(Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное
учреждение высшего образования
**«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ» (ТУСУР)**

Кафедра автоматизации обработки информации (АОИ)

ЭКОНОМЕТРИКА

Методические указания к лабораторным работам
и организации самостоятельной работы
для студентов направлений
«Государственное и муниципальное управление»
(уровень бакалавриата)
«Бизнес-информатика» (уровень бакалавриата)

2018

Потахова Ирина Владимировна
Эконометрика: Методические указания к лабораторным работам и организации самостоятельной работы для студентов направлений «Государственное и муниципальное управление» (уровень бакалавриата), «Бизнес-информатика» (уровень бакалавриата) / И.В. Потахова. – Томск, 2018. – 61 с.

© Томский государственный университет систем управления и радиоэлектроники, 2018
© Потахова И.В., 2018

Оглавление

1 Введение

«Эконометрика» – учебная дисциплина, в рамках которой студенты изучают методы проверки, обоснования, оценивания количественных закономерностей и качественных утверждений на основе анализа статистических данных.

Целью проведения лабораторных работ и организации самостоятельной деятельности по дисциплине «Эконометрика» является закрепление теоретического материала, формирование и развитие навыков анализа и интерпретации статистических данных.

Для выполнения лабораторных работ и осуществления самостоятельной деятельности от студента требуются знания, умения и навыки, полученные при изучении дисциплины «Теория вероятностей и математическая статистика».

Лабораторные работы предусматривают выполнение ряда практических заданий. Перед началом работы студент должен ознакомиться с теоретическим материалом по соответствующей теме.

Работы рекомендуется выполнять в порядке их следования.

Итогом выполнения лабораторных работ является: ответ на тест по теме лабораторной работы; результаты расчетов с пояснениями; ответы на вопросы преподавателя.

Методические указания по проведению лабораторных работ

2.1 Общие положения

Основными целями лабораторных работ являются:

закрепление теоретического материала, посредством решения индивидуальных задач, связанных с вопросами построения и анализа эконометрических моделей;

интерпретация полученных результатов.

Основной формой проведения лабораторных работ является решение задач с использованием средств вычислительной техники (рекомендуется использовать любой из доступных табличных процессоров).

Контроль формирования компетенций осуществляется посредством индивидуальной и (или) групповой защиты отчетов по лабораторным работам и (или) опроса студентов по вопросам, образующим тематическое наполнение каждого из заданий. По итогам или в ходе защиты отчетов (опроса) производится разбор типичных трудностей и ошибок, допущенных студентами в ходе выполнения заданий.

2.2 Лабораторная работа « Построение и анализ модели линейной парной регрессии».

Цель работы: закрепление теоретического материала по вопросам построения и исследования уравнения парной линейной регрессии.

Форма проведения: выполнение индивидуального задания.

Форма отчетности: выполнение теста, защита отчета.

Теоретические основы

<u>Спецификация модели парной регрессии</u>

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

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

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 $\hat{y} = f(x)$.

Различают линейные и нелинейные регрессии.

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

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

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

<u>Параметризация модели парной линейной регрессии. Метод наименьших квадратов (МНК)</u>

Коэффициенты (a, b), определяемые на основе метода наименьших квадратов, являются решением системы уравнений:

$$\begin{cases} a + b \cdot \overline{x} = \overline{y} \\ a \cdot \overline{x} + b \cdot \overline{x^2} = \overline{x \cdot y} \end{cases}$$

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

где

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i; \qquad \overline{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i; \qquad \overline{x \cdot y} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \cdot y_i;$$

$$\overline{x_i^2} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2,$$

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

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

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

<u>Оценка качества уравнения линейной регрессии</u>

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

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

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

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

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

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

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

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

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

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.

Оценка значимости уравнения линейной регрессии и существенности параметров линейной регрессии.

Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Оценка значимости уравнения регрессии в целом осуществляется с помощью F-критерия и заключается в проверке гипотезы H_0 о статистической незначимости уравнения регрессии.

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

$$F = \frac{S^2_{факт}}{S^2_{ост}} > F_{\alpha;1;n-2}$$

где

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

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

$F_{\alpha;m;n-2}$ – значения квантиля уровня α F-распределения с числами степеней свободы $k_1 = m$; $k_2 = n-2$.

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

Оценка значимости коэффициента линейной регрессии.

Для проверки значимости 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 вычисляется по формуле:

$$t_b = \frac{b}{m_b}.$$

Стандартная ошибка коэффициента (b) регрессии определяется по формуле:

$$m_b = \frac{S_{ост}}{\sigma_x \cdot \sqrt{n}},$$

где $S_{ост}^2 = \dfrac{\sum_{i=1}^{n} (y - \hat{y}_x)^2}{n - m - 1}$, — 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.

Построение интервальных оценок для параметров регрессии, функции парной линейной регрессии

Интервальная оценка (доверительный интервал) для коэффициента b с надежностью (доверительной вероятностью) равной $\gamma = 1 - \alpha$ определяется выражением:

$$b \pm t_{табл} \cdot m_b$$

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

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

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

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

Стандартная ошибка вычисленного значения $\hat{y}_i$ определяется по формуле: $m_{\hat{y}_i} = S_{ост} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_i - \overline{x})^2}{n \cdot \sigma_x^2}}$

Индивидуальные задания

1. Построить уравнение регрессии, описывающее зависимость средней заработной платы за один рабочий день (y) от величины среднедушевого прожиточного минимума в день на одного работающего (x). Поясните экономический смысл коэффициента регрессии.

2. Оценить степень тесноты связи между переменными с помощью коэффициента корреляции.

3. Оценить качество модели с помощью коэффициента детерминации и средней ошибки аппроксимации.

4. Рассчитать:

a. стандартную ошибку регрессии;
b. стандартные ошибки оценок параметров уравнения регрессии.

Проверить существенность параметров уравнения регрессии и построить их интервальные оценки на уровне значимости 0,05.

5. Оценить статистическую значимость уравнения регрессии с помощью F-критерия Фишера.

6. Найти стандартную ошибку и доверительный интервал для уравнения регрессии в целом, а также для индивидуального прогнозного значения y_p для $x_p = 1{,}5 \cdot \overline{x}$.

7. Оформить отчет.

Варианты задания

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 7		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

2.3. Лабораторная работа. Построение и анализ модели нелинейной парной регрессии

Цель: закрепление теоретического материала по вопросам построения и исследования уравнения нелинейной регрессии.

Форма проведения: выполнение индивидуального задания.

Форма отчетности: выполнение теста, защита отчета.

Теоретические основы

<u>Спецификация модели парной регрессии</u>

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

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

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 $\hat{y} = f(x)$.

В данной работе изучается вопрос параметризации нелинейной регрессии с применением метода наименьших квадратов (МНК).

Различают два вида нелинейных регрессий:

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

- равносторонняя гипербола: $y = a + \frac{b}{x} + \varepsilon$.
- полулогарифмическая: $y = a + b \cdot \ln(x) + \varepsilon$
- с квадратным корнем: $y = a + b \cdot \sqrt{x} + \varepsilon$

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

- логарифмическая модель (степенная): $y = a \cdot x^b \cdot \varepsilon$

- показательная: $y = a \cdot b^x \cdot \varepsilon$;
- экспоненциальная: $y = a \cdot e^{b \cdot x} \cdot \varepsilon$.

Оценка параметров нелинейной регрессии

При оценивании параметров регрессий, нелинейных относительно включенных в анализ объясняющих переменных, 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	$\hat{y}_x = a + b \cdot \dfrac{1}{x}$	$X = \dfrac{1}{x}$	$x \neq 0$	$a = a$	$b = b$
3	$y = a + b \cdot \sqrt{x} + \varepsilon$	$X = \sqrt{x}$	$x > 0$	$a = a$	$b = b$

Оценка параметров регрессий, нелинейных относительно оцениваемых параметров, 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	$\hat{y}_x = a \cdot e^{b \cdot x}$	$Y = \ln y,\ \ A = \ln a$		$a = a$	$b = b$

<u>Оценка качества уравнения нелинейной регрессии</u>

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

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

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

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

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

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

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

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

Следует обратить внимание на то, что разности в соответствующих суммах $\sum(y - \bar{y})^2$, $\sum(\hat{y} - \bar{y})^2$ *и* $\sum(y - \hat{y}_x)^2$ *берутся не в преобразованных, а* ***в исходных значениях результативного признака, если уравнение регрессии нелинейно относительно объясняющих переменных****. Иначе говоря, при вычислении этих сумм следует использовать не преобразованные (линеаризованные) зависимости, а исходные нелинейные уравнения* *регрессии.* ***В случае регрессий, нелинейных относительно параметров соответствующие суммы вычисляются в преобразованных данных***.

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

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

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:

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

Таблица. Формулы для расчета $\bar{Э}$ средних коэффициентов эластичности.

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

$y = a + b \cdot \sqrt{x} + \varepsilon$	$\dfrac{b}{2 \cdot \sqrt{x}}$	$\dfrac{b \cdot \sqrt{\bar{x}}}{2 \cdot (a + b \cdot \sqrt{\bar{x}})}$
$y = a \cdot b^x \cdot \varepsilon$	$a \cdot \ln b \cdot b^x$	$\bar{x} \cdot \ln b$
$y = a + b \cdot \ln x + \varepsilon$	$\dfrac{b}{x}$	$\dfrac{b}{a + b \cdot \ln \bar{x}}$
$y = a \cdot e^{b \cdot x} \cdot \varepsilon$	$a \cdot b \cdot e^{b \cdot x}$	$b \cdot \bar{x}$

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.

Индивидуальные задания

1. На основании таблиц 4-8 построить предложенные уравнения регрессий.

2. Вычислить индексы парной корреляции и индексы детерминации для каждого уравнения. Дать сравнительную оценку силы связи фактора с результатом.

3. Вычислить среднюю ошибку аппроксимации для каждого уравнения.

4. Вычислить средний коэффициент эластичности. Дать смысловую интерпретацию.

5. Оценить с помощью F-критерия Фишера статистическую надежность результатов регрессионного моделирования.

6. Рассчитать прогнозное значение результата, если прогнозное значение фактора увеличится на *15%* от его среднего уровня. Определить доверительный интервал прогноза для уровня значимости α=0,05.

7. Свести данные анализа уравнений регрессий в таблицу, дополнив вычислениями по парной регрессии. Сделать выводы по проделанной работе:

   a. Определить лучшее уравнение на основе средней ошибки аппроксимации.

   b. Определить лучшее уравнение на основе совместного анализа значений индекса детерминации и средней ошибки аппроксимации.

Таблица 4. Варианты заданий

Вариант	Виды кривых аппроксимации					
	Гиперболи-ческая	Полулогариф-мическая	С квадратным корнем	Степенная	Показательная	Экспоненци-альная
1	*			*		
2		*		*		
3			*	*		
4	*					*
5		*			*	
6			*		*	
7	*				*	
8		*				*
9			*			*
10	*			*		
11		*				*
12			*		*	

Таблица 5 Данные для построения уравнений регрессий

Варианты 1-3	
Личный располагаемый доход, тыс.	*Расходы на моторное масло, тыс. руб.*

руб.	
x	***y***
622,9	*4,9*
658	*5,2*
700,4	*5,5*
740,6	*5,6*
774,4	*5,6*
816,2	*5,3*
853,5	*5*
876,8	*4,7*
900	*4,6*
951,4	*5*
1007,9	*5,4*
1004,8	*4,2*
1010,8	*4,2*
1056,2	*4,6*
1105,4	*4,4*
1162,3	*4,7*
1200,7	*4,7*
1209,5	*3,9*
1248,6	*3,6*
1254,4	*3,6*
1284,6	*4*

Таблица 6 Данные для построения уравнений регрессий

Варианты 4-6	
Фактическое конечное потребление домашних хозяйств (в текущих ценах), млрд. руб.	Среднедушевые денежные доходы населения (в месяц), руб.
y	***x***
872	515,9
3813	2281,1
5014	3062,0
6400	3947,2
7708	5170,4

9848	6410,3
12455	8111,9
15284	10196,0
18928	12602,7
23695	14940,6
25151	16857,0

Таблица 7 Данные для построения уравнений регрессий

Варианты 7-9	
Вес грузов	Заказы
x (кг)	***y (тыс.шт)***
554	6,1
719	9,1
608	7,2
526	7,5
666	6,9
955	11,5
865	10,3
767	9,5
931	9,2
971	10,6
1022	12,5
1075	12,9
1215	14,5
1092	13,6
1035	12,8
1393	16,5
1443	17,1
1278	15,0
1336	16,2
1266	15,8
1584	19,0
1706	19,4
1524	19,1
1590	18,0
1644	20,2

Таблица 8 Данные для построения уравнений регрессий

Варианты 10-12	
Прожиточный минимум	Заработная плата
y (руб)	***x (руб)***
234	445
246	484
261	518
237	457
267	524
318	623
201	396
264	517
219	434
261	517
228	449
345	685
207	419
252	526
276	553

2.4. Лабораторная работа. Построение и анализ модели множественной линейной регрессии

Цель работы: построение модели с большим числом факторов и определение влияния каждого фактора в отдельности, а также их совместного воздействия на моделируемый показатель (результат).

Форма проведения: выполнение индивидуального задания.

Форма отчетности: выполнение теста, защита отчета.

Теоретические основы

<u>Спецификация модели множественной линейной регрессии</u>

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.

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.

<u>Вычисление коэффициентов линейной множественной регрессии. Матричный метод.</u>

Пример. Данные о сменной добыче угля на одного рабочего (переменная Y – измеряется в тоннах), мощности пласта (переменная X_1 – измеряется в метрах) и уровнем механизации работ в шахте (переменная X_2 – измеряется в процентах), характеризующие процесс добычи угля в 10 шахтах приведены в таблице.

Предполагая, что между переменными Y, X_1, X_2 *существует линейная зависимость*, необходимо найти аналитическое выражение для этой зависимости, т.е. построить уравнение линейной регрессии.

Таблица. Данные для построения

Номер шахты $\boldsymbol{i}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	$\boldsymbol{y}$
1	8	5	5
2	11	8	10
3	12	8	10
4	9	5	7
5	8	7	5
6	8	8	6
7	9	6	6
8	9	4	5
9	8	5	6
10	12	7	8

Используя пространственную выборку таблицы необходимо вычислить вектор коэффициентов $B=\begin{vmatrix} b_0 \\ b_1 \\ b_2 \end{vmatrix}$ уравнения регрессии

$$y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \varepsilon$$

Вектор коэффициентов, найденный методом наименьших квадратов является решением следующей системы уравнений:

$$X^T XB = X^T Y \ ,$$

где X - матрица размера 10×3, первый столбец которой составлен из 1, а другие два столбца составлены из значений x_{i1}, x_{i2} , т.е. матрица X имеет следующую структуру (символы … означают не отображенные элементы)

$$X = \begin{vmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ \cdots & \cdots & \cdots \\ 1 & 12 & 7 \end{vmatrix},$$

а Y - вектор, составленный из 10 значений y_i , т.е.

$$Y = \begin{vmatrix} 5 \\ 10 \\ \dots \\ 8 \end{vmatrix}$$

Матрица $X^T X$ имеет обратную матрицу $\left(X^T X\right)^{-1}$ и тогда вектор

коэффициентов вычисляется в виде: $B = \left(X^T X\right)^{-1} (X^T Y)$.

В итоге получен следующий результат:

$$B = \begin{vmatrix} -3.5393 \\ 0.8539 \\ 0.3670 \end{vmatrix}$$

Тогда уравнение регрессии примет вид: $\hat{y}(x_1, x_2) = -3.54 + 0.854x_1 + 0.367x_2$

<u>Исследование модели множественной линейной регрессии. Расчетные соотношения.</u>

Уравнение множественной линейной регрессии в стандартизованных переменных:

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

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

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

Частные коэффициенты корреляции.

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

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 $r_{x_1 x_2} = \frac{\mathbf{cov}(x_1, x_2)}{\sigma_{x_1} \cdot \sigma_{x_2}}$ – парный коэффициент корреляции.

Оценка статистической значимости модели в целом. F-критерий.

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

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.

Оценка статистической значимости параметров модели множественной регрессии. Фактическое значение t-критерия и доверительные интервалы.:

$$t_{b_i} = \frac{b_i}{m_{b_i}}, \quad (i = 1,2,...,m), \qquad t_a = \frac{a}{m_a}.$$

Здесь m_{b_i}, m_a — стандартные ошибки параметров уравнения регрессии.

Стандартные ошибки параметров уравнения множественной регрессии определяются соотношениями:

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

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.

$S^2_{ост} = \dfrac{\sum_{i=1}^{n}(y_i - \hat{y}_{x_i})^2}{n-m-1}$ — несмещенная оценка остаточной дисперсии.

Доверительные интервалы для параметров b_i уравнения линейной множественной регрессии указывают границы, в которых с заданной долей вероятности находятся значения соответствующих параметров и определяются соотношениями:

$$b_i - t(\alpha, n-m-1) \cdot m_{b_i} < b_i < b_i + t(\alpha, n-m-1) \cdot m_{b_i}$$

$$a - t(\alpha, n-m-1) \cdot m_a < a < a + t(\alpha, n-m-1) \cdot m_a$$

Оцениваемый параметр значим, если в границы доверительного интервала не попадает нуль.

Частные F- критерии:

$$F_{x_i} = \frac{R^2_{yx_1x_2...x_m} - R^2_{yx_1...x_{i-1}x_{i+1}...x_m}}{1 - R^2_{yx_1x_2...x_m}} \cdot \frac{n-m-1}{1}$$

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

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

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. На основании данных в таблицах соответствующих вариантам, построить линейную модель множественной регрессии (матричный метод). Записать уравнение регрессии.

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

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

4. С помощью F – критерия оценить статистическую надежность уравнения регрессии и коэффициента детерминации $R^2_{yx_1x_2}$ при уровнях значимости 0,05 и 0,01.

5. Проверить статистическую надежность параметров уравнения регрессии. (анализ *t*-статистики и интервальных оценок)

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

7. Записать уравнение парной линейной регрессии, оставив лишь один значащий фактор.

8. Оформить отчет по проделанной работе.

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

2.5 Лабораторная работа. Анализ случайных остатков в модели регрессии.

Цель: исследование случайных остатков регрессионной модели (выявление эффекта гетероскедастичности)

Форма проведения: выполнение индивидуального задания.

Форма отчетности: выполнение теста, защита отчета.

Теоретические основы

<u>Тест Парка</u>

Тест Парка основан на предположении, что дисперсия $\sigma_i^2 = \sigma^2(\varepsilon_i)$ является функцией i-го значения объясняющей переменной X. Парк предложил следующую зависимость.

$$\ln \varepsilon_i^2 = a + b \cdot \ln x_{ij} + v_i, \tag{4.1}$$

$$где \quad x_{ij} \; - \; i-е \; значение \; о-го \;\; фактора$$

$$v_i \; - случайный \;\; остаток$$

Выдвигаются гипотеза $H_0 : b = 0$, что соответствует гомоскедастичности остатков, и гипотеза $H_1 : b \neq 0$, которая выявляет наличие связи между $\ln \varepsilon_i^2$ и $\ln x_{ij}$. Отсюда следует, что гетероскедастичность остатков имеет место.

Условие принятия гипотезы H_1: $t_b > t_{\alpha, n-2}$

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

<u>Тест ранговой корреляции Спирмена</u>

Использование данного теста основано на предположении, **что дисперсия отклонений увеличиваются либо уменьшаются с ростом значения какого-либо фактора.** Поэтому для регрессии, построенной по МНК, абсолютные величины отклонений ε_i и значения x_i будут коррелированы.

Для проведения проверки по этому тесту выполняются следующие действия.

1. Ранжируются (упорядочиваются по величинам) значения модулей остатков ε_i и значения выбранного фактора x_i

2. Определяется коэффициент ранговой корреляции Спирмена:

$$r_{x,e} = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)},$$

где x — одна из объясняющих переменных,

d_i — разность между рангом *i*-го наблюдения x и рангом модуля остатка в *i*-м наблюдении. $i = 1, 2, \ldots, n$, n — число наблюдений.

3. Проверяется значимость вычисленного коэффициента ранговой корреляции.

Гипотеза $H_0 : r_{x,\varepsilon} = 0$ — гомоскедастичность остатков.

Гипотеза $H_1 : r_{x,\varepsilon} \neq 0$ — гетероскедастичность остатков.

Для проверки гипотезы H_0 рассчитывается фактическое значение t-критерия:

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

Если значение, рассчитанное по формуле, превышает табличное $t_{табл} = t_{\alpha, n-2}$, гипотеза H_0 о гомоскедастичности остатков отклоняется. В противном случае гипотеза о гомоскедастичности принимается.

Если в модели регрессии больше, чем одна объясняющая переменная, то проверка гипотезы может осуществляться с помощью t-статистики для каждой из них отдельно.

<u>Тест Голдфельда — Квандта</u>

Тест Голдфельда-Квандта применяется если случайные остатки предполагаются нормально распределенными случайными величинами и стандартное отклонение $\sigma_i = \sigma(\varepsilon_i)$ пропорционально значению x_i переменной X в этом наблюдении, т.е. $\sigma_i^2 = \sigma^2 x_i^2$, $i = 1,2,\ldots,n$.

Процедура проверки состоит в следующем:

1. Все наблюдения упорядочиваются по возрастанию фактора X.
2. Упорядоченная совокупность разбивается на три группы размерностей $k,\ (n-2\cdot k),\ k$ соответственно. Причем k должно быть больше чем число параметров модели. На практике рассчитывается k из соотношения $\frac{k}{n} = \frac{4}{15}$.
3. Оцениваются отдельные регрессии для первой группы (k первых наблюдений) и для третьей группы (k последних наблюдений). Если предположение о пропорциональности дисперсий отклонений значениям фактора X верно, то дисперсия регрессии по первой группы (рассчитываемая как $S_1 = \sum_{i=1}^{k} \varepsilon_i^2$) будет существенно меньше дисперсии регрессии по третьей группе (рассчитываемой как $S_3 = \sum_{i=n-k+1}^{n} \varepsilon_i^2$).
4. Формулируются:

Основная гипотеза, предполагающая постоянство дисперсий случайных ошибок модели регрессии, т. е. присутствие в модели условия гомоскедастичности: $H_0 : S_1 = S_3$.

Альтернативная гипотеза, предполагающая непостоянство дисперсий случайных ошибок в различных наблюдениях, т. е. присутствие в модели условия гетероскедастичности: $H_1 : S_1 \neq S_3$

5. Для сравнения соответствующих дисперсий вычисляется фактическое значение F-критерия:

$$F_{факт} = \frac{S_3/(k-m-1)}{S_1/(k-m-1)} = \frac{S_3}{S_1}.$$

Здесь $(k-m-1)$ — число степеней свободы соответствующих выборочных дисперсий (m — количество объясняющих переменных в уравнении регрессии).

Если $F_{факт}=\dfrac{S_3}{S_1}>F_{табл}$ (где $F_{табл}=F_{\alpha,k_1,k_2}$, α – выбранный уровень значимости), то гипотеза H_0 об отсутствии гетероскедастичности отклоняется.

Этот же тест может использоваться при предположении об обратной пропорциональности между σ_i и значениями объясняющей переменной. При этом статистика Фишера имеет вид:

$$F=\frac{S_1}{S_3}.$$

Для множественной регрессии данный тест обычно проводится для той объясняющей переменной, которая в наибольшей степени связана с σ_i. При этом k должно быть больше, чем $(m+1)$. Если нет уверенности относительно выбора переменной X_j, то данный тест может осуществляться для каждой из объясняющих переменных.

Индивидуальное задание

1. Построить линейное уравнение множественной регрессии со всеми факторами, оценить его значимость.
2. Построить матрицу парных коэффициентов корреляции. Установить, какие факторы (среди всех) коллинеарны, исключить зависимые факторы, если подтверждается взаимозависимость факторов модели.
3. Построить уравнение множественной регрессии со статистически значимыми факторами
4. Оценить статистическую значимость полученного уравнения множественной регрессии.
5. Вычислить остатки и построить графики остатков. Сделайть выводы.
6. Провести тестирование ошибок уравнения множественной регрессии на гетероскедастичность.
7. Оформить отчет

Варианты заданий:

	тест Голдфельда-Квандта	тест Спирмена	тест Парка
Таблица 1	1	2	3
Таблица 2	4	5	6
Таблица 3	7	8	9
Таблица 4	10	11	12

Таблица 1

Номер наблюдения	Цена x_1 (р.)	Цена на первый товар x_2 (р.)	Цена на второй товар x_3 (р.)	Средний доход населения x_4 (т. р.)	Спрос y (тыс.шт.)
1	15,09р.	24,30р.	12,85р.	5,09	125,1779
2	15,21р.	26,65р.	12,26р.	5,03	123,8094
3	15,28р.	25,22р.	13,42р.	4,80	121,175
4	15,49р.	26,59р.	12,05р.	4,95	116,9143
5	15,54р.	26,88р.	12,70р.	4,88	119,8643
6	15,62р.	24,74р.	12,41р.	4,96	118,0681
7	15,70р.	24,42р.	13,83р.	5,10	123,5887
8	15,91р.	25,79р.	13,10р.	4,90	117,0877
9	15,92р.	24,14р.	13,07р.	4,72	116,1699
10	15,95р.	26,70р.	12,40р.	4,81	118,3436
11	16,31р.	24,66р.	12,82р.	4,95	116,2008
12	16,33р.	24,04р.	12,48р.	4,88	111,4565
13	16,60р.	25,15р.	13,20р.	5,02	115,1026
14	16,69р.	24,10р.	12,40р.	4,80	110,1056
15	16,76р.	24,49р.	12,01р.	4,85	110,0231

Таблица 2

Номер наблюдения	Цена x_1 (р.)	Цена на первый товар x_2 (р.)	Цена на второй товар x_3 (р.)	Средний доход населения x_4 (т. р.)	Спрос y (тыс.шт.)
1	93,82р.	75,98р.	124,14р.	5,093176	135,5503
2	91,01р.	74,26р.	130,59р.	5,031956	171,1037
3	95,96р.	75,51р.	119,98р.	4,79949	106,8294
4	98,99р.	72,00р.	117,02р.	4,95135	82,65364
5	98,85р.	71,31р.	124,77р.	4,877776	107,9813
6	99,58р.	73,48р.	119,30р.	4,961544	87,08134
7	90,14р.	72,40р.	119,50р.	5,098437	140,1648
8	94,07р.	77,10р.	116,61р.	4,899322	107,1392
9	98,63р.	70,25р.	121,64р.	4,722132	96,99346
10	91,39р.	79,16р.	133,30р.	4,810111	175,2817
11	92,45р.	72,07р.	121,98р.	4,948579	134,0831
12	90,45р.	75,34р.	112,02р.	4,884198	111,3925
13	90,32р.	72,00р.	116,68р.	5,01854	128,979
14	91,64р.	70,07р.	118,36р.	4,801321	124,8081
15	92,20р.	74,12р.	121,55р.	4,848674	132,8335

Таблица 3

Номер наблюдения	Цена x_1 (р.)	Цена на первый товар x_2 (р.)	Цена на второй товар x_3 (р.)	Средний доход населения x_4 (т. р.)	Спрос y (тыс.шт.)
1	31,91р.	22,99р.	43,98р.	5,09р.	74,73
2	30,50р.	22,13р.	44,91р.	5,03р.	77,73
3	32,98р.	22,76р.	42,49р.	4,80р.	66,26
4	34,50р.	21,00р.	41,38р.	4,95р.	55,45

5	34,42р.	20,65р.	44,15р.	4,88р.	51,00
6	34,79р.	21,74р.	42,22р.	4,96р.	53,51
7	30,07р.	21,20р.	42,30р.	5,10р.	84,14
8	32,04р.	23,55р.	41,24р.	4,90р.	76,03
9	34,32р.	20,13р.	43,14р.	4,72р.	54,02
10	30,69р.	24,58р.	44,98р.	4,81р.	86,40
11	31,23р.	21,03р.	43,27р.	4,95р.	73,19
12	30,23р.	22,67р.	40,28р.	4,88р.	82,31
13	30,16р.	21,00р.	41,27р.	5,02р.	77,21
14	30,82р.	20,04р.	41,86р.	4,80р.	67,53
15	31,10р.	22,06р.	43,11р.	4,85р.	74,15

Таблица 4

Номер наблюдения	Цена x_1 (р.)	Цена на первый товар x_2 (р.)	Цена на второй товар x_3 (р.)	Средний доход населения x_4 (т. р.)	Спрос y (тыс.шт.)
1	43,22р.	75,98р.	30,90р.	5,09	166,95
2	44,44р.	60,87р.	38,75р.	5,03	114,21
3	43,87р.	69,06р.	36,93р.	4,80	141,13
4	44,81р.	69,51р.	33,18р.	4,95	139,50
5	41,19р.	62,12р.	34,06р.	4,88	131,81
6	41,28р.	65,40р.	46,54р.	4,96	142,16
7	40,48р.	73,34р.	41,61р.	5,10	178,34
8	40,63р.	62,76р.	35,35р.	4,90	140,91
9	40,95р.	64,90р.	37,86р.	4,72	145,70
10	43,00р.	66,99р.	47,44р.	4,81	142,50
11	44,27р.	66,19р.	32,51р.	4,95	133,54
12	41,37р.	69,24р.	30,30р.	4,88	156,54
13	40,41р.	76,86р.	44,63р.	5,02	183,50

14	40,53р.	77,57р.	48,57р.	4,80	182,00
15	41,13р.	73,05р.	35,98р.	4,85	167,27

2.6 Взвешенный метод наименьших квадратов

Цель: применение взвешенного метода наименьших квадратов с целью смягчения проблемы гетероскедастичности

Форма проведения: выполнение индивидуального задания.

Форма отчетности: выполнение теста, защита отчета.

Теоретические основы

Взвешенный метод наименьших квадратов (ВМНК)

ВМНК используется для корректировки гетероскедастичности, за счет преобразования данных, позволяющего получать оценки, которые обладают не только свойством несмещенности, но и имеют меньшие выборочные дисперсии.

Пусть σ_{ε_i} — стандартное отклонение случайной ошибки ε_i в i–м наблюдении. В случае если σ_{ε_i} известно, гетероскедастичность можно корректировать, разделив каждое наблюдение на соответствующее ему значение σ_{ε_i}. Так для парной регрессии $y_i = a + b \cdot x_i + \varepsilon_i$ соответствующее преобразование данных будет иметь вид:

$$\frac{y_i}{\sigma_{\varepsilon_i}} = \frac{a}{\sigma_{\varepsilon_i}} + b \cdot \frac{x_i}{\sigma_{\varepsilon_i}} + \frac{\varepsilon_i}{\sigma_{\varepsilon_i}} \cdot$$

Тогда дисперсия остатков представляется в виде:

$$D\left(\frac{\varepsilon_i}{\sigma_{\varepsilon_i}}\right) = \frac{1}{\sigma^2_{\varepsilon_i}} \cdot \sigma^2(\varepsilon_i) = \frac{\sigma^2_{\varepsilon_i}}{\sigma^2_{\varepsilon_i}} = 1$$

В результате этой процедуры каждое наблюдение будет иметь случайную ошибку с единичной дисперсией. Следовательно, для преобразованной модели выполняется предпосылка МНК о гомоскедастичности дисперсии остатков, а оценки параметров регрессии, полученные по МНК будут наилучшими несмещенными оценками.

Рассмотрим практические решения этой проблемы (2 случая).

1. Применение ВМНК основано на преобразовании данных, в котором что фактические значения σ_{ε_i} заменяются оценкой $\hat{\sigma}_{\varepsilon}$

Алгоритм вычисления оценки $\hat{\sigma}_{\varepsilon}$ рассмотрим на примере:

№	x_1	x_2	y	**№**	x_1	x_2	y
1.	13	43	79	11.	58	161	207
2.	28	56	110	12.		108	152
3.	33	24	97	13.	69	86	199
4.	42	98	171	14.	8	143	144
5.	12	176	204	15.	60	42	140
6.	44	124	174	16.	11	199	183
7.	36	130	184	17.	26	145	178
8.	33	291	311	18.	61	115	185
9.	34	141	206	19.	18	111	152
10.	21	95	128	20.	30	192	204

По таблице исходных данных получена модель $\hat{Y} = 36.78 + 1.19 \cdot x_1 + 0.76 \cdot x_2$.

Упорядочим модули остатков данной модели и выделим группы, содержащие остатки, близкие по значениям модулей.

Остатки	*модуль*
-0,02531	0,025315
-0,22768	0,227681
2,639116	2,639116
-2,73419	2,734187
5,462125	5,462125
5,684687	5,684687
-5,97146	5,971459
-6,04654	6,046542
9,362728	9,362728
-9,51027	9,510271
9,643585	9,643585

-11,0514	11,05136
-11,9281	11,92808
13,61468	13,61468
-14,5311	14,53107
14,58852	14,58852
-18,2088	18,20879
19,08836	19,08836
-21,3305	21,33054
21,48151	21,48151

По каждой группе вычисляем оценку среднеквадратичного отклонения:

1 группа

-0,02531	0,000641
-0,22768	0,051839
2,639116	6,964934
-2,73419	7,475778
Дисперсия	14,49319/4
$\hat{\sigma}_{\varepsilon_1}$	**1,903496**

2 группа

5,462125	29,8348
5,684687	32,31567
-5,97146	35,65833
-6,04654	36,56067
9,362728	87,66067
-9,51027	90,44526
9,643585	92,99873
Дисперсия	405,4741/7
$\hat{\sigma}_{\varepsilon_2}$	**7,610839**

3 группа

-11,0514	122,1326
-11,9281	142,2791
13,61468	185,3594
-14,5311	211,152
14,58852	212,825
-18,2088	331,5602
19,08836	364,3655
-21,3305	454,9921
21,48151	461,4552
Дисперсия	2486,121/9
$\hat{\sigma}_{\varepsilon_3}$	**16,62034**

Преобразуем исходные данные, разделив на соответствующий коэффициент пропорциональности $\hat{\sigma}_{\varepsilon_j}$, j=1,2,3 и стром модель по преобразованным данным.

x_1/σ	x_2/σ	y/σ	x_1/σ	x_2/σ	y/σ
6,829538	22,59001	41,50258	3,4897	9,686926	12,45462
14,70977	29,41955	57,7884	12,08303	56,7377	79,85306
17,33652	12,60838	50,95886	4,15154	5,174383	11,97328
5,518445	12,87637	22,46796	0,481338	8,603916	8,664083
0,722007	10,58944	12,27412	31,52095	22,06466	73,54887
5,781228	16,29255	22,86213	0,66184	11,97328	11,01061
18,91257	68,29538	96,66423	13,65908	76,17562	93,51214
1,985519	17,50867	18,71201	3,670202	6,919233	11,13094
2,045686	8,483582	12,39445	2,365048	14,58446	19,97152
11,03233	49,90816	67,24468	1,805017	11,55211	12,27412

2. ***Применение ВМНК основано на предположении, что среднее значение остаточных величин равно нулю, а вот дисперсия их представлена в виде произведения некоторой величины K_i либо K_i^2 на постоянную величину σ^2:***

$$\sigma_{\varepsilon_i}^2 = \sigma^2 \cdot K_i \qquad \text{либо} \qquad \sigma_{\varepsilon_i}^2 = \sigma^2 \cdot K_i^2$$

В эконометрических исследованиях довольно часто выдвигается гипотеза, что остатки ε_i пропорциональны значениям какого-либо фактора. Так, если в уравнении

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

предположить, что $K_i = x_{1_i}$ и $\sigma_{\varepsilon_i}^2 = \sigma^2 \cdot x_{1_i}^2$, то ВМНК предполагает оценку параметров следующего трансформированного уравнения:

$$\frac{y_i}{x_{1_i}} = a \cdot \frac{1}{x_{1_i}} + b_1 + b_2 \cdot \frac{x_{2_i}}{x_{1_i}} + \ldots + b_m \cdot \frac{x_{m_i}}{x_{1_i}} + \frac{\varepsilon_i}{x_{1_i}}.$$

Применение в этом случае ВМНК приводит к тому, что наблюдения с меньшими значениями преобразованных переменных $\frac{x}{K}$ имеют при определении параметров регрессии относительно больший вес, чем с первоначальными переменными. Вместе с тем, следует иметь в виду, что новые преобразованные переменные получают новое экономическое содержание и их регрессия имеет иной смысл, чем регрессия по исходным данным.

Индивидуальные задания

1. Используя исходные данные предыдущей работы найти значения параметров регрессии, используя взвешенный метод наименьших квадратов, если

a) коэффициент пропорциональности K равен значению регрессора x_j

b) коэффициент пропорциональности K равен оценке $\hat{\sigma}_{\varepsilon}$

2. Тестом Гольдфелда-Квандта исследовать вновь полученную регрессионную модель.

3. Оформить отчет.

2.7 Модели регрессии с фиктивными переменными.

Цель работы: принятие решения о необходимости включения качественного фактора в модель, построение и анализ регрессионной модели с фиктивными переменными.

Форма проведения: выполнение индивидуального задания.

Форма отчетности: выполнение теста, защита отчета.

Теоретические основы

<u>Спецификация модели регрессии с фиктивными переменными</u>

1. Модели регрессии с фиктивными переменными сдвига

Общий вид модели:

$y = a + bx + \delta \cdot z + \varepsilon$.

y — зависимая переменная;

x — независимая количественная переменная;

z — рассматривается как бинарная переменная, принимающая всего два значения: 1 и 0, которые определяют два состояния качественной переменной, например:

$$z = \begin{cases} \mathbf{1} & \text{– **мужской пол,**} \\ \mathbf{0} & \text{– **женский пол.**} \end{cases}$$

2. Модели регрессии с фиктивными переменными наклона.

Модели регрессии с фиктивными переменными наклона отражают зависимость количественного фактора от значения фиктивной переменной.

Общий вид модели: $\hat{y} = a + b \cdot x + \delta \cdot x \cdot z$

В таком случае говорят, что имеет место структурные изменения в исследуемой зависимости. Для их учета в уравнение регрессии вводят фиктивную переменную z как сомножитель при количественной переменной.

<u>Исследование структурных изменений с помощью теста Чоу.</u>

Тест Чоу в случае регрессионного анализа с фиктивными переменными позволяет решить задачу о необходимости введения в уравнение регрессии соответствующей фиктивной переменной.

Алгоритм теста

Пусть имеется две подвыборки: одна объемом n_1, а другая объемом n_2 (в соответствие значению качественного фактора).

1. По каждой подвыборке строятся линейные регрессионные модели с m переменными:

$$y_1 = b_{1_0} + \sum_{j=1}^{m} b_{1_j} \cdot x_j + \varepsilon_1 \text{, для первой подвыборки,}$$

$$y_2 = b_{2_0} + \sum_{j=1}^{m} b_{2_j} \cdot x_j + \varepsilon_2 \text{, для второй подвыборки.}$$

Рассчитываются суммы квадратов остатков для этих регрессий SS_1 и SS_2.

2. Строится линейная регрессия по объединенной выборке:

$$y = b_0 + \sum_{j=1}^{m} b_j \cdot x_j + \varepsilon \; .$$

Вычисляется ее сумма квадратов остатков SS.

3. Формулируется нулевая гипотеза:

$H_0 : b_{1_j} = b_{2_j}, \; j = \overline{0, m},$ где b_{1_j}, b_{2_j} — параметры моделей.

Очевидно, что при совпадении параметров регрессии выполняется равенство $SS = SS_1 + SS_2$. Чем сильнее различие в поведении y для двух подвыборок, тем больше значение SS будет превосходить значение суммы $SS_1 + SS_2$.

4. Для проверки гипотезы вычисляется фактическое значение F -статистики по формуле:

$$F = \frac{(SS - (SS_1 + SS_2)}{(SS_1 + SS_2)} \cdot \frac{n - 2 \cdot m - 2}{m + 1}.$$

Здесь m — количество параметров уравнений регрессий, n — число наблюдений по всей совокупности.

В случае, если $F < F_{табл}(\alpha,(m+1),(n-2 \cdot m-2))$, то считается, что различие между SS и $SS_1 + SS_2$ статистически незначимо и возможно построение уравнение регрессии по объединенной выборке объема $n = n_1 + n_2$ без учета качественного фактора.

Если, $F > F_{табл}(\alpha,(m+1),(n-2 \cdot m-2))$ то различие между SS и $SS_1 + SS_2$ статистически значимо, что определяет и существенность различия поведения наблюдаемой переменной y для двух подвыборок.

<u>Пример</u>

Используем тест Чоу для выявления целесообразности рассмотрения общей выборки и введения фиктивной переменной на примере данных таблиц 1, 2.

Таблица 1 Изменение заработной платы мужчин в зависимости от стажа

N набл.	1	2	3	4	5	6	7	8	9
y_1	9	10	12	13	15	18	20	22	25
x	2	3	4	5	7	8	10	12	15

Таблица 2 Изменение заработной платы женщин в зависимости от стажа.

N набл.	1	2	3	4	5	6	7	8	9
y_2	6	7	8,5	9	9	9,5	12	15	16
x	2	3	4	5	7	8	10	12	15

Построим по каждой из них линейные модели зависимости заработной платы (y) от стажа (X):

$$\hat{y}_1 = 6{,}689 + 1{,}269 \cdot x\,; \quad R^2 = 0{,}99; \quad SS_1 = 2{,}94\,;$$

$$\hat{y}_2 = 4{,}61 + 0{,}765 \cdot x\,; \quad R^2 = 0{,}95; \quad SS_2 = 5{,}019\,.$$

Построим линейную модель по объединенной выборке:

$$\hat{y} = 5{,}649 + 1{,}018 \cdot x\,; \quad R^2 = 0{,}639; \quad SS = 177{,}52\,.$$

Рассчитаем статистику

$$F = \frac{(177{,}52 - 2{,}94 - 5{,}019) \cdot 14}{(2{,}94 + 5{,}019) \cdot 2} = 80{,}4\,.$$

$F_{табл}(\alpha = 0{,}05;\ 2,\ 14) = 3{,}74$. Так как вычисленное значение $F > F_{табл}$, то следует признать существенность различия роста заработной платы от стажа в зависимости от пола. Следовательно, для построения общего уравнения регрессии целесообразно ввести фиктивную переменную, определяющую пол работника.

Индивидуальные задания.

Исследовать, насколько большим или меньшим спросом(Y, тыс. шт) пользуется новый товар в зависимости от — цены товара (X , руб) и вида (новый или старый).

1. Для данных своего варианта подобрать наилучшее воздействие фиктивной переменной (влияние на наклон или сдвиг, для

этого построить две модели: с фиктивной переменной сдвига, с фиктивной переменной наклона)

2. Сделать сравнительный анализ моделей. Определить, насколько изменяется спрос при переходе от одной категории к другой (новый либо старый).

3. Провести исследование структурных изменений с помощью теста Чоу. Сделать выводы о целесообразности построения модели переменной структуры.

4. Оформить отчет.

Структура отчета:

- Постановка задачи
- Таблица исходных данных
- Результаты построения моделей с выводами. (корреляции, детеминация, статистика Фишера, статистика Стьюдента)
- Результаты проведения теста Чоу.

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

Вариант 2, 6		
Вид	Цена x (р.)	Спрос y (тыс.шт.)
новый	93,82	135,5503
новый	91,01	171,1037
старый	95,96	106,8294
старый	98,99	82,65364
новый	98,85	107,9813
новый	99,58	87,08134
старый	90,14	140,1648
старый	94,07	107,1392
новый	98,63	96,99346
новый	91,39	175,2817
старый	92,45	134,0831
старый	90,45	111,3925
новый	90,32	128,979
старый	91,64	124,8081
старый	92,20	132,8335

Вариант 3, 7			
Номер наблюдения	Вид	Цена x (р.)	Спрос y (тыс.шт.)
1	новый	31,91	74,73
2	новый	30,50	77,73
3	старый	32,98	66,26
4	старый	34,50	55,45
5	новый	34,42	51,00
6	новый	34,79	53,51
7	старый	30,07	84,14
8	старый	32,04	76,03
9	новый	34,32	54,02
10	новый	30,69	86,40
11	старый	31,23	73,19
12	старый	30,23	82,31
13	новый	30,16	77,21
14	старый	30,82	67,53
15	старый	31,10	74,15

Вариант 4, 8		
Вид	Цена x (р.)	Спрос y (тыс.шт.)
новый	43,22	166,95
новый	44,44	114,21
старый	43,87	141,13
старый	44,81	139,50
новый	41,19	131,81
новый	41,28	142,16
старый	40,48	178,34
старый	40,63	140,91
новый	40,95	145,70
новый	43,00	142,50
старый	44,27	133,54
старый	41,37	156,54
новый	40,41	183,50
старый	40,53	182,00
старый	41,13	167,27

Вариант 5, 9			
Номер наблюдения	Вид	Цена x (р.)	Спрос y (тыс.шт.)
1	новый	38,37	422,9505
2	новый	31,30	436,158
3	старый	33,97	534,9022
4	старый	49,18	295,2635
5	новый	41,26	251,038
6	новый	39,58	339,7103
7	старый	47,66	371,124
8	старый	44,20	234,9415
9	новый	32,62	543,2922

Вариант 6, 10		
Вид	Цена x (р.)	Спрос y (тыс.шт.)
новый	25,73	132,775
новый	21,51	170,105
старый	28,95	100,926
старый	33,49	60,666
новый	33,27	60,598
новый	34,38	49,481
старый	20,22	188,878
старый	26,11	128,269
новый	32,95	69,717

10	новый	42,10	397,1158
11	старый	32,90	494,8871
12	старый	42,54	458,4898
13	новый	38,34	283,1757
14	старый	34,20	439,8954
15	старый	34,42	478,3711

новый	22,08	170,243
старый	23,68	153,002
старый	20,68	175,995
новый	20,49	177,151
старый	22,46	156,506
старый	23,29	152,925

2.8 Системы одновременных уравнений (Структурная форма модели)

Цель работы: приобретение навыков исследования структурной формы модели на разрешимость (идентификация модели) и оценивании параметров модели с помощью двухшагового метода наименьших квадратов

Форма проведения: выполнение индивидуального задания.

Форма отчетности: выполнение теста, защита отчета.

Теоретические основы

Идентификация модели

Процедуру идентификация рассмотрим на примере.

Изучается модель вида

$$\begin{cases} КП_t = a_1 + b_{11}ВВП_t + \varepsilon_1 \\ ВН_t = a_2 + b_{21}ВВП_{t-4} + \varepsilon_2 \\ ВВП_t = КП_t + ВН_t + Э_t \end{cases}$$

где $ВВП_{t-4}$ — объем $ВВП$ за аналогичный квартал предыдущего года.

Выполним проверку системы на идентификацию с помощью необходимого условия.

Для первого уравнения имеем:

- количество эндогенных переменных, входящих в это уравнение, равно двум ($КП_t$ и $ВВП_t$ t), $H = 2$;
- количество предопределенных переменных, не входящих в это уравнение, равно двум ($Э_t$ и $ВВП_{t-4}$), $D = 2$.

$H < D + 1$, следовательно, уравнение сверхидентифицировано.

Для второго уравнения имеем:

- количество эндогенных переменных, входящих в это уравнение, – одна ($ВН_t$), $H = 1$;
- количество предопределенных переменных, не входящих в это уравнение, – одна ($Э_t$), $D = 1$

$H < D+1$, следовательно, уравнение сверхидентифицировано.

Третье уравнение является тождеством и на идентификацию не проверяется.

Поскольку оба уравнений системы сверхидентифицированы, то и система сверхидентифицирована.

Выполним проверку системы на идентификацию с помощью достаточного условия.

Определим матрицу коэффициентов при переменных, отсутствующих в первом уравнении ($ВН_t$, $ВВП_{t-4}$, $Э_t$):

$$\begin{pmatrix} -1 & b_{21} & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Определитель данной матрицы не равен нулю, ранг матрицы равен двум, что на единицу меньше количества эндогенных переменных в системе. Таким образом, достаточное условие соблюдается.

Определим матрицу коэффициентов при переменных, отсутствующих во втором уравнении ($КП_t$, $ВВП_t$, $Э_t$):

$$\begin{pmatrix} -1 & b_{11} & 0 \\ 1 & -1 & 1 \end{pmatrix}.$$

Ранг данной матрицы равен 2, следовательно, достаточное условие соблюдается.

<u>Двухшаговый метод наименьших квадратов.</u>

.

Применение метода предусматривает следующие шаги:

1. Построение приведенной формы модели (ПФМ).
2. Для каждого уравнения структурной формы модели выполняются следующие действия:

- находят эндогенные переменные, являющиеся факторными признаками (стоят в правой части уравнения);

- для этих переменных определяют их выровненные значения $\hat{y}_i = A_i + \delta_{i1}x_1 + \delta_{i2}x_2 + ... + \delta_{im}x_m$, используя соответствующее уравнение ПФМ;
- находят параметры рассматриваемого уравнения структурной формы модели обычным МНК, заменяя исходные значения эндогенных переменных-факторов их выровненными значениями.

<u>Задание.</u>

1. Указать эндогенные и экзогенные переменные в системе.
2. Определить тип каждого уравнения системы.
3. Оценить параметры системы.
4. Оформить отчет

Варианты задания 1-2

$$\begin{cases} Y_1 = a_{13} \cdot Y_3 + \varepsilon_1 \\ Y_2 = a_{21} \cdot Y_1 + b_{21} \cdot X_1 + \varepsilon_2 \\ Y_3 = a_{32} \cdot Y_2 + b_{32} \cdot X_2 + \varepsilon_3 \end{cases}.$$

Вариант 1						Вариант 2					
n	Y1	Y2	Y3	X1	X2	n	Y1	Y2	Y3	X1	X2
1	1,70	2,10	3,70	0,71	1,10	1	3,59	4,11	5,10	0,61	1,48
2	2,40	2,00	4,50	0,69	1,40	2	3,40	3,78	4,50	0,69	1,40
3	3,00	2,50	4,00	0,54	1,30	3	3,10	2,50	4,90	0,54	1,71
4	2,60	3,10	4,20	0,50	1,35	4	2,64	3,12	4,22	0,51	1,35
5	4,30	2,70	3,90	0,64	1,42	5	3,95	3,70	3,90	0,64	1,46
6	2,70	3,20	3,80	0,57	1,47	6	2,73	3,33	4,17	0,72	1,47
7	1,90	2,60	4,60	0,59	1,32	7	3,49	2,60	4,60	0,59	1,52
8	2,30	3,00	5,00	0,62	1,38	8	2,81	3,00	5,12	0,64	1,38
9	4,10	2,80	4,90	0,67	1,41	9	4,13	2,80	4,85	0,67	1,72
10	3,30	1,90	3,60	0,60	1,43	10	3,30	2,96	3,60	0,60	1,43

Варианты задания 3-4

$$\begin{cases} Y_1 = a_{12} \cdot Y_2 + b_{12} \cdot X_2 + \varepsilon_1 \\ Y_2 = a_{23} \cdot Y_3 + b_{21} \cdot X_1 + \varepsilon_2 \\ Y_3 = a_{31} \cdot Y_1 + \varepsilon_3 \end{cases}$$

Вариант 3						Вариант 4					
n	Y_1	Y_2	Y_3	X_1	X_2	n	Y_1	Y_2	Y_3	X_1	X_2
1	11,63	3,11	5,00	0,57	1,42	1	2,31	2,44	4,15	0,58	1,52
2	10,78	2,78	4,50	0,69	1,40	2	2,40	2,00	4,50	0,69	1,40
3	11,09	2,50	4,00	0,54	1,49	3	3,00	2,50	4,00	0,54	1,30
4	10,67	3,12	4,22	0,51	1,35	4	2,60	3,12	4,20	0,50	1,35
5	11,83	2,70	3,90	0,64	1,46	5	3,95	2,70	3,90	0,64	1,42
6	10,88	3,33	3,80	0,72	1,47	6	2,70	3,20	3,80	0,57	1,47
7	11,42	2,60	4,60	0,59	1,52	7	2,49	2,60	4,60	0,59	1,32
8	10,77	3,00	5,12	0,64	1,38	8	2,30	3,00	5,00	0,64	1,38
9	11,46	2,80	4,85	0,67	1,72	9	4,10	2,80	4,85	0,67	1,41
10	10,92	1,98	3,60	0,60	1,57	10	3,30	1,98	3,60	0,60	1,43

Варианты задания 5-6

$$\begin{cases} Y_1 = a_{12} \cdot Y_2 + \varepsilon_1 \\ Y_2 = a_{23} \cdot Y_3 + b_{22} \cdot X_2 + \varepsilon_2 \\ Y_3 = a_{31} \cdot Y_1 + b_{31} \cdot X_1 + \varepsilon_3 \end{cases}$$

Вариант 5						Вариант 6					
n	Y_1	Y_2	Y_3	X_1	X_2	n	Y_1	Y_2	Y_3	X_1	X_2
1	2,31	2,44	4,15	0,58	1,52	1	3,31	2,54	4,37	0,63	1,54
2	2,40	2,00	4,50	0,69	1,40	2	2,40	2,17	4,50	0,69	1,40
3	3,00	2,50	4,00	0,54	1,30	3	3,00	2,50	4,00	0,54	1,30
4	2,60	3,12	4,20	0,50	1,35	4	2,60	3,12	4,20	0,50	1,35
5	3,95	2,70	3,90	0,64	1,42	5	3,95	2,70	3,90	0,64	1,42
6	2,70	3,20	3,80	0,57	1,47	6	2,70	3,20	3,80	0,57	1,47
7	2,49	2,60	4,60	0,59	1,32	7	2,49	2,60	4,60	0,59	1,52
8	2,30	3,00	5,00	0,64	1,38	8	2,74	3,00	5,12	0,64	1,38
9	4,10	2,80	4,85	0,67	1,41	9	4,10	2,80	4,85	0,67	1,41
10	3,30	1,98	3,60	0,60	1,43	10	3,30	1,98	3,60	0,60	1,43

Варианты задания 7-8

$$\begin{cases} Y_1 = a_{13} \cdot Y_3 + \varepsilon_1 \\ Y_2 = a_{21} \cdot Y_1 + b_{22} \cdot X_2 + \varepsilon_2 \\ Y_3 = a_{32} \cdot Y_2 + b_{31} \cdot X_1 + \varepsilon_3 \end{cases}$$

Вариант 7						Вариант 8					
n	Y_1	Y_2	Y_3	X_1	X_2	n	Y_1	Y_2	Y_3	X_1	X_2
1	1,70	2,10	7,98	0,71	1,10	1	2,59	3,11	5,00	0,57	1,42

2	2,40	2,00	7,09	0,69	1,40	2	2,40	2,78	4,50	0,69	1,40
3	3,00	2,50	8,12	0,54	1,30	3	3,00	2,50	4,00	0,66	1,30
4	2,60	3,10	8,04	0,50	1,35	4	2,64	3,12	4,22	0,51	1,35
5	4,30	2,70	7,44	0,64	1,42	5	3,95	2,70	3,90	0,64	1,46
6	2,70	3,20	7,51	0,57	1,47	6	2,73	3,33	3,80	0,72	1,47
7	1,90	2,60	8,34	0,59	1,32	7	2,49	2,60	4,60	0,59	1,52
8	2,30	3,00	8,51	0,62	1,38	8	2,81	3,00	5,12	0,64	1,38
9	4,10	2,80	6,85	0,67	1,41	9	4,13	2,80	4,85	0,67	1,72
10	3,30	1,90	7,73	0,60	1,43	10	3,30	1,98	3,60	0,61	1,43

Варианты задания 9-10

$$\begin{cases} Y_1 = a_{13} \cdot Y_3 + b_{12} \cdot X_2 + \varepsilon_1 \\ Y_2 = a_{21} \cdot Y_1 + \varepsilon_2 \\ Y_3 = a_{32} \cdot Y_2 + b_{31} \cdot X_1 + \varepsilon_3 \end{cases}$$

Вариант 9						Вариант 10					
n	Y_1	Y_2	Y_3	X_1	X_2	n	Y_1	Y_2	Y_3	X_1	X_2
1	3,59	4,11	5,10	0,61	1,48	1	9,57	3,11	4,99	0,57	1,42
2	3,40	3,78	4,50	0,69	1,40	2	10,07	2,78	4,50	0,69	1,53
3	3,10	2,50	4,90	0,54	1,71	3	9,93	2,50	4,17	0,54	1,30
4	2,64	3,12	4,22	0,51	1,35	4	10,13	3,12	4,22	0,51	1,35
5	3,95	3,70	3,90	0,64	1,46	5	9,84	2,70	3,90	0,64	1,46
6	2,73	3,33	4,17	0,72	1,47	6	10,66	3,33	3,80	0,72	1,47
7	3,49	2,60	4,60	0,59	1,52	7	9,69	2,60	4,60	0,59	1,52
8	2,81	3,00	5,12	0,64	1,38	8	10,37	3,00	5,12	0,64	1,38
9	4,13	2,80	4,85	0,67	1,72	9	9,88	2,80	4,85	0,67	1,77
10	3,30	2,96	3,60	0,60	1,43	10	10,36	1,98	4,01	0,60	1,72

3 Методические указания для организации самостоятельной работы

3.1 Общие положения

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

Самостоятельная работа студента по дисциплине «Эконометрика» включает следующие виды деятельности:

1) проработка лекционного материала;
2) подготовка к лабораторным работам;
3) самостоятельное изучение тем (вопросов) теоретической части курса.

В ходе самостоятельной работы студент, ориентируясь на изложенные рекомендации, планирует свое время и перечень необходимых работ в зависимости от индивидуальных психофизических особенностей. Формат самостоятельной работы студентов может отличаться в зависимости от формы обучения и объема аудиторной работы.

3.2 Проработка лекционного материала и подготовка к лабораторным работам

Для качественного усвоения учебного материала целесообразно осуществлять проработку лекционного материала, которая направлена как на систематизацию имеющегося материала, так и на подготовку к освоению практических аспектов, связанных с содержанием дисциплины.

Проработка лекционного материала включает деятельность, связанную с изучением рекомендуемых преподавателем источников, в которых отражены основные моменты, затрагиваемые в ходе лекций. Кроме того, важное место отведено работе с собственноручно составленным конспектом лекций. При конспектировании во время лекции помните, что не следует записывать все, что говорит и/или демонстрирует лектор: старайтесь выявить главное и записать только это. Цель конспекта – формирование целостного логически выстроенного взгляда на круг вопросов, затрагиваемых в ходе изучения соответствующей темы, а не механическая фиксация текстовой и графической информации.

Во внеаудиторное время проработка лекционного материала может быть выстроена в двух основных форматах:

а) отработка прослушанной лекции (прочтение конспекта и рекомендованных преподавателем источников с сопоставлением записей) и восполнение пробелов, если они имелись (например, если студент не понял чего-то, не успел записать);

б) прочтение перед каждой последующей лекцией предыдущей, дабы не тратилось много времени на восстановление контекста изучения дисциплины при продолжающейся или связанной теме.

В ходе проработки лекционного материала обращайте внимание на контрольные вопросы, которые, как правило, имеются в конце каждой темы учебника (учебного пособия). Отвечая на них, можно сделать вывод о степени понимания материала. Если ответы на какие-то вопросы вызвали затруднения, то следует предпринять еще одну попытку изучения отдельных вопросов.

При подготовке к лабораторным занятиям необходимо заранее изучить методические рекомендации по его проведению, обратить внимание на цель, формат и содержание занятия. Если какие-то моменты вызвали дополнительные вопросы, целесообразно обратиться к содержанию лекционного материала, рекомендациям преподавателя по изучению теоретической части курса (рекомендуемым источникам) или за личной консультацией. В ходе подготовки к лабораторным работам может потребоваться обращение к различным источникам. Проявляйте инициативу и самостоятельность в данном вопросе. При этом следует пользоваться только авторитетными изданиями, как печатными, так и электронными.

3.3 Самостоятельное изучение тем теоретической части курса

В ходе изучения дисциплины некоторые из тем курса выносятся исключительно на самостоятельное изучение. Следует обратить внимание на то, что работа по этим темам включает как подбор источников, так и изучение их содержания.

В зависимости от особенностей усвоения учебного материала студентами и объема аудиторной работы некоторые из вопросов, рассматриваемые в ходе проведения лекций и лабораторных работ, могут быть также вынесены в формат самостоятельного изучения.

Студент самостоятельно изучает дополнительные вопросы, связанные с построением и анализом моделей множественной регрессии и эконометрических моделей по временным рядам. Для достижения этой цели сформулированы следующие задания:

- Мультиколлинеарность.
- Автокорреляция. Обнаружение и методы устранения автокорреляции.
- Изучение взаимосвязей по временным рядам. Построение аддитивной и мультипликативной модели временного ряда.

Тема «Мультиколлинеарность»

Дать определениея мультиколлинеарности. Охарактеризовать причины возникновения мультиколлинеарности, последствия мультиколлинеарности. Изучить методы обнаружения и методы устранения мультиколлинеарности.

Тема «Автокорреляция. Обнаружение и методы устранения автокорреляции»

Рассмотреть суть и причины возникновения автокорреляции, последствия автокорреляции. Охарактеризовать виды автокорреляции, рассматриваемой в эконометрическом моделировании. Изучить методы обнаружения автокорреляции и методы устранения автокорреляции.

Тема «Изучение взаимосвязей по временным рядам. Построение аддитивной и мультипликативной модели временного ряда»

Охарактеризовать задачи, связанные с применением временных рядов в эконометрическом моделировании. Выделить и описать основные составляющие временных рядов. Изучить вопросы, связанные с построением аддитивной и мультипликативной моделей.

4 Рекомендуемые источники

1. Потахова, И. В. Эконометрика: Учебное пособие [Электронный ресурс] / И. В. Потахова. — Томск: ТУСУР, 2015. — 110 с. — Режим доступа: https://edu.tusur.ru/publications/5205
2. 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**)