Министерство образования Республики Беларусь

Учреждение образования
«Гомельский государственный технический
университет имени П. О. Сухого»

Кафедра «Экономика и управление в отраслях»

Н. В. Ермалинская

ЭКОНОМЕТРИКА
(ПРОДВИНУТЫЙ УРОВЕНЬ)

ПРАКТИКУМ
для магистрантов специальности 1-25 80 04
«Экономика и управление народным хозяйством»
дневной и заочной форм обучения

Электронный аналог печатного издания

Гомель 2017

УДК 519.862.6(075.8)
ББК 65в631я73
Е72

Рекомендовано к изданию научно-методическим советом гуманитарно-экономического факультета ГГТУ им. П. О. Сухого (*протокол № 10 от 27.06.2016 г.*)

Рецензенты: зав. каф. «Коммерческая деятельность и информационные технологии в экономике» Гомельского государственного университета имени Ф. Скорины канд. экон. наук, доц. *С. Н. Говейко*;
доц. каф. «Маркетинг и логистика» ГФ УО ФПБ «Международный университет МИТСО» канд. экон. наук, доц. *М. Н. Ковалев*

Ермалинская, Н. В.
Е72 Эконометрика (продвинутый уровень) : практикум для магистрантов специальности 1-25 80 04 «Экономика и управление народным хозяйством» днев. и заоч. форм обучения / Н. В. Ермалинская. – Гомель : ГГТУ им. П. О. Сухого, 2017. – 72 с. – Систем. требования: PC не ниже Intel Celeron 300 МГц ; 32 Mb RAM ; свободное место на HDD 16 Mb ; Windows 98 и выше ; Adobe Acrobat Reader. – Режим доступа: https://elib.gstu.by. – Загл. с титул. экрана.

ISBN 978-985-535-339-4.

Содержит методические рекомендации, алгоритмы и примеры проведения эконометрических расчетов (в том числе с применением функций табличного процессора MS Excel), а также варианты индивидуальных заданий для выполнения расчетных работ. Способствует выработке навыков использования эконометрических моделей и методов в научно-исследовательской и организационно-управленческой деятельности.
Для магистрантов специальности 1-25 80 04 «Экономика и управление народным хозяйством».

УДК 519.862.6(075.8)
ББК 65в631я73

ISBN 978-985-535-339-4

Оглавление

ПРЕДИСЛОВИЕ

Получение обоснованных и достоверных выводов о состоянии и тенденциях развития исследуемых социально-экономических процессов и явлений на основе анализа эмпирических наблюдений обусловливает научную значимость и практическую применимость предлагаемых исследователями решений. В такой ситуации получение знаний и развитие умений использования классического и современного эконометрического инструментария при проведении микро- и макроэкономического моделирования и прогнозирования является, бесспорно, актуальной задачей при формировании профессиональных компетенций в процессе подготовки магистров экономических наук.

Практикум по дисциплине «Эконометрика (продвинутый уровень» для магистрантов специальности 1-25 80 04 «Экономика и управление народным хозяйством» призван обеспечить закрепление теоретико-методических знаний и формирование практических навыков использования эконометрических моделей и методов изучения экономических явлений в процессе осуществления научно-исследовательской и организационно-управленческой деятельности.

В настоящем практикуме содержится краткий теоретический и основной методический материал по темам, представлены алгоритмы и примеры проведения эконометрических расчетов (в том числе с применением функций табличного процессора MS Excel), размещена справочная информация и рекомендуемая для более углубленного изучения литература. Для закрепления базовых знаний и выработки у магистрантов навыков их практического использования предложен комплекс индивидуальных расчетных заданий.

Разделы книги содержат описание алгоритмов построения и анализа качества эконометрических моделей (парных и множественных); методов выявления, смягчения и устранения нарушений модельных предположений (гетероскедастичности, автокорреляции, мультиколлинеарности), методов анализа временных рядов и пр.

Материал практикума базируется на работах таких авторов, как И. В. Белько, С. А. Бородич, Н. Ш. Кремер, Н. И. Шевченко и др.

Практикум может быть использован как в процессе аудиторной, так и самостоятельной работы магистрантов.

1. ПАРНЫЕ ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ: ПОСТРОЕНИЕ, ОЦЕНКА ЗНАЧИМОСТИ И АНАЛИЗ ПАРАМЕТРОВ УРАВНЕНИЯ

1.1. Построение уравнения регрессии

Парной регрессией называется уравнение связи двух переменных y и x вида [см. формулу (1.1)]:

$$y = f(x), \tag{1.1}$$

где y – зависимая переменная (результативный признак); x – независимая, объясняющая переменная (признак-фактор).

Построение уравнения регрессии осуществляется в два этапа: 1) спецификация модели (определение вида зависимости $\hat{y} = f(x)$; 2) оценка параметров (определение значений) выбранной модели.

Для выбора вида зависимости применяется три метода: графический (на основе анализа поля корреляции); аналитический (исходя из теории изучаемой взаимосвязи); экспериментальный (путем сравнения величины остаточной дисперсии $D_{\text{ост}}$ или средней ошибки аппроксимации $\overline{A}$, рассчитанных для разных моделей регрессии).

Для оценки параметров регрессии, линейных по этим параметрам, используется *метод наименьших квадратов* (МНК).

В случае *линейной регрессии* параметры a и b находятся из следующей системы нормальных уравнений (1.2) МНК:

$$\begin{cases} na + b\sum x = \sum y, \\ a\sum x + b\sum x^2. \end{cases} \Rightarrow \; a = \bar{y} - bx, \quad b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}. \tag{1.2}$$

Нелинейные уравнения регрессии предварительно приводятся к линейному виду путем преобразования переменных. Для нелинейных уравнений, приводимых к линейным, система нормальных уравнений имеет вид (1.3) в преобразованных переменных x', y'. Далее для нелинейных моделей проводятся линеаризующие преобразования.

Гиперболическая регрессия: $y_x = a_0 + a_1 / x$.

Линеаризующее преобразование: $x' = 1/x$; $y' = y$.

Формулы (1.2) принимают вид:

$$a_0 = \frac{1}{n}\sum y - \frac{1}{n} a_1 \sum \frac{1}{x}, \quad a_1 = \frac{n\sum \frac{y}{x} - \sum \frac{1}{x} \sum y}{n\sum \frac{1}{x^2} - \left(\sum \frac{1}{x}\right)^2}. \tag{1.3}$$

Экспоненциальная регрессия: $y_x = e^{a_0 + a_1 x}$.
Линеаризующее преобразование: $x' = x$; $y' = \ln y$.

$$a_0 = \frac{1}{n}\sum \ln y - \frac{1}{n} a_1 \sum x, \quad a_1 = \frac{n\sum x \ln y - \sum x \sum \ln y}{n\sum x^2 - \left(\sum x\right)^2}. \tag{1.4}$$

Степенная функция: $y_x = a_0 x^{a_1}$, $(a_0 > 0)$.
Линеаризующее преобразование: $x' = \ln x$; $y' = \ln y$.

$$\ln a_0 = \frac{1}{n}\sum \ln y - \frac{1}{n} a_1 \sum \ln x, \quad a_1 = \frac{n\sum (\ln x \ln y) - \sum \ln x \sum \ln y}{n\sum (\ln x)^2 - \left(\sum \ln x\right)^2}. \tag{1.5}$$

Показательная функция: $y_x = a_0 a_1^x$.
Линеаризующее преобразование: $x' = x$; $y' = \ln y$.

$$\ln a_0 = \frac{1}{n}\sum \ln y - \frac{1}{n} \ln a_1 \sum x, \quad \ln a_1 = \frac{n\sum x \ln y - \sum x \sum \ln y}{n\sum x^2 - \left(\sum x\right)^2}. \tag{1.6}$$

Логарифмическая функция: $y_x = a_0 + a_1 \ln x$.
Линеаризующее преобразование: $x' = \ln x$; $y' = y$.

$$a_0 = \frac{1}{n}\sum y - \frac{1}{n} a_1 \sum \ln x, \quad a_1 = \frac{n\sum \ln xy - \sum \ln x \sum y}{n\sum (\ln x)^2 - \left(\sum \ln x\right)^2}. \tag{1.7}$$

1.2. Оценка качества и точности построенной модели регрессии

Качество построенной модели регрессии оценивается с помощью линейного коэффициента корреляции r_{xy}^2 (в случае линейной зависимости); индекса корреляции R или коэффициента детерминации $R^2 = (R)^2$ (в случае нелинейной зависимости).

Линейный коэффициент корреляции r_{xy} определяется (1.8):

$$r_{xy} = \frac{\frac{1}{n}\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y} = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} \quad \text{или} \quad r_{xy} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y}, \qquad (1.8)$$

где n – количество наблюдений; x_i, y_i – данные наблюдений; $\bar{x}$, $\bar{y}$ – средние значения переменных x, y; σ_x, σ_y – средние квадратические отклонения переменных x, y.

$$\sigma_x = \sqrt{\frac{1}{n}\sum_i (x_i - \bar{x})^2} = \sqrt{\overline{x^2} - \bar{x}^2}, \quad \sigma_y = \sqrt{\frac{1}{\mathrm{n}}\sum_i (y_i - \bar{y})^2} = \sqrt{\overline{y^2} - \bar{y}^2}. \qquad (1.9)$$

Коэффициент r_{xy} принимает значение в диапазоне $-1 \le r_{xy} \le 1$.

При $r_{xy} > 0$ связь является прямой, при $r_{xy} < 0$ – обратной.

Чем ближе величина $|r_{xy}|$ к единице, тем теснее линейная связь и тем лучше линейная зависимость согласуется с данными наблюдений.

Градация степени тесноты связи приведена в табл. 1.1.

Таблица 1.1

Количественные критерии оценки тесноты связи

Коэффициент корреляции $|r_{xy}|$	**Характер связи**	**Коэффициент корреляции $|r_{xy}|$**	**Характер связи**
$|r_{xy}| < 0{,}3$	Отсутствует	$0{,}5 \le |r_{xy}| \le 0{,}7$	Умеренная
$0{,}3 < |r_{xy}| < 0{,}5$	Слабая	$0{,}7 \le |r_{xy}|$	Сильная

Тесноту нелинейной связи оценивают с помощью индекса корреляции R по следующей формуле (1.10):

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}}, \qquad (1.10)$$

где n – количество наблюдений; x_i, y_i – данные наблюдений; $\bar{x}$, $\bar{y}$ – средние значения переменных x, y; $\hat{y}_i$ – расчетные значения переменной y, вычисленные по уравнению $\hat{y}_i = f(x_i)$.

Коэффициент детерминации R^2 принимает значение в диапазоне $0 \le R^2 \le 1$. Чем ближе величина R^2 к единице, тем лучше уравнение регрессии $\hat{y} = f(x)$ согласуется с данными наблюдений.

Величина R^2 показывает, какая доля общей дисперсии (вариации) результативного признака *y* объясняется уравнением регрессии. Например, значение $R^2 = 0{,}75$ означает, что уравнение регрессии объясняет 75 % общей дисперсии (вариации) результативного признака *y*.

Точность построенной модели регрессии оценивается с помощью *средней квадратической ошибки* $(\varepsilon_{\text{кв}})$ либо *средней ошибки аппроксимации* $(\overline{A})$, которая представляет собой среднее относительное отклонение расчетных значений от наблюдаемых:

$$\varepsilon_{\text{кв}} = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{n}}, \tag{1.11}$$

или

$$\overline{A} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \hat{y}_i}{y_i}100\ \%. \tag{1.12}$$

Чем ниже средняя ошибка аппроксимации $\overline{A}$, тем лучше модель регрессии описывает исходные данные.

Построенное уравнение регрессии можно считать удовлетворительным, если значение $\overline{A}$ не превышает 10–12 %.

1.3. Оценка значимости уравнения регрессии и точности его коэффициентов

Оценка *статистической значимости уравнения регрессии* осуществляется с помощью *F*-критерия Фишера. Если выполняется условие $F_{\text{факт}} > F_{\text{табл}}$, то гипотеза о статистической незначимости уравнения регрессии отвергается и уравнение считается значимым. Если $F_{\text{факт}} \le F_{\text{табл}}$, то признается статистическая незначимость или ненадежность уравнения регрессии. Величина $F_{\text{факт}}$ определяется:

$$F_{\text{факт}} = \frac{R^2}{1 - R^2} \cdot \frac{n - p - 1}{p}, \tag{1.13}$$

где *n* – число единиц совокупности; *p* – число параметров при факторных переменных. Для парной линейной регрессии *p* = 1.

$F_{\text{табл}}$ – табличное значение F-критерия Фишера при уровне значимости α и числе степеней свободы $k_1 = p,\ k_2 = n - p - 1$.

Для оценки *статистической значимости коэффициентов линейной регрессии* применяется t-критерий Стьюдента. Для этого вычисляется фактическое значение t-критерия $t_{\text{факт}}$ и определяется критическое (табличное) значение t-критерия $t_{\text{табл}}$. Далее проверяется условие $t_{\text{факт}} > t_{\text{табл}}$. Если условие выполняется, то гипотеза о статистической незначимости коэффициента уравнения регрессии отвергается и коэффициент регрессии считается статистически значимым.

Величины $t_{b\,\text{факт}}$, $t_{a\,\text{факт}}$ определяются по формулам (1.14):

$$t_{b\text{факт}} = \frac{b}{s_b};\ t_{a\text{факт}} = \frac{a}{s_a}, \tag{1.14}$$

где s_b и s_a – стандартные ошибки коэффициентов регрессии.

Величина $t_{\text{крит}} = t_{1-\alpha,\, n-2}$ представляет собой табличное значение t-критерия Стьюдента при уровне значимости α и числе степеней свободы $k = n - p - 1$ (определяется по таблицам).

Представление о *точности полученных оценок* коэффициентов регрессии (на сколько далеко они могут отклониться от истинных значений коэффициентов) можно получить, используя «стандартные ошибки» коэффициентов регрессии.

Стандартные ошибки коэффициентов регрессии (s_a), (s_b) определяются отношениями (1.15)–(1.16):

$$s_b = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \Big/ (n-2)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{\frac{s_{\text{ост}}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \frac{s_{\text{ост}}}{\sigma_x \sqrt{n}}; \tag{1.15}$$

$$s_a = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{(n-2)} \times \frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}} = \sqrt{s_{\text{ост}}^2 \frac{\sum_{i=1}^{n} x_i^2}{n^2 \sigma_x^2}} = s_{\text{ост}} \frac{\sqrt{\sum_{i=1}^{n} x_i^2}}{n\sigma_x}, \tag{1.16}$$

где $s_{\text{ост}}^2$ – представляет собой несмещенную оценку остаточной дисперсии [см. формулу (1.17)].

$$s_{\text{ост}}^2 = \frac{\sum_{i=1}^{n}\left(\widehat{y}_i - y_i\right)^2}{(n-2)}. \tag{1.17}$$

Для определения величины отклонений полученных оценок от точных значений осуществляется построение доверительных интервалов. Под доверительным интервалом понимаются пределы, в которых лежит точное значение определенного показателя с заданной вероятностью ($P = 1 - \alpha$).

Доверительные интервалы для точных значений параметров $\widetilde{a}$ и $\widetilde{b}$ уравнения линейной регрессии определяются:

$$a - t_{1-\alpha,\, n-2} \cdot s_a < \widetilde{a} < a + t_{1-\alpha,\, n-2} \cdot s_a; \tag{1.18}$$

$$b - t_{1-\alpha,\, n-2} \cdot s_b < \widetilde{a} < a + t_{1-\alpha,\, n-2} \cdot s_b. \tag{1.19}$$

Величина $t_{1-\alpha,\, n-2}$ – это табличное значение t-критерия Стьюдента при уровне значимости α и числе степеней свободы $n - 2$.

Если в границы доверительного интервала попадает нуль, т. е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается равным нулю.

1.4. Точечный и интервальный прогноз по уравнению линейной регрессии. Коэффициент эластичности

Точечный прогноз заключается в получении прогнозного значения y_p, которое определяется путем подстановки в уравнение регрессии $\widehat{y}_x = a + bx$ соответствующего (прогнозного) значения x_p.

Интервальный прогноз заключается в построении доверительного интервала прогноза, т. е. нижней и верхней границ $y_{p\,\min}$, $y_{p\,\max}$, интервала, содержащего точную величину для прогнозного значения y_p ($y_{p\,\min} < y_p < y_{p\,\max}$) с заданной вероятностью.

При построении доверительного интервала прогноза используется стандартная ошибка значения прогноза s_{y_p}, связанная с дисперсией ошибки прогноза $s_{y_p}^2$ соотношением $s_{y_p} = \sqrt{s_{y_p}^2}$.

$$s_{y_p} = s_{\text{ост}} \sqrt{1 + \frac{1}{n} + \frac{(x_{\text{р}} - \bar{x})^2}{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2}}. \qquad (1.20)$$

Доверительный интервал для индивидуального значения прогноза $y_{\text{р}}$ определяется соотношением (1.21):

$$y_p - t_{1-\alpha,\, n-2} \cdot s_{y_p} \le \hat{y}_p \le y_p + t_{1-\alpha,\, n-2} \cdot s_{y_p}, \qquad (1.21)$$

где $t_{1-\alpha,\, n-2}$ представляет собой табличное значение t-критерия Стьюдента при уровне значимости α и числе степеней свободы $n - 2$.

В исследованиях применение находит коэффициент эластичности (Э). Для линейной регрессии рассчитывают средний коэффициент эластичности [см. формулу (1.22)]:

$$\overline{Э} = f'(\bar{x}) \frac{\bar{x}}{\bar{y}} = b \frac{\bar{x}}{\bar{y}}. \qquad (1.22)$$

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

Расчетная работа № 1
Парный регрессионный анализ: построение модели и проверка ее качества

Задание. На основании данных Приложения (см. табл. П.1.1 и П.1.2) для соответствующего варианта (см. табл. 1.2):

1. Построить предложенные в табл. 1.2 уравнения регрессии, включая линейную регрессию, используя формулы (1.4)–(1.9).
2. Вычислить показатели качества и точности уравнений.
3. Проверить значимость уравнений регрессии при уровнях значимости 0,05 и 0,01.
4. Определить лучшее уравнение регрессии на основе средней ошибки аппроксимации.
5. Проверить значимость коэффициентов линейной регрессии и построить доверительные интервалы для значений $\tilde{a}$ и $\tilde{b}$ уравнения линейной регрессии с уровнем значимости 0,05.
6. Построить точечный и интервальный прогноз для значения $x = x_{\max}$ по уравнению линейной регрессии с уровнем значимости 0,05.

7. Определить средний коэффициент эластичности по уравнению линейной регрессии.

Таблица 1.2

Варианты кривых выравнивания к работе № 1

Вари-ант	Графы из табл. П.1.1 и П.1.2	Вари-ант	Графы из табл. П.1.1 и П.1.2	Вари-ант	Графы из табл. П.1.1 и П.1.2	Вари-ант	Графы из табл. П.1.1 и П.1.2	Вари-ант	Графы из табл. П.1.1 и П1.2	Виды кривых выравнивания
1	1, 14	**6**	11, 14	**11**	8, 15	**16**	6, 17	**21**	2, 19	линейная, степенная
2	2, 14	**7**	12, 14	**12**	12, 15	**17**	9, 17	**22**	4, 19	линейная, экспоненциальная
3	4, 14	**8**	2, 15	**13**	1, 17	**18**	11, 17	**23**	6, 19	линейная, показательная
4	6, 14	**9**	3, 15	**14**	2, 17	**19**	12, 17	**24**	9, 19	линейная, логарифмическая
5	9, 14	**10**	7, 15	**15**	4, 17	**20**	1, 19	**25**	11, 19	линейная, гиперболическая

Пример выполнения работы. Исходные данные представлены в табл. 1.3. Необходимо построить линейную и степенную модели.

Таблица 1.3

Исходные данные для примера выполнения работы № 1

Номер наблюдения	Значения x	Значения y	Номер наблюдения	Значения x	Значения y	Номер наблюдения	Значения x	Значения y	Номер наблюдения	Значения x	Значения y
1	113	39	**7**	117	31	**13**	125	39	**19**	138	34
2	124	37	**8**	113	40	**14**	118	37	**20**	124	48
3	124	36	**9**	122	48	**15**	122	35	**21**	123	30
4	122	36	**10**	139	64	**16**	133	54	**22**	149	59
5	128	26	**11**	126	39	**17**	136	36	–		
6	140	43	**12**	120	34	**18**	136	35			

1. Определим коэффициенты a и b линейной регрессии, используя результаты промежуточных расчетов (табл. 2.3).

Таблица 1.4

Промежуточные расчеты для линейной регрессии (фрагмент)

№	x	y	x^2	y^2	xy	$\hat{y}$	$\left\lvert\frac{(\hat{y}-y)}{y}\right\rvert$	$(\hat{y}-y)^2$	$(y-\bar{y})^2$
1	113	39	12769	1521	4407	33,4	0,15	31,57	1,00
…	..	..	…	…	…	…	…	…	…
22	149	59	22201	3481	8791	50,5	0,21	72,04	361,00
Сумма	2792	880	356192	37038	112566	880	3,28	1416,4	1838
Среднее значение	126,9	40	16191	1683,5	5117	40	0,15	64,38	83,5

$$b = \frac{5116,6 - 126,9 \cdot 40}{\overline{16190,6} - 126,9^2} = 0,476, \quad a = 40 - 126,9 \cdot 0,476 = -20,39.$$

Уравнение линейной регрессии: $y = -20,39 + 0,476 \cdot x$.

2. Для построения степенной функции $y_x = a_0 x^{a_1}$ введем новые переменные $x' = \ln x$; $y' = \ln y$, вычислим значения новых переменных и выполним промежуточные расчеты (табл. 1.5).

Получим: $a_1 = \dfrac{\overline{\ln x \cdot \ln y} + \overline{\ln x} \cdot \overline{\ln y}}{\overline{\ln^2 x} - \overline{\ln x}^2} = \dfrac{39,231 - 4,84 \cdot 3,67}{23,439 - 4,84^2} = 1,263.$

$$\ln a_0 = \overline{\ln y} - a_1 \overline{\ln x} = 3,367 - 1,263 \cdot 4,841 = 2,451;$$

$$a_0 = \exp(2,451) = 0,0862.$$

Уравнение степенной регрессии: $y = 0,0862 x^{1,263}$.

Таблица 1.5

Промежуточные расчеты для степенной регрессии (фрагмент)

№	$\ln x$	$\ln y$	$\ln^2 x$	$\ln^2 y$	$\ln x \cdot \ln y$	$\hat{y}$	$\left\lvert\frac{(\hat{y}-y)}{y}\right\rvert$	$(\hat{y}-y)^2$	$(y-\bar{y})^2$
1	4,73	3,66	22,35	13,42	17,32	33,85	0,14	26,52	1,00
…									
22	5,00	4,08	25,04	16,63	20,41	48,01	0,14	120,86	361
Сумма	107	80,64	515,67	296,6	390,5	863,08	3,60	1437,9	1838
Среднее значение	4,84	3,67	23,44	13,48	17,75	39,23	0,16	65,36	83,55

3. Вычислим показатели качества: индекс корреляции R, индекс детерминации R^2, среднюю квадратическую ошибку $\varepsilon_{\text{кв}}$, среднюю ошибку аппроксимации $\overline{A}$. Для *линейной* модели (табл. 1.4):

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}} = \sqrt{1 - \frac{1416{,}37}{1838}} = 0{,}479, \quad R^2 = 0{,}479^2 = 0{,}229.$$

$$\varepsilon_{\text{кв}} = \sqrt{\frac{1416{,}371}{22}} = \sqrt{64{,}381} = 8{,}024, \quad \overline{A} = \frac{1}{22} \cdot 3{,}282 \cdot 100\,\% = 14{,}9\,\%.$$

Для *степенной* модели (табл. 1.5):

$$R = \sqrt{1 - \frac{1437{,}85}{1838}} = 0{,}467, \quad R^2 = 0{,}467^2 = 0{,}218;$$

$$\varepsilon_{\text{кв}} = \sqrt{\frac{1437{,}85}{22}} = \sqrt{65{,}357} = 8{,}084, \quad \overline{A} = \frac{1}{22} \cdot 3{,}596 \cdot 100\,\% = 16{,}3\,\%.$$

4. Проверим значимость уравнения регрессии. Для *линейной* модели (табл. 1.4):

$$F_{\text{факт}} = \frac{0{,}229}{1 - 0{,}229} \cdot \frac{22 - 1 - 1}{1} = 5{,}95;$$

$$F_{\text{крит } 0{,}05} = \text{FРРАСПОБ}(0{,}05; 1; 20) = 4{,}35;$$

$$F_{\text{крит } 0{,}01} = \text{FРРАСПОБ}(0{,}01; 1; 20) = 8{,}10.$$

Таким образом, при $\alpha = 0{,}05$ линейное уравнение значимо, при $\alpha = 0{,}01$ – не значимо.

Для *степенной* модели (табл. 1.5):

$$F_{\text{факт}} = \frac{0{,}218}{1 - 0{,}218} \cdot \frac{22 - 1 - 1}{1} = 5{,}57, \quad F_{\text{крит } 0{,}05} \text{ и } F_{\text{крит } 0{,}01} \text{ – те же.}$$

Таким образом, при $\alpha = 0{,}05$ степенное уравнение значимо, при $\alpha = 0{,}01$ – не значимо.

5. Определим лучшее уравнение регрессии по средней ошибке аппроксимации. Так как $\overline{A}_{\text{лин}} = 14{,}9\,\% < \overline{A}_{\text{степ}} = 16{,}3\,\%$, то линейная модель дает меньшую погрешность.

6. Проверка значимости коэффициентов линейной регрессии.

Определим оценку остаточной дисперсии $S^2_{\text{ост}}$, используя данные табл. 1.4: $S^2_{\text{ост}} = \dfrac{1416{,}37}{22-2} = 70{,}8.$

Определим стандартные ошибки коэффициентов регрессии (s_a), (s_b), используя значение σ_x:

$$S_b = \frac{\sqrt{70{,}8}}{22 \cdot 9{,}199} = 0{,}0416,\ \ S_a = \sqrt{70{,}8} \cdot \frac{\sqrt{356192}}{22 \cdot 9{,}199} = 24{,}81.$$

Вычислим: $t_a = \dfrac{-20{,}39}{24{,}81} = -0{,}82, \quad t_b = \dfrac{0{,}476}{0{,}0416} = 11{,}44.$

Значение $t_{1-\sigma,\, n-2}$ определим из Приложения (табл. П.1.8).

При $\sigma = 0{,}05$ и степени свободы $k = n - 2 = 22 - 2 = 20$ получаем $t_{1-\sigma,\, n-2} = 2{,}09.$ Так как $|\, t_a \,| = 0{,}82 < t_{1-\sigma,\, n-2} = 2{,}09,$ то делаем вывод о значимости коэффициента *a*. Так как $|\, t_b \,| = 11{,}44 > t_{1-\sigma,\, n-2} = 2{,}09,$ то делаем вывод о значимости коэффициента *b*.

7. Определение доверительных интервалов для точных значений параметров $\tilde{a}$ и $\tilde{b}$ уравнения линейной регрессии.

Для точного значения параметра $\tilde{a}$ доверительный интервал определять не нужно, так как значение коэффициента *a* не значимо.

Доверительный интервал для точного значения параметра $\tilde{b}$:

(0,476 – 2,09 · 0,0416; 0,476 + 2,09 · 0,0416) = (0,389; 0,563).

8. Построение точечного и интервального прогноза для значения для значения $x = x_{\max}$ по уравнению линейной регрессии $x_{\max} = 149.$

Точечный прогноз y_p: $y_p = -20{,}39 + 0{,}476 x_{\max} = -20{,}39 +$ $+ 0{,}476 \cdot 149 = 50{,}53.$

Вычислим: $S_{y_p} = \sqrt{70{,}8} \cdot \sqrt{1 + \dfrac{1}{22} + \dfrac{(149 - 126{,}9)^2}{1861{,}7}} = 9{,}62.$

Погноз: (50,53 – 2,09 · 9,62; 50,53 + 2,09 · 9,62) = (30,4; 70,6).

2. МНОЖЕСТВЕННЫЕ ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ: ПОСТРОЕНИЕ И АНАЛИЗ ПОКАЗАТЕЛЕЙ КАЧЕСТВА УРАВНЕНИЯ

2.1. Понятие множественной регрессии. Алгоритм выбора факторов и формы уравнения

Множественная регрессия представляет собой уравнение связи с несколькими независимыми переменными [см. формулу (2.1)]:

$$y = f(x_1, x_2, ..., x_m) + \varepsilon, \tag{2.1}$$

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

Для оценки параметров уравнения множественной регрессий применяют метод наименьших квадратов (МНК). Для линейных уравнений строится система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии:

$$b_1 = \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2}, \quad b_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{1 - r_{x_1x_2}^2},$$
$$a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2. \tag{2.2}$$

Спецификация модели включает в себя решение двух задач: 1) отбор p факторов x_j, наиболее влияющих на величину y; 2) выбор вида уравнения регрессии $\hat{y} = f(x_1, x_2, ..., x_m) + \varepsilon$.

Требования к отбору факторов во множественную регрессию:

– факторы не должны быть *коррелированны* и находиться в точной функциональной связи;

– факторы должны оказывать влияние на *вариацию независимой переменной*, т. е. быть статистически значимыми для модели (R^2).

Наличие высокой корреляции выявляется по значению линейного коэффициента корреляции. Если выполняется условие $r_{x_ix_j} \geq 0{,}8$, то факторные переменные x_i, x_j называются явно коллинеарными.

В уравнение регрессии включается только один из коллинеарных факторов, при этом предпочтение отдается тому фактору, который имеет наименьшую тесноту связи с другими факторами.

В уравнении *линейной* множественной регрессии (25) параметры b_i при x_i называются коэффициентами «чистой» регрессии и характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизмененном значении других факторов, закрепленных на среднем уровне:

$$\hat{y}_x = a + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p. \quad (2.3)$$

В уравнении *степенной* множественной регрессии (26) показатели степеней b_j являются коэффициентами эластичности:

$$\hat{y}_x = a x_1^{b_1} x_2^{b_2} \ldots x_p^{b_p}. \quad (2.4)$$

Они показывают, на сколько процентов изменяется в среднем результат с изменением соответствующего фактора на 1 % при неизменности действия других факторов.

2.2. Уравнение множественной регрессии в стандартизированном масштабе

На практике бывает необходимым сравнение влияния на зависимую переменную различных объясняющих переменных, когда эти переменные имеют разные единицы измерений. В этом случае используют *стандартизированные коэффициенты регрессии* β_i. Достоинство данных коэффициентов β_i в том, что их можно сравнивать межу собой (коэффициенты «чистой» регрессии b_i несравнимы), а также ранжировать факторы по силе их воздействия на результат.

Уравнение множественной регрессии в стандартизированном масштабе имеет вид [см. формулу (2.5)]:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \ldots + \beta_m t_{x_m} + \varepsilon, \quad (2.5)$$

где $t_y, t_{x_1}, ..., t_{x_m}$ – стандартизированные переменные; β_i – стандартизированные коэффициенты регрессии.

Коэффициенты «чистой» регрессии b_i связаны со стандартизированными коэффициентами регрессии β_i следующим образом:

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}} \text{ или } \beta_i = b_i \frac{\sigma_{x_1}}{\sigma_y}. \quad (2.6)$$

Средние коэффициенты эластичности рассчитываются (2.7):

$$\overline{Э}_{yx_i} = b_i \frac{\overline{x}_i}{\overline{y}}. \quad (2.7)$$

Данные коэффициенты показывают, на сколько процентов в среднем измениться результат, при изменении соответствующего фактора на 1 %. Средние показатели эластичности можно сравнивать.

2.3. Проверка значимости уравнения множественной регрессии и его коэффициентов

Тесноту совместного влияния факторов на результат оценивает *коэффициент множественной корреляции* [см. формулу (2.8)].

Значение индекса множественной корреляции лежит в пределах от 0 до 1 и должно соответствовать условию: $R_{yx_1x_2\ldots x_m} \geq r_{yx_i}$:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1 - \frac{\sigma^2_{y_{\text{ост}}}}{\sigma^2_y}}. \tag{2.8}$$

Также при линейной зависимости признаков формула коэффициента множественной корреляции (2.9) может быть представлена:

$$R_{yx_1x_2\ldots x_m} \geq \sqrt{\Sigma\beta_i \cdot r_{yx_i}}, \tag{2.9}$$

где β_i – стандартизированные коэффициенты регрессии; r_{yx_i} – парные коэффициенты результата с каждым фактором.

Качество модели в целом оценивает коэффициент множественной детерминации, который определяется как квадрат $R^2_{yx_1x_2\ldots x_m}$.

Для того чтобы не допустить преувеличения тесноты связи, применяется скорректированный коэффициент множественной детерминации, который содержит поправку на число степеней свободы и рассчитывается по формуле (2.10):

$$\hat{R}^2 = 1 - \left(1 - R^2\right) \cdot \frac{(n-1)}{(n-m-1)}, \tag{2.10}$$

где n – число наблюдений; m – число факторов.

Частные коэффициенты корреляции, измеряющие влияние на y фактора x_i при исключении влияния других факторов, могут быть определены через множественные коэффициенты корреляции или по рекуррентной формуле. В случае построения уравнения с двумя факторами данные формулы получают вид:

$$r_{yx_1x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_2}}}, \quad r_{yx_2x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_1}}}; \qquad (2.11)$$

$$r_{yx_1x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\left(1 - r^2_{yx_2}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}, \quad r_{yx_2x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}. \qquad (2.12)$$

Рассчитанные по рекуррентной формуле (2.11)–(2.12) частные коэффициенты корреляции изменяются в пределах от –1 до +1, а по формулам (2.10) через множественные коэффициенты детерминации – от 0 до 1. Частные коэффициенты корреляции дают меру тесноты связи каждого фактора с результатом в чистом виде.

Значимость уравнения множественной регрессии в целом оценивается с помощью *F*-критерия Фишера. Частный *F*-критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении. С его помощью можно оценить наиболее целесообразную последовательность включения факторов в уравнение множественной регрессии. В случае уравнения с двумя факторами формулы частного *F*-критерия для x_1 и x_2 следующие:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - R^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{n - m - 1}{1}, \quad F_{x_2} = \frac{R^2_{yx_1x_2} - R^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot \frac{n - m - 1}{1}. \qquad (2.13)$$

Фактическое значение частного *F*-критерия сравнивается с табличным при уровне значимости α и числе степеней свободы: $k_1 = 1$ и $k_2 = n - m - 1$. Если фактическое значение F_{x_i} превышает $F_{\text{табл}}\,(\alpha, k_1, k_2)$, то дополнительное включение x_i в модель статистически оправдано и коэффициент частной регрессии b_i при x_i статистически значим.

Оценка значимости коэффициентов чистой регрессии проводится по *t*-критерию Стьюдента. Для каждого фактора, как и в парной регрессии, используется формула (2.14):

$$t_{b_i} = \frac{b_i}{m_{b_i}}. \qquad (2.14)$$

Для уравнения регрессии в случае двухфакторной модели средняя квадратическая ошибка коэффициента регрессии определяется:

$$m_{b_1} = \frac{\sigma_y \sqrt{1 - R^2_{yx_1x_2}}}{\sigma_{x_1} \sqrt{1 - r^2_{x_1x_2}}} \cdot \frac{1}{\sqrt{n-m-1}}, \quad m_{b_2} = \frac{\sigma_y \sqrt{1 - R^2_{yx_1x_2}}}{\sigma_{x_2} \sqrt{1 - r^2_{x_1x_2}}} \cdot \frac{1}{\sqrt{n-m-1}}. \quad (2.15)$$

Между t-критерием Стьюдента и частным F-критерием Фишера существует следующая связь: $|t_{b_i}| = \sqrt{F_{x_i}}$.

Расчетная работа № 2
Построение множественной регрессионной модели и проверка ее качества

Задание. На основании данных Приложения (см. табл. П.1.1 и П.1.2) соответствующего варианта (см. табл. 2.1):

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

Таблица 2.1

Варианты выполнения расчетной работы № 2

Вариант	Графы из табл. П.1.1 и П.1.2 для переменных		Уровень значимости α	Вариант	Графы из табл. П.1.1 и П.1.2 для переменных		Уровень значимости α	Вариант	Графы из табл. П1.1 и П.1.2 для переменных		Уровень значимости α
	y	x_i			y	x_i			y	x_i	
1	14	1,2,3	0,05	**10**	15	3,4,5	0,01	**19**	17	7,8,9	0,05
2	15	1,2,3	0,01	**11**	19	3,4,5	0,05	**20**	18	7,8,9	0,01
3	16	1,2,3	0,05	**12**	14	6,7,8	0,01	**21**	15	8,9,10	0,05
4	17	1,2,3	0,01	**13**	15	6,7,8	0,05	**22**	19	8,9,10	0,01
5	18	1,2,3	0,05	**14**	16	6,7,8	0,01	**23**	16	11,12,13	0,05
6	14	2,3,4	0,01	**15**	17	6,7,8	0,05	**24**	17	11,12,13	0,01
7	15	2,3,4	0,05	**16**	18	6,7,8	0,01	**25**	18	11,12,13	0,05
8	17	2,3,4	0,01	**17**	14	7,8,9	0,05	–			
9	18	2,3,4	0,05	**18**	15	7,8,9	0,01				

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

3. Найти скорректированный коэффициент множественной детерминации. Сравнить его с нескорректированным (общим) коэффициентом детерминации.

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

5. С помощью t-критерия оценить статистическую значимость коэффициентов чистой регрессии;

6. 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.

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

8. Провести проверку полученных результатов с помощью надстройки «Анализ данных» табличного процессора MS Excel.

Пример выполнения работы. По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (млн р.) 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 (%). Исходные данные представлены в табл. 2.2.

Таблица 2.2

Исходные данные для примера выполнения работы № 2

№	Значения			№	Значения			№	Значения		
	y	x_1	x_2		y	x_1	x_2		y	x_1	x_2
1	7,0	3,9	10,0	**8**	8,0	4,4	20,0	**15**	12,0	8,0	28,0
2	7,0	3,9	14,0	**9**	8,0	5,3	20,0	**16**	12,0	8,2	29,0
3	7,0	3,7	15,0	**10**	10,0	6,8	20,0	**17**	12,0	8,1	30,0
4	7,0	4,0	16,0	**11**	9,0	6,0	21,0	**18**	12,0	8,5	31,0
5	7,0	3,8	17,0	**12**	11,0	6,4	22,0	**19**	14,0	9,6	32,0
6	8,0	4,8	19,0	**13**	9,0	6,8	22,0	**20**	14,0	9,0	36,0
7	8,0	5,4	19,0	**14**	11,0	7,2	25,0	–			

1. Найдем средние квадратические отклонения признаков. Для удобства проведения расчетов поместим результаты промежуточных расчетов в табл. 2.3.

Получаем:

$$\sigma_y = \sqrt{\overline{y^2} - \overline{y}^2} = \sqrt{97{,}9 - 9{,}6^2} = \sqrt{5{,}74} = 2{,}396;$$

$$\sigma_{x_1} = \sqrt{\overline{x_1^2} - \overline{x_1}^2} = \sqrt{41{,}887 - 6{,}19^2} = \sqrt{3{,}571} = 1{,}890;$$

$$\sigma_{x_2} = \sqrt{\overline{x_2^2} - \overline{x_2}^2} = \sqrt{541{,}4 - 22{,}3^2} = \sqrt{44{,}11} = 6{,}642.$$

Определим параметры линейного уравнения множественной регрессии. Для этого рассчитаем парные коэффициенты корреляции:

$$r_{yx_1} = \frac{\overline{yx_1} - \overline{y} \cdot \overline{x_1}}{\sigma_y \sigma_{x_1}} = \frac{63{,}815 - 6{,}19 \cdot 9{,}6}{1{,}890 \cdot 2{,}396} = 0{,}970;$$

$$r_{yx_2} = \frac{\overline{yx_2} - \overline{y} \cdot \overline{x_2}}{\sigma_y \sigma_{x_2}} = \frac{229{,}05 - 22{,}3 \cdot 9{,}6}{6{,}642 \cdot 2{,}396} = 0{,}941;$$

$$r_{x_1x_2} = \frac{\overline{x_1x_2} - \overline{x_1} \cdot \overline{x_2}}{\sigma_{x_1} \sigma_{x_2}} = \frac{149{,}87 - 6{,}19 \cdot 22{,}3}{1{,}890 \cdot 6{,}642} = 0{,}943.$$

Таблица 2.3

Промежуточные расчеты для среднего квадратического отклонения признаков (фрагмент)

№	y	x_1	x_2	yx_1	yx_2	x_1x_2	x_1^2	x_2^2	x^2
1	7,0	3,9	10,0	27,3	70,0	39,0	15,21	100,0	49,0
2	7,0	3,9	14,0	27,3	98,0	54,6	15,21	196,6	49,0
…	…	…	…	…	…	…	…	…	…
19	14,0	9,6	32,0	134,4	448,0	307,2	92,16	1024,0	196,0
20	14,0	9,0	36,0	126,0	504,0	324,0	81,0	1296,0	196,0
Сумма	192	123,8	446	1276,3	4581	2997,4	837,7	10828,0	1958,0
Среднее значение	9,6	6,19	22,3	63,82	229,05	149,87	41,89	541,4	97,9

Далее находим коэффициенты чистой регрессии и параметр a:

$$b_1 = \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{2{,}396}{1{,}890} \cdot \frac{0{,}970 - 0{,}941 \cdot 0{,}943}{1 - 0{,}943^2} = 0{,}946;$$

$$b_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{2{,}396}{6{,}642} \cdot \frac{0{,}941 - 0{,}970 \cdot 0{,}943}{1 - 0{,}943^2} = 0{,}0856;$$

$$a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 = 9{,}6 - 0{,}946 \cdot 6{,}19 - 0{,}0856 \cdot 22{,}3 = 1{,}835.$$

Получаем уравнение: $\hat{y} = 1{,}835 + 0{,}946x_1 + 0{,}0856x_2$.

Уравнение регрессии показывает, что при увеличении ввода в действие основных фондов на 1 % (при неизменном уровне удельного веса рабочих высокой квалификации) выработка продукции на одного рабочего увеличится в среднем на 0,946 млн р., а при увеличении удельного веса рабочих высокой квалификации в общей численности рабочих на 1 % выработка продукции на одного рабочего увеличивается в среднем на 0,086 млн р.

Далее проведем расчет остаточной дисперсии, средней ошибки аппроксимации и параметров стандартизированного уравнения регрессии. Промежуточные расчеты представим в виде табл. 2.4.

Таблица 2.4

Промежуточные расчеты остаточной дисперсии и ошибки апроксимации (фрагмент)

№	y	x_1	x_2	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$	A_i, %
1	7,0	3,9	10,0	6,380	0,620	0,384	8,851
2	7,0	3,9	14,0	6,723	0,277	0,077	3,960
…	…	…	…	…	…	…	…
19	14,0	9,6	32,0	13,656	0,344	0,118	2,459
20	14,0	9,0	36,0	13,431	0,569	0,324	4,067
Сумма	192	123,8	446	191,992	0,008	6,093	95,437
Среднее значение	9,6	6,19	22,3	9,6	–	0,305	4,77

Остаточная дисперсия: $\sigma^2_{\text{ост}} = \dfrac{\sum(y-\hat{v})^2}{n} = \dfrac{6{,}093}{20} = 0{,}305.$

Средняя ошибка аппроксимации: $\overline{\text{A}} = \dfrac{1}{n}\sum\left|\dfrac{y-\hat{y}}{y}\right| \cdot 100\ \% =$

$= \dfrac{95{,}44\ \%}{20} = 4{,}77\ \%.$

Таким образом, качество модели признается хорошим, так как средняя ошибка аппроксимации не превышает 10 %.

Далее определяем коэффициенты β_1 и β_2 стандартизованного уравнения регрессии:

$$\beta_1 = b_1\frac{\sigma_{x_1}}{\sigma_y} = 0{,}946\frac{1{,}890}{2{,}396} = 0{,}746;\ \beta_2 = b_2\frac{\sigma_{x_2}}{\sigma_y} = 0{,}0856\frac{6{,}642}{2{,}396} = 0{,}237.$$

Стандартизированное уравнение: $\hat{t}_y = 0{,}746 t_{x_1} + 0{,}237 t_{x_2}$.

Таким образом, можно сказать, что ввод в действие новых основных фондов оказывает большое влияние на выработку продукции, чем удельный вес рабочих высокой квалификации.

Далее оцениваем влияние факторов при помощи средних коэффициентов эластичности:

$$\overline{Э}_1 = b_1 \frac{\overline{x}_1}{\overline{y}_{x_1}} = 0{,}946 \frac{6{,}19}{9{,}6} = 0{,}61; \quad \overline{Э}_2 = b_2 \frac{\overline{x}_2}{\overline{y}_{x_2}} = 0{,}0856 \frac{22{,}3}{9{,}6} = 0{,}20.$$

То есть увеличение только основных фондов (от своего среднего значения) или только удельного веса рабочих высокой квалификации на 1 % увеличивает в среднем выработку продукции на 0,61 % или 0,20 % соответственно. Таким образом, подтверждается большое влияние на результат y фактора x_1, чем фактора x_2.

2. Из предыдущего расчета видно, что значения коэффициентов парной корреляции равны: $r_{yx_1} = 0{,}970$; $r_{yx_2} = 0{,}941$; $r_{x_1x_2} = 0{,}943$.

Они указывают на сильную связь каждого фактора с результатом, а также высокую межфакторную зависимость (факторы x_1 и x_2 явно коллинеарны, так как $r_{x_1x_2} = 0{,}943 > 0{,}7$). При такой сильной межфакторной зависимости рекомендуется один из факторов исключить.

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии.

При двух факторах частные коэффициенты корреляции рассчитываются следующим образом:

$$r_{yx_1x_2} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right)\left(1 - r^2_{x_1x_2}\right)}} = \frac{0{,}970 - 0{,}941 \cdot 0{,}943}{\sqrt{\left(1 - 0{,}941^2\right) \cdot \left(1 - 0{,}943^2\right)}} = 0{,}734;$$

$$r_{yx_2x_1} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right)\left(1 - r^2_{x_1x_2}\right)}} = \frac{0{,}941 - 0{,}970 \cdot 0{,}943}{\sqrt{\left(1 - 0{,}970^2\right) \cdot \left(1 - 0{,}943^2\right)}} = 0{,}325.$$

Если сравнивать коэффициенты парной и частной корреляции, то можно увидеть, что из-за высокой межфакторной зависимости коэффициенты парной корреляции дают завышенные оценки тесноты связи. Именно по этой причине рекомендуется при наличии сильной коллинеарности (взаимосвязи) факторов исключать из исследования тот фактор, у которого теснота парной зависимости меньше, чем теснота межфакторной связи.

Далее рассчитаем коэффициент множественной корреляции:

$$R_{yx_1x_2} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}} = \sqrt{1 - \frac{0{,}305}{5{,}74}} = 0{,}973;$$

$$R_{yx_1x_2} = \sqrt{\sum \beta_i r_{yx_i}} = \sqrt{0{,}746 \cdot 0{,}970 + 0{,}237 \cdot 0{,}941} = 0{,}973.$$

Коэффициент множественной корреляции указывает на весьма сильную связь всего набора факторов с результатом.

3. Нескорректированный коэффициент множественной детерминации $R^2_{yx_1x_2} = 0{,}947$ указывает на то, что доля дисперсии результата за счет представленных в уравнении факторов в общей вариации результата составляет 94,7 %.

Скорректированный коэффициент множественной детерминации позволяет оценить тесноту связи, которая не зависит от числа факторов и может сравниваться по разным моделям с разным числом факторов. Для данной модели он равен:

$$\hat{R}^2 = 1 - \left(1 - R^2\right)\frac{(n-1)}{(n-m-1)} = 1 - (1 - 0{,}947)\frac{20-1}{20-2-1} = 0{,}941.$$

Коэффициент указывает на весьма высокую (более 94 %) детерминированность результата y в модели факторами x_1 и x_2.

4. Оценку надежности уравнения регрессии в целом и показателя тесноты связи $R_{yx_1x_2}$ дает F-критерий Фишера:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n-m-1}{m} = \frac{0{,}973^2}{1 - 0{,}973^2} \cdot \frac{20-2-1}{2} = 151{,}88.$$

Получаем, что $F_{\text{факт}} = 151{,}88 > F_{\text{табл}} = 3{,}59$ (при n = 20), т. е. вероятность случайно получить такое значение F-критерия не превышает допустимый уровень значимости – 5 %. Следовательно, полученное значение сформировалось не случайно, а под влиянием выбранных факторов. Это подтверждает статистическую значимость всего уравнения и показателя тесноты связи $R_{yx_1x_2}$.

5. Проведем оценку статистической значимости параметров чистой регрессии с помощью t-критерия Стьюдента.

$$m_{b_1} = \frac{\sigma_y \sqrt{1 - R^2_{yx_1x_2}}}{\sigma_{x_1}\sqrt{1 - r^2_{x_1x_2}}} \cdot \frac{1}{\sqrt{n-m-1}} = \frac{2{,}396\sqrt{1-0{,}973^2}}{1{,}890\sqrt{1-0{,}943^2}} \cdot \frac{1}{\sqrt{20-3}} = 0{,}2132;$$

$$m_{b_2} = \frac{\sigma_y \sqrt{1 - R^2_{yx_1x_2}}}{\sigma_{x_2}\sqrt{1 - r^2_{x_1x_2}}} \cdot \frac{1}{\sqrt{n-m-1}} = \frac{2{,}396\sqrt{1-0{,}973^2}}{6{,}642\sqrt{1-0{,}943^2}} \cdot \frac{1}{\sqrt{20-3}} = 0{,}0607.$$

Тогда фактическое значение t-критерия Стьюдента равен:

$$t_{b_1} = \frac{b_1}{m_{b_1}} = \frac{0{,}946}{0{,}2132} = 4{,}44; \quad t_{b_2} = \frac{b_2}{m_{b_2}} = \frac{0{,}0856}{0{,}0607} = 1{,}41.$$

Табличное значение критерия при $\alpha = 0{,}05$ и числе степеней свободы k =17 составит $t_{\text{табл}}$ $(\alpha = 0{,}05;\ k = 17) = 2{,}11$. Таким образом, признается статистическая значимость параметра b_1, так как $t_{b_1} > t_{\text{табл}}$, и случайная природа формирования параметра b_2, так как $t_{b_2} < t_{\text{табл}}$.

Доверительные интервалы для параметров чистой регрессии согласно неравенству $b_1 - m_{b_1} \cdot t_{\text{табл}} \le b_1^* \le b_1 + m_{b_1} \cdot t_{\text{табл}}$ составят:

$$0{,}496 \le b_1^* \le 1{,}396 \ \text{ и } \ -0{,}0425 \le b_2^* \le 0{,}2137.$$

6. С помощью частных F-критериев Фишера оценим целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1 согласно нижеприведенному расчету.

В начале определим:

$$R^2_{yx_1} = r^2_{yx_1} = 0{,}970^2 = 0{,}941; \quad R^2_{yx_2} = r^2_{yx_2} = 0{,}941^2 = 0{,}885.$$

Тогда имеем:

$$F_{x_1} = \frac{0{,}947 - 0{,}885}{1 - 0{,}947} \cdot \frac{20-2-1}{1} = 19{,}89;$$

$$F_{x_2} = \frac{0{,}947 - 0{,}941}{1 - 0{,}947} \cdot \frac{20-2-1}{1} = 1{,}924.$$

Таким образом, $F_{x_2} = 0{,}89 < F_{\text{табл}}$ $(\alpha = 0{,}05;\ k_1 = 1;\ k_2 = 17) = 4{,}45$. Следовательно, включение в модель фактора x_2 после того, как в модель включен фактор x_1 статистически нецелесообразно: прирост факторной дисперсии за счет дополнительного признака x_2 оказывается незначительным, несущественным; фактор x_2 включать в уравнение после фактора x_1.

Если поменять порядок включения факторов в модель и рассмотреть вариант включения x_1 после x_2, то результат расчета частного F-критерия для x_1 будет иным. $F_{x_1} = 17{,}86 > F_{\text{табл}} = 4{,}45$, т. е. вероятность его случайного формирования меньше принятого стандарта $\alpha = 0{,}05$ (5 %). Следовательно, значение частного F-критерия для дополнительно включенного фактора x_1 не случайно, является статистически значимым, надежным и достоверным: прирост факторной дисперсии за счет дополнительного фактора x_1 является существенным. Фактор x_1 должен присутствовать в уравнении, в числе в варианте, когда он дополнительно включается после фактора x_2.

7. Общий вывод состоит в том, что множественная модель с факторами x_1 и x_2 с $R^2_{yx_1x_2} = 0{,}947$ содержит неинформативный фактор x_2. Если исключить фактор x_2, то можно ограничиться уравнением парной регрессии. Определим его параметры.

$$\beta = \frac{\operatorname{cov}(y, x_1)}{\sigma_{x_2}} = \frac{63{,}815 - 6{,}19 \cdot 9{,}6}{3{,}571} = 1{,}23; \;\; \alpha = \bar{y} - \beta \cdot \bar{x} = 9{,}6 - 1{,}23 \cdot 6{,}19 = 1{,}99.$$

Таким образом, получаем:

$$\hat{y}_{x_1} = 1{,}99 + 1{,}23x_1; \quad r^2_{yx_1} = 0{,}941.$$

8. Для ведения расчетов следует использовать возможности надстройки «Анализ данных» процессора MS Excel, реализованные в таких функциях, как: «КОРРЕЛЯЦИЯ», «РЕГРЕССИЯ», «СТАНДОТКЛОНП()» «СТЬЮДРАСПОБР()», «FРАСПОБР()».

3. МЕТОДЫ ОБНАРУЖЕНИЯ И УСТРАНЕНИЯ МУЛЬТИКОЛЛИНЕАРНОСТИ

3.1. Признаки определения мультиколлинеарности

Существует несколько признаков, по которым может быть установлено наличие мультиколлинеарности:

1. Коэффициент детерминации R^2 достаточно высок, но некоторые из коэффициентов регрессии имеют низкие t-статистики.

2. Парная корреляция между двумя малозначимыми объясняющими переменными достаточно высока.

3. Высокие частные коэффициенты корреляции.

4. Сильная дополнительная регрессия. Мультиколлинеарность может иметь место вследствие того, что какая-либо из объясняющих переменных является линейной комбинацией (или близкой к линейной), комбинацией других объясняющих переменных. Для анализа строятся уравнения регрессии каждой из объясняющих переменных x_i на оставшиеся объясняющие переменные. Вычисляются соответствующие коэффициенты детерминации R^2 и рассчитывается их статистическая значимость на основе F-статистики. Если коэффициент R_j^2 статистически незначим, то x_i не является линейной комбинацией других переменных и ее можно оставить в уравнении.

3.2. Методы корректировки модели и выборки данных

К методам устранения мультиколлинеарности относятся:

1. Исключение из модели ряда коррелированных переменных.

2. Получение дополнительных данных или новой выборки. Например, при использовании ежегодных данных можно перейти к поквартальным данным. Увеличение количества данных сокращает дисперсии коэффициентов регрессии и тем самым увеличивает их статистическую значимость.

3. Изменение спецификации модели. Проблема мультиколлинеарности может быть решена путем изменения спецификации модели: либо изменяется форма модели, либо добавляются объясняющие переменные, не учтенные в первоначальной модели, но существенно влияющие независимую переменную.

3.3. Метод устранения мультиколлинеарности на основе предварительной информации

Для построения модели множественной регрессии можно воспользоваться предварительной информацией, в частности, известными значениями некоторых коэффициентов регрессии.

Значения коэффициентов, рассчитанные для предварительных моделей, могут быть использованы для разрабатываемой модели.

Предположим, что переменные x_1 и x_2 коррелированны. Для ранее построенной модели парной регрессии $y = b_0 + b_1 x_1 + \varepsilon$ был определен статистически значимый коэффициент b_1 (для определенности пусть $b_1 = 0{,}8$), связывающий y с x_1, если есть основания считать, что связь между y с x_1 остается неизменной, то можно предположить $\beta_1 = b_1 = 0{,}8$, тогда уравнение примет вид:

$$y = \beta_0 + 0{,}8x_1 + \beta_2 x_2 + \varepsilon \Rightarrow y - 0{,}8x_1 = \beta_0 + \beta_2 x_2 + \varepsilon. \qquad (3.1)$$

Уравнение (3.1) фактически является уравнением парной регрессии, для которого проблема мультиколлинеарности не существует.

3.4. Метод устранения мультиколлинеарности путем преобразования переменных

В ряде случаев минимизировать проблему мультиколлинеарности можно с помощью преобразования переменных.

Например, эмпирическое уравнение регрессии имеет вид (2.12):

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2, \qquad (3.2)$$

где x_1 и x_2 – коррелированны переменные. В этой ситуации можно попытаться определять зависимости (2.13) относительных величин:

$$\frac{\hat{y}}{x_1} = b_0 + b_2 \frac{x_2}{x_1}, \quad \frac{\hat{y}}{x_2} = b_0 + b_1 \frac{x_1}{x_2}. \qquad (3.3)$$

В данной модели проблема мультиколлинеарности отсутствует.

Расчетная работа № 3
Реализация методов установления и устранения мультиколлинеарности

Задание. На основании данных Приложения (см. табл. П.1.1 и П.1.2) для соответствующего варианта (см. табл. 2.1):

1. Проверить наличие коллинеарности и мультиколлинеарности. Отобрать неколлинеарные факторы.
2. Построить уравнение линейной регрессии.
3. Определить коэффициент множественной корреляции.
4. Проверить значимость уравнения при уровнях значимости 0,05 и 0,01.
5. Построить частные уравнения регрессии.
6. Определить средние частные коэффициенты эластичности.

Указания к решению. При выполнении лабораторной работы использовать возможности надстройки «Анализ данных» табличного процессора MS Excel (для расчета корреляционной матрицы, нахождения уравнений регрессии и коэффициентов координации и др.).

Пример выполнения работы. Исходные данные представлены в табл. 3.1.

Таблица 3.1

Исходные данные для примера выполнения работы № 3

№	y	x_1	x_2	x_3	$y_{расч}$	ε	№	y	x_1	x_2	x_3	$y_{расч}$	ε
1	113	10	1	77	124,92	–11,92	**16**	133	15	5	79	132,75	0,25
2	124	5	2	64	121,52	2,49	**17**	136	12	4	71	127,82	8,18
3	124	10	2	77	126,68	–2,68	**18**	146	16	9	68	135,43	10,57
4	122	13	2	66	122,31	–0,31	**19**	148	23	5	78	132,36	15,64
5	128	9	1	71	122,53	5,47	**20**	136	16	8	74	136,05	–0,05
6	140	14	6	81	135,31	4,69	**21**	138	10	3	64	123,28	14,72
7	117	12	1	58	117,37	–0,37	**22**	124	12	7	74	134,29	–10,29
8	113	15	3	66	124,07	–11,07	**23**	123	8	3	71	126,06	–3,06
9	122	13	2	73	125,09	–3,09	**24**	149	29	8	87	141,21	7,79
10	139	27	14	81	149,40	–10,40	**25**	130	9	4	56	121,86	8,14
11	126	8	6	73	132,13	–6,13	**26**	117	91	3	65	123,67	–6,67
12	120	8	3	65	123,67	–3,67	**27**	126	12	1	61	118,56	7,44
13	125	24	6	66	129,35	–4,35	**28**	110	7	1	35	108,24	1,76
14	118	8	1	74	123,72	–5,72	**29**	98	6	0	26	102,91	–4,91
15	122	8	4	64	125,04	–3,04	–						

1. Проверка наличия коллинеарности или мультиколлинеарности. Отбор неколлинеарных факторов.

Построим корреляционную матрицу, используя функцию «Сервис. Анализ данных. Корреляция» табличного процессора MS Excel.

Таблица 3.2

Корреляционная матрица

	y	x_1	x_2	x_3
y	1			
x_1	0,638	1		
x_2	0,680	0,710	1	
x_3	0,661	0,513	0,506	1

Из матрицы следует, что наблюдается коллинеарность между факторами x_1 и x_2, так как $r_{x_1x_2} = 0{,}710$. Для дальнейшего рассмотрения оставляем фактор x_2, так как он меньше коррелирует с фактором x_3 $(r_{x_2x_3} = 0{,}506 < r_{x_1x_3} = 0{,}513)$. Таким образом, далее будет строиться регрессия y на факторы x_2 и x_3.

2. Для построения уравнения линейной регрессии используем функцию «Сервис. Анализ данных. Регрессия» MS Excel: 1) вызов функции осуществляется через пункты меню: Сервис → Анализ данных → → Регрессия; 2) указываются ячейки, содержащие исходные значения переменных y и x_i; 3) если отсутствует свободный член в уравнении регрессии – установить флажок «Константа-ноль»; 4) указать место, где будут представлены результаты; 5) искомые значения коэффициентов линейного уравнения регрессии (a, b_i) берутся из столбца «Коэффициенты» таблицы результатов регрессии (табл. 3.5).

Результаты работы функции приведены в табл. 3.3–3.5.

Таблица 3.3

Результаты корреляционного анализа

Множественный *R*	0,773
R-квадрат	0,597
Нормированный *R*-квадрат	0,566
Стандартная ошибка	7,768
Наблюдения	29

Таблица 3.4

Результаты дисперсионного анализа

Пояснения	Число степеней свободы *df*	Сумма квадратов отклонений *SS*	Дисперсия на 1 степень свободы *MS*	Статистика Фишера *F*	Уровень значимости *F*
Регрессия	22	2326,1	1163,1	19,3	7,35E-06
Остаток	26	1569,1	60,3	–	–
Итого	28	3895,2	–	–	–

Таблица 3.5

Результаты регрессионного анализа

Пояснения	Коэффициенты уравнения регрессии	Стандартная ошибка определения коэффициентов	*t*-статистика	Вероятность ошибки α	Нижние 95%-ные пределы	Верхние 95%-ные пределы
Показатели	**Коэффициенты**	**Стандартная ошибка**	***t*-статистика**	***P*-значения**	**Нижние 95%-ные пределы**	**Верхние 95%-ные пределы**
Y-пересечение	92,585	8,351	11,087	0,0000	75,420	109,750
Переменная *X*1	1,761	0,547	3,219	0,0030	0,637	2,886
Переменная *X*2	0,397	0,134	2,952	0,0070	0,120	0,673

Из табл. 3.5 следует, что уравнение регрессии имеет вид:

$$y = 92{,}585 + 1{,}761x_1 + 0{,}397x_2.$$

3. Коэффициент множественной корреляции на основе данных табл. 3.3 принимаем равным $R_{yx_1x_2} = 0{,}773$.

4. Проверка значимости уравнения регрессии основана на использовании F-критерия Фишера. Фактическое значение критерия согласно табл. 3.4 равно $F_{\text{факт}} = 19{,}3$.

Для определения табличных значений используем функцию MS Excel «FРАСПОБР()», задав следующие параметры: $k_1 = 2$, $k_2 = 29 - 2 - 1 = 26$, $\alpha = 0{,}05$ и $\alpha = 0{,}01$.

В результате получаем $F_{\text{факт } 0{,}05} = 3{,}369$, $F_{\text{факт } 0{,}01} = 5{,}526$. Откуда следует, что уравнение регрессии значимо и при $\alpha = 0{,}05$ и $\alpha = 0{,}01$.

5. Частные уравнения регрессии. Предварительно определим средние значения переменных: $\bar{y} = 126{,}448$; $\bar{x}_2 = 3{,}966$; $\bar{x}_3 = 67{,}76$.

С учетом средних значений построим частные уравнения:

$$\hat{y}_{x_2x_3} = 92{,}585 + 1{,}761x_2 + 0{,}397 \cdot 67{,}76 = 119{,}486 + 1{,}761x_2;$$

$$\hat{y}_{x_3x_2} = 92{,}585 + 1{,}761 \cdot 3{,}966 + 0{,}397x_3 = 99{,}569 + 0{,}397x_3.$$

6. Средние частные коэффициенты эластичности:

$$\bar{Э}_{yx_2} = b_2\frac{\bar{x}_2}{\bar{y}} = 1{,}761\frac{3{,}966}{126{,}448} = 0{,}0552;\ \bar{Э}_{yx_3} = b_3\frac{\bar{x}_3}{\bar{y}} = 0{,}397\frac{67{,}76}{126{,}448} = 0{,}213.$$

4. МЕТОДЫ ОБНАРУЖЕНИЯ И СМЯГЧЕНИЯ ГЕТЕРОСКЕДАСТИЧНОСТИ

4.1. Обнаружение гетероскедастичности по методу Гольдфельда–Квандта

Для того чтобы метод наименьших квадратов давал надежные оценки параметров линейной регрессии, требуется чтобы дисперсии остатков ε в модели $y = a + b_1x_1 + b_2x_2 + ... + b_px_p + \varepsilon$ для каждого наблюдения были одинаковыми. Остатки, обладающие таким свойством, называются *гомоскедастичными*, а не обладающие – *гетероскедастичными*.

Для оценки гетероскедастичности можно использовать *метод Гольдфельда–Квандта*, который проверяет наличие зависимости остатков ε от одной из факторных переменных x_i. Алгоритм применения теста Гольдфельда–Квандта состоит из следующих шагов:

1) исходные данные наблюдений упорядочиваются по мере возрастания выбранной переменной x_i;

2) выделяются первые n_0 и последние n_0 наблюдений и исключаются из рассмотрения $C = n - 2n_0$ центральных наблюдений. При этом должно выполняться условие $n_0 > p$, где p – число оцениваемых параметров;

3) для каждой из групп наблюдений (верхней и нижней) оцениваются уравнения регрессии остатков ε по значимым факторам:

$$\varepsilon = a + b_1 x_1 + b_2 x_2 + \ldots + b_p x_p + u; \tag{4.1}$$

4) для каждого уравнения определяются остаточные суммы квадратов (S_1) и (S_2) остатков u_i и находится их отношение:

$$R = \frac{\max(S_2, S_1)}{\min(S_2, S_1)}. \tag{4.2}$$

Если выполняется условие $R > F_{\text{табл}}$, где $F_{\text{табл}}$ представляет собой табличное значение F-критерия Фишера при уровне значимости α и числе степеней свободы $k_1 = n_0 - p$, $k_2 = n_0 - p$, то предпосылка о равенстве дисперсий остаточных величин отвергается с уровнем α.

Чем больше величина R превышает табличное значение критерия $F_{\text{табл}}$, тем более нарушена предпосылка о равенстве дисперсий остаточных величин.

Авторами рекомендовано для случая одного фактора $n = 20$ принимать $C = 4$, при $n = 30$ принимать $C = 8$, при $n = 60$ принимать $C = 16$.

4.2. Тест ранговой корреляции Спирмена

При использовании данного теста предполагается, что дисперсия отклонений будет либо увеличиваться, либо уменьшаться с увеличением значений x. Поэтому для регрессии, построенной по методу наименьших квадратов, абсолютные величины отклонений ε_i и значения x_i будут коррелированны. Значения x_i и ε_i ранжируются (упорядочиваются по величинам). Затем определяется коэффициент ранговой корреляции [см. формулу (4.3)]:

$$r_{x,\varepsilon} = 1 - 6\frac{\sum d_i^2}{n(n^2 - 1)}, \tag{4.3}$$

где d_i – разность между рангами x_i и ε_i, $i = 1, 2, \ldots, n$; n – число наблюдений.

Например, если x_{20} является 25-м по величине среди всех наблюдений x, а ε_{20} является 32-м, то $d_i = 25 - 32 = -7$.

Доказано, что если коэффициент корреляции $r_{x,\varepsilon}$ для генеральной совокупности равен нулю, то статистика [см. формулу (4.4)] имеет распределение Стьюдента с числом степеней свободы $n - 2$:

$$t = \frac{r_{x,\varepsilon}\sqrt{n-2}}{\sqrt{1 - r_{x,\varepsilon}^2}}. \tag{4.4}$$

Следовательно, если наблюдаемое значение t-статистики, вычисляемое по формуле (4.4), превышает $t_{\text{кр}} = t_{\alpha/2,\, n-2}$ (определяемое блице критических точек распределения Стьюдента), то необходимо отклонить гипотезу о равенстве нулю коэффициента корреляции $r_{x,\varepsilon}$, а следовательно, и об отсутствии гетероскедастичности.

Если в модели больше одной объясняющей переменной, то проверка осуществляется с помощью t-статистики для каждой из них.

4.3. Смягчение проблемы гетероскедастичности по методу взвешенных наименьших квадратов

Данный метод применяется при известных для каждого наблюдения значениях σ_i^2. В этом случае можно устранить гетероскедастичность, разделив каждое наблюдаемое значение на соответствующее ему значение среднеквадратического отклонения.

Разделив обе части уравнения $y = \beta_0 + b_1 x_2 + \varepsilon_i$ на известное $\sigma_i = \sigma_i^2$, получаем:

$$\frac{y_i}{\sigma_i} = \beta_0\frac{1}{\sigma_i} + \beta_i\frac{1}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}. \tag{4.5}$$

Положив $\frac{y_i}{\sigma_i} = y_i^*$, $\frac{x_i}{\sigma_i} = x_i^*$, $\frac{\varepsilon_i}{\sigma_i} = v_i$, $\frac{1}{\sigma_i} = z_i$, получим уравнение регрессии без свободного члена, но с дополнительной объясняющей переменной *z* и с «преобразованным» отклонением *v*:

$$y_i^* = \beta_0 z_i + \beta_1 x_i^* + v_i . \tag{4.6}$$

Для преобразованной модели [см. формулу(4.6)] выполняются все предпосылки МНК.

Таким образом, метод взвешенных наименьших квадратов включает следующие этапы:

1. Значения каждой пары наблюдений (x_i, y_i) (x_i, y_i) делятся на известную величину σ_i. Тем самым наблюдениям с наименьшими дисперсиями придаются наибольшие «веса», а с максимальными дисперсиями – наименьшие «веса». Учет этого фактора увеличивает вероятность получения более точных оценок.

2. По МНК для преобразованных значений $\frac{y_i}{\sigma_i}$, $\frac{x_i}{\sigma_i}$, $\frac{1}{\sigma_i}$ строится уравнение регрессии без свободного члена с гарантированными качествами оценок.

Расчетная работа № 4
Реализация методов обнаружения и смягчения гетероскедастичности

Задание. На основании данных Приложения (см. табл. П.1.1 и П.1.2) для соответствующего варианта (см. табл. 2.1):

1. Построить уравнение линейной регрессии в стандартизированном масштабе.
2. Оценить информативность факторов на основе уравнения линейной регрессии в стандартизированном масштабе.
3. Вычислить частные коэффициенты корреляции.
4. Оценить их значимость при уровнях значимости 0,05 и 0,01.
5. Оценить информативность факторов на основе частных коэффициентов корреляции.
6. Построить уравнение регрессии с учетом только информативных факторов.
7. Проверить гипотезу о гомоскедастичности ряда остатков с уровнем значимости $\alpha = 0{,}05$.

8. Построить уравнение регрессии с помощью метода взвешенных наименьших квадратов.

Указания к решению. При выполнении работы использовать возможности надстройки «Анализ данных» табличного процессора MS Excel (для расчета корреляционной матрицы, нахождения уравнений регрессии, нахождения коэффициентов координации и др.).

Пример выполнения работы. Исходные данные возьмем из расчетной работы № 3 (см. табл. 3.1).

1. Построим уравнение линейной регрессии в стандартизированном масштабе. Его коэффициенты связаны с коэффициентами обычного уравнения регресии через среднеквадратические отклонения признаков. Определим средние квадратические отклонения σ_y, σ_{x_i}, использую функцию MS Excel «СТАНДОТКЛОНП()»:

$$\sigma_y = 11{,}59;\ \sigma_{x_2} = 3{,}057;\ \sigma_{x_3} = 12{,}44.$$

Следовательно, получаем:

$$\beta_2 = b_2 \frac{\sigma_{x_2}}{\sigma_y} = 1{,}761 \frac{3{,}057}{11{,}59} = 0{,}464; \quad \beta_3 = b_3 \frac{\sigma_{x_3}}{\sigma_y} = 0{,}397 \frac{12{,}44}{11{,}59} = 0{,}426.$$

Уравнение линейной регрессии в стандартизированном масштабе имеет вид: $t_y = 0{,}464 \cdot t_{x_2} + 0{,}426 \cdot t_{x_3}$.

2. Информативность факторов. Так как $\beta_1 = 0{,}464$ и $\beta_2 = 0{,}426$, то делаем вывод, что факторы практически одинаково информативны.

3. Частные коэффициенты корреляции. Для их вычисления воспользуемся формулой (4.7), в соответствии с которой необходимо вычислить $R^2_{yx_2x_3}$, $R^2_{yx_2}$, $R^2_{yx_3}$:

$$r_{yx_1x_2\ldots x_p} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2\ldots x_p}}{1 - R^2_{yx_2\ldots x_p}}}. \tag{4.7}$$

В данном примере величину $R^2_{yx_2x_3}$ можно взять из табл. 3.3, а величины $R^2_{yx_2}$, $R^2_{yx_3}$ вычислить, используя соответствующие коэффициенты линейной корреляции r_{yx_2} и r_{yx_3} из корреляционной матрицы в примере расчетной работы № 3:

$$R^2_{yx_2x_3} = 0{,}597; \quad R^2_{yx_2} = 0{,}462; \quad R^2_{yx_3} = 0{,}437.$$

В результате получим частные коэффициенты корреляции:

$$r_{yx_2x_3} = \sqrt{1-\frac{1-0{,}597}{1-0{,}437}} = 0{,}533; \quad r_{yx_3x_2} = \sqrt{1-\frac{1-0{,}597}{1-0{,}462}} = 0{,}500.$$

4. Оценим их значимость. Вычислим фактические значения частного F-критерия Фишера:

$$F_{\text{част}x_2} = \frac{R^2_{yx_2x_3} - R^2_{yx_3}}{1-R^2_{yx_2x_3}} \cdot \frac{n-2-1}{1} = \frac{0{,}597-0{,}437}{1-0{,}597} \cdot \frac{29-2-1}{1} = 10{,}33;$$

$$F_{\text{част}x_3} = \frac{R^2_{yx_2x_3} - R^2_{yx_2}}{1-R^2_{yx_2x_3}} \cdot \frac{n-2-1}{1} = \frac{0{,}597-0{,}462}{1-0{,}597} \cdot \frac{29-2-1}{1} = 8{,}68.$$

Для определения табличных значений используем встроенную функцию MS Excel «FРАСПОБР», задавая параметры $k_1 = 1$, $k_2 = 29 - 2 - 1 = 26$, $\alpha = 0{,}05$ и $\alpha = 0{,}01$. В результате получаем $F_{\text{факт }0{,}05} = 4{,}225$, $F_{\text{факт }0{,}01} = 7{,}721$. Откуда следует, что оба коэффициента корреляции значимы и при $\alpha = 0{,}05$, и при $\alpha = 0{,}01$.

5. Информативность факторов. Так как оба частных коэффициента значимы, то оба фактора x_2 и x_3 информативны и должны быть включены в уравнение регрессии.

6. Уравнение регрессии (из расчетной работы № 3):

$$y = 92{,}585 + 1{,}761x_2 + 0{,}397x_3.$$

7. Проверка гомоскедастичности. Вычисляем расчетные значения результативного признака по уравнению регрессии и определим остатки (см. табл. 3.1). Согласно методу Гольфельда–Квандта, упорядочим ряд остатков отдельно по фактору x_2 и по фактору x_3. Результаты приведены в табл. 4.1. Зачеркнутым шрифтом отмечены данные, не участвующие в рассмотрении. Согласно рекомендациям, их число равно $C = 7$.

Таблица 4.1

Упорядочивание ряда остатков по факторам x_2 и x_3

Остатки, упорядоченные по x_2								Остатки, упорядоченные по x_3							
№	x_2	x_3	ε	**№**	x_2	x_3	ε	**№**	x_2	x_3	ε	**№**	x_2	x_3	ε
1	0	26	–4,91	**16**	~~3~~	~~65~~	~~–6,67~~	**1**	0	26	–4,91	**16**	~~4~~	~~71~~	~~8,18~~
2	1	77	–11,92	**17**	~~4~~	~~64~~	~~–3,04~~	**2**	1	35	1,76	**17**	~~3~~	~~71~~	~~–3,06~~
3	1	71	5,47	**18**	~~4~~	~~71~~	~~8,18~~	**3**	4	56	8,14	**18**	~~2~~	~~73~~	~~–3,09~~

Окончание табл. 4.1

Остатки, упорядоченные по x_2								Остатки, упорядоченные по x_3							
№	x_2	x_3	ε	**№**	x_2	x_3	ε	**№**	x_2	x_3	ε	**№**	x_2	x_3	ε
4	1	58	–0,37	**19**	4	56	8,14	**4**	1	58	–0,37	**19**	6	73	–6,13
5	1	74	–5,72	**20**	5	79	0,25	**5**	1	61	7,44	**20**	1	74	–5,72
6	1	61	7,44	**21**	5	78	15,64	**6**	2	64	2,49	**21**	8	74	–0,05
7	1	35	1,76	**22**	6	81	4,69	**7**	4	64	–3,04	**22**	7	74	–10,29
8	2	64	2,49	**23**	6	73	–6,13	**8**	3	64	14,72	**23**	1	77	–11,92
9	2	77	–2,68	**24**	6	66	–4,35	**9**	3	65	–3,67	**24**	2	77	–2,68
10	2	66	–0,31	**25**	7	74	–10,29	**10**	3	65	–6,67	**25**	5	78	15,64
11	2	73	–3,09	**26**	8	74	–0,05	**11**	2	66	–0,31	**26**	5	79	0,25
12	~~3~~	~~66~~	~~–11,07~~	**27**	8	87	7,79	**12**	~~3~~	~~66~~	~~–11,07~~	**27**	6	81	4,69
13	~~3~~	~~65~~	~~–3,67~~	**28**	9	68	10,57	**13**	~~6~~	~~66~~	~~–4,35~~	**28**	14	81	–10,40
14	~~3~~	~~64~~	~~14,72~~	**29**	14	81	–10,40	**14**	~~9~~	~~68~~	~~10,57~~	**29**	8	87	7,79
15	~~3~~	~~71~~	~~–3,06~~	–				**15**	~~1~~	~~71~~	~~5,47~~	–			

7.1. Проверка гомоскедастичности по фактору x_2. Построим уравнение регрессии на основе верхней части (№ 1–11) данных с остатками, упорядоченными по x_2 (левая часть таблицы). Для этого используем функцию «Сервис. Анализ данных. Регрессия».

Результаты работы функции приведены в табл. 4.2–4.4.

Таблица 4.2

Результаты корреляционного анализа

Множественный *R*	0,330
R-квадрат	0,109
Нормированный *R*-квадрат	–0,114
Стандартная ошибка	5,756
Наблюдения	11

Таблица 4.3

Результаты дисперсионного анализа

Пояснения	***df***	***SS***	***MS***	***F***	**Значимости *F***
Регрессия	2	32,405	16,203	0,489	0,630
Остаток	8	265,051	33,131		
Итого	10	297,457			

Таблица 4.4

Результаты регрессионного анализа

Показатели	**Коэффициенты**	**Стандартная ошибка**	***t*-статистика**	***P*-значение**	**Нижние 95%-ные пределы**	**Верхние 95%-ные пределы**
Y-пересечение	2,709	6,936	0,391	0,706	–13,284	18,703
Переменная *X*1	3,123	3,595	0,869	0,410	–5,166	11,413
Переменная *X*2	–0,125	0,138	–0,910	0,389	–0,442	0,192

Из табл. 4.3 остаточная дисперсия (графа «*SS*»): S_1 = 265,05.

Аналогично определим остаточную дисперсию для уравнения регрессии на основе нижней части (№ 19–29) данных с остатками, упорядоченными по x_2 (левая часть табл. 4.1). Получаем S_2 = 624,21.

Далее найдем отношение S_2 / S_1 = 624,21/265,05 = 2,36.

Определим критическое значение для теста Гольдфельда–Квандта как значение *F*-критерия со степенями свободы $k_1 = (n - C - 2p)/2,\ k_2 = (n - C - 2p).$ В нашем случае *C* = 7 и *p* = 2 (переменные x_1 и x_2):

$$k_1 = k_2 = 29 - 7 - 2 \cdot 2 = 18.$$

Соответствующее значение критерия при α = 0,05 равно $F_{\text{факт } 0,05}$ = 2,217. Так как $S_2 / S_1 = 2{,}36 > F_{\text{факт } 0,05} = 2{,}217,$ то нарушается предпосылка о равенстве дисперсий, т. е. о гомоскедастичности остатков по переменной x_2.

7.2. Проверка гомоскедастичности по фактору x_3. Действуя аналогично и используя правую часть табл. 5.1 (данные с остатками, упорядоченными по x_3), получим следующие величины остаточных дисперсий S_1 = 403,39 и S_2 = 563,0.

Так как отношение $S_2 / S_1 = 563{,}0 / 403{,}39 = 1{,}40 < F_{\text{факт } 0,05} = 2{,}217,$ то предпосылка о равенстве дисперсий, т. е. о гомоскедастичности остатков по переменной x_3, не нарушается.

8. Построить уравнение регрессии с помощью метода взвешенных наименьших квадратов.

Так как предпосылка о гомоскедастичности остатков нарушается по переменной x_2, то корректировку исходных данных для расчета взвешенного уравнения регрессии будем проводить по результатам расчетов в пункте 7.1.

Необходимо оценить среднеквадратические ошибки по остаткам, упорядоченным по x_2, на основе верхней I_1 (№ 1–11) и нижней части I_2 (№ 19–29) данных в табл. 4.1 (левая часть таблицы). Для этого следует воспользоваться формулой (4.8):

$$\sigma_1 = \sqrt{\frac{1}{n}\sum_{i \in I_1} \varepsilon_i^2} \text{ и } \sigma_2 = \sqrt{\frac{1}{n}\sum_{i \in I_2} \varepsilon_i^2}. \quad (49)$$

Таким образом, получим: $\sigma_1 = 5{,}312$ и $\sigma_2 = 8{,}398$.

Далее проводим преобразование исходных данных путем деления независимой и зависимых переменных каждого наблюдения в верхней части данных табл. 5.1 на I_1 (№ 1–11) $\sigma_1 = 5{,}312$, в нижней I_2 (№ 19–29) – $\sigma_2 = 8{,}398$.

Таблица 4.5

Преобразованные данные для построения взвешенного уравнения регрессии

№	y	x_2	x_3	y/σ	x_2/σ	x_3/σ	№	y	x_2	x_3	y/σ	x_2/σ	x_3/σ
1	98	0	26	18,45	0,00	4,89	**16**	117	3	65	–	–	–
2	110	1	77	20,71	0,19	14,50	**17**	122	4	64	–	–	–
3	113	1	71	21,27	0,19	13,37	**18**	136	4	71	–	–	–
4	128	1	58	24,10	0,19	10,92	**19**	130	4	56	15,48	0,48	6,67
5	117	1	74	22,03	0,19	13,93	**20**	133	5	79	15,84	0,60	9,41
6	118	1	61	22,21	0,19	11,48	**21**	148	5	78	17,62	0,60	9,29
7	126	1	35	23,72	0,19	6,59	**22**	140	6	81	16,67	0,71	9,65
8	124	2	64	23,34	0,38	12,05	**23**	126	6	73	15,00	0,71	8,69
9	124	2	77	23,34	0,38	14,50	**24**	125	6	66	14,88	0,71	7,86
10	122	2	66	22,97	0,38	12,42	**25**	124	7	74	14,77	0,83	8,81
11	122	2	73	22,97	0,38	13,74	**26**	136	8	74	16,19	0,95	8,81
12	113	3	66	–	–	–	**27**	149	8	87	17,74	0,95	10,36
13	120	3	65	–	–	–	**28**	146	9	68	17,39	1,07	8,10
14	138	3	64	–	–	–	**29**	139	14	81	16,55	1,67	9,65
15	123	3	71	–	–	–	–						

Примечание. По данным с № 12 и по № 18 преобразование не осуществляется, однако они участвуют в построении взвешенного уравнения регрессии.

Далее осуществляется построение регрессионной модели по преобразованным данным в табл. 4.5. Для этого используем функцию «Сервис. Анализ данных. Регрессия».

Результаты работы функции приведены в табл. 4.6–4.8.

Таблица 4.6

Результаты корреляционного анализа

Множественный *R*	0,992
R-квадрат	0,985
Нормированный *R*-квадрат	0,984
Стандартная ошибка	5,893
Наблюдения	29

Таблица 4.7

Результаты дисперсионного анализа

Пояснения	*df*	*SS*	*MS*	*F*	Значимости *F*
Регрессия	2	58310,76	29155,38	839,540	2,410E–24
Остаток	26	902,922	34,728		
Итого	28	59213,68			

Таблица 4.8

Результаты регрессионного анализа

Показатели	Коэффициенты	Стандартная ошибка	*t*-статистика	*P*-значение	Нижние 95%-ные пределы	Верхние 95%-ные пределы
Y-пересечение	0,345	1,543	0,224	0,825	–2,826	3,517
Переменная X1	1,916	2,507	0,764	0,452	–3,237	7,070
Переменная X2	1,757	0,128	13,699	2,11E–13	1,493	2,021

Таким образом, уравнение регрессии, построенное с учетом гетероскедастичности по независимой переменной x_2 имеет вид:

$$y = 0{,}345 + 1{,}916x_2 + 1{,}757x_3.$$

5. МЕТОДЫ ОБНАРУЖЕНИЯ И УСТРАНЕНИЯ АВТОКОРРЕЛЯЦИИ

5.1. Обнаружение автокорреляции методом рангов

Этот метод достаточно прост: последовательно определяются знаки отклонений e_t. Например, (– – – – –)(+ + + + + + +)(– – –)(+ + + +)(–), т. е. 5 «–», 7 «+», 3 «–», 4 «+», 1 «–» при 20 наблюдениях.

Ряд определяется как непрерывная последовательность одинаковых знаков. Количество знаков в ряду называется *длиной ряда*.

Если рядов слишком мало по сравнению с количеством наблюдений *n*, то вполне вероятна положительная автокорреляция. Если же рядов слишком много, то вероятна отрицательная автокорреляция.

Для более детального анализа предлагается следующая процедура. Пусть n – объем выборки; n_1 – общее количество знаков «+» при n наблюдениях (количество положительных отклонений e_t); n_2 – общее количество знаков «–» при n наблюдениях (количество отрицательных отклонений e_t); k – количество рангов.

При достаточно большом количестве наблюдений $(n_1 > 10,$ $n_2 > 20)$ k имеет асимптоматически нормальное распределение с:

$$M\,(k) = \frac{2n_1n_2}{n_1+n_2} + 1; \tag{5.1}$$

$$D\,(k) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2)^2(n_1+n_2-1)}. \tag{5.2}$$

Для небольшого числа наблюдений $(n_1 < 10,\ n_2 < 20)$ Свед и Эйзенхарт разработали таблицы критических значений количества ряда при n наблюдениях (см. Приложение, табл. П.1.3 и П.1.4). Суть таблиц в следующем.

На пересечении строки n_1 и столбца n_2 определяются нижнее k_1 и верхнее k_2 значения при уровне значимости α = 0,05. Если $k_1 < k < k_2$, то говорят об отсутствии автокорреляции. Если $k \le k_1$, – положительной автокорреляции. Если $k \le k_2$, – отрицательной автокорреляции.

В нашем примере n = 20, n_1 = 11, n_2 = 9, k = 5. По табл. П.1.3 и П.1.4 из Приложения определяем k_1 = 6, k_2 = 16. Поскольку $k = 5 < 6 = k_1$, то принимается предположение о наличии положительной автокорреляции при уровне значимости α = 0,05.

5.2. Обнаружение автокорреляции по критерию Дарбина–Уотсона

Наиболее известным критерием обнаружения автокорреляции первого порядка является критерий Дарбина–Уотсона. Суть метода состоит в том, что на основе вычисленной статистики DW Дарбина–Уотсона делается вывод об автокорреляции:

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}. \tag{5.3}$$

Согласно формуле (5.3), статистика Дарбина–Уотсона тесно связана с выборочным коэффициентом корреляции $r_{e_t e_{t-1}}$:

$$DW \approx 2(1 - r_{e_t e_{t-1}}). \qquad (5.4)$$

Таким образом, $0 \le DW \le 4$, и ее значения могут указать на наличие либо отсутствие автокорреляции. Действительно, если $r_{e_t e_{t-1}} \approx 0$ (автокорреляция отсутствует), то $DW \approx 2$. Если $r_{e_t e_{t-1}} \approx 1$ (положительная автокорреляция), то $DW \approx 0$. Если $r_{e_t e_{t-1}} \approx -1$ (отрицательная автокорреляция), то $DW \approx 4$.

Для более точного определения, какое значение *DW* свидетельствует об отсутствии автокорреляции, а какое – об ее наличии, была построена таблица критических точек распределения Дарбина–Уотсона. По ней для заданного уровня значимости α, числа наблюдений *n* и количества объясняющих переменных *m* определяются два значения: d_1 – нижняя граница и d_u – верхняя граница.

Схема использования критерия Дарбина–Уотсона следующая:

1. По построенному эмпирическому уравнению регрессии $\hat{y}_t = b_0 + b_1 x_{t1} + ... + b_m x_{tm}$ определяются значения отклонений $e_t = y_t - \hat{y}_t$ для каждого наблюдения *t*, *t* = 1, 2, …, *T*.

2. По формуле (5.3) рассчитывается статистика *DW*.

По таблице критических точек Дарбина–Уотсона (см. Приложение, табл. П.1.5 и П.1.6) определяются два числа d_1 и d_u.

Далее делаются выводы по правилу:

$0 \le DW < d_1$ – существует положительная автокорреляция;

$d_1 < DW < d_u$ – вывод о наличии автокорреляции не определен;

$d_u \le DW < 4 - d_u$ – автокорреляция отсутствует;

$4 - d_u \le DW < 4 - d_1$ – наличие автокорреляции не определено;

$4 - d_1 \le DW \le 4$ – существует отрицательная автокорреляция.

Отметим, что при использовании критерия Дарбина–Уотсона необходимо учитывать следующие ограничения:

1. Критерий *DW* применяется лишь для тех моделей, которые содержат свободный член.

2. Предполагается, что случайные отклонения ε_t определяются по итерационной схеме: $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$, называемой авторегрессионной схемой первого порядка *AR*(1). Здесь v_t – случайный член.

3. Статистические данные должны иметь одинаковую периодичность (т. е. не должно быть пропусков в наблюдениях).

4. Критерий Дарбина–Уотсона не применим для регрессионных моделей, содержащих в составе объясняющих переменных зависимую переменную с временным лагом в один период, т. е. для так называемых авторегрессионных моделей вида:

$$y_t = \beta_0 + \beta_0 x_{t1} + \ldots + \beta_m x_{tm} + \gamma y_{t-1} + \varepsilon_t. \tag{5.5}$$

Для авторегрессионных моделей (5.5) разработаны специальные тесты обнаружения автокорреляции, в частности h-статистика Дарбина, которая определяется по формуле (5.6):

$$h = \hat{\rho}\sqrt{\frac{n}{1 - nD(g)}}, \tag{5.6}$$

где $\hat{\rho}$ – оценка ρ авторегрессии первого порядка; n – число наблюдений; $D(g)$ – выборочная дисперсия коэффициента при переменной y_{t-1}.

Обычно значение $\hat{\rho}$ рассчитывается по формуле (5.7):

$$\hat{\rho} = 1 - 0{,}5D(g). \tag{5.7}$$

Поэтому h легко вычисляется на основе оцененной регрессии.

Основная проблема при использовании этого теста заключается в невозможности вычисления h при $nD(g) > 1$.

5.3. Авторегрессионное преобразование

Возможно, автокорреляция вызвана отсутствием в модели некоторой важной объясняющей переменной. Следует попытаться определить данный фактор и учесть его в уравнении регрессии. Также можно попробовать изменить формулу зависимости.

Если все процедуры изменения спецификации модели исчерпаны, а автокорреляция имеет место, то можно предположить, что она обусловлена внутренними свойствами ряда $\{e_t\}$. В этом случае можно воспользоваться авторегрессионным преобразованием. В линейной регрессионной модели наиболее простым преобразованием является *авторегрессионная схема первого порядка* $AR(1)$.

Рассмотрим модель парной линейной регрессии (5.8):

$$y = \beta_0 + \beta_1 x + \varepsilon. \tag{5.8}$$

Наблюдениям t и $(t-1)$ соответствуют формулы (5.9) и (5.10):

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t; \tag{5.9}$$

$$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1}. \tag{5.10}$$

Пусть отклонения подвержены воздействию авторегрессии первого порядка $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$, где v_t – отклонения, удовлетворяющие предпосылкам МНК, а коэффициент ρ известен.

Вычтем из (5.9) соотношение (5.10), умноженное на ρ:

$$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \beta_1(x_t - \rho x_t) + (\varepsilon_t - \rho\varepsilon_t). \tag{5.11}$$

Если $y_t^* = y_t - \rho y_{t-1}$, $x_t^* = x_t - \rho x_{t-1}$, $\beta_0^* = \beta_0(1-\rho)$, получаем:

$$y_t^* = \beta_t^* + \beta_1 x_t^* + v_t. \tag{5.12}$$

Так как по предположению коэффициент ρ известен, то очевидно y_t^*, x_t^*, v_t^* вычисляются достаточно просто. В силу того, что случайные отклонения v_t удовлетворяют предпосылкам МНК, оценки β_0^* и β_1 будут обладать свойствами наилучших линейных несмещенных оценок.

Однако способ вычисления y_1^*, x_1^* приводит к потере первого наблюдения. Число степеней свободы уменьшится на единицу, что при малых выборках может привести к потере эффективности. Эта проблема преодолевается с помощью *поправки Прайса–Винстена*:

$$x_1^* = \sqrt{1-\rho^2} \cdot x_1, \quad y_1^* = \sqrt{1-\rho^2} \cdot y_1. \tag{5.13}$$

Авторегрессионное преобразование первого порядка $AR(1)$ может быть обобщено на преобразования более высоких порядков $AR(2)$, $AR(3)$ и т. д.: $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + v_t$, $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \rho_3\varepsilon_{t-3} + v_t$.

Однако на практике значение коэффициента ρ обычно неизвестно. Существуют различные методы его оценивания. Рассмотрим подход к определению ρ на основе статистики Дарбина–Уотсона.

Статистика Дарбина–Уотсона тесно связана с коэффициентом корреляции между соседними отклонениями через соотношение $DW \approx 2(1 - r_{e_t e_{t-1}})$. Тогда в качестве оценки коэффициента ρ может быть взят коэффициент $r = r_{e_t e_{t-1}}$. Тогда имеем:

$$r \approx 1 - \frac{DW}{2}. \quad (63)$$

Этот метод оценивания хорош при большом числе наблюдений. В этом случае оценка r параметра ρ будет достаточно точной.

Расчетная работа № 5
Реализация методов обнаружения и устранения автокорреляции

Задание. На основании данных Приложения (см. табл. П.1.1 и П.1.2) для соответствующего варианта (см. табл. 2.1):

1. Проверить наличие автокорреляции методом рангов.
2. Проверить наличие автокорреляции.
3. Провести авторегрессионное преобразование с учетом поправки Прайса–Винстена.

Указания к решению. При выполнении лабораторной работы использовать возможности надстройки «Анализ данных» табличного процессора MS Excel.

6. МОДЕЛИ И МЕТОДЫ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

6.1. Автокорреляция уровней временного ряда

Временной ряд (ряд динамики) – это совокупность значений показателя за несколько последовательных периодов времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы: 1) факторы, формирующие тенденцию ряда; 2) факторы, формирующие циклические колебания ряда; 3) случайные факторы.

Фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется *аддитивной моделью временного ряда*. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется мультипликативной моделью временного ряда. Основная задача исследования временного ряда – выявление и придание количественного выражения каждой из перечисленных выше компонент с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда.

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют *автокорреляцией уровней ряда*.

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Формула для расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^{n}\left(y_t - \bar{y}_1\right)\left(y_{t-1} - \bar{y}_2\right)}{\sqrt{\sum_{t=2}^{n}\left(y_t - \bar{y}_1\right)^2 \sum_{t=2}^{n}\left(y_{t-1} - \bar{y}_2\right)^2}}, \quad (6.1)$$

где $\bar{y}_1 = \frac{1}{n-1}\sum_{t=2}^{n} y_t, \; \bar{y}_2 = \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1}.$ (6.2)

Эту величину называют *коэффициентом автокорреляции уровней ряда первого порядка*, так как он измеряет зависимость между соседними уровнями ряда y_t и y_{t-1}.

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-2} и определяется по формуле (6.3):

$$r_2 = \frac{\sum_{t=3}^{n}\left(y_t - \bar{y}_3\right)\left(y_{t-2} - \bar{y}_4\right)}{\sqrt{\sum_{t=3}^{n}\left(y_t - \bar{y}_3\right)^2 \sum_{t=3}^{n}\left(y_{t-2} - \bar{y}_4\right)^2}}, \quad (6.3)$$

где $\bar{y}_3 = \frac{1}{n-2}\sum_{t=3}^{n} y_t, \; \bar{y}_4 = \frac{1}{n-2}\sum_{t=3}^{n} y_{t-2}.$ (6.4)

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют *лагом*. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается целесообразным использовать правило – максимальный лаг должен быть не больше $n/4$.

Последовательность коэффициентов автокорреляции уровней первого, второго и т. д. порядков называют *автокорреляционной функцией временного ряда*. График зависимости ее значений от величины лага называется *коррелограммой.*

Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, а следовательно, и лаг, при котором связь между текущим и предыдущими уровнями ряда наиболее тесная.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка k, то ряд содержит циклические колебания с периодичностью в k моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений относительно структуры этого ряда:

– либо ряд не содержит тенденции и циклических колебаний;

– либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

Поэтому коэффициент автокорреляции уровней и автокорреляционную функцию целесообразно использовать для выявления во временном ряде наличия или отсутствия трендовой компоненты и циклической (сезонной) компоненты.

6.2. Моделирование тенденций 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: $\widehat{y}_t = a + bt;$ (6.5)

– гипербола: $\widehat{y}_t = a + \dfrac{b}{t};$ (6.6)

– экспоненциальный тренд: $\widehat{y}_t = e^{a+bt}$ (или $\widehat{y}_t = ab^t$); (6.7)

– степенная функция $\hat{y}_t = at^b$; (6.8)

– полиномы: $\hat{y}_t = a + b_1 t + b_1 t^2 + ... + b_m t^m$. (6.9)

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 $\hat{y}_t$. Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

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 $\hat{y}_t$ и $\hat{y}_{t-1}$ 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.

6.3. Моделирование сезонных колебаний

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:

$$Y = T + S + E. \quad (6.10)$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма *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 T, S и E для каждого уровня временного 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 (TE) \в мультипликативной модели.

4. Аналитическое выравнивание уровней ($T + E$) или (TE) и расчет значений T с использованием полученного уравнения тренда.

5. Расчет полученных по модели значений ($T + E$) или (TE).

6. Прогноз будущих значений уровней временного ряда на основе построенной модели.

6.4. Автокорреляция в остатках. Критерий Дарбина–Уотсона

Автокорреляция в остатках может быть вызвана несколькими причинами, имеющими различную природу.

1. Она может быть связана с исходными данными и вызвана наличием ошибок измерения в значениях результативного признака.

2. В ряде случаев автокорреляция может быть следствием неправильной спецификации модели. Модель может не включать фактор, который оказывает существенное воздействие на результат и влияние которого отражается в остатках, вследствие чего последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени t.

От истинной автокорреляции остатков следует отличать ситуации, когда причина автокорреляции заключается в неправильной спецификации функциональной формы модели. В этом случае следует изменить форму модели, а не использовать специальные методы расчета параметров уравнения регрессии при наличии автокорреляции в остатках.

Один из более распространенных методов определения автокорреляции в остатках – это расчет критерия Дарбина–Уотсона (6.11):

$$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2}. \qquad (6.11)$$

Величина d – отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели.

Можно показать, что при больших значениях n существует следующее соотношение между критерием Дарбина–Уотсона d и коэффициентом автокорреляции остатков первого порядка r_1:

$$d \cong 2 \times (1 - r_1). \qquad (6.12)$$

Таким образом, если в остатках существует полная положительная автокорреляция и $r_1 = 1$, то $d = 0$. Если в остатках полная отрицательная автокорреляция, то $r_1 = -1$ и, следовательно, $d = 4$. Если автокорреляция остатков отсутствует, то $r_1 = 0$, то $d = 2$. То есть $0 \le d \le 4$.

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина–Уотсона следующий. Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H_1^* состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам (см. Приложение, табл. П.1.5 и П.1.6) определяются критические значения критерия Дарбина–Уотсона d_L и d_U для заданного числа наблюдений n, числа независимых переменных модели m и уровня значимости α. По этим значениям числовой промежуток [0; 4] разбивают на пять отрезков.

Принятие или отклонение каждой из гипотез с вероятностью $1-\alpha$ осуществляется следующим образом:

$0 < d < d_L$ – есть положительная автокорреляция остатков, H_0 отклоняется, с вероятностью $P = 1-\alpha$ принимается H_1;

$d_L < d < d_U$ – зона неопределенности;

$d_U < d < 4 - d_U$ – нет оснований отклонять H_0, т. е. автокорреляция остатков отсутсвует;

$4–d_U < d < 4–d_L$ – зона неопределенности;

$4–d_L < d < 4$ – есть отрицательная автокорреляция остатков, H_0 отклоняется, с вероятностью $P = 1 - \alpha$ принимается H_1^*.

Если фактическое значение критерия Дарбина–Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу H_0.

Существует несколько ограничений на применение критерия Дарбина–Уотсона:

1. Он неприменим к моделям, включающим в качестве независимых переменных лаговые значения результативного признака.
2. Методика расчета и использования критерия Дарбина–Уотсона направлена только на выявление автокорреляции остатков первого порядка.
3. Критерий Дарбина–Уотсона дает достоверные результаты только для больших выборок.

Расчетная работа № 6
Анализ временных рядов

Задание. Имеются условные данные об объемах потребления электроэнергии (y_t) жителями региона за 16 кварталов. На основании данных табл. 6.1 для соответствующего варианта:

1. Рассчитать автокорреляцию уровней ряда. Построить автокорреляционную функцию и сделать выводы о наличии сезонных колебаний.
2. Построить аддитивную модель временного ряда.
3. Построить мультипликативную модель временного ряда.
4. Сделать прогноз уровней временного ряда на два квартала вперед по аддитивной и мультипликативной моделям.
5. Проверить наличие автокорреляции в остатках по критерию Дарбина–Уотсона.

Таблица 6.1

Исходные данные для выполнения работы № 6

t	Значения y_t для соответствующих вариантов									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	5,8	5,5	5,3	5,5	5,6	5,4	5,7	5,4	5,3	5,5
2	4,5	4,6	4,7	4,8	4,7	4,5	4,6	4,5	4,6	4,8
3	5,1	5,0	5,2	5,1	5,2	5,1	5,1	5,0	5,0	5,3

Окончание табл. 6.1

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

Указания к решению. При выполнении лабораторной работы использовать возможности надстройки «Анализ данных» табличного процессора MS Excel.

Пример выполнения работы. Пусть имеются некоторые условные данные об общем количестве правонарушений на таможне.

Таблица 6.2

Исходные значения временного ряда

Год	Квартал	t	Количество возбужденных дел, y_t	Год	Квартал	t	Количество возбужденных дел, y_t
2012	I	**1**	375	2014	I	**9**	390
	II	**2**	371		II	**10**	355
	III	**3**	869		III	**11**	992
	IV	**4**	1015		IV	**12**	905
2013	I	**5**	357	2015	I	**13**	461
	II	**6**	471		II	**14**	454
	III	**7**	992		III	**15**	920
	IV	**8**	1020		IV	**16**	927

1. Рассчитаем несколько последовательных коэффициентов автокорреляции. Составляем первую вспомогательную табл. 6.3.

Таблица 6.3

Расчет коэффициентов корреляции первого порядка

t	y_t	y_{t-1}	$y_t-\bar{y}_1$	$y_{t-1}-\bar{y}_2$	$(y_t-\bar{y}_1)\times$ $\times(y_{t-1}-\bar{y}_2)$	$(y_t-\bar{y}_1)^2$	$(y_{t-1}-\bar{y}_2)^2$
1	375	–	–	–	–	–	–
2	371	375	–328,93	–288,13	94774,60	108194,94	83018,90
3	869	371	169,7	–292,13	–49390,42	28584,66	85339,94
4	1015	869	315,07	205,87	64863,46	99269,10	42382,46
5	357	1015	–342,93	351,87	–120666,78	117600,98	12381250
6	471	357	–228,93	–306,13	70082,34	52408,94	93715,58
7	992	471	292,07	–192,13	–56115,41	85304,88	36913,94
8	1020	992	320,07	328,87	105261,42	102444,80	108155,48
9	390	1020	–309,93	356,87	–110604,72	95056,60	127356,20
10	355	390	–344,93	–273,13	94210,73	118976,70	74600,00
11	992	355	292,07	–308,13	–89995,53	85304,88	94944,10
12	905	992	205,07	328,87	67441,37	42053,70	108155,48
13	461	905	–238,93	241,87	–57790,00	57087,54	58501,10
14	454	461	–245,93	–202,13	49709,83	60481,56	40856,54
15	920	454	220,07	–209,13	–46023,24	48430,80	43735,36
16	927	920	227,07	256,87	58327,47	51560,78	65982,20
Сумма	10499	9947	0,05	0,05	74085,13	1153760,93	1187469,73
Среднее значение	699,93	663,13	–	–	–	–	–

Следует заметить, что среднее значение получается путем деления не на 16, а на 15, так как у нас теперь на одно наблюдение меньше.

Вычисляем коэффициент автокорреляции первого порядка:

$$r=\frac{74085}{\sqrt{1153760{,}39\cdot 1187469{,}73}}=0{,}063294.$$

Составляем вспомогательную табл. 6.4 для расчета коэффициента автокорреляции второго порядка.

Следовательно: $r=\dfrac{-1034792{,}71}{\sqrt{1037835{,}43\cdot 1116776{,}36}}=-0{,}961183.$

Таблица 6.4

Расчет коэффициентов корреляции второго порядка

t	y_t	y_{t-2}	$y_t - \bar{y}_3$	$y_{t-2} - \bar{y}_4$	$(y_t - \bar{y}_3) \times$ $\times (y_{t-2} - \bar{y}_4)$	$(y_t - \bar{y}_3)^2$	$(y_{t-2} - \bar{y}_4)^2$
1	375	–	–	–	–	–	–
2	371	–	–	–	–	–	–
3	869	375	145,57	–269,79	–39273,23	21190,62	72786,64
4	1015	371	291,57	–273,79	–79828,95	85013,06	74960,96
5	357	869	–366,43	224,21	–82157,27	134270,94	50270,12
6	471	1015	–252,43	370,21	–93452,11	63720,90	137055,44
7	992	357	268,57	–287,29	–77291,76	72129,84	82823,08
8	1020	471	296,57	–173,79	–51540,90	87953,76	30202,96
9	390	992	–333,43	347,21	–115770,23	111175,56	120554,78
10	355	1020	–368,43	375,21	–138,238,62	135740,66	140782,54
11	992	390	268,57	–254,79	–68428,95	72129,84	64917,94
12	905	355	181,57	–289,79	–52617,17	32967,66	83978,24
13	461	992	–262,43	347,21	–91118,32	68869,50	120554,78
14	454	905	–269,43	260,21	–70108,38	75592,52	67709,24
15	920	461	196,57	–183,79	–37127,60	38639,76	33778,76
16	927	454	203,57	–190,79	–38839,12	41440,74	36400,82
Сумма	10499	9027	–0,02	–0,06	–1034792,71	1037835,43	1116776,36
Среднее значение	699,93	644,79	–	–	–	–	–

Аналогично находим коэффициенты автокорреляции более высоких порядков, полученные значения заносим в табл. 6.5.

Таблица 6.5

Коэффициенты автокорреляции уровней ряда

Лаг	**Коэффициент автокорреляции уровней**	**Лаг**	**Коэффициент автокорреляции уровней**	**Лаг**	**Коэффициент автокорреляции уровней**
1	0,063294	**5**	0,050594	**9**	0,162064
2	–0,961183	**6**	–0,976516	**10**	–0,972918
3	–0,036290	**7**	0,069444	**11**	–0,065323
4	0,964735	**8**	0,964629	**12**	0,985761

Анализ табл. 6.5 позволяет сделать вывод о наличии в изучаемом ряде сезонных колебаний периодичностью в четыре квартала.

2. Построение аддитивной модели временного ряда. Обратимся к данным об объеме правонарушений на таможне за четыре года, представленным в табл. 6.2.

Было показано, что данный временной ряд содержит сезонные колебания периодичностью 4, так как количество правонарушений в первый–второй квартал ниже, чем в третий–четвертый. Рассчитываем компоненты аддитивной модели временного ряда.

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней. Для этого:

1.1. Просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени и определим условные годовые объемы потребления электроэнергии (гр. 3 табл. 6.6.).

1.2. Разделив полученные суммы на 4, найдем скользящие средние (гр. 5 табл. 6.6). Полученные таким образом выровненные значения уже не содержат сезонной компоненты.

1.3. Приведем эти значения в соответствии с фактическими моментами времени, для чего найдем средние значения из двух последовательных скользящих средних – центрированные скользящие средние (гр. 5 табл. 6.6).

Таблица 6.6

Выравнивание уровней временного ряда (аддитивная модель)

Номер квартала, t	**Количество правонарушений, y_t**	**Итого за четыре квартала**	**Скользящая средняя за четыре квартала**	**Центрированная скользящая средняя**	**Оценка сезонной компоненты**
1	375	–	–	–	–
2	371	2630	657,5	–	–
3	869	2612	653	655,25	213,75
4	1015	2712	678	665,5	349,5
5	357	2835	708,75	693,375	–336,375
6	471	2840	710	709,375	–238375
7	992	2873	718,25	714,125	277,875
8	1020	2757	689,25	703,75	316,25
9	390	2757	689,25	689,25	–299,25
10	355	2642	660,5	674,875	–319,875
11	992	2713	678,25	669,375	322,625
12	905	2812	703	690,625	214,375
13	461	2740	685	694	–233
14	454	2762	690,5	687,75	–233,75
15	920	–	–	–	–
16	927	–	–	–	–

Шаг 2. Найдем оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними (гр. 6 табл. 6.6). Используем эти оценки для расчета значений сезонной компоненты S_i (табл. 6.7). Для этого найдем средние за каждый квартал (по всем годам) оценки сезонной компоненты S_i.

В моделях с сезонной компонентой обычно предлагается, что сезонные воздействия за период взаимопогашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю.

Таблица 6.7

Оценка сезонной компоненты уровней временного ряда (аддитивная модель)

Показатели	Год	Номер квартала, *i*			
		I	II	III	IV
Период	2012	–	–	213,75	349,5
	2013	–336,375	–238,375	277,875	316,25
	2014	–299,25	–319,875	322,625	214,375
	2015	–233	–233,75	–	–
Всего за *i*-й квартал		–868,625	–792	814,25	880,125
Средняя оценка сезонной компоненты для *i*-го квартала, S_i		–289,542	–264	271,417	293,375
Скорректированная компонента, S_i		–292,355	–266,813	268, 604	290,562

Для модели имеем: –289,542 – 264 + 271,417 + 293,375 = 11,25.

Корректирующий коэффициент: k = 11,25/4 = 2,813.

Рассчитываем скорректированные значения сезонной компоненты $\left(S_i = \overline{S}_i - k\right)$ и заносим полученные данные в табл. 6.7.

Проверим равенство нулю суммы значений сезонной компоненты: –292,355 – 266,813 + 268, 604 + 290,562 = 0,00.

Шаг 3. Исключим влияние сезонной компоненты, вычитая ее значения из каждого уровня исходного временного ряда. Получим величины $T + E = Y - S$ (гр. 4 табл. 6.8). Эти значения рассчитываются за каждый момент времени и содержат только тенденцию и случайную компоненту.

Шаг 4. Определим компоненту T модели. Для этого проведем аналитическое выравнивание ряда ($T + E$) с помощью линейного тренда. Результаты аналитического выравнивания следующие:

$$T = 671{,}759 + 0{,}9255t.$$

Подставляя в это уравнение значения t = 1, 2, …, 16, найдем уровни T для каждого момента времени (гр. 5 табл. 6.8).

Таблица 6.8

Исключение сезонной компоненты из уровней временного ряда (аддитивная модель)

t	y_t	S_i	$y_t - S_i$	T	$T + S$	$E = y_t - (T + S)$	E^2
1	375	–292,355	667,355	672, 684	380,329	–5,329	28,3982
2	371	–266,813	637,813	673,610	406,797	–35,797	1281,425
3	869	268,604	600,396	674,535	943,139	–74,139	5496,591
4	1015	290,562	724,438	675,461	966,023	48,977	2398,747
5	357	–292,355	649,355	676,386	384,031	–27,031	730,675
6	471	–266,813	737,813	677,312	410,499	60,501	3660,371
7	992	268,604	723,396	678,237	946,841	45,159	2039,335
8	1020	290,562	729,438	679,163	969,725	50,275	2527,576
9	390	–292,355	682,355	680,088	387,733	2,267	5139,289
10	355	–266,813	621,813	681,014	414,201	–59,201	3504,758
11	992	268,604	723,396	681,939	950,543	41,457	1718,683
12	905	290,562	614,438	682, 865	973,427	–68,427	4682,254
13	461	–292,355	753,355	683,790	391,435	69,565	4839,289
14	454	–266,813	720,813	684,716	417,903	36,097	1302,993
15	920	268,604	651,396	685,641	954,245	–34,245	1172,720
16	927	290,562	636,438	686,567	977,129	–50,129	2512,917

Шаг 5. Найдем значения уровней ряда, полученные по аддитивной модели. Для этого прибавим к уровням T значения сезонной компоненты для соответствующих кварталов (гр. 6 табл. 6.8).

Для оценки качества построенной модели применим сумму квадратов полученных абсолютных ошибок:

$$R^2 = 1 - \frac{\sum E^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{37901,872}{1252743,75} = 0,970.$$

Следовательно, можно сказать, что аддитивная модель объясняет 97 % общей вариации уровней временного ряда количества правонарушений по кварталам за 4 года.

Шаг 6. Прогнозирование по аддитивной модели. Предположим, что по нашему примеру необходимо дать прогноз об общем объеме правонарушений на 1 и 2 кварталы 2016 года. Прогнозное значение F_t уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда: $T = 671,777 + 0,9233t$.

Получим: $T_{17} = 671,777 + 0,9233 \cdot 17 = 687,473$; $T_{18} = 671,777 + 0,9233 \cdot 18 = 688,396$.

Значения сезонных компонент за соответствующие кварталы равны: $S_1 = -292,448$ и $S_2 = -266,781$. Таким образом, $F_{17} = T_{17} + S1 = 687,473 - 292,448 \approx 395$; $F_{18} = T_{18} + S_2 = 688,396 - 266,781 \approx 422$.

То есть в первые два квартала 2016 года следует ожидать около 395 и 422 правонарушений соответственно.

Рассмотрим построение мультипликативной модели.

Шаг 1. Методика, применяемая на этом шаге, полностью совпадает с методикой построения аддитивной модели (табл. 6.9).

Таблица 6.9

Выравнивание уровней ряда (мультипликативная модель)

Номер квартала, t	Количество правонарушений, y_t	Итого за четыре квартала	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	375	–	–	–	–
2	371	2630	657,5	–	–
3	869	2612	653	655,25	1,3262
4	1015	2712	678	665,5	1,5252
5	357	2835	708,75	693,375	0,5149
6	471	2840	710	709,375	0,6640
7	992	2873	718,25	714,125	1,3891
8	1020	2757	689,25	703,75	1,4494
9	390	2757	689,25	689,25	0,5658
10	355	2642	660,5	674,875	0,5260
11	992	2713	678,25	669,375	1,4820
12	905	2812	703	690,625	1,3104
13	461	2740	685	694	0,6643
14	454	2762	690,5	687,75	0,6601
15	920	–	–	–	–
16	927	–	–	–	–

Шаг 2. Найдем оценки сезонной компоненты как частное от деления фактических уровней ряда на центрированные скользящие средние (гр. 6 табл. 6.9). Эти оценки используются для расчета сезонной компоненты S (табл. 6.10). Для этого найдем средние за каждый квартал оценки сезонной компоненты S_i. Так же, как и в аддитивной модели считается, что сезонные воздействия за период взаимопогашаются. В мультипликативной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна

быть равна числу периодов в цикле. В нашем случае число периодов одного цикла равно 4.

Имеем: 0,5816 + 0,6167 + 1,3991 + 1,4283 = 4,0257. Определяем корректирующий коэффициент: k = 4/4,0257 = 0,9936.

Скорректированные значения сезонной компоненты S_i получаются при умножении ее средней оценки на коэффициент k.

Проверяем условие равенства «4» сумме значений сезонной компоненты: 0,5779 + 0,6128 + 1,3901 +1,4192 = 4.

Таблица 6.10

Оценка сезонной компоненты уровней временного ряда (мультипликативная модель)

Показатели	Год	Номер квартала, i			
		I	II	III	IV
Период	1999	–	–	1,3262	1,5252
	2000	0,5149	0,6640	1,3891	1,4494
	2001	0,5658	0,5260	1,4820	1,3104
	2002	0,6643	0,6601	–	–
Всего за i-й квартал		1,7450	1,8501	4,1973	4,2850
Средняя оценка сезонной компоненты для i-го квартала, S_i		0,5816	0,6167	1,3991	1,4283
Скорректированная компонента, S_i		0,5779	0,6128	1,3901	1,4192

Шаг 3. Разделим каждый уровень ряда на соответствующие значения сезонной компоненты. Получим величины $TE = Y / S$ (гр. 4 табл. 6.11), которые содержат тенденцию и случайную компоненту.

***Шаг 4*.** Определим компоненту T в мультипликативной модели.

Для этого рассчитаем параметры линейного тренда, используя уровни TE. В результате получим уравнение тренда: T = 651,6354 + + 3,2809t.

Подставляя в это уравнение значения t = 1, 2, …, 16, найдем уровни T для каждого момента времени (гр. 5 табл. 6.11).

***Шаг 5*.** Найдем уровни ряда, умножив значения T на соответствующие значения сезонной компоненты (гр. 5 табл. 6.11). На одном графике откладываем фактические значения уровней временного ряда и теоретические, полученные по мультипликативной модели.

Расчет ошибки в мультипликативной модели и других моделей временного ряда можно, по аналогии с аддитивной моделью, использовать сумму квадратов абсолютных ошибок $(y_t - TS)^2$:

$$R^2 = 1 - \frac{\sum (y_t - TS)^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{43065{,}02}{1252743{,}75} = 0{,}966.$$

Таблица 6.11

Исключение сезонной компоненты из уровней временного ряда (мультипликативная модель)

t	y_t	S_i	y_t / S_i	T	TS	$E = y_t /(T + S)$
1	375	0,5779	648,9012	654,9173	378,4767	0,9908
2	371	0,6128	605,4178	658,1982	403,3439	0,9198
3	869	1,3901	625,1349	661,4791	919,5221	0,9451
4	1015	1,4192	715,1917	664,7600	943,4274	1,0759
5	357	0,5779	617,7539	668,0409	386,0608	0,9247
6	471	0,6128	768,6031	671,3218	411,3860	1,1449
7	992	1,3901	713,6177	674,6027	937,7652	1,0578
8	1020	1,4192	718,7148	677,8836	962,0524	1,0602
9	390	0,5779	674,8572	681,1645	393,6450	0,9907
10	355	0,6128	579,3081	684,4454	419,4281	0,8464
11	992	1,3901	713,6177	687,7263	956,0083	1,0377
12	905	1,4192	637,6832	691,0072	980,6774	0,9228
13	461	0,5779	797,7159	694,2881	401,2291	1,1490
14	454	0,6128	740,8616	697,5690	427,4703	1,0621
15	920	1,3901	661,8229	700,8499	974,2515	0,9443
16	927	1,4192	653,1849	704,1308	999,3024	0,9277

Сравнивая показатели детерминации аддитивной и мультипликативной моделей, делаем вывод, что они примерно одинаково аппроксимируют исходные данные.

Шаг 6. Прогнозирование по мультипликативной модели. Прогнозное значение F_t уровня временного ряда в мультипликативной модели есть произведение трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда $T = 651{,}6364 + 3{,}2809t$. Получим: $T_{17} = 651{,}6364 + 3{,}2809 \cdot 17 = 707{,}4117$; $T_{18} = 651{,}6364 + 3{,}2809 \cdot 18 = 710{,}6926$.

Значения сезонных компонент за соответствующие кварталы равны: S_1 = 0,5779 и S_2 = 0,6128. Таким образом $F_{17} = T_{17} \cdot S_1 =$ $= 707{,}4117 \cdot 0{,}5779 \approx 409$; $F_{18} = T_{18} \cdot S_2 = 710{,}6926 \cdot 0{,}6128 \approx 436$.

То есть в первые 2 квартала 2016 г. следовало ожидать около 409 и 436 правонарушений соответственно.

Таким образом, аддитивная и мультипликативная модели дают примерно одинаковый результат по прогнозу.

РЕКОМЕНДОВАННАЯ ЛИТЕРАТУРА

Основная

1. Белько, И. В. Эконометрика. Практикум : учеб. пособие / И. В. Белько, Е. А. Криштапович. – Минск : Изд-во Гревцова, 2011. – 224 с.

2. Бородич, С. А. Эконометрика : учеб. пособие / С. А. Бородич. – 2-е изд., испр. – Минск : Новое знание, 2004. – 416 с.

3. Кремер, Н. Ш. Эконометрика : учебник / Н. Ш. Кремер, Б. А. Путко ; под ред. Н. Ш. Кремера. – 2-е изд. – М. : ЮНИТИ, 2008. – 311 с.

4. Магнус, Я. Р. Эконометрика: Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – М. : Дело, 2005. – 247 с.

Дополнительная

5. Артамонов, Н. В. Введение в эконометрику : курс лекций. – 2010. – 204 с. – Режим доступа: http://www. gaudeamus.omskcity.com/ lib–pdf/econom/ArtamonovNV_–_Vvedenie VEkonometriku_KursLekcij_–_2010_204 _PDF.zip.

6. Замков, О. О. Математические методы в экономике : учебник / О. О. Замков, А. В. Толстопятенко, Ю. Н. Черемных. – М. : МГУ им. М. В. Ломоносова : ДИС, 1998. – 368 с.

7. Минюк, С. А. Математические методы и модели в экономике : учеб. пособие / С. А. Минюк, Е. А. Ровба, К. К. Кузьмич – Минск : ТетраСистемс, 2002. – 432 с.

8. Экономико-математические методы и прикладные модели : учеб. пособие для вузов / В. В. Федосеев [и др.] ; под ред. В. В. Федосеева. – М. : ЮНИТИ, 2001. – 391 с.

9. Шанченко, Н. И. Эконометрика: лабораторный практикум : учеб. пособие / Н. И. Шевченко. – Ульяновск : УлГТУ, 2011. – 117 с. – Режим доступа: http://www.gaudeamus. omskcity.com/lib–pdf/econom/ShanchenkoEkono–metrika_ LabPraktikum_ –_UP_–_2011_117_PDF.zip.

Учебно-методические комплексы

10. Кожевников, Е. А. Экономико–математические методы и модели : учеб.-метод. комплекс / Е. А. Кожевников. – Минск : ГИУСТ, 2004. – 148 с.

ПРИЛОЖЕНИЕ

Таблица П.1.1

Исходные данные к расчетным работам № 1 и 2 (значения факторных переменных (x_i))

Факторные переменные (x_i)												
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
0,16	0,11	2,40	0,16	14,99	0,80	0,57	12,01	0,81	74,96	3,35	2,39	50,46
0,80	0,19	5,44	0,24	14,72	4,01	0,96	27,20	1,20	73,61	16,83	4,02	114,26
0,94	0,20	5,87	0,32	4,55	4,72	1,00	29,33	1,58	22,77	19,80	4,19	123,17
0,26	0,15	9,65	0,48	11,57	1,29	0,75	48,27	2,39	57,86	5,44	3,13	202,72
0,27	0,05	8,11	0,13	1,64	1,36	0,23	40,53	0,66	8,18	5,73	0,99	170,22
0,47	0,05	8,23	0,10	16,75	2,35	0,25	41,17	0,50	83,73	9,87	1,04	172,90
0,34	0,16	3,89	0,22	17,56	1,71	0,80	19,47	1,12	87,81	7,17	3,35	81,79
0,31	0,06	7,97	0,13	18,92	1,55	0,28	39,87	0,66	94,60	6,53	1,18	167,45
0,65	0,12	0,97	0,11	6,11	3,27	0,61	4,83	0,53	30,57	13,74	2,55	20,29
0,06	0,08	9,05	0,30	14,92	0,31	0,40	45,24	1,48	74,62	1,28	1,66	190,01
0,14	0,05	9,20	0,18	19,72	0,70	0,25	46,02	0,90	98,58	2,95	1,05	193,29
0,10	0,09	5,00	0,14	19,90	0,52	0,46	25,01	0,68	99,52	2,20	1,94	105,02
0,81	0,15	5,69	0,24	9,36	4,04	0,75	28,47	1,18	46,81	16,95	3,14	119,58
0,19	0,11	7,32	0,36	3,61	0,93	0,53	36,59	1,81	18,07	3,89	2,22	153,66
0,39	0,05	2,99	0,17	1,02	1,97	0,26	14,93	0,85	5,08	8,26	1,10	62,71
0,91	0,11	3,05	0,19	3,11	4,55	0,55	15,27	0,94	15,53	19,12	2,32	64,15
0,64	0,02	4,17	0,48	1,89	3,20	0,11	20,87	2,39	9,45	13,44	0,46	87,66
0,87	0,01	9,54	0,03	9,53	4,33	0,06	47,72	0,16	47,64	18,17	0,24	200,41
0,33	0,14	6,88	0,15	0,01	1,64	0,69	34,41	0,76	0,03	6,90	2,92	144,51
0,92	0,16	8,51	0,04	11,08	4,62	0,78	42,57	0,20	55,39	19,39	3,26	178,80
0,49	0,11	0,94	0,09	2,98	2,44	0,53	4,69	0,45	14,92	10,24	2,24	19,70
0,17	0,01	7,51	0,10	1,88	0,83	0,03	37,56	0,52	9,39	3,48	0,14	157,75
0,47	0,07	0,81	0,17	7,65	2,33	0,35	4,06	0,84	38,25	9,80	1,45	17,06
0,79	0,15	5,16	0,44	0,02	3,97	0,75	25,80	2,18	0,08	16,67	3,17	108,35
0,22	0,08	6,21	0,33	4,25	1,08	0,38	31,06	1,66	21,25	4,53	1,60	130,45
0,39	0,04	9,38	0,47	0,60	1,95	0,21	46,88	2,34	3,01	8,20	0,86	196,90
0,57	0,04	4,28	0,10	2,30	2,84	0,22	21,38	0,52	11,51	11,94	0,93	89,80
0,24	0,05	3,42	0,30	10,11	1,20	0,24	17,10	1,52	50,53	5,03	1,00	71,83
0,08	0,20	3,90	0,06	0,10	0,39	1,00	19,52	0,28	0,48	1,64	4,19	81,98
0,53	0,08	4,38	0,11	17,98	2,66	0,40	21,90	0,56	89,90	11,18	1,69	91,97
0,24	0,11	5,30	0,28	1,34	1,19	0,55	26,51	1,41	6,70	5,01	2,29	111,35
0,12	0,13	1,63	0,39	6,40	0,58	0,65	8,15	1,95	32,01	2,45	2,74	34,25
0,76	0,09	5,71	0,47	1,86	3,78	0,44	28,53	2,37	9,32	15,87	1,85	119,84

Окончание табл. П.1.1

Факторные переменные (x_i)												
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
0,61	0,01	7,65	0,45	3,49	3,07	0,03	38,25	2,24	17,47	12,90	0,12	160,64
0,85	0,13	0,82	0,41	15,02	4,23	0,63	4,10	2,05	75,10	17,78	2,63	17,24
0,39	0,20	4,50	0,38	10,15	1,95	0,99	22,50	1,89	50,76	8,17	4,18	94,51
0,23	0,17	1,17	0,09	14,31	1,17	0,85	5,87	0,47	71,55	4,90	3,58	24,66
0,77	0,10	5,71	0,28	6,39	3,85	0,51	28,53	1,42	31,96	16,15	2,15	119,82
0,29	0,15	8,93	0,48	11,19	1,46	0,76	44,64	2,38	55,95	6,13	3,20	187,48
0,99	0,15	1,63	0,12	0,30	4,97	0,76	8,17	0,59	1,51	20,89	3,19	34,32
0,83	0,17	8,58	0,08	17,06	4,17	0,86	42,91	0,42	85,30	17,50	3,61	180,22
0,44	0,13	4,19	0,46	1,50	2,19	0,65	20,96	2,28	7,50	9,20	2,73	88,02
0,03	0,16	6,48	0,34	12,22	0,14	0,81	32,40	1,69	61,08	0,60	3,38	136,08
0,07	0,11	2,37	0,34	5,36	0,33	0,57	11,85	1,71	26,79	1,41	2,40	49,77
0,75	0,16	2,80	0,10	3,24	3,74	0,80	14,00	0,50	16,22	15,69	3,36	58,78

Таблица П.1.2

Исходные данные к расчетным работам № 1 и 2
(значения зависимой переменной (y))

Значения зависимой переменной (y)											
Начало перечня						**Окончание перечня**					
14	**15**	**16**	**17**	**18**	**19**	**14**	**15**	**16**	**17**	**18**	**19**
21,1	15,0	–0,6	20,7	23,0	12,7	20,9	2,7	25,3	20,7	25,8	30,4
20,6	10,3	43,0	20,5	24,4	45,7	21,2	1,2	31,1	20,4	22,9	57,3
20,7	4,5	49,2	20,6	25,7	51,0	21,0	–7,6	46,4	20,1	21,0	83,4
22,1	–0,5	51,7	21,0	28,5	101,7	19,5	3,5	19,9	19,4	18,9	23,3
20,2	–4,5	27,6	19,9	20,3	49,5	20,8	9,2	27,5	20,1	21,2	36,4
19,8	2,9	41,7	19,6	19,4	62,4	21,6	6,2	1,0	21,5	24,0	18,0
21,3	14,2	26,2	20,8	24,8	34,8	19,9	11,8	19,6	19,8	20,1	24,5
20,2	4,6	23,7	20,0	20,4	41,4	21,2	2,2	30,2	20,6	23,9	50,3
20,0	13,6	24,2	19,9	20,6	5,9	22,0	12,5	1,4	20,9	26,7	14,7
21,4	0,9	36,8	20,6	23,4	80,1	20,6	1,3	40,0	20,3	22,6	49,6
20,7	2,4	42,4	20,2	21,1	83,9	20,2	–3,4	54,4	19,6	17,9	81,0
20,9	11,8	32,0	20,6	22,3	51,6	20,5	18,4	29,6	20,4	23,8	10,7
20,3	6,3	36,4	20,2	22,7	38,9	21,9	10,1	22,8	21,3	29,1	36,9
216	–0,7	17,4	20,7	24,8	42,5	21,2	18,2	–2,8	21,0	23,4	4,0
20,1	5,6	33,0	19,8	20,2	30,8	20,1	3,8	50,4	19,9	21,5	57,1
19,6	7,7	45,1	19,6	20,2	29,7	22,1	0,8	49,5	21,0	28,5	93,8
20,3	3,0	31,6	19,8	18,8	38,6	19,6	9,9	37,4	19,6	20,2	12,4
18,5	–4,1	70,8	18,4	16,8	94,3	20,0	4,8	44,1	19,9	20,9	47,8

Окончание табл. П.1.2

Значения зависимой переменной (y)											
Начало перечня						**Окончание перечня**					
14	**15**	**16**	**17**	**18**	**19**	**14**	**15**	**16**	**17**	**18**	**19**
20,9	–1,0	26,3	20,6	230	44,8	21,5	5,0	28,8	20,8	26,2	40,9
19,5	1,6	71,4	19,5	19,4	88,9	22,2	6,4	18,1	21,3	27,9	48,8
20,1	11,8	23,8	20,0	20,7	6,2	21,8	10,2	20,0	20,8	25,5	27,2
20,0	–3,9	34,6	19,8	19,5	64,1	20,1	9,2	14,5	20,0	21,2	11,5
20,1	13,6	4,7	19,9	20,4	3,6	–					

Таблица П.1.3

Критические значения количества рядов для определения наличия автокорреляции по методу рядов ($\alpha = 0,05$) (нижняя граница k_1)

		Значения n_2																		
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Значения n_1	**2**											2	2	2	2	2	2	2	2	2
	3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
	4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
	5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
	6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
	7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
	8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
	9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
	10		2	3	4	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
	11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
	12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
	13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
	14	2	2	3	4	5	5	6	6	7	7	8	8	9	9	10	10	10	11	11
	15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
	16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
	17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
	18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
	19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
	20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

Таблица П.1.4

Критические значения количества рядов для определения наличия автокорреляции по методу рядов (α = 0,05) (верхняя граница k_2)

		Значения n_2																		
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Значения n_1	**2**																			
	3																			
	4				9	9														
	5			9	10	10	11	11												
	6			9	10	11	12	12	13	13	13	13								
	7				11	12	13	13	14	14	14	14	15	15	15					
	8				11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
	9					13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
	10					13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
	11					13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
	12					13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
	13						15	16	17	18	19	19	20	20	21	21	22	22	23	23
	14						15	16	17	18	19	20	20	21	22	22	23	23	23	24
	15						15	16	18	18	19	20	21	22	22	23	23	24	24	25
	16							17	18	19	20	21	21	22	23	23	24	25	25	25
	17							17	18	19	20	21	22	23	23	24	25	25	26	26
	18							17	18	19	20	21	22	23	24	25	25	26	26	27
	19							17	18	20	21	22	23	23	24	25	26	26	27	27
	20							17	18	20	21	22	23	24	25	25	26	27	27	28

Пример. Пусть при n = 20 будет 11 знаков «+» (= n_1) и 9 знаков «–» (= n_2).Тогда при $\alpha = 0{,}05$ нижняя граница $k_1 = 6$, верхняя граница $k_2 = 16$. Если $k_{\text{набл}} \le 6$ или $k_{\text{набл}} \ge 16$, то гипотеза об отсутствии автокорреляции должна быть отклонена.

Таблица П.1.5

Распределение Дарбина–Уотсона. Критические точки d_1 и d_u при уровне значимости α = 0,05 (n – объем выборки, m – число объясняющих переменных в уравнении регрессии)

n	m = 1		m = 2		m = 3		m = 4		m = 5		m = 6	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
6	0,610	1,400										
7	0,700	1,356	**0,467**	**1,896**								
8	0,763	1,332	**0,359**	**1,777**	0,368	2,287						
9	0,824	1,320	**0,629**	**1,699**	0,435	2,128	**0,296**	**2,388**	0,243	2,822		

Окончание табл. П.1.5

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$		$m = 6$	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,414	0,316	2,645	0,203	3,005
11	0,927	1,324	0,658	1,604	0,595	1,928	0,444	2,283	0,379	2,506	0,268	2,832
12	0,971	1,331	0,812	1,579	0,58	1,864	0,512	2,177	0,445	2,390	0,328	2,692
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094	0,505	2,296	0,389	2,572
14	1,045	1,330	0,905	1,551	0,767	1,779	0,632	2,030	0,562	2,220	0,447	2,472
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977	0,615	2,157	0,502	2,388
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935	0,664	2,104	0,554	2,318
17	1,133	1,381	1,015	1,536	0,897	1,710	0,779	1,900	0,710	2,060	0,603	2,257
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872	0,752	2,023	0,649	2,206
19	1,180	1,401	1,074	1,536	0,967	1,685	0,859	1,848	0,792	1,991	0,692	2,162
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828	0,829	1,964	0,732	2,124
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812	0,863	1,940	0,769	2,090
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797	0,895	1,920	0,804	2,061
23	1,257	1,437	1,168	1,543	1,078	1,660	0,986	1,785	0,925	1,902	0,837	2,035
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775	0,953	1,886	0,868	2,012
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767	0,979	1,873	0,897	1,992
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759	1,004	1,861	0,925	1,974
27	1,316	1,469	1,240	1,556	1,162	1,651	1,084	1,753	1,028	1,850	0,951	1,958
28	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747	1,050	1,841	0,975	1,944
29	1,341	1,483	1,270	1,563	1198	1,650	1,124	1,743	1,071	1,833	0,998	1,931
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,090	1,825	1,020	1,920
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732	1,127	1,813	1,061	1,900
34	1,393	1,514	1,333	1,580	1,271	1,652	1,208	1,728	1,160	1,803	1,097	1,884
36	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724	1,190	1,795	1,131	1,870
38	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722	1,218	1,789	1,161	1,859
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,230	1,786	1,175	1,854
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720	1,287	1,776	1,238	1,835
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721	1,335	1,771	1,291	1,822
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724	1,374	1,768	1,334	1,814
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727	1,408	1,767	1,372	1,808
65	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731	1,438	1,767	1,404	1,805
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735	1,464	1,768	1,433	1,802
75	1,598	1,650	1,571	1,680	1,543	1,709	1,515	1,739	1,487	1,770	1,458	1,801
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743	1,507	1,772	1,480	1,801
85	1,624	1,671	1,600	1,696	1,575	1,721	1,550	1,747	1,525	1,774	1,500	1,801
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751	1,542	1,776	1,518	1,801

Таблица П.1.6

Распределение Дарбина–Уотсона. Критические точки d_1 и d_u при уровне значимости α = 0,01 (*n* – объем выборки, *m* – число объясняющих переменных в уравнении регрессии)

n	*m* = 1		*m* = 2		*m* = 3		*m* = 4		*m* = 5		*m* = 6	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
6	0,390	1,142										
7	0,433	1,036	**0,294**	**1,676**								
8	0,497	1,003	**0,343**	**1,489**	0,229	2,102						
9	0,554	0,998	**0,408**	**1,389**	0,279	1,873	**0,183**	**2,433**				
10	0,604	1,001	**0,466**	**1,333**	0,340	1,733	**0,230**	**2,193**	0,130	2,690		
11	0,633	1,010	**0,319**	**1,297**	0,396	1,640	**0,286**	**2,030**	0,193	2,433	**0,124**	**2,892**
12	0,697	1,023	**0,369**	**1,274**	0,449	1,373	**0,339**	**1,913**	0,244	2,280	**0,164**	**2,663**
13	0,738	1,038	**0,616**	**1,261**	0,499	1,326	**0,391**	**1,826**	0,294	2,130	**0,211**	**2,490**
14	0,776	1,034	**0,660**	**1,234**	0,347	1,490	**0,441**	**1,737**	0,343	2,049	**0,237**	**2,334**
15	0,811	1,070	**0,700**	**1,232**	0,391	1,464	**0,488**	**1,704**	0,391	1,967	**0,303**	**2,244**
16	0,844	1,086	**0,737**	**1,232**	0,633	1,446	**0,332**	**1,663**	0,437	1,900	**0,349**	**2,133**
17	0,874	1,102	**0,772**	**1,233**	0,672	1,432	**0,374**	**1,630**	0,480	1,847	**0,393**	**2,078**
18	0,902	1,118	**0,803**	**1,239**	0,708	1,422	**0,613**	**1,604**	0,322	1,803	**0,433**	**2,013**
19	0,928	1,132	**0,833**	**1,263**	0,742	1,413	**0,630**	**1,384**	0,361	1,767	**0,476**	**1,963**
20	0,932	1,147	**0,863**	**1,271**	0,773	1,411	**0,683**	**1,367**	0,398	1,737	**0,313**	**1,918**
21	0,973	1,161	**0,890**	**1,277**	0,803	1,408	**0,718**	**1,334**	0,633	1,712	**0,332**	**1,881**
22	0,997	1,174	**0,914**	**1,284**	0,831	1,407	**0,748**	**1,343**	0,667	1,691	**0,387**	**1,849**
23	1,018	1,187	**0,938**	**1,291**	0,838	1,407	**0,777**	**1,334**	0,698	1,673	**0,620**	**1,821**
24	1,037	1,199	**0,960**	**1,298**	0,882	1,407	**0,803**	**1,328**	0,728	1,638	**0,632**	**1,797**
25	1,033	1,211	**0,981**	**1,303**	0,906	1,409	**0,831**	**1,323**	0,736	1,643	**0,682**	**1,776**
26	1,072	1,222	**1,001**	**1,312**	0,928	1,411	**0,833**	**1,318**	0,783	1,633	**0,711**	**1,739**
27	1,089	1,233	**1,019**	**1,319**	0,949	1,413	**0,878**	**1,313**	0,808	1,626	**0,738**	**1,743**
28	1,104	1,244	**1,037**	**1,323**	0,969	1,413	**0,900**	**1,313**	0,832	1,618	**0,764**	**1,729**
29	1,119	1,234	**1,034**	**1,332**	0,988	1,418	**0,921**	**1,312**	0,833	1,611	**0,788**	**1,718**
30	1,133	1,263	**1,070**	**1,339**	1,006	1,421	**0,941**	**1,311**	0,877	1,606	**0,812**	**1,707**
32	1,160	1,282	**1,100**	**1,332**	1,040	1,428	**0,979**	**1,310**	0,917	1,397	**0,836**	**1,690**
34	1,184	1,299	**1,128**	**1,364**	1,070	1,433	**1,012**	**1,311**	0,934	1,391	**0,896**	**1,677**
36	1,206	1,313	**1,133**	**1,376**	1,098	1,442	**1,043**	**1,313**	0,988	1,388	**0,932**	**1,666**
38	1,227	1,330	**1,176**	**1,388**	1,124	1,449	**1,072**	**1,313**	1,019	1,383	**0,966**	**1,638**
40	1,246	1,344	**1,198**	**1,398**	1,148	1,437	**1,098**	**1,318**	1,048	1,384	**0,997**	**1,632**
45	1,288	1,376	**1,243**	**1,423**	1,201	1,474	**1,136**	**1,328**	1,111	1,384	**1,063**	**1,643**
50	1,324	1,403	**1,283**	**1,446**	1,243	1,491	**1,203**	**1,338**	1,164	1,387	**1,123**	**1,639**
55	1,336	1,427	**1,320**	**1,466**	1,284	1,306	**1,247**	**1,348**	1,209	1,392	**1,172**	**1,638**
60	1,383	1,449	**1,330**	**1,484**	1,317	1,320	**1,283**	**1,338**	1,249	1,398	**1,214**	**1,639**
65	1,407	1,468	**1,377**	**1,300**	1,346	1,334	**1,313**	**1,368**	1,283	1,604	**1,231**	**1,642**
70	1,429	1,483	**1,400**	**1,313**	1,372	1,346	**1,343**	**1,378**	1,313	1,611	**1,283**	**1,643**

Окончание таб. П.1.6

n	*m* = 1		*m* = 2		*m* = 3		*m* = 4		*m* = 5		*m* = 6	
	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u	d_1	d_u
75	1,448	1,301	**1,422**	**1,329**	1,393	1,337	**1,368**	**1,387**	1,340	1,617	**1,313**	**1,649**
80	1,466	1,313	**1,441**	**1,341**	1,416	1,368	**1,390**	**1,393**	1,364	1,624	**1,338**	**1,633**
85	1,482	1,328	**1,438**	**1,333**	1,433	1,378	**1,411**	**1,603**	1,386	1,630	**1,362**	**1,637**
90	1,496	1,340	**1,474**	**1,363**	1,432	1,387	**1,429**	**1,611**	1,406	1,636	**1,383**	**1,661**

Таблица П.1.7

Значения *F*-критерия Фишера при уровне значимости α = 0,05

		Значения k_1									
		1	**2**	**3**	**4**	**5**	**6**	**8**	**12**	**24**	**∞**
Значения k_2	**1**	161,5	199,5	215,7	224,6	230,2	234,0	238,9	243,9	249,0	254,3
	2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
	3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
	11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
	12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
	13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
	21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
	23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
	25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
	30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
	35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
	40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
	45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
	50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44

Окончание табл. П.1.7

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

Таблица П.1.8

Критические значения *t*-критерия Стьюдента при уровне значимости α = 0,10, α = 0,05, α = 0,01 (двухсторонний)

Число степеней свободы *df*	Уровень значимости α			Число степеней свободы *df*	Уровень значимости α		
	0,10	**0,05**	**0,01**		**0,10**	**0,05**	**0,01**
1	6,3138	12,706	63,657	**18**	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	**19**	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	**20**	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,5041	**21**	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	**22**	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	**23**	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	**24**	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	**25**	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	**26**	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	**27**	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	**28**	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	**29**	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,1023	**30**	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	**40**	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	**60**	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	**120**	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Учебное электронное издание комбинированного распространения

Учебное издание

Ермалинская Наталья Васильевна

ЭКОНОМЕТРИКА (ПРОДВИНУТЫЙ УРОВЕНЬ)

Практикум для магистрантов специальности 1-25 80 04 «Экономика и управление народным хозяйством» дневной и заочной форм обучения

Электронный аналог печатного издания

Редактор *Н. Г. Мансурова*
Компьютерная верстка *Н. Б. Козловская*

Подписано в печать 20.04.17.

Формат 60x84/$_{16}$. Бумага офсетная. Гарнитура «Таймс».
Ризография. Усл. печ. л. 4,42. Уч.-изд. л. 4,5.
Изд. № 99.
http://www.gstu.by

Издатель и полиграфическое исполнение
Гомельский государственный
технический университет имени П. О. Сухого.
Свидетельство о гос. регистрации в качестве издателя
печатных изданий за № 1/273 от 04.04.2014 г.
пр. Октября, 48, 246746, г. Гомель