,269	2,181	2,112	2,055	2,008
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	2,040	1,993
65	3,989	3,138	2,746	2,513	2,356	2,242	2,154	2,084	2,027	1,980
70	3,978	3,128	2,736	2,503	2,346	2,231	2,143	2,074	2,017	1,969
80	3,960	3,111	2,719	2,486	2,329	2,214	2,126	2,056	1,999	1,951
90	3,947	3,098	2,706	2,473	2,316	2,201	2,113	2,043	1,986	1,938
100	3,936	3,087	2,696	2,463	2,305	2,191	2,103	2,032	1,975	1,927
110	3,927	3,079	2,687	2,454	2,297	2,182	2,094	2,024	1,966	1,918
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,959	1,910

УДК 330
ББК 65

Сергей Владимирович Павликов
Алексей Геннадьевич Исавнин

ЭКОНОМЕТРИКА

Электронный образовательный ресурс

Подписано в печать 25.05.2014 г.
Формат 60x84 1/16   Гарнитура Times New Roman
Уч.-изд.л. 3;   Усл.-печ.л. 5,46;   Тираж 350 экз.

Заказ 238
Отпечатано с готового оригинал-макета.

Копировальный участок НЧФ КНИТУ-КАИ им. А.Н. Туполева
г. Набережные Челны, РТ.
kaicopy@outlook.com, +7(987) 292-43-54