Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Приморская государственная сельскохозяйственная академия»
Институт лесного и лесопаркового хозяйства

Кафедра философии и
социально-гуманитарных
дисциплин

Жуплей И.В., Дьяков И.И.

ЭКОНОМЕТРИКА

Учебное пособие для подготовки к практическим занятиям
и выполнения самостоятельной работы
для обучающихся по направлению подготовки
38.03.01Экономика (уровень бакалавриата)
ФГБОУ ВО Приморская ГСХА

Электронное издание

Уссурийск 2017

УДК 514.12
ББК 22.18 я73

Рецензенты: Т.И. Еременко, ведущий специалист-эксперт отдела государственной статистики г. Владивостоке (г. Уссурийск) Приморскстата;

Е.В. Савельева, к.т.н., зав. кафедрой физики и высшей математики ФГБОУ ВО Приморская ГСХА

Эконометрика: Учебное пособие для подготовки к практическим занятиям и выполнения самостоятельной работы для обучающихся по направлению подготовки 38.03.01Экономика (уровень бакалавриата) ФГБОУ ВО Приморская ГСХА [Электронный ресурс] / ФГБОУ ВО Приморская ГСХА; сост. И.В. Жуплей, И.И. Дьяков. – Электрон. текст. дан. - Уссурийск: ФГБОУ ВО Приморская ГСХА 2017. – 83 с. - Режим доступа: www.de.primacad.ru

Учебное пособие предназначено для подготовки к практическим занятиям и выполнения самостоятельной работы по дисциплине (модулю) Эконометрика для обучающихся по направлению подготовки 38.03.01 Экономика (уровень бакалавриата).

Содержит необходимый теоретический минимум, решение основных типов задач, вопросы и тест для самопроверки, а также задания для самостоятельного решения.

Учебное пособие может быть использовано для самостоятельной работы обучающихся, а также на практических занятиях и при проведении контрольных мероприятий.

Для обучающихся очной и заочной форм обучения по направлению подготовки 38.03.01 Экономика (уровень бакалавриата).

Электронное издание

Издается по решению методического совета ФГБОУ ВО Приморская ГСХА

ПРЕДИСЛОВИЕ

Эконометрика входит в число важнейших дисциплин (модулей) современного экономического образования, в связи с чем, данный курс играет важную роль в подготовке бакалавров экономического профиля вообще, и в сфере агропромышленного комплекса в частности.

Целью изучения курса эконометрики является получение обучающимися базовых знаний и навыков в области эконометрического анализа. Они должны уметь применять их в исследовании экономических процессов. Обучающиеся должны приобрести навыки построения и оценки качества моделей парной и множественной линейной регрессии, познакомиться с некоторыми видами нелинейных моделей и систем уравнений, понимая область и границы их применения в экономике.

Предполагается, что обучающиеся, изучающие эконометрику, уже освоили базовые курсы по математике и статистике, и при изучении курса эконометрики необходимо лишь восстановить знания основных положений, без которых невозможно понимание материала.

Объем аудиторных занятий по курсу эконометрики в сельскохозяйственных вузах крайне мал, а у обучающихся заочного отделения вообще сведен к минимуму. Соответственно увеличено время, предусмотренное для самостоятельной учебной работы.

В связи с этим для более глубокого понимания и усвоения курса эконометрики необходимо наличие систематизированного учебного материала, изложенного в доступной форме и предназначенного, прежде всего, для самостоятельного изучения.

Учебное пособие структурно состоит из трех глав («Парная регрессия и корреляция», «Множественная регрессия и корреляция», «Системы эконометрических уравнений»), в каждой из которых содержится необходимый теоретический минимум, примеры решения типовых задач, вопросы для самопроверки и задачи для самостоятельного решения. При составлении задач использовались условные данные.

Для контроля усвоения материала в методических указаниях предусмотрен итоговый тест.

Важной частью самостоятельной работы по изучению дисциплины (модуля) является выполнение заданий вычислительного практикума.

Вычислительный практикум содержит десять вариантов индивидуальных заданий по основным разделам эконометрики: парная регрессия и корреляция, множественная регрессия и корреляция.

По результатам выполнения заданий вычислительного практикума можно судить о том, насколько обучающийся овладел методологией построения эконометрических моделей и каковы его возможности применить полученные знания на практике.

Значение вычислительного практикума состоит в том, что в процессе его выполнения обучающийся не только закрепляет, но и углубляет полученные теоретические знания и практические навыки.

1. ПАРНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

1.1 Необходимый теоретический минимум

Основные эконометрические понятия

Эконометрика – наука, которая дает количественное выражение теоретическим моделям, описывающим реальные экономические явления и процессы.

Экономические явления и процессы, как правило, выражаются в массовых, а не в единичных фактах (массовые – повторяющиеся во времени и в пространстве), в связи с этим предметом исследования эконометрики являются массовые экономические явления и процессы.

Для количественной характеристики массовых варьирующих явлений и связей между ними используются математико-статистические методы исследования и статистические показатели.

Основная задача эконометрики состоит в построении моделей специфического типа (эконометрических моделей), описывающих состояние и развитие экономических процессов. Эти модели используются в анализе и прогнозировании общих закономерностей и конкретных количественных характеристик рассматриваемых процессов, определении управляющих воздействий.

Решение конкретных задач, стоящих перед эконометрикой осуществляется в ходе эконометрических исследований. При этом любое эконометрическое исследование состоит из следующих основных этапов:

1) постановка проблемы;
2) теоретический анализ сущности изучаемого явления;
3) выбор общего вида модели (спецификация);
4) получение исходных данных и анализ их качества;
5) оценка параметров модели (параметризация);
6) оценка качества модели (верификация);
7) интерпретация полученных результатов.

На первом этапе решается вопрос о необходимости исследования того или иного экономического явления, а также определяется конечная цель исследования. Как правило, целью эконометрического исследования является выявление зависимости между признаками экономического явления или процесса (экономическими показателями) и построение соответствующей модели связи.

На втором этапе проводится сбор и изучение теоретической информации о моделируемом явлении или процессе. Важность данного этапа для исследования в целом заключается в том, что без знания природы и экономической сущности явления или процесса невозможно построить модель, достоверно воспроизводящую это явление или процесс.

Третий этап, спецификация модели, заключается в выборе аналитической формы зависимости (математической модели), исходя из соответствующей природы связи между признаками.

Четвертый этап, по сути, представляет собой статистическое наблюдение, подробно рассмотренное в курсе статистики. На данном этапе осуществляется сбор необходимой статистической информации – наблюдаемых (реальных) значений экономических показателей.

Пятый этап, параметризация модели, заключается в определении числовых значений параметров модели на основе имеющихся статистических данных.

На шестом этапе, в процессе верификации модели, выясняется насколько удачно решены задачи предыдущих этапов, какова точность расчетов по данной модели, насколько соответствует построенная модель моделируемому реальному экономическому явлению или процессу.

Если модель удовлетворяет всем необходимым требованиям качества, то она может быть использована либо для прогнозирования, либо для объяснения внутренних механизмов исследуемого явления или процесса. В этом случае параметры модели должны иметь экономический смысл. Поэтому заключительным этапом исследования является экономическая интерпретация параметров модели и других результатов исследования.

При моделировании экономических явлений и процессов используют два типа данных: пространственные и временные данные.

Пространственные данные (пространственный срез) – набор сведений по разным однотипным экономическим объектам, взятый за один и тот же период или момент времени. Например, объем производства, выручка от реализации продукции, прибыль и т.п. разных сельскохозяйственных предприятий за один и тот же период (год, квартал, месяц); численность работников перерабатывающих предприятий района на 1 января текущего года и т.д.

Временные данные (временной срез) – набор сведений, характеризующий один и тот же объект, но в разные периоды или моменты времени. Например, ежедневный курс валюты на бирже, средняя урожайность зерновых в Приморском крае за каждый год в течение последних десяти лет и т.д. Отличительной особенностью временных данных является их естественная упорядоченность во времени.

Любой набор экономических сведений (данных) представляет собой множество признаков, характеризующих объект исследования. При моделировании взаимосвязи между признаками предполагается, что они могут выступать в качестве либо результативного, либо факторного признака.

Результативный признак – признак, значение которого зависит от значения других признаков.

Факторный признак – признак, значение которого определяет значение признака-результата.

В эконометрической модели результативный признак называют объясняемой переменной, а факторный признак – объясняющей переменной. Переменные, участвующие в эконометрической модели, также разделяют на:

- экзогенные (независимые) – значение которых задается вне модели;
- эндогенные (зависимые) – значение которых определяется внутри модели;

- лаговые – экзогенные или эндогенные переменные, определенные в предыдущий период или момент времени и находящиеся в модели с текущими переменными.

Выделяют три основных класса эконометрических моделей:

1. Регрессионные модели с одним уравнением – модели, в которых результативный признак представлен в виде функции одного или нескольких факторных признаков;

2. Модели временных данных – модели, в которых результативный признак является функцией переменной времени или переменных, относящихся к другим моментам (периодам) времени;

3. Системы уравнений – модели, состоящие из тождеств и регрессионных уравнений, в которых в качестве независимых переменных могут включаться не только факторные, но и результативные признаки из других уравнений системы.

В зависимости от количества факторов, включенных в уравнение, различают модели парной и множественной регрессии.

Парная регрессия представляет собой зависимость между двумя переменными – y и x, т.е. модель вида:

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

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

Знак «^» означает, что между переменными x и y нет строгой функциональной зависимости. При этом величина y может быть представлена в виде суммы:

$$y_i = \hat{y}_{x_i} + \varepsilon_i, \tag{1.2}$$

где y_i – фактическое значение результативного признака;
$\hat{y}_{x_i}$ – теоретическое значение результативного признака (т.е. найденное исходя из уравнения регрессии);
ε_i– ошибка модели, (т.е. случайная величина, характеризующая отклонения фактических значений результативного признака от теоретических).

Случайная величина ε_i включает ошибки спецификации, ошибки выборки и ошибки измерения. Величина случайных ошибок тем меньше, чем в большей мере теоретические значения результативного признака, подходят к фактическим данным.

Основное внимание в эконометрических исследованиях уделяется ошибкам спецификации модели. К ошибкам спецификации относятся неправильный выбор математической функции и недоучет в уравнении регрессии какого-либо существенного фактора.

В парной регрессии выбор вида математической функции может быть осуществлен тремя методами:

1) графическим;
2) аналитическим;
3) экспериментальным.

В настоящее время, при обработке информации на компьютере, выбор вида уравнения регрессии обычно осуществляется экспериментальным методом, т. е. путем сравнения величины остаточной дисперсии $\sigma^2_{\text{ост}}$, рассчитанной для разных моделей.

В основе расчета остаточной дисперсии лежит величина отклонений фактических значений результативного признака от теоретических:

$$\sigma^2_{\text{ост}} = \frac{1}{n}\sum_{i=1}^{n}(y - \hat{y}_{x_i})^2 \quad (1.3)$$

● Замечание. Для упрощения индексы в значках сумм будем опускать.

Чем меньше величина остаточной дисперсии, тем меньше влияние не учитываемых в уравнении регрессии факторов и тем лучше уравнение регрессии подходит к фактическим данным.

Парная линейная регрессия и корреляция

Наиболее широкое применение в эконометрике находит линейная регрессия. Это связано с возможностью четкой экономической интерпретации ее параметров.

Линейная регрессия сводится к нахождению уравнения вида:

$$\hat{y}_{x_i} = a + bx_i, \quad (1.4)$$

или в общем виде:

$$y_i = a + bx_i + \varepsilon_i. \quad (1.5)$$

где *a, b* – параметры регрессии.

Параметр *b* называется коэффициентом регрессии. Его величина показывает среднее изменение результата с изменением фактора на одну единицу.

Параметр *a* – значение *y* при *x* = 0. Если факторный признак (*x*) не имеет и не может иметь нулевого значения, то интерпретация параметра *a* не имеет смысла, т.е. параметр *a* может не иметь экономического содержания.

Построение уравнения линейной регрессии сводится к оценке его параметров – *a* и *b*. Основным подходом к оцениванию параметров линейной регрессии является метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров *a* и *b*, при которых сумма квадратов отклонений фактических значений результативного признака от теоретических минимальна:

$$\sum(y_i - \hat{y}_{x_i})^2 = \sum \varepsilon_i^2 \to \min . \qquad (1.6)$$

Как известно из курса математики, чтобы найти экстремум функции зависящей от двух переменных, надо вычислить частные производные по каждому из параметров a и b и приравнять их к нулю.

Обозначим $\sum_{i=1}^{n} \varepsilon_i^2$ через $S(a,\ b)$, тогда:

$$S(a,\ b) = \sum(y - a - b \cdot x)^2 .$$

$$\begin{cases} \dfrac{\partial S}{\partial a} = -2\sum(y - a - b \cdot x) = 0; \\ \dfrac{\partial S}{\partial b} = -2\sum x(y - a - b \cdot x) = 0. \end{cases}$$

После преобразований, получим следующую систему нормальных уравнений для оценки параметров a и b:

$$\begin{cases} a \cdot n \quad + b \cdot \sum x = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 = \sum x \cdot y. \end{cases} \qquad (1.7)$$

Решение данной системы имеет вид:

$$b = \frac{n \cdot \Sigma xy - \Sigma x \cdot \Sigma y}{n \cdot \Sigma x^2 - (\Sigma x)^2} ; \qquad (1.8)$$

$$a = \frac{1}{n}(\Sigma y - b \cdot \Sigma x) . \qquad (1.9)$$

Оценки параметров a и b также можно найти по формулам:

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x{}^2}; \tag{1.10}$$

$$a = \bar{y} - b \cdot \bar{x}. \tag{1.11}$$

Применение корреляционного анализа предполагает определение показателя тесноты связи. При использовании линейной регрессии в качестве такого показателя выступает линейный коэффициент корреляции r_{xy}, который можно рассчитать по следующим формулам:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y}; \tag{1.12}$$

$$r_{xy} = \frac{\Sigma(y-\bar{y})(x-\bar{x})}{\sqrt{\Sigma(y-\bar{y})^2 \cdot \Sigma(x-\bar{x})^2}}; \tag{1.13}$$

$$r_{xy} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y}. \tag{1.14}$$

где σ_x, σ_y – средние квадратические отклонения факторного и результативного признака;

$\bar{x}, \bar{y}$ - средние значения факторного и результативного признака;

Линейный коэффициент корреляции находится в пределах: $-1 \le r_{xy} \le 1$. Чем ближе абсолютное значение r_{xy} к единице, тем сильнее линейная связь между факторами (при $r_{xy} = \pm 1$ имеем строгую функциональную зависимость). Но следует иметь в виду, что близость абсолютной величины линейного коэффициента корреляции к нулю еще не означает отсутствия связи между признаками. При другой (нелинейной) спецификации модели связь между признаками может оказаться достаточно тесной.

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции r_{xy}^2, называемый коэффициентом детерминации. Коэффициент детерминации характеризует долю вариации результативного признака y, объясняемую регрессией, в общей вариации результативного признака:

$$r_{xy}^2 = 1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}, \tag{1.15}$$

где $\sigma_{\text{ост}}^2 = \frac{1}{n}\sum\left(y - y_x\right)^2$, $\sigma_y^2 = \frac{1}{n}\sum\left(y - \bar{y}\right)^2 = \overline{y^2} - \bar{y}^2$.

Соответственно величина $(1 - r_{xy}^2)$ характеризует долю вариации y, вызванную влиянием не учтенных в модели факторов.

Для оценки качества модели также определяют среднюю ошибку аппроксимации:

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y - y_x}{y}\right| \cdot 100\%. \tag{1.16}$$

Средняя ошибка аппроксимации не должна превышать 8–10%.

Оценка статистической значимости уравнения регрессии

После того как найдено уравнение линейной регрессии, проводится оценка значимости как уравнения в целом, так и отдельных его параметров.

Оценка значимости уравнения регрессии в целом производится на основе F-критерия Фишера. Расчету F-критерия предшествует дисперсионный анализ. В математической статистике дисперсионный анализ рассматривается как самостоятельный инструмент статистического анализа. В эконометрике он применяется как вспомогательное средство для изучения качества регрессионной модели.

Согласно основной идее дисперсионного анализа, общая вариация переменной y вокруг среднего значения $\bar{y}$ раскладывается на две части – «объясненную» и «необъясненную».

Схема дисперсионного анализа имеет вид, представленный в таблице 1.1 (n – число наблюдений, m – число параметров при переменной x).

Таблица 1.1

Вариация	Сумма квадратов отклонений	Число степеней свободы	Дисперсия на одну степень свободы
Общая	$\sum\left(y - \bar{y}\right)^2$	$n-1$	$S_{\text{общ}}^2 = \frac{\sum\left(y - \bar{y}\right)^2}{n-1}$
Факторная	$\sum\left(y_x - \bar{y}\right)^2$	m	$S_{\text{факт}}^2 = \frac{\sum\left(y_x - \bar{y}\right)^2}{m}$
Остаточная	$\sum\left(y - y_x\right)^2$	$n-m-1$	$S_{\text{ост}}^2 = \frac{\sum\left(y - y_x\right)^2}{n-m-1}$

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F-критерия Фишера:

$$F = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2}. \qquad (1.17)$$

Фактическое значение F-критерия Фишера сравнивается с табличным значением $F_{табл} = F_{\alpha}(k_1; k_2)$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$. При этом, если фактическое значение F-критерия больше табличного, то признается статистическая значимость уравнения в целом.

Для парной линейной регрессии $m = 1$, поэтому

$$F = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2} = \frac{\sum\left(y_x - \bar{y}\right)^2}{\sum\left(y - y_x\right)^2} \cdot \left(n - 2\right). \qquad (1.18)$$

Величина F-критерия связана с коэффициентом детерминации r_{xy}^2, и ее можно рассчитать по следующей формуле:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot \left(n - 2\right). \qquad (1.19)$$

В парной линейной регрессии оценивается значимость не только уравнения в целом, но и отдельных его параметров.

Для оценки существенности коэффициента регрессии используется *t*-критерий Стьюдента.

Алгоритм оценки:

1. Найти стандартную ошибку коэффициента регрессии:

$$m_b = \sqrt{\frac{S_{\text{ост}}^2}{\sum\left(x - \bar{x}\right)^2}} = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}}, \qquad (1.20)$$

где $S_{\text{ост}}^2$ – остаточная дисперсия в расчете на одну степень свободы (дисперсия ошибок $\hat{\sigma}^2$):

$$S^2_{\text{ост}} = \hat{\sigma}^2 = \frac{\Sigma(y-\hat{y})^2}{n-m-1} \tag{1.21}$$

2. Определить фактическое значение t-критерия Стьюдента:

$$t_b = \frac{b}{m_b}\ . \tag{1.22}$$

3. Найти критическое (табличное) значение t-критерия по таблице распределения Стьюдента (Приложение1):

$$t_{табл} = t_\alpha(k),$$

где $\alpha = 1 - p$ –уровень значимости (величина, связанная с вероятностью суждения p);

$k = n - m - 1$ –остаточное число степеней свободы.

4. Сравнить фактическое значение t-критерия Стьюдента с критическим.

Если $t_b > t_{\text{табл}}$, то коэффициент регрессии признается существенным, статистически значимым.

Величина стандартной ошибки совместно с t -распределением Стьюдента также применяется для расчета доверительного интервала коэффициента регрессии.

Доверительный интервал для коэффициента регрессии определяется выражением ($b \pm t_{\text{табл}} \cdot m_b$). При этом, границы доверительного интервала для коэффициента регрессии не должны содержать противоречивых результатов, например, $-1{,}5 \le b \le 0{,}8$. Такого рода запись указывает, что истинное значение коэффициента регрессии одновременно содержит положительные и отрицательные величины и даже ноль, чего не может быть.

Алгоритм оценки существенности параметра a аналогичен рассмотренному выше для коэффициента регрессии:

1. Определить стандартную ошибку параметра a по формуле:

$$m_a = \sqrt{S^2_{\text{ост}} \cdot \frac{\sum x^2}{n \cdot \sum\left(x-\bar{x}\right)^2}} = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n}. \tag{1.23}$$

2. Рассчитать фактическое значение t-критерия по формуле:

$$t_a = \frac{a}{m_a}. \tag{1.24}$$

3. Найти критическое значение t-критерия по таблице распределения Стьюдента (Приложение1);

4. Сравнить фактическое значение t-критерия Стьюдента с критическим. Если $t_a > t_{табл}$, то параметр a признается существенным, статистически значимым.

Аналогично определяется значимость линейного коэффициента корреляции. При этом, ошибка коэффициента корреляции m_r определяется по формуле:

$$m_r = \sqrt{\frac{1-r^2}{n-2}}. \quad (1.25)$$

Фактическое значение t-критерия Стьюдента определяется следующим образом:

$$t_r = \frac{r}{m_r}. \quad (1.26)$$

Связь между t-критерием Стьюдента и F-критерием Фишера имеет вид:

$$t_b = t_r = \sqrt{F}.$$

В прогнозных расчетах по уравнению регрессии определяется предсказываемое $Y_{пр}$ значение как точечный прогноз $\hat{Y}_{пр}$ при $x_{пр}$, т.е. путем подстановки в уравнение регрессии $y_x = a + b \cdot x$ соответствующего значения $x_{пр}$. Далее точечный прогноз дополняется расчетом стандартной ошибки $\hat{Y}_{пр}$, т.е. $m_{\hat{y}_{пр}}$:

$$m_{\hat{y}_{пр}} = S_{ост} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{пр} - \bar{x})^2}{\Sigma(x - \bar{x})^2}}, \quad (1.27)$$

и соответственно интервальной оценкой прогнозного значения $Y_{пр}$:

$$\hat{Y}_{пр} - \Delta_{\hat{y}_{пр}} \le Y_{пр} \le \hat{Y}_{пр} + \Delta_{\hat{y}_{пр}},$$

где $\Delta_{\hat{y}_{пр}}$ – предельная ошибка прогнозируемого значения:

$$\Delta_{\hat{y}_{пр}} = m_{\hat{y}_{пр}} \cdot t_{табл}. \quad (1.28)$$

Большинство реальных экономических процессов можно описать с помощью линейной регрессии, однако в некоторых случаях наибольшее приближение к исходным данным дают другие виды зависимостей (нелинейные регрессии).

Нелинейная регрессия и корреляция

Нелинейные регрессии делятся на два класса:

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

Например:

- полиномы разных степеней $y = a + b_1 x + b_2 x^2 + b_3 x^3 + \varepsilon$;
- равносторонняя гипербола $y = a + \dfrac{b}{x} + \varepsilon$.

2. Регрессии, нелинейные по оцениваемым параметрам. Например:

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

Для оценки параметров регрессий нелинейных по объясняющим переменным используется подход, суть которого состоит в замене нелинейных переменных линейными и приведении регрессии к линейному виду (линеаризация).

Модели нелинейные по параметрам делятся на два типа:

1. Внутренне линейные (могут быть приведены к линейному виду);
2. Внутренне нелинейные (не могут быть приведены к линейному виду).

К линеаризованным моделям может быть применен обычный МНК.

Теснота связи для нелинейных регрессий оценивается с помощью индекса корреляции R_{xy}:

$$R_{xy} = \sqrt{1 - \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2}}. \qquad (1.29)$$

Индекс корреляции может находиться в пределах $0 \le R_{xy} \le 1$.

Индекс детерминации для нелинейной регрессии рассчитывается как квадрат индекса корреляции R_{xy}^2.

1.2 Решение типовых задач

Задача 1.1. Имеются данные о себестоимости молока (y, руб.) и средней продуктивности коров (x, ц) по 16 сельскохозяйственным предприятиям региона (таблица 1.2).

Требуется:

1. Рассчитать параметры линейного уравнения регрессии;
2. Оценить тесноту связи с помощью коэффициента корреляции;
3. Рассчитать коэффициент детерминации, среднюю ошибку аппроксимации и оценить качество модели.
4. Рассчитать средний коэффициент эластичности и сделать вывод.
5. С помощью F-критерия Фишера (при α = 0,05) оценить надежность уравнения регрессии.
6. Рассчитать прогнозное значение $\hat{Y}_{пр}$ если прогнозное значение фактора увеличится на 5% от его среднего значения. Определить доверительный интервал прогноза для α = 0,05.

Таблица 1.2

Предприятия	Продуктивность, ц	Себестоимость, руб.
А	13,85	885
Б	11,25	905
В	13,60	845
Г	16,75	814
Д	15,28	875
Е	16,48	795
Ж	19,24	764
З	20,50	720
И	21,30	812
К	20,85	790
Л	22,30	760
М	25,20	703
Н	14,15	910
О	13,60	920
П	20,16	798
Р	20,15	760

Решение:

1. Параметры линейного уравнения регрессии найдем по формулам (1.10)-(1.11):

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^{\ 2}} = \frac{14294{,}8025 - 17{,}7913 \cdot 816{,}00}{14{,}5825} = -15{,}28$$

$$a = \bar{y} - b \cdot \bar{x} = 816{,}00 - (-15{,}285) \cdot 17{,}7913 = 1087{,}9$$

Расчетные данные приведены в таблице 1.3.

Таким образом, линейное уравнение регрессии имеет вид:

$$\hat{y} = 1087{,}9 - 15{,}28x\,.$$

Из уравнения регрессии можно сделать вывод, что при увеличении продуктивности коров на 1 ц себестоимость молока снижается в среднем на 15,28 руб.

2. Рассчитаем линейный коэффициент парной корреляции по формуле (1.12):

$$r_{xy} = -15{,}28 \cdot \frac{3{,}82}{65{,}71} = -0{,}888\,.$$

Рассчитанный коэффициент корреляции говорит о том, что связь между себестоимостью молока и продуктивностью коров сильная и обратно пропорциональная.

3. Коэффициент детерминации равен:

$$r_{xy}^2 = (-0{,}888)^2 = 0{,}789\,,$$

следовательно, 78,9% вариации себестоимости молока объясняется продуктивностью коров, а 21,1% - влияние не учтенных в модели факторов.

Средняя ошибка аппроксимации рассчитывается по формуле (1.16):

По данным таблицы 1.3 средняя ошибка аппроксимации $\overline{A}$ =3,22% находится в допустимых пределах (8–10 %), что говорит о хорошем качестве модели регрессии.

4. Рассчитаем средний коэффициент эластичности:

$$\overline{Э} = b \cdot \frac{\overline{x}}{\overline{y}}\,,$$

$$\overline{Э} = -15{,}28 \cdot \frac{17{,}79}{816{,}00} = -0{,}33\,.$$

Коэффициент эластичности показывает, что при увеличении продуктивности коров на 1% себестоимость молока снижается на 0,33%.

5. Рассчитаем F-критерий Фишера по формуле (1.19):

$$F_{расч} = \frac{0{,}789}{1 - 0{,}789} \cdot 14 = 52{,}35\,.$$

Критическое значение F-критерия найдем по таблице распределения Фишера (Приложение 2):

$$F_{табл} = F_{0,05}(1;14) = 4,60 .$$

Так как $F_{расч} > F_{табл}$, то уравнение регрессии признается статистически значимым.

Таким образом, можно сделать вывод, что рассмотренная модель удовлетворяет требованиям качества и может использоваться для объяснения поведения показателей, а также для прогнозирования.

6. Найдем прогнозное значение себестоимости молока $\hat{Y}_{пр}$ при условии, что значение фактора увеличилось на 5% от среднего уровня:

$$x_{пр} = \bar{x} \cdot 1,05 = 17,7913 \cdot 1,05 = 18,68 \text{ (ц).}$$

Для этого сначала построим точечное значение прогноза:

$$\hat{Y}_{пр} = 1087,9 - 15,28 \cdot x_{пр} = 1087,9 - 15,28 \cdot 18,68 = 802,47 \text{ (руб).}$$

Затем рассчитаем среднюю ошибку прогноза по формуле (1.27):

$$m_{\hat{y}_{пр}} = 32,276 \cdot \sqrt{1 + \frac{1}{16} + \frac{(18,68 - 17,79)^2}{233,32}} = 33,322$$

где $S_{ост} = \sqrt{\frac{\Sigma(y - \hat{y})^2}{n - m - 1}} = \sqrt{\frac{114584,723}{14}} = 32,276$.

Предельная ошибка прогноза:

$$\Delta_{\hat{y}_{пр}} = m_{\hat{y}_{пр}} \cdot t_{табл} = 33,322 \cdot 2,145 = 71,48 ,$$

при вероятности $p = 0,95$ получим $t_{табл} = t_{0,05}(14) = 2,145$.

Строим доверительный интервал:

$$\hat{Y}_{пр} - \Delta_{\hat{y}_{пр}} \le Y_{пр} \le \hat{Y}_{пр} + \Delta_{\hat{y}_{пр}}$$

$$802,47 - 71,48 \le Y_{пр} \le 802,47 + 71,48$$

$$730,99 \le Y_{пр} \le 873,95$$

Выполненный прогноз является надежным (p = 0,95), и достаточно точным, так как:

$$\frac{Y_{np\,\max}}{Y_{np\,\min}} = \frac{873{,}95}{730{,}99} = 1{,}20.$$

Задача 1.2. На основе данных таблицы 1.2 требуется:

1. Рассчитать параметры уравнения регрессии $y = a + b\sqrt{x} + \varepsilon$;
2. Оценить тесноту связи с помощью индекса корреляции;
3. Рассчитать индекс детерминации, среднюю ошибку аппроксимации и оценить качество модели.
4. С помощью F-критерия Фишера (при α = 0,05) оценить надежность уравнения регрессии.

Решение:

Для расчета параметров уравнения воспользуемся методом «замены переменных»: $z = \sqrt{x}$.

Тогда $\hat{y} = a + bz$.

Для расчетов используем данные таблицы 1.4.

Значения параметров регрессии составили:

$$b = \frac{\overline{zy} - \bar{z} \cdot \bar{y}}{\sigma_z^{\ 2}} = \frac{3394{,}920 - 4{,}1932 \cdot 816{,}00}{0{,}2086} = -128{,}15;$$

$$a = \bar{y} - b \cdot \bar{z} = 816{,}00 - (-128{,}15) \cdot 4{,}1932 = 1353{,}36.$$

Получили уравнение регрессии:

$$\hat{y} = 1353{,}36 - 128{,}15\sqrt{x}.$$

Для оценки тесноты связи рассчитаем индекс корреляции по формуле (1.29):

$$R_{xy} = \sqrt{1 - \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2}} = \sqrt{1 - \frac{14384{,}326}{69078}} = 0{,}8898.$$

Значение индекса корреляции близко к единице, связь между признаками сильная.

Индекс детерминации равен:

$$R_{xy}^2 = (0{,}8898)^2 = 0{,}7917,$$

следовательно, данная модель объясняет 79,17% вариации себестоимости молока, а 20,83% - влияние не учтенных в модели факторов.

В данном случае средняя ошибка аппроксимации $\overline{A}$ =3,27%, что меньше допустимого предела, т.е. качество данной модели также хорошее.

Найдем расчетное значение F-критерия Фишера:

$$F_{расч} = \frac{R_{xy}^2}{1 - R_{xy}^2} \cdot (n-2),$$

$$F_{расч} = \frac{0{,}7917}{1 - 0{,}7917} \cdot 14 = 53{,}21.$$

Критическое значение F-критерия Фишера найдем по таблице (Приложение 2):

$$F_{табл} = F_{0{,}05}(1;14) = 4{,}60.$$

Так как $F_{расч}$ = 53,21 > $F_{табл}$ = 4,60, то уравнение регрессии признается статистически значимым.

Характеристики рассмотренной модели позволяют сделать вывод, что она несколько лучше линейной функции описывает взаимосвязь себестоимости молока и продуктивности коров. Однако, качество данной модели незначительно выше, поэтому как более простую целесообразно использовать линейную функцию.

1.3 Вопросы для самопроверки и задания для самостоятельной работы

I. Вопросы для самопроверки

1. Определение эконометрики. Связь эконометрики с другими науками.
2. Предмет исследования и методы эконометрики.
3. Основные этапы эконометрического исследования.
4. Типы данных, используемые в эконометрическом исследовании и их характеристика.
5. Основные виды эконометрических моделей.

6. Методы выбора общего вида регрессионной модели.
7. Уравнение парной линейной регрессии.
8. Сущность метода наименьших квадратов для определения параметров линейной регрессии.
9. Интерпретация параметров линейной регрессии.
10. Оценка силы связи, свойства коэффициента корреляции.
11. Показатели качества регрессионной модели
12. Суть дисперсионного анализа.
13. Оценка статистической значимости уравнения регрессии и его параметров.
14. Построение доверительного интервала для коэффициента регрессии.
15. Построение доверительного интервала прогноза по уравнению линейной регрессии.

II. Задания для самостоятельной работы

1. По данным 5 наблюдений (таблица 1.5) требуется:

1) построить линейную модель регрессии, описывающую зависимость У от Х;
2) оценить дисперсию ошибок;
3) рассчитать коэффициент корреляции и оценить тесноту связи;
4) оценить качество модели с помощью коэффициента детерминации;
5) сделать выводы.

Таблица 1.5

i	X_i	Y_i
1	1,0	1,25
2	1,5	1,40
3	3,0	1,50
4	4,5	1,75
5	5,0	2,25

2. Имеются данные о себестоимости и затратах труда на 1 ц производства овощей по 10 сельскохозяйственным предприятиям. Данные представлены в таблице 1.6.
Требуется:

1) построить линейную модель регрессии, описывающую зависимость себестоимости от затрат труда;
2) рассчитать коэффициент корреляции и оценить тесноту связи;
3) рассчитать коэффициент детерминации, среднюю ошибку аппроксимации и оценить качество модели;
4) оценить статистическую значимость уравнения регрессии в целом по F-критерию Фишера;
5) оценить значимость параметров уравнения регрессии по t-критерию Стьюдента;

6) оценить значимость коэффициента корреляции по t-критерию Стьюдента;
7) сделать выводы.

Таблица 1.6

Предприятие	Себестоимость, тыс.руб.	Затраты труда на 1 ц, чел.-час.
А	0,280	14,3
Б	0,276	9,9
В	0,270	8,9
Г	0,245	8,8
Д	0,246	7,8
Е	0,258	9,8
Ж	0,280	10,1
З	0,275	11,2
И	0,268	12,3
К	0,274	13,2

3. На основе данных бюджетного обследования случайно выбранных семей (таблица 1.7) требуется:

1) построить линейную модель регрессии, описывающую зависимость накоплений семьи от величины дохода;
2) отобразить на графике исходные данные и результаты моделирования:
3) оценить качество построенной модели;
4) спрогнозировать накопления семьи, имеющей доход 42 тыс.руб., дать интервальную оценку прогноза;
5) сделать выводы.

Таблица 1.7

Наблюдение	Накопления, тыс.руб.	Доход, тыс.руб.
1	3,0	40
2	6,0	55
3	5,0	45
4	3,5	30
5	1,5	30
6	4,5	50
7	2,0	35

. В таблице 1.8 приведены данные о производительности труда (Y) и уровне механизации работ (X) для 14 предприятий.

Требуется:

1) построить линейную модель регрессии, выражающую зависимость между производительностью труда и уровнем механизации работ;
2) оценить качество модели;
3) сделать прогноз о производительности труда при 58-процентном уровне механизации работ;
4) отобразить на графике исходные данные и результаты моделирования.

Таблица 1.8

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14
X, %	32	30	36	40	41	47	56	54	60	55	61	67	69	76
Y, т/ч	20	24	28	30	31	33	34	37	38	40	41	43	45	48

5. В результате наблюдения получены 16 пар значений (X;У) и определены следующие выборочные моменты: $\Sigma y^2 = 526$; $\Sigma x^2 = 657$; $\Sigma xy = 492$; $\Sigma y = 64$; $\Sigma x = 96$. Оцените регрессию $Y_i = a + b \cdot X_i + \varepsilon_i$

6. Зависимость прироста урожайности зерновых от дозы внесения удобрения по данным 13 предприятий характеризуется следующим образом:
уравнение регрессии: $\hat{y}_x = -0{,}135 + 9{,}858 \cdot x$
сумма квадратов отклонений: $\sum(y - \bar{y})^2 = 590{,}540$; $\sum(\hat{y} - \bar{y})^2 = 584{,}053$
Требуется: провести дисперсионный анализ и оценить статистическую значимость уравнения регрессии по F-критерию Фишера. Результаты оформить в виде таблицы 1.9.

Таблица 1.9

Вариация результата	Число степеней свободы	Сумма квадратов отклонений	Дисперсия на одну степень свободы	$F_{расч}$	$F_{табл}$
Общая					
Факторная					
Остаточная					

7. Проведена серия опытов по определению влияния дозы внесения органических удобрений на повышение урожайности пшеницы (таблица 1.10).
Требуется: 1)построить линейную, гиперболическую и параболическую модели регрессии прироста урожайности от дозы внесения удобрений; 2) дать сравнительную оценку качества построенных моделей с помощью коэффициента детерминации и ошибки аппроксимации.

Таблица 1.10

Наблюдение	Внесённая доза, ц/га	Прирост урожайности, ц/га
1	0,342	2,10
2	0,417	4,70
3	0,675	6,05
4	0,867	8,65
5	1,000	10,00
6	1,158	12,60
7	1,283	12,08
8	1,500	14,68
9	1,733	16,65
10	2,008	19,25
11	2,083	19,27
12	2,242	19,28
13	2,508	18,32

2. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

2.1 Необходимый теоретический минимум

Модель множественной регрессии – модель, содержащая более одной объясняющей переменной, т.е. модель вида:

$$y = \hat{f}(x_1, x_2, x_3, \ldots, x_n), \tag{2.1}$$

где y – зависимая переменная (результативный признак);

$x_1, x_2, x_3, \ldots, x_n$ – независимые переменные (факторные признаки).

Основная цель множественной регрессии – построить модель с оптимальным числом факторов, определив влияние каждого фактора в отдельности, а также совокупное их воздействие на моделируемый показатель.

Непосредственное построение уравнения множественной регрессии начинается с этапа спецификации, который включает в себя:

1) отбор факторов, которые целесообразно включить в модель;
2) выбор вида уравнения регрессии.

Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям:

1) все факторы должны быть количественно измеримы;
2) факторы должны быть независимыми.

Если между факторами существует высокая корреляционная или функциональная связь, то нельзя определить их изолированное влияние на результат, а параметры уравнения регрессии оказываются неинтерпретируемыми.

Отбор факторов производится в два этапа:

1) анализ экономической сущности явления и выделение факторов, влияющих на результат;
2) исключение высококоррелируемых между собой факторов.

Как уже отмечалось, модель множественной регрессии должна содержать оптимальное число факторов.

Если построена модель с набором p факторов, то для нее можно определить показатель детерминации R^2, который показывает долю вариации результата под влиянием p факторов. При дополнительном включении в модель $(p+1)$ фактора коэффициент детерминации должен возрастать, т.е.

$$R^2_{p+1} > R^2_p .$$

Если этого не происходит и данные показатели мало отличаются друг от друга, то включаемый в модель фактор x_{p+1} не улучшает модель и практически является лишним. Включение в модель такого фактора приводит к статистической незначимости параметров регрессии.

Для оценки степени влияния факторов друг на друга используется матрица парных коэффициентов корреляции:

$$M = \begin{pmatrix} r_{x_1x_1} & r_{x_2x_1} & \dots & r_{x_nx_1} \\ r_{x_1x_2} & r_{x_2x_2} & \dots & r_{x_nx_2} \\ \dots & \dots & \dots & \dots \\ r_{x_1x_n} & r_{x_2x_n} & \dots & r_{x_nx_n} \end{pmatrix}. \qquad (2.2)$$

Где $r_{x_ix_j}$ - парные коэффициенты корреляции факторов.

При этом, если факторы не коррелируют между собой, то определитель матрицы *det M* = 1, а если между факторами существует функциональная связь – *det M* = 0. Таким образом, чем ближе значение *det M* к нулю, тем сильнее зависимость между факторами и ненадежнее результаты множественной регрессии.

Если $r_{x_ix_j} \geq 0{,}7$, то две переменные явно коллинеарны (т.е. находятся между собой в линейной зависимости).

Если более чем два фактора связаны между собой линейной зависимостью, то имеет место мультиколлинеарность факторов.

Коллинеарность факторов нарушает условие их независимости. Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из модели. Исключают обычно тот фактор, который при достаточно тесной связи с результатом имеет наибольшую тесноту связи с другими факторами.

Оценка параметров уравнения множественной регрессии

В практических приложениях часто используются:

1) линейная модель множественной регрессии:

$$\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_nx_n. \qquad (2.3)$$

2) модели на основе производственных функций:

$$\hat{y} = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \ldots \cdot x_n^{b_n}. \qquad (2.4)$$

Функция (2.4) является нелинейной по параметрам. Однако она может быть линеаризована посредством логарифмирования обеих ее частей.

Параметры линейного уравнения множественной регрессии, также как и в случае парной регрессии, оцениваются МНК. Применение МНК позволяет получить следующую систему нормальных уравнений:

$$\begin{cases} \sum y = n \cdot a + b_1 \cdot \sum x_1 + b_2 \sum x_2 + ... + b_n \cdot \sum x_n; \\ \sum yx_1 = a \cdot \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + ... + b_n \sum x_1 x_n; \\ \sum yx_2 = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 + ... + b_n \sum x_2 x_n; \\ .... \\ \sum yx_n = a \sum x_n + b_1 \sum x_n x_1 + b_2 \sum x_n x_2 + ... + b_n \sum x_n^2. \end{cases} \quad (2.5)$$

В случае с двумя независимыми переменными соответствующая система нормальных уравнений примет вид:

$$\begin{cases} \sum y = n \cdot a + b_1 \sum x_1 + b_2 \sum x_2; \\ \sum yx_1 = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2; \\ \sum yx_2 = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2. \end{cases} \quad (2.6)$$

Систему нормальных уравнений можно решить любым способом, предназначенным для решения системы линейных уравнений (метод подстановки, метод обратной матрицы, метод Крамера, Гаусса, Жордано-Гаусса).

Линейную регрессию, заданную уравнением вида (2.3), называют регрессией в естественной форме (естественном масштабе). Параметры уравнения регрессии в естественной форме можно определить методом стандартизации, т.е. на основе уравнения регрессии в стандартизованном масштабе.

Уравнение регрессии в стандартизованном масштабе – уравнение, построенное на основе матрицы парных коэффициентов корреляции и имеющее вид:

$$t_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2} + ... + \beta_n \cdot t_{x_n} + \varepsilon, \quad (2.7)$$

где t_y, t_{x_1}, t_{x_2}, …, t_{x_n} – стандартизованные переменные;

β_1, β_2,…, β_n – стандартизованные коэффициенты регрессии.

Стандартизованные переменные регрессии – это такие переменные, для которых среднее значение равно нулю, а среднее квадратическое отклонение - единице, т.е.:

$$\bar{t}_y = \bar{t}_{x_1} = \bar{t}_{x_2} = ... = \bar{t}_{x_n} = 0,$$

$$\sigma_{t_y} = \sigma_{t_{x_1}} = ... = \sigma_{t_{x_n}} = 1.$$

Стандартизованные переменные вычисляются по формулам:

$$t_y = \frac{y - \bar{y}}{\sigma_y}, \quad t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}} (i = 1; n).$$

К уравнению регрессии в стандартизованном масштабе также может быть применен МНК, который дает следующую систему нормальных уравнений:

$$\begin{cases} R_{yx_1} = \beta_1 + \beta_2 R_{x_2 x_1} + \beta_3 R_{x_3 x_1} + ... + \beta_n R_{x_n x_1}; \\ R_{yx_2} = \beta_1 R_{x_2 x_1} + \beta_2 + \beta_3 R_{x_3 x_2} + ... + \beta_n R_{x_n x_2}; \\ R_{yx_n} = \beta_1 R_{x_1 x_n} + \beta_2 R_{x_2 x_n} + \beta_3 R_{x_n x_3} + ... + \beta_n. \end{cases} \quad (2.8)$$

Решив данную систему, получим β–коэффициенты.

Решение системы (2.8) для случая с двумя факторами имеет вид:

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{1 - r^2_{x_1 x_2}}, \quad (2.9)$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{1 - r^2_{x_1 x_2}}. \quad (2.10)$$

Связь между коэффициентами множественной регрессии в естественной форме и стандартизованными коэффициентами выражается формулой:

$$b_i = \beta_i \cdot \frac{\sigma_y}{\sigma_{x_i}}. \quad (2.11)$$

Параметр a находится по формуле:

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - ... - b_n \bar{x}_n. \quad (2.12)$$

Для множественной регрессии тесноту совместного влияния факторов на результат оценивает индекс множественной корреляции:

$$R_{yx_1x_2\ldots x_n} = \sqrt{1-\frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}}\ . \qquad (2.13)$$

Если известны значения парных коэффициентов корреляции, то может быть рассчитан коэффициент множественной корреляции:

$$R_{yx_1x_2} = \sqrt{\frac{r^2{}_{yx_1} + r^2{}_{yx_2} - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1-r^2{}_{x_1x_2}}}\ . \qquad (2.14)$$

Для регрессии в стандартизованной форме формула для нахождения индекса множественной корреляции примет вид:

$$R_{yx_1x_2\ldots x_n} = \sqrt{\sum \beta_i \cdot r_{yx_i}}\ . \qquad (2.15)$$

Свойства индекса множественной корреляции:

1. Значения индекса множественной корреляции изменяются на отрезке [0;1]:

$$R_{yx_1\ldots x_n} \in [0;1].$$

2. Значение индекса множественной корреляции должно быть больше или равно максимальному парному индексу корреляции:

$$R_{yx_1\ldots x_n} \geq r_{yx_i}\ (i=1;n),$$

где r_{yx_i} - наибольший по величине коэффициент парной корреляции.

Индекс множественной детерминации характеризует долю вариации результативного признака, объясняемую включенными в модель факторами, и рассчитывается по формуле:

$$R^2{}_{yx_1x_2} = 1-\frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2} \qquad (2.16)$$

Если число факторов *m*, включенных в модель, приближается к числу наблюдений n, то необходимо скорректировать индекс детерминации на потерю числа степеней свободы.

Скорректированный индекс детерминации рассчитывается следующим образом:

$$\overline{R}^2 = 1-(1-R^2)\cdot\frac{n-1}{n-m-1} \tag{2.17}$$

Частная регрессия и корреляция

На основе линейного уравнения множественной регрессии (2.3) могут быть найдены частные уравнения регрессии:

$$\begin{cases} y_{x_1\cdot x_2, x_3,\dots, x_p} = f(x_1), \\ y_{x_2\cdot x_1, x_3,\dots, x_p} = f(x_2), \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots, \\ y_{x_p\cdot x_1, x_2,\dots, x_{p-1}} = f(x_p). \end{cases} \tag{2.18}$$

Данные уравнения регрессии связывают результативный признак с соответствующим фактором x при закреплении других учитываемых во множественной регрессии факторов на среднем уровне. Частные уравнения регрессии имеют следующий вид:

$$y_{x_1\backslash x_2, x_3,\dots, x_p} = a + b_1\cdot x_1 + b_2\cdot\overline{x}_2 + b_3\cdot\overline{x}_3 + \dots + b_p\cdot\overline{x}_p + \varepsilon;$$

$$y_{x_2\backslash x_1, x_3,\dots, x_p} = a + b_1\cdot\overline{x}_1 + b_2\cdot x_2 + b_3\cdot\overline{x}_3 + \dots + b_p\cdot\overline{x}_p + \varepsilon;$$

$$y_{x_p\backslash x_1, x_2,\dots, x_{p-1}} = a + b_1\cdot\overline{x}_1 + b_2\cdot\overline{x}_2 + \dots + b_{p-1}\cdot\overline{x}_{p-1} + b_p\cdot x_p + \varepsilon.$$

При подстановке в эти уравнения средних значений соответствующих факторов они принимают вид парных уравнений линейной регрессии, т.е. имеем:

$$\begin{cases} \hat{y}_{x_1\backslash x_2 x_3\dots x_p} = A_1 + b_1 x_1, \\ \hat{y}_{x_2\backslash x_1 x_3\dots x_p} = A_2 + b_2 x_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots, \\ \hat{y}_{x_p\backslash x_1 x_2\dots x_{p-1}} = A_p + b_p x_p, \end{cases} \tag{2.19}$$

где

$$\begin{cases} A_1 = a + b_2 \cdot \bar{x}_2 + b_3 \cdot \bar{x}_3 + \ldots + b_p \cdot x_p, \\ A_2 = a + b_1 \cdot \bar{x}_1 + b_3 \cdot \bar{x}_3 + \ldots + b_p \cdot \bar{x}_p, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ A_p = a + b_1 \cdot \bar{x}_1 + b_2 \cdot \bar{x}_2 + \ldots + b_{p-1} \cdot \bar{x}_{p-1}. \end{cases}$$

В отличие от парной регрессии частные уравнения регрессии характеризуют изолированное влияние фактора на результат, ибо другие факторы закреплены на неизменном уровне. Эффекты влияния других факторов присоединены в них к свободному члену уравнения множественной регрессии. Это позволяет на основе частных уравнений регрессии определять частные коэффициенты эластичности:

$$Э_{y_{xi}} = b_i \cdot \frac{x_i}{\hat{y}_{x_i / x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_p}}, \tag{2.20}$$

где b_i – коэффициент регрессии для фактора x_i в уравнении множественной регрессии;

$\hat{y}_{x_i / x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_p}$ - значение результативного признака, найденное из частного уравнения регрессии для определенного значения x_i.

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

Показатели частной корреляции представляют собой отношение сокращения остаточной дисперсии за счет дополнительного включения в анализ нового фактора к остаточной дисперсии, имевшей место до введения его в модель.

Остаточную дисперсию можно выразить через показатель детерминации, например:

$$r_{yx_1 / x_2} = \sqrt{\frac{s^2_{yx_2} - s^2_{yx_1x_2}}{s^2_{yx_2}}} = \sqrt{1 - \frac{s^2_{yx_1x_2}}{s^2_{yx_2}}} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_2}}}, \tag{2.21}$$

$$r_{yx_2 / x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_1}}}. \tag{2.22}$$

Рассмотренные показатели частной корреляции называют коэффициентами (индексами) частной корреляции первого порядка, так как они фиксируют тесноту связи двух переменных при закреплении (элиминировании влияния) одного фактора.

Если рассматривается регрессия с числом факторов p, то возможны частные коэффициенты корреляции не только первого, но и второго, третьего, …, (p-1) порядка, т.е. влияние фактора x_1 можно оценить при разных условиях независимости действия других факторов:

r_{yx_1/x_2} - при постоянном действии фактора x_2;

r_{yx_1/x_2x_3} - при постоянном действии факторов x_2 и x_3;

$r_{yx_1/x_2...x_p}$ - при неизменном действии всех факторов, включенных в уравнение регрессии.

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Соответственно коэффициенты парной корреляции называются коэффициентами нулевого порядка. Коэффициенты частной корреляции более высоких порядков можно определить через коэффициенты частной корреляции более низких порядков по рекуррентной формуле. Например, при двух факторах и $i=1$ данная формула имеет вид:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1-r^2_{yx_2}) \cdot (1-r^2_{x_1x_2})}}. \tag{2.23}$$

Соответственно при двух факторах и $i=2$ частный коэффициент корреляции y с фактором x_2 можно определить по формуле:

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1-r^2_{yx_1}) \cdot (1-r^2_{x_1x_2})}}. \tag{2.24}$$

Для уравнения регрессии с тремя факторами частные коэффициенты корреляции второго порядка определяются на основе частных коэффициентов корреляции первого порядка.

Например, при $i=1$ получаем формулу для расчета r_{yx_1/x_2x_3}:

$$r_{yx_1/x_2x_3} = \frac{r_{yx_1/x_2} - r_{yx_3/x_2} \cdot r_{x_1x_3/x_2}}{\sqrt{(1 - r^2_{yx_3/x_2}) \cdot (1 - r^2_{x_1x_3/x_2})}} . \qquad (2.25)$$

В практических исследованиях предпочтение отдают показателям частной корреляции самого высокого порядка, так как именно эти показатели являются дополнением к уравнению множественной регрессии.

В общем виде при наличии *p* факторов для уравнения (2.3) коэффициент частной корреляции, измеряющий влияние на *y* фактора x_i при неизменном уровне других факторов, можно определить по формуле:

$$r_{yx_i/x_1x_2...x_{i-1}x_{i+1}...x_p} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2...x_i...x_p}}{1 - R^2_{yx_1x_2...x_{i-1}x_{i+1}...x_p}}} , \qquad (2.26)$$

где $R^2_{yx_1x_2...x_i...x_p}$ - множественный коэффициент детерминации всего комплекса *p* факторов с результатом;

$R^2_{yx_1x_2...x_{i-1}x_{i+1}...x_p}$ - тот же показатель детерминации, но без введения в модель фактора x_i

При *i*=1 формула коэффициента частной корреляции примет вид:

$$r_{yx_1/x_2...x_p} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2...x_p}}{1 - R^2_{yx_2...x_p}}} . \qquad (2.27)$$

Данный коэффициент частной корреляции позволяет измерить тесноту связи между *y* и x_1 при неизменном уровне всех других факторов, включенных в уравнение регрессии.

Включение в модель неколичественных факторов

Как уже говорилось выше, все факторы, включаемые в модель множественной регрессии, должны быть количественно измеримы. Вместе с тем, часто возникает необходимость учесть в модели влияние качественного признака.

Качественными факторами аграрного производства, например, являются такие признаки, как природная зона, форма собственности предприятий, преобладающее производственное направление (отрасль) и другие.

Если игнорировать влияние качественных признаков, то их влияние либо уйдет в остаточную вариацию, ухудшив модель, либо станет смешиваться с влиянием тех или иных количественных факторов, искажая меру их влияния.

Для введения качественных признаков в модель, их следует преобразовать в количественные. В таких случаях качественные градации признака можно закодировать специальными переменными, называемыми в статистической и эконометрической литературе «фиктивными» или «структурными», так как они отражают неоднородность качественной структуры совокупности.

Обычно в моделях фиктивная переменная отражает два противоположных состояния качественного фактора. В этом случае фиктивная переменная принимает два значения: 0 – если фактор действует, 1 – если фактор не действует. Если же число значений качественного фактора больше двух, то в модель вводится несколько фиктивных переменных. Их число должно быть на единицу меньше числа значений качественного фактора.

Например, необходимо построить модель рентабельности по данным n предприятий различных форм собственности. При этом, форма собственности является неколичественным (качественным) фактором, который необходимо включить в модель вместе с количественными факторами. Исходные данные для моделирования, включая закодированные значения фиктивных переменных(U_j), представлены в таблице 2.1.

Таблица 2.1

Форма собственности	Единица совокупности	Количественные признаки					Структурные переменные	
		y	x_1	x_2	…	x_k	U_1	U_2
Государственная	1 2 3 . .	Значения количественных признаков					0 0 . . 0	0 0 . . 0
Частная	. . . . . .	Значения количественных признаков					1 1 . . 1 1	0 0 . . 0 0
Кооперативная	. . . . . n	Значения количественных признаков					0 0 . . 0 0	1 1 . . 1 1

В результате решения данной задачи обычным МНК будет получена модель вида:

$$\hat{y} = a + b_1 x_1 + ... + b_k x_k + b_{k+1} x_{k+1} + b_{k+2} x_{k+2} \; ,$$

где x_{k+1} соответствует переменной U_1,

x_{k+2} - переменной U_2.

Перепишем модель в специальных обозначениях для понимания ее сути:

$$\hat{y} = a + b_1 x_1 + ... + b_k x_k + c_1 U_1 + c_2 U_2 \qquad (2.28)$$

Значение коэффициентов при структурных переменных таково: коэффициент c_1 означает, что предприятия частной формы собственности при тех же значениях количественных факторов $x_1 ... x_k$ имеют рентабельность на c_1 больше, чем предприятия государственные, которые приняты за базу сравнения (не имеют структурных переменных U_1 и U_2). Предприятия кооперативной формы собственности имеют рентабельность, на c_2 большую, чем государственные. Конечно, величины c_1 и c_2 могут быть как положительными, так и отрицательными.

Вместо общей модели можно записать три частные модели для предприятий отдельных групп по формам собственности, присоединяя коэффициент при структурной переменной к свободному члену уравнения:

а) для предприятий государственного сектора

$$\hat{y} = a + b_1 x_1 + ... + b_k x_k \; ;$$

б) для предприятий частного сектора

$$\hat{y} = (a + c_1) + b_1 x_1 + ... + b_k x_k \; ;$$

в) для предприятий кооперативного сектора

$$\hat{y} = (a + c_2) + b_1 x_1 + ... + b_k x_k \; .$$

Оценка значимости уравнения множественной регрессии

Для оценки значимости уравнения множественной регрессии в целом используют общий F-критерий Фишера. Фактическое (расчетное) значение F-критерия находят по формуле:

$$F_{факт} = \frac{R^2_{yx_1x_2...x_m}}{1 - R^2_{yx_1x_2...x_m}} \cdot \frac{n - m - 1}{m}, \qquad (2.29)$$

где $R^2_{yx_1x_2...x_m}$ - множественный коэффициент (индекс) детерминации;

n – число единиц совокупности;

m – число независимых переменных (факторов).

Фактическое значение F-критерия Фишера сравнивается с табличным значением $F_{табл} = F_{\alpha}(k_1; k_2)$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$. При этом, если фактическое значение F-критерия больше табличного, то признается статистическая значимость уравнения регрессии.

Оценка значимости дополнительного включения фактора

Включение в модель дополнительного фактора целесообразно в том случае, когда фактор, вошедший в модель, может существенно увеличить долю объясненной вариации результативного признака. Для оценки значимости улучшения качества модели, после включения дополнительного фактора, служит частный F-критерий.

Например, для модели множественной регрессии с двумя факторами, значимость фактора x_1 как дополнительно включенного в модель после фактора x_2 оценивается следующим образом:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{n - m - 1}{1}. \qquad (2.30)$$

Соответственно, для оценки значимости включения фактора x_2 после x_1 частный F-критерий будет вычисляться по формуле:

$$F_{x_2} = \frac{R^2_{yx_1x_2} - r_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot \frac{n - m - 1}{1}. \qquad (2.26)$$

Если фактическое значение F-критерия больше критического, то дополнительное введение в модель фактора статистически оправданно (значимо).

Применение многофакторных регрессионных моделей для анализа деятельности предприятий и прогнозирования

Прогнозирование на основе регрессионной модели исходит из предположения (гипотезы), что факторы управляемы и могут принять то или иное плановое, ожидаемое значение, а прочие неизвестные условия сохранятся на среднем по совокупности уровне. Управляемость факторов не означает, что при прогнозе в модель можно подставлять любые их значения. Уравнение регрессии отражает те условия, которые существовали в совокупности, по данным которой уравнение получено. Поэтому рекомендуется при прогнозировании по уравнению регрессии не выходить за пределы реально наблюдаемых значений факторов в совокупности или выходить за эти границы не более чем на 10-15% средних величин.

Точечный прогноз представляет собой математическое ожидание (среднюю) возможных значений прогнозируемого признака. Необходимо дополнить точечный прогноз расчетом доверительных границ с достаточно большой вероятностью.

2.2 Решение типовых задач

Задача 2.1. Имеются данные 5 наблюдений (таблица 2.2).

Требуется:

1. Построить линейную модель множественной регрессии.
2. Оценить дисперсию ошибок.
3. Оценить адекватность модели с помощью коэффициента множественной детерминации.

Таблица 2.2

№пп	y	x_1	x_2
1	1,0	0,6	1,2
2	2,0	0,5	1,1
3	0,5	0,4	1,0
4	1,5	0,5	1,3
5	3,0	0,7	1,4

Решение:

1. Линейная модель множественной регрессии, описывающая зависимость y от x_1 и x_2, имеет вид:

$$\hat{y} = a + b_1 x_1 + b_2 x_2 .$$

Параметры $a; b_1; b_2$ - есть решение следующей системы нормальных уравнений:

$$\begin{cases} \sum y = n \cdot a + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum y x_1 = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \\ \sum y x_2 = a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 . \end{cases}$$

Для расчета параметров модели составим вспомогательную таблицу 2.3.

Таблица 2.3 - Вспомогательная таблица для расчета параметров модели

№	y	x_1	x_2	yx_1	x_1^2	x_1x_2	yx_2	x_2^2
1	1,0	0,6	1,20	0,60	0,36	0,72	1,20	1,44
2	2,0	0,5	1,10	1,00	0,25	0,55	2,20	1,21
3	0,5	0,4	1,00	0,20	0,16	0,40	0,50	1,00
4	1,5	0,5	1,30	0,75	0,25	0,65	1,95	1,69
5	3,0	0,7	1,40	2,10	0,49	0,98	4,20	1,96
Σ	**8,0**	**2,7**	**6,00**	**4,65**	**1,51**	**3,30**	**10,05**	**7,30**

Подставим значения, рассчитанные в таблице 2.3, и получим систему уравнений:

$$\begin{cases} 8 = 5a + 2{,}7b_1 + 6b_2 \\ 4{,}65 = 2{,}7a + 1{,}51b_1 + 3{,}3b_2 \\ 10{,}05 = 6a + 3{,}3\,b_1 + 7{,}3b_2 \end{cases}$$

Решим данную систему по формулам Крамера:

$$a = \frac{\Delta a}{\Delta};\ b_1 = \frac{\Delta b_1}{\Delta};\ b_2 = \frac{\Delta b_2}{\Delta},$$

где $$\Delta = \begin{vmatrix} 5 & 2{,}7 & 6 \\ 2{,}7 & 1{,}51 & 3{,}3 \\ 6 & 3{,}3 & 7{,}3 \end{vmatrix} = 0{,}008,$$

$$\Delta a = \begin{vmatrix} 8 & 2{,}7 & 6 \\ 4{,}65 & 1{,}51 & 3{,}3 \\ 10{,}05 & 3{,}3 & 7{,}3 \end{vmatrix} = -0{,}025,$$

$$\Delta b_1 = \begin{vmatrix} 5 & 8 & 6 \\ 2{,}7 & 4{,}65 & 3{,}3 \\ 6 & 10{,}05 & 7{,}3 \end{vmatrix} = 0{,}030,$$

$$\Delta b_2 = \begin{vmatrix} 5 & 2{,}7 & 8 \\ 2{,}7 & 1{,}51 & 4{,}65 \\ 6 & 3{,}3 & 10{,}05 \end{vmatrix} = 0{,}018.$$

Подставляя найденные значения определителей в формулы Крамера, получим:

$$a = \frac{-0{,}025}{0{,}008} = -3{,}125,$$

$$b_1 = \frac{0{,}03}{0{,}008} = 3{,}75,$$

$$b_2 = \frac{0{,}018}{0{,}008} = 2{,}25.$$

Таким образом, получим следующую регрессионную модель:

$$\hat{y} = -3{,}125 + 3{,}75x_1 + 2{,}25x_2.$$

2. Для расчета дисперсии ошибок необходимо найти сумму квадратов отклонений фактических значений результативного признака от теоретических (таблица 2.4).

Таблица 2.4

№	ŷ	y-ŷ	(y-ŷ)^2	y-y$_{ср}$	(y-y$_{ср}$)^2
1	1,825	-0,825	0,681	-0,6	0,36
2	1,225	0,775	0,601	0,4	0,16
3	0,625	-0,125	0,016	-1,1	1,21
4	1,675	-0,175	0,031	-0,1	0,01
5	2,650	0,350	0,123	1,4	1,96
Σ	**8,000**	**0,000**	**1,452**	**0,0**	**3,70**

Оценка дисперсии ошибок ε_i представляет собой остаточную дисперсию в расчете на одну степень свободы и находится по формуле (1.21):

$$\hat{\sigma}^2 = \frac{1{,}452}{5-2-1} = 0{,}726.$$

3. Оценим адекватность модели с помощью коэффициента множественной детерминации по формуле (2.14):

Подставляя расчетные данные из таблицы 2.4, получим:

$$R^2 yx_1x_2 = 1 - \frac{1{,}452}{3{,}70} = 0{,}608.$$

Тогда коэффициент множественной корреляции:

$$R_{yx_1x_2} = \sqrt{R^2 yx_1x_2} = \sqrt{0{,}608} = 0{,}78.$$

Таким образом, модель в целом адекватна, так как связь между признаками высокая, причем 60,8% вариации зависимой переменной объясняется факторами, учтенными в модели.

Задача 2.2. Имеются данные о себестоимости (y, руб.), урожайности (x_1, ц/га) и трудоемкости (x_2, чел.-час.) производства зерновых по 19 сельскохозяйственным предприятиям (таблица 2.5).

Требуется:

1. Рассчитать параметры линейного уравнения множественной регрессии. Сделать вывод.
2. Оценить тесноту связи результативного признака с каждым из факторов, а также тесноту связи между факторами.
3. Оценить тесноту связи между всеми признаками.
4. Оценить качество модели с помощью коэффициента множественной детерминации.

Таблица 2.5

Предприятия	Себестоимость, руб.	Урожайность, ц/га	Трудоемкость, чел.-час.
1	336	6,1	3,1
2	213	10,9	4,5
3	138	16,4	2,7
4	624	3,5	5,1
5	162	12,6	2,0
6	253	6,6	1,9
7	617	2,9	8,3
8	260	6,2	3,7
9	553	4,7	2,1
10	270	3,7	2,5
11	178	9,3	2,4
12	260	11,1	3,1
13	270	7,2	2,8
14	322	9,3	2,4
15	269	11,0	3,5
16	153	10,4	4,1
17	308	7,1	5,0
18	1912	2,9	4,9
19	208	7,8	2,7

Решение:

1. Линейная модель множественной регрессии для случая с двумя факторными признаками имеет вид:

$$\hat{y} = a + b_1 x_1 + b_2 x_2 \,.$$

Параметры данной модели найдем, решив систему нормальных уравнений (2.6).

Для расчета коэффициентов регрессии и корреляции составим вспомогательную таблицу 2.6.

Таблица 2.6 - Вспомогательная таблица для расчета коэффициентов регрессии и корреляции

№ предприятия	Себестоимость 1ц, руб.	Урожайность, ц/га	Трудоемкость, чел.-час.	Расчетные величины					
	y	x_1	x_2	y	x_1	x_2	yx_1	yx_2	x_1x_2
1	336	6,1	3,1	112896	37,2	9,6	2049,6	1041,6	18,9
2	213	10,9	4,5	45369	118,8	20,3	2321,7	958,5	49,1
3	138	16,4	2,7	19044	269,0	7,3	2263,2	372,6	44,3
4	624	3,5	5,1	389376	12,3	26,0	2184,0	3182,4	17,9
5	162	12,6	2,0	26244	158,8	4,0	2041,2	324,0	25,2
6	253	6,6	1,9	64009	43,6	3,6	1669,8	480,7	12,5
7	617	2,9	8,3	380689	8,4	68,9	1789,3	5121,1	24,1
8	260	6,2	3,7	67600	38,4	13,7	1612,0	962,0	22,9
9	553	4,7	2,1	305809	22,1	4,4	2599,1	1161,3	9,9
10	270	3,7	2,5	72900	13,7	6,3	999,0	675,0	9,3
11	178	9,3	2,4	31684	86,5	5,8	1655,4	427,2	22,3
12	260	11,1	3,1	67600	123,2	9,6	2886,0	806,0	34,4
13	270	7,2	2,8	72900	51,8	7,8	1944,0	756,0	20,2
14	322	9,3	2,4	103684	86,5	5,8	2994,6	772,8	22,3
15	269	11,0	3,5	72361	121,0	12,3	2959,0	941,5	38,5
16	153	10,4	4,1	23409	108,2	16,8	1591,2	627,3	42,6
17	308	7,1	5,0	94864	50,4	25,0	2186,8	1540,0	35,5
18	1912	2,9	4,9	3655744	8,4	24,0	5544,8	9368,8	14,2
19	208	7,8	2,7	43264	60,8	7,3	1622,4	561,6	21,1
Итог	**7306**	**149,7**	**66,8**	**5649446**	**1419,03**	**278,34**	**42913,10**	**30080,40**	**485,1**

Подставим значения, рассчитанные в таблице 2.6, и получим систему уравнений:

$$\begin{cases} 19 \cdot a + 149{,}7 \cdot b_1 + 66{,}8 \cdot b_2 = 7306 \\ 149{,}7 \cdot a + 1419{,}03 \cdot b_1 + 485{,}1 \cdot b_2 = 42913{,}10 \\ 66{,}8 \cdot a + 485{,}1 \cdot b_1 + 278{,}34 \cdot b_2 = 30080{,}4 \end{cases}$$

Решение данной системы нормальных уравнений:

$a = 615{,}71$; $b_1 = -52{,}30$; $b_2 = 51{,}45$

Тогда уравнение регрессии примет вид:

$$\hat{y} = 615{,}71 - 52{,}30x_1 + 51{,}45x_2$$

Из данного уравнения регрессии следует, что при увеличении урожайности на 1ц/га, себестоимость зерновых снижается на 52,30 руб., а при увеличении трудоемкости на 1 чел.-час. себестоимость увеличивается на 51,45 руб.

2. Парные линейные коэффициенты корреляции рассчитаем по формуле (1.14).

Для расчета коэффициента корреляции между себестоимостью и урожайностью зерновых ($r_{x_1 y}$) необходимо найти соответствующие средние значения и среднеквадратические отклонения:

$$\overline{x_1 y} = \frac{\Sigma x_1 y}{n} = \frac{42913{,}1}{19} = 2258{,}6;$$

$$\bar{x}_1 = \frac{\Sigma x_1}{n} = \frac{149{,}7}{19} = 7{,}88;$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{7306}{19} = 384{,}53;$$

$$\sigma_{x_1} = \sqrt{\frac{\Sigma x_1^2}{n} - \left(\bar{x}_1^2\right)} = \sqrt{\frac{1419{,}03}{19} - \left(7{,}88^2\right)} = 3{,}55;$$

$$\sigma_y = \sqrt{\frac{\Sigma y^2}{n} - \left(\bar{y}^2\right)} = \sqrt{\frac{5649446}{19} - \left(384{,}53^2\right)} = 386{,}62.$$

Подставляя полученные значения в формулу (1.14), получим:

$$r_{x_1 y} = \frac{\overline{x_1 y} - \bar{x}_1 \cdot \bar{y}}{\sigma_{x_1} \cdot \sigma_y} = \frac{2258{,}6 - 7{,}88 \cdot 384{,}53}{3{,}55 \cdot 386{,}62} = -0{,}562.$$

Аналогично рассчитаем коэффициент корреляции между себестоимостью и трудоемкостью производства зерновых ($r_{x_2 y}$), а также коэффициент межфакторной корреляции ($r_{x_1 x_2}$):

$$\overline{x_2 y} = \frac{\Sigma x_2 y}{n} = \frac{30080,4}{19} = 1583,18;$$

$$\bar{x}_2 = \frac{\Sigma x_2}{n} = \frac{66,8}{19} = 3,52;$$

$$\sigma_{x_2} = \sqrt{\frac{\Sigma x_2{}^2}{n} - \left(\bar{x}_2{}^2\right)} = \sqrt{\frac{278,34}{19} - \left(3,52^2\right)} = 1,503;$$

$$\overline{x_1 x_2} = \frac{\Sigma x_1 x_2}{n} = \frac{485,1}{19} = 25,53;$$

$$r_{x_2 y} = \frac{\overline{x_2 y} - \bar{x}_2 \cdot \bar{y}}{\sigma_{x_2} \cdot \sigma_y} = \frac{1583,18 - 3,52 \cdot 384,53}{1,503 \cdot 386,62} = 0,395;$$

$$r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \bar{x}_1 \cdot \bar{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{25,53 - 7,88 \cdot 3,52}{3,55 \cdot 1,503} = -0,414.$$

Таким образом, теснота связи между себестоимостью и урожайностью заметная и обратная, между себестоимостью и трудоемкостью прямая и умеренная, между урожайностью и трудоемкостью обратная и умеренная.

3. Тесноту связи между всеми признаками определим с помощью коэффициента множественной корреляции по формуле (2.14):

$$R_{yx_1x_2} = \sqrt{\frac{(-0,562)^2 + 0,395^2 - 2 \cdot (-0,562) \cdot 0,395 \cdot (-0,414)}{1 - (0,414)^2}} = 0,590.$$

То есть связь между себестоимостью, урожайностью и трудоемкостью производства зерновых является заметной.

4. Определим показатель детерминации(R^2):

$R^2 = (0,590)^2 = 0,348.$

Коэффициент детерминации показывает долю вариации результативного признака под воздействием изучаемых факторов, поэтому в нашей модели 34,8% изменения в себестоимости вызывано влиянием урожайности и трудоемкости, а 65,2% -доля влияния неучтенных факторов.

Таким образом, можно отметить, что построенная модель не адекватна, так как она плохо объясняет поведение результативного признака (себестоимости зерновых) в исследуемой совокупности. Это связано с тем, что выбранные факторные признаки имеют не высокую связь с результатом. Следовательно, в модель включены не основные факторы.

Задача 2.3. По 30 сельскохозяйственным предприятиям имеются данные (таблица 2.7) о средних значениях и вариации урожайности картофеля, количестве внесённых органических удобрений и доли посадок картофеля после лучших предшественников, а также о значениях коэффициентов парной корреляции между этими признаками.

Требуется:

1. Построить уравнение множественной линейной регрессии, описывающее зависимость урожайности картофеля от количества внесённых органических удобрений и доли посадок картофеля по лучшим предшественникам в стандартизованном масштабе и в естественной форме.
2. Определить линейный коэффициент множественной корреляции и коэффициент множественной детерминации. Сделать выводы.
3. Рассчитать общий F – критерий Фишера (при уровне значимости α=0,05). Сделать вывод.

Таблица 2.7

Показатель	Переменная	Среднее значение	Среднее квадратическое отклонение	Линейные коэффициенты парной корреляции
Урожайность картофеля с 1 га, ц	y	167,9	28,773	–
Внесено органических удобрений на 1 га посадки картофеля, т	x_1	29,5	2,769	$r_{yx_1} = 0{,}395$
Доля посадок картофеля по лучшим предшественникам, %	x_2	72,5	14,636	$r_{yx_2} = 0{,}546$ $r_{x_1x_2} = 0{,}046$

Решение.

1.Уравнение множественной линейной регрессии имеет вид:

$$\hat{y} = a + b_1x_1 + b_2x_2 .$$

где $\hat{y}_{x_1x_2}$ – урожайность картофеля с 1 га, ц.;

x_1 – внесено органических удобрений на 1 га посадки картофеля, т.;

x_2 – доля посадок картофеля по лучшим предшественникам, %;

a, $b_{1,}$, b_2 – неизвестные параметры регрессии.

Для оценки параметров $a,\ b_{1},\ b_2$ применим метод стандартизации, т.е. сначала построим уравнение множественной регрессии в стандартизованном масштабе (2.7):

$$t_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2}\ ,$$

где $t_y,\ t_{x_1}, t_{x_2}$ – стандартизованные переменные;

β_1, β_2 – стандартизованные коэффициенты регрессии.

Стандартизованные коэффициенты регрессии вычислим по формулам (2.9)-(2.10). Подставляя значения коэффициентов корреляции из условия задачи, получим:

$$\beta_1 = \frac{0{,}395 - 0{,}546 \cdot 0{,}046}{1 - 0{,}046^2} = 0{,}37;$$

$$\beta_2 = \frac{0{,}546 - 0{,}395 \cdot 0{,}046}{1 - 0{,}046^2} = 0{,}529.$$

Таким образом, получим уравнение множественной линейной регрессии урожайности картофеля от удобрений и предшественников в стандартизованном масштабе:

$$t_y = 0{,}37 \cdot t_{x_1} + 0{,}529 \cdot t_{x_2}\ .$$

Стандартизованные коэффициенты регрессии позволяют сравнить степень влияния каждого независимого фактора на результирующую переменную. В нашем случае получаем, что наиболее существенное влияние на урожайность имеет доля посадок картофеля по лучшим предшественникам $(\beta_2 = 0{,}529)$, а количество внесённых органических удобрений оказывает меньшее воздействие $(\beta_1 = 0{,}37)$.

Для построения уравнения в естественной форме рассчитаем $b_{1},\ b_2$, используя формулу перехода (2.11).

Взяв соответствующие значения средних квадратических отклонений из условия задачи, по формуле (2.11) находим:

$$b_1 = \beta_1 \cdot \frac{\sigma_y}{\sigma_{x_1}} = 0{,}37 \cdot \frac{28{,}773}{2{,}769} = 3{,}845;$$

$$b_2 = \beta_2 \cdot \frac{\sigma_y}{\sigma_{x_2}} = 0{,}529 \cdot \frac{28{,}773}{14{,}636} = 1{,}04.$$

Параметр *a* найдём, используя средние значения признаков (см. условие) и найденные значения $b_{1,}$, b_2 по формуле (2.12):

$$a = 167{,}9 - 3{,}845 \cdot 29{,}5 - 1{,}04 \cdot 72{,}5 = -20{,}928 \, .$$

В итоге получаем уравнение множественной линейной регрессии в естественной форме:

$$\hat{y}_{x_1 x_2} = -20{,}928 + 3{,}845 \cdot x_1 + 1{,}04 \cdot x_2 .$$

Каждый из коэффициентов уравнения регрессии определяет среднее изменение урожайности за счёт изменения соответствующего фактора при фиксированном уровне другого. Так, коэффициент при x_1 показывает, что увеличение (снижение) количества внесения органических удобрений на 1 т/га ведёт к повышению (снижению) урожайности картофеля на 3,845 ц/га. Соответственно коэффициент при x_2 определяет меру зависимости урожайности картофеля от доли высадки его по лучшим предшественникам: рост (снижение) доли посадок по лучшим предшественникам приведёт к росту (снижению) урожайности на 1,04 ц/га.

2. Линейный коэффициент множественной корреляции найдем по формуле 2.15.

В нашем случае применение формулы (2.15) даёт:

$$R_{yx_1x_2} = \sqrt{r_{yx_1} \cdot \beta_1 + r_{yx_2} \cdot \beta_2} = \sqrt{0{,}395 \cdot 0{,}37 + 0{,}546 \cdot 0{,}529} = 0{,}659.$$

Коэффициент множественной корреляции показывает заметную зависимость между выделенными признаками.

Определим коэффициент множественной детерминации:

$$R^2{}_{yx_1x_2} = (R_{yx_1x_2})^2 = 0{,}659^2 = 0{,}434.$$

Коэффициент множественной детерминации свидетельствует, что примерно 43,4% изменения урожайности связано с анализируемыми факторными признаками.

3. Статистическую значимость уравнения регрессии в целом и показателя тесноты связи оценим с помощью общего F – критерия Фишера.

Фактическое (расчётное) значение F – критерия найдём по формуле (2.29). В нашем случае: n = 30, m = 2, R^2 = 0,434.

Тогда получим:

$$F_{факт} = \frac{0{,}434}{1-0{,}434} \cdot \frac{30-2-1}{2} = 10{,}383.$$

Табличное (критическое) значение F – критерия определим по таблице распределения Фишера (Приложение 2) при заданном уровне значимости α = 0,05 и степенях свободы:

$$k_1 = m = 2, \quad k_2 = n - m - 1 = 30 - 2 - 1 = 27;$$

$$F_{кр.} = F_{0{,}05}(2;\ 27) = 3{,}35.$$

Сравним $F_{\text{факт.}}$ и $F_{\text{кр}}$:

$$F_{\text{факт}} = 10{,}383 > F_{\text{кр}} = 3{,}35.$$

Фактическое значение критерия больше табличного, что свидетельствует о статистической значимости построенного уравнения множественной регрессии в целом и показателя тесноты связи $R_{yx_1x_2}$, которые сформировались под неслучайным воздействием факторов x_1 и x_2.

Задача 2.4. По данным о 20 рабочих предприятия (таблица 2.8) оценивается зависимость заработной платы одного рабочего за месяц (Y) от возраста (X_1) и пола (X_2).

Требуется:

1. Построить модель множественной линейной регрессии;
2. Оценить качество построенной модели с помощью множественных коэффициентов корреляции и детерминации;
3. Оценить статистическую значимость уравнения регрессии в целом;
4. Оценить статистическую значимость коэффициентов регрессии;
5. Построить частные уравнения регрессии для двух частей совокупности: мужчин и женщин;
6. Сделать выводы.

Таблица 2.8

№ наблюдения	Заработная плата рабочего за месяц, тыс.руб.	Возраст рабочего, лет	Пол, м/ж
1	6,84	21	м
2	14,76	40	м
3	14,04	41	м
4	9,00	28	ж
5	14,04	39	м
6	14,22	46	м
7	10,80	28	ж
8	14,76	47	м
9	11,34	42	м
10	11,16	35	ж
11	11,52	30	ж
12	12,60	46	м
13	12,24	37	ж
14	13,68	51	м
15	12,60	30	м
16	7,02	25	м
17	7,92	29	ж
18	8,28	23	м
19	9,36	30	ж
20	12,96	38	м

Решение.

1. В данной задаче оценивается зависимость заработной платы рабочего за месяц от количественного фактора x_1 – возраст рабочего и качественного фактора x_2 – пол. В данном случае качественный признак должен быть преобразован в количественный. Для этого введем фиктивную переменную z, которая принимает 2 значения: 1 – если пол рабочего мужской; 0 – если пол женский.

Таким образом, построим модель вида: $\hat{y} = a + b_1 \cdot x_1 + c \cdot z$

Исходные данные для построения модели представлены в таблице 2.9.

Таблица 2.9

№ п/п	y	x_1	z
1	6,84	21	1
2	14,76	40	1
3	14,04	41	1
4	9,00	28	0
5	14,04	39	1
6	14,22	46	1
7	10,80	28	0
8	14,76	47	1
9	11,34	42	1
10	11,16	35	0
11	11,52	30	0

12	12,60	46	1
13	12,24	37	0
14	13,68	51	1
15	12,60	30	1
16	7,02	25	1
17	7,92	29	0
18	8,28	23	1
19	9,36	30	0
20	12,96	38	1

Для оценки параметров модели в качестве программного средства воспользуемся инструментом «Регрессия» программы «Анализ данных» в ППП «Excel». Результаты представлены в таблице 2.10.

Таблица 2.10

ВЫВОД ИТОГОВ				
Регрессионная статистика				
Множественный R	0,843946			
R-квадрат	0,712245			
Нормированный R-квадрат	0,678392			
Стандартная ошибка	1,470815			
Наблюдения	20			
Дисперсионный анализ				
	df	*SS*	*MS*	*F*
Регрессия	2	91,02738	45,51369	21,03905
Остаток	17	36,77604	2,163296	
Итого	19	127,8034		
	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	2,564235	1,414389	1,812963	0,087534
X1	0,24908	0,041954	5,937045	1,62E-05
Z	0,154218	0,743288	0,207481	0,838099

Итак, модель зависимости заработной платы рабочего от возраста и пола имеет вид:

$$\hat{y} = 2{,}564 + 0{,}249 \cdot x_1 + 0{,}154 \cdot z \; .$$

Интерпретируется параметр c = 0,154 при фиктивной переменной следующим образом: зарплата мужчин по сравнению с женщинами при одном и том же возрасте выше в среднем на 0,154 тыс.руб.

2. Коэффициент множественной корреляции R, равный 0,844, свидетельствует о тесной связи между признаками.

Коэффициент множественной детерминации R^2, равный 0,712, показывает, что около 71% вариации результативного признака (заработной платы рабочего) учтено в модели и обусловлено влиянием включенных факторов (возрастом и полом). Соответственно, 29% - влияние других факторов, не учтенных в модели.

3. Проверку значимости уравнения регрессии осуществим на основе – критерия Фишера. Расчетное значение $F_{расч}$ равно 21,04 (таблица 2.10).

Табличное значение критерия при уровне значимости α = 0,05 и числе степеней свободы k_1 = 2 и k_2 = 17 составляет 3,59 (Приложение 2).

Поскольку $F_{расч} > F_{табл}$, уравнение регрессии следует признать статистически значимым, а связь между признаками существенной.

4. Значимость коэффициентов регрессии оценим с помощью t–критерия Стьюдента. Расчетные значения критерия Стьюдента представлены в таблице 2.10: $t_{b_1} = 5{,}94$; $t_z = 0{,}21$.

Табличное значение критерия при уровне значимости α = 0,05 и числе степеней свободы k = 17 составляет 2,11(Приложение 1).

Так как $t_z < t_{табл}$, то коэффициент регрессии c статистически незначим. Следовательно, связь между заработной платой и полом рабочего в данной совокупности является несущественной и данный факторный признак нужно исключить из модели.

Задача 2.5. Имеются данные о производстве продукции растениеводства в сельскохозяйственном предприятии (таблица 2.11).

Требуется:

1. Построить модель множественной регрессии, описывающую зависимость объема выпущенной продукции от объема используемых основных фондов и объема используемых трудовых ресурсов, на основе производственной функции.

2. Оценить адекватность модели с помощью индекса множественной детерминации. Скорректировать индекс детерминации на потерю степеней свободы.

3. Рассчитать общий F – критерий Фишера (при уровне значимости α=0,05). Сделать вывод.

Таблица 2.11

Год	Стоимость валовой продукции, тыс. руб., Y	Среднегодовая стоимость ОПФ, тыс. руб. , K	Среднегодовая численность работников, чел., L
2001	1608	9317	12
2002	2124	10290	13
2003	3312	11980	18
2004	2640	13570	20
2005	3281	13405	22
2006	1997	10740	16
2007	1014	8927	11

Решение.

1. Построим модель на основе производственной функции, которая имеет вид:

$$Y = A \cdot K^{\alpha} \cdot L^{\beta} \cdot \varepsilon \; ,$$

где Y – объем выпущенной продукции;

K – объем используемых основных фондов;

L – объем используемых трудовых ресурсов (затраты живого труда);

A, α, β – параметры модели.

Для приведения данной функции к линейному виду необходимо ее прологарифмировать. В результате чего получим выражение:

$$\lg Y = \lg A + \alpha \lg K + \beta \lg L \; ,$$

которое сводится к линейному заменами:

$$y = \lg Y \; , \quad a = \lg A \; , \quad k = \lg K \; , \quad l = \lg L \; .$$

Таким образом, получаем линейное уравнение регрессии:

$$y = a + \alpha \cdot k + \beta \cdot l \; .$$

Промежуточные расчеты для определения параметров данного уравнения представлены в таблице 2.12.

Таблица 2.12

Год	***Y***	***K***	***L***	***y = lgY***	***k = lgK***	***l = lgL***
2001	1608	9317	12	3,206286	3,969276	1,079181
2002	2124	10290	13	3,327155	4,012415	1,113943
2003	3312	11980	18	3,520090	4,078457	1,255273
2004	2640	13570	20	3,421604	4,132580	1,301030
2005	3281	13405	22	3,516006	4,127267	1,342423
2006	1997	10740	16	3,300378	4,031004	1,204120
2007	1014	8927	11	3,006038	3,950706	1,041393

Параметры полученного линейного уравнения регрессии определяются обычным МНК.

В результате получаем следующее уравнение:

$$y = -1{,}0637 + 0{,}8207k + 0{,}9014l \; ,$$

или

$$lgY = -1{,}0637 + 0{,}8207lgK + 0{,}9014lgL\ .$$

Представленное уравнение приведем к естественной форме путем потенцирования обеих частей уравнения:

$$\hat{Y} = 10^{-1{,}0637} \cdot K^{0{,}8207} \cdot L^{0{,}9014}\ .$$

Таким образом, модель объема производства продукции растениеводства имеет вид:

$$\hat{Y} = 0{,}0863 \cdot K^{0{,}8207} \cdot L^{0{,}9014}\ .$$

Коэффициенты регрессии α и β в построенной модели представляют собой коэффициенты эластичности результата Y от K и L, на основании которых можно сделать вывод, что с ростом стоимости основных средств, используемых в растениеводстве, на 1% производство продукции растениеводства увеличивается на 0,8207%. Влияние среднегодовой численности работников, занятых в растениеводстве, на производство продукции в данной отрасли несколько сильнее, то есть при увеличении данного фактора на 1% результат увеличивается на 0,9014%.

2. Для оценки адекватности построенной модели фактическим данным рассчитаем индекс детерминации по формуле (2.16):

Для расчета индекса детерминации используем данные таблицы 2.13.

$$R^2 = 1 - \frac{1100238{,}359}{4355353{,}429} = 0{,}7474\ .$$

В связи с тем, что в данной работе модель строилась на основе достаточно небольшого числа наблюдений необходимо скорректировать коэффициент детерминации на потерю степеней свободы.

Таблица 2.13

Год	***Y***	***K***	***L***	$\hat{Y}$	$(Y-\hat{Y})^2$	$(Y-Y_{ср})^2$
2001	1608	9317	12	1466,796	19938,622	454661,224
2002	2124	10290	13	1710,443	171029,309	25054,367
2003	3312	11980	18	2598,382	509251,134	1060311,510
2004	2640	13570	20	3164,955	275577,563	127959,510
2005	3281	13405	22	3414,432	17804,120	997430,224
2006	1997	10740	16	2136,239	19387,393	81387,939
2007	1014	8927	11	1309,381	87250,220	1608548,653
Σ	**15976**				**1100238,359**	**4355353,429**

Индекс детерминации с учетом корректировки на число степеней свободы рассчитывается по формуле (2.17):

$$\overline{R}^2 = 1-(1-0{,}7474)\cdot\frac{7-1}{7-2-1} = 0{,}6211.$$

Таким образом. 62,11% вариации стоимости валовой продукции растениеводства объясняются вариацией стоимости основных фондов и среднегодовой численности работников данной отрасли, а 37,89% - объясняются прочими, не входящими в модель причинами.

В данном случае связь можно считать тесной, а модель достаточно пригодной, так как скорректированный индекс детерминации превышает 0,5, т.е. объясняется более половины вариации результативного признака.

3. Статистическую значимость полученной модели оценим с помощью общего F – критерия Фишера.

Фактическое (расчётное) значение F – критерия найдём по формуле (2.29):

$$F_{факт} = \frac{0{,}6211}{1-0{,}6211}\cdot\frac{7-2-1}{2} = 3{,}278.$$

Табличное (критическое) значение F – критерия определим по таблице распределения Фишера (Приложение 2): $F_{кр.} = F_{0,05}(2;4) = 6{,}94$.

Фактическое значение критерия меньше табличного ($F_{факт}=3{,}278 < F_{кр} =6{,}94$), следовательно, построенное уравнение множественной регрессии является статистически незначимым и не может быть использовано на практике. Это связано с тем, что данная модель строилась на основе достаточно небольшого числа наблюдений.

2.3 Вопросы для самопроверки и задания для самостоятельной работы

I. Вопросы для самопроверки

1. Основная цель множественной регрессии. Области применения множественной регрессии в экономике.

2. Требования, которым должны отвечать факторы, включаемые во множественную регрессию.

3. Понятие мультиколлинеарности, её признаки и методы устранения.

4. Множественная линейная регрессия в естественной форме и в стандартизованном масштабе.

5. Оценка параметров уравнения множественной линейной регрессии.

6. Применение фиктивных (структурных) переменных во множественной регрессии.

7. Индекс множественной корреляции: вычисление и интерпретация.

8. Коэффициент множественной детерминации: вычисление и интерпретация.

9. Частные уравнения регрессии. Частные коэффициенты (индексы) корреляции: вычисление и интерпретация.
10. Оценка статистической значимости коэффициентов регрессии с использованием t- критерия Стьюдента.
11. Оценка статистической значимости уравнения множественной регрессии в целом с использованием F-критерия Фишера.
12. Оценка статистической значимости присутствия факторов в уравнении множественной регрессии с помощью частных F-критериев Фишера.
13. Проверка качества подбора уравнения множественной регрессии через среднюю ошибку аппроксимации.

II. Задания для самостоятельной работы

1. Имеются данные 5 наблюдений (таблица 2.14).

Таблица 2.14

№пп	y	x_1	x_2
1	1,3	3,2	3,2
2	1,4	5,7	5,5
3	1,6	7,8	8,0
4	1,2	2,0	2,5
5	1,5	7,0	7,2

Требуется:
1) построить линейную модель множественной регрессии;
2) оценить дисперсию ошибок;
3) оценить адекватность модели с помощью коэффициента множественной детерминации.

2. Имеются данные об урожайности (y, ц/га), количестве внесенных органических удобрений (x_1, т/га), и энергообеспеченности (x_2, л.с.) производства зерновых по 10 сельскохозяйственным предприятиям (таблица 2.15).

Таблица 2.15

Предприятия	Урожайность, ц/га	Внесено органических удобрений на 1 га посева, т	Энергообеспеченность на 100 га пашни, л.с.
1	12,4	0,12	282
2	13,5	0,24	380
3	11,3	0,22	320
4	12,3	0,15	310
5	13,2	0,13	400
6	10,5	0,18	270
7	9,4	0,21	265
8	9,2	0,17	249
9	8,3	0,26	230
10	7,8	0,27	253

Требуется: 1) рассчитать параметры линейного уравнения множественной регрессии; сделать вывод; 2) оценить силу связи факторов с результатом с помощью средних коэффициентов эластичности; 3) оценить статистическую значимость параметров и уравнения регрессии в целом с помощью соответственно параметров Стьюдента и Фишера ($\alpha = 0{,}01$); 4) рассчитать среднюю ошибку аппроксимации; сделать вывод; 5) составить матрицы парных и частных коэффициентов корреляции и указать информативные факторы.

3. По данным о прибыли (Y), затратах на 1 руб. произведенной продукции (X_1) и стоимости основных фондов (X_2) (таблица 2.16) построена модель:

$$Y = 4308{,}53 - 41{,}58X_1 - 26{,}95X_2 .$$

Требуется:

1) оценить качество построенной модели с помощью множественных коэффициентов корреляции и детерминации. Сделать вывод;
2) оценить статистическую значимость уравнения регрессии в целом;
3) оценить статистическую значимость присутствия факторов X_1 и X_2 в уравнении множественной регрессии.

Таблица 2.16

Y	X_1	X_2
1070	76	5,9
1000	76	5,9
700	80	4,7
750	83	4,0
602	90	3,5
225	94	4,0

4. По 50 сельскохозяйственным предприятиям имеются данные (таблица 2.17) о средних значениях и вариации себестоимости, урожайности и затратах на амортизацию в расчете на 1га посадок картофеля, а также о значениях коэффициентов парной корреляции между этими признаками.

Таблица 2.17

Показатель	Переменная	Среднее значение	Среднее квадратическое отклонение	Линейные коэффициенты парной корреляции
Себестоимость 1ц, руб.	y	365,9	47,2	–
Урожайность, ц/га	x_1	101,6	9,9	$r_{yx_1} = -0{,}44$
Затраты на амортизацию в расчете на 1га, тыс.руб.	x_2	7,27	2,70	$r_{yx_2} = 0{,}24$ $r_{x_1x_2} = 0{,}04$

Требуется:

1) построить уравнение множественной регрессии в стандартизованном масштабе и в естественной форме;
2) определить показатели частной и множественной корреляции;
3) определить средние коэффициенты эластичности и сравнить их с β- коэффициентами;
4) рассчитать общий и частные F – критерии Фишера (при уровне значимости α=0,05);
5) сделать выводы.

5. По 30 хозяйствам региона имеются данные об урожайности картофеля и факторах, которые могут оказывать на нее влияние (таблица 2.18).

Таблица 2.18

№ хозяйства	Урожайность, ц/га	Норма высева, ц на 1 га	Внесено органических удобрений, т на 1 га	Внесено минеральных удобрений, ц на 1 га	Доля минеральных удобрений в подкормке	Доля посадок картофеля по лучшим предшественникам, %
	y	x_1	x_2	x_3	x_4	x_5
1	174	29,5	24	3,7	21	88
2	187	36,2	24	4,1	25	100
3	190	31,0	29	3,7	22	76
4	183	25,5	22	3,7	24	70
5	130	25,6	12	2,4	25	63
6	208	30,4	34	4,0	29	64
7	165	34,7	22	3,8	21	54
8	209	27,9	29	5,2	33	99
9	120	28,8	10	1,9	23	61
10	141	24,6	15	2,6	21	59
11	131	27,8	14	2,5	23	60
12	194	25,5	25	4,0	27	100
13	190	32,6	23	4,1	31	100
14	182	28,7	26	4,3	30	87
15	130	29,9	11	2,2	17	59
16	146	31,9	17	3,3	19	43
17	148	24,1	18	3,1	19	76
18	181	26,2	24	3,4	22	82
19	136	23,1	13	2,8	20	71
20	143	28,4	14	2,4	21	46

Требуется:
1) построить матрицу парных коэффициентов факторной корреляции и на ее основе выбрать факторы для построения модели множественной регрессии;
2) построить модель множественной линейной регрессии;
3) оценить статистическую значимость коэффициентов регрессии с помощью t –критерия Стьюдента. Выявить несущественные коэффициенты регрессии;

4) построить модель множественной линейной регрессии, включив в нее факторы, имеющие существенную связь с результатом;
5) сделать выводы.

6. Зависимость урожайности зерновых культур (y, ц/га) от числа колесных тракторов на 100 га (x_1, шт) и дозы внесения удобрения на 1 га (x_2, ц) по данным 20 предприятий характеризуется следующим образом:

уравнение регрессии: $\hat{y}_x = 7{,}298 + 0{,}365 \cdot x_1 + 3{,}355 \cdot x_2$

суммы квадратов отклонений:

$$\sum (y - \bar{y})^2 = 74{,}270, \ \sum (y - \hat{y})^2 = 37{,}247.$$

Требуется:
1) оценить качество построенного уравнения регрессии с помощью множественного коэффициента детерминации;
2) провести дисперсионный анализ и оценить статистическую значимость уравнения регрессии по F-критерию Фишера. Результаты оформить в виде таблицы 1.9.

7. На основе исходных данных о 20 условных предприятиях (таблица 2.19) оценивается зависимость объема реализации продукции за месяц (Y) от расходов на рекламу (X_1), цены единицы продукции (X_2) и наличия отдела маркетинга в предприятии. Требуется:
1) построить модель множественной линейной регрессии;
2) оценить качество построенной модели с помощью множественных коэффициентов корреляции и детерминации;
3) оценить статистическую значимость уравнения регрессии в целом;
4) оценить статистическую значимость коэффициентов регрессии с помощью t –критерия Стъюдента и отобрать существенные факторы;
5) построить модель зависимости объема реализации от выбранных факторов и оценить качество полученного уравнения регрессии;
6) построить частные уравнения регрессии для предприятий, в которых есть отдел маркетинга, и для предприятий, в которых отдела маркетинга нет;
7) сделать выводы.

Таблица 2.19

№ предприятия	Объем реализации продукции, млн.руб. (У)	Расходы на рекламу, тыс.руб. (X1)	Цена единицы продукции, руб. (X2)	Наличие отдела маркетинга, 1-есть, 0-нет (X3)
1	1,27	109	140	нет
2	1,34	134	141	есть
3	1,25	106	126	нет
4	1,28	107	149	есть
5	1,43	127	154	есть
6	1,25	125	143	нет

7	1,53	116	155	есть
8	1,57	95	165	есть
9	1,27	145	151	нет
10	1,46	135	154	есть
11	1,28	113	147	нет
12	1,55	109	151	есть
13	1,35	145	144	нет
14	1,49	104	156	есть
15	1,46	132	152	есть
16	1,25	88	141	нет
17	1,29	163	148	есть
18	1,28	139	141	нет
19	1,33	211	139	есть
20	1,51	98	147	есть

8. Имеются данные об объеме выпуска продукции Q, затратах капитала K и труда L в условном предприятии за 10 лет (таблица 2.20).

Таблица 2.20

Год	Объем выпуска продукции (Q), усл.ден.ед.	Затраты капитала (K), усл.ден.ед.	Затраты труда (L), тыс. чел.-час.
1	46,0	1,8	2
2	37,5	2	4
3	130,0	2,3	5
4	154,0	5,6	6
5	146,5	10,4	6
6	70,5	3,8	4
7	108,0	6,5	4
8	74,0	9,8	2
9	225,0	11,4	5
10	88,5	5,6	4

Требуется:

1) построить модель множественной регрессии, описывающую зависимость объема выпущенной продукции от затрат капитала и труда, используя производственную функцию Кобба-Дугласа.
2) оценить адекватность модели с помощью индекса множественной детерминации. Скорректировать индекс детерминации на потерю степеней свободы.
3) Рассчитать общий F – критерий Фишера (при уровне значимости α = 0,05). Сделать вывод.

3. СИСТЕМЫ ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

3.1 Необходимый теоретический минимум

Виды систем эконометрических уравнений

Для изучения сложных экономических явлений в эконометрике, как правило, применяют не отдельные уравнения регрессии, а системы уравнений. Системы уравнений подразделяются на виды:

- системы независимых уравнений;
- системы рекурсивных уравнений;
- системы взаимозависимых (совместных, одновременных) уравнений.

Система независимых уравнений – это система, в которой каждый результативный признак (объясняемая переменная) y_j рассматривается как функция одного и того же набора факторов (объясняющих переменных) x_i:

$$\begin{cases} y_1 = a_{11} \cdot x_1 + a_{12} \cdot x_2 + \ldots + a_{1m} \cdot x_m + \varepsilon_1, \\ y_2 = a_{21} \cdot x_1 + a_{22} \cdot x_2 + \ldots + a_{2m} \cdot x_m + \varepsilon_{2,} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \ldots + a_{nm} \cdot x_m + \varepsilon_n \end{cases} \tag{3.1}$$

Набор факторов в каждом уравнении данной системы может варьировать в зависимости от изучаемого явления.

Система рекурсивных уравнений – это система, в которой результативный признак (объясняемая переменная) y_j одного уравнения системы в каждом последующем уравнении выступает в виде фактора наряду с одной и той же совокупностью объясняющих переменных (факторов) x_i:

$$\begin{cases} y_1 = a_{11} \cdot x_1 + a_{12} \cdot x_2 + \ldots + a_{1m} \cdot x_m + \varepsilon_1, \\ y_2 = b_{21} \cdot y_1 + a_{21} \cdot x_1 + a_{22} \cdot x_2 + \ldots + a_{2m} \cdot x_m + \varepsilon_{2,} \\ y_3 = b_{31} \cdot y_1 + b_{32} \cdot y_2 +_{31} \cdot x_1 + a_{32} \cdot x_2 + \ldots + a_{3m} \cdot x_m + \varepsilon_{3,} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1} \cdot y_1 + b_{n2} \cdot y_2 + \ldots + b_{nn-1} \cdot y_{n-1} + a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \ldots \\ + a_{nm} \cdot x_m + \varepsilon_n \end{cases} \tag{3.2}$$

Система взаимозависимых (совместных, одновременных) уравнений – это система, в которой результативный признак y_j одного уравнения системы входит во все другие уравнения системы в качестве фактора наряду с одной и той же совокупностью объясняющих переменных (факторов) x_i:

$$
\begin{cases}
y_1 = b_{12} \cdot y_2 + b_{13} \cdot y_3 + \ldots + b_{1m} \cdot y_n + a_{11} \cdot x_1 + a_{12} \cdot x_2 + \ldots + \\
+ a_{1m} \cdot x_m + \varepsilon_1, \\
y_2 = b_{21} \cdot y_1 + b_{23} \cdot y_3 + \ldots + b_{2n} \cdot y_n + a_{21} \cdot x_1 + a_{22} \cdot x_2 + \ldots + \\
+ a_{2m} \cdot x_m + \varepsilon_2, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
y_n = b_{n1} \cdot y_1 + b_{n2} \cdot y_2 + \ldots + b_{nn-1} \cdot y_{n-1} + a_{n1} \cdot x_1 + a_{n2} \cdot x_2 + \ldots + \\
+ a_{nm} \cdot x_m + \varepsilon_n
\end{cases} \tag{3.3}
$$

Структурная и приведенная формы модели

Системы совместных уравнений представляют наибольший практический интерес. Такие системы эффективны в эконометрических исследованиях и наиболее широко применяются в экономике. В силу этого под системой эконометрических уравнений часто понимают систему совместных уравнений. При этом, систему (3.3) называют структурной формой модели.

Системы независимых или рекурсивных уравнений решают с помощью МНК. Для решения системы взаимозависимых уравнений обычный МНК сразу применить нельзя. При решении системы взаимозависимых уравнений сначала необходимо структурную форму модели преобразовать в приведенную форму:

$$
\begin{cases}
y_1 = \alpha_{10} + \alpha_{11} \cdot x_1 + \alpha_{12} \cdot x_2 + \ldots + \alpha_{1m} \cdot x_m + \varepsilon_1, \\
y_2 = \alpha_{20} + \alpha_{21} \cdot x_1 + \alpha_{22} \cdot x_2 + \ldots + \alpha_{2m} \cdot x_m + \varepsilon_2, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots . \\
y_n = \alpha_{n0} + \alpha_{n1} \cdot x_1 + \alpha_{n2} \cdot x_2 + \ldots + \alpha_{nm} \cdot x_m + \varepsilon_n .
\end{cases} \tag{3.4}
$$

Далее параметры приведенной формы модели оцениваются с помощью МНК. Затем приведенная форма модели преобразуется в структурную, т.е. на основе коэффициентов приведенной формы оцениваются структурные коэффициенты модели.

Переход от приведенной формы модели к ее структурной форме связан с решением проблемы идентификации.

Идентификация – это установление соответствия между приведенной и структурной формами модели.

Классы структурных моделей с точки зрения идентификации:

- идентифицируемая – модель, в которой все структурные коэффициенты однозначно определяются через приведенные коэффициенты;
- неидентифицируемая – модель, в которой структурные коэффициенты невозможно найти по приведенным коэффициентам;
- сверхидентифицируемая – модель, структурные коэффициенты ко-

торой, выраженные через приведенные коэффициенты, имеют два и более числовых значения.

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

Обозначим через H – число эндогенных переменных в уравнении, а через D – число экзогенных переменных, отсутствующих в уравнении, но присутствующих в системе. Тогда необходимое условие идентификации отдельного уравнения принимает вид:

- уравнение идентифицируемо, если $D + 1 = H$;
- уравнение неидентифицируемо, если $D + 1 < H$;
- уравнение сверхидентифицируемо, если $D + 1 > \text{H}$.

Если необходимое условие выполнено, то далее проверяется достаточное условие идентификации**.**

Достаточное условие идентификации: уравнение идентифицируемо, если определитель матрицы, составленной из коэффициентов при переменных, отсутствующих в исследуемом уравнении, не равен нулю, и ранг этой матрицы не менее числа эндогенных переменных системы без единицы.

Оценивание параметров структурной модели

Для решения идентифицируемых уравнений применяется косвенный метод наименьших квадратов, для решения сверхидентифицируемых – двухшаговый метод наименьших квадратов.

Косвенный МНК состоит в следующем:

1) составляют приведенную форму модели и определяют численные значения параметров для каждого ее уравнения в отдельности с помощью обычного МНК;

2) путем алгебраических преобразований переходят от приведенной формы к уравнениям структурной формы модели, получая тем самым численные оценки структурных параметров.

Двухшаговый МНК заключается в следующем:

1) составляют приведенную форму модели и определяют численные значения параметров каждого ее уравнения в отдельности с помощью обычного МНК;

2) выявляют эндогенные переменные, находящиеся в правой части структурного уравнения и находят расчетные (теоретичесие) значения по полученным на первом этапе соответствующим уравнениям приведенной формы модели;

3) с помощью обычного МНК определяют параметры каждого структурного уравнения в отдельности, используя в качестве исходных данных фактические значения предопределенных переменных и расчетные значения эндогенных переменных, полученных на втором этапе.

3.2 Решение типовых задач

Задача 3.1. Модель имеет вид:

$$Y_1 = a_1 + b_{11}x_1 + b_{13}x_3 + C_{12}Y_2 + \varepsilon_1;$$
$$Y_2 = a_2 + b_{22}x_2 + C_{21}Y_1 + \varepsilon_2;$$
$$Y_3 = a_3 + b_{32}x_2 + b_{33}x_3 + \varepsilon_3.$$

Требуется:

1. Используя необходимое и достаточное условие идентификации, определить, идентифицируемо ли каждое уравнение модели.
2. Определить тип модели.
3. Определить метод оценки параметров модели.

Решение:

1. Модель имеет три эндогенные (Y_1, Y_2, Y_3) и три экзогенные переменные (x_1, x_2, x_3).

Проверим каждое уравнение на идентифицируемость:

1-е уравнение:

D = 1 (x_2), H = 2 (Y_1, Y_2),

1 + 1 = 2 – уравнение идентифицируемо;

2-е уравнение:

D = 2 (x_1, x_3), H = 2 (Y_1, Y_2),

2 + 1 > 2 – уравнение сверхидентифицируемо;

3-е уравнение:

D = 1 (x_1), H = 1 (Y_3),

1 + 1 > 1 – уравнение сверхидентифицируемо.

Таким образом, необходимое условие идентификации выполняется только в первом уравнении, второе и третье являются сверхидентифицируемыми.

Проверим достаточное условие идентификации для первого уравнения:

В 1-м уравнении нет переменных x_2, Y_3. Из коэффициентов при них в других уравнениях системы строим матрицу:

Уравнение	Отсутствующие переменные	
	x_2	Y_3
2-е	b_{22}	0
3-е	b_{32}	-1

$$\det M = \det\begin{pmatrix} b_{22} & 0 \\ b_{32} & -1 \end{pmatrix} = \begin{vmatrix} b_{22} & 0 \\ b_{32} & -1 \end{vmatrix} \neq 0, \ \text{rang } M = 2.$$

Следовательно, достаточное условие идентификации выполняется, т.е. 1-е уравнение является точно идентифицируемым.

2. Представленная в задаче система является системой совместных (одновременных) уравнений или структурной формой модели. Так как в данной системе только одно уравнение точно идентифицируемо, а два других сверхидентифицируемы, то система в целом является сверхидентифицируемой.

3. Для оценки параметров сверхидентифицируемой системы используется двухшаговый метод наименьших квадратов (ДМНК).

Так как в системе, представленной в задаче, есть точно идентифицируемое уравнение, то его структурные коэффициенты находятся из системы приведенных уравнений.

Задача 3.2. Структурная форма модели имеет вид:

$$\begin{cases} y_1 = b_{13} \cdot y_3 + a_{11} \cdot x_1 + a_{13} \cdot x_3 + \varepsilon_1, \\ y_2 = b_{21} \cdot y_1 + b_{23} \cdot y_3 + a_{22} \cdot x_2 + \varepsilon_2, \\ y_3 = b_{32} \cdot y_2 + a_{31} \cdot x_1 + a_{33} \cdot x_3 + \varepsilon_3. \end{cases}$$

Для данной модели составлена приведенная форма и определены значения ее параметров:

$$\begin{cases} y_1 = 2x_1 + 4x_2 + 10x_3, \\ y_2 = 3x_1 - 6x_2 + 2x_3, \\ y_3 = -5x_1 + 8x_2 + 5x_3. \end{cases}$$

Требуется:

1. Оценить структурную модель на идентификацию;
2. Исходя из приведенной формы модели, найти структурные коэффициенты модели.

Решение:

1. Исследование модели на идентифицируемость. Модель имеет три эндогенные (y_1, y_2, y_3) и три экзогенные (x_1, x_2, x_3) переменные.

Проверим каждое уравнение системы на необходимое и достаточное условия идентификации.

Первое уравнение:

эндогенных переменных - 2 (y_1, y_3),

отсутствующих экзогенных - 1 (x_2).

Выполняется необходимое условие: 2 = 1+1, следовательно, уравнение идентифицируемо.

В первом уравнении отсутствуют y_2 и x_2.

Построим матрицу из коэффициентов при них в других уравнениях системы:

Уравнение	Отсутствующие переменные	
	y_2	x_2
Второе	– 1	a_{22}
Третье	b_{32}	0

$$Det\ A = -1 \cdot 0 - b_{32} \cdot a_{22} \neq 0.$$

Определитель матрицы не равен 0, ранг матрицы равен 2; следовательно, выполняется достаточное условие идентификации, и первое уравнение точно идентифицируемо.

Второе уравнение:

эндогенных переменных - 3 (y_1, y_2, y_3),

отсутствующих экзогенных - 2 (x_1, x_3).

Выполняется необходимое условие: 3 = 2+1, следовательно, уравнение идентифицируемо.

Во втором уравнении отсутствуют x_1 и x_3. Построим матрицу из коэффициентов при них в других уравнениях системы:

Уравнение	Отсутствующие переменные	
	x_1	x_3
Первое	a_{11}	a_{13}
Третье	a_{31}	a_{33}

$$Det\ A = a_{11} \cdot a_{33} - a_{31} \cdot a_{13} \neq 0.$$

Определитель матрицы не равен 0, ранг матрицы равен 2, следовательно, выполняется достаточное условие идентификации, и второе уравнение точно идентифицируемо.

Аналогично доказывается, что и третье уравнение точно идентифицируемо.

Следовательно, исследуемая система точно идентифицируема и может быть решена косвенным методом наименьших квадратов.

2. Вычислим структурные коэффициенты модели:

1) Из третьего уравнения приведенной формы выразим x_2 (так как его нет в первом уравнении структурной формы):

$$x_2 = \frac{y_3 + 5 \cdot x_1 - 5 \cdot x_3}{8}.$$

Данное выражение содержит переменные y_3, x_1 и x_3, которые нужны для первого уравнения структурной формы модели (СФМ). Подставим полученное выражение x_2 в первое уравнение приведенной формы модели (ПФМ):

$$y_1 = 2 \cdot x_1 + 4 \cdot \frac{y_{3.} + 5 \cdot x_1 - 5 \cdot x_3}{8} + 10 \cdot x_3 \Rightarrow$$

$$y_1 = 0{,}5 \cdot y_3 + 4{,}5 \cdot x_1 + 7{,}5 \cdot x_3$$

2) Во втором уравнении СФМ нет переменных x_1 и x_3. Структурные параметры второго уравнения СФМ можно будет определить в два этапа:

Первый этап: выразим x_1 в данном случае из первого или третьего уравнения ПФМ. Например, из первого уравнения:

$$x_1 = \frac{y_1 - 4 \cdot x_2 - 10 \cdot x_3}{2} = 0{,}5 \cdot y_1 - 2 \cdot x_2 - 5 \cdot x_3$$

Подстановка данного выражения во второе уравнение ПФМ не решило бы задачу до конца, так как в выражении присутствует x_3 , которого нет в СФМ.

Выразим x_3 из третьего уравнения ПФМ:

$$x_3 = \frac{y_3 + 5 \cdot x_1 - 8 \cdot x_2}{5}.$$

Подставим его в выражение x_1:

$$x_1 = 0{,}5 \cdot y_1 - 2 \cdot x_2 - 5 \cdot \left(\frac{y_3 + 5 \cdot x_1 - 8 \cdot x_2}{5} \right) = 0{,}5 \cdot y_1 - y_3 + 6 \cdot x_2 - 5 \cdot x_1;$$

$$x_1 = \frac{0{,}5 \cdot y_1 - y_3 + 6 \cdot x_2}{6}.$$

Второй этап: аналогично, чтобы выразить x_3 через искомые y_1, y_3 и x_2, заменим в выражении x_3 значение x_1 на полученное из первого уравнения ПФМ:

$$x_3 = \frac{y_{3.} + 5 \cdot (0{,}5 \cdot y_1 - 2 \cdot x_2 - 5 \cdot x_3) - 8 \cdot x_2}{5} =$$

$$= 0{,}2 \cdot y_3 + 0{,}5 \cdot y_1 - 3{,}6 \cdot x_2 - 5 \cdot x_3.$$

Следовательно,

$$x_3 = 0{,}033 \cdot y_3 + 0{,}083 \cdot y_1 - 0{,}6 \cdot x_2.$$

Подставим полученные x_1 и x_3 во второе уравнение ПФМ:

$$y_2 = 3 \cdot \frac{0{,}5 \cdot y_1 - y_3 + 6 \cdot x_2}{6} - 6 \cdot x_2 + 2 \cdot \left(0{,}033 \cdot y_3 + 0{,}083 \cdot y_1 - 0{,}6 \cdot x_2\right) \Rightarrow$$

$$y_2 = 0{,}416 \cdot y_1 - 0{,}434 \cdot y_3 - 4{,}2 \cdot x_2 \text{ – второе уравнение СФМ;}$$

3) Из второго уравнения ПФМ выразим x_2, так как его нет в третьем уравнении СФМ:

$$x_2 = \frac{-y_2 + 3 \cdot x_1 + 2 \cdot x_3}{6} = -0{,}167 \cdot y_2 + 0{,}5 \cdot x_1 + 0{,}333 \cdot x_3.$$

Подставим полученное выражение в третье уравнение ПФМ:

$$y_3 = -5 \cdot x_1 + 8 \cdot \left(-0{,}167 \cdot y_2 + 0{,}5 \cdot x_1 + 0{,}333 \cdot x_3\right) + 5 \cdot x_3 \Rightarrow$$

$$y_3 = -1{,}336 \cdot y_2 - x_1 + 7{,}664 \cdot x_3 \text{ – третье уравнение СФМ.}$$

Таким образом, структурная форма модели имеет вид:

$$\begin{cases} y_1 = 0{,}5 \cdot y_3 + 4{,}5 \cdot x_1 + 7{,}5 \cdot x_3, \\ y_2 = 0{,}416 \cdot y_1 - 0{,}434 \cdot y_3 - 4{,}2 \cdot x_2, \\ y_3 = -1{,}336 \cdot y_2 - x_1 + 7{,}664 \cdot x_3. \end{cases}$$

Задача 3.3. Структурная форма модели имеет вид:

$$\begin{cases} y_1 = c_{10} + b_{12} \cdot y_2 + a_{11} \cdot x_1 + \varepsilon_1, \\ y_2 = c_{20} + b_{21} \cdot y_1 + a_{22} \cdot x_2 + a_{23} \cdot x_3 + \varepsilon_2. \end{cases}$$

Условные исходные данные приведены в таблице 3.1.

Таблица 3.1

№ п/п	y_1	y_2	x_1	x_2	x_3
1	5	6	3	4	3
2	4	8	2	2	3
3	3	10	1	2	2
4	3	7	2	5	1
5	7	4	4	6	6
6	8	3	3	8	9
7	9	4	6	9	7
8	11	2	8	10	9
9	10	2	7	8	11
10	11	1	9	9	10
Сумма	**71**	**47**	**45**	**63**	**61**

Требуется:

1. Проверить структурную модель на идентификацию;
2. Оценить параметры структурной модели.

Решение:

1. Проверим каждое уравнение системы на необходимое и достаточное условия идентификации.

Первое уравнение: эндогенных переменных - 2 (y_1, y_2), отсутствующих экзогенных - 2 (x_2, x_3). В данном случае выполняется условие: 2 < 2+1, следовательно, уравнение сверхидентифицируемо.

Второе уравнение: эндогенных переменных - 2 (y_1, y_2), отсутствующих экзогенных - 1 (x_1). Выполняется необходимое условие: 2 = 1+1, следовательно, уравнение идентифицируемо.

Во втором уравнении отсутствует только x_1. Коэффициент при нем в первом уравнении a_{11}.

$$Det\ M = a_{11} \neq 0.$$

Следовательно, выполняется достаточное условие идентификации, и второе уравнение точно идентифицируемо.

2. Для оценки параметров сверхидентифицируемой структурной модели используем двухшаговый МНК:

a) От структурной формы модели переходим к приведенной форме модели:

$$\begin{cases} y_1 = \alpha_{10} + \alpha_{11}x_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \eta_1, \\ y_2 = \alpha_{20} + \alpha_{21}x_1 + a_{22} \cdot x_2 + a_{23} \cdot x_3 + \eta_2. \end{cases}$$

b) Оценку α-коэффициентов осуществим с помощью МНК. Для этого воспользуемся инструментом «Регрессия» программы «Анализ данных» в ППП «Excel».

Получаем уравнения приведенной формы модели:

$$y_1 = 1{,}301 + 0{,}443x_1 + 0{,}223x_2 + 0{,}393 \cdot x_3 + \eta_1;$$

$$y_2 = 10{,}235 - 0{,}182x_1 - 0{,}417x_2 - 0{,}342 \cdot x_3 + \eta_2.$$

c) На основе приведенной формы модели определяем теоретические значения эндогенных переменных y_1 и y_2 (Таблица 3.2).

Таблица 3.2

№ п/п	$\hat{y}_1$	$\hat{y}_2$
1	4,70	6,99
2	3,81	8,01
3	2,98	8,53
4	3,70	7,44
5	6,77	4,95
6	7,96	3,27
7	8,72	2,99
8	10,62	1,53
9	10,51	1,86
10	11,23	1,42

d) С помощью МНК на основе фактических значений экзогенных переменных и теоретических значений эндогенных переменных оценим параметры уравнений структурной формы модели. Для этого снова воспользуемся инструментом «Регрессия» программы «Анализ данных» в ППП «Excel».

Таким образом, получаем несмещенные и состоятельные оценки параметров системы структурных уравнений:

$$\begin{cases} y_1 = 9{,}892 - 0{,}857 y_2 + 0{,}275 x_1 + \varepsilon_1, \\ y_2 = 10{,}769 - 0{,}410 y_1 - 0{,}326 x_2 - 0{,}181 x_3 + \varepsilon_2. \end{cases}$$

3.3 Вопросы для самопроверки и задания для самостоятельной работы

I. Вопросы для самопроверки

1. Виды систем уравнений и их характеристика.
2. Структурная и приведенная формы модели.
3. Идентификация модели. Виды структурных моделей с точки зрения идентифицируемости.
4. Необходимое и достаточное условие идентифицируемости уравнений.
5. В каком случае применяется и что представляет собой косвенный МНК?
6. В каком случае применяется и что представляет собой двухшаговый МНК?

II. Задания для самостоятельной работы

1. Модель спроса и предложения кейнсианского типа имеет вид:

$$\begin{cases} Y_{1t} = c_{10} + b_{13} Y_{3t} + a_{11} X_{1t} + \varepsilon_1; \\ Y_{2t} = c_{20} + b_{23} Y_{3t} + a_{23} Y_{3.t-1} + \varepsilon_2; \\ Y_{1t} = Y_{2t}, \end{cases}$$

где Y_{1t} – спрос на товар в момент времени t;
Y_{2t} – предложение товара в момент t;
Y_{3t} – цена товара в момент t;
$Y_{3.t-1}$ – цена товара в момент $t-1$;
X_{1t} – доход в момент t;
t – текущий период;
(t -1) – предыдущий период.

Требуется:
1) оценить структурную модель на идентификацию;
2) записать приведенную форму модели.

2. Имеется следующая модель денежного рынка:

$$\begin{cases} Y_{1t} = c_{10} + b_{12}Y_{2t} + a_{11}X_{1t} + \varepsilon_1; \\ Y_{2t} = c_{20} + b_{21}Y_{1t} + \varepsilon_2, \end{cases}$$

где Y_{1t} – процентные ставки в период времени t;

Y_{2t} – ВВП в период t;

X_{1t} – денежная масса в период t.

Требуется:

1) оценить структурную модель на идентификацию;
2) ввести во второе уравнение системы дополнительную экзогенную переменную X_{2t} – внутренние инвестиции в период t. Как это изменение повлияет на идентификацию модели?

3. Имеется следующая структурная модель:

$$Y_1 = b_{12}Y_2 + a_{11}X_1 + a_{12}X_2,$$
$$Y_2 = b_{21}Y_1 + b_{23}Y_3 + a_{22}X_2,$$
$$Y_3 = b_{32}Y_2 + a_{31}X_1 + a_{33}X_3.$$

Приведенная форма исходной модели имеет вид:

$$Y_1 = 3X_1 - 6X_2 + 2X_3,$$
$$Y_2 = 2X_1 + 4X_2 + 10X_3,$$
$$Y_3 = -5X_1 + 6X_2 + 5X_3.$$

Требуется:

1) оценить структурную модель на идентификацию;
2) определить структурные коэффициенты модели.

4. Имеется следующая структурная модель:

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

Условные исходные данные приведены в таблице 3.3.

Таблица 3.3

№ п/п	y_1	y_2	x_1	x_2
1	2	4	1	6
2	3	6	3	4
3	4	8	2	2
4	5	7	5	3
5	7	2	4	10
6	8	7	3	8
7	9	4	6	9
8	10	2	8	7

Требуется:
1) оценить структурную модель на идентификацию;
2) определить структурные коэффициенты модели.

5. Структурная модель спроса и предложения некоторого товара имеет вид:

$$\begin{cases} Y_{1t} = c_{10} + b_{13}Y_{3t} + a_{11}X_{1t} + a_{12}X_{1.t-1} + \varepsilon_1; \\ Y_{2t} = c_{20} + b_{23}Y_{3t} + a_{23}Y_{3.t-1} + \varepsilon_2; \\ Y_{1t} = Y_{2t} = Y_t, \end{cases}$$

где Y_{1t} – спрос на товар в момент времени t;
Y_{2t} – предложение товара в момент t;
Y_{3t} – цена товара в момент t;
$Y_{3.t-1}$ – цена товара в момент $t-1$;
X_{1t} – доход в момент t;
$X_{1.t-1}$ – доход в момент $t-1$.

Условные исходные данные приведены в таблице 3.4.

Таблица 3.4

t	Y_t	Y_{3t}	$Y_{3.t-1}$	X_{1t}	$X_{1.t-1}$
1	105	14	12	14	13
2	130	12	14	16	14
3	100	13	12	12	16
4	120	15	13	14	12
5	125	14	15	15	14
6	120	15	14	14	15

Требуется:
1) оценить структурную модель на идентификацию;
2) определить структурные коэффициенты модели.

ИТОГОВЫЙ ТЕСТ

1. Парная регрессия – это уравнение вида:

   а) $y = f(x)$; б) $y = f(x_1, x_2)$; в) $y = f(x_1, x_2, \ldots, x_n)$.

2. Регрессией, нелинейной по объясняющим переменным, является:

   а) степенная $y = a \cdot x^b \cdot \varepsilon$;
   б) показательная
   в) равносторонняя гипербола
   г) а,б;
   д) б,в;
   е) все перечисленное.

3. К регрессиям, нелинейным по оцениваемым параметрам, относят:

   а) степенная $y = a \cdot x^b \cdot \varepsilon$;
   б) показательная
   в) равносторонняя гипербола
   г) а,б;
   д) б,в;
   е) все перечисленное.

4. Случайная величина ε включает:

   а) влияние фактора, учтенного в модели;
   б) влияние не учтенных в модели факторов;
   в) влияние наиболее существенных факторов;
   г) влияние всех возможных факторов.

5. Неправильный выбор общего вида модели относится к ошибкам:

   а) измерения;
   б) выборки;
   в) спецификации;
   г) нет верного ответа.

6. При исключении из совокупности единиц с аномальными значениями исследуемых признаков возникают ошибки:

   а) измерения;
   б) выборки;
   в) спецификации;
   г) нет верного ответа.

7. Ошибки спецификации возможно уменьшить:

   а) изменяя форму модели;
   б) увеличивая объем исходных данных;
   в) уменьшая объем выборки;
   г) невозможно уменьшить.

8. Экспериментальный метод выбора вида математической функции основан на:

   а) построении поля корреляции;
   б) изучении теории связи;
   в) сравнении величины остаточной дисперсии, рассчитанной для разных моделей.

9. Аналитический метод выбора вида математической функции основан на:

   а) построении поля корреляции;

б) изучении теории связи;
в) сравнении величины остаточной дисперсии, рассчитанной для разных моделей.

10. Если вариация результативного признака полностью обусловлена влиянием фактора, учтенного в модели, то остаточная дисперсия:
а) равна нулю;
б) не равна нулю;
в) больше нуля;
г) меньше нуля;
д) равна единице.

11. При выборе общего вида модели предпочтение отдается той математической функции, у которой:
а) остаточная дисперсия является наименьшей;
б) остаточная дисперсия является наибольшей;
в) нет верного ответа.

12. МНК позволяет получить такие оценки параметров линейной регрессии, при которых:
а) $\sum(y-\bar{y})^2 \to min$;
б) $\sum(\hat{y}-\bar{y})^2 \to min$;
в) $\sum(y-\hat{y})^2 \to max$;
г) $\sum(y-\hat{y})^2 \to min$.

13. Показателем тесноты связи является:
а) коэффициент вариации;
б) коэффициент корреляции;
в) коэффициент регрессии;
г) нет верного ответа.

14. Значение линейного коэффициента парной корреляции находится в границах:
а) $-\infty \le r_{xy} \le \infty$; б) $0 \le r_{xy} \le 1$;
в) $-1 \le r_{xy} \le 1$; г) $r_{xy} \ge 0$.

15. Близость линейного коэффициента парной корреляции к нулю означает:
а) отсутствие связи между признаками;
б) отсутствие линейной формы связи между признаками;
в) наличие высокой линейной зависимости;
г) наличие обратно пропорциональной зависимости

16. Близость линейного коэффициента парной корреляции к единице означает:
а) отсутствие связи между признаками;
б) отсутствие линейной формы связи между признаками;
в) наличие высокой линейной зависимости;
г) наличие обратно пропорциональной зависимости.

17. Отрицательная величина линейного коэффициента парной означает:
    г) отсутствие связи между признаками;
    д) отсутствие линейной формы связи между признаками;
    е) наличие высокой линейной зависимости;
    ж) наличие обратно пропорциональной зависимости.

18. Долю вариации результативного признака, объясненную регрессией, в общей вариации результативного признака характеризует:
    а) коэффициент ковариации;
    б) коэффициент вариации;
    в) коэффициент детерминации;
    г) нет верного ответа.

19. F-критерий Фишера используется для:
    а) оценки значимости уравнения регрессии в целом;
    б) оценки значимости параметра *a*;
    в) оценки тесноты связи между признаками;
    г) выбора вида математической функции.

20. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим:
    а) t-критерий Стьюдента;
    б) F-критерий Фишера;
    в) коэффициент вариации;
    г) коэффициент корреляции.

21. Для признания уравнения регрессии статистически значимым необходимо, чтобы:
    а) остаточная дисперсия превышала факторную в несколько раз;
    б) факторная дисперсия превышала остаточную в несколько раз;
    в) факторная дисперсия равнялась нулю;
    г) нет верного ответа.

22. Индекс корреляции может быть рассчитан:
    а) для нелинейной регрессии;
    б) для линейной регрессии;
    в) для регрессии любого вида;
    г) нет верного ответа.

23. Величина $\bar{A} = \frac{1}{n}\sum\left|\frac{y - y_x}{y}\right| \cdot 100\%$ является:
    а) ошибкой регрессии;
    б) ошибкой аппроксимации;
    в) дисперсией ошибок;
    г) не имеет смысла.

24. Верно ли утверждение, что если между факторами существует высокая корреляция, то можно определить их изолированное влияние на результативный признак:
    а) верно;
    б) неверно.

25. Уравнения регрессии, которые связывают результативный признак с соответствующими факторами при закреплении других учитываемых во множественной регрессии факторов на среднем уровне, называются:
   а) парными уравнениями регрессии;
   б) частными уравнениями регрессии;
   в) уравнениями множественной регрессии;
   г) нет верного ответа.

26. Показатель множественной корреляции:
   а) оценивает тесноту связи между всеми факторами;
   б) оценивает тесноту совместного влияния факторов на результат;
   в) характеризует вариацию результативного признака;
   г) нет верного ответа.

27. Определитель матрицы парных коэффициентов межфакторной корреляции используется для:
   а) оценки существенности уравнения регрессии;
   б) оценки мультиколлинеарности факторов;
   в) оценки параметров модели;
   г) нет верного ответа.

28. Для того, чтобы ввести в уравнение регрессии качественные факторы используются:
   а) объясняющие переменные; в) лаговые переменные;
   б) объясняемые переменные; г) фиктивные переменные.

29. Для оценки значимости дополнительно включенного в модель фактора используется:
   а) дополнительный F-критерий;
   б) частный F-критерий;
   в) последовательный F-критерий;
   г) t-критерий.

30. Сколько фиктивных переменных вводится в модель для учета признака имеющего два значения:
   а) одна; б) две; в) три; г) четыре.

31. Если коэффициент эластичности составляет -0,53, то это означает, что при увеличении фактора на 1% результат:
   а) уменьшится на 0,53%; в) уменьшится на 0,53 единицы;
   б) уменьшится на 53%; г) увеличится на 0,53%.

32. Коэффициент регрессии b = 10,35 показывает, что:
   а) при увеличении фактора на 1 единицу результат возрастет на 10,35 единиц;

б) при увеличении фактора на 10,35 единиц результат возрастет на 1 единицу;
в) при увеличении фактора на 1 % результат возрастет на 10,35 %;
г) при увеличении фактора на 1 единицу результат снизится на 10,35 единиц.

33. Система уравнений, в которой каждая эндогенная переменная является функцией одной и той же совокупности экзогенных переменных, называется стстемой:
а) одновременных уравнений;
б) рекурсивных уравнений;
в) нормальных уравнений;
г) независимых уравнений.

34. Если структурные коэффициенты модели выражены через приведенные коэффициенты и имеют более одного числового значения, то такая модель:
а) сверхидентифицируемая;
б) неидентифицируемая;
в) идентифицируемая.

35. Определите, для какого уравнения структурной модели

$$\begin{cases} Y_1 = c_{10} + b_{13}Y_3 + \varepsilon_1; \\ Y_2 = c_{20} + a_{21}X_1 + \varepsilon_2; \\ Y_3 = c_{30} + b_{32}Y_2 + a_{32}X_2 + \varepsilon_3. \end{cases}$$

выполняется необходимое условие идентифицируемости:
а) 1-е уравнение;
б) 2-е уравнение;
в) 3-е уравнение.

36. Экзогенные переменные модели характеризуются тем, что они:
а) относятся к предыдущим моментам времени;
б) являются независимыми и определяются вне системы;
в) являются зависимыми и определяются внутри системы.

37. В уравнении структурной модели находятся две эндогенные переменные и отсутствуют три предопределенные переменные, следовательно, данное уравнение:
а) сверхидентифицируемо;
б) неидентифицируемо;
в) идентифицируемо.

ВЫЧИСЛИТЕЛЬНЫЙ ПРАКТИКУМ

Индивидуальное задание №1

По 10 предприятиям оптовой торговли одинаковой площади изучается зависимость цены объекта (*Y*, тыс. долл.) от расстояния до центра города (*X*, км) (смотри таблицу своего варианта).

Требуется:

1. Определить параметры модели парной линейной регрессии зависимости цены объекта (*Y*) от расстояния до центра города (*X*).
2. Оценить качество построенной модели с помощью средней ошибки аппроксимации.
3. Найти средний коэффициент эластичности. Сделать вывод.
4. Оценить тесноту связи с помощью коэффициентов корреляции и детерминации.
5. Оценить значимость коэффициента корреляции через *t*–критерий Стьюдента (при уровне значимости α=0,05).
6. Оценить статистическую надёжность модели в целом через *F*–критерий Фишера (при уровне значимости α=0,05).
7. Рассчитать прогнозное значение результата при *X*=*M* и найти для этого прогнозного значения доверительный интервал. (при α=0,05).
8. Отобразить на графике исходные данные и результаты моделирования.

Вариант 1
M = 8

Цена объекта, тыс. долл.	71	65	56	61	49	35	33	30	25	23
Расстояние до центра города, км	5	7	9	11	14	15	15	18	20	21

Вариант 2
M = 7

Цена объекта, тыс. долл.	72	65	57	60	48	35	33	30	26	25
Расстояние до центра города, км	5	8	9	11	14	15	15	18	21	22

Вариант 3
M = 13

Цена объекта, тыс. долл.	71	64	56	59	47	35	34	30	25	23
Расстояние до центра города, км	5	8	10	11	14	15	15	18	19	20

Вариант 4
M = 16

Цена объекта, тыс. долл.	70	63	57	53	48	35	33	30	26	25
Расстояние до центра города, км	5	8	9	12	14	15	17	18	21	22

Вариант 5, *M = 11*

Цена объекта, тыс. долл.	62	55	47	45	38	35	33	20	16	15
Расстояние до центра города, км	5	8	9	10	14	15	16	18	21	22

Вариант 6
***M* = 17**

Цена объекта, тыс. долл.	61	55	47	45	38	35	33	22	18	15
Расстояние до центра города, км	7	8	9	10	12	14	16	18	19	22

Вариант 7
***M* = 20**

Цена объекта, тыс. долл.	82	65	63	60	58	55	43	40	36	29
Расстояние до центра города, км	3	5	9	11	14	15	15	18	21	27

Вариант 8
***M* = 19**

Цена объекта, тыс. долл.	72	65	57	60	48	35	33	30	26	25
Расстояние до центра города, км	5	8	9	11	14	15	15	18	21	22

Вариант 9
***M* = 9**

Цена объекта, тыс. долл.	72	65	57	60	48	35	33	30	26	25
Расстояние до центра города, км	6	8	13	15	24	25	26	28	31	32

Вариант 10, *M* = 14

Цена объекта, тыс. долл.	80	75	63	60	58	55	53	40	36	33
Расстояние до центра города, км	4	5	8	10	13	15	15	18	21	25

Индивидуальное задание №2

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

Требуется:

1. Построить линейную модель множественной регрессии. Записать стандартизованное уравнение множественной регрессии. Сделать выводы.
2. Рассчитать средние коэффициенты эластичности. Сделать выводы.
3. Найти коэффициенты парной, частной и множественной корреляции. Проанализировать их.
4. Найти скорректированный коэффициент множественной детерминации. Сравнить его с нескорректированным (общим) коэффициентом детерминации.
5. С помощью F-критерия Фишера оценить статистическую надежность уравнения 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.

Вариант 1

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

Вариант 2

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

Вариант 3

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,7	9	11	11	6,3	22
2	7	3,7	11	12	11	6,4	22
3	7	3,9	11	13	11	7,2	23
4	7	4,1	15	14	12	7,5	25
5	8	4,2	17	15	12	7,9	27
6	8	4,9	19	16	13	8,1	30
7	8	5,3	19	17	13	8,4	31
8	9	5,1	20	18	13	8,6	32
9	10	5,6	20	19	14	9,5	35
10	10	6,1	21	20	15	9,5	36

Вариант 4

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

Вариант 5

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

Вариант 6

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

Вариант 7

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

Вариант 8

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7	3,8	9	11	11	7,1	22
2	7	4,1	14	12	11	7,5	23
3	7	4,3	16	13	12	7,8	25
4	7	4,1	17	14	12	7,6	27
5	8	4,6	17	15	12	7,9	29
6	8	4,7	18	16	13	8,1	30
7	9	5,3	20	17	13	8,5	32
8	9	5,5	20	18	14	8,7	32
9	11	6,9	21	19	14	9,6	33
10	10	6,8	21	20	15	9,8	36

Вариант 9

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

Вариант 10

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

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

Основная литература

1. Заяц, О.А. Эконометрика [Электронный ресурс] : учебное пособие / О.А. Заяц. — Электрон. дан. — Волгоград : Волгоградский ГАУ, 2016. — 96 с. — Режим доступа: https://e.lanbook.com/book/76670 . — Загл. с экрана.
2. Литвинова, И.А. Эконометрика [Электронный ресурс] : учебное пособие / И.А. Литвинова. — Электрон. дан. — Кемерово : КемГУ, 2016. — 112 с. — Режим доступа: https://e.lanbook.com/book/99564 . — Загл. с экрана.

Яковлев, В.П. Эконометрика [Электронный ресурс] : учебник / В.П. Яковлев. — Электрон. дан. — Москва : Дашков и К, 2016. — 384 с. — Режим доступа: https://e.lanbook.com/book/70602 . — Загл. с экрана.

Дополнительная литература

1.Дьяков, И.И. Основы эконометрики [Электронный ресурс] : учебное пособие / И.И. Дьяков, И.В. Жуплей. — Электрон. дан. — Уссурийск : Приморская ГСХА, 2013. — 103 с. — Режим доступа: https://e.lanbook.com/book/69558. — Загл. с экрана.

2. Тимофеев, В.С. Эконометрика /В.С. Тимофеев.- М.: Юрайт, 2014.- 328 с.

3. Новиков, А.И. Эконометрика [Электронный ресурс] : учебное пособие / А.И. Новиков. — Электрон. дан. — Москва : Дашков и К, 2017. — 224 с. — Режим доступа: https://e.lanbook.com/book/93399 . — Загл. с экрана.

Интернет-ресурсы

1. http://www.nsu.ru/ef/tsy/ecmr/index.htm
2. http://www.nsu.ru/ef/tsy/ecmr/study.htm
3. http://elar.usu.ru/bitstream/1234.56789/1479/4/1324633_methodinst.pdf

ПРИЛОЖЕНИЕ 1

Критические точки распределения Стьюдента

(t-распределение)

Число степеней свободы k	**Уровень значимости α (односторонняя критическая область)**					
	0,05	**0,025**	**0,01**	**0,005**	**0,001**	**0,0005**
1	6,31	12,71	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,3
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2.90	3,65	3,95
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73.	2,09	2,53	2,85	3.55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3.39	3.65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29

ПРИЛОЖЕНИЕ 2

Критические точки распределения Фишера
(F –распределение)

Уровень значимости α=0,01												
k_2 \ k_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3.55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Уровень значимости α=0,05												
k_2 \ k_1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19;30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6.26	6,16	6,09	6,04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3.88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

СОДЕРЖАНИЕ

Жуплей Ирина Викторовна
Дьяков Иван Иванович

Эконометрика: Учебное пособие для подготовки к практическим занятиям и выполнения самостоятельной работы для обучающихся по направлению подготовки 38.03.01Экономика (уровень бакалавриата) ФГБОУ ВО Приморская ГСХА

ЭЛЕКТРОННОЕ ИЗДАНИЕ

ФГБОУ ВО Приморская ГСХА

Адрес: 692510, г.Уссурийск, пр-т Блюхера, 44