МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Государственное образовательное учреждение высшего профессионального образования

«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР))

Кафедра автоматизации обработки информации (АОИ)

УТВЕРЖДАЮ
Заведующий кафедрой АОИ
____________________Ю.П.Ехлаков
«___» ____________________ 2015 г.

Методические указания
к лабораторным и самостоятельным работам
по дисциплине

Эконометрика

Направление подготовки: 38.03.05(**080500.62) «Бизнес-информатика»**

Форма обучения: **очная**

Факультет систем управления (ФСУ)

Кафедра автоматизации обработки информации (АОИ)

Курс 3 Семестр 6

Разработчик
Старший преподаватель кафедры АОИ И.В. Потахова

Лабораторные работы	36 часов
Самостоятельная работа	54 часа

2015

СОДЕРЖАНИЕ

Методические указания	3
Построение и анализ модели линейной парной регрессии	4
Построение и анализ модели нелинейной парной регрессии	11
Построение и анализ модели множественной линейной регрессии	17
Анализ случайных остатков в модели регрессии	19
Модели регрессии с фиктивными переменными	31
Идентификация модели	36
Оценивание параметров структурной модели	41
Изучение взаимосвязей по временным рядам	43
Методические указания по выполнению самостоятельной работы	47
Список литературы	48

Методические указания

Лабораторные работы выполняются в рамках курса «Эконометрика», предусматривающего изучение методов проверки, обоснования, оценивания количественных закономерностей и качественных утверждений на основе анализа статистических данных. Кроме этого рассматриваются возможности применения Excel для решение означенных задач. В работах предусмотрено выполнение ряда практических заданий.

Работы рекомендуется выполнять в порядке их следования.

По выполненным лабораторным работам студент отчитывается перед преподавателем. Отчет студента должен быть представлен выполненными заданиями и пояснениями по ходу их выполнения.

Лабораторная работа №1. Построение и анализ модели линейной парной регрессии.

<u>**Цель:**</u> построение и исследование уравнения линейной регрессии.

Простая (парная) регрессия представляет собой регрессию между двумя переменными — y и x вида:

$$y = f(x) + \varepsilon, \qquad (1.1)$$

где y — зависимая переменная (результативный признак);

x — независимая, или объясняющая, переменная (признак-фактор);

ε — случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии $\hat{y} = f(x) + \varepsilon$. (1.2)

Различают линейные и нелинейные регрессии.

Линейная регрессия описывается уравнением: $y = a + b \cdot x + \varepsilon$. (1.3)

На практике построение линейной регрессии сводится к оценке параметров уравнения $\hat{y}_x = a + b \cdot x$. (1.4)

Классический подход к оцениванию параметров линейной регрессии основан на методе наименьших квадратов (МНК).

<u>**Расчетные соотношения.**</u>

1. ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТОВ УРАВНЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ

Коэффициенты, определяемые на основе метода наименьших квадратов, являются решением системы уравнений:

$$\begin{cases} a + b \cdot \bar{x} = \bar{y} \\ a \cdot \bar{x} + b \cdot \overline{x^2} = \overline{x \cdot y} \end{cases} \qquad (1.5)$$

где

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i; \quad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i; \quad \overline{x \cdot y} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \cdot y_i; \quad \overline{x_i^2} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2. \qquad (1.6)$$

Коэффициенты уравнений (1.3), (1.4) получаем, решив систему (1.5).

$$b = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{\operatorname{cov}(x, y)}{\sigma_x^2}); \qquad a = \bar{y} - b \cdot \bar{x} \qquad (1.7),$$

где $\operatorname{cov}(x, y)$ — выборочное значение корреляционного момента (ковариация), определенного по формуле: $\operatorname{cov}(x, y) = \overline{x \cdot y} - \bar{x} \cdot \bar{y}$, (1.8)

σ_x^2 – выборочное значение дисперсии величины x, определяемой по формуле:

$$\sigma_x^2 = \overline{x^2} - (\bar{x})^2 \qquad (1.9)$$

2. ИССЛЕДОВАНИЕ УРАВНЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ

2.1. Оценка тесноты связи

Линейный коэффициент корреляции. При линейной регрессии в качестве показателя тесноты связи выступает линейный коэффициент корреляции. Его значение находится в границах $-1 \le r_{xy} \le 1$.

$$r_{xy} = \frac{\operatorname{cov}(x,y)}{\sigma_x \cdot \sigma_y}, \tag{1.10}$$

где $$\sigma_x^2 = \overline{x^2} - \bar{x}^2, \quad \sigma_y^2 = \overline{y^2} - \bar{y}^2, \quad \overline{y_i^2} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i^2 . \tag{1.11}$$

Коэффициент детерминации. Он характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака.

$$r^2{}_{xy} = \frac{\sigma^2_{объясн}}{\sigma_y^2} = 1 - \frac{\sigma^2_{ост}}{\sigma_y^2} \tag{1.12}$$

где $$\sigma_y^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2 = \overline{y^2} - \bar{y}^2,$$

$$\sigma^2_{объясн} = \frac{1}{n} \cdot \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2, \quad \sigma^2_{ост} = \frac{1}{n} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{1.13}$$

Средняя ошибка аппроксимации. Средняя ошибка аппроксимации – среднее относительное отклонение расчетных значений $\hat{y}$ от фактических y.

$$\bar{A} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_{x_i}}{y_i} \right| \cdot 100\% \tag{1.14}$$

Построенное уравнение регрессии считается удовлетворительным, если значение $\bar{A}$ не превышает 10–12 %.

Чем выше показатель детерминации или чем ниже средняя ошибка аппроксимации, тем лучше модель описывает исходные данные.

2.2. Оценка значимости уравнения линейной регрессии и существенности параметров линейной регрессии.

2.2.1. Вычисление оценок дисперсий парной линейной регрессии

Оценки для дисперсий определяются формулами:

$$S^2_{общ} = \frac{\sum_{i=1}^{n} (y - \bar{y})^2}{n-1}$$ — общая дисперсия результативного признака. (1.15)

$$S^2_{факт} = \frac{\sum_{i=1}^{n} (\hat{y}_x - \bar{y})^2}{m}$$ — факторная (объясненная) дисперсия результативного признака. (1.16)

$S_{ост}^2 = \frac{\sum_{i=1}^{n}(y - \hat{y}_x)^2}{n - m - 1}$ остаточная (необъясненная) дисперсия результативного признака. (1.17)

Здесь n — количество наблюдений, m = 1 для парной регрессии.

2.2.2. Оценка значимости уравнения регрессии в целом дается с помощью F-критерия *Фишера*

Выдвигается нулевая гипотеза $H_0 : b = 0$ о том, что коэффициент регрессии равен нулю и, следовательно, фактор *x* не оказывает влияния на результат *y*.

Уравнение парной регрессии значимо с уровнем значимости α, если выполняется следующее неравенство:

$$F = \frac{S_{факт}^2}{S_{ост}^2} > F_{\alpha;1;n-2} \quad (1.18)$$

где $F_{\alpha;1;n-2}$ – значения квантиля уровня α *F*-распределения с числами степеней свободы $k_1 = 1$; $k_2 = n - 2$. Для вычисления квантиля можно использовать таблицу или функцию *Excel*: $F_{\alpha;1;n-2}$ = FРАСПОБР($\alpha;1;n-2$).

2.2.3. Оценка существенности коэффициента линейной регрессии.

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

Фактическое значение *t* – критерия Стьюдента вычисляется по формуле:

$$t_b = \frac{b}{m_b}. \quad (1.19)$$

Стандартная ошибка коэффициента (b) регрессии определяется по формуле:

$$m_b = \frac{S_{ост}}{\sigma_x \cdot \sqrt{n}}, \quad (1.20)$$

где

где $S_{ост}^2 = \frac{\sum_{i=1}^{n}(y - \hat{y}_x)^2}{n - m - 1}$, — остаточная дисперсия на одну степень свободы,

$\sigma_x^2 = \overline{x^2} - \bar{x}^2$ — дисперсия признака x.

Вычисленное значение (t_b) сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы $(n-2)$. Здесь проверяется нулевая гипотеза $H_0 : b = 0$, предполагающая несущественность статистической связи между y и x, но только учитывающая значение b, а не соотношение между факторной и остаточной дисперсиями в общем балансе дисперсии результативного признака. Но общий смысл гипотез один и тот же: проверка наличия статистической связи между y и

x или её отсутствия.

Если $t_b > t_{табл}(\alpha, n-2)$, то гипотеза $H_0 : b = 0$ должна быть отклонена, а статистическая связь y и x считается установленной. В случае $t_b < t_{табл}(\alpha, n-2)$ нулевая гипотеза не может быть отклонена, и влияние y на x признается несущественным.

2.2.4. Построение интервальных оценок для параметров регрессии, функции парной линейной регрессии

- Интервальная оценка (доверительный интервал) для коэффициента b с надежностью (доверительной вероятностью) равной γ определяется выражением:

$$b \pm t_{табл} \cdot m_b \tag{1.30}$$

Аналогично строится интервальная оценка параметра a. При этом используются следующая расчетная формула вычисления стандартной ошибки коэффициента a:

$$m_a = \frac{S_{ост} \cdot \sqrt{\sum_{i=1}^{n} x_i^2}}{\sigma_x \cdot n} \tag{1.31}$$

- Интервальная оценка (доверительный интервал) для вычисленного значения $\hat{y}_i$ при заданном значении x_i с надежностью (доверительной вероятностью) равной $\gamma = 1 - \alpha$ определяется выражением

$$\hat{y}_i \pm t_{табл} \cdot m_{\hat{y}_i} \tag{1.32}$$

Стандартная ошибка вычисленного значения $\hat{y}_i$ определяется по формуле:

$$m_{\hat{y}_i} = S_{ост} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{n \cdot \sigma_x^2}} \tag{1.33}$$

Таким образом, в (1.30) входят две величины $m_{\hat{y}_i}$ (зависит от x_i) и $t(\gamma, n-2)$, вычисляемая с помощью функции Excel:

$$t(\gamma, n-2) = СТЬЮДРАСПОБР(1-\gamma, n-2).$$

3. Задания на лабораторную работу

(1 – 9 варианты). В таблице приведены данные о среднедушевом прожиточном минимуме в день на одного работающего x (в рублях) и данные о средней заработной плате за один рабочий день y (в рублях) в 15-ти регионах.

1. Постройте уравнение парной регрессии y от x.
2. Рассчитайте коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.
3. Оцените статистическую значимость параметров регрессии и уравнения регрессии с помощью F-критерия Фишера и t-критерия Стьюдента.
4. Найдите доверительные интервалы для коэффициентов регрессии и уравнения регрессии на уровне значимости $\alpha = 0{,}05$. Отразите на графике
5. Найдите и удалите из выборки две точки, наиболее удалённые от линии регрессии. Постройте линию регрессии для этой выборки. Сравните результаты.

Вариант 1		Вариант 2		Вариант 3	
Прожиточный минимум	Заработная плата	Прожиточный минимум	Заработная плата	Прожиточный минимум	Заработная плата
234	445	215	486	206	369
246	484	226	531	216	410
261	518	239	569	229	441
237	457	217	500	208	383
267	524	245	574	235	443
318	623	292	682	279	526
201	396	184	433	177	335
264	517	242	566	232	436
219	434	201	476	192	369
261	517	239	567	229	439
228	449	209	492	200	380
345	685	316	752	303	583
207	419	190	460	182	360
252	526	231	579	221	459
276	553	253	607	242	472

Вариант 4		Вариант 5		Вариант 6	
Прожиточный минимум	Заработная плата	Прожиточный минимум	Заработная плата	Прожиточный минимум	Заработная плата
223	448	312	628	234	472
233	467	342	685	261	524
246	494	367	735	282	565
225	451	322	644	243	487
252	504	371	742	283	567
296	594	443	887	338	678
194	388	277	555	211	423
249	499	365	732	279	558
209	420	306	612	234	469
246	494	366	733	281	563
217	436	316	633	242	484
320	641	489	979	377	754
199	399	295	591	228	457
238	478	374	749	294	589
259	520	393	786	303	607

Вариант 7		Вариант 8		Вариант 9	
Прожиточный минимум	Заработная плата	Прожиточный минимум	Заработная плата	Прожиточный минимум	Заработная плата
287	577	406	816	302	608
299	600	445	890	337	675
317	635	478	957	365	730
289	579	417	836	313	627
324	649	483	967	366	733
384	769	579	1160	440	881
247	495	358	717	270	541

321	642	476	953	360	721
268	537	396	793	301	603
317	635	477	955	363	727
278	558	410	822	311	622
415	832	641	1283	491	983
254	509	382	764	293	586
307	614	488	976	381	763
335	670	512	1025	393	786

(10 – 18 варианты). В таблице приведены данные о весе грузов x (в килограммах) и количестве заказов соответствующего груза y (в тысячах).

1. Постройте уравнение парной регрессии y от x.

2. Рассчитайте коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.

3. Оцените статистическую значимость параметров регрессии и уравнения регрессии с помощью F-критерия Фишера и t-критерия Стьюдента.

4. Найдите доверительные интервалы для коэффициентов регрессии и уравнения регрессии на уровне значимости $\alpha = 0{,}05$. Отразите на графике.

5. Найдите и удалите из выборки две точки, наиболее удалённые от линии регрессии. Постройте линию регрессии для этой выборки. Сравните результаты.

Вариант 10		Вариант 11		Вариант 12	
Вес	Заказы	Вес	Заказы	Вес	Заказы
540	6.1	554	6.7	547	6.7
708	9.1	719	11.5	713	11.5
593	7.2	608	8.0	600	8.0
508	7.5	526	5.6	517	5.6
648	6.9	666	2.9	657	2.9
935	11.5	955	9.8	945	9.8
855	10.3	865	9.6	860	9.6
753	9.5	767	9.7	760	9.7
913	9.2	931	6.7	922	6.7
960	10.6	971	8.8	966	8.8
1010	12.5	1022	7.6	1016	7.6
1065	12.9	1075	14.5	1070	14.5
1205	14.5	1215	10.4	1210	10.4
1080	13.6	1092	11.9	1086	11.9
1023	12.8	1035	13.9	1029	13.9
1383	16.5	1393	16.2	1388	16.2
1430	17.1	1443	18.8	1436	18.8
1265	15.0	1278	14.9	1272	14.9
1320	16.2	1336	14.3	1328	14.3
1253	15.8	1266	17.1	1259	17.1
1570	19.0	1584	18.0	1577	18.0
1693	19.4	1706	19.8	1699	19.8
1505	19.1	1524	21.4	1515	21.4

1575	18.0	1590	15.6	1582	15.6
1630	20.2	1644	15.2	1637	15.2

Вариант 13		Вариант 14		Вариант 15	
Вес	Заказы	Вес	Заказы	Вес	Заказы
554	6.1	561	12.8	1108	6.1
719	9.1	724	20.6	1437	9.1
608	7.2	616	15.2	1217	7.2
526	7.5	536	13.1	1053	7.5
666	6.9	676	9.8	1333	6.9
955	11.5	964	21.3	1909	11.5
865	10.3	870	19.9	1730	10.3
767	9.5	774	19.2	1533	9.5
931	9.2	940	15.9	1862	9.2
971	10.6	977	19.4	1943	10.6
1022	12.5	1029	20.1	2045	12.5
1075	12.9	1081	27.4	2151	12.9
1215	14.5	1220	24.9	2431	14.5
1092	13.6	1097	25.5	2183	13.6
1035	12.8	1041	26.7	2069	12.8
1393	16.5	1398	32.7	2785	16.5
1443	17.1	1449	35.9	2886	17.1
1278	15.0	1285	29.9	2557	15.0
1336	16.2	1343	30.5	2671	16.2
1266	15.8	1273	32.9	2532	15.8
1584	19.0	1591	37.0	3167	19.0
1706	19.4	1713	39.2	3412	19.4
1524	19.1	1534	40.5	3048	19.1
1590	18.0	1597	33.6	3179	18
1644	20.2	1651	35.4	3289	20.2

Лабораторная работа №2. Построение и анализ модели нелинейной парной регрессии

Цель: построение и исследование уравнения нелинейной регрессии.

Простая (парная) регрессия представляет собой регрессию между двумя переменными — y и x вида:

$$y = f(x) + \varepsilon, \qquad (1.1)$$

где y — зависимая переменная (результативный признак);

x — независимая, или объясняющая, переменная (признак-фактор);

ε — случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии $\hat{y} = f(x) + \varepsilon$.

Построение и анализ нелинейных моделей имеют свою специфику. В рамках данной лабораторной работы рассматриваются нелинейные модели, допускающие сведения их к линейному типу.

Расчетные соотношения.

1. ОЦЕНКА ПАРАМЕТРОВ НЕЛИНЕЙНОЙ РЕГРЕССИИ

1.1. Регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам.

- равносторонняя гипербола: $y = a + \frac{b}{x} + \varepsilon$. (1.2)
- полулогарифмическая: $y = a + b \cdot \ln(x) + \varepsilon$. (1.3)

При оценке параметров регрессий (1.1), (1.2) используется метод замены переменных. Суть его состоит в замене нелинейных объясняющих переменных, в результате чего нелинейные функции регрессии сводятся к линейным. К новой, преобразованной функции регрессии может быть применен обычный метод наименьших квадратов (МНК), рассмотренный в лабораторной работе №1.

Таблица 1. Возможные замены переменных

	Вид модели	Линеаризующие преобразования	Ограничения	Обратная замена переменных	
1	$\hat{y}_x = a + b \cdot \ln x$	$X = \ln x$	$x > 0$	$a = a$	$b = b$
2	$\hat{y}_x = a + b \cdot \frac{1}{x}$	$X = \frac{1}{x}$	$x \neq 0$	$a = a$	$b = b$

1.2. Регрессии, нелинейные относительно оцениваемых параметров, но линейных относительно включенных в анализ объясняющих переменных.

- логарифмическая модель (степенная): $y = a \cdot x^b + \varepsilon$ (1.4)
- показательная: $y = a \cdot b^x + \varepsilon$; (1.5)
- экспоненциальная: $y = e^{a + b \cdot x} + \varepsilon$ (1.6).

Оценка параметров регрессий (1.3), (1.4), (1,5) выполняется по следующему алгоритму:

1. уравнения приводятся к линейному виду с помощью логарифмирования и последующей замены переменных;
2. оцениваются параметры преобразованного уравнения с использованием метода наименьших квадратов;
3. выполняется обратная замена переменных и записывается исходное уравнение.

Таблица 2. Возможные замены переменных

	Вид модели	Линеаризующие преобразования	Ограничения	Обратная замена переменных	
1	$\hat{y}_x = a \cdot x^b$	$Y = \ln y,\ X = \ln x,$ $A = \ln a$	$x > 0,\ y > 0,$ $a > 0$	$a = e^A$	$b = b$
2	$\hat{y}_x = a \cdot b^x$	$Y = \ln y,\ B = \ln b,$ $A = \ln a$	$b > 0,\ y > 0,$ $a > 0$	$a = e^A$	$b = e^B$
3	$\hat{y}_x = e^{a + b \cdot x}$	$Y = \ln y$	$y > 0$	$a = a$	$b = b$

2. ИССЛЕДОВАНИЕ УРАВНЕНИЯ НЕЛИНЕЙНОЙ РЕГРЕССИИ

2.1. Оценка тесноты связи

Индекс корреляции. При нелинейной регрессии в качестве показателя тесноты связи выступает индекс корреляции. Его значение находится в границах $0 \le \rho_{xy} \le 1$.

$$\rho_{xy} = \sqrt{1 - \frac{\sigma^2_{ост}}{\sigma^2_y}}, \qquad (1.7)$$

где $\sigma^2_y = \frac{1}{n}\sum (y - \bar{y})^2, \quad \sigma^2_{ост} = \frac{1}{n}\sum (y - \hat{y}_x)^2 \qquad (1.8)$

Индекс детерминации. Он характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака.

$$\rho^2{}_{xy} = \frac{\sigma^2_{объясн}}{\sigma^2_y} = 1 - \frac{\sigma^2_{ост}}{\sigma^2_y} \qquad (1.9)$$

где $\sigma^2_y = \frac{1}{n} \cdot \sum_{i=1}^{n} (y_i - \bar{y})^2 = \overline{y^2} - \bar{y}^2,$

$$\sigma^2_{объясн} = \frac{1}{n} \cdot \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2, \quad \sigma^2_{ост} = \frac{1}{n} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \qquad (1.10)$$

Следует обратить внимание на то, что разности в соответствующих суммах $\sum (y - \bar{y})^2$, $\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ и $\sum (y - \hat{y}_x)^2$ берутся не в преобразованных, а в исходных значениях результативного признака. Иначе говоря, при вычислении этих сумм следует использовать не преобразованные (линеаризованные) зависимости, а исходные нелинейные уравнения регрессии.

2.2. Оценка качества уравнения регрессии

Средняя ошибка аппроксимации. Средняя ошибка аппроксимации – среднее относительное отклонение расчетных значений $\hat{y}$ от фактических y.

$$\bar{A} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_{x_i}}{y_i} \right| \cdot 100\% \qquad (1.11)$$

Построенное уравнение регрессии считается удовлетворительным, если значение $\bar{A}$ не превышает 8–10 %.

Чем выше показатель детерминации или чем ниже средняя ошибка аппроксимации, тем лучше модель описывает исходные данные.

Коэффициент эластичности. Коэффициент эластичности. показывает, на сколько процентов измениться в среднем результат, если фактор изменится на 1%.

Формула для расчета коэффициента эластичности имеет вид:

$$Э = f'(x) \cdot \frac{x}{y}.$$

Так как для большинства функций коэффициент эластичности не является постоянной величиной, а зависит от соответствующего значения фактора x, то обычно рассчитывается средний коэффициент эластичности:

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

Таблица3. Формулы для расчета $\overline{Э}$ средних коэффициентов эластичности.

Вид модели, y	Первая производная, y'	Средний коэффициент эластичности, $\overline{Э}$
$y = a + \frac{b}{x} + \varepsilon$	$-\frac{b}{x^2}$	$-\frac{b}{a \cdot \overline{x} + b}$
$y = a \cdot x^b \cdot \varepsilon$	$y = a \cdot b \cdot x^{b-1}$	b
$y = a \cdot b^x \cdot \varepsilon$	$a \cdot \ln b \cdot b^x$	$\overline{x} \cdot \ln b$
$y = a + b \cdot \ln x + \varepsilon$	$\frac{b}{x}$	$\frac{b}{a + b \cdot \ln \overline{x}}$
$y = e^{a+b \cdot x} + \varepsilon$	$b \cdot e^{a+b \cdot x}$	$b \cdot \overline{x}$

2.3. Оценка значимости уравнения нелинейной регрессии.

$$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} \cdot \frac{n - m - 1}{m},$$

где ρ_{xy}^2 – индекс детерминации, n – число наблюдений, m– число параметров при переменной x. Фактическое значение F-критерия сравнивается с табличным при уровне значимости α и числе степеней свободы $k_2 = n - m - 1$ (для остаточной суммы квадратов) и $k_1 = m$ (для факторной суммы квадратов). Вычисленное значение F-критерия признается достоверным, если оно больше табличного при заданном уровне значимости α. В этом случае делается вывод о существенности уравнения регрессии в целом.

3. ЗАДАНИЕ НА ЛАБОРАТОРНУЮ РАБОТУ

1. На основании таблиц 3,4 построить предложенные уравнения регрессий.
2. Вычислить индексы парной корреляции и индексы детерминации для каждого уравнения.
3. Проверить значимость уравнения регрессии в целом для каждой кривой выравнивания.

4. Определить лучшее уравнение на основе средней ошибки аппроксимации.
5. Определить лучшее уравнение на основе совместного анализа значений индекса детерминации и средней ошибки аппроксимации.
6. Вычислить средний коэффициент эластичности.
7. С помощью встроенной функции Excel построить линии тренда указанных функций, выведя соответствующее уравнение тренда и значение R^2. Сравнить с расчетными данными.
8. Сделать выводы по проделанной работе.

Таблица 3. Варианты заданий

Вариант	Графы табл. 4	Виды кривых аппроксимации				
		Гиперболи-ческая	Полулогариф-мическая	Степенная	Показательная	Экспоненци-альная
1	**1, 2**	*		*		*
2	**2, 3**	*	*		*	
3	**3, 4**		*	*		*
4	**5, 6**	*		*	*	
5	**7, 8**		*		*	*
6	**9, 10**	*	*	*		
7	**1, 3**			*	*	*
8	**1, 4**		*	*	*	
9	**1, 5**	*			*	*
10	**1, 6**	*	*	*		
11	**2, 4**	*	*			*
12	**2, 6**	*			*	*
13	**2, 7**		*		*	*
14	**3, 6**	*		*		*
15	**3, 8**		*	*	*	

Таблица 4. Наличие предметов длительного пользования в домашних хозяйствах (шт)

Области и республики	Телевизоры	Видеомагнитофоны	Магнитофоны	Музыкальные центры	Персональные компьютеры	Холодильники	Стиральные машины	Пылесосы	Швейные машины	Легковые автомобили
	1	2	3	4	5	6	7	8	9	10
Белгородская область	113	39	65	10	1	103	93	77	58	26
Брянская область	124	37	47	5	2	99	72	64	50	18
Владимирская область	124	36	61	10	2	105	90	77	68	24

Воронежская область	122	36	44	13	2	102	96	66	60	25
Ивановская область	128	26	47	9	1	106	92	71	71	9
Калужская область	140	43	61	14	6	106	88	81	66	28
Костромская область	117	31	45	12	1	100	85	58	58	14
Курская область	113	40	47	15	3	100	78	66	51	28
Липецкая область	122	48	58	13	2	113	95	73	66	34
Московская область	139	64	57	27	14	106	87	81	62	22
Орловская область	126	39	69	8	6	11	93	73	72	27
Рязанская область	120	34	46	8	3	106	80	65	51	21
Смоленская область	125	39	55	24	6	115	93	66	49	21
Тамбовская область	118	37	59	8	1	108	99	74	65	23
Тверская область	122	35	52	8	4	102	87	64	65	12
Тульская область	133	54	58	15	5	102	93	79	36	16
Ярославская область	136	36	47	12	4	110	88	71	69	14
Республика Карелия	146	49	65	16	9	106	87	68	55	32
Республика Коми	148	58	59	23	5	11	92	78	69	34
Архангельская область	136	35	58	16	8	103	95	74	71	15
Вологодская область	138	34	51	10	3	104	95	64	60	19
Калининградская область	124	48	53	12	7	105	85	74	38	29
Ленинградская область	123	30	65	8	3	102	84	71	52	10
Мурманская область	149	59	63	29	8	107	92	87	74	21
Новгородская область	130	26	63	9	4	96	76	56	45	14
Псковская область	117	26	44	91	3	99	82	65	60	20

Краснодарский край	114	44	60	14	4	109	90	74	67	28
Ставропольский край	114	40	63	12	2	104	91	78	49	29
Астраханская область	126	54	55	11	5	116	87	76	66	29
Волгоградская область	109	41	55	8	1	106	93	74	70	23
Ростовская область	120	43	57	20	8	109	91	73	60	26
Республика Башкортостан	115	40	52	16	4	116	94	75	67	29
Республика Марий Эл	134	28	64	8	1	108	87	72	73	22
Республика Мордовия	130	33	52	10	1	109	89	77	58	18
Республика Татарстан	120	52	61	19	5	119	90	76	55	22
Удмуртская республика	123	32	53	7	4	111	97	69	74	26
Чувашская республика	128	31	60	11	0	105	85	76	78	19
Кировская область	144	27	60	7	2	120	109	74	86	22
Нижегородская область	125	36	52	6	3	114	101	81	75	21
Оренбургская область	124	47	57	17	6	119	105	82	66	36
Пензенская область	121	36	47	7	2	109	94	70	69	15
Пермская область	123	40	50	15	6	113	98	73	75	27
Самарская область	128	62	56	24	14	121	100	76	68	35
Саратовская область	118	38	51	12	4	124	87	65	62	27
Ульяновская область	116	37	51	9	4	109	96	77	67	25

Лабораторная работа №3. Построение и анализ модели множественной линейной регрессии

<u>Цель:</u> построение и исследование уравнения множественной линейной регрессии, используя режим ***Регрессия программы Excel***.

Множественная регрессия представляет собой модель вида: $y = f(x_1, x_2, ..., x_m) + \varepsilon$,

где y — зависимая переменная (результат);

$x_1, x_2, ..., x_m$ — независимые переменные (факторы);

ε — случайная ошибка регрессионной зависимости;

f — некоторая математическая функция.

Простая (парная) регрессия представляет собой регрессию между двумя переменными — y и x вида:

Линейная модель множественной регрессии — зависимость вида:

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_m \cdot x_m + \varepsilon,$$

где $a, b_1, b_2, ... + b_m$ — параметры функции.

Параметр a называется свободным членом и определяет значение результирующей переменной y в случае, когда все объясняющие переменные $x_1, x_2, ..., x_m$ равны нулю. Если же факторы по своему экономическому содержанию не могут принимать нулевых значений, то значение параметра a может не иметь экономического смысла.

Параметры b_j называются **коэффициентами «чистой» регрессии**. Они характеризуют среднее изменение результата y с изменением соответствующего фактора x_j на единицу при неизменном значении других факторов, закрепленном на среднем уровне.

Основной целью множественной регрессии является построение модели с большим числом факторов и определение влияния каждого фактора в отдельности, а также их совместного воздействия на моделируемый показатель (результат).

<u>1. Расчетные соотношения.</u>

1. Уравнение множественной линейной регрессии в стандартизованных переменных:

$$t_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2} + ... + \beta_m \cdot t_{x_m} + \varepsilon,$$

где $\beta_i = b_i \cdot \dfrac{\sigma_{x_i}}{\sigma_y}$ — стандартизованные коэффициенты

2. Средние коэффициенты эластичности $\overline{Э}_i = b_i \cdot \dfrac{\overline{x}_i}{\overline{y}_{x_i}}$, $(i = 1,2,\ldots,m)$

3. Частные коэффициенты корреляции.

Если факторы x_i, x_j, находятся в корреляционной связи, то с помощью частных коэффициентов корреляции $r_{yx_i \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_m}$ можно оценить тесноту связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии.

Коэффициенты частной корреляции можно рассчитать по формуле:

$$r_{yx_i \cdot x_1 x_2 \ldots x_p} = \frac{r_{yx_i \cdot x_1 x_2 \ldots x_{p-1}} - r_{yx_p \cdot x_1 x_2 \ldots x_{p-1}} \cdot r_{x_i x_p \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_{p-1}}}{\sqrt{(1 - r^2_{yx_p \cdot x_1 x_2 \ldots x_{p-1}}) \cdot (1 - r^2_{x_i x_p \cdot x_1 x_2 \ldots x_{p-1}})}}.$$

Коэффициенты частной корреляции для двух факторов:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r^2_{yx_2}) \cdot (1 - r^2_{x_1 x_2})}} \quad r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1 - r^2_{yx_1}) \cdot (1 - r^2_{x_1 x_2})}},$$

где $r_{x_1 x_2} = \dfrac{\operatorname{cov}(x_1, x_2)}{\sigma_{x_1} \cdot \sigma_{x_2}}$

4. Критерий Фишера.

$$F = \frac{S^2_{факт}}{S^2_{ост}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m},$$

где $S^2_{факт}$ — факторная сумма квадратов на одну степень свободы; $S^2_{ост}$ – остаточная сумма квадратов на одну степень свободы; R^2 — коэффициент множественной детерминации; m — число параметров при переменных x (в линейной регрессии совпадает с числом включенных в модель факторов); n — число наблюдений.

Частные критерии Фишера могут быть вычислены по следующей формуле:

$$F_{x_i} = \frac{R^2_{yx_1 x_2 \ldots x_m} - R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_m}}{1 - R^2_{yx_1 x_2 \ldots x_m}} \cdot \frac{n - m - 1}{1}$$

Для модели с двумя факторами частные F-критерии вычисляются по формулам:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot (n-3); \qquad F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot (n-3).$$

2 Режим Регрессия модуля Анализ данных.

Табличный процессор Excel содержит модуль ***Анализ данных***. Этот модуль позволяет выполнить статистический анализ выборочных данных (построение гистограмм, вычисление числовых характеристик и т.д.). Режим работы Регрессия этого модуля осуществляет вычисление коэффициентов линейной множественной регрессии с k переменными, построение доверительные интервалы и проверку значимости уравнения регрессии.

Для вызова режима *Регрессия* модуля ***Анализ данных*** необходимо:

Для вызова режима *Регрессия* модуля ***Анализ данных*** необходимо:

- обратиться к пункту меню **Сервис (Excel 2000)**; **Данные (Excel 2007)**
- в появившемся меню выбрать команду ***Анализ данных;***
- в списке режимов работы модуля ***Анализ данных*** выбрать режим *Регрессия* и щелкнуть на кнопке ***Ok***.

После вызова режима *Регрессия* на экране появляется диалоговое окно (см. рис. 3.1), в котором задаются следующие параметры:

1. *Входной интервал Y* – вводится диапазон адресов ячеек, содержащих значения y_i (ячейки должны составлять один столбец).
2. *Входной интервал X* – вводится диапазон адресов ячеек, содержащих значения независимых переменных. Значения каждой переменной представляются одним столбцом. Количество переменных не более 16 (т.е. $k \le 16$).

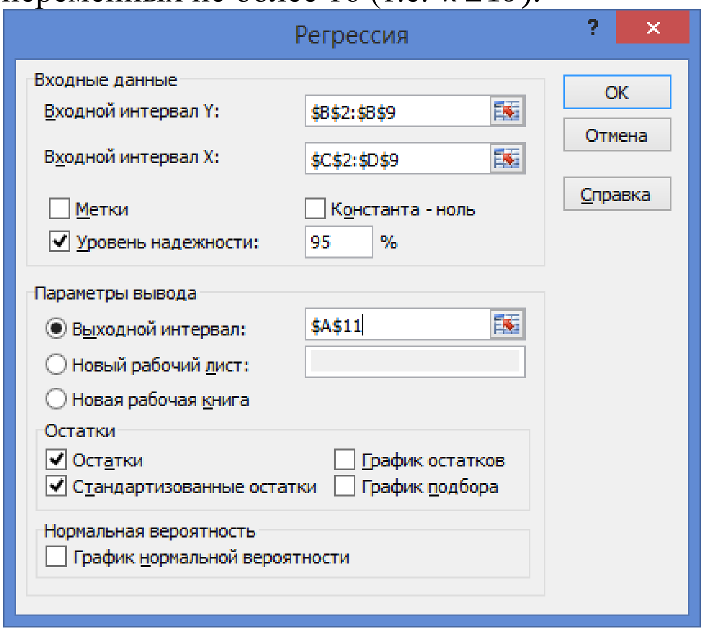

Рис. 3.1. Диалоговое окно режима *Регрессия*

3. Метки – включается, если первая строка во входном диапазоне содержит заголовок. В этом случае автоматически будут созданы стандартные названия.

4. Уровень надежности – при включении этого параметра задается надежность γ при построении доверительных интервалов.

5. Константа-ноль – при включении этого параметра коэффициент $b_0 = 0$.

6. Выходной интервал – при включении активизируется поле, в которое необходимо ввести адрес левой верхней ячейки выходного диапазона, который будет содержать ячейки с результатами вычислений режима *Регрессия.*

7. Новый рабочий лист – при включении этого параметра открывается новый лист, в который начиная с ячейки $A1$ вставляются результаты работы режима *Регрессия.*

8. Новая рабочая книга - при включении этого параметра открывается новая книга на первом листе которой начиная с ячейки $A1$ вставляются результаты работы режима *Регрессия.*

9. Остатки – при включении вычисляется столбец, содержащий невязки $y_i - \hat{y}_i, i = 1,...,n$.

10. Стандартизованные остатки – при включении вычисляется столбец, содержащий стандартизованные остатки.

11. График остатков – при включении выводятся точечные графики невязки $y_i - \hat{y}_i, i = 1,...,n$, в зависимости от значений переменных $x_j, j = 1,...,k$. Количество графиков равно числу k переменных x_j.

12. График подбора – при включении выводятся точечные графики предсказанных по построенной регрессии значений $\hat{y}_i$ от значений переменных $x_j, j = 1,...,k$. Количество графиков равно числу k переменных x_j.

3. Пример решения типовой задачи.

По данным таблицы (см. рис. 3.2) построить и оценить модель множественной линейной регрессии.

	A	B	C	D
1		Внешнеторговый оборот, в %(y)	Экспорт, в % (x_1)	Импорт, в % (x_2)
2	1	17,4	12,6	23,1
3	2	21,3	16,8	26,5
4	3	26,4	20,4	32,7
5	4	39,1	31,5	47,7
6	5	47,3	40,7	54,5
7	6	47	41	54,6
8	7	48,7	44	54,7
9	8	48,4	45,5	51,8

Рис. 3.2 Исходные данные для построения модели

Первоначально заполним таблицу, как показано на рисунке 3.2.

После этого вызовем режим *Регрессия* и в диалоговом окне зададим необходимые параметры (см. рис 3.1). Результаты работы приводятся на рис. 3.3 – 3.5.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множествен-ный R	0,99990
R-квадрат	0,99979
Нормирован-ный R-квадрат	0,99971
Стандартная ошибка	0,22622
Наблюдения	8

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значи-мость F*
Регрессия	2	1220,084	610,042	11920,166	6,37E-10
Остаток	5	0,256	0,051		
Итого	7	1220,340			

Рис. 3.3. Результаты работы режима *Регрессия*

Дадим краткую интерпретацию показателям, значения которых вычисляются в режиме *Регрессия.* Первоначально рассмотрим показатели, объединенные названием *Регрессионная статистика* (см. рис. 3.3).

Множественный R - корень квадратный из коэффициента детерминации.

$R-квадрат$ – коэффициент детерминации R^2.

Нормированный $R-квадрат$ – приведенный коэффициент детерминации $\hat{R}^2$.

Стандартная ошибка – оценка s для среднеквадратического отклонения σ.

Наблюдения – число наблюдений n.

Перейдем к показателям, объединенным названием *Дисперсионный анализ* (см. рис. 3.3).

Столбец df — число степеней свободы. Для строки *Регрессия* показатель равен количеству коэффициентов регрессии $k_r = m$; для строки *Остаток* соответствующий показатель $k_e = n - m - 1$; для строки *Итого* число степеней свободы равно $n - 1$.

Столбец SS – сумма квадратов отклонений. Для строки *Регрессия* показатель равен величине факторной суммы квадратов

$$SS_r = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2;$$

для строки *Остаток* - равен величине остаточной суммы квадратов

$$SS_e = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2;$$

для строки *Итого* – $SS = SS_r + SS_e$— общая сумма квадратов отклонений переменной y от среднего значения $\bar{y}$.

Столбец MS – дисперсии, вычисленные по формуле

$$MS = \frac{SS}{df},$$

т.е. дисперсия на одну степень свободы.

Столбец F – значение F_c, равное F – критерию Фишера, вычисленного по формуле:

$$F_c = \frac{SS_r / k_r}{SS_e / k_e}.$$

Столбец значимость F - значение уровня значимости, соответствующее вычисленной величине F – критерия и равное вероятности $P(F(k_r, k_e) \geq F_c)$, где $F(k_r, k_e)$ - случайная величина, подчиняющаяся распределению Фишера с k_r, k_e степенями свободы. Эту вероятность можно также определить с помощью функции FРАСП($F_c; k_r; k_e$). *Если вероятность меньше уровня значимости* α *(обычно* $\alpha = 0.05$*), то построенная регрессия является значимой..*

Перейдем к следующей группе показателей, объединенных в таблице, показанной на рис. 3.4.

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	0,0092	0,3983	0,0232	0,9824	-1,0145	1,0330
Переменная X 1	0,5179	0,0289	17,9504	0,0000	0,4437	0,5921
Переменная X 2	0,4767	0,0282	16,8818	0,0000	0,4041	0,5493

Рис. 3.4. Продолжение результатов работы режима *Регрессия*

Столбец Коэффициенты – вычисленные значения коэффициентов a, b_1, b_2, расположенных сверху-вниз.

Столбец Стандартная ошибка – значения m_{b_i}, $(i = 0,1,2,\ldots,m)$, вычисленные по формуле

$$m_{b_i} = \sqrt{S_{ост}^2 \cdot \left[(X' \cdot X)^{-1}\right]_{ii}} \qquad (i = 0,1,2,\ldots,m),$$

где $\left[(X' \cdot X)^{-1}\right]_{ii}$ — элемент (ii) матрицы $(X' \cdot X)^{-1}$. Значение $i = 0$ соответствует номеру элемента матрицы $(X' \cdot X)^{-1}$ для вычисления стандартной ошибки параметра a.

$S_{ост}^2 = \dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_{x_i})^2}{n-m-1}$ — несмещенная оценка остаточной дисперсии (столбец *MS*, рис 3.3).

Столбец $t-$*статистика* – значения статистик T_{b_j}.

Столбец $P-$ *значение* – содержит вероятности случайных событий $P(t(n-m) \geq T_{b_j})$, где $t(n-m)-$случайная величина, подчиняющаяся распределению Стьюдента с $n-m$ степенями свободы.

Если эта вероятность меньше уровня значимости α, *то принимается гипотеза о значимости соответствующего коэффициента регрессии.*

Столбцы Нижние 95% и Верхние 95% - соответственно нижние и верхние интервалы для оцениваемых коэффициентов a, b_1, b_2.

Перейдем к следующей группе показателей, объединенных в таблице, показанной на рис. 3.5.

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное Y*	*Остатки*	*Стандартные остатки*
1	17,547	-0,147	-0,770
2	21,343	-0,043	-0,226
3	26,163	0,237	1,238
4	39,063	0,037	0,194
5	47,069	0,231	1,207
6	47,272	-0,272	-1,424
7	48,874	-0,174	-0,909
8	48,268	0,132	0,690

Рис. 3.5. Продолжение результатов работы режима *Регрессия*

Столбец Наблюдение – содержит номера наблюдений.

Столбец Предсказанное Y – значения $\hat{y}_i$, вычисленные по построенному уравнению регрессии.

Столбец Остатки – значения невязок $y_i - \hat{y}_i$

4. Индивидуальное задание

По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%) (смотри таблицу своего варианта).

Требуется:

1. Построить линейную модель множественной регрессии. Выполнить анализ результатов.

2. Записать стандартизованное уравнение множественной регрессии. На основе стандартизованных коэффициентов регрессии и средних коэффициентов эластичности ранжировать факторы по степени их влияния на результат.

3. С помощью F – критерия Фишера оценить статистическую надежность уравнения регрессии и коэффициента детерминации $R^2_{yx_1x_2}$.

4. Найти коэффициенты парной, частной и множественной корреляции. Проанализировать их.

5. С помощью частных F– критериев Фишера оценить целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1.

6. Составить уравнение линейной парной регрессии, оставив лишь один значащий фактор.

Вариант 1

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,6	9	11	9	6,3	21
2	6	3,6	12	12	11	6,4	22
3	6	3,9	14	13	11	7	24
4	7	4,1	17	14	12	7,5	25
5	7	3,9	18	15	12	7,9	28
6	7	4,5	19	16	13	8,2	30
7	8	5,3	19	17	13	8	30
8	8	5,3	19	18	13	8,6	31
9	9	5,6	20	19	14	9,5	33
10	10	6,8	21	20	14	9	36

Вариант 2

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	6	3,5	10	11	10	6,3	21
2	6	3,6	12	12	11	6,4	22
3	7	3,9	15	13	11	7	23
4	7	4,1	17	14	12	7,5	25
5	7	4,2	18	15	12	7,9	28
6	8	4,5	19	16	13	8,2	30
7	8	5,3	19	17	13	8,4	31
8	9	5,3	20	18	14	8,6	31
9	9	5,6	20	19	14	9,5	35
10	10	6	21	20	15	10	36

Вариант 3

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,7	9	11	11	6,3	22
2	7	3,7	11	12	11	6,4	22
3	7	3,9	11	13	11	7,2	23
4	7	4,1	15	14	12	7,5	25
5	8	4,2	17	15	12	7,9	27

6	8	4,9	19	16	13	8,1	30
7	8	5,3	19	17	13	8,4	31
8	9	5,1	20	18	13	8,6	32
9	10	5,6	20	19	14	9,5	35
10	10	6,1	21	20	15	9,5	36

Вариант 4

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,5	9	11	10	6,3	22
2	7	3,6	10	12	10	6,5	22
3	7	3,9	12	13	11	7,2	24
4	7	4,1	17	14	12	7,5	25
5	8	4,2	18	15	12	7,9	27
6	8	4,5	19	16	13	8,2	30
7	9	5,3	19	17	13	8,4	31
8	9	5,5	20	18	14	8,6	33
9	10	5,6	21	19	14	9,5	35
10	10	6,1	21	20	15	9,6	36

Вариант 5

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,6	9	11	10	6,3	21
2	7	3,6	11	12	11	6,9	23
3	7	3,7	12	13	11	7,2	24
4	8	4,1	16	14	12	7,8	25
5	8	4,3	19	15	13	8,1	27
6	8	4,5	19	16	13	8,2	29
7	9	5,4	20	17	13	8,4	31
8	9	5,5	20	18	14	8,8	33
9	10	5,8	21	19	14	9,5	35
10	10	6,1	21	20	14	9,7	34

Вариант 6

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,5	9	11	10	6,3	21
2	7	3,6	10	12	10	6,8	22
3	7	3,8	14	13	11	7,2	24
4	7	4,2	15	14	12	7,9	25
5	8	4,3	18	15	12	8,1	26
6	8	4,7	19	16	13	8,3	29
7	9	5,4	19	17	13	8,4	31
8	9	5,6	20	18	13	8,8	32
9	10	5,9	20	19	14	9,6	35
10	10	6,1	21	20	14	9,7	36

Вариант 7

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,8	11	11	10	6,8	21
2	7	3,8	12	12	11	7,4	23
3	7	3,9	16	13	11	7,8	24
4	7	4,1	17	14	12	7,5	26
5	7	4,6	18	15	12	7,9	28
6	8	4,5	18	16	12	8,1	30
7	8	5,3	19	17	13	8,4	31
8	9	5,5	20	18	13	8,7	32
9	9	6,1	20	19	13	9,5	33
10	10	6,8	21	20	14	9,7	35

Вариант 8

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,8	9	11	11	7,1	22
2	7	4,1	14	12	11	7,5	23
3	7	4,3	16	13	12	7,8	25
4	7	4,1	17	14	12	7,6	27
5	8	4,6	17	15	12	7,9	29
6	8	4,7	18	16	13	8,1	30

7	9	5,3	20	17	13	8,5	32
8	9	5,5	20	18	14	8,7	32
9	11	6,9	21	19	14	9,6	33
10	10	6,8	21	20	15	9,8	36

Вариант 9

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,9	12	11	11	7,1	22
2	7	4,2	13	12	12	7,5	25
3	7	4,3	15	13	13	7,8	26
4	7	4,4	17	14	12	7,9	27
5	8	4,6	18	15	13	8,1	30
6	8	4,8	19	16	13	8,4	31
7	9	5,3	19	17	13	8,6	32
8	9	5,7	20	18	14	8,8	32
9	10	6,9	21	19	14	9,6	34
10	10	6,8	21	20	14	9,9	36

Вариант 10

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,6	12	11	10	7,2	23
2	7	4,1	14	12	11	7,6	25
3	7	4,3	16	13	12	7,8	26
4	7	4,4	17	14	11	7,9	28
5	7	4,5	18	15	12	8,2	30
6	8	4,8	19	16	12	8,4	31
7	8	5,3	20	17	12	8,6	32
8	8	5,6	20	18	13	8,8	32
9	9	6,7	21	19	13	9,2	33
10	10	6,9	22	20	14	9,6	34

Лабораторная работа №4. Анализ случайных остатков в модели регрессии

Цель: научиться оценивать наличие эффекта гетероскедастичности.

Основные формулы и понятия:

Тест Парка

$$\ln e_i^2 = a + b \cdot \ln x_{ij} + v_i,$$

где x_{ij} *– i – е значение о – го фактора*

v_i *– случайный остаток*

Условие принятия гипотезы: $t_b > t_{\alpha,n-2}$

Если данное условие выполняется, то нулевая гипотеза о наличии гетероскедастичности будет принята при уровне значимости α.

Тест ранговой корреляции Спирмена

$$r_{x,e} = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot \left(n^2 - 1\right)}$$ — коэффициент ранговой корреляции Спирмена,

где x — одна из объясняющих переменных,

d_i — разность между рангом *i-го* наблюдения *x* и рангом модуля остатка *в i-м* наблюдении.

$$t_r = \frac{r_{x,e} \cdot \sqrt{n-1}}{\sqrt{1 - r_{x,e}^2}}$$ — статистика.

Если в модели регрессии имеется более одной объясняющей переменной, то проверка гипотезы может выполняться с использованием каждой из них.

Условие принятия гипотез: $t_r > t_{\alpha,n-2}$.

Если данное условие выполняется, то нулевая гипотеза об отсутствии гетероскедастичности отклоняется при уровне значимости α.

Тест Голдфельда — Кванта

В этом случае все наблюдения необходимо упорядочить по мере возрастания значений х. Затем построить регрессионную модель для первых ***k*** и последних ***k*** наблюдений. Соответственно обозначим через $SS_{ост}^{(1)}$ и $SS_{ост}^{(3)}$ необъясненную сумму квадратов отклонений в каждой регрессии. Тогда статистика имеет вид

$$F = \frac{SS_{ост}^{(3)}}{SS_{ост}^{(1)}}.$$

Если выполняется условие $F > F_\gamma(k-m-1, k-m-1)$, *то гипотеза об отсутствии гетероскедастичности отвергается.*

Для проведения теста ранговой корреляции Спирмена необходимо выполнить следующие действия:

1. Отсортировать данные в таблице по возрастанию значений *x*;
2. Придать каждому наблюдению ранг, для чего необходимо добавить новый столбец, в котором задать числа от 1 до *n*;
3. Вызвать из пакета анализа надстройку ***Регрессия***, указав в диалоговом окне опцию ***Остатки***. После выполнения данной надстройки появится дополнительная таблица, в которой содержатся номера наблюдений, прогнозы и остатки. Тот столбец таблицы, в котором находятся остатки, необходимо перенести к исходным данным. После выполнения этих действий наша таблица будет содержать четыре столбца: ранг наблюдения, упорядоченные значения регрессора *x*, значения *y* и значения остатков;
4. Отсортировать данные по возрастанию модулей остатков и добавить новый столбец рангов остатков, аналогичным образом задав значения от 1 до *n*;
5. В дополнительном столбце вычислить значения разности между двумя полученными рангами (это и будет значение d_i);
6. На основании формул подсчитать коэффициент ранговой корреляции и статистику;
7. Проверить гипотезу.

Вид таблицы для проведения теста ранговой корреляции Спирмена

Ранг по х	Ценах1(р.)	Спрос у (тыс. шт.)	Остатки	Ранг по остаткам	Разность рангов D_i	D_i* D_i
8	15,91р.	117,088	-0,34387	1	7	49
5	15,54р.	119,864	-0,39014	2	3	9
15	16,76р.	110,023	-0,84306	3	12	144
2	15,21р.	123,809	1,019821	4	-2	4
3	15,28р.	121,175	-1,11646	5	-2	4
9	15,92р.	116,17	-1,12322	6	3	9
10	15,95р.	118,344	1,257187	7	3	9
14	16,69р.	110,106	-1,31194	8	6	36
1	15,09р.	125,178	1,426776	9	-8	64
6	15,62р.	118,068	-1,5813	10	-4	16
11	16,31р.	116,201	1,847847	11	0	0
12	16,33р.	111,457	-2,67328	12	0	0
13	16,60р.	115,103	3,003645	13	0	0
4	15,49р.	116,914	-3,7319	14	-10	100
7	15,70р.	123,589	4,559903	15	-8	64
					Сумма	508

Следовательно, значение ранговой корреляции Спирмена будет равно

$$r_{x,e} = 1 - \frac{6 \cdot \sum_{i=1}^{n} D_i^2}{n \cdot \left(n^2 - 1\right)} = 1 - \frac{6 \cdot 508}{15 \cdot (225 - 1)} = 0.0928$$

А значение статистики будет $t = 0.0928 \cdot \sqrt{15 - 1} = 0.028$

Выбрав уровень значимости 5 %, получаем критическую точку $t_{0.05,13} = 2.16$. Данное значение получено формулой СТЬЮДРАСПОБР(0,05;13).

Поскольку условие $t < t_{\alpha,n-2}$ не выполняется, то гипотеза о наличии гетероскедастичности будет принята.

Для проверки подобной гипотезы на основании теста Гольдфельда — Кванта необходимо подобным образом отсортировать наблюдения по возрастанию значения *x*, а затем отдельно оценить каждую регрессионную модель для первой трети и для последней трети наблюдений. Просчитать соответствующую статистику и проверить гипотезу об отсутствии гетероскедастичности.

Задание для самостоятельной работы

Провести исследование табличных данных практической работы «множественная регрессия» на наличие гетероскедастичности, между значением ***y*** и каждым регрессором отдельно

a) Тестом Парка
b) Тестом ранговой корреляции Спирмена;
c) Тестом Гольдфельда — Кванта.

Сделать выводы.

Лабораторная работа №5. Модели регрессии с фиктивными переменными.

Цель: научиться использовать в модели фиктивные переменные сдвига и наклона, а также различные категории.

Основные формулы и понятия:

Фиктивная переменная необходима для описания качественного изменения и может принимать два значения 0 и 1.

$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot D + u$ — *модель с фиктивной переменной сдвига*;

$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot D \cdot x + u$ — *модель с фиктивной переменной наклона*;

$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot D \cdot x + \beta_3 \cdot D + u$ — *модель с фиктивной переменной наклона и сдвига*.

Категория — событие, про которое для каждого наблюдения можно определенно сказать, произошло оно в этом наблюдении или нет.

Набор категорий — конечный набор взаимоисключающих событий, полностью исчерпывающий все возможности.

Для описания категорий необходимо ввести совокупность фиктивных переменных.

Электронная таблица Excel

До сих пор нами рассматривался только случай количественных регрессоров, поскольку значение цен и спроса являются числами. Однако может возникнуть ситуация, когда необходимо учесть некоторую специфическую информацию. Рассматривая модель спроса, можно предположить, что продаются два одинаковых продукта по одной цене, но имеющие некоторые различия. Например, наряду с уже давно продающимся чистящим порошком, поступает в продажу такой же порошок, но с новым ароматом. И имеется задача исследовать, насколько большим или меньшим спросом пользуется новая продукция. Конечно, можно построить две различные модели, и посмотреть разницу между ними, однако нас будет интересовать общая модель. В этом случае в модель необходимо вносить качественный регрессор, для чего нужно использовать фиктивную переменную. Данная переменная может принимать только два значение 0 или 1, в зависимости от отсутствия или наличия нового качества. В этом случае можно строить модель с фиктивной переменной наклона и сдвига. Работа с фиктивными переменными ни чем не отличается от построения регрессионной модели.

Поэтому рассмотрим задачу. Значение цены x и спроса y на два различных товара, которые мы условно назовем «обычный» и «новый», представлены в таблице 17.

Таблица 1

Номер наблюдения	Вид	Цена x^1(р.)	Спрос y (тыс. шт.)
1	новый	15,09р.	125,1779
2	новый	15,21р.	123,8094
3	старый	15,28р.	121,175
4	старый	15,49р.	116,9143
5	старый	15,54р.	119,8643
6	старый	15,62р.	118,0681
7	новый	15,70р.	123,5887
8	новый	15,91р.	117,0877
9	старый	15,92р.	116,1699
10	новый	15,95р.	118,3436
11	новый	16,31р.	116,2008
12	старый	16,33р.	111,4565
13	новый	16,60р.	115,1026
14	старый	16,69р.	110,1056
15	старый	16,76р.	110,0231

В электронной таблице Excel имеются возможности для быстрого задания значений фиктивной переменой. Для этого необходимо вставить столбец между колонками с названиями *Вид* и *Цена*. Озаглавим этот столбец как *Фиктивная переменная*, и для определения значений будем использовать логическую функцию ЕСЛИ. Данная функция имеет три аргумента. Первый — это логическое выражение, которое может принимать истинное или ложное значение. Вторым аргументом идет то значение, которое появляется в ячейке при истинности условия, а соответственно в третьем аргументе — значение, которое появляется в противном случае.

Выполнив данные действия, получим первые две строки таблицы 18.

Таблица 2

Номер наблюдения	Вид	Фиктивная переменная	Цена x^1 (р.)	Спрос y (тыс. шт.)
1	новый	=ЕСЛИ(B2="новый";1;0)	15,09р.	125,1779

В столбце фиктивной переменной появится значение 1, если в предыдущем столбце находилось слово «новый», и 0 в противоположном случае. После этого необходимо значение функции, находящейся в столбце ***С,*** скопировать во все нижние ячейки, а поскольку адресация относительная, то адрес будет меняться. Необходимо отметить, что логическая функция может иметь и другой вид:

ЕСЛИ(B2 = "обычный";0;1).

Теперь наша задача заключается в определении степени влияния фиктивной переменной. А именно, влияет ли это значение на свободный член (в этом случае при изменении качества можно говорить о том, что спрос изменится на какое-то количество) или на наклон линии регрессии (спрос изменится во сколько-то), или на оба эти значения сразу.

Вначале оценим регрессию, при условии, что фиктивная переменная влияет только на значение свободного члена. В этом случае итоговая таблица после выполнения надстройки **Регрессии**, при условии, что *Входной интервал Y* задан в виде ***E1:E16***, а *Входной интервал X* в виде ***C1:D16,*** имеет вид, изображенный в таблице 19.

Таблица 3

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,963696
R-квадрат	0,928711

Нормированный R-квадрат	0,916830
Стандартная Ошибка	1,363084
Наблюдения	15

Продолжение табл. 4

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	2	290,4628387	145,231419	78,16547142	1,31E–07
Остаток	12	22,29599593	1,85799966		
Итого	14	312,7588347			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-значение*	*Нижние 95 %*	*Верхние 95 %*
Y-пересечение	232,0028	10,78827	21,5051052	5,9691E-11	208,49	255,508
Фиктивная переменная	3,474500	0,7109700	4,8869856	0,00037407	1,9254	5,02357
Цена x(р.)	-7,30442	0,675558	–10,8124125	1,5303E–07	–8,77634	–5,83251

Регрессионная модель имеет вид: $y = 232 + 3{,}47D - 7{,}304x$

Поскольку значение фиктивной переменной D равно 1 для «нового» вида и 0 для «обычного», то данную модель можно отдельно расписать для каждого случая.
$y = 232 - 7{,}304x$ — обычный вид,
$y = 235{,}47 - 7{,}304x$ — новый вид.

Следовательно, спрос на новый вид продукции приблизительно на 3,47 тыс. ед. больше. Коэффициент детерминации равен 0,928, что намного больше, чем данное значение для парного случая.

Рассмотрим теперь возможность построения модели с фиктивной переменной наклона, для чего в качестве регрессоров значения необходимо использовать переменные x и Dx. Следовательно, необходимо добавить дополнительный столбец между фиктивной переменной и значениями x, в который надо записать их произведения.

Опустим таблицу, которая генерируется надстройкой **Регрессия**. Однако, самостоятельно выполнив данные операции, можно получить следующую модель: $y = 233{,}52 + 0{,}21Dx - 7{,}403x$.

Аналогичным образом интерпретируя значение фиктивной переменной, можно расписать два случая:

$y = 233{,}52 - 7{,}4x$ — для обычного вида продукции;

$y = 233{,}52 - 7{,}19x$ — для нового вида продукции.

Выводы из полученных моделей совершенно очевидны, поскольку видна разница во влиянии цены на спрос для каждого вида продукции. Коэффициент детерминации в этом случае равен 0,929, что не намного больше соответствующего значения для фиктивной переменной сдвига, а следовательно, они обе пригодны для прогнозирования. Однако результаты использования моделей будут во многом различными. В первом случае спрос на «новый» вид продукции на 3,47 тыс. ед. больше, чем на «старый», во втором случае цена сильнее влияет на «старый» вид продукции.

При необходимости можно построить модель, в которой фиктивная переменная влияет как на наклон, так и на сдвиг.

До сих пор нами рассматривался случай, когда имеются всего два значения качества, то есть два вида продукции. Однако нередки случаи, когда необходимо проанализировать спрос для различных продуктов. Тогда необходимо вводить *набор категорий* — как конечный набор взаимоисключающих событий, полностью описывающий все возможности. Предположим, что исследуется влияние цены на спрос при наличии «старой», «обычной», «новой» и «самой новой» продукции.

В этом случае для описания этих категорий необходимо вводить набор фиктивных переменных по следующему правилу.

1. Число фиктивных переменных должно быть на единицу меньше, чем число категорий. В данном случае имеется четыре категории, а следовательно, необходимо ввести три фиктивные переменные, которые мы обозначим D1, D2, D3.
2. Выбрать произвольную категорию в качестве эталонной. Именно с этой категорий в последствии будут сравниваться все остальные. Для эталонной категории необходимо, чтобы значения всех фиктивных переменных равнялись нулю.
3. Для всех остальных категорий необходимо, чтобы одна из фиктивных переменных равнялась 1, в то время как значение всех остальных равно 0.

Достаточно легко можно расставить значения фиктивных переменных, используя ту же условную функцию ЕСЛИ. При наличии четырёх различных видов продукции необходимо вставить три дополнительных столбца, в которых будут находиться фиктивные переменных. Задать логические функции можно так, как показано в таблице 20.

Таблица 5

Номер наблюдения	Вид	Фиктивная переменная D1	Фиктивная переменная D2	Фиктивная переменная D3	Цена x^1 (р.)	Спрос y (тыс.шт.)
1		=ЕСЛИ(B2= «обычный»;1;0)	=ЕСЛИ(B2= «новой»;1;0)	=ЕСЛИ(B2= «самой новый»;1;0)	15,09р.	125,1779

После копирования данных функций вниз для значения старой все фиктивные переменные будут равны нулю, для обычной — только значение первой фиктивной переменной будет равно 1 и т. д.

После этого можно вызвать надстройку **Регрессия**, у которой в качестве входного интервала X, необходимо указать значения всех фиктивных переменных D и нефиктивной переменной *X*, то есть задать *Входной интервал X* в виде ***C1:F16***.

Полученные результаты поддаются достаточно простой интерпретации. Значение, находящееся напротив фиктивной переменной $D1$, показывает, насколько изменился спрос при переходе от эталонной к первой категории, то есть насколько различен спрос между «обычной» и «новой» продукцией. Аналогично интерпретируются значения, стоящие напротив других фиктивных переменных.

Задания для самостоятельной работы

2. Для данных своего варианта подобрать наилучшее воздействие фиктивной переменной (влияние на наклон или сдвиг). При этом категории «старый» и «обычный» воспринимать как одно значение, а категории «новый» и «самый новый» — как другое.
3. Определить, насколько изменяется спрос при переходе от одной категории к другой.

Лабораторная работа №6. Идентификация модели.

Пример решения типовой задачи

Рассмотрим **пример.** Изучается модель вида

$$\begin{cases} C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + \varepsilon_2, \\ r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C_t – расходы на потребление в период t, Y_t – совокупный доход в период t, I_t – инвестиции в период t, r_t – процентная ставка в период t, M_t – денежная масса в период t, G_t – государственные расходы в период t, C_{t-1} – расходы на потребление в период $t-1$, I_{t-1} инвестиции в период $t-1$.

Первое уравнение – функция потребления, второе уравнение – функция инвестиций, третье уравнение – функция денежного рынка, четвертое уравнение – тождество дохода.

Модель представляет собой систему одновременных уравнений. Проверим каждое ее уравнение на идентификацию.

Модель включает четыре эндогенные переменные (C_t, I_t, Y_t, r_t) и четыре предопределенные переменные (две экзогенные переменные – M_t и G_t и две лаговые переменные – C_{t-1} и I_{t-1}).

1. Проверим необходимое условие идентификации для каждого из уравнений модели.

Первое уравнение: $C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1$. Это уравнение содержит две эндогенные переменные C_t и Y_t и одну предопределенную переменную C_{t-1}. Таким образом, $H = 2$, а $D = 4 - 1 = 3$, т.е. выполняется условие $D + 1 > H$. Уравнение сверхидентифицируемо.

Второе уравнение: $I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + \varepsilon_2$. Оно включает две эндогенные переменные I_t и r_t и одну экзогенную переменную I_{t-1}. Выполняется условие $D + 1 = 3 + 1 > H = 2$. Уравнение сверхидентифицируемо.

Третье уравнение: $r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + \varepsilon_3$. Оно включает две эндогенные переменные Y_t и r_t и одну экзогенную переменную M_t. Выполняется условие $D + 1 = 3 + 1 > H = 2$. Уравнение сверхидентифицируемо.

Четвертое уравнение: $Y_t = C_t + I_t + G_t$. Оно представляет собой тождество, параметры которого известны. Необходимости в идентификации нет.

2. Проверим для каждого уравнения достаточное условие идентификации. Для этого составим матрицу коэффициентов при переменных модели.

	C_t	I_t	r_t	Y_t	C_{t-1}	I_{t-1}	M_t	G_t
I уравнение	–1	0	0	b_{11}	b_{12}	0	0	0
II уравнение	0	–1	b_{21}	0	0	b_{22}	0	0
III уравнение	0	0	–1	b_{31}	0	0	b_{32}	0
Тождество	1	1	0	–1	0	0	0	1

В соответствии с достаточным условием идентификации ранг матрицы коэффициентов при переменных, не входящих в исследуемое уравнение, должен быть равен числу эндогенных переменных модели без одного.

Первое уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	I_t	r_t	I_{t-1}	M_t	G_t
II уравнение	–1	b_{21}	b_{22}	0	0
III уравнение	0	–1	0	b_{32}	0
Тождество	1	0	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{22} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{22} b_{32} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Второе уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	C_t	Y_t	C_{t-1}	M_t	G_t
I уравнение	–1	b_{11}	b_{12}	0	0
III уравнение	0	b_{31}	0	b_{32}	0
Тождество	1	–1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12} b_{32} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Третье уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	C_t	I_t	C_{t-1}	I_{t-1}	G_t
I уравнение	–1	0	b_{12}	0	0
II уравнение	0	–1	0	b_{22}	0
Тождество	1	1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12} b_{22} \neq 0 \cdot$$

Достаточное условие идентификации для данного уравнения выполняется.

Таким образом, все уравнения модели сверхидентифицируемы. Приведенная форма модели в общем виде будет выглядеть следующим образом:

$$\begin{cases} C_t = A_1 + \delta_{11}C_{t-1} + \delta_{12}I_{t-1} + \delta_{13}M_t + \delta_{14}G_t + u_1, \\ I_t = A_2 + \delta_{21}C_{t-1} + \delta_{22}I_{t-1} + \delta_{23}M_t + \delta_{24}G_t + u_2, \\ r_t = A_3 + \delta_{31}C_{t-1} + \delta_{32}I_{t-1} + \delta_{33}M_t + \delta_{34}G_t + u_3, \\ Y_t = A_4 + \delta_{41}C_{t-1} + \delta_{42}I_{t-1} + \delta_{43}M_t + \delta_{44}G_t + u_1. \end{cases}$$

Варианты индивидуальных заданий

Даны системы эконометрических уравнений.

Требуется

1. Применив необходимое и достаточное условие идентификации, определите, идентифицируемо ли каждое из уравнений модели.

2. Определите метод оценки параметров модели.

3. Запишите в общем виде приведенную форму модели.

Вариант 1

Модель протекционизма Сальватора (упрощенная версия):

$$\begin{cases} M_t = a_1 + b_{12}N_t + b_{13}S_t + b_{14}E_{t-1} + b_{15}M_{t-1} + \varepsilon_1, \\ N_t = a_2 + b_{21}M_t + b_{23}S_t + b_{26}Y_t + \varepsilon_2, \\ S_t = a_3 + b_{31}M_t + b_{32}N_t + b_{36}X_t + \varepsilon_3. \end{cases}$$

где M – доля импорта в ВВП; N – общее число прошений об освобождении от таможенных пошлин; S – число удовлетворенных прошений об освобождении от таможенных пошлин; E – фиктивная переменная, равная 1 для тех лет, в которые курс доллара на международных валютных рынках был искусственно завышен, и 0 – для всех остальных лет; Y – реальный ВВП; X – реальный объем чистого экспорта; t – текущий период; $t-1$ – предыдущий период.

Вариант 2

Макроэкономическая модель (упрощенная версия модели Клейна):

$$\begin{cases} C_t = a_1 + b_{12}Y_t + b_{13}T_t + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + b_{24}K_{t-1} + \varepsilon_2, \\ Y_t = C_t + I_t, \end{cases}$$

где C – потребление; I – инвестиции; Y – доход; T – налоги; K – запас капитала; t – текущий период; $t-1$ – предыдущий период.

Вариант 3

Макроэкономическая модель экономики США (одна из версий):

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + b_{23}r_t + \varepsilon_2, \\ r_t = a_3 + b_{31}Y_t + b_{34}M_t + b_{35}r_{t-1} + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C – потребление; Y – ВВП; I – инвестиции; r – процентная ставка; M – денежная масса; G – государственные расходы; t – текущий период; $t-1$ –

предыдущий период.

Вариант 4

Модель Кейнса (одна из версий):

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}Y_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + \varepsilon_2, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C – потребление; Y – ВВП; I – валовые инвестиции; G – государственные расходы; t – текущий период; $t-1$ – предыдущий период.

Вариант 5

Модель денежного и товарного рынков:

$$\begin{cases} R_t = a_1 + b_{12}Y_t + b_{14}M_t + \varepsilon_1, \\ Y_t = a_2 + b_{21}R_t + b_{23}I_t + b_{25}G_t + \varepsilon_2, \\ I_t = a_3 + b_{31}R_t + \varepsilon_3, \end{cases}$$

где R – процентные ставки; Y – реальный ВВП; M – денежная масса; I – внутренние инвестиции; G – реальные государственные расходы.

Вариант 6

Модифицированная модель Кейнса:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + b_{22}Y_{t-1} + \varepsilon_2, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C – потребление; Y – доход; I – инвестиции; G – государственные расходы; t – текущий период; $t-1$ – предыдущий период.

Вариант 7

Макроэкономическая модель:

$$\begin{cases} C_t = a_1 + b_{11}D_t + \varepsilon_1, \\ I_t = a_2 + b_{22}Y_t + b_{23}Y_{t-1} + \varepsilon_2, \\ Y_t = D_t + T_t, \\ D_t = C_t + I_t + G_t, \end{cases}$$

где C – расходы на потребление; Y – чистый национальный продукт; D – чистый национальный доход; I – инвестиции; T – косвенные налоги; G – государственные расходы; t – текущий период; $t-1$ – предыдущий период.

Вариант 8

Гипотетическая модель экономики:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}J_t + \varepsilon_1, \\ J_t = a_2 + b_{21}Y_{t-1} + \varepsilon_2, \\ T_t = a_3 + b_{31}Y_t + \varepsilon_3, \\ Y_t = C_t + J_t + G_t, \end{cases}$$

где C – совокупное потребление в период t; Y – совокупный доход в период t; J – инвестиции в период t; T – налоги в период t; G – государственные доходы в период t.

Вариант 9

Модель денежного рынка:

$$\begin{cases} R_t = a_1 + b_{11}M_t + b_{12}Y_t + +\varepsilon_1, \\ Y_t = a_2 + b_{21}R_t + b_{22}I_t + \varepsilon_2, \\ I_t = a_3 + b_{33}R_t + \varepsilon_3, \end{cases}$$

где R – процентные ставки; Y – ВВП; M – денежная масса; I – внутренние инвестиции.

Вариант 10

Конъюнктурная модель имеет вид:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21}r_t + b_{22}I_{t-1} + \varepsilon_2, \\ r_t = a_3 + b_{31}Y_t + b_{32}M_t + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C – расходы на потребление; Y – ВВП; I – инвестиции; r – процентная ставка; M – денежная масса; G – государственные расходы; t – текущий период; $t-1$ – предыдущий период.

Лабораторная работа №7. Оценивание параметров структурной модели.

Двухшаговый МНК основан на использовании, так называемых, «инструментальных» переменных и является универсальным методом. Как уже отмечалось, в системе одновременных уравнений нарушаются предпосылки о независимости факторов (выражаемых эндогенными переменными) и ошибок уравнений. Для преодоления этой трудности можно использовать замену эндогенных переменных y_i в правых частях уравнений модели на вспомогательные «инструментальные» переменные $\hat{y}_i$, которые были бы близки к исходным эндогенным переменным и при этом не зависели бы от ошибок уравнений. В качестве таких переменных предлагается использовать переменные, определяемые уравнениями приведенной формы модели

Согласно *двухшаговому МНК*, численные значения структурных параметров определяются в следующей последовательности:

1) Исходная система уравнений преобразуется в приведенную форму модели и определяются численные значения параметров δ_{ij} для каждого ее уравнения в отдельности с помощью традиционного МНК;

2) По полученным уравнениям приведенной формы находятся расчетные значения инструментальных переменных $\hat{y}_i$, соответствующих эндогенным переменным y_i для каждого наблюдения;

3) С помощью обычного МНК определяются параметры каждого структурного уравнения в отдельности, используя в качестве факторов фактические значения предопределенных переменных и полученные расчетные значения инструментальных переменных $\hat{y}_i$.

Более эффективным, но требующим существенно больших вычислительных затрат, является трехшаговый метод наименьших квадратов (ТМНК).
Он заключается в том, что двухшаговый метод наименьших квадратов применяется не к исходным уравнениям модели, а к уравнениям, преобразованным согласно обобщенному методу наименьших квадратов. Трехшаговый МНК является итерационной процедурой:

1) Параметры модели определяются обычным или двухшаговым МНК.

2) Вычисляются ошибки модели и определяется оценка корреляционной матрицы ошибок.

3) Уравнения преобразуются согласно обобщенному МНК.

4) Применяется двухшаговый МНК к преобразованным уравнениям и получается улучшенная модель (с улучшенными параметрами).

5) Процесс повторяется, начиная со второго шага, пока не будет достигнута заданная точность (либо превышено заданное количество итераций). Если случайные члены структурной модели не коррелируют, то трехшаговый метод сводится к двухшаговому.

<u>Задание</u>

1. Дана модифицированная модель Кейнса:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + \varepsilon_1, \\ I_t = a_2 + b_{21}Y_t + b_{22}Y_{t-1} + \varepsilon_2, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где

C – потребление; Y – доход; I – инвестиции; G – государственные расходы;
t – текущий период; $t-1$ – предыдущий период.

Годы	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Y	95,75	98,55	103,55	109	108,25	107,4	112,7	117,75	123,45	126,55	125,85	128,1	125,35	130,25	138,3	142,65	146,80	151,3	157,4	161,25
C	60,45	62,45	65,9	68,9	68,45	70	73,55	76,55	79,7	81,6	81,55	82,55	83,45	87,35	91,55	95,50	99	101,75	105,4	107,45
I	14,3	15,85	17,75	19,7	18,1	14,6	17,35	20	22,15	22,3	19,8	21	18	20	25,25	24,85	24,5	25	25,8	26,15

а) В предположении, что потребление зависит линейно от дохода (первое уравнение модели), оцените по МНК параметры a_1 и b_{11} функции потребления.

б) Оцените те же параметры по ДМНК и по ТМНК

в) Сравните полученные результаты. Сделайте выводы по качеству оценок.

Лабораторная работа №7. Изучение взаимосвязей по временным рядам. Построение аддитивной и мультипликативной модели временного ряда.

Основные понятия.

Временной ряд (динамический ряд, или ряд динамики) последовательность наблюдений некоторого признака (случайной величины) Y в последовательные моменты времени

$$y_t = u_t + v_t + c_t + \varepsilon_t \ \ (t = 1,2,3,\dots, n)$$

u_t* – *тренд, плавно меняющаяся компонента, описывающая чистое влияние долговременных факторов, то есть длительную тенденцию изменения признака

v_t* – *сезонная компонента, отражающая повторяемость экономических процессов в течение не очень длительного периода (года, месяца, недели и т.д.)

c_t* – *циклическая компонента, отражающая повторяемость экономических процессов в течение длительных периодов

e_t* – *случайная компонента, отражающая влияние не поддающихся учету и регистрации случайных факторов.

Автокорреляция уровней ряда – корреляционная зависимость между последовательными уровнями временного ряда

Коэффициент автокорреляции уровней ряда первого порядка

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}},$$

Коэффициент автокорреляции уровней ряда второго порядка

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}},$$

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют ***автокорреляционной функцией*** временного ряда.

Распространенным способом моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда. Этот способ называют *аналитическим выравниванием временного ряда*.

Поскольку зависимость от времени может принимать разные формы, для ее формализации можно использовать различные виды функций. Для построения трендов чаще всего применяются следующие функции:

линейный тренд: $y_t = a + b \cdot t$;

гипербола: $y_t = a + b/t$;

экспоненциальный тренд: $y_t = e^{a+b\cdot t}$ (или $y_t = a \cdot b^t$);

степенная функция: $y_t = a \cdot t^b$;

полиномы различных степеней: $y_t = a + b_1 \cdot t + b_2 \cdot t^2 + ... + b_m \cdot t^m$.

Параметры каждого из перечисленных выше трендов можно определить обычным МНК, используя в качестве независимой переменной время $t = 1, 2, ..., n$, а в качестве зависимой переменной – фактические уровни временного ряда y_t. Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Существует несколько способов определения типа тенденции. К числу наиболее распространенных способов относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени. В этих же целях можно использовать и коэффициенты автокорреляции уровней ряда. Тип тенденции можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда. Если временной ряд имеет линейную тенденцию, то его соседние уровни y_t и y_{t-1} тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким. Если временной ряд содержит нелинейную тенденцию, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражена нелинейная тенденция в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

Выбор наилучшего уравнения в случае, когда ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации и средней ошибки аппроксимации. Этот метод легко реализуется при компьютерной обработке данных.

Простейший подход к моделированию сезонных колебаний – это расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Общий вид аддитивной модели следующий:

$$Y = T + S + E.$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой (*T*), сезонной (*S*) и случайной (*E*) компонент.

Общий вид мультипликативной модели выглядит так:

$$Y = T \cdot S \cdot E.$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой (*T*), сезонной (*S*) и случайной (*E*) компонент.

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений трендовой (*T*), сезонной (*S*) и случайной (*E*) компонент для каждого уровня ряда.

Процесс построения модели включает в себя следующие шаги.

1) Выравнивание исходного ряда методом скользящей средней.
2) Расчет значений сезонной компоненты S.
3) Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных $(T + E)$ в аддитивной или $(T \cdot E)$ в мультипликативной модели.
4) Аналитическое выравнивание уровней $(T + E)$ или $(T \cdot E)$ и расчет значений T с использованием полученного уравнения тренда.
5) Расчет полученных по модели значений $(T + E)$ или $(T \cdot E)$.
6) Расчет абсолютных и/или относительных ошибок. Если полученные значения ошибок не содержат автокорреляции, ими можно заменить исходные уровни ряда и в дальнейшем использовать временной ряд ошибок E для анализа взаимосвязи исходного ряда и других временных рядов.

Задание

Имеются условные данные об объемах потребления электроэнергии (y_t) жителями региона за 16 кварталов.

Требуется:

1. Построить автокорреляционную функцию и сделать вывод о наличии сезонных колебаний.

2. Построить аддитивную модель временного ряда (для нечетных вариантов) или мультипликативную модель временного ряда (для четных вариантов).

3. Сделать прогноз на 2 квартала вперед.

Варианты 1, 2

t	y_t	t	y_t
1	5,8	9	7,9
2	4,5	10	5,5
3	5,1	11	6,3
4	9,1	12	10,8
5	7,0	13	9,0
6	5,0	14	6,5
7	6,0	15	7,0
8	10,1	16	11,1

Варианты 3, 4

t	y_t	t	y_t
1	5,5	9	8,0
2	4,6	10	5,6
3	5,0	11	6,4
4	9,2	12	10,9
5	7,1	13	9,1
6	5,1	14	6,4
7	5,9	15	7,2
8	10,0	16	11,0

Варианты 5, 6

t	y_t	t	y_t
1	5,3	9	8,2
2	4,7	10	5,5
3	5,2	11	6,5
4	9,1	12	11,0
5	7,0	13	8,9
6	5,0	14	6,5
7	6,0	15	7,3
8	10,1	16	11,2

Варианты 7, 8

t	y_t	t	y_t
1	5,5	9	8,3
2	4,8	10	5,4
3	5,1	11	6,4
4	9,0	12	10,9
5	7,1	13	9,0
6	4,9	14	6,6
7	6,1	15	7,5
8	10,0	16	11,2

Варианты 9, 10

t	y_t	t	y_t
1	5,6	9	8,2
2	4,7	10	5,6
3	5,2	11	6,4
4	9,1	12	10,8
5	7,0	13	9,1
6	5,1	14	6,7
7	6,0	15	7,5
8	10,2	16	11,3

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Введение

Самостоятельная работа предусмотрена учебным планом. Цель самостоятельной работы студента в рамках курса «Эконометрика» — закрепление и расширение знаний, полученных во время проведения аудиторных занятий.

Содержание самостоятельной работы

1. Проработка лекционного материала осуществляется студентом с использованием конспекта лекций и рекомендуемых учебников. Цель — подготовка к восприятию очередной темы, рассматриваемой на лекции.

2. Подготовка к лабораторным работам. В соответствии с темой лабораторной работы студент должен изучить теоретический материал, подготовить решение задания к реализации на компьютере.
   Темы лабораторных (соответственно, самостоятельных) работ:
   1. Парная регрессия
   2. Множественная регрессия
   3. Различные аспекты множественной регрессии
   4. Системы эконометрических уравнений
   5. Временные ряды

3. В рамках раздела «Изучение дополнительных тем курса» студент самостоятельно изучает дополнительные вопросы, связанные с построением и анализом моделей множественной регрессии, систем эконометрических уравнений и эконометрических моделей по временным рядам. Для достижения этой цели сформулированы следующие задания:
   - Построение и анализ множественной нелинейной модели.
   - Трехшаговый метод наименьших квадратов.
   - Автокорреляция. Обнаружение и методы устранения автокорреляции.
   - Авторегрессионные модели.

Список литературы

1. Тихомиров, Николай Петрович. Эконометрика : учебник для вузов / Н. П. Тихомиров, Е. Ю. Дорохина . — М. : ЭКЗАМЕН, 2007 – 510[2] с. : ил., табл. (в библиотеке 11 экз.) (**Гриф**)

2. Яновский, Леонид Петрович. Введение в эконометрику : учебное пособие для вузов / Л. П. Яновский, А. Г. Буховец ; ред. Л. П. Яновский. - 2-е изд., доп. — М. : КноРус, 2009. - 254[2] с. : ил., табл. (в библиотеке 10 экз.)

3. Эконометрика : учебник для вузов / И. И. Елисеева [и др.] ; ред. И. И. Елисеева. - 2-е изд., перераб. и доп. - М. : Финансы и статистика, 2008. - 574[2] с. : ил., табл. (в библиотеке 5 экз.) (**Гриф**)

3.2. Дополнительная литература

1. Орлов, Александр Иванович. Эконометрика: Учебник для вузов/ А. И. Орлов. — 3-е изд., перераб и доп.. — М.: Экзамен, 2004. - 573[3] с.. (в библиотеке 1 экз.)

2. Практикум по эконометрике: Учебное пособие для вузов / Ирина Ильинична Елисеева, Светлана Владимировна Курышева, Нелли Михайловна Гордеенко и др; Ред. И. И. Елисеева. - М.: Финансы и статистика, 2001. - 192 с. (в библиотеке 2 экз.)

3. Бородич, Сергей Аркадьевич. Эконометрика: Учебное пособие для вузов. — Минск: Новое знание, 2001. - 408[8] с. : ил. (в библиотеке 4 экз.) (**Гриф**)

4. Кремер, Наум Шевелевич. Эконометрика: Учебник для вузов. — М.: ЮНИТИ-ДАНА, 2003. - 311 с. : ил. (в библиотеке 2 экз.) (**Гриф**)