

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Калининградский государственный технический университет»

Ю. Я. Настин

ЭКОНОМЕТРИКА (магистерский курс)

Учебно-методическое пособие по освоению дисциплины и выполнению
контрольной работы для студентов заочной формы обучения, обучающихся в
магистратуре по направлению подготовки 38.04.01 «Экономика»

Калининград
Изд-во ФГБОУ ВО «КГТУ»
2020

330.43(075)

Рецензенты:

д-р техн. наук, профессор, заведующий кафедрой инструментальных методов в экономике и управлении ФГБОУ ВО «КГТУ» **А.М. Карлов**,

д-р экон. наук, профессор, заведующий кафедрой экономики сельского и рыбного хозяйства ФГБОУ ВО «КГТУ» **Т.Е. Степанова**

Настин, Ю. Я. Эконометрика (магистерский курс): учеб.-методич. пособие по освоению дисциплины и выполнению контрольной работы для студ. заочной формы обучения, обучающихся в магистратуре по направлению подгот. 38.04.01 «Экономика» / **Ю. Я. Настин.** – Калининград: Изд-во ФГБОУ ВО «КГТУ», 2020. – 81 с.

В учебно-методическое пособие по освоению дисциплины вошёл практически весь материал, который предусмотрен рабочей программой. В первом разделе излагаются общие сведения о дисциплине, во втором – методические указания по освоению дисциплины и кратко – её содержание, в третьем разделе – методические указания по практическим занятиям. В разделе четвертом в развёрнутой форме даются методические указания по написанию контрольной работы с применением MS Excel.

Рис. 24, табл. 23

Учебно-методическое пособие рассмотрено и одобрено кафедрой инструментальных методов в экономике и управлении Института отраслевой экономики и управления ФГБОУ ВО «КГТУ», протокол № 11 от 30 июня 2020 г.

Учебно-методическое пособие рассмотрено и одобрено методической комиссией Института отраслевой экономики и управления ФГБОУ ВО «КГТУ», протокол № 13 от 03 июля 2020 г.

330.43(075)

© Федеральное государственное
бюджетное образовательное
учреждение высшего образования
«Калининградский государственный
технический университет», 2020 г.
© Настин Ю. Я., 2020 г.

СОДЕРЖАНИЕ

| | |
|--|----|
| ПРЕДИСЛОВИЕ | 4 |
| 1 Общие сведения о дисциплине | 7 |
| 1.1 Цель и задачи дисциплины | 7 |
| 1.2 Требования к результатам освоения дисциплины | 7 |
| 1.3 Общая характеристика и структура дисциплины | 8 |
| 2 Содержание дисциплины и методические указания по её освоению | 9 |
| 2.1 Тема 1. Современная теория и практика эконометрического моделирования..... | 9 |
| 2.2 Тема 2. Построение парных и множественных регрессионных моделей..... | 16 |
| 2.3 Тема 3. Техника построения эконометрических моделей | 28 |
| 2.4 Тема 4. Временные ряды и регрессионные модели..... | 34 |
| 3 Содержание и методические указания по подготовке к практическим занятиям | 38 |
| 4 Задания и методические указания по выполнению контрольной работы..... | 43 |
| 4.1 Общие сведения, выбор варианта и исходных данных..... | 43 |
| 4.2 Пример ручного решения задачи 1. Моделирование магазина на основе линейной троичной регрессии | 45 |
| 4.3 Пример решения задачи 2 в MS Excel. Моделирование магазина на основе линейной троичной регрессии | 51 |
| 4.4 Пример решения задачи 3 в MS Excel. Выделение тренда временного ряда методами скользящей средней..... | 55 |
| 4.5 Пример решения задачи 4 в MS Excel. Выделение тренда временного ряда методом экспоненциального сглаживания | 59 |
| 4.6 Пример решения задачи 5 в MS Excel. Прогнозирование на основе авторегрессий 1-го и 2-го порядков | 61 |
| 5 Вопросы для контрольной (самостоятельной) работы и зачёта | 65 |
| СПИСОК РЕКОМЕНДУЕМЫХ ИСТОЧНИКОВ | 67 |
| ПРИЛОЖЕНИЕ А Исходные данные для контрольных (самостоятельных) работ | 68 |
| ПРИЛОЖЕНИЕ Б Пример страницы СОДЕРЖАНИЕ для контрольных (самостоятельных) работ | 70 |
| ПРИЛОЖЕНИЕ В – Построение графика линейной парной регрессии в Wordе через Excel..... | 71 |
| ПРИЛОЖЕНИЕ Г Значения $t_{\alpha,k}$ - критерия Стьюдента | 72 |
| ПРИЛОЖЕНИЕ Д Значения $\chi^2_{\alpha,k}$ - критерия Пирсона..... | 73 |
| ПРИЛОЖЕНИЕ Е Значения $F_{\alpha,k1,k2}$ - критерия Фишера-Снедекора | 74 |
| ПРИЛОЖЕНИЕ Ж – Краткий словарь терминов | 75 |

ПРЕДИСЛОВИЕ

Настоящее пособие содержит материал, который предусмотрен рабочей программой по «Эконометрике (магистерский курс)» для направления подготовки 38.04.01 «Экономика». Несколько слов о содержании пособия и его структуре.

Первый раздел посвящён общим сведениям о дисциплине, например, он содержит таблицу с темами, с делением по видам занятий (лекции, практические, контрольная (самостоятельная) работа).

Второй раздел посвящён методическим указаниям – рекомендациям по освоению дисциплины, кратко разбирается её содержание.

Третий раздел содержит методические указания по подготовке к практическим занятиям согласно тематическому плану.

В четвёртом разделе приводятся методические указания по выполнению контрольной (самостоятельной) работы с рекомендациями по выбору варианта и исходных данных для решения задач, подробно рассматриваются примеры решения этих задач в среде MS Excel.

В пятом разделе даны по темам вопросы для зачёта, для контрольных (самостоятельных) работ.

В приложениях приведены исходные данные для контрольных (самостоятельных) работ по вариантам, статистические таблицы, словарь терминов, другие материалы.

Особенности дисциплины и её изучения. Впервые с подобной дисциплиной студенты сталкиваются при обучении в бакалавриате. Поэтому магистрантам (логично предположить) мы должны предложить продолжение курса, «2-ю часть» эконометрики. Однако не всё так просто: 1) магистранты из других вузов могли ранее не изучать этот предмет; 2) выпускники бакалавриата по прошествии времени забывают материал (ведь его основа – математическая статистика). Поэтому магистерский курс эконометрики носит двойственный характер, он состоит из первой части - введения в предмет - и второй части - выхода за рамки бакалаврской программы. Именно так построены пособие и рабочая программа дисциплины для магистрантов.

Эконометрика – это наука об измерении и анализе экономических явлений и их взаимосвязей, её зарождение является следствием междисциплинарного подхода к изучению экономики. Эконометрика возникла в результате взаимодействия и объединения трех частей экономической теории, статистических и математических методов. Во 2-й половине XX в. к ним присоединилась вычислительная техника как необходимое условие развития эконометрики.

Задачи эконометрики – построение эконометрических моделей, оценка их параметров, проверка гипотез о свойствах экономических показателей, установление видов их взаимосвязей.

Высокий уровень математизации, применение компьютерных технологий делает эту дисциплину нелёгкой в освоении. Поэтому рекомендуем серьёзно отнестись к ней, в первую очередь это касается осмысленного самостоятельно-

го выполнения контрольной (самостоятельной) работы, большая часть которой основана на применении MS Excel.

Вот какую оценку роли эконометрики дал директор Центрального экономико-математического института Российской академии наук (ЦЭМИ РАН) академик В.Л. Макаров: «Эконометрика сейчас - это та же таблица умножения в экономике. Ею все пользуются. Есть множество разных групп, ... которые используют эконометрику для составления различных экономических прогнозов. Она является одной из самых простых экономических моделей, работающих на практике».

Приведём мысли «великих», собрать которые нам помогла умная книга Б.Я. Брусиловского «Теория систем и система теорий». Эти мысли относятся к методологии науки в целом, но к нашей области знаний – эконометрике, особенно. Некоторые из них в чём то и спорные. Их полезно знать, полезно осмысливать.

1) Из «Правил для руководства ума»: а) задача любого вида сводится к математической задаче; б) математическая задача любого вида сводится к алгебраической задаче... [Рене Декарт, философ и математик, 1596-1650].

2) Не следует умножать сущее без необходимости - принцип «Бритва Оккама» [монах-философ Уильям из Оккама, XIV в.].

3) В каждой науке ровно столько науки, сколько в ней математики [Блез Паскаль, математик и физик, 1623-1662].

4) В обучении примеры важнее правил [Исаак Ньютон, физик и математик, 1643-1727].

5) В природе математики не заложена необходимость заниматься идеями числа и величины [Джордж Буль, математик и логик, 1815-1864].

6) Задача математики состоит в том, чтобы узаконить интуицию, и никакой другой задачи у неё нет [Жак Адамар, математик и механик, 1865-1963].

Наконец, на примерах мировых корифеев-экономистов (малая выборка из неохватной генеральной совокупности) покажем связь экономики с математикой, математической статистикой.

Карл Маркс (1818-1883), великий немецкий экономист, автор «Капитала». Оставил после себя математическую рукопись объёмом около тысячи страниц, в основном по дифференциальному исчислению. Из его письма к Ф. Энгельсу: «... ты знаешь таблицы, в которых цены, учетный процент и т.д. представлены в их движении .. в виде зигзагообразных линий. Я неоднократно пытался — для анализа кризисов — вычислить эти up and downs и думал .. математически вывести из этого главные законы кризисов».

Джон Кейнс (1883-1946), великий английский экономист, лауреат Нобелевской премии, автор «Общей теории занятости, процента и денег». Четыре года работал над «Трактатом по теории вероятностей». Занимаясь проблемами экономики, внёс в неё знания из формальной логики, математики.

Леонид Канторович (1912-1986), известный советский математик и экономист, лауреат Нобелевской премии по экономике, пионер и один из основоположников линейного программирования.

Пол Самуэльсон (1915-2009), известный американский экономист, лауреат Нобелевской премии, автор знаменитого во всём мире учебника «Экономика». Был президентом эконометрического общества, соавтор книги «Линейное программирование и экономическая деятельность».

Более подробное введение в эконометрику, различные определения этой области знаний см. в [1, с. 6-8].

При подготовке настоящего пособия автор использовал источники [1-11], приведенные в конце пособия. Особенно отметим работы [1, 9 и 10], откуда заимствованы фрагменты теоретического материала и некоторых задач.

Автор будет благодарен всем, кто пришлёт свои замечания на адрес издательства университета.

Сокращения

| | |
|-----|---|
| ЗР | - закон распределения; |
| МНК | - метод наименьших квадратов; |
| МО | - математическое ожидание; |
| НЗР | - нормальный закон распределения; |
| СВ | - случайная величина; |
| СКО | - среднее квадратическое (квадратичное) отклонение, стандартное отклонение. |

Обозначения

| | |
|------------------------------------|---|
| $D(X)$ | - дисперсия СВ X ; |
| $M_x(Y)$ | - регрессия, условное математическое ожидание СВ Y как функции от X ; |
| m | - число параметров в уравнении регрессии; |
| n | - длина выборки; |
| p | - число факторных переменных в уравнении регрессии; |
| $\sum x_i$ | - оператор суммирования по индексу i от 1 до n ; |
| X и Y | - случайные величины; если по тексту - вектор, то это вектор-столбец $n \times 1$; |
| X | - матрица плана, размерность $n \times (p+1)$; |
| X' | - транспонированная матрица плана X (штрих - оператор транспонирования) ; |
| $M(X)$ | - оператор: математическое ожидание СВ X , |
| $\beta_0, \beta_1, \dots, \beta_p$ | - истинные неизвестные параметры регрессии; |
| b_0, b_1, \dots, b_p | - оценки истинных значений параметров регрессии; |
| r | - парный коэффициент корреляции; |
| σ и s | - средние квадратические отклонения: истинное и его оценка; |
| $Y=M_x(Y)+\varepsilon$ | - уравнение регрессионной модели (в строгом смысле). |

Синонимы (по убыванию предпочтения)

- X, x - фактор; переменная факторная, объясняющая, входная; регрессор;
Y, y - переменная результирующая, объясняемая, зависимая, выходная;
 ε , e - СВ - отклонение, ошибка, возмущение, остаток.

1 Общие сведения о дисциплине

1.1 Цель и задачи дисциплины

Дисциплина «Эконометрика (магистерский курс)» в учебном плане входит в часть, формируемую участниками образовательных отношений.

Целью освоения дисциплины является развитие способности к подготовке аналитических материалов для количественной оценки мероприятий в области экономической политики и принятия экономических решений на микро- и макроуровнях.

Задачи изучения дисциплины:

- освоение современных эконометрических методов и моделей процессов, относящихся к сфере профессиональной деятельности экономиста-аналитика;
- приобретение навыков построения эконометрических моделей процессов, в том числе с применением компьютерной техники, в частности, программы MS EXCEL;
- формирование базовых знаний, умений и навыков для успешного (в т.ч. самостоятельного) применения эконометрических моделей процессов на микро- и макроуровнях.

1.2 Требования к результатам освоения дисциплины

Результатом освоения дисциплины «Эконометрика (магистерский курс)» должны быть сформированные у обучающегося следующие способности:

- использовать методы эконометрики для анализа и обработки данных, необходимых для решения профессиональных задач;
- выбрать и использовать методы эконометрики для обработки экономических данных, анализа и обоснования полученных результатов;
- применять методы математического моделирования, позволяющие получить количественное выражение закономерностей экономического развития в профессиональной деятельности.

Другими словами, в результате освоения дисциплины студент должен:

- знать: базовые понятия эконометрического моделирования, относящиеся к сфере профессиональной деятельности в области экономических процессов;
- уметь: разрабатывать эконометрические модели исследуемых процессов, относящихся к области экономических процессов;

- владеть: навыками применения эконометрических моделей в области экономических процессов.

1.3 Общая характеристика и структура дисциплины

Трудоемкость дисциплины составляет 2 зачетных единицы (ЗЕ) - 72 академических часа. Сюда входит контактная (лекционные и практические занятия) и самостоятельная учебная работа студента.

Распределение трудоемкости освоения дисциплины по темам и видам учебной работы студента приведено в таблице 1.1.

Форма промежуточной аттестации по дисциплине:

- очная форма, первый семестр – зачёт (в основном в форме защиты выполненной расчётно-аналитической самостоятельной работы с применением MS Excel);
- заочная форма, первый семестр – зачёт (в основном в форме защиты выполненной контрольной работы, носящей расчётно-аналитический характер, с применением MS Excel);

Таблица 1.1 – Трудоемкость освоения дисциплины в заочной форме обучения, структура дисциплины

| Номер и наименование темы, вид учебной работы | Объем учебной работы, час | | | | |
|---|---------------------------|-----|----|-------|-------|
| | Контактн. работа | | | СРС | Всего |
| | Лекц | ЛЗ* | ПЗ | | |
| Семестр – 1, трудоемкость – 2 ЗЕТ (72 час.) | | | | | |
| 1) Современная теория и практика эконометрического моделирования | 1 | - | 1 | 10 | 12 |
| 2) Построение и исследование парных и множественных регрессионных моделей | 1 | - | 2 | 14 | 17 |
| 3) Техника построения эконометрических моделей | 1 | - | 2 | 16 | 19 |
| 4) Временные ряды и динамические регрессионные модели | 1 | - | 1 | 15.25 | 17.25 |
| Учебные занятия | 4 | - | 6 | 55,25 | 65.25 |
| РЭ+КА+ контроль (зачёт)) | 2+1.15+3.6 | | | | 6.75 |
| Итого по дисциплине | | | | | 72 |

*ЛЗ – лабораторные занятия, ПЗ – практические занятия, СРС – самостоятельная работа студентов

2 Содержание дисциплины и методические указания по её освоению

2.1 Тема 1. Современная теория и практика эконометрического моделирования

Содержание темы

- 1) Предмет и метод эконометрики, её место в системе экономических наук.
- 2) Примеры эконометрических моделей: функции спроса в моделях потребления, функции прогнозирования вероятности банкротства предприятия.
- 3) Математические и программные инструменты эконометрического моделирования экономических процессов.
- 4) Специфика эконометрических данных.

Методические указания по освоению темы

Цель освоения темы - получить представление: 1) о характере математических моделей, применяемых в эконометрике; 2) о математических методах, используемых при моделировании; 3) о специфике решаемых экономических задач.

2.1.1 Предмет и метод эконометрики, её место в системе экономических наук

Цель освоения этого материала – получить представление о понятиях, вынесенных в заголовок. Следует иметь в виду, что рассматриваемый здесь материал носит дискуссионный характер.

Вопросы, на которые нужно стараться отвечать по мере работы с материалом.

- 1) Что вы можете добавить к сказанному ниже про общую тенденцию XX века - математизацию знаний?
- 2) Что является предметом и методом эконометрики?
- 3) Как вам представляется симбиоз эконометрики и искусственного интеллекта?

Рекомендуемые источники: [1, с. 6-8; 6; 11].

История эконометрики началась только в начале XX века. *Важно понять:*

появление эконометрики – это отражение общей тенденции XX века - математизации знания, в том числе гуманитарного и в том числе - экономического.

Яркий организационный пример математизации экономики: в 1963 г. в Москве был открыт Центральный экономико-математический институт - ЦЭМИ АН СССР, который успешно работает и поныне.

Предмет эконометрики – массовые экономические явления, в том числе в менеджменте и финансах.

Метод эконометрики – это симбиоз методов из различных дисциплин: экономики, статистики, математики, информатики.

Эконометрика становится частью экономической теории наряду с микро- и макроэкономикой, мировой экономикой.

Будущее эконометрики, несомненно, в первую очередь будет определяться развитием систем искусственного интеллекта.

2.1.2 Два примера эконометрических моделей

Цель освоения этого материала - получить представление об экономических явлениях, которые и составляют предмет эконометрики. Это можно уяснить на приводимых ниже примерах.

Вопросы, на которые нужно стараться отвечать по мере работы с материалом.

- 1) В чём нестандартность эконометрической модели в первом примере?
- 2) В чём нестандартность эконометрической модели во втором примере?

Рекомендуемые источники: [1, с. 9-17; 4, глава 1; 6; 11].

Пример 1. Модель потребления домохозяйства как функция его дохода (Л. Торнквист). *Пример интересен тем, что эконометрическая модель здесь состоит не из одного алгебраического выражения, а из четырех: три функции регрессии и одно уравнение баланса.*

Главная идея здесь - разделение потребляемых товаров на 3 группы, см. рисунок 2.1. Ниже - составные части модели (формулы для функций y_2 и y_3 аналогичны (но не тождественны) формуле (2.1), здесь мы их не приводим):

| | |
|----------------------------|--|
| x | – доход на одного члена домохозяйства, руб/(чел*год), |
| $y_1(x)$ | – объём потребления предметов первой необходимости, руб/(чел*год) – хлеб, колбаса, одежда и т.п., $y_1 = \frac{\beta_0 x}{\beta_1 + x} \quad (2.1)$ |
| $y_2(x)$ | – то же, объём потребления предметов длительного пользования – холодильник, телевизор и т.п., |
| $y_3(x)$ | – объём потребления предметов роскоши – драгоценности, яхта и т.п. |
| $y_1(x)+y_2(x)+y_3(x) = x$ | – уравнение баланса: сумма расходов в левой части равна доходу x в правой части. |

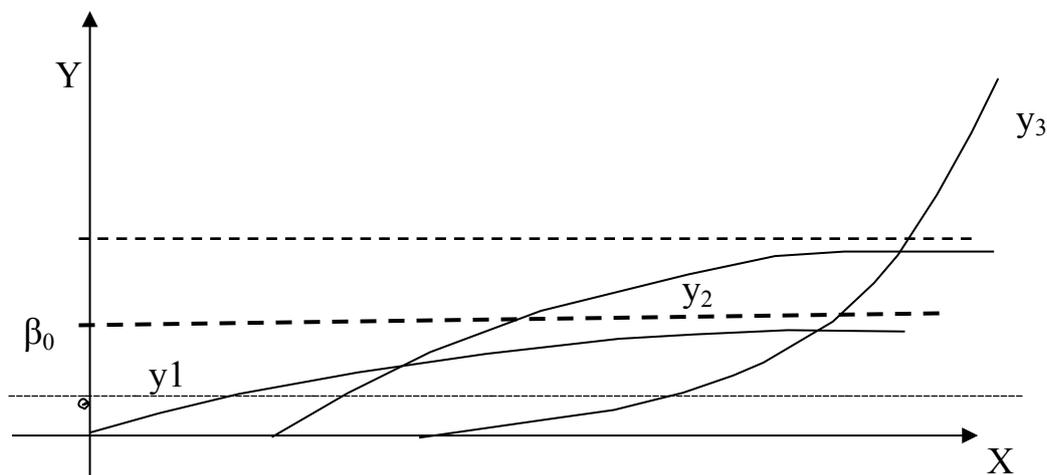


Рисунок 2.1 – Эконометрическая модель зависимости потребления Y от дохода X

Пример 2. Прогнозирование вероятности банкротства предприятия. В 1968 году Э. Альтман (США) предложил пятифакторную модель:

$$Z = 1.2 \cdot X_1 + 1.4 \cdot X_2 + 3.3 \cdot X_3 + 0.6 \cdot X_4 + X_5 \rightarrow \max \quad (2.2)$$

Обращаем внимание: эта модель относится к дискриминантному анализу, хотя с виду - обычное уравнение регрессии (дискриминировать – значит разделять). В чём отличие? В регрессионном анализе результирующая переменная наблюдается и фиксируется непосредственно. В дискриминантном анализе результирующая переменная «конструируется» искусственно на основе результатов наблюдения. Дадим определение.

Дискриминантный анализ — это раздел статистического анализа, который решает задачи распознавания образов для принятия решения о том, какие переменные разделяют (дискриминируют) возникающие наборы данных (группы); в отличие от кластерного анализа здесь группы известны априори (здесь – это 3 группы, см. ниже).

В нашем примере искусственной результирующей переменной является показатель Z (см. таблицу 2.1). Если для анализируемого предприятия

- $Z > 2,9$, оно находится в зоне финансовой устойчивости (в «зеленой» зоне): чем больше Z , тем выше финансовая устойчивость;
- $1,8 < Z < 2,9$, оно находится в зоне неопределенности (в «серой» зоне);
- $Z < 1,8$, оно находится в зоне финансового риска (в «красной» зоне).

Таблица 2.1 – Расчет значений факторных переменных в модели Альтмана

| Фактор | Расчёт значения фактора | | |
|--------|---|--|---|
| | формула | РСБУ* | МСФО |
| X1 | X1 = Оборотный капитал/Активы | (стр.1200*-стр.1500)/ стр.1600 | (Working Capital) / Total Assets |
| X2 | X2 = Нераспределенная прибыль/Активы | стр.2400/ стр.1600 | Retained Earnings / Total Assets |
| X3 | X3 = Операционная прибыль/Активы | стр.2300/ стр.1600 | ЕВIT / Total Assets |
| X4 | X4 = Рыночная стоимость акций / Обязательства | рыночная стоимость акций/ (стр.1400+стр.1500) | Market value of Equity/ Book value of Total Liabilities |
| X5 | X5 = Выручка/Активы | стр.2110 /стр.1600 | Sales/Total Assets |

*Примечания к таблице 2.1: 1) «стр» - строка документа; 2) цифра «1» в начале номера – бухгалтерский баланс; «2» - отчёт о финансовых результатах; 3) РСБУ – российская система бухгалтерского учёта; 4) МСФО – международная система финансовой отчётности.

2.1.3 Математические и программные инструменты эконометрического моделирования экономических процессов

Цель освоения этого материала – получить представление о том инструментарии, который применяется для построения регрессионных моделей, их исследования и применения.

Вопросы, на которые нужно стараться отвечать по мере работы с материалом.

1) Какие задачи исследования функции могут возникать, помимо указанных ниже?

2) Какие два значения имеет термин χ^2 (хи-квадрат)?

3) Что означают два числа в записи стандартного нормального закона распределения: НЗР $N(0, 1)$, каков его графический образ?

4) Как можно назвать значения X1 и X2 на оси абсцисс, см. рисунок 2.2?

Рекомендуемые источники: [1, с. 24-39; 4, глава 2; 6; 8; 11].

Итак, рассмотрим инструменты эконометрического моделирования.

1) **Инструменты для решения задачи исследования функций.** Сюда входит, например: определение области определения, выявление точек экстремума и перегиба (стационарные точки), выявление асимптот. Рассмотрим

функцию $U_1(X)$ в (2.1), где X – доход на одного члена домохозяйства, руб/год*чел, U_1 – расход на приобретение предметов 1-й необходимости, руб/чел*год. Рассмотрим несколько приёмов исследования на примере этой функции, см. рисунок 2.1:

- находим область определения функции (руб): $0 \leq x \leq \infty$;
- исследуем особые точки: если доход $x=0$, то и расход $y_1=0$;
- определяем, как ведёт себя функция при росте дохода ($x \rightarrow \infty$): возникает неопределённость: $y_1 = \infty/\infty$. Применяем правило Лопиталя: дифференцируем отдельно числитель и знаменатель, подставляем значение $X=0$ и получаем уравнение асимптоты $y_1 = \beta_0$ - пунктир на рисунке 2.2. Таким образом, с ростом дохода расходы на предметы первой необходимости практически не растут.

Теперь характер функции $y_1(X)$ ясен: она начинается в начале координат, плавно растёт, приближаясь к горизонтальной асимптоте $y_1 = \beta_0$ – своему «потолку».

2) Набор «статистических» инструментов. Здесь нельзя обойтись без такого базового понятия, как закон распределения случайной величины (СВ). Законы распределения (ЗР) имеют две формы: дифференциальный ЗР (синоним - плотность вероятности) и интегральный ЗР (синоним - функция распределения). Пример дифференциального ЗР см. на рисунке 2.2, где заштрихованная площадь равна вероятности попадания СВ X на интервал $[X_1, X_2]$. В таблице 2.2 даны те законы распределения и типы случайных величин, с которыми нам придётся работать.

Таблица 2.2 – Законы распределения и типы случайных величин

| Закон распределения, СВ | Формульное выражение | Параметры |
|-------------------------------------|---|---|
| Нормальный закон распределения | $\varphi_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$ | a, σ |
| Распределение χ^2 (хи-квадрат) | СВ $\chi^2 = \sum_{i=1}^k Z_i^2$ | СВ Z_i распределены по стандартному НЗР $N(0, 1)$ |
| t- распределение Стьюдента | СВ $t = \frac{Z}{\sqrt{\chi^2/k}}$ | Z - СВ с НЗР $N(0, 1)$ |
| F-распределение Фишера-Снедекора | СВ $F = \frac{\chi^2(k_1)/k_1}{\chi^2(k_2)/k_2}$ | k_1 и k_2 - степени свободы |

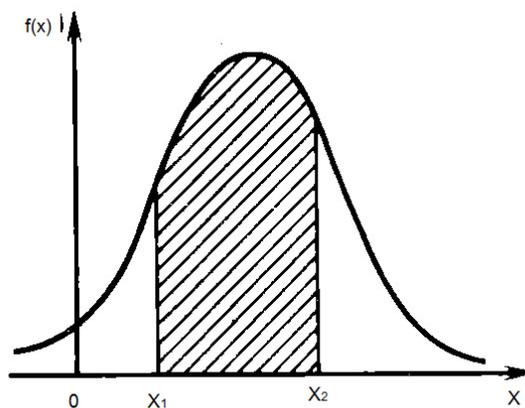


Рисунок 2.2 – Дифференциальный закон распределения СВ X (плотность вероятности)

3) Пакеты прикладных программ - мощный и современный инструмент. Для решения эконометрических задач наибольший интерес представляют пакеты программ: STATISTICA 6.0, SPSS 12.00, Stata 8, STADIA 6.0, Eviews 4.1, S-Plus, StatGrathics, MathCad 12 и другие.

а) Пакеты STATISTICA 6.0, SPSS 12.00 и STADIA 6.0 отличает большое количество методов, начиная от простых дескриптивных (описательных) статистик, заканчивая такими сложными статистико-математическими методами, как нейронные сети и многомерное шкалирование. Основным недостатком данных пакетов программ является низкая гибкость алгоритмов построения эконометрических моделей.

б) Статистико-эконометрические программы (Eviews 4.1, Stata 8) рассчитаны в основном на профессионалов в области эконометрического моделирования. Важной особенностью данных пакетов программ является практически полное отсутствие оконного меню и наличие командной строки в качестве основного рабочего инструмента.

в) Сюда можно отнести программы S-Plus и StatGrathics, которые хотя и обладают довольно большим набором математико-статистических методов, но не пригодны для аналитических изысканий, они скорее всего наилучшим образом подойдут для начинающих исследователей, студентов и аспирантов.

г) Пакеты Mathcad, Maple и другие являются программами, которые предназначены для математических вычислений, рассчитанных на исследователя с обширными математическими знаниями. Данные пакеты имеют возможность реализации статистических методов, но при этом требуют от исследователя знания алгоритмов построения процедур.

д) С ростом типов информации возникает необходимость в так называемых табличных редакторах, таких как MS Excel, Lotus 1-2-3, Quattro Pro, StarOffice Calc и 602Tab, предназначенных для быстрого ввода числовых данных, их редактирования и преобразования в необходимый формат. К достоинствам данных программ можно отнести возможность импорта информации из баз данных, редактирование информации и проведение разведочного анализа. Недостатком, с точки зрения исследователя, можно назвать незначительное ко-

личество статистических алгоритмов и процедур обработки данных, неудобство их применения.

Мы в дальнейшем будем применять MS Excel.

2.1.4 Специфика эконометрических данных

Цель освоения этого материала – получить представление о том, что в некоторых отношениях экономические данные отличаются от технических или астрономических, и эти отличия необходимо учитывать при выборе методов анализа.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом.

1) У компании 7 магазинов, на 1 января их годовая выручка составила ($У_1, У_2, \dots, У_7$), это пространственная выборка?

2) Число постоянных жителей пяти самых крупных городов области по состоянию на 1 января составило (X_1, X_2, \dots, X_5), это временная выборка?

3) Среднемесячная выручка магазина «Тюлька» за последние пять месяцев составила (B_1, B_2, \dots, B_5), это панельные данные?

Рекомендуемые источники: [1, с. 13-17; 8, Орлов, главы 8 и 9].

Выделяют три типа выборочных данных, кратко рассмотрим их.

1) Пространственная выборка (данные). Под ней понимают набор показателей экономических переменных, полученный в данный момент времени. Более точно: пространственная выборка есть серия из n независимых наблюдений r -мерной случайной величины (X_1, \dots, X_r). Два типа таких выборок: моментная и интервальная.

2) Временной (динамический) ряд – это выборка наблюдений с однозначно установленным порядком их следования друг за другом. Например, одна и та же переменная измеряется в разные моменты времени.

3) Панельные данные – это данные, которые сочетают в себе пространственные выборки и временные ряды.

В работе [8, Орлов] рассматривается более сложная проблема: специфика экономических данных по сравнению с данными из других областей, например, их законы распределения. Для анализа экономических данных могут применяться все 4 раздела прикладной статистики, см. таблицу 2.3.

Объекты нечисловой природы – это элементы пространств, в которых нет операций сложения и умножения на число. Примеры таких объектов: значения качественных признаков, бинарные отношения (ранжировки, разбиения, толерантности), последовательности из 0 и 1, множества, нечеткие множества, интервалы, тексты.

Таблица 2.3 – Четыре раздела прикладной статистики

| Название раздела статистики | Математическая природа элементов выборки |
|---|--|
| 1) Статистика случайных величин | Числа |
| 2) Многомерный статистический анализ | Вектора |
| 3) Статистика временных рядов и случайных процессов | Функции |
| 4) Статистика объектов нечисловой природы, в том числе статистика интервальных данных | Объекты нечисловой природы |

Как и для применений статистических методов в иных областях, в эконометрике решаются задачи описания данных (в том числе усреднения), оценивания, проверки гипотез, восстановления зависимостей, классификации объектов и признаков, прогнозирования, принятия статистических решений и др.

2.2 Тема 2. Построение и исследование парных и множественных регрессионных моделей

Содержание темы

- 1) Уравнение парной регрессии на примере. Основные предпосылки регрессионного анализа и метод наименьших квадратов.
- 2) Качество оценок для параметров β_0 , β_1 и σ^2 парной регрессионной модели.
- 3) Интервальная оценка точности прогноза по функции регрессии.
- 4) Классическая нормальная линейная модель множественной регрессии и оценка её параметров.
- 5) Метод наименьших квадратов в матричной форме.
- 6) Оценка ковариационной матрицы и остаточной дисперсии.
- 7) Доверительные интервалы для коэффициентов и функции регрессии.
- 8) Оценка значимости уравнений регрессии (3 показателя).

Методические указания по освоению темы

Цель освоения темы - получить представление о простой регрессионной модели, о предпосылках «правильного» регрессионного анализа и методе наименьших квадратов; о задачах анализа регрессионных моделей. Заметим, что этот материал является ядром классической эконометрики.

2.2.1 Уравнение парной регрессии на примере. Основные предпосылки регрессионного анализа и метод наименьших квадратов

Цель освоения этого материала – научиться строить регрессионные модели, анализировать их свойства, понимать существо метода наименьших квадратов.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом.

1) В чём отличие показателей $\beta_0, \beta_1, \varepsilon_i$ от соответствующих им переменных b_0, b_1, e_i ? Создайте их графические образы на рисунке 2.3.

2) Параметры регрессии b_0, b_1 называют оптимальными (наилучшими), какой смысл вкладывают в это?

3) Если судить по рисунку 2.3, то с какой ситуацией мы имеем дело: гомо- или гетероскедастичности? Создайте на этом рисунке графический образ статистики σ .

4) Создайте на рисунке 2.3 графический образ тех самых квадратов, о которых идёт речь в методе наименьших квадратов.

Рекомендуемые источники: [1, с. 50-62; 2; 6; 11].

Пример постановки эконометрической задачи. Компания располагает семью магазинами «Морепродукты» и планирует создать восьмой магазин площадью 1100 м². Для этого разрабатывается бизнес-план и как его часть – модель магазина. На ней сотрудники хотят исследовать зависимость суточной выручки Y от фактора X – размера площади магазина. Исходные данные известны и представлены в таблице 2.4.

Таблица 2.4 – Исходные данные для модели магазина – парной регрессии

| Номер магазина | i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------------------------|---|---|---|---|---|---|---|----|----|
| Площадь *100 м ² | X | 1 | 1 | 2 | 3 | 4 | 5 | 8 | 11 |
| Выручка *10 ⁵ р/сут | Y | 2 | 3 | 4 | 5 | 5 | 7 | 14 | ? |

Зависимость выручки от площади предполагается искать в форме линейной функции и случайного аддитивного отклонения – «ошибки» (обращаем внимание - используются буквы греческого алфавита):

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (2.3)$$

где - β_0 и β_1 – параметры регрессии, неизвестные нам; «находятся» по генеральной выборке;

- ε_i – отклонение (ошибка) – случайная величина,

- i – номер наблюдения (магазина).

На практике мы используем *выборочную* регрессионную модель (2.4) (обращаем внимание – используются уже буквы латинского алфавита):

$$\begin{aligned} \text{а) } y_i &= b_0 + b_1 x_i + e_i, \quad i=1, 2, \dots, n \\ \text{б) } \hat{y} &= b_0 + b_1 x, \end{aligned} \quad (2.4)$$

где b_0 и b_1 – оценки для неизвестных нам параметров регрессии, находятся по выборке, в которой семь магазинов;
 - e_i – отклонение (ошибка) – значение случайной величины ε_i .

На рисунке 2.3 представлено корреляционное поле (диаграмма рассеяния) и линия регрессии (2.4) для данных из таблицы 2.4.

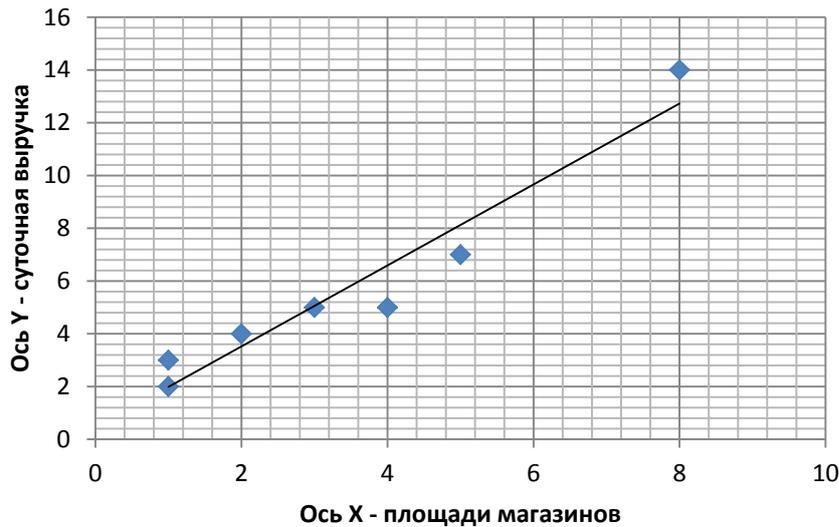


Рисунок 2.3 - Корреляционное поле (диаграмма рассеяния) и линия регрессии

Продолжим разговор о теории. Требуемые свойства модели определяют предпосылки, выполнение которых позволяет проводить исследование модели. Ниже – пять предпосылок для классического регрессионного анализа.

- 1) Отклонение ε_i (и y_i) есть величина случайная, а фактор x_i - неслучайная.
- 2) ε_i – центрированная СВ: $M(\varepsilon_i) = 0$ и, следовательно:

$$M(y_i) = \beta_0 + \beta_1 x_i. \quad (2.5)$$

3) Дисперсия СВ ε_i (и СВ y_i) постоянна, это условие создаёт явление *гомоскедастичности*:

$$D(\varepsilon_i) = \sigma^2 = D(y_i) = \text{const}. \quad (2.6)$$

Невыполнение этого условия приводит к *гетероскедастичности*.

- 4) Отклонения ε_i и ε_j (и переменные y_i и y_j) *некоррелированы*:

$$r = M(\varepsilon_i \cdot \varepsilon_j) = 0 \quad \text{для всех } i, j \quad (i \neq j). \quad (2.7)$$

5) Отклонение ε_i (и переменная y_i) есть нормально распределённая случайная величина (НРСВ).

Модель, для которой выполняются все пять предпосылок, называется *нормальной классической линейной регрессионной моделью* (НКЛРМ).

Для получения качественной регрессии достаточно предпосылок 1-4 (КЛРМ). Предпосылка 5 нужна для оценки точности уравнения и его параметров.

Несколько слов о методе наименьших квадратов (МНК). Его суть *рекомендуем заучить наизусть*:

Суть МНК состоит в отыскании таких значений параметров b_0 и b_1 , которые доставляют минимум сумме квадратов отклонений наблюдаемых значений y_i от теоретических значений \hat{y}_i :

$$S(b_0, b_1) = \sum (\hat{y}_i - y_i)^2 = \sum (b_0 + b_1 x_i - y_i)^2 \rightarrow \min. \quad (2.8)$$

Для отыскания оптимальных b_0 и b_1 приравняем к нулю частные производные S по b_0 и b_1 , продифференцируем функции и получим систему из двух линейных уравнений, которые после преобразований примут вид:

$$\begin{aligned} \text{а) } & b_0 n + b_1 \sum x_i = \sum y_i, \\ \text{б) } & \dots b_0 \sum x_i + b_1 \sum x_i^2 = \sum x_i y_i. \end{aligned} \quad (2.9)$$

Из решения системы уравнений (2.9) получим формулы для расчета неизвестных оптимальных значений параметров регрессии:

$$\begin{aligned} \text{а) } & b_0 = \bar{y} - b_1 \bar{x}, \\ \text{б) } & b_1 = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{C\hat{o}v(X, Y)}{s_x^2}, \end{aligned} \quad (2.10)$$

где s_x^2 - выборочная дисперсия переменной X : $s_x^2 = \sum x_i^2/n - (\bar{x})^2$.

2.2.2 Качество оценок для параметров β_0 , β_1 и σ^2 парной регрессионной модели

Цель освоения этого материала – получить представление о том, каким критериям качества удовлетворяют наши выборочные параметры – оценки.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом.

1) Покажите на рисунке 2.3 отрезки $\hat{y}_2, \hat{y}_3, \hat{y}_6, y_2, y_3, y_6$.

2) Что означают фразы: оценка несмещённая, оценка имеет наименьшую дисперсию?

Рекомендуемые источники: [1, с. 62-63; 2; 6].

Речь пойдёт о параметрах регрессии β_0, β_1 и остаточной дисперсии σ^2 . Оценки b_0, b_1 для β_0, β_1 получают методом наименьших квадратов. Оценку s^2 для остаточной дисперсии (дисперсии отклонений) σ^2 вычисляют по формуле:

$$s^2 = \frac{\sum (\hat{y}_i - y_i)^2}{n - 2} = \frac{\sum e_i^2}{n - 2}, \quad (2.11)$$

где s^2 – выборочная несмещенная оценка остаточной дисперсии, служит показателем качества регрессии;

$e_i = \hat{y}_i - y_i$ – выборочная оценка отклонения ε_i ;

n и m – длина выборки и число параметров (здесь $m=2$).

Вопрос: являются ли оценки b_0, b_1 и s^2 наилучшими? Ответ даёт теорема.

Теорема (Гаусса-Маркова). Если регрессионная модель (2.4) удовлетворяет предпосылкам 1-4 из 2.2.1, то оценки b_0, b_1 несмещённые и имеют наименьшую дисперсию (эффективные). Оценка s^2 несмещённая и состоятельная.

2.2.3 Интервальная оценка точности прогноза по функции регрессии

Цель освоения этого материала – получить представление о том, что такое доверительный интервал и как его построить для регрессии при произвольном значении X . Напомним: до сих пор мы имели дело с точечными оценками.

Вопросы, на которые нужно стараться отвечать при работе с материалом

- 1) Что означают жирно выделенные линии-границы на рисунке 2.4?
- 2) Что означают пунктиром выделенные границы на рисунке 2.4?
- 3) Что можно сказать о доверительном интервале на основе (2.13)?

Рекомендуемые источники: [1, с. 64-67; 2; 6].

Точечный прогноз на основе выборочной регрессии показывает одно из случайных значений результирующей переменной $\hat{y}(x)$. Грамотный экономист-практик скажет: покажите мне тот интервал, за пределы которого с вероятностью, например 0,95, не выйдет значение прогноза \hat{y} . Другими словами, тот доверительный интервал, на котором с вероятностью 0,95 лежит точное неизвестное нам значение регрессии $M_x(Y)$. На рисунке 2.4 такой доверительный интервал (отрезок I_2) обозначен вертикальной стрелкой \leftrightarrow для $X_0=11$.

Напомним: параметры регрессии b_0 и b_1 – случайные величины (они найдены по выборке), а значит и значения \hat{y} для каждого X также СВ. Формула для определения дисперсии СВ \hat{y} как функции от X :

$$s_{\hat{y}}^2(x) = s^2 \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right]. \quad (2.12)$$

Заметим: в (2.12) минимума дисперсия достигает при $X = \bar{x}$ (числитель обращается в 0), доверительный интервал I_1 ; по мере удаления X от среднего значения в обе стороны дисперсия возрастает пропорционально квадрату X (на рисунке 2.4 границы доверительных интервалов выделены жирно).

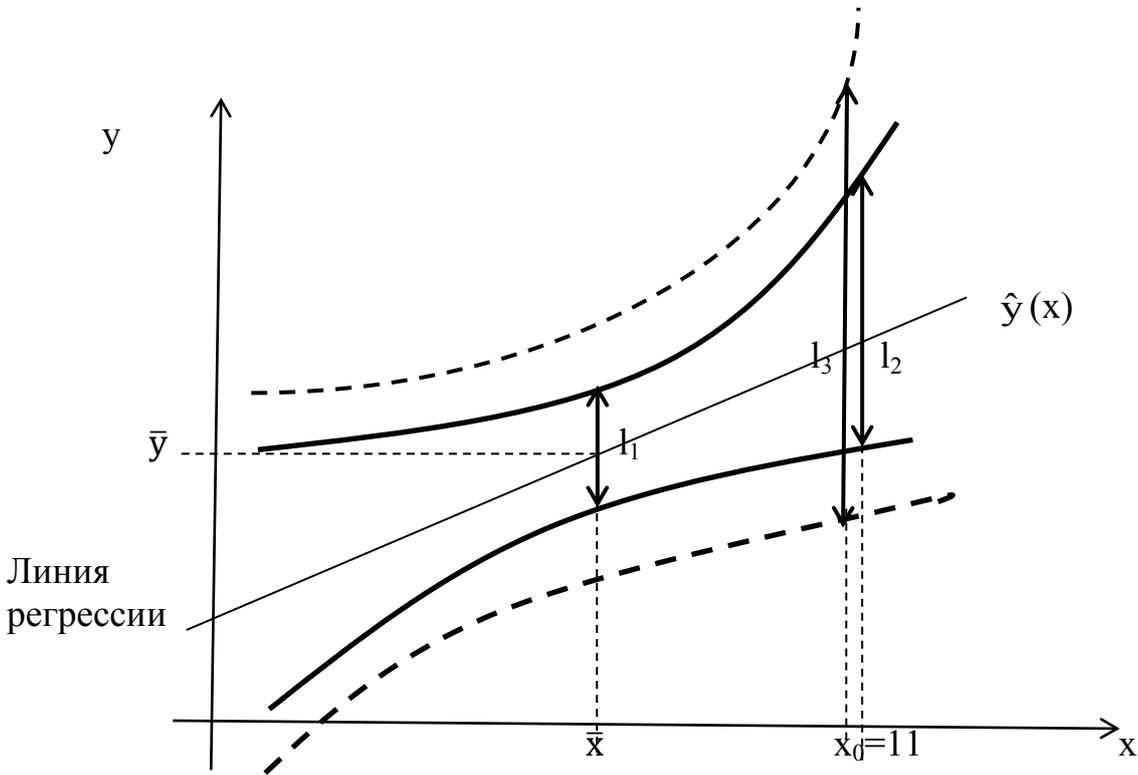


Рисунок 2.4 - Линии границ доверительных интервалов l_1, l_2, l_3 – доверительные интервалы для $M(Y|\bar{x})$, для $M(Y|x_0)$ и для индивидуального значения $y_0(x_0)$

Искомый доверительный интервал ищем в такой общей форме:

$$\hat{y} - \Delta \leq M_x(Y) \leq \hat{y} + \Delta \quad (2.13)$$

Пусть выполняются предпосылки регрессионного анализа 1-5. Тогда статистика $t = (\hat{y} - M_x(Y)) / s_{\hat{y}}$ имеет t -распределение Стьюдента с $k=n-2$ степенями свободы. Искомый доверительный интервал (напомним, $\gamma=1-\alpha$):

$$\hat{y} - t_{1-\alpha/2, k} s_{\hat{y}} \leq M_x(Y) \leq \hat{y} + t_{1-\alpha/2, k} s_{\hat{y}}, \quad (2.14)$$

где значения $t_{1-\alpha, k}$ находятся по таблице из Приложения Г. На рисунке 2.4 доверительный интервал показан в виде вертикального отрезка I_2 для точки $X_0=11$.

2.2.4 Классическая нормальная линейная модель множественной регрессии и оценка её параметров.

Цель освоения этого материала – получить представление о матричной форме записи формул в эконометрике, о шести предпосылках – условиях – линейного регрессионного анализа, которые записаны в матричной форме.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом

- 1) Какие признаки характеризуют функцию (2.15) как линейную?
- 2) Дайте определение каждого термина из названия модели: *классическая, нормальная, линейная, модель, множественная, регрессия.*
- 3) $(y_1, y_2, \dots, y_n)'$ – это матрица-строка или матрица-столбик?
- 4) Что является неизвестным в системе уравнений, записанных в матричной форме: $X'X \cdot b = X'Y$?

Рекомендуемые источники: [1, с. 82-87; 2; 6].

Сложное экономическое явление требует использовать не одну, а несколько факторных переменных. Общий вид такой модели множественной линейной регрессии (ММЛР):

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.15)$$

где ε_i – СВ, отклонения для i -объекта;
 p – число факторных переменных.

Введем обозначения:

$Y=(y_1, y_2, \dots, y_n)'$ – матрица (здесь - вектор) значений зависимой наблюдаемой переменной, размерность $n \times 1$. Штрих - операция транспонирования;

X – матрица значений объясняющих переменных (матрица плана), размерность $n \times (p+1)$. Её первый – единичный - столбец – это значения фиктивной объясняющей переменной x_{i0} :

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}, \quad (2.16)$$

$$\beta = (\beta_0 \ \beta_1 \dots \ \beta_p)',$$

$\varepsilon = (\varepsilon_1 \ \dots \ \varepsilon_n)'$ – матрицы-столбцы параметров регрессии и отклонений.

Шесть предпосылок для уравнения множественной регрессии и теорема Гаусса-Маркова в матричной форме.

- 1) В модели (2.15) ε – случайный вектор, X - неслучайная матрица.
- 2) $M(\varepsilon) = 0_n$, где 0_n , – нулевой вектор-столбец.
- 3) и 4)*. Матрица $\Sigma_\varepsilon = M(\varepsilon\varepsilon') = \sigma^2 E_n$, где E_n - единичная матрица $n \times n$.
- 5) ε - центрированный случайный вектор с n -мерным НЗР $N_n(0, \sigma^2 E_n)$.
- 6) Ранг матрицы $r(X) = p+1$ (столбцы матрицы-плана X должны быть линейно независимыми).

Модель (2.15), которая удовлетворяет предпосылкам 1-6, называется *классической нормальной линейной моделью множественной регрессии (КНЛММР)*. Если из шести предпосылок не выполняется лишь пятая (о НЗР отклонения ε), то модель называется *классической линейной моделью множественной регрессии (КЛММР)*.

Примечание к пп. 3 и 4. Обозначение Σ_ε присвоено матрице размерностью $n \times n$; она имеет всюду нули, кроме главной диагонали, где всюду дисперсия отклонений одинакова и равна σ^2 .

Теорема Гаусса-Маркова. При выполнении предпосылок анализа 1-4 и 6 оценка $b = (X'X)^{-1}X'Y$ является эффективной, т.е. обладает наименьшей дисперсией в классе линейных несмещенных оценок.

2.2.5 Метод наименьших квадратов в матричной форме

Цель освоения этого материала – получить представление о способе записи метода наименьших квадратов в матричной форме.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом

- 1) Чему равно выражение $e' \cdot e$ при $e' = (2 \ 3)$?
- 2) Чему равно выражение $e \cdot e'$ при $e' = (2 \ 3)$?
- 3) Что собой представляет матрица-столбец Y на рисунке 2.5?

Рекомендуемые источники [1, с. 83-87; 2; 6; 11].

Уже знакомый нам критерий оптимальности оценки параметров в методе наименьших квадратов в матричной форме запишется так:

$$S = \sum (\hat{y}_{xi} - y_i)^2 = \sum e_i^2 = e'e = (Y-Xb)'(Y-Xb) \rightarrow \min. \quad (2.17)$$

Как и ранее, для отыскания оптимального значения вектора b составим систему уравнений, приравняем к нулю частные производные от S по b_i , запишем её в матричной форме (читается "набла S по b равно 0-вектору $m \times 1$ "):

$$\nabla_b S = 0_m,$$

где m – число параметров в регрессии ($m=p+1$).

Окончательно система уравнений в матричной форме относительно искомого (неизвестного) вектора b :

$$X'X \cdot b = X'Y. \quad (2.18)$$

Выражение (2.18) представлено на рисунке 2.5 в виде прямоугольников:

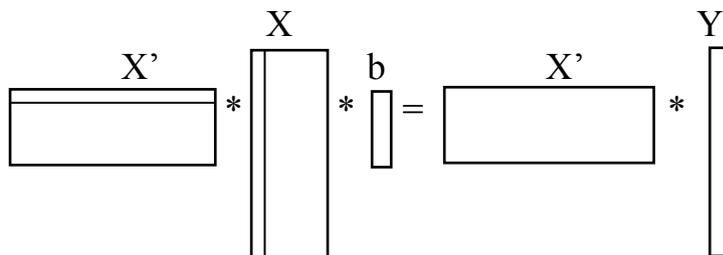


Рисунок 2.5 – Графический образ матричного выражения (2.18)

Разрешим уравнение (2.18) относительно b , для этого умножим *слева* обе его части на обратную матрицу $(X'X)^{-1}$ (при $p=2$ за этим стоит обычное решение системы из трёх линейных уравнений с тремя неизвестными). Получим искомое решение:

$$b = (X'X)^{-1}X'Y. \quad (2.19)$$

При $p=2$ решение (2.19) может выглядеть, например, так:

$$b = \begin{pmatrix} 18 \\ 5 \\ 3 \end{pmatrix}$$

2.2.6 Оценка ковариационной матрицы и остаточной дисперсии

Цель освоения этого материала – получить представление о структуре ковариационной матрицы для случайного вектора параметров регрессии b . Нас будет интересовать в первую очередь «состояние» главной диагонали.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом:

- 1) Какую размерность имеет матрица Σ_b для тройственной регрессии?
- 2) Что означает дословно термин «ковариация»?
- 3) Что за выражение (переменная) σ^2 в (2.20)?

Рекомендуемые источники: [1, с. 91-93; 2; 6; 11].

В множественном регрессионном анализе матричным аналогом дисперсии одной случайной величины является ковариационная матрица Σ_b (сигма бэтая) для случайного вектора b . Введём её:

$$\Sigma_b = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \dots & \sigma_{0p} \\ \sigma_{10} & \sigma_{11} & \dots & \sigma_{1p} \\ \dots & \dots & \dots & \dots \\ \sigma_{p0} & \sigma_{p1} & \dots & \sigma_{pp} \end{pmatrix},$$

где $\sigma_{ij} = M[(b_i - M(b_i)) \cdot (b_j - M(b_j))]$ - ковариации оценок для параметров β_i и β_j .
 На главной диагонали матрицы Σ_b находятся дисперсии оценок параметров регрессии:

$$\sigma_{ii} = M[(b_i - M(b_i)) \cdot (b_i - M(b_i))] = \sigma_{b_i}^2.$$

Путем преобразований можно получить выражение для выборочной оценки ковариационной матрицы:

$$\Sigma_b = \sigma^2 (X'X)^{-1}. \quad (2.20)$$

Обратим внимание: с помощью обратной матрицы $(X'X)^{-1}$ определяется как сам вектор оценок b , так и дисперсии-ковариации его компонент.

Точечная оценка s^2 для остаточной дисперсии σ^2 :

$$s^2 = \frac{e' e}{n - p - 1} = \frac{\sum e_i^2}{n - p - 1}. \quad (2.21)$$

Доверительный интервал для σ^2 – остаточной дисперсии множественной регрессии – со своим числом степеней свободы для χ^2 (хи квадрат):

$$\frac{ns^2}{\chi_{\alpha/2, n-p-1}^2} \leq \sigma^2 \leq \frac{ns^2}{\chi_{1-\alpha/2, n-p-1}^2}. \quad (2.22)$$

Примечание: индексы при χ^2 « $\alpha/2, n-p-1$ » и « $1-\alpha/2, n-p-1$ ».

2.2.7 Доверительные интервалы для коэффициентов и функции регрессии

Цель освоения этого материала – получить представление о точности оценки коэффициентов регрессии. Особое значение имеет оценка точности регрессии (прогноза) через её доверительный интервал.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом

1) При расчёте дисперсии СВ b_1 какой диагональный элемент обратной матрицы нужно выбрать, см. (2.23)?

2) Если нулевая гипотеза $H_0: \beta_j = 0$ отвергается, то какой из этого следует вывод?

3) Как будет выглядеть графический образ доверительного интервала для β_1 на рисунке 2.3?

Рекомендуемые источники: [1, с. 97-101; 2; 6; 11].

Формула для расчёта дисперсии СВ - коэффициента b_j :

$$s_{b_j}^2 = s^2 [(X'X)^{-1}]_{j+1, j+1}, \quad (2.23)$$

где выражение в квадратных скобках есть $(j+1, j+1)$ -й диагональный элемент обратной матрицы.

Примечание. Из выражения (2.23) следует, что, например, для b_1 мы должны взять элемент матрицы с индексами $(2; 2)$, так как параметры b_j имеют индексацию: $0, 1, 2, \dots$, а не $1, 2, 3, \dots$

Конец примечания.

Нулевая гипотеза « $H_0: \beta_j = 0$ » отвергается с уровнем значимости α , если:

$$|t| = |b_j - \beta_j| / s_{bj} > t_{1-\alpha, n-p-1}.$$

Отсюда доверительный интервал для β_j :

$$b_j - t_{1-\alpha, n-p-1} s_{bj} \leq \beta_j \leq b_j + t_{1-\alpha, n-p-1} s_{bj}.$$

Аналогично доверительному интервалу парной регрессии доверительный интервал для условного МО $M_x(Y)$ множественной регрессии:

$$\hat{y} - t_{1-\alpha, k} s_{\hat{y}} \leq M_x(Y) \leq \hat{y} + t_{1-\alpha, k} s_{\hat{y}}, \quad (2.24)$$

где \hat{y} - результирующая переменная, определяемая по уравнению регрессии;
 $s_{\hat{y}} = s \sqrt{(X_0' (X'X)^{-1} X_0)}$ - СКО для результирующей \hat{y} ;
 $X_0' = (1 \ x_{10} \ x_{20} \ \dots \ x_{p0})$ - вектор значений факторных переменных. (2.25)

2.2.8 Оценка значимости уравнений регрессии (3 показателя)

Цель освоения этого материала – получить представление о самом главном – значимости (адекватности) регрессии и о методах её оценки.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом

1) Как определить критическое значение статистики Стьюдента $t_{1-\alpha, n-2}$ в (2.26)?

2) Справедливо ли суждение по (2.27): чем больше Q_R , тем более значима регрессия? И если «да», то какой в этом смысл?

3) Справедливо ли суждение по (2.29): за введение в регрессию каждой дополнительной факторной переменной уплачивается штраф?

Рекомендуемые источники: [1, с.102-105; 2; 6; 11].

1) Оценка *парной* регрессии по критерию Стьюдента. Её значимость можно *равносильно* оценить по коэффициенту b_1 . Нулевая гипотеза $H_0: b_1$ незначимо отличается от 0. Если она верна, то линия регрессии параллельна оси ОХ, фактор Х ничего не объясняет. Обратная гипотеза $H_1: b_1$ значимо различает-

ся от 0; если это так, то регрессия значима. Критерий на основе статистики Стьюдента и гипотезы H_1 :

$$H_0: \beta_1=0 \text{ отвергается, если } t = \left| \frac{b_1 - 0}{s} \cdot \sqrt{\sum (x_i - \bar{x})^2} \right| > t_{1-\alpha, n-2}. \quad (2.26)$$

2) Оценка по критерию Фишера-Снедекора. Гипотеза H_0 о равенстве нулю всех коэффициентов модели ($\beta_1=\beta_2= \dots = \beta_p = 0$) отвергается, если фактическое значение статистики Фишера-Снедекора F больше ее табличного значения:

$$F = \frac{Q_R (n - p - 1)}{Q_e p} > F_{\alpha, p, n-p-1}. \quad (2.27)$$

где $Q_R = \sum (\hat{y}_i - \bar{y})^2 =$ – сумма квадратов отклонений, обусловленная регрессией;
 $= Y'Y - b'X'Y$
 $Q_e = \sum (y_i - \hat{y}_i)^2 =$ – остаточная сумма квадратов отклонений, обусловленная влиянием неучтенных факторов.
 $= b'X'Y - n\bar{y}^2$

3) Оценка по коэффициенту детерминации. Для множественной регрессии коэффициент R^2 можно рассчитать по одной из нескольких формул:

$$R^2 = Q_R / (Q_R + Q_e) = \frac{b'X'Y - n\bar{y}^2}{Y'Y - n\bar{y}^2} \quad (2.28)$$

Скорректированный (адаптированный) коэффициент детерминации \hat{R}^2 :

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1} (1 - R^2) = 1 - \frac{(n-1)e'e}{(n-p-1)y'y}. \quad (2.29)$$

Обращаем внимание на (2.29): чем больше переменных p , тем меньше \hat{R}^2 в сравнении с R^2 . Это заставляет включать в модель только те переменные, которые действительно информативны и влияют на переменную Y .

2.3 Тема 3. Техника построения эконометрических моделей

Содержание темы

- 1) Мультиколлинеарность и отбор наиболее значимых факторов.
- 2) Модели с переменной структурой. Фиктивные переменные.
- 3) Нелинейные модели регрессии, классификация и линейаризация.
- 4) Производственная функция Кобба-Дугласа и функции эластичности.
- 5) Проблема гетероскедастичности выборки.

Методические указания по освоению темы.

Цель освоения темы - получить представление: 1) о практических трудностях, с которыми сталкивается эконометрист при построении и исследовании моделей, например, проблема отбора наиболее значимых факторов; 2) о проблемах применения нелинейных моделей; 3) о явлении гетероскедастичности, её выявлении и устранении.

2.3.1 Мультиколлинеарность и отбор наиболее значимых факторов

Цель освоения этого материала – получить представление о важнейшей с практической точки зрения проблеме: отборе наиболее значимых факторов. Отбор тесно связан с явлением мультиколлинеарности – взаимозависимостью факторов.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом.

- 1) Почему при мультиколлинеарности оценки очень чувствительны к исходным данным?
- 2) Сколько всего регрессий нужно построить, чтобы реализовать алгоритм, показанный на рисунке 2.6?
- 3) Какой будет траектория решения, если в задаче на рисунке 2.6 использовать не скорректированный, а обычный коэффициент детерминации?
- 4) О мультиколлинеарности каких векторов идёт речь, назовите пару таких векторов.

Рекомендуемые источники: [1, с.108-115; 2; 6; 11].

Мультиколлинеарность – это взаимная зависимость факторных переменных X_1, X_2, \dots, X_r . Образ полной мультиколлинеарности: несколько векторов лежат на параллельных прямых. Образ частичной мультиколлинеарности: вектора лежат на неортогональных прямых (в том числе близких к параллельным).

При мультиколлинеарности определитель матрицы $(X' \cdot X)$ очень близок к нулю - произведение матриц $X' \cdot X$ см в (2.18) и (2.19). При обращении $(X' \cdot X)$ определитель $|(X' \cdot X)|$ стоит в знаменателе. Поэтому:

- оценки становятся очень чувствительными к малым изменениям наблюдаемых данных, к объему выборки;

- уравнение регрессии содержательно не интерпретируется, так как коэффициенты могут иметь неверные по смыслу знаки и неоправданно большие значения;
- факторные переменные неинформативны, так как они дублируют друг друга.

Рассмотрим два метода выявления и снижения мультиколлинеарности.

Первый метод основан на анализе корреляционной матрицы R с размерностью $(p+1) \times (p+1)$, матрица состоит из парных коэффициентов корреляции между переменными Y, X_1, \dots, X_p . Признаком мультиколлинеарности здесь — наличие коэффициентов корреляции со значениями, например, от $|0,7|$ и выше. Пусть между X_1 и X_4 коэффициент корреляции $0,8$. Тогда одну из переменных можно исключить из регрессии. Например ту, у которой ниже корреляция с Y .

Второй метод основан на «вращении факторов» - их полном переборе. В качестве критерия для отбора используют скорректированный коэффициент детерминации \hat{R}^2 (см. (2.29)).

Пошаговый алгоритм метода «вращения» факторов, см. рисунок 2.6:

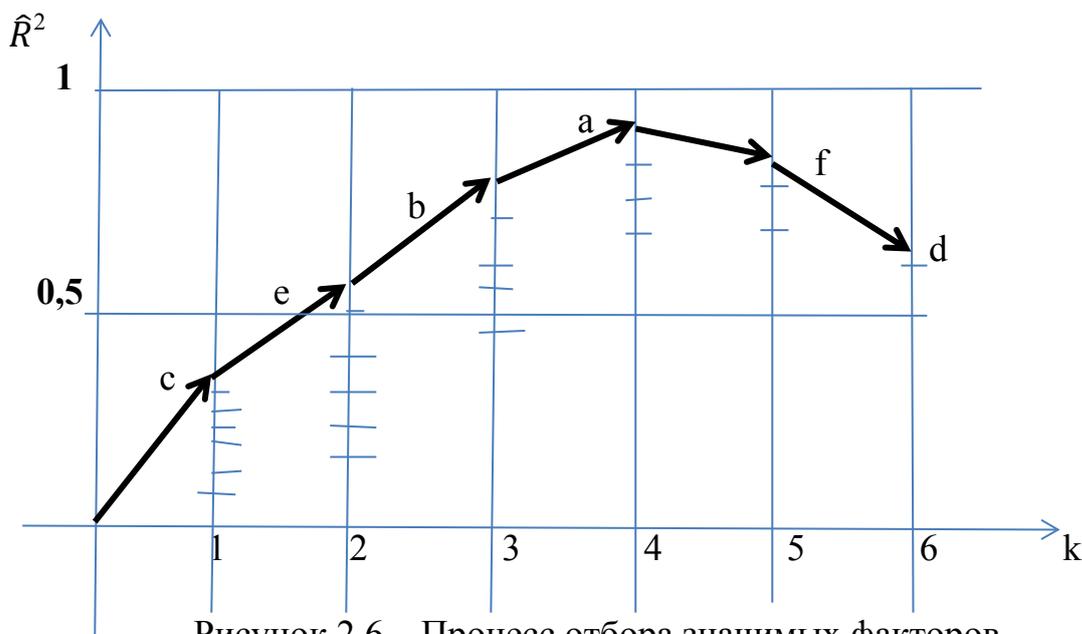


Рисунок 2.6 – Процесс отбора значимых факторов методом «вращения факторов»

- 1) Отбираем, например, шесть объясняющих факторов-претендентов: $X_a, X_b, X_c, X_d, X_e, X_f$.
- 2) Строим шесть парных линейных регрессий с каждым из факторов, для каждой регрессии находим значение \hat{R}^2 .
- 3) Выбираем фактор, который даёт наибольшее значение \hat{R}^2 , например, X_c , он — самый значимый фактор.
- 4) Строим пять регрессий с двумя факторами: $(X_c, X_a), (X_c, X_b)$, и т.д., и для каждой регрессии находим \hat{R}^2 .
- 5) Выбираем тот фактор, который даёт наибольшее значение \hat{R}^2 , например, X_e — он второй по значимости фактор.

- 6) Строим четыре регрессии с факторами: (X_c, X_e, X_a) , (X_c, X_e, X_b) и т.д.
- 7) Процесс останавливается, когда очередной скорректированный коэффициент детерминации будет меньше предыдущего; в нашем примере - после 4-го шага. Итого, в регрессию включаются четыре фактора: X_c, X_e, X_b, X_a . Они обладают минимальной мультиколлинеарностью и дают максимальный скорректированный коэффициент детерминации (≈ 0.8).

2.3.2 Модели с переменной структурой. Фиктивные переменные

Цель освоения этого материала – получить представление о возможности включения в регрессию качественных переменных. Обычно качественная переменная допускает отношение порядка на ее значениях (успеваемость: плохая, удовлетворительная, хорошая).

Вопросы, на которые рекомендуем отвечать в процессе работы с материалом:

- 1) Составляется таблица с графами: ФИО, Стаж, Пол, Зарплата. Как будет выглядеть графа Пол с булевой переменной z ?
- 2) Верно ли утверждение: для атрибутивной переменной с 4-мя уровнями нужно ввести 3 булевских переменных?

Рекомендуемые источники: [1, с. 115-122; 2; 6; 11].

До сих пор мы рассматривали регрессионные модели с факторными *количественными* переменными. Однако нередко возникает необходимость учёта качественных (атрибутивных) факторов, имеющих несколько уровней (градаций). Например, пол (мужской, женский), образование (среднее, среднее специальное, высшее).

Качественные факторы могут существенно влиять на структуру линейных связей между переменными, скачкообразно меняя параметры регрессионной модели. В этом случае говорят о *регрессионных моделях с переменной структурой*.

Атрибутивные переменные включают в уравнение регрессии, наделив их условными (фиктивными) числовыми значениями. Например, для женщин $z_1=0$, для мужчин $z_1=1$ (фиктивная булевская переменная). Для показателя “образование” можно взять одну переменную z с тремя значениями: 1 – среднее, 2 – среднее специальное, 3 - высшее. Однако для анализа удобнее ввести две булевских переменных и использовать их комбинацию, см. таблицу 2.5.

Таблица 2.5 – Включение двух булевских переменных в регрессию

| Переменная | Вид образования | | | Итого |
|------------|-----------------|---------------------|--------|-------|
| | Среднее | Среднее специальное | Высшее | |
| Z_1 | 0 | 0 | 1 | 1 |
| Z_2 | 0 | 1 | 0 | 1 |

2.3.3 Нелинейные модели регрессии, их классификация и линеаризация

Цель освоения этого материала – получить представление о дереве классификации типов регрессий по признакам нелинейности параметров и факторов.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом:

- 1) Что означает термин: функция, линейная по параметрам?
- 2) За счёт чего достигается «внутренняя линейность»?

Рекомендуемые источники: [1, с. 124-128; 2; 6; 11].

В таблице 2.6 представлены три класса нелинейных моделей и примеры.

Класс 1. Эти модели определяются с помощью МНК путем замены переменных. Например, в уравнении $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ заменим переменные: $x = x_1, x^2 = x_2$ и получим двухфакторное уравнение линейной регрессии:
 $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$.

Класс 2. Нелинейные по параметрам.

Подкласс 2.1. Модели этого подкласса после преобразования к линейному виду также решаются с помощью МНК. Например, прологарифмируем степенную функцию, получим: $\ln y = \ln \beta_0 + \beta_1 \ln x + \ln \varepsilon$. Заменим переменные: $\ln y = z, \ln \beta_0 = \beta_0', \ln x = u, \ln \varepsilon = \varepsilon'$. Получим линейное уравнение регрессии:
 $z = \beta_0' + \beta_1 u + \varepsilon'$.

Таблица 2.6 – Классификация нелинейных уравнений регрессии и примеры

| 1 - Нелинейные по переменным, но линейные по параметрам | 2 - Нелинейные по параметрам | |
|---|--|--|
| | 2.1 - и внутренне линейные | 2.2 - и внутренне нелинейные |
| $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ | степенная: $y = \beta_0 x^{\beta_1} \varepsilon$ | степенная: $y = \beta_0 x^{\beta_1} + \varepsilon$ |
| $y = \beta_0 + \beta_1/x + \varepsilon$ | показательная: $y = \beta_0 \beta_1^x \varepsilon$ | гиперболическая $y = \beta_0 (1 - \frac{1}{1 - x^b}) + \varepsilon$ |
| $y = \beta_0 + \beta_1 \ln x + \varepsilon$; | экспоненциальная: $y = e^{(\beta_0 + \beta_1 x)} \varepsilon$. | |

Подкласс 2.2. Здесь МНК может быть применен только в форме специальных поисковых процедур. Например – метод наискорейшего спуска Коши. Причина: система теперь уже нелинейных уравнений $\nabla S = 0$ (читается так: градиент эс равно ноль-вектору) не разрешается аналитически.

2.3.4 Функции эластичности. Производственная функция Кобба-Дугласа

Цель освоения этого материала – получить представление: 1) о таком важном для экономиста понятии, как эластичность, 2) о нелинейной регрессионной модели предприятия в форме функции Кобба-Дугласа; здесь факторные переменные: K – капитал и L – труд.

Вопросы, на которые нужно стараться отвечать по мере работы с материалом.

1) Как выглядит график функции эластичности для линейной парной регрессии (см. таблицу 2.7, следует применить правило Лопиталья)?

2) Можно ли в качестве фактора L – труд - использовать показатель (чел*смен/год)?

Рекомендуемые источники: [1, с.126-128; 2; 6; 11].

Функцией эластичности $E_{x_i}(\hat{y})$ множественной регрессии $\hat{y}=f(x_1, x_2, \dots, x_p)$ по x_i называется функция:

$$E_{x_i}(\hat{y}) = \frac{\partial \hat{y}}{\partial x_i} \cdot \frac{x_i}{\hat{y}}$$

В таблице 2.7 даны два примера функций эластичности. Коэффициент эластичности – это значение функции эластичности для данного значения X .

Таблица 2.7 - Примеры функций эластичности

| Функция регрессии | Производная регрессии | Функция эластичности |
|--|-----------------------|---|
| Линейная $\hat{y} = \beta_0 + \beta_1 x$ | β_1 | $E = \beta_1 x / (\beta_0 + \beta_1 x)$ |
| Гипербола $\hat{y} = \beta_0 + \beta_1 / x$ | $-\beta_1 / x^2$ | $E = -\beta_1 / (\beta_0 x + \beta_1)$ |

Смысл функции эластичности $E(x)$ в том, что она показывает: на сколько процентов изменится \hat{y} при изменении X на 1%.

Производственная функция Кобба-Дугласа имеет вид: $\hat{Y} = AK^\alpha L^\beta$,

где Y – объем производства, руб/год;

K – затраты капитала, руб;

L – затраты труда, руб/год (фонд оплаты труда);

α и β - коэффициенты эластичности Y по K и Y по L : $E_K(Y) = \alpha$, $E_L(Y) = \beta$.

С учетом влияния отклонений функция Кобба-Дугласа может выглядеть как мультипликативная степенная модель: $Y = AK^\alpha L^\beta \varepsilon$. Она нелинейная относительно параметров и её можно свести к линейной путем логарифмирования:

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L + \ln \varepsilon.$$

Введём переменные: $Y1 = \ln Y$, $A1 = \ln A$, $K1 = \ln K$, $L1 = \ln L$, $\varepsilon 1 = \ln \varepsilon$. Теперь нетрудно записать линейное уравнение регрессии:

$$Y1 = A1 + \alpha K1 + \beta L1 + \varepsilon 1.$$

Примечание. Если СВ ε будет входить в функцию как слагаемое, то линейаризация станет невозможной. Тогда для решения системы уравнений с неизвестными $b_0, b_1 \dots$ пришлось бы применять методы вычислительной математики, например, метод покоординатного или градиентного спуска.

2.3.5 Проблема гетероскедастичности выборки

Цель освоения этого материала – получить представление о «неприятном» явлении – гетероскедастичности. Её противоположность – гомоскедастичность – одна из предпосылок классического регрессионного анализа. Поэтому нужно уметь определять наличие гетероскедастичности, применять соответствующий тест-критерий.

Вопросы, на которые рекомендуем отвечать в процессе работы с материалом

- 1) Как будет выглядеть числовой пример для выражения: $\sum \varepsilon = \sigma^2 E_n$?
- 2) Можно ли применять обычный МНК в случае гетероскедастичности?
- 3) В чём отличие парного коэффициента корреляции от рангового?
- 4) В чём суть теста ранговой корреляции Спирмена? Можно ли его применить в случае, показанном на рисунке 2.7.

Рекомендуемые источники [1, с. 155-160; 2; 6; 11].

Явление гомоскедастичности записывается так: $\sum \varepsilon = \sigma^2 E_n$ (это - матрица с размерностью $n \times n$, у которой на главной диагонали всюду одинаковая дисперсия σ^2 , а все остальные клетки – нули.). Пример гомоскедастичности - корреляционное поле на рисунке 2.3. На практике мы чаще имеем дело с гетероскедастичностью. На рисунке 2.7 показано, как с ростом фактора X растёт разброс точек.

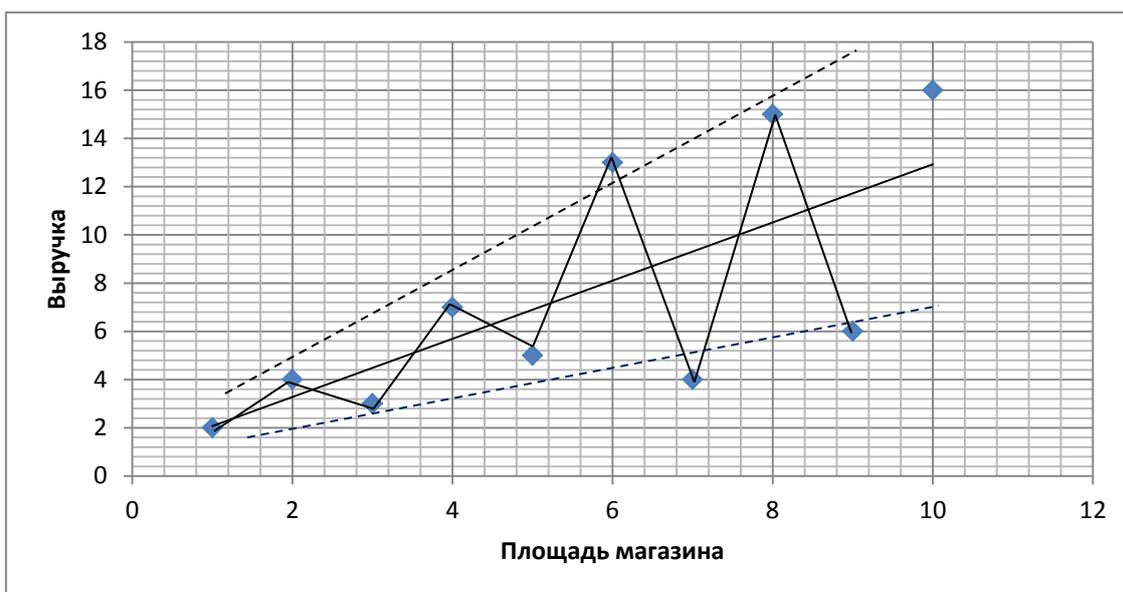


Рисунок 2.7 – Явление гетероскедастичности

Последствия гетероскедастичности. Пусть имеет место гетероскедастичность и мы для оценки регрессии применили обычный МНК, получили вектор $b=(X'X)^{-1}X'Y$. Вместо Y подставим его модель $Y=X\beta+\varepsilon$, получим вектор-оценку:

$$b=(X'X)^{-1}X'Y=\beta+(X'X)^{-1}X'\varepsilon.$$

Эта оценка несмещенная и состоятельная, вывод: обычный МНК применим и в случае гетероскедастичности.

Но если мы хотим исследовать модель (оценить точность, значимость, получить интервальные оценки коэффициентов), то *результаты окажутся непригодными*. И ещё: несмещенная и состоятельная оценка в случае гетероскедастичности не будет оптимальной в смысле теоремы Гаусса-Маркова. Значит, оценка b будет значительно отличаться от истинного значения β .

Тест ранговой корреляции Спирмена на гетероскедастичность. Суть теста: модули отклонений $|e_i|$ являются оценками СКО σ_i , поэтому при гетероскедастичности эти модули и значения факторов x_i будут коррелированы. Эта связь хорошо видна на рисунке 2.7. Воспользуемся ранговым коэффициентом корреляции:

$$\rho_{x,e} = \frac{1 - 6 \sum d_i^2}{(n^3 - n)}, \quad (2.30)$$

где d_i – разность между рангами значений x_i и e_i .

Коэффициент ранговой корреляции значим на уровне α , другими словами, имеет место гетероскедастичность, если выполняется условие:

$$|t| = \frac{|\rho_{x,e}| \sqrt{n-2}}{\sqrt{1-\rho_{x,e}^2}} > t_{1-\alpha, n-2}, \quad (2.31)$$

где $t_{1-\alpha, n-2}$ – табличное значение t-критерия Стьюдента с уровнем значимости α и числом степеней свободы $(n-2)$.

И в заключение: гетероскедастичность можно устранить, но для этого придётся применить *не обычный МНК, а взвешенный*.

2.4 Тема 4. Временные ряды и регрессионные модели

Содержание темы

- 1) Стационарные временные ряды и их характеристики.
- 2) Автокорреляционная функция.
- 3) Прогнозирование на основе моделей временных рядов.
- 4) Авторегрессионные модели и модели скользящей средней.

Методические указания по освоению темы

Цель освоения темы - получить представление: 1) о типах временных рядов (стационарные и нестационарные); 2) о показателях и функции, которые характеризуют ряд; 3) о прогнозировании временных рядов различными методами.

2.4.1 Временные ряды и их характеристики

Цель освоения этого материала – получить представление о числовых характеристиках временного ряда, его структуре.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом:

- 1) Из каких компонент может состоять временной ряд?
- 2) Какой ряд называется стационарным?
- 3) О каких числовых характеристиках ряда идёт речь?

Рекомендуемые источники [1, с. 133-136; 4; 6; 7; 10].

Временным (динамическим) рядом называют упорядоченный набор значений СВ Y в последовательные моменты времени. При исследовании такого ряда выделяют несколько компонент – слагаемых:

$$\bullet \quad y_t = u_t + v_t + c_t + \varepsilon_t, \quad \bullet \quad (2.32)$$

где u_t – тренд, описывает влияние медленно изменяющихся факторов;

v_t – сезонная компонента, отражает колебания с периодом до года;

c_t – циклическая компонента, отражает колебания с периодом более года;

ε_t – случайная компонента, отражает влияние случайных факторов.

Временной ряд y_t ($t=1, 2, \dots, n$) называется стационарным в узком смысле, если совместное распределение вероятностей n наблюдений y_1, y_2, \dots, y_n такое же, как и n наблюдений $y_{1+\tau}, y_{2+\tau}, \dots, y_{n+\tau}$ при любых n, t, τ . Отсюда следует возможность вычисления числовых характеристик ряда: $MO[y(t)] = a$, $СКО[y(t)] = \sigma$, их можно оценить по известным формулам:

$$\bar{y} = \sum y_t / n \quad s_t^2 = \sum (y_t - \bar{y})^2 / n \quad (2.33)$$

Примером стационарного ряда может служить «белый шум», у него математическое ожидание равно нулю, а отклонения ε_t некоррелированы.

2.4.2 Автокорреляционная функция

Цель освоения вопроса – получить представление о более сложной характеристике временного ряда – автокорреляционной функции

Вопросы, на которые нужно стараться отвечать при работе с материалом:

- 1) Что означает термин «авто» в нашем случае?

2) Есть ли связь между терминами «парный коэффициент корреляции» и «корреляционная функция»?

Рекомендуемые источники [1, с. 136-139; 4; 6; 7; 9]

Для описания стационарного временного ряда широко используется автокорреляционная функция, которая выражает тесноту связи между членами одного и того же ряда, сдвинутыми на τ единиц времени:

$$\rho(\tau) = M[(y_t - a)(y_{t+\tau} - a)] / \sigma^2 \quad (2.34)$$

Выборочная автокорреляционная функция определяется по формуле:

$$r(\tau) = \frac{\frac{1}{n - \tau} \sum_{t=1}^{n-\tau} (y(t) - \bar{y})(y(t + \tau) - \bar{y})}{\frac{1}{n} \sum_{t=1}^n (y(t) - \bar{y})^2} \quad (2.35)$$

При увеличении τ (лага) модуль $r(\tau)$ стремится к нулю. У быстро меняющегося временного ряда $r(\tau)$ быстро затухает, у медленно меняющегося ряда – медленно затухает. Размер лага ограничивают выражением $\tau \leq n/4$.

2.4.3 Прогнозирование на основе моделей временных рядов

Цель освоения вопроса – получить представление о форме модели временного ряда.

Вопросы, на которые нужно стараться отвечать в процессе работы с материалом:

- 1) Как время может выступать в роли факторной переменной?
- 2) Верно ли утверждение: моделью временного ряда является его тренд?

Рекомендуемые источники [1, с. 139-149; 6; 7; 9; 10].

Здесь мы рассмотрим модели временных рядов, у которых имеется факторная переменная «время»: $y_t = f(t) + \varepsilon_t$.

Для прогнозирования на основе временного ряда главное - выявить тенденцию (тренд) процесса. *При этом исходят из предположения: тенденция развития в прошлом периоде может быть распространена (экстраполирована) на будущий период.*

Для выявления тренда используют два метода: 1) метод наименьших квадратов; 2) метод выравнивания (сглаживания) ряда.

Метод наименьших квадратов. Здесь сначала выбирают вид функции, например, линейную $f(t) = b_0 + b_1 t$; квадратичную $f(t) = b_0 + b_1 t + b_2 t^2$. Отбирается функция на основе компромисса между тремя критериями: 1) она должна лучше других подходить по смыслу; 2) для неё сумма квадратов отклонений фактических данных от расчетных ($\sum_{t=1}^n e^2$) должна быть наименьшей; 3) она

должна быть наиболее простой. После подбора функции применяется собственно метод наименьших квадратов, роль факторной переменной играет время t (1, 2, ..., n). Получаем модель в виде уравнения регрессии.

Метод скользящих средних. Этот метод выравнивает (сглаживает) временной ряд, выделяя неслучайную составляющую. Он основан на переходе от исходных значений членов ряда к их средним значениям на рабочем интервале времени, длина которого m определена заранее, например, $m=3$ или 5. При этом рабочий интервал времени «скользит» вдоль ряда. Получаемый таким образом ряд скользящих средних выглядит более гладко. Для усреднения можно использовать среднюю арифметическую (простую или с некоторыми весами), медиану и др. В конце этой процедуры может быть применён метод наименьших квадратов.

2.4.4 Авторегрессионные модели и модели скользящей средней

Цель освоения вопроса – *получить представление о трёх типах моделей временных рядов без факторной переменной «время».*

Вопросы, на которые нужно стараться отвечать при работе с материалом:

1) Как называется модель, у которой факторы – лаговые результирующие переменные?

2) Как называется модель, у которой факторы – лаговые переменные отклонения?

Рекомендуемые источники [1, с. 146-149; 7; 9; 10].

Первый тип моделей временных рядов без фактора «время» - это линейные модели, у которых факторы – лаговые результирующие переменные – авторегрессионная модель p -го порядка, модель $AR(p)$:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \varepsilon_t. \quad (2.36)$$

При $p=1$ имеем простую авторегрессионную модель 1-го порядка:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t. \quad (2.37)$$

Пример такой прогностической модели: курс рубля на завтра y_t зависит от его курса y_{t-1} на сегодня.

Второй тип моделей временных рядов без фактора «время» - это линейные модели, у которых факторы – лаговые переменные отклонения - модель скользящей средней q -го порядка, модель $MA(q)$:

$$y_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \dots + \gamma_q \varepsilon_{t-q} \quad (2.38)$$

Третий тип моделей временных рядов без фактора «время» - гибридные линейные модели, у которых факторы – лаговые переменные - результирующие и отклонения, модель ARMA(p,q):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \dots + \gamma_q \varepsilon_{t-q} \quad (2.39)$$

Это – авторегрессионная модель скользящей средней порядков p и q.

3 Содержание и методические указания по подготовке к практическим занятиям

Цель проведения практических занятий – закрепление теоретических знаний, полученных на лекциях и – главное - при самостоятельном изучении дисциплины, для выработки профессиональных умений и навыков, сформулированных в рабочей программе.

Учебным планом 2020 г. для заочной формы обучения на 1-4 темы отведено 1, 2, 2, 1 академических часов (40 мин/час) соответственно. Таким образом, на занятиях можно успеть показать решения основных типовых задач, ответить на вопросы, разобрать по несколько тестовых заданий. Поэтому главное внимание следует уделить самостоятельной работе, её объём $55,25/4 \approx 14$ акад. час/тема, что не так уж и мало.

Итак, основное внимание следует уделить самостоятельной работе, подготовке к лекционным и практическим занятиям. Перед практическим занятием следует постараться ответить на вопросы по теме, привлекая конспект лекции, настоящее пособие (раздел 2), рекомендованные преподавателем источники.

При подготовке следует обращать внимание на практическое применение теории и на методику решения типовых задач. Наиболее целесообразная форма проведения занятия - интерактивное практическое занятие: в процессе выполнения заданий приветствуются вопросы со стороны студентов и коллективный поиск ответов на них, обсуждение сложных моментов.

Напомним: параллельно должна выполняться контрольная работа. А если она уже выполнена к началу проведения практических занятий, то её следует снова и снова осмысливать.

Итак, методические указания по подготовке к практическим занятиям.

Для того, чтобы каждое практическое занятие прошло с максимальной пользой, настоятельно рекомендуем готовиться к ним. В самом кратком варианте такая подготовка может состоять в следующем. Ищется ответ на каждый вопрос, который представлен здесь. При этом не следует стремиться охватить слишком большой объём материала. Постарайтесь ответить кратко и точно на поставленный вопрос. Ответы желательно законспектировать, вдумчиво несколько раз повторить их для себя.

С полученным запасом знаний вам будет интересно участвовать в опросе по теме, работать с тестом и решать задачи.

Примечание. Приведенные ниже вопросы взяты без изменения из 2-го раздела настоящего учебно-методического пособия, они разбиты по 4-м темам и по «порциям» - пунктам - материала внутри тем, их нумерация сохранена, все ссылки в них – на 2-й раздел.

Рекомендуемые источники: [1, 2; 9; 10], раздел 2 настоящего пособия.

Занятие по теме 1. Современная теория и практика эконометрического моделирования

Форма занятия: интерактивное практическое занятие.

План занятия:

- a) Краткий опрос по теме.
- b) Работа с тестом.
- c) Решение основных типовых задач.

Вопросы по теме 1 (левая графа – номер пункта с ответами):

- 1) Что вы можете добавить к сказанному ниже про общую тенденцию XX века - математизацию знаний?
- 2.1.1 2) Что является предметом и методом эконометрики?
- 3) Как вам представляется симбиоз эконометрики и искусственного интеллекта?
- 2.1.2 1) В чём нестандартность эконометрической модели в первом примере?
- 2) В чём нестандартность эконометрической модели во втором примере?
- 2.1.3 1) Какие задачи исследования функции могут возникать, помимо указанных ниже?
- 2) Какие два значения имеет термин χ^2 (хи-квадрат)?
- 3) Что означают два числа в записи стандартного нормального закона распределения: НЗР $N(0, 1)$, каков его графический образ?
- 4) Как можно назвать значения X_1 и X_2 на оси абсцисс, см. рисунок 2.2?
- 2.1.4 1) У компании 7 магазинов, на 1 января их годовая выручка составила (Y_1, Y_2, \dots, Y_7), это пространственная выборка?
- 2) Число постоянных жителей пяти самых крупных городов области по состоянию на 1 января составило (X_1, X_2, \dots, X_5), это временная выборка?
- 3) Среднемесячная выручка магазина «Тюлька» за последние пять месяцев составила (B_1, B_2, \dots, B_5), это панельные данные?

Занятие по теме 2. Построение и исследование парных и множественных регрессионных моделей

Форма занятия: интерактивное практическое занятие.

План занятия:

а) Опрос по материалам лекций.

б) Работа с тестом.

в) Решение задач.

Вопросы по теме 2 (левая графа – номер пункта с ответами):

- 2.2.1 1) В чём отличие переменных $\beta_0, \beta_1, \varepsilon_7$ от соответствующих им переменных b_0, b_1, e_7 ? Создайте их графические образы на рисунке 2.3.
- 2) Параметры регрессии b_0, b_1 называют оптимальными (наилучшими), какой смысл вкладывают в это?
- 3) Если судить по рисунку 2.3, то с какой ситуацией мы имеем дело: гомо- или гетероскедастичности? Создайте на этом рисунке графический образ статистики σ .
- 4) Создайте на рисунке 2.3 графический образ тех самых квадратов, о которых идёт речь в методе наименьших квадратов.
- 2.2.2 1) Покажите на рисунке 2.3 отрезки $\hat{y}_2, \hat{y}_3, \hat{y}_6, y_2, y_3, y_6$.
- 2) Что означают фразы: оценка несмещённая, оценка имеет наименьшую дисперсию?
- 2.2.3 1) Что означают жирно выделенные линии-границы на рисунке 2.4?
- 2) Что означают пунктиром выделенные границы на рисунке 2.4?
- 3) Что можно сказать о доверительном интервале на основе (2.13)?
- 2.2.4 1) Какие признаки характеризуют функцию (2.15) как линейную?
- 2) Дайте определение каждого термина из названия модели: *классическая, нормальная, линейная, модель, множественная, регрессия*.
- 3) $(y_1, y_2, \dots, y_n)'$ – это матрица-строка или матрица-столбик?
- 4) Что является неизвестным в системе уравнений, записанных в матричной форме: $X'X \cdot b = X'Y$?
- 2.2.5 1) Чему равно выражение $e' \cdot e$ при $e' = (2 \ 3)$?
- 2) Чему равно выражение $e \cdot e'$ при $e' = (2 \ 3)$?
- 3) Что собой представляет матрица-столбец Y на рисунке 2.5?
- 2.2.6 1) Какую размерность имеет матрица Σ_b для тройственной регрессии?
- 2) Что означает дословно термин «ковариация»?
- 3) Что за выражение (переменная) σ^2 в (2.20)?
- 2.2.7 1) При расчёте дисперсии СВ b_1 какой диагональный элемент об-

ратной матрицы нужно выбрать, см. (2.23)?

2) Если нулевая гипотеза $H_0: \beta_j = 0$ отвергается, то какой из этого следует вывод?

3) Как будет выглядеть графический образ доверительного интервала для β_1 на рисунке 2.3?

2.2.8 1) Как определить значение $t_{1-\alpha, n-2}$ в (2.26)?

2) Справедливо ли суждение по (2.27): чем больше Q_R , тем более значима регрессия? Какой в этом смысл?

3) Справедливо ли суждение по (2.29): за введение в регрессию каждой дополнительной факторной переменной уплачивается штраф?

Занятие по теме 3. Техника построения эконометрических моделей.

Форма занятия: интерактивное практическое занятие.

План занятия:

- a) Опрос по материалам лекций.
- b) Работа с тестом.
- c) Решение задач.

Вопросы по теме 3 (левая графа – номер пункта с ответами):

2.3.1 1) Почему при мультиколлинеарности оценки очень чувствительны к исходным данным?

2) Сколько всего регрессий нужно построить, чтобы реализовать алгоритм, показанный на рисунке 2.6?

3) Какой будет траектория решения, если в задаче на рисунке 2.6 использовать не скорректированный, а обычный коэффициент детерминации?

2.3.2 1) Составляется таблица с графами: ФИО, Стаж, Пол, Зарплата. Как будет выглядеть графа Пол с булевской переменной z ?

2) Верно ли утверждение: для атрибутивной переменной с 4-мя уровнями нужно ввести 3 булевских переменных?

2.3.3 1) Что означает термин: функция, линейная по параметрам?
2) За счёт чего достигается «внутренняя линейность»?

2.3.4 1) Как выглядит график функции эластичности для линейной парной регрессии (см. таблицу 2.7, следует применить правило Лопиталля)?

2) Можно ли в качестве фактора L – труд - использовать показатель (чел*смен/год)?

- 2.3.5 1) Как будет выглядеть числовой пример для выражения: $\sum \varepsilon = \sigma^2 E_n$?
- 2) Можно ли применять обычный МНК в случае гетероскедастичности?
- 3) В чём отличие парного коэффициента корреляции от рангового?
- 4) В чём суть теста ранговой корреляции Спирмена? Можно ли его применить в случае, показанном на рисунке 2.7.

Занятие по теме 4. Временные ряды и динамические регрессионные модели

Форма занятия: интерактивное практическое занятие.

План занятия:

- a) Опрос по материалам лекций.
- b) Работа с тестом.
- c) Решение задач.

Вопросы по теме 4 (левая графа – номер пункта с ответами):

- 1) Из каких компонент может состоять временной ряд?
- 2.4.1 2) Какой ряд называется стационарным?
- 3) О каких числовых характеристиках ряда идёт речь?
- 1) Что означает термин «авто» в нашем случае?
- 2.4.2 2) Есть ли связь между терминами «парный коэффициент корреляции» и «корреляционная функция»?
- 1) Как время может выступать в роли факторной переменной?
- 2.4.3 2) Верно ли утверждение: моделью временного ряда является его тренд?
- 1) Как называется модель, у которой факторы - лаговые результирующие переменные?
- 2.4.4 2) Как называется модель, у которой факторы - лаговые переменные отклонения?

4 Задания и методические указания по выполнению контрольной работы

4.1 Общие сведения, выбор варианта, теоретических вопросов, исходных данных для задач

Контрольная работа (для дневной формы обучения – самостоятельная работа) состоит из шести разделов: 1) два теоретических вопроса; 2) задача 1; 3) задача 2; 4) задача 3; 5) задача 4; 6) задача 5. И ещё - список использованных источников. В приложении Б дана структура работы в форме листа «СОДЕРЖАНИЕ».

Объём работы не должен превышать 22 страницы, не следует без нужды увеличивать его, например, путём включения статистических таблиц и т.п.

Работа должна быть оформлена строго по правилам, изложенным в /5, Мнацаканян и др./, где имеется образец титульного листа, правила написания всего текста с его формулами, рисунками, таблицами, списком использованных источников.

Трудоёмкость выполнения контрольной работы снижается в 10 раз за счёт использования ворд-файла настоящего раздела пособия (обращаться к преподавателю). Студент должен, во-первых, вставить свои исходные данные в формулы и таблицы, поправить готовый текст и рисунки, выполнить вычисления - вручную и на ЭВМ, во-вторых, разобраться с экономическим и эконометрическим смыслом каждого элемента, части работы. *Ниже для решения каждой из пяти задач даются подробнейшие инструкции и наглядные примеры.*

Выбор варианта и номера вопроса. Вариант выбирается по первой подходящей гласной букве полного имени и первой подходящей (гласной либо согласной) букве фамилии. Например, если Вы – тАтьяна сОколова, то **ваш вариант АО**, ваши теоретические вопросы 11 и 31, см. таблицу 4.1.

Таблица 4.1 – Распределение номеров теоретических вопросов по вариантам

| Имя* | Первая подходящая буква фамилии | | | | | | | | | | | | |
|-------------|---------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-------------------|-----------|-----------|
| | А | Б | В | Г | Д | Е | К | Л | М | Н | О | П | Р |
| А, Е | 1, 21 | 2, 22 | 3, 23 | 4, 24 | 5, 25 | 6, 26 | 7, 27 | 8, 28 | 9, 29 | 10, 30 | 11, 31 | 12, 32 | 13, 33 |
| И, О | 14, 34 | 15, 35 | 16, 36 | 17, 37 | 18, 38 | 19, 39 | 20, 1 | 21, 2 | 22, 3 | 23, 4 | 24, 5 | 25, 6 | 26, 7 |
| У, Я | 27, 8 | 28, 9 | 29, 10 | 30, 11 | 31, 12 | 32, 13 | 33, 14 | 34, 15 | 35, 16 | 36, 17 | 37, 18 | 38, 19 | 39, 20 |

*первая подходящая гласная буква полного имени

Список вопросов см. в разделе 5. Вопросы узкие, поэтому и ответы на 2 вопроса должны занимать не более одной-двух страниц.

Рекомендации по решению задач

Задача 1. Пример решения задачи 1 для варианта А-Ъ дан в 4.2. Задача решается вручную – это очень важно для понимания теории и овладения техникой расчётов - на основе данных, приведенных в приложении А, таблицах А.1-А.4. Для проверки правильности решения на отдельных шагах можно пользоваться, например, матричным калькулятором. *Здесь требуется только внимание и терпение, ничего сложного нет!!!*

Задача 2. Пример решения задачи 2 для варианта А-Ъ дан в 4.3. Задача решается с помощью табличного редактора MS Excel на основе тех же исходных данных, что и задача 1. Смысл решения: а) проверить правильность ручного решения задачи 1; б) освоить инструменты Excel для решения эконометрических задач.

У задач 1 и 2 одна и та же содержательная постановка, одинаковый экономический смысл, одинаковые исходные данные: столбцы X1, X2, Y, см приложение А. Сначала по первой подходящей гласной имени выбирается таблица из набора А.1-А.3. Из выбранной таблицы:

1) считывается столбец X1 (он слева),

2) по первой подходящей букве фамилии выбирается и считывается столбец Y. Например, если Вы – тАтьяна сОколова, то ваш вариант АО, Ваша таблица А.1, а в ней графа для X1 и графа для Y с буквой О - выделены жирно.

Затем из таблицы А.4 выбирается столбец X2 по первой подходящей гласной букве полного имени. Столбцы X1, X2, Y содержат по 7 целых чисел.

Для автоматического ежегодного изменения исходных данных в таблицы введена переменная «Г» - год, последняя цифра календарного года получения задания. Например, если год получения задания 2021, то Г=1.

У задач 3 и 4 одна и та же содержательная постановка. Они решаются с помощью MS Excel на основе одних и тех же данных: объём выпуска продукции предприятием за последние 20 кварталов в стоимостном выражении.

Исходные данные для решения задач 3 и 4 одни и те же. В части, общей для всех вариантов, они приведены в таблице А.5. Там имеются всего 20 клеток для временного ряда. Числа в 13-ти клетках (с 5-й по 17-ю) для всех вариантов одни и те же, а 7 клеток (1-4 и 18-20) пустые. Они заполняются каждым студентом самостоятельно на основе столбца Y задачи 1 для своего варианта.

Правило заполнения:

1) очередное число из столбца Y умножается на 100;

2) если произведение меньше 300, то в соответствующую клетку таблицы А.5 записывается 300;

3) если произведение больше 1200, то в соответствующую клетку таблицы А.5 записывается 1200;

4) иначе произведение без изменения записывается в соответствующую клетку таблицы А.5.

В таблице 4.2 см. пример сформированного временного ряда.

Таблица 4.2 – Формирование поквартального временного ряда для варианта АО

| Квартал | Год наблюдения | | | | |
|---------|--------------------|-------|-------|-------|----------------------|
| | 1* | 2 | 3 | 4 | 5* |
| I | $2*100 \div 300,0$ | 660,0 | 830,0 | 750,0 | 750,0 |
| II | $3*100=300,0$ | 670,0 | 810,0 | 710,0 | $9*100=900,0$ |
| III | $6*100=600,0$ | 680,0 | 750,0 | 660,0 | $9*100=900,0$ |
| IV | $6*100=600,0$ | 690,0 | 780,0 | 690,0 | $20*100 \div 1200,0$ |

*год 2020, поэтому $\Gamma=0$

Задача 5 решается с помощью MS Excel. Исходные данные - временной ряд из 10 чисел – см. в таблице А.6. В этом ряду первые 4 числа отсутствуют, они вычисляются по правилу, приведенному выше. Другими словами: в задаче 5 используются первые 10 чисел временного ряда из задач 3 и 4. В таблице 4.3 дан пример ряда для задачи 5, вариант АО.

Таблица 4.3 – Валовой региональный продукт по годам за 10 лет для варианта АО, млрд. руб.

| Год | ВРП | Год | ВРП | Год | ВРП |
|-----|-------|-----|-------|-----|-------|
| 1 | 300,0 | 5 | 660,0 | 9 | 830,0 |
| 2 | 300,0 | 6 | 670,0 | 10 | 810,0 |
| 3 | 600,0 | 7 | 680,0 | - | - |
| 4 | 600,0 | 8 | 690,0 | - | - |

4.2 Пример ручного решения задачи 1. Моделирование магазина на основе линейной троичной регрессии

Постановка задачи 1 (общая для всех вариантов). Торговая компания располагает семью магазинами типа «Морепродукты». Компания планирует открыть восьмой магазин, для этого её экономисты разрабатывают бизнес-план и в том числе — эконометрическую модель магазина. На этой модели они хотят исследовать зависимость среднегодовой суточной выручки (Y руб/день) от площади магазина (X_1 м²) и площади паркинга (X_2 ед.авто), чтобы спрогнозировать среднесуточную выручку 8-го магазина (проектные данные: площадь 1100 м², паркинг 80 ед авто).

1) Выберем данные нашего варианта «А-Ъ» для x_1 , x_2 , y из приложения А - таблицы А.1 и А.4 и перенесём их в таблицу 4.4.

Таблица 4.4 – Исходные данные для задачи 1 вариант «А-Ъ»

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|-----------------------------------|
| • | • | • | • | • | • | • | • | • | • м ² *1 00 |
| • | • | • | • | • | • | • | • | • | • ед *1 0 |
| • | • | • | • | • | • | • | • | • | • ру б/ су т* 10 5 |

*курсив – проектные данные.

Как видно из таблицы 4.4 зависимость $\hat{y}(x_1, x_2)$ прямая. Проверим, можно ли аппроксимировать её плоскостью – линейной функцией:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 \quad (4.1)$$

Для этого положим $x_2 = 0$ и получим функцию $\hat{y} = b_0 + b_1 x_1$ – след от пересечения плоскости (4.1) с плоскостью $x_2 = 0$. Проверим: будет ли этот след хорошей аппроксимацией соответствующего корреляционного поля. Для этого построим график зависимости y от x_1 , используя табличный редактор Excel (см. приложение В, где дана инструкция). На рисунке 4.1 видно, что корреляционное поле может быть хорошо аппроксимировано прямой линией. Так же можно поступить и с зависимостью $\hat{y} = b_0 + b_2 x_2$.

Вывод: зависимость $\hat{y}(x_1, x_2)$ хорошо аппроксимируется линейной функцией (4.1). Заметим: линию на рисунке 4.1 компьютер строит по методу наименьших квадратов, она является парной линейной регрессией.

2) Запишем матрицу X значений объясняющих переменных (матрицу плана), см. ниже среднюю матрицу. Запишем слева от неё транспонированную матрицу плана X' (здесь штрих – оператор транспонирования). Найдём матрицу A как произведение матриц $X'X$:

$$A = X'X = \begin{pmatrix} 11 & 11 & 11 \\ 1 & 1 & 2 & 3 & 4 & 5 & 8 \\ 1 & 2 & 2 & 3 & 3 & 4 & 6 \end{pmatrix} \cdot \begin{pmatrix} 11 & 1 \\ 11 & 2 \\ 12 & 2 \\ 13 & 3 \\ 14 & 3 \\ 15 & 4 \\ 18 & 6 \end{pmatrix} = \begin{pmatrix} 7 & 24 & 21 \\ 24 & 120 & 96 \\ 21 & 96 & 79 \end{pmatrix}$$

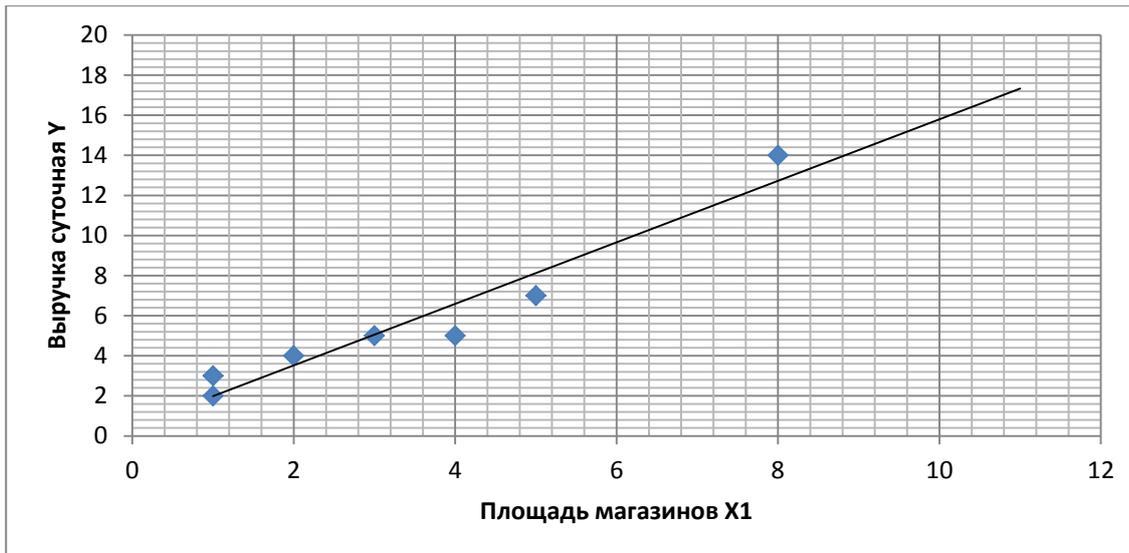


Рисунок 4.1 – Корреляционное поле и линия регрессии $\hat{y} = b_0 + b_1 x_1$

3) Найдём обратную матрицу $A^{-1} = (X'X)^{-1}$ по формуле:

$$A^{-1} = \frac{1}{|A|} (A_{ij})'$$

где: $|A|$ - определитель матрицы A ,

$(A_{ij})'$ - транспонированная матрица, составленная из алгебраических дополнений матрицы A .

Определитель: $|A| = 7 \cdot 120 \cdot 79 + 24 \cdot 96 \cdot 21 + 21 \cdot 96 \cdot 24 - 21 \cdot 120 \cdot 21 - 96 \cdot 96 \cdot 7 - 79 \cdot 24 \cdot 24 = 192$.

Алгебраические дополнения (слева сумма индексов чётная, справа - нечётная):

$$\begin{aligned} A_{11} &= 120 \cdot 79 - 96 \cdot 96 = 264; & A_{12} &= -(24 \cdot 79 - 96 \cdot 21) = 120; \\ A_{13} &= 24 \cdot 96 - 120 \cdot 21 = -216; & A_{21} &= -(24 \cdot 79 - 21 \cdot 96) = 120; \\ A_{22} &= 7 \cdot 79 - 21 \cdot 21 = 112; & A_{23} &= -(7 \cdot 96 - 24 \cdot 21) = -168; \\ A_{31} &= 24 \cdot 96 - 21 \cdot 120 = -216; & A_{32} &= -(7 \cdot 96 - 21 \cdot 24) = -168; \\ A_{33} &= 7 \cdot 120 - 24 \cdot 24 = 264. \end{aligned}$$

Искомая обратная матрица:

$$A^{-1} = \frac{1}{192} \begin{pmatrix} 264 & 120 & -216 \\ 120 & 112 & -168 \\ -216 & -168 & 264 \end{pmatrix} = \begin{pmatrix} 1,375 & 0,625 & -1,125 \\ 0,625 & 0,584 & -0,875 \\ -1,125 & -0,875 & 1,375 \end{pmatrix}.$$

Проверим, выполняется ли равенство: $A^{-1}A = E$?

$$\begin{aligned} AA^{-1} &= \frac{1}{192} \begin{pmatrix} 7 & 24 & 21 \\ 24 & 120 & 96 \\ 21 & 96 & 79 \end{pmatrix} \begin{pmatrix} 264 & 120 & -216 \\ 120 & 112 & -168 \\ -216 & -168 & 264 \end{pmatrix} = \\ &= \frac{1}{192} \begin{pmatrix} 192 & 0 & 0 \\ 0 & 192 & 0 \\ 0 & 0 & 192 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Вывод: равенство выполняется, расчет обратной матрицы выполнен верно.

4) Найдём произведение матриц $X'Y$:

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 4 & 5 & 8 \\ 1 & 2 & 2 & 3 & 3 & 4 & 6 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \\ 5 \\ 7 \\ 14 \end{pmatrix} = \begin{pmatrix} 40 \\ 195 \\ 158 \end{pmatrix}.$$

5) Найдём параметры уравнения троичной регрессии (4.1) методом наименьших квадратов (ниже b – это вектор параметров регрессии):

$$b = (X'X)^{-1} X'Y = (1/192) \begin{pmatrix} 264 & 120 & -216 \\ 120 & 112 & -168 \\ -216 & -168 & 264 \end{pmatrix} \cdot \begin{pmatrix} 40 \\ 195 \\ 158 \end{pmatrix} = (1/192) \cdot \begin{pmatrix} -168 \\ 96 \\ 312 \end{pmatrix} = \begin{pmatrix} -0,88 \\ 0,50 \\ 1,63 \end{pmatrix}.$$

Ответ: $b_0 = -0,88$; $b_1 = 0,50$; $b_2 = 1,63$. Наше уравнение троичной регрессии имеет вид:

$$\hat{y} = -0,88 + 0,50x_1 + 1,63x_2. \quad (4.2)$$

6) Найдём точечный прогноз выручки – регрессионное значение - для проектируемого магазина-8 с площадью $x_1=11$ (1100 м²) и площадью паркинга $x_2 = 8$ (80 автомашин): $\hat{y} = -0,88 + 0,50 \cdot 11 + 1,63 \cdot 8 = 17,66$.

7) Найдём 95%-ый доверительный интервальный прогноз выручки для регрессионного значения магазина-8 при значении факторов: $x_1=11$, $x_2=8$. Для этого нужно найти дисперсию s^2 СВ отклонений e и дисперсию $s_{\hat{y}}$ СВ \hat{y} по формулам:

$$s^2 = \frac{e'e}{n-(p+1)} = \frac{\sum e_i^2}{n-p-1}, \quad s_{\hat{y}} = s\sqrt{X_0'(X'X)^{-1}X_0};$$

Для удобства вычислений составим таблицу 4.5.

Таблица 4.5 – Расчётная таблица для задачи 1.1

| №п/п | x_{i1} | x_{i2} | y_i | $x_{i1} \cdot x_{i2}$ | \hat{y}_i | e_i | e_i^2 |
|------------------|----------|----------|-------|-----------------------|-------------|-------|---------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 1 | 1 | 2 | 1 | 1,25 | 0,75 | 0,56 |
| 2 | 1 | 2 | 3 | 2 | 2,88 | 0,12 | 0,02 |
| 3 | 2 | 2 | 4 | 4 | 3,38 | 0,62 | 0,39 |
| 4 | 3 | 3 | 5 | 9 | 5,51 | -0,51 | 0,26 |
| 5 | 4 | 3 | 5 | 12 | 6,01 | -1,01 | 1,02 |
| 6 | 5 | 4 | 7 | 20 | 8,14 | -1,14 | 1,30 |
| 7 | 8 | 6 | 14 | 48 | 12,90 | 1,10 | 1,21 |
| Итого Σ | 24 | 21 | 40 | 96 | 40,07 | -0,07 | 4,76 |
| Среднее значение | 3,43 | 3,0 | 5,71 | 13,71 | | | |

На основе табличных данных рассчитываем дисперсию и СКО:

$$s^2 = 4,76/(7-2-1) = 1,19; \quad s = \sqrt{1,19} = 1,09,$$

$$X_0'(X'X)^{-1} = (1 \ 11 \ 8) \begin{pmatrix} 1,375 & 0,625 & -1,125 \\ 0,625 & 0,584 & -0,875 \\ -1,125 & -0,875 & 1,375 \end{pmatrix} = (-0,750 \ 0,049 \ 0,250),$$

$$(X_0'(X'X)^{-1})X_0 = (-0,750 \ 0,049 \ 0,250) \begin{pmatrix} 1 \\ 11 \\ 8 \end{pmatrix} = 1,789 \approx 1,79,$$

$$s_{\hat{y}} = 1,09\sqrt{1,79} = 1,46.$$

По таблице из приложения Г находим критическое значение статистики Стьюдента $t_{0,95; 7-2-1=4} = 2,78$. Полуинтервал $\Delta = t_{0,95; 4} \cdot s_{\hat{y}} = 2,78 \cdot 1,46 = 4,05$.

Нижняя граница доверительного интервала: $\hat{y}_{min} = \hat{y}_{X_0} - \Delta = 17,66 - 4,05 = 13,61$.

Верхняя граница доверительного интервала: $\hat{y}_{max} = \hat{y}_{X_0} + \Delta = 17,66 + 4,05 = 21,71$.

Ответ: искомый доверительный интервал для прогнозного значения \hat{y}_{X_0} - рекомендуем построить его на рисунке 4.1 - задаётся системой неравенств:

$$13,61 \leq M_{X_0}(Y) \leq 21,71. \quad (4.3)$$

8) **Рассчитаем** коэффициент корреляции между факторными переменными – проверка на мультиколлинеарность:

$$r_{x_1x_2} = \frac{\overline{X_1 \cdot X_2} - \overline{X_1} \cdot \overline{X_2}}{s_{x_1} \cdot s_{x_2}}$$

Здесь известно всё, кроме СКО s_{x_2} . Вычислим это стандартное отклонение:

$$\text{дисперсия } s_{x_2}^2 = \overline{x_2^2} - \bar{x}_2^2 = 120/7 - 3,0 \cdot 3,0 = 8,14; \quad \text{СКО } s_{x_2} = \sqrt{8,14} = 2,85.$$

Искомый коэффициент корреляции:

$$r_{x_1x_2} = \frac{13,71 - 3,43 \cdot 3,0}{2,32 \cdot 2,85} = 0,52.$$

Вывод: мультиколлинеарность факторов X_1 и X_2 достаточно велика, поэтому в дальнейшем анализе можно ожидать противоречивых результатов.

9) **Проверим** значимость коэффициентов регрессии. СКО коэффициента рассчитывается по формуле:

$$s_{b_j} = s \sqrt{[(X'X)^{-1}]_{jj}}$$

где $j=0, 1, 2$, а выражение под корнем есть диагональный элемент обратной матрицы $(X'X)^{-1}$, см п.2. Отсюда:

$$s_{b_1} = 1,09 \sqrt{0,584} = 0,83; \quad s_{b_2} = 1,09 \sqrt{1,375} = 1,28.$$

Имеем: а) так как $t = |b_1| / s_{b_1} = 0,50/0,83 = 0,60 < t_{0,95;4} = 2,78$, то **коэффициент b_1 незначим** (незначимо отличается от нуля).

б) так как $t = |b_2| / s_{b_2} = 1,63/1,28 = 1,27 < t_{0,95;4} = 2,78$, то **коэффициент b_2 незначим**.

Выводы: а) на такой результат могли повлиять: мультиколлинеарность факторов, см. п. 9, и слишком малые объёмы выборок,

б) полученный результат влечёт незначимость уравнения регрессии, см. п.10.

10) **Оценим** значимость уравнения регрессии на уровне $\alpha=0,05$ по коэффициенту детерминации:

$$R^2 = Q_R / Q = (b'X'Y - n\bar{y}^2) / (Y'Y - n\bar{y}^2)$$

$$b'(X'Y) = (-0,88 \quad 0,50 \quad 1,63) \begin{pmatrix} 40 \\ 195 \\ 158 \end{pmatrix} = 319,84.$$

$\bar{y} = 40/7 = 5,72$ (см. табл. 4.1) ; $Y'Y = \sum y_i^2 = 324$. Окончательно:

$$R^2 = (319,84 - 7 \cdot 5,72 \cdot 5,72) / (324 - 7 \cdot 5,72 \cdot 5,72) = 90,81 / 94,97 = 0,96.$$

Вывод: значимость регрессии здесь очень высокая, хотя в п.9 была установлена незначимость обоих коэффициентов регрессии. Противоречие.

11) Оценим значимость уравнения регрессии на уровне $\alpha=0,05$ по критерию Фишера. Регрессия значима, если справедливо неравенство:

$$F = R^2 (n-p-1) / ((1-R^2) p) > F_{\alpha; k_1; k_2}.$$

$$\text{Отсюда } F = 0,96^2 (7-2-1) / ((1-0,96^2) 2) = 23,0 > (F_{0,05; 2; 4} = 6,94).$$

Вывод: уравнение значимо. Снова противоречие с п.9.

Вывод по решению задачи 1: полученная модель магазина (4.2) несовершенна. План её доработки:

- 1) рассчитать коэффициенты корреляции $r_{x_1, \hat{y}}$ и $r_{x_2, \hat{y}}$;
- 2) если первый коэффициент больше второго, то фактор x_2 исключить из регрессии, а вместо него включить, например, фактор x_3 – средняя ширина ассортимента магазина (ширина ассортимента – это общее количество категорий товаров);
- 3) если второй коэффициент больше первого, то фактор x_1 исключить из регрессии, а вместо него включить, например, фактор x_3 .

4.3 Пример решения задачи 2 в MS Excel. Моделирование магазина на основе линейной троичной регрессии

Постановка задачи 2 и исходные данные те же, что и для задачи 1, выбор исходных данных см. в 4.2.

Для решения используется функция «Регрессия» из пакета анализа данных.

- 1) Внесём** в лист MS Excel исходные данные своего варианта.

| | A | B | C | D | E |
|---|-------|----------|----------|---|---|
| 1 | Y_i | X_{1i} | X_{2i} | | |
| 2 | 2 | 1 | 1 | | |
| 3 | 3 | 1 | 2 | | |
| 4 | 4 | 2 | 2 | | |
| 5 | 5 | 3 | 3 | | |
| 6 | 5 | 4 | 3 | | |
| 7 | 7 | 5 | 4 | | |
| 8 | 14 | 8 | 6 | | |

Рисунок 4.2 – Исходные данные для построения троичной регрессии

2) Выберем пункт «Анализ данных» во вкладке «Данные». В появившемся окне найдём инструмент «Регрессия».

3) Зададим входные интервалы переменных X, Y и укажем интервал вывода итогов регрессионного анализа (верхнюю левую ячейку диапазона вывода данных). Выбор пункта «Остатки» позволит получить данные об остатках (ошибках) регрессии.

4) Проведём расчёты и представим результаты с помощью Excel в виде нескольких таблиц, см. рисунок 4.4. Расшифруем полученные данные.

а) Таблица «Регрессионная статистика» содержит:

«Множественный R» – коэффициент множественной корреляции;

«R-квадрат» – коэффициент детерминации;

«Нормированный R-квадрат» – коэффициент детерминации, скорректированный на количество степеней свободы;

«Стандартная ошибка» – квадратный корень из остаточной дисперсии;

«Наблюдения» – количество наблюдений n.

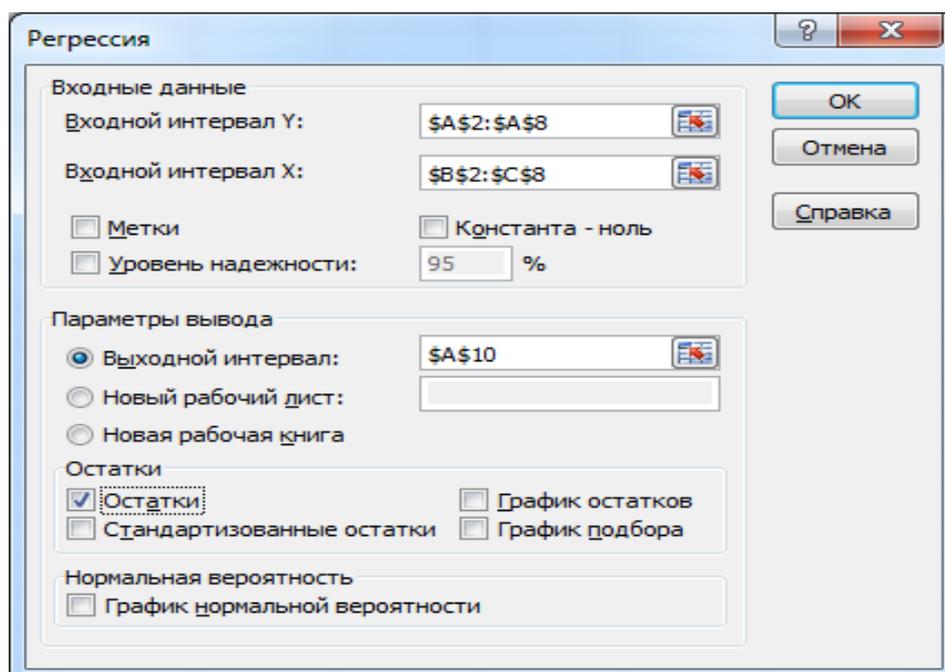


Рисунок 4.3 – Параметры построения троичной регрессии

б) Таблица «Дисперсионный анализ» содержит данные:

- столбец « df » - три числа степеней свободы: для регрессии, остатка и общее;
- столбец « SS » - три суммы квадратов отклонений: для регрессии, остатка и общее;
- столбец « MS » - два числа дисперсии на одну степень свободы: для регрессии и остатка;
- столбец « F » - значение СВ F , распределённой по ЗР Фишера;
- столбец «*Значимость F* » - p -значение для СВ F , т.е. вероятность того, что выполняется нулевая гипотеза о случайном отклонении коэффициента Фишера от нуля. В нашем примере p -значение ничтожно мало, поэтому гипотеза H_0 отклоняется.

в) Следующая таблица содержит информацию о параметрах уравнения регрессии:

- столбец «Коэффициенты» - значения параметров уравнения регрессии. Строка « Y -пересечение» содержит значения параметра b_0 , строка «Переменная X_1 » - параметра b_1 , строка «Переменная X_2 » - параметра b_2 ;
- столбец «Стандартная ошибка» – стандартные ошибки параметров уравнения регрессии. Строка « Y -пересечение» - стандартная ошибка для параметра b_0 , строка «Переменная X_1 » – стандартная ошибка для параметра b_1 , строка «Переменная X_2 » – стандартная ошибка для параметра b_2 ;
- столбец « t -статистика» - значения t -критериев Стьюдента для параметров уравнения регрессии. Строка « Y -пересечение» – значение t -критерия Стьюдента для параметра b_0 , строка «Переменная X_1 » – значение t -критерия Стьюдента для параметра b_1 , строка «Переменная X_2 » – значение t -критерия Стьюдента для параметра b_2 ;
- столбец « p -значение» – p -значения t -критерия Стьюдента для параметров регрессии. Строка « Y -пересечение» – p -значение t -критерия Стьюдента для параметра b_0 , строка «Переменная X_1 » – p -значение t -критерия Стьюдента для параметра b_1 , строка «Переменная X_2 » – p -значение t -критерия Стьюдента для параметра b_2 ;
- столбцы «Верхние 95 %» и «Нижние 95 %» – интервалы для параметров уравнения регрессии, построенные с доверительной вероятностью 95 %.

г) Таблица «Вывод остатка» содержит данные:

«Предсказанное Y » - значения y , рассчитанные по уравнению регрессии (\hat{y}).

«Остатки» - разница между предсказанными значениями y и наблюдаемыми значениями y (е).

5) Конец решения. Сохраним полученный результат для использования в отчётной работе.

| Вывод итогов | | | | | | | | |
|--------------------------|-----------------|--------------------|--------------|-------------|--------------|-------------|--------------|---------------|
| Регрессионная статистика | | | | | | | | |
| Множественный R | 0,97479462 | | | | | | | |
| R-квадрат | 0,950224551 | | | | | | | |
| Нормированный R-квадрат | 0,925336826 | | | | | | | |
| Стандартная ошибка | 1,089724736 | | | | | | | |
| Наблюдения | 7 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | df | SS | MS | F | Значимость F | | | |
| Регрессия | 2 | 90,67857143 | 45,33928571 | 38,18045113 | 0,002477595 | | | |
| Остаток | 4 | 4,75 | 1,1875 | | | | | |
| Итого | 6 | 95,42857143 | | | | | | |
| | Коэффициенты | Стандартная ошибка | t-статистика | P-Значение | Нижние 95% | Верхние 95% | Нижние 95,0% | Верхние 95,0% |
| Y-пересечение | -0,875 | 1,277815519 | -0,684762383 | 0,531111679 | -4,422784642 | 2,672784642 | -4,422784642 | 2,672784642 |
| Переменная X 1 | 0,5 | 0,832291015 | 0,600751409 | 0,580386945 | -1,810810314 | 2,810810314 | -1,810810314 | 2,810810314 |
| Переменная X 2 | 1,625 | 1,277815519 | 1,271701569 | 0,27238898 | -1,922784642 | 5,172784642 | -1,922784642 | 5,172784642 |
| Вывод остатка | | | | | | | | |
| Наблюдение | Предсказанное Y | Остатки | | | | | | |
| 1 | 1,25 | 0,75 | | | | | | |
| 2 | 2,875 | 0,125 | | | | | | |
| 3 | 3,375 | 0,625 | | | | | | |
| 4 | 5,5 | -0,5 | | | | | | |
| 5 | 6 | -1 | | | | | | |
| 6 | 8,125 | -1,125 | | | | | | |
| 7 | 12,875 | 1,125 | | | | | | |

Рисунок 4.4 – Вывод итогов решения задачи 2

Ниже в таблице 4.6 приведён количественный анализ совпадения ручного и компьютерного решений.

Таблица 4.6 – Сравнение ручного и компьютерного решений

| Обозначение и решение ручное, см. пункт 4.2.1 | | Обозначение и решение в MS Excel | | Относит. отклоне- ние, % |
|---|-------|-------------------------------------|--------|--------------------------------|
| 1 | 2 | 3 | 4 | |
| а) Коэффициент детерминации R^2 , п.10 | 0.96 | R-квадрат | 0.951 | 0.9 |
| б) СКО отклонений s , п.7 | 1.09 | Стандартная ошибка | 1.090 | 0.0 |
| в) Свободный член b_0 , п.5 | -0,88 | Y-пересечение - Коэффициенты | -0.875 | 0.6 |
| г) Коэффициент b_1 , п.5 | 0.50 | Переменная X1 - | 0.500 | 0.0 |

| | | | | |
|--------------------------------------|------|------------------------------|-------|-----|
| | | Коэффициенты | | |
| 9 Коэффициент регрессии b_2 , п. 5 | 1.63 | Переменная X1 - Коэффициенты | 1.625 | 0.3 |

*(графа 5: $\frac{(2)-(4)}{(4)}100\%$).

Вывод: относительное отклонение ручного и компьютерного решений не превысило 0,9 %, следовательно, ручное решение было выполнено верно.

4.4 Пример решения задачи 3 в MS Excel. Выделение тренда временного ряда методами скользящей средней

В таблице 4.7 представлен временной ряд поквартального объема выпуска консервов и пресервов на ООО «Морепродукты», принятый для нашего примера - варианта А-Б.

Таблица 4.7 – Поквартальный объём выпуска морепродуктов в сопоставимых ценах (млн руб./квартал)

| Квартал | Год | | | | |
|---------|-------|-------|-------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 |
| I | 601,2 | 666,8 | 826,2 | 754,1 | 688,8 |
| II | 639,7 | 665,3 | 812,0 | 711,8 | 658,3 |
| III | 647,9 | 678,4 | 752,3 | 656,0 | 635,8 |
| IV | 623,0 | 693,7 | 774,1 | 690,0 | 660,0 |

1) Ввод временного ряда, построение его графика, анализ структуры ряда по его графику. Расположим временной ряд в ячейках A1-A21, ряд содержит 20 уровней, $n=20$. Для построения графика временного ряда выделим ячейки A2-A21. Во вкладке «Вставка» выберем вид графика «Точечная с маркерами», получим график временного ряда, см. рисунок 4.5. Из рисунка видно, что уровни ряда в начальном периоде от $t=1$ до $t=10$ имеют тенденцию возрастания, а затем – убывания. Уровни ряда содержат колебания, но их характер не говорит о наличии сезонной компоненты, связанной с кварталами года.

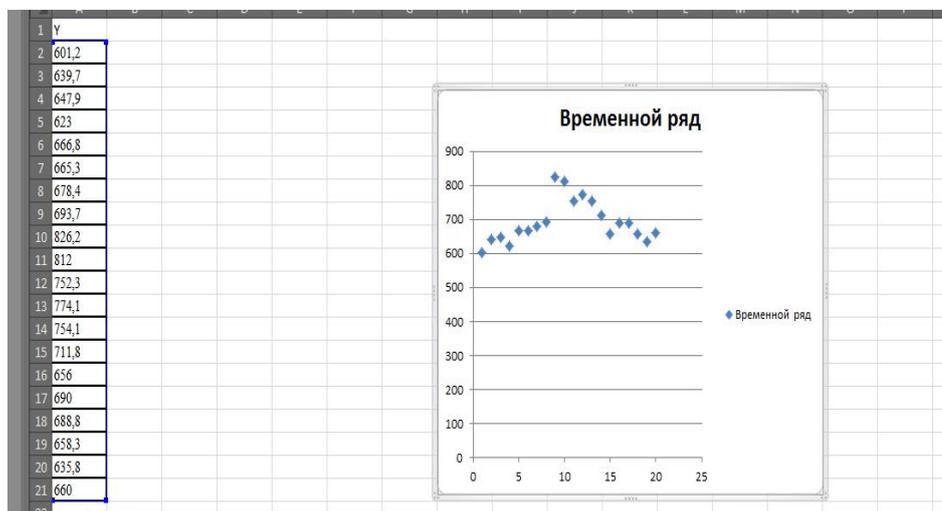


Рисунок 4.5 - Временной ряд и его график

2) Выделение тренда временного ряда методом простой скользящей средней с длиной окна скольжения, равной 4. По вкладке «Данные» выберем «Анализ данных», здесь в окне выберем «Скользящее среднее», заполнение полей его окна:

- в поле *входной интервал* укажем ячейки **A1-A21**, содержащие временной ряд;
- выберем метки в первой строке, так как в ячейке **A1** имя временного ряда;
- в поле *интервал* укажем длину окна скольжения, равную 4;
- в поле *выходной интервал*, для удобства построения графиков ряда и скользящих средних, укажем ячейку **I2**;
- выберем «Вывод графика»; можно ещё выбрать «Стандартные погрешности», в этом случае наряду со сглаженным рядом выводятся и вычисленные по окну скольжения средние квадратические отклонения уровней ряда от сглаженных значений:

$$e_t = \sqrt{\frac{\sum_{i=0}^{T-1} (y_{t-i} - \bar{y}_{t-i})^2}{T}}, \quad (4.4)$$

где T – длина окна скольжения,

\bar{y}_{t-i} – скользящее среднее для момента времени $t-i$.

По «ОК» в ячейках **I2-I21** получим сглаженные значения временного ряда, а в ячейках **J2-J21** значения средних квадратических отклонений, см. рисунок 4.6. Внимание: процедура сглаживания относит сглаженные значения уровней, попавших в окно, к последнему моменту времени в окне скольжения. В данном случае сглаженные значения располагаются, начиная с ячейки **I5**. Необходимо рассчитать сглаженные значения для моментов времени, находящихся в середине окна скольжения (длина окна 4). Значения скользящей средней для 3-го момента времени в окне скольжения находятся как среднее арифметическое скользящего среднего для данного и следующего положения окна скольжения на временной оси. Для этого выделим ячейку **B4** и в строке формул введем **=СРЗНАЧ(I5:I6)**, по «ОК» получим значение скользящей средней для момента времени $t=3$. Аналогично в ячейках **B5-B19** получаем значения скользящей средней для последующих моментов времени, до момента $n-2$ включительно, см. рисунок 4.6.

3) Построение скользящей средней с нечётной длиной окна скольжения, равной пяти, производится так же, результат см. в ячейках **S2-S21**, см. рисунок 4.6. Среднее значение уровней, попавших в интервал сглаживания, приписывается последнему моменту времени, вошедшему в интервал сглаживания. По-

этому их нужно переместить в середины интервалов. Для этого выделим ячейку С4 и в строке формул введём =S6, тем самым среднее первых пяти уровней относим к моменту времени $t=3$. Аналогично в ячейках С5-С19 получаем значения скользящей средней для последующих моментов времени, см. рисунок 4.6.

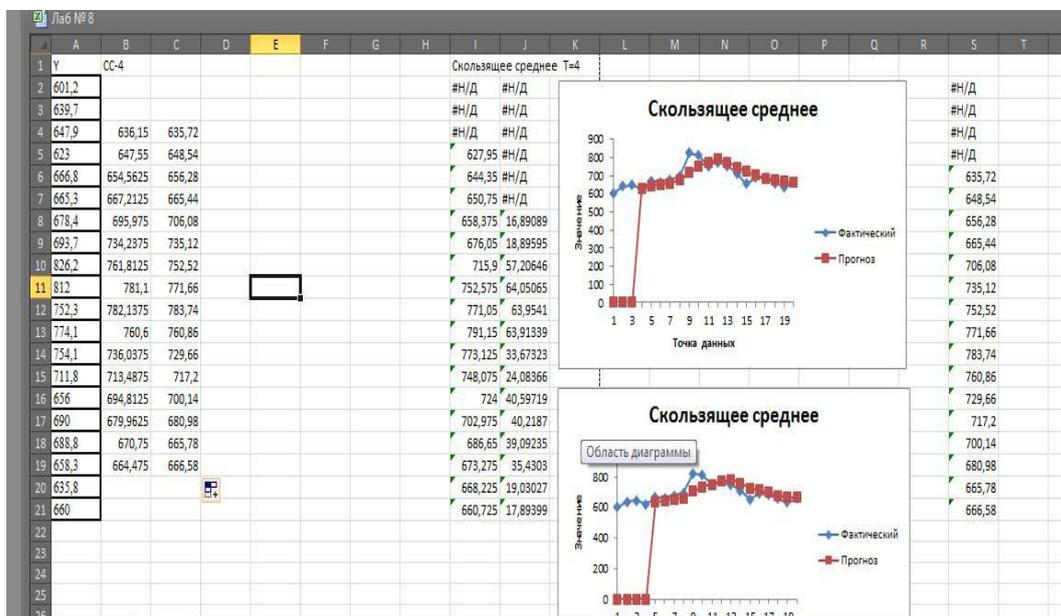


Рисунок 4.6 – Построение скользящей средней с длиной окна, равной 4 и 5

4) *Графики скользящих средних и исходного ряда.* Для их построения во вкладке «Вставка» выберем вид графика «График с маркерами» и по «Вводу данных» в окне «Выбор источника данных» укажем **A2-C21**. Средствами MS Excel изменим названия легенд рядов. На рисунке 4.7 приведены графики исходного временного ряда и скользящих средних с длиной окна скольжения 4 и 5.

5) *Выделение тенденции ряда методом взвешенной скользящей средней для длины окна сглаживания, равной 5, с аппроксимирующим полиномом второй степени.* Выделим ячейку **D4** и введём в строке формул:

$$=(-3*A2+12*A3+17*A4+12*A5-3*A6)/35$$

Получим в этой ячейке первое значение взвешенной скользящей средней. В ячейках **D5-D19** аналогичным образом получаем значения взвешенной скользящей средней при смещении окна скольжения вдоль оси времени, см. рисунок 4.8. Взвешенная скользящая средняя лучше, чем простые скользящие средние, согласуется с исходным временным рядом, показывает изменение тенденции в момент времени $t=11$ и хуже сглаживает колебания уровней временного ряда.

Графики скользящих средних не содержат колебаний и показывают изменение тенденции с возрастания на убывание в момент времени $t=11$.

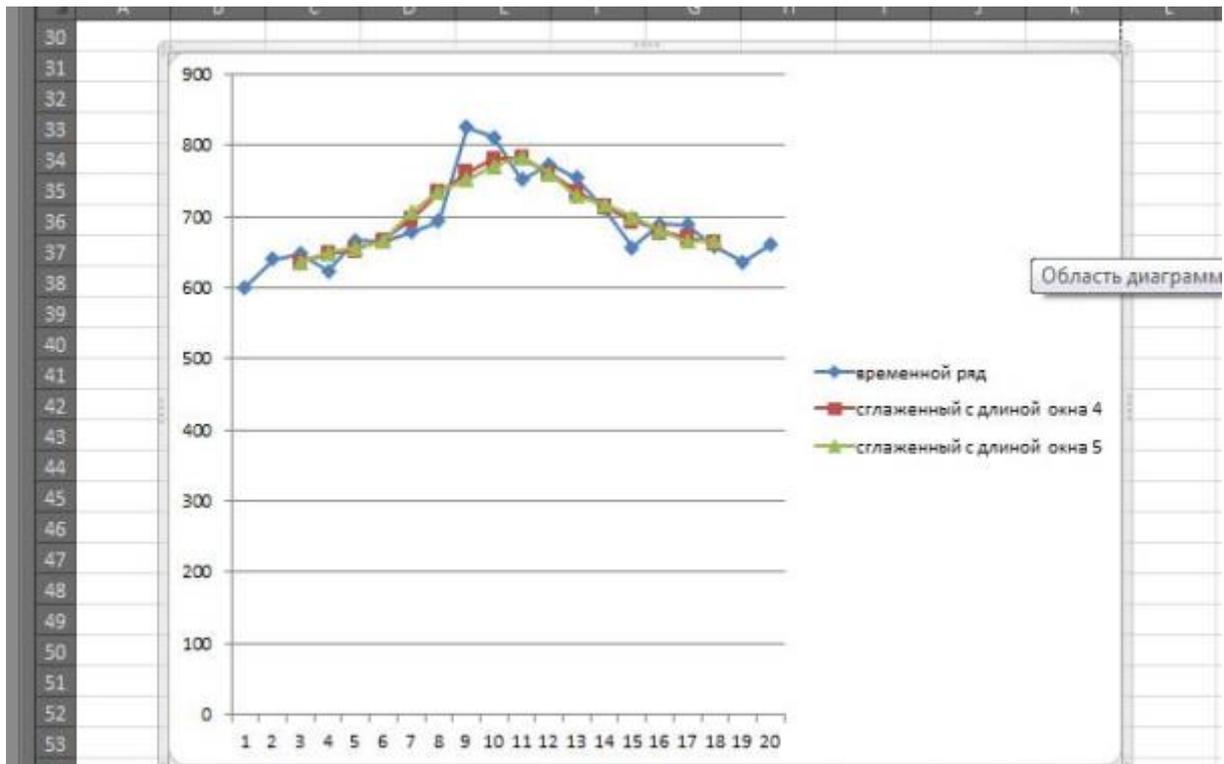


Рисунок 4.7 – Графики временного ряда и скользящих средних

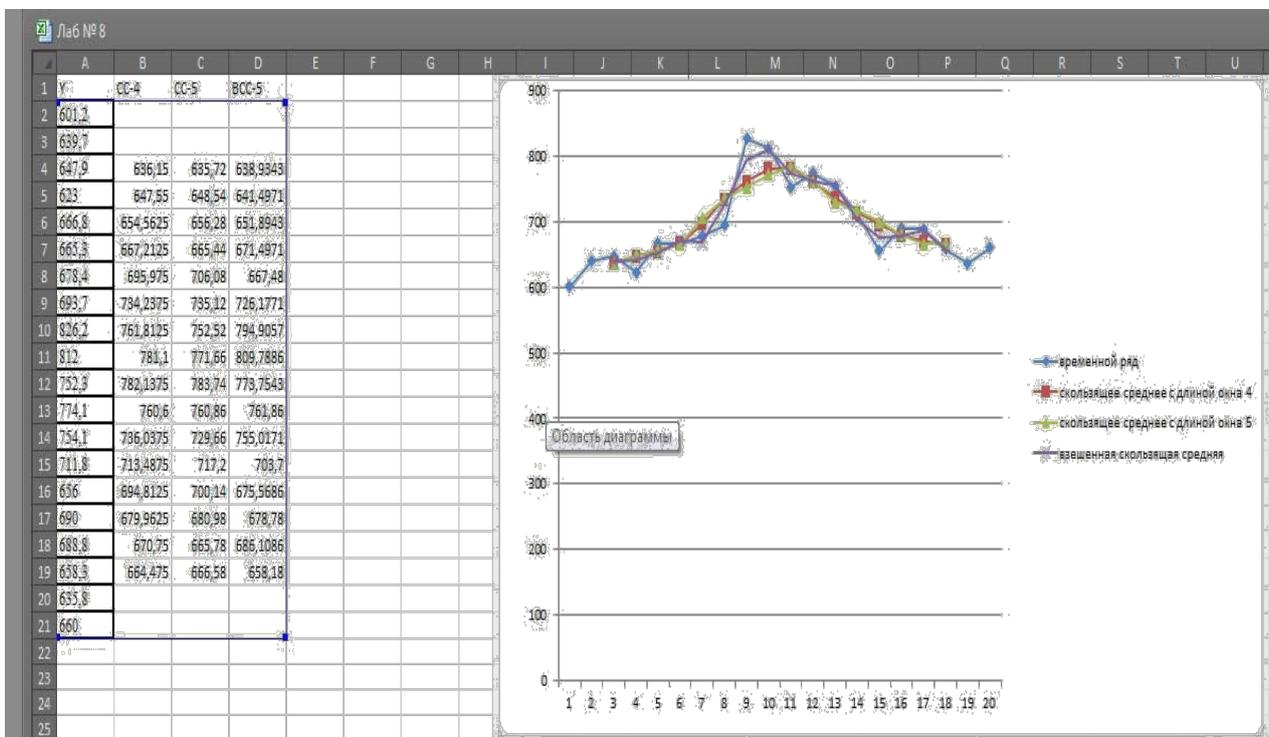


Рисунок 4.8 – Графики скользящих средних со взвешенной линией (чёрная)

4.5 Пример решения задачи 4 в MS Excel. Выделение тренда временного ряда методом экспоненциального сглаживания

Экспоненциальное сглаживание проведем для двух значений параметра сглаживания α : 0,1 и 0,3. В окне «Анализ данных» выберем инструмент анализа «Экспоненциальное сглаживание». В полях окна этого инструмента введем:

- в поле *входной интервал* укажем ячейки **A1-A21**, содержащие временной ряд;
- выберем метки в первой строке, так как в ячейке **A1** имя временного ряда;
- в поле *фактор затухания* зададим число, равное $1-\alpha$ (в примере соответственно 0,9 и 0,7);
- в поле *выходной интервал* укажем, например, ячейку **F2** для $\alpha=0,1$ и **G2** для $\alpha=0,3$;
- выберем вывод графика.

При необходимости можно выбрать также «Стандартные погрешности», в этом случае наряду со сглаженным рядом выводятся и средние квадратические отклонения уровней ряда от сглаженных значений:

$$\bullet e_t = \sqrt{\frac{\sum_{i=1}^3 (y_{t-i} - \bar{y}_{t-i})^2}{3}} \quad \bullet (4.5)$$

где \bar{y}_{t-i} – экспоненциальное скользящее среднее для момента времени $t-i$.

Результаты экспоненциального сглаживания приведены на рисунке 4.9. Прогноз уровней временного ряда по формуле экспоненциального сглаживания $S_{t+1} = S_t + \alpha(y_t - S_t)$ для последующих моментов времени $t=21$ и $t=22$ рассчитаем в ячейках F22-F23 и G22-G23, см. рисунок 4.9.

Общее заключение. Проведенное сглаживание уровней временного ряда квартальных объемов выпуска продукции говорит о тенденции возрастания в начальном периоде и тенденции убывания в конечном периоде. Изменение тенденции происходит в момент времени $t=10$. Графики сглаженных уровней показывают, что простые скользящие средние с окном скольжения 4 и 5 хорошо отражают тенденцию возрастания квартальных объемов выпуска продукции в начальном периоде и тенденцию их убывания в конечном периоде, но сдвигают на шаг вправо точку изменения тенденции.

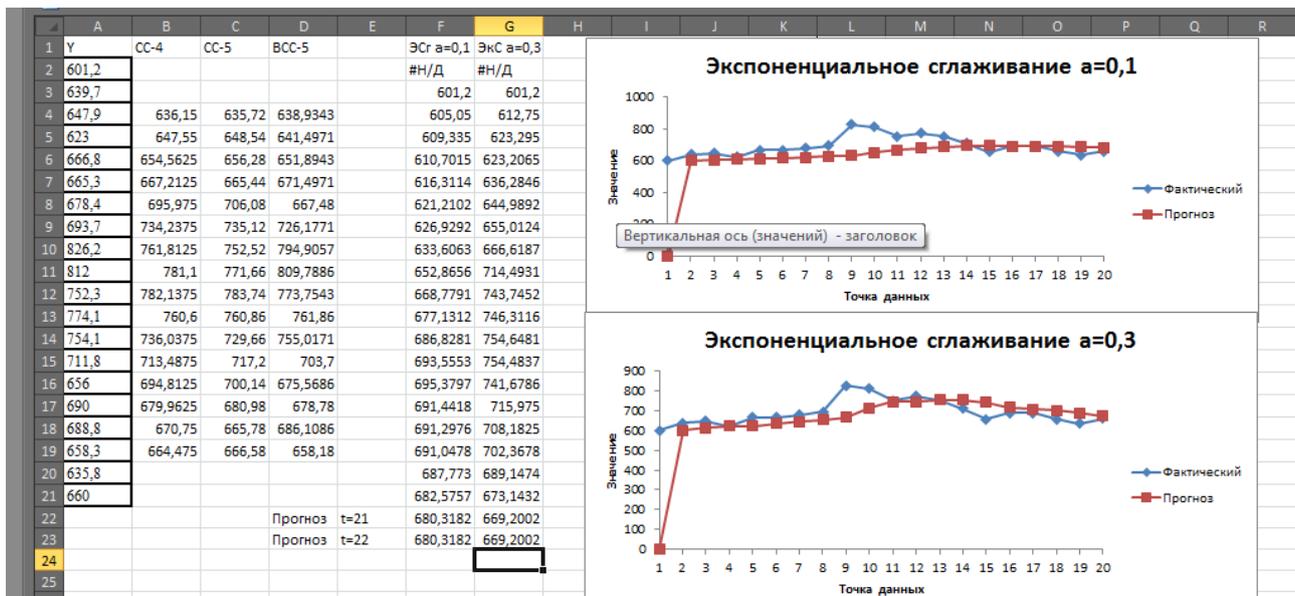


Рисунок 4.9 – Графики экспоненциального сглаживания

Взвешенная скользящая средняя лучше согласуется с исходным временным рядом, также сдвигает вправо на шаг точку изменения тенденции и хуже, чем простые скользящие средние, сглаживает колебания уровней временного ряда.

Экспоненциальное сглаживание с параметром сглаживания $\alpha=0,1$ занижает тенденцию возрастания в начальном периоде и тенденцию убывания в конечном периоде, а также значительно сдвигает вправо точку изменения характера тенденции. Экспоненциальное сглаживание с параметром сглаживания $\alpha=0,3$ также занижает тенденцию возрастания в начальном периоде и значительно сдвигает вправо точку изменения характера тенденции. В обоих случаях экспоненциальное сглаживание хорошо сглаживает колебания уровней временного ряда.

Из рассмотренных методов сглаживания и фильтрации тенденцию временного ряда квартальных объемов выпуска продукции лучше представляют простые скользящие средние с окном скользящего 4 или 5.

Колебания уровней относительно тенденции в начальный период, от $t=1$ до $t=8$, имеют сезонный характер с возрастающей амплитудой колебаний. При $t > 8$ характер колебаний изменяется, поэтому нельзя говорить о наличии сезонных колебаний во всем временном ряде.

4.6 Пример решения задачи 5 в MS Excel. Прогнозирование на основе авторегрессий 1-го и 2-го порядков

Авторегрессия первого порядка, сокращённое обозначение AR I:

$$Y_i = a_0 + a_1 * Y_{i-1} + \varepsilon_i.$$

Как видно из этой формулы, линейная модель авторегрессии 1-го порядка состоит только из одного влияющего фактора, а именно из Y_{i-1} , то есть, изучается зависимость (наиболее тесная) только от того, каким был итоговый показатель на один шаг назад.

Рассмотрим построение модели с помощью "Пакета анализа" в Excel на примере некоего ряда динамики валового регионального продукта (ВРП) с 2004 по 2012 гг., рисунок 4.10.

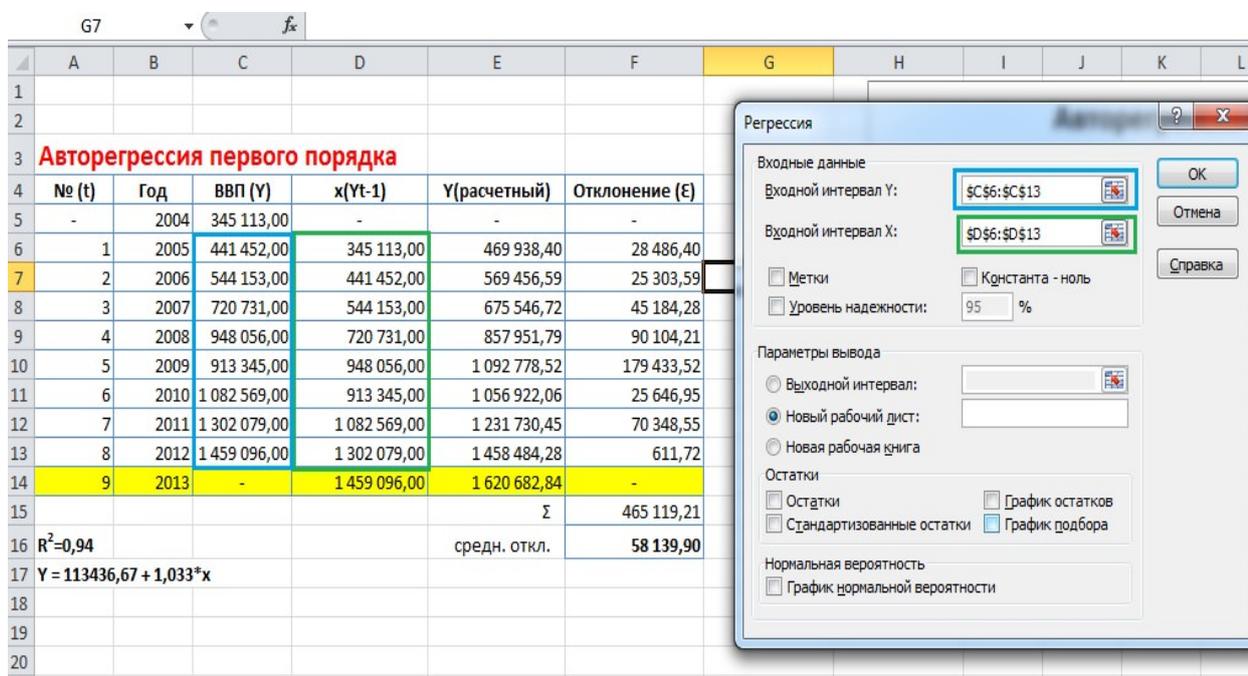


Рисунок 4.10 – Построение авторегрессии 1-го порядка

Исходные данные предварительно нужно подготовить, а именно прописать в четвёртой графе Y_{i-1} – это и будет фактором. Таким образом, наша совокупность, то есть динамический ряд, который будет использоваться для регрессии, сократился на одну позицию, обратите внимание на графы 1 и 2, с 2005 по 2012 гг. Далее в Excel активируем вкладку "Данные" и нажимаем "Анализ данных", указываем диапазон исходных данных (см. пример на скриншоте) и жмем кнопку ОК. Если по указанному пути нет кнопки "Анализ данных", то «Пакет анализа» нужно активировать, это несложно (инструкцию см. в интернете).

| | A | B | C | D | E | F | G | H | I |
|----|---------------------------------|---------------------|---------------------------|---------------------|-------------------|---------------------|--------------------|---------------------|---------------------|
| 1 | Вывод ИТОГОВ | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | <i>Регрессионная статистика</i> | | | | | | | | |
| 4 | Множественный R | 0,971302904 | | | | | | | |
| 5 | R-квадрат | 0,943429332 | | | | | | | |
| 6 | Нормированный R-квадрат | 0,934000887 | | | | | | | |
| 7 | Стандартная ошибка | 90751,79091 | | | | | | | |
| 8 | Наблюдения | 8 | | | | | | | |
| 9 | | | | | | | | | |
| 10 | <i>Дисперсионный анализ</i> | | | | | | | | |
| 11 | | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Значимость F</i> | | | |
| 12 | Регрессия | 1 | 8,241E+11 | 8,241E+11 | 100,0620322 | 5,78175E-05 | | | |
| 13 | Остаток | 6 | 49415325316 | 8235887553 | | | | | |
| 14 | Итого | 7 | 8,73515E+11 | | | | | | |
| 15 | | | | | | | | | |
| 16 | | <i>Коэффициенты</i> | <i>Стандартная ошибка</i> | <i>t-статистика</i> | <i>P-Значение</i> | <i>Нижние 95%</i> | <i>верхние 95%</i> | <i>Нижние 95,0%</i> | <i>Срхние 95,0%</i> |
| 17 | Y-пересечение | 113436,6764 | 87378,79114 | 1,298217507 | 0,241878233 | -100371,5232 | 327244,9 | -100372 | 327244,9 |
| 18 | Переменная X 1 | 1,032789147 | 0,103246897 | 10,00310113 | 5,78175E-05 | 0,780153093 | 1,285425 | 0,780153 | 1,285425 |
| 19 | | | | | | | | | |
| 20 | | | | | | | | | |
| 21 | Y = 113436,67 + 1,033*X | | | | | | | | |
| 22 | | | | | | | | | |

Рисунок 4.11 – Результаты расчётов для авторегрессии 1-го порядка

Результаты расчетов – см. рисунок 4.11 – «Пакет анализа» выдает нам на новом листе (если в настройках не было указано иначе), первоочередные по важности ячейки выделены тёмным. Из этих данных собираем авторегрессию с числовыми параметрами:

$$Y = 113436,67 + 1,033 * X$$

или

$$Y_i = 113436,67 + 1,033 * Y_{i-1}$$

Возвращаемся в нашу табличку с исходными данными и подставляем полученное уравнение в столбец в качестве формулы, таким образом, получаем расчетные значения по модели, что мы и сделали в столбце Y(расчетный), протягивая формулу на период ниже, получаем прогноз (в табличке строка выделена тёмным).

Соотношение между фактическими данными и модельными наглядно представлено на рисунке 4.12.

Авторегрессия второго порядка, сокращённое обозначение AR II:

$$Y_i = a_0 + a_1 * Y_{i-1} + a_2 * Y_{i-2} + \varepsilon_i$$

Модель авторегрессии 2-го порядка отличается от первой тем, что она включает в себя еще один влияющий фактор Y_{i-2} . Это может повысить значимость регрессии и привести к более точному прогнозу.



Рисунок 4.12 – Фактические и расчётные значения показателя Y

Все расчеты проводятся аналогично описанию в авторегрессии 1-го порядка с той лишь разницей, что теперь мы имеем два столбца с влияющими факторами. Заметим, что на этот раз используемый диапазон динамического ряда исходных данных сократится не на один период - лаг, а на два (см. графу t). На рисунке 4.13 показаны результаты расчёта для авторегрессии 2-го порядка.

| Авторегрессия второго порядка | | | | | | | |
|-------------------------------|------|--------------|----------------|----------------|-----------------------------------|----------------|---|
| № (t) | Год | ВВП (Y) | $x_1(Y_{t-1})$ | $x_2(Y_{t-2})$ | Y(расчетный) | Отклонение (E) | |
| - | 2004 | 345 113,00 | - | - | - | - | - |
| - | 2005 | 441 452,00 | 345 113,00 | - | - | - | - |
| 1 | 2006 | 544 153,00 | 441 452,00 | 345 113,00 | 581 443,40 | 37 290,40 | |
| 2 | 2007 | 720 731,00 | 544 153,00 | 441 452,00 | 686 627,40 | 34 103,60 | |
| 3 | 2008 | 948 056,00 | 720 731,00 | 544 153,00 | 847 334,19 | 100 721,81 | |
| 4 | 2009 | 913 345,00 | 948 056,00 | 720 731,00 | 1 068 422,45 | 155 077,45 | |
| 5 | 2010 | 1 082 569,00 | 913 345,00 | 948 056,00 | 1 116 035,69 | 33 466,69 | |
| 6 | 2011 | 1 302 079,00 | 1 082 569,00 | 913 345,00 | 1 227 446,34 | 74 632,66 | |
| 7 | 2012 | 1 459 096,00 | 1 302 079,00 | 1 082 569,00 | 1 440 523,26 | 18 572,74 | |
| 8 | 2013 | - | 1 459 096,00 | 1 302 079,00 | 1 624 446,77 | - | |
| | | | | | Σ | 453 865,34 | |
| | | | | | средн. откл. | 64 837,91 | |
| | | | | | $R^2=0,927$ | | |
| | | | | | $Y=151395,987+0,724*x_1+0,32*x_2$ | | |

Рисунок 4.13 – Результаты расчёта для авторегрессии 2-го порядка

Полученная модель $Y=151395,987+0,724*x_1+0,32*x_2$ или – в других обозначениях $Y_i=151395,987+0,724*Y_{i-1}+0,32*Y_{i-2}$ имеет:

- показатель детерминации R^2 ниже, чем у модели 1-го порядка (0,927 против 0,94);
- среднее отклонение у модели 2-го порядка больше (64 837,91 против 58 139,90).

На рисунке 4.14 показано соотношение фактических и модельных значений показателя Y .

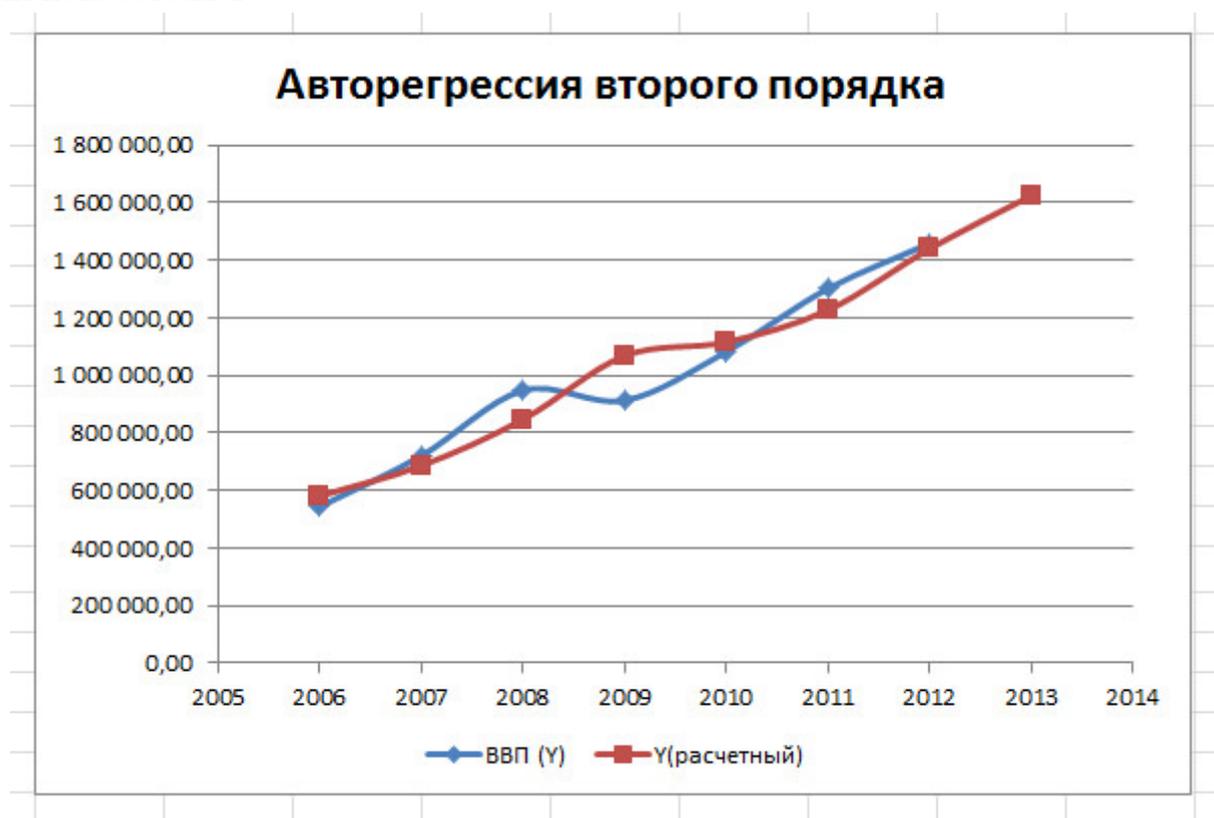


Рисунок 4.14 – Фактические и расчётные значения показателя Y

Вывод: модель 1-го порядка более точная. Это может быть связано с тем, что диапазон исходных данных мал, его сокращение ещё на один период имело значительные последствия для точности модели.

5 Вопросы для контрольной (самостоятельной) работы и зачёта

Краткие ответы на приводимые ниже вопросы – почти на все – имеются в разделе 2, там же даются ссылки на источники, где можно найти более обстоятельные ответы на все вопросы.

Тема 1

- 1) В чём состоит общая тенденция XX века – математизация знаний, привести примеры.
- 2) Что является предметом и методом эконометрики?
- 3) Как вам представляется симбиоз эконометрики и искусственного интеллекта?
- 4) В чём нестандартность эконометрической модели в первом примере?
- 5) В чём нестандартность эконометрической модели во втором примере?
- 6) Какие приёмы исследования функции вы знаете?
- 7) Какие два значения имеет термин χ^2 (хи-квадрат)?
- 8) Что означают два числа в записи стандартного нормального закона распределения: НЗР $N(0, 1)$, нарисуйте его.
- 9) Как можно назвать значения X_1 и X_2 на оси абсцисс, см. рисунок 2.2?
- 10) У компании 7 магазинов, на 1 января их годовая выручка составила млн руб. (12, 13, ..., 8), какая это выборка?

Тема 2

- 11) В чём отличие переменных $\beta_0, \beta_1, \varepsilon_7$ от соответствующих им переменных b_0, b_1, e_7 ? Создайте их графические образы на рисунке 2.3.
- 12) Параметры регрессии b_0, b_1 называют оптимальными (наилучшими), какой смысл вкладывают в это?
- 13) Если судить по рисунку 2.3, то с какой ситуацией мы имеем дело: гомо- или гетероскедастичности?
- 14) Создайте на рисунке 2.3 графический образ тех самых квадратов, о которых идёт речь в методе наименьших квадратов.
- 15) Покажите на рисунке 2.3 отрезки $\hat{y}_2, \hat{y}_3, \hat{y}_6, y_2, y_3, y_6$.
- 16) Что означают жирно выделенные линии на рисунке 2.4?
- 17) Что можно сказать о доверительном интервале на основе (2.13)?
- 18) Дайте определение каждого термина из названия модели: *классическая, нормальная, линейная, модель, множественная, регрессия*.
- 19) Что является неизвестным в системе уравнений, записанных в матричной форме: $X'X \cdot b = X'Y$?
- 20) Что за выражение (переменная) σ^2 в (2.20)?
- 21) При расчёте дисперсии СВ b_1 какой диагональный элемент обратной матрицы нужно выбрать, см. (2.23)?
- 22) Если нулевая гипотеза $H_0: \beta_j = 0$ отвергается, то какой из этого следует вывод для парной регрессии?

- 23) Как будет выглядеть графический образ доверительного интервала для β_1 на рисунке 2.3?

Тема 3

- 24) Сколько всего регрессий нужно построить, чтобы реализовать алгоритм отбора, показанный на рисунке 2.6?
- 25) Составляется таблица с графами: ФИО, Стаж, Пол, Зарплата. Как будет выглядеть графа Пол с булевской переменной z ?
- 26) Верно ли утверждение: для атрибутивной переменной с 4-мя уровнями нужно ввести 3 булевских переменных?
- 27) Что означает термин: функция, линейная по параметрам?
- 28) Как выглядит график функции эластичности для линейной парной регрессии (см. таблицу 2.7, следует применить правило Лопиталья)?
- 29) В каких единицах можно измерять фактор L – труд?
- 30) Составьте числовой пример для выражения: $\sum \varepsilon = \sigma^2 E_n$.
- 31) Можно ли применять обычный МНК в случае гетероскедастичности?
- 32) В чём суть теста ранговой корреляции Спирмена? Можно ли его применить в случае, показанном на рисунке 2.7.

Тема 4

- 33) Из каких компонент может состоять временной ряд?
- 34) Какой ряд называется стационарным?
- 35) Есть ли связь между терминами «парный коэффициент корреляции» и «корреляционная функция»?
- 36) Как время может выступать в роли факторной переменной?
- 37) Верно ли утверждение: моделью временного ряда является его тренд?
- 38) Как называется модель, у которой факторы – лаговые результирующие переменные?
- 39) Как называется модель, у которой факторы – лаговые переменные отклонений?

СПИСОК РЕКОМЕНДУЕМЫХ ИСТОЧНИКОВ

Основные

- 1) Кремер, Н.Ш. Эконометрика: учебник / Н.Ш. Кремер, Б.А. Путко. – Москва: ЮНИТИ-ДАНА, 2012. – 328 с. (ЭБС «Электронная библиотека онлайн»).
- 2) Настин, Ю.Я. Эконометрика: учебно-методическое пособие по практическим занятиям для студентов, обучающихся по направлениям «Экономика» и «Торговое дело» / Ю.Я. Настин, Д.К. Тылик. - Калининград: Изд-во ФГБОУ ВО «КГТУ», 2017. – 86 с.
- 3) Настин, Ю.Я. Эконометрика: учеб.-методич. пособие по освоению дисциплины и выполнению контрольной работы для студентов заочной формы обучения, обучающихся в бакалавриате по направлению подгот. 38.03.01 «Экономика» / Ю.Я. Настин. – Калининград: Изд-во ФГБОУ ВО «КГТУ», 2019. – 71 с.
- 4) Карлов, А.М. Теория вероятностей и математическая статистика для экономистов: учебное пособие / А.М. Карлов. – Москва: КНОРУС, 2015. – 264 с.

Дополнительные

- 5) Мнацаканян, А.Г. Методические указания по оформлению учебных текстовых работ (рефератов, контрольных, курсовых, выпускных квалификационных) / А.Г. Мнацаканян, Ю.Я. Настин, Э.С. Круглова. – Калининград: Изд-во КГТУ, 2017. – 22 с.
- 6) Балдин, К.В. Эконометрика: учеб пособие / К.В. Балдин, О.Ф. Быстров, М.М. Соколов. – Москва: ЮНИТИ-ДАНА, 2015. – 254 с. (ЭБС «Электронная библиотека онлайн»).
- 7) Дубова, Т.А. Статистические методы прогнозирования в экономике: учебное пособие, практикум, тесты, программа курса, руководство по изучению дисциплины / Т.А. Дубова, М.Ю. Архипова. – Москва: Изд-во МЭСИ, 2004. - 136
- 8) Орлов, А.И. Эконометрика: учебник / А.И. Орлов. – Москва: «Экзамен», 2002. – 413 с.
- 9) Абдуллин, Р.З. Эконометрика в MS EXCEL: практикум / Р.З. Абдуллин, В.Р. Абдуллин. - Иркутск, Изд-во БГУ, 2016. – 135 с. Режим доступа: <http://lib-catalog.isea.ru>
- 10) Гудвин, А. Авторегрессия – моделирование и прогнозирование в Excel, http://archie-goodwin.net/load/specializirovannye_blogi/ms_office/avtoregressija_modelirovanie_i_prognozirovanie_v_excel/28-1-0-422
- 11) Мхитарян, В.С. Эконометрика: учебно-практическое пособие / В.С. Мхитарян, М.Ю. Архипова, В.П. Сиротин. – Москва: АНО «Евразийский открытый университет», 2012. – 221 с. (ЭБС «Электронная библиотека онлайн»).

ПРИЛОЖЕНИЕ А

Исходные данные для контрольных (самостоятельных) работ

А.1 Исходные данные для задач 1 и 2

Таблица А.1 - Данные для переменных X1 и Y, первая подходящая гласная буква **А** или **Е** полного имени

| X1_i | Y_i – выручка (100 тыс. руб./сутки) | | | | | | | | | | | | | |
|-----------------------|--|----|----|----|----|----|----|----|----|----|----|-----------|----|----|
| А, Е | Ъ | А | Б | В | Г | Д | Е | К | Л | М | Н | О | П | Р |
| 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 |
| 1 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 |
| 2 | 4 | 4 | 4 | 5 | 5 | 6 | 6 | 4 | 4 | 5 | 5 | 6 | 6 | 4 |
| 3 | 5 | 6 | 7 | 8 | 6 | 7 | 7 | 7 | 7 | 4 | 5 | 6 | 8 | 7 |
| 4 | 5 | 6+ | 7+ | 8+ | 7+ | 8+ | 8+ | 9 | 9 | 7 | 8 | 9 | 10 | 9 |
| | | Г | Г | Г | Г | Г | Г | | | | | | | |
| 5 | 7 | 8 | 9 | 7 | 8 | 9 | 7 | 8+ | 9+ | 7+ | 8+ | 9+ | 11 | 12 |
| | | | | | | | | Г | Г | Г | Г | Г | +Г | +Г |
| 8 | 14 | 16 | 18 | 20 | 14 | 16 | 18 | 19 | 20 | 16 | 18 | 20 | 22 | 23 |

Таблица А.2 - Данные для переменных X1 и Y, первая подходящая гласная буква **И** или **О** полного имени

| X1_i | Y_i – выручка (100 тыс. руб./сутки) | | | | | | | | | | | | | |
|-----------------------|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| И, О | А | Б | В | Г | Д | Е | К | Л | М | Н | О | П | Р | |
| 1 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | |
| 2 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | |
| 4 | 4 | 4 | 4 | 5 | 5 | 6 | 4 | 4 | 5 | 5 | 6 | 6 | 4 | |
| 4 | 5 | 5 | 6 | 6 | 6 | 6 | 7 | 7 | 6 | 5 | 6 | 7 | 7 | |
| 5 | 7 | 8 | 9 | 8 | 8 | 9 | 8+Г | 7+Г | 6+Г | 7+Г | 8+Г | 7+Г | 6+Г | |
| 6 | 7+Г | 8+Г | 9+Г | 9+Г | 8+Г | 9+Г | 10 | 9 | 11 | 9 | 10 | 11 | 12 | |
| 8 | 14 | 16 | 18 | 20 | 14 | 16 | 19 | 20 | 16 | 18 | 20 | 22 | 23 | |

Таблица А.3 - Данные для переменных X1 и Y, первая подходящая гласная буква **У** или **Я** полного имени

| X1_i | Y_i – выручка (100 тыс. руб./сутки) | | | | | | | | | | | | | |
|-----------------------|--|-----|-----|-----|-----|----------|-----|-----|-----|-----|-----|-----|-----|--|
| У, Я | А | Б | В | Г | Д | Е | К | Л | М | Н | О | П | Р | |
| 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | |
| 4 | 5 | 5 | 5 | 5 | 5 | 6 | 5 | 5 | 5 | 5 | 6 | 6 | 5 | |
| 5 | 7 | 7 | 7 | 8 | 6 | 6 | 7 | 7 | 7 | 5 | 6 | 7 | 7 | |
| 7 | 5+Г | 6+Г | 7+Г | 8+Г | 9 | 9 | 9 | 9 | 6 | 7+Г | 8+Г | 5+Г | 6+Г | |
| 7 | 7 | 8 | 9 | 7 | 8+Г | 9+Г | 8+Г | 9+Г | 7+Г | 9 | 10 | 11 | 12 | |
| 9 | 14 | 16 | 18 | 20 | 14 | 16 | 19 | 20 | 16 | 18 | 20 | 22 | 23 | |

Таблица А.4 - Исходные данные для переменной X_2 , выбор по первой подходящей гласной букве полного имени

| X_2 – площадь паркинга (шт. автомашин*10) | | | | | | |
|---|----------|----------|----------|----------|----------|----------|
| Ъ | А | Е | И | О | У | Я |
| 1 | 2 | 1 | 3 | 1 | 1 | 1 |
| 2 | 3 | 4 | 2 | 3 | 1 | 2 |
| 2 | 4 | 5 | 4 | 4 | 3 | 3 |
| 3 | 5 | 6 | 5 | 5 | 4 | 6 |
| 3 | 7 | 5 | 5 | 7 | 4 | 5 |
| 4 | 8 | 7 | 7 | 8 | 7 | 6 |
| 6 | 6+г | 6+Г | 6+г | 6+Г | 6+Г | 6+Г |

А.2 Исходные данные для задач 3 и 4

Таблица А.5 – Поквартальный объём выпуска морепродуктов в сопоставимых ценах (млн руб./квартал)

| Квартал | Год | | | | |
|---------|-----|-------|-------|-------|-------|
| | 1* | 2 | 3 | 4 | 5* |
| I | ? | 660,0 | 830,0 | 750,0 | 750,0 |
| II | ? | 670,0 | 810,0 | 710,0 | ? |
| III | ? | 680,0 | 750,0 | 660,0 | ? |
| IV | ? | 690,0 | 780,0 | 690,0 | ? |

*правило заполнения пустых затемнённых клеток индивидуальными данными см в подразделе 4.1.

А.3 Исходные данные для задачи 5

Таблица А.6 – Валовой региональный продукт по годам за 10 лет, млрд руб.

| Год | ВРП | Год | ВРП | Год | ВРП |
|-----|-----|-----|-------|-----|-------|
| 1 | ? | 5 | 660,0 | 9 | 830,0 |
| 2 | ? | 6 | 670,0 | 10 | 810,0 |
| 3 | ? | 7 | 680,0 | - | - |
| 4 | ? | 8 | 690,0 | - | - |

*правило заполнения пустых затемнённых клеток индивидуальными данными см. в подразделе 4.1.

ПРИЛОЖЕНИЕ Б

Пример страницы СОДЕРЖАНИЕ для контрольных (самостоятельных) работ

СОДЕРЖАНИЕ

| | | |
|-----|---|----|
| 1 | Ответы на вопросы | 3 |
| 1.1 | Вопрос 11. <Текст вопроса>..... | 3 |
| 1.2 | Вопрос 31. <Текст вопроса>..... | 3 |
| 2 | Задача 1. Моделирование магазина на основе линейной троичной регрессии – решение вручную..... | 4 |
| 3 | Задача 2. Моделирование магазина на основе линейной троичной регрессии – решение в MS Excel..... | 9 |
| 4 | Задача 3. Выделение тренда временного ряда. Решение методом скользящей средней в MS Excel | 12 |
| 5 | Задача 4. Выделение тренда временного ряда. Решение методом экспоненциального сглаживания в MS Excel..... | 15 |
| 6 | Задача 5. Прогнозирование на основе авторегрессий 1-го и 2-го порядков в MS Excel..... | 18 |
| | СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ..... | 21 |

Примечания:

- 1) Страницы проставлены ориентировочно
- 2) Границы после заполнения таблицы убрать
- 3) Можно эту страницу формировать иначе – средствами редактора

ПРИЛОЖЕНИЕ В

Построение графика линейной парной регрессии в Word через Excel

Работа в Excel

- 1) Создать таблицу с исходными данными в левом верхнем углу от A1 до B8 (слева X1, справа Y).

| | |
|----|----|
| 1 | 2 |
| 2 | 3 |
| 2 | 4 |
| 3 | 5 |
| 4 | 5 |
| 5 | 7 |
| 8 | 14 |
| 11 | |

Значение 11 в нижней строке мы вставили, чтобы увеличить нумерацию по осям.

- 2) Выделить таблицу.
- 3) Нажать «Вставка».
- 4) Нажать «Точечная», затем 1-ю иконку (точечная с маркерами), получим рисунок с полем корреляции.
- 5) Нажать сверху иконку 3-ю слева: сетка с прямой линией, получим искомый график с регрессией – прямой линией и точками.
- 6) Курсор в белое поле рисунка, правой клавишей вызвать меню.
- 7) Вырезать рисунок.

Работа в Word

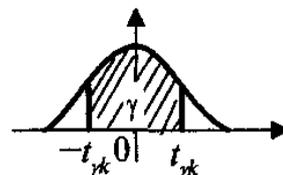
- 8) Вставить рисунок в нужное место ворд-файла.
- 9) Вписать названия координат, легенду лучше удалить.
- 10) Можно поэкспериментировать: раздвинуть границы, применить заливку и цвета и т.п.

ПРИЛОЖЕНИЕ Г

Значения критерия Стьюдента

Таблица II

Значения $t_{\gamma,k}$ -критерия Стьюдента



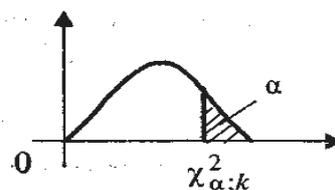
| Число степеней свободы k | Вероятность γ | | | | | | | | | | | |
|----------------------------|----------------------|------|------|------|------|------|------|------|------|-------|-------|-------|
| | 0,1 | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 | 0,9 | 0,95 | 0,98 | 0,99 |
| 1 | 0,16 | 0,32 | 0,51 | 0,73 | 1,00 | 1,38 | 1,96 | 3,08 | 6,31 | 12,71 | 31,82 | 63,66 |
| 2 | 14 | 29 | 44 | 62 | 0,82 | 06 | 34 | 1,89 | 2,92 | 4,30 | 6,96 | 9,92 |
| 3 | 14 | 28 | 42 | 58 | 76 | 0,98 | 25 | 64 | 35 | 3,18 | 4,54 | 5,84 |
| 4 | 13 | 27 | 41 | 57 | 74 | 94 | 19 | 53 | 13 | 2,78 | 3,75 | 4,60 |
| 5 | 13 | 27 | 41 | 56 | 73 | 92 | 16 | 48 | 01 | 57 | 36 | 03 |
| 6 | 0,13 | 0,26 | 0,40 | 0,55 | 1,72 | 1,91 | 1,13 | 1,44 | 1,94 | 2,45 | 3,14 | 3,71 |
| 7 | 13 | 26 | 40 | 55 | 71 | 90 | 12 | 41 | 89 | 36 | 00 | 50 |
| 8 | 13 | 26 | 40 | 55 | 70 | 89 | 11 | 40 | 86 | 31 | 2,90 | 35 |
| 9 | 13 | 26 | 40 | 54 | 70 | 88 | 10 | 38 | 83 | 26 | 82 | 25 |
| 10 | 13 | 26 | 40 | 54 | 70 | 88 | 09 | 37 | 81 | 23 | 76 | 17 |
| 11 | 0,13 | 0,26 | 0,40 | 0,54 | 0,70 | 0,88 | 1,09 | 1,36 | 1,80 | 2,20 | 2,72 | 3,11 |
| 12 | 13 | 26 | 39 | 54 | 69 | 87 | 08 | 36 | 78 | 18 | 68 | 05 |
| 13 | 13 | 26 | 39 | 54 | 69 | 87 | 08 | 35 | 77 | 16 | 65 | 01 |
| 14 | 13 | 26 | 39 | 54 | 69 | 87 | 08 | 34 | 76 | 14 | 62 | 2,98 |
| 15 | 13 | 26 | 39 | 54 | 69 | 87 | 07 | 34 | 75 | 13 | 60 | 95 |
| 16 | 0,13 | 0,26 | 0,39 | 0,53 | 0,69 | 0,86 | 1,07 | 1,34 | 1,75 | 2,12 | 2,58 | 2,92 |
| 17 | 13 | 26 | 39 | 53 | 69 | 86 | 07 | 33 | 74 | 11 | 57 | 90 |
| 18 | 13 | 26 | 39 | 53 | 69 | 86 | 07 | 33 | 73 | 10 | 55 | 88 |
| 19 | 13 | 26 | 39 | 53 | 69 | 86 | 07 | 33 | 73 | 09 | 54 | 86 |
| 20 | 13 | 26 | 39 | 53 | 69 | 86 | 06 | 32 | 72 | 09 | 53 | 84 |
| 21 | 0,13 | 0,26 | 0,39 | 0,53 | 0,69 | 0,86 | 1,06 | 1,32 | 1,72 | 2,08 | 2,52 | 2,83 |
| 22 | 13 | 26 | 39 | 53 | 69 | 86 | 06 | 32 | 72 | 07 | 51 | 82 |
| 23 | 13 | 26 | 39 | 53 | 68 | 86 | 06 | 32 | 71 | 07 | 50 | 81 |
| 24 | 13 | 26 | 39 | 53 | 68 | 86 | 06 | 32 | 71 | 06 | 49 | 80 |
| 25 | 13 | 26 | 39 | 53 | 68 | 86 | 06 | 32 | 71 | 06 | 48 | 79 |
| 26 | 0,13 | 0,26 | 0,39 | 0,53 | 0,68 | 0,86 | 1,06 | 1,31 | 1,71 | 2,06 | 2,48 | 2,78 |
| 27 | 13 | 26 | 39 | 53 | 68 | 85 | 06 | 31 | 70 | 05 | 47 | 77 |
| 28 | 13 | 26 | 39 | 53 | 68 | 85 | 06 | 31 | 70 | 05 | 47 | 76 |
| 29 | 13 | 26 | 39 | 53 | 68 | 85 | 05 | 31 | 70 | 04 | 46 | 76 |
| 30 | 13 | 26 | 39 | 53 | 68 | 85 | 05 | 31 | 70 | 04 | 46 | 75 |
| 40 | 0,13 | 0,25 | 0,39 | 0,53 | 0,68 | 0,85 | 1,05 | 1,30 | 1,68 | 2,02 | 2,42 | 2,70 |
| 60 | 13 | 25 | 39 | 53 | 68 | 85 | 05 | 30 | 67 | 00 | 39 | 66 |
| 120 | 0,13 | 0,25 | 0,39 | 0,53 | 0,68 | 0,84 | 1,04 | 1,29 | 1,66 | 1,98 | 2,36 | 2,62 |
| ∞ | 13 | 25 | 38 | 52 | 67 | 84 | 04 | 28 | 64 | 96 | 33 | 58 |

ПРИЛОЖЕНИЕ Д

Значения критерия Пирсона

Таблица III

Значения $\chi^2_{\alpha;k}$ критерия Пирсона



| Число степеней свободы k | Вероятность α | | | | | | | | | | | | |
|----------------------------|----------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 0,99 | 0,98 | 0,95 | 0,90 | 0,80 | 0,70 | 0,50 | 0,30 | 0,20 | 0,10 | 0,05 | 0,02 | 0,01 |
| 1 | 0,00 | 0,00 | 0,00 | 0,02 | 0,06 | 0,15 | 0,45 | 1,07 | 1,64 | 2,71 | 3,84 | 5,41 | 6,64 |
| 2 | 0,02 | 0,04 | 0,10 | 0,21 | 0,45 | 0,71 | 1,39 | 2,41 | 3,22 | 4,60 | 5,99 | 7,82 | 9,21 |
| 3 | 0,11 | 0,18 | 0,35 | 0,58 | 1,00 | 1,42 | 2,37 | 3,66 | 4,64 | 6,25 | 7,82 | 9,84 | 11,3 |
| 4 | 0,30 | 0,43 | 0,71 | 1,06 | 1,65 | 2,20 | 3,36 | 4,88 | 5,99 | 7,78 | 9,49 | 11,7 | 13,3 |
| 5 | 0,55 | 0,75 | 1,14 | 1,61 | 2,34 | 3,00 | 4,35 | 6,06 | 7,29 | 9,24 | 11,1 | 13,4 | 15,1 |
| 6 | 0,87 | 1,13 | 1,63 | 2,20 | 3,07 | 3,83 | 5,35 | 7,23 | 8,56 | 10,6 | 12,6 | 15,0 | 16,8 |
| 7 | 1,24 | 1,56 | 2,17 | 2,83 | 3,82 | 4,67 | 6,35 | 8,38 | 9,80 | 12,0 | 14,1 | 16,6 | 18,5 |
| 8 | 1,65 | 2,03 | 2,73 | 3,49 | 4,59 | 5,53 | 7,34 | 9,52 | 11,0 | 13,4 | 15,5 | 18,2 | 20,1 |
| 9 | 2,09 | 2,53 | 3,32 | 4,17 | 5,38 | 6,39 | 8,34 | 10,7 | 12,2 | 14,7 | 16,9 | 19,7 | 21,7 |
| 10 | 2,56 | 3,06 | 3,94 | 4,86 | 6,18 | 7,27 | 9,34 | 11,8 | 13,4 | 16,0 | 18,3 | 21,2 | 23,2 |
| 11 | 3,05 | 3,61 | 4,58 | 5,58 | 6,99 | 8,15 | 10,3 | 12,9 | 14,6 | 17,3 | 19,7 | 22,6 | 24,7 |
| 12 | 3,57 | 4,18 | 5,23 | 6,30 | 7,81 | 9,03 | 11,3 | 14,0 | 15,8 | 18,5 | 21,0 | 24,1 | 26,2 |
| 13 | 4,11 | 4,76 | 5,89 | 7,04 | 8,63 | 9,93 | 12,3 | 15,1 | 17,0 | 19,8 | 22,4 | 25,5 | 27,7 |
| 14 | 4,66 | 5,37 | 6,57 | 7,79 | 9,47 | 10,8 | 13,3 | 16,2 | 18,1 | 21,1 | 23,7 | 26,9 | 29,1 |
| 15 | 5,23 | 5,98 | 7,26 | 8,55 | 10,3 | 11,7 | 14,3 | 17,3 | 19,3 | 22,3 | 25,0 | 28,3 | 30,6 |
| 16 | 5,81 | 6,61 | 7,96 | 9,31 | 11,1 | 12,6 | 15,3 | 18,4 | 20,5 | 23,5 | 26,3 | 29,6 | 32,0 |
| 17 | 6,41 | 7,26 | 8,67 | 10,1 | 12,0 | 13,5 | 16,3 | 19,5 | 21,6 | 24,8 | 27,6 | 31,0 | 33,4 |
| 18 | 7,02 | 7,91 | 9,39 | 10,9 | 12,9 | 14,4 | 17,3 | 20,6 | 22,8 | 26,0 | 28,9 | 32,3 | 34,8 |
| 19 | 7,63 | 8,57 | 10,1 | 11,6 | 13,7 | 15,3 | 18,3 | 21,7 | 23,9 | 27,2 | 30,1 | 33,7 | 36,2 |
| 20 | 8,26 | 9,24 | 10,8 | 12,4 | 14,6 | 16,3 | 19,3 | 22,8 | 25,0 | 28,4 | 31,4 | 35,0 | 37,6 |
| 21 | 8,90 | 9,92 | 11,6 | 13,2 | 15,4 | 17,2 | 20,3 | 23,9 | 26,2 | 29,6 | 32,7 | 36,3 | 38,9 |
| 22 | 9,54 | 10,6 | 12,3 | 14,0 | 16,3 | 18,1 | 21,3 | 24,9 | 27,3 | 30,8 | 33,9 | 37,7 | 40,3 |
| 23 | 10,2 | 11,3 | 13,1 | 14,8 | 17,2 | 19,0 | 22,3 | 26,0 | 28,4 | 32,0 | 35,2 | 39,0 | 41,6 |
| 24 | 10,9 | 12,0 | 13,8 | 15,7 | 18,1 | 19,9 | 23,3 | 27,1 | 29,6 | 33,2 | 36,4 | 40,3 | 43,0 |
| 25 | 11,5 | 12,7 | 14,6 | 16,5 | 18,9 | 20,9 | 24,3 | 28,2 | 30,7 | 34,4 | 37,7 | 41,7 | 44,3 |
| 26 | 12,2 | 13,4 | 15,4 | 17,3 | 19,8 | 21,8 | 25,3 | 29,2 | 31,8 | 35,6 | 38,9 | 42,9 | 45,6 |
| 27 | 12,9 | 14,1 | 16,1 | 18,1 | 20,7 | 22,7 | 26,3 | 30,3 | 32,9 | 36,7 | 40,1 | 44,1 | 47,0 |
| 28 | 13,6 | 14,8 | 16,9 | 18,9 | 21,6 | 23,6 | 27,3 | 31,4 | 34,0 | 37,9 | 41,3 | 45,4 | 48,3 |
| 29 | 14,3 | 15,6 | 17,7 | 19,8 | 22,5 | 24,6 | 28,3 | 32,5 | 35,1 | 39,1 | 42,6 | 46,7 | 49,6 |
| 30 | 14,9 | 16,3 | 18,5 | 20,6 | 23,4 | 25,5 | 29,3 | 33,5 | 36,2 | 40,3 | 43,8 | 48,0 | 50,9 |

ПРИЛОЖЕНИЕ Е

Значения F-критерия Фишера при уровне значимости $\alpha = 0,05$

| k_1 | | | | | | | | | | |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------|
| k_2 | 1 | 2 | 3 | 4 | 5 | 6 | 8 | 12 | 24 | ∞ |
| 1 | 161,45 | 199,50 | 215,72 | 224,57 | 230,17 | 233,97 | 238,89 | 243,91 | 249,04 | 254,32 |
| 2 | 18,51 | 19,00 | 19,16 | 19,25 | 19,30 | 19,33 | 19,37 | 19,41 | 19,45 | 19,50 |
| 3 | 10,13 | 9,55 | 9,28 | 9,12 | 9,01 | 8,94 | 8,84 | 8,74 | 8,64 | 8,53 |
| 4 | 7,71 | 6,94 | 6,59 | 6,39 | 6,26 | 6,16 | 6,04 | 5,91 | 5,77 | 5,63 |
| 5 | 6,61 | 5,79 | 5,41 | 5,19 | 5,05 | 4,95 | 4,82 | 4,68 | 4,53 | 4,36 |
| 6 | 5,99 | 5,14 | 4,76 | 4,53 | 4,39 | 4,28 | 4,15 | 4,00 | 3,84 | 3,67 |
| 7 | 5,59 | 4,74 | 4,35 | 4,12 | 3,97 | 3,87 | 3,73 | 3,57 | 3,41 | 3,23 |
| 8 | 5,32 | 4,46 | 4,07 | 3,84 | 3,69 | 3,58 | 3,44 | 3,28 | 3,12 | 2,93 |
| 9 | 5,12 | 4,26 | 3,86 | 3,63 | 3,48 | 3,37 | 3,23 | 3,07 | 2,90 | 2,71 |
| 10 | 4,96 | 4,10 | 3,71 | 3,48 | 3,33 | 3,22 | 3,07 | 2,91 | 2,74 | 2,54 |
| 11 | 4,84 | 3,98 | 3,59 | 3,36 | 3,20 | 3,09 | 2,95 | 2,79 | 2,61 | 2,40 |
| 12 | 4,75 | 3,88 | 3,49 | 3,26 | 3,11 | 3,00 | 2,85 | 2,69 | 2,50 | 2,30 |
| 13 | 4,67 | 3,80 | 3,41 | 3,18 | 3,02 | 2,92 | 2,77 | 2,60 | 2,42 | 2,21 |
| 14 | 4,60 | 3,74 | 3,34 | 3,11 | 2,96 | 2,85 | 2,70 | 2,53 | 2,35 | 2,13 |
| 15 | 4,54 | 3,68 | 3,29 | 3,06 | 2,90 | 2,79 | 2,64 | 2,48 | 2,29 | 2,07 |
| 16 | 4,49 | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,59 | 2,42 | 2,24 | 2,01 |
| 17 | 4,45 | 3,59 | 3,20 | 2,96 | 2,81 | 2,70 | 2,55 | 2,38 | 2,19 | 1,96 |
| 18 | 4,41 | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,51 | 2,34 | 2,15 | 1,92 |
| 19 | 4,38 | 3,52 | 3,13 | 2,90 | 2,74 | 2,63 | 2,48 | 2,31 | 2,11 | 1,88 |
| 20 | 4,35 | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,45 | 2,28 | 2,08 | 1,84 |
| 21 | 4,32 | 3,47 | 3,07 | 2,84 | 2,68 | 2,57 | 2,42 | 2,25 | 2,05 | 1,81 |
| 22 | 4,30 | 3,44 | 3,05 | 2,82 | 2,66 | 2,55 | 2,40 | 2,23 | 2,03 | 1,78 |
| 23 | 4,28 | 3,42 | 3,03 | 2,80 | 2,64 | 2,53 | 2,38 | 2,20 | 2,00 | 1,76 |
| 24 | 4,26 | 3,40 | 3,01 | 2,78 | 2,62 | 2,51 | 2,36 | 2,18 | 1,98 | 1,73 |
| 25 | 4,24 | 3,38 | 2,99 | 2,76 | 2,60 | 2,49 | 2,34 | 2,16 | 1,96 | 1,71 |

ПРИЛОЖЕНИЕ Ж

Краткий словарь терминов

Алгебраическое дополнение элемента a_{ij} — произведение соответствующего минора, умноженного на минус единицу в степени $(i+j)$, где i — номер строки матрицы, j — номер столбца матрицы.

Альтернативная (конкурирующая) гипотеза — гипотеза, которая является логическим отрицанием нулевой гипотезы.

Априорный этап — начальный этап эконометрического моделирования, на котором проводится анализ сущности изучаемого объекта, формирование и формализация априорной (известной до начала моделирования) информации.

«Белый шум» — простейший пример временного ряда, у которого математическое ожидание равно нулю, а ошибки ε_t некоррелированы. В случае нормального распределения возмущения (ошибки) ε_t образуют нормальный (гауссовский) «белый шум».

Выборочная дисперсия — характеристика рассеяния случайной величины для выборочной совокупности. Исправленная выборочная дисперсия является состоятельной, несмещенной и эффективной точечной оценкой генеральной дисперсии.

Выборка — совокупность объектов, случайным образом отобранная из генеральной совокупности с целью исследования.

Временной (динамический) ряд, или ряд динамики — выборка наблюдений, в которой важны не только сами наблюдаемые значения случайных величин, но и порядок их следования друг за другом. Чаще всего упорядоченность обусловлена тем, что экспериментальные данные представляют собой серию наблюдений одной и той же случайной величины в последовательные моменты времени. В этом случае динамический ряд называется временным рядом. При этом предполагается, что тип распределения наблюдаемой случайной величины остается одним и тем же (например, нормальным), но параметры его могут меняться в зависимости от времени. Отдельные наблюдения называют уровнями ряда.

Взвешенный метод наименьших квадратов — применяется для отыскания параметра b и представляет собой обобщенный метод наименьших квадратов для модели с гетероскедастичностью, когда ковариационная матрица возмущений есть диагональная матрица.

Вероятностная (стохастическая или статистическая) зависимость — зависимость между двумя случайными величинами, причем каждому значению одной из них соответствует определенное (условное) распределение другой.

Гетероскедастичность — нарушение равенства дисперсий ошибок регрессии.

Генеральная совокупность — совокупность всех мыслимых результатов наблюдения, которые могут быть получены в данных условиях. Различают конечные, содержащие конечное число элементов, и бесконечные, содержащие бесконечное число элементов, генеральные совокупности.

Гомоскедастичность модели — свойство постоянства дисперсий ошибок регрессии ε для каждого значения x_i , равенство дисперсий возмущений (ошибок) регрессии.

Дискретная случайная величина — множество возможных значений случайной величины, число которых конечно или счетно.

Дисперсия $D(X)$ случайной величины X — математическое ожидание квадрата ее отклонения от математического ожидания. Дисперсия характеризует отклонение (разброс, рассеяние, вариацию) значений случайной величины относительно среднего значения.

Доверительная вероятность — достоверность (надежность) определения неизвестного значения параметра с помощью оценки параметра.

Доверительный интервал (при интервальной оценке неизвестного параметра генеральной совокупности) – числовой интервал, который с заданной доверительной вероятностью покрывает неизвестное значение параметра.

Зависимая переменная — в регрессионной модели некоторая переменная Y , являющаяся функцией регрессии с точностью до случайного возмущения.

Задачи регрессионного анализа — установление формы зависимости между переменными, оценка функции регрессии, оценка неизвестных значений (прогноз значений) зависимой переменной.

Закон распределения случайной величины — всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

Значимость на уровне - для коэффициента уравнения регрессии это означает, что гипотезу о равенстве его нулю надо отбросить на уровне значимости.

Интервальная оценка параметра — числовой интервал, который с заданной вероятностью покрывает неизвестное значение параметра. Интервал называется доверительным, а вероятность — доверительной (или надежностью оценки).

Классическая нормальная линейная модель множественной регрессии — модель, в которой зависимая переменная возмущения и объясняющие переменные удовлетворяют предпосылкам регрессионного анализа и предпосылке о невырожденности матрицы значений объясняющих переменных.

Ковариация (корреляционный момент) двух случайных величин – математическое ожидание произведения отклонений этих величин от своих математических ожиданий.

Коррелированность – наличие линейной зависимости между двумя случайными величинами. Случайные величины X и Y называются коррелированными, если коэффициент корреляции отличен от нуля и некоррелированными в противном случае. Из независимости случайных величин X и Y следует их некоррелированность (равенство нулю коэффициента корреляции), но из некоррелированности не следует их независимость, т.е. равенство нулю коэффициента корреляции указывает на отсутствие линейной связи между переменными, но не на отсутствие связи между ними вообще.

Коррелограмма — график выборочной автокорреляционной функции.

Корреляционная статистическая зависимость — соотношение, соответствие, зависимость между двумя случайными переменными, при этом каждому значению одной случайной переменной соответствует определенное условное математическое ожидание (среднее значение) другой.

Коэффициент авторегрессии — коэффициент корреляции между соседними возмущениями или коэффициент автокорреляции.

Коэффициент вариации — показатель относительной колеблемости признака, отношение среднего квадратического отклонения случайной величины к ее математическому ожиданию.

Коэффициент детерминации — одна из эффективных оценок адекватности регрессионной модели, мера качества уравнения регрессии. Он показывает, какая часть вариации зависимой переменной обусловлена вариацией объясняющей переменной. Чем ближе коэффициент детерминации к единице, тем лучше регрессия аппроксимирует эмпирические данные. Если коэффициент детерминации равен единице, то эмпирические точки лежат на линии регрессии и между переменными X и Y существует линейная функциональная связь. Если коэффициент детерминации равен нулю, то вариация зависимой переменной полностью обусловлена воздействием неучтенных в модели переменных и линия регрессии параллельна оси абсцисс. В случае парной линейной регрессионной модели коэффициент детерминации равен квадрату коэффициента корреляции.

Коэффициент корреляции двух случайных величин — отношение их ковариации к произведению средних квадратических отклонений этих величин. Коэффициент корреляции определяет тесноту линейной связи двух случайных величин. Чем ближе коэффициент корреляции по модулю к единице, тем теснее линейная связь. В случае равенства нулю коэффициента корреляции линейная корреляционная связь отсутствует и линия регрессии параллельна оси абсцисс.

Лаг — смещение во времени изменения одного показателя по сравнению с изменением другого.

Лаг временной — направление и продолжительность отставания уровней одного из взаимосвязанных временных рядов от уровней другого ряда.

Лаговые переменные — переменные, взятые в предыдущий момент времени и выступающие в качестве эндогенных и экзогенных переменных.

Линии регрессии (кривые регрессии) — графики функций регрессии, или просто регрессии Y по X и X по Y .

Линии регрессии нормально распределенных случайных величин — прямые линии, т.е. нормальные регрессии случайных величин всегда линейны.

Модель — идеальный или реальный объект, который заменяет другой объект в процессе его познания, сохраняет его существенные черты и позволяет получить новые знания о нём.

Математическое ожидание — важнейшая числовая характеристика случайных величин. Математическим ожиданием дискретной случайной величины называется сумма произведений всех ее возможных значений на соответствующие им вероятности.

Матрица — прямоугольная таблица чисел, содержащая m строк и n столбцов. Различают виды матриц: вектор-столбец, вектор-строка, квадратная, диагональная, нулевая, единичная, обратная и транспонированная.

Метод наименьших квадратов — метод обработки статистических наблюдений, применяется в корреляционном и регрессионном анализе; неизвестные значения параметров уравнения регрессии определяются так, чтобы сумма квадратов отклонений эмпирических значений от значений, найденных по уравнению регрессии, была минимальной.

Метод скользящих средних — метод выравнивания (сглаживания) временного ряда, т.е. выделения неслучайной составляющей. Основан на переходе от начальных значений членов ряда к их средним значениям на интервале времени, длина которого определена заранее. При этом сам выбранный интервал времени «скользит» вдоль ряда.

Метод уменьшения мультиколлинеарности — из двух объясняющих переменных, имеющих высокий коэффициент корреляции (больше 0,8), одну переменную исключают из рассмотрения. Обычно оставляют ту из двух переменных, которая имеет больший коэффициент корреляции с зависимой переменной.

Минор элемента a_{ij} — определитель матрицы $(n-1)$ -го порядка, полученной из матрицы A вычеркиванием i -й строки и j -го столбца.

Модели временных рядов — к этому классу моделей относятся модели тренда и модели сезонности.

Модель сезонности — характеризует устойчивые внутригодовые колебания уровня показателя.

Мультиколлинеарность — высокая взаимная коррелированность объясняющих переменных; может проявляться в функциональной (явной) и стохастической (скрытой) формах.

Обобщенная регрессионная модель — модель, в которой ковариации и дисперсии объясняющих переменных могут быть произвольными.

Обратная матрица — матрица, умноженная на исходную матрицу, в результате дающая единичную.

Объясняемая и объясняющая переменные — см. Односторонняя зависимость случайной переменной от одной (или нескольких) неслучайной независимой переменной.

Односторонняя зависимость случайной переменной от одной (или нескольких) неслучайной независимой переменной — возникает, например, в случае, когда при каждом фиксированном значении неслучайной независимой переменной соответствующие значения случайной переменной подвержены случайному разбросу за счет действия ряда неконтролируемых факторов. При этом зависимую переменную называют функцией отклика, объясняемой, выходной, результирующей, эндогенной переменной, результативным признаком, а независимую переменную — объясняющей, входной, предсказывающей, предикторной, экзогенной переменной, фактором, регрессором, факторным признаком.

Оценка параметра — всякая функция результатов наблюдений над случайной величиной X (иначе статистика), с помощью которой судят о значениях параметра.

Оценка несмещенная — оценка, математическое ожидание которой равно истинному значению искомого параметра.

Оценка эффективная — несмещенная оценка, имеющая минимальную дисперсию.

Оценка состоятельная — оценка, сходящаяся по вероятности к оцениваемому параметру с увеличением числа испытаний.

Парная регрессия – зависимость между переменными в генеральной совокупности вида, где X – неслучайная величина, Y и ε - случайные величины. Наиболее распространенный вид парной регрессии – линейная парная регрессия.

Пространственные данные (пространственная выборка) – в экономике под пространственной выборкой понимают набор показателей экономических переменных, полученных в данный момент времени. В эконометрике о пространственной выборке имеет смысл говорить в том случае, если все наблюдения получены примерно в неизменных условиях, т.е. представляют собой набор независимых выборочных данных из некоторой генеральной совокупности.

Процедура сглаживания экспериментальных данных — состоит из двух этапов: 1) определяется параметрическое семейство, к которому принадлежит искомая функция $Mx(Y)$ (рассматриваемая как функция от значений объясняющих переменных X). Это может быть множество линейных функций, показательных функций и т.д.; 2) находятся оценки параметров этой функции с помощью одного из методов математической статистики.

Ранг матрицы – наивысший порядок ее миноров, отличных от нуля.

Регрессия (Y по X) – зависимость условного математического ожидания $Mx(Y)$ случайной величины Y (при $X=x$) от x ; аналогично, регрессия (X по Y) – зависимость условного математического ожидания $Mu(Y)$ случайной величины X (при $Y=y$) от y .

Регрессии уравнение — уравнение линии, вокруг которой группируются точки корреляционного поля; указывает основное направление и тенденцию связи.

Регрессионный анализ — 1) раздел математической статистики, изучающий характер связи между случайными переменными; 2) совокупность статистических методов обработки результатов экспериментов, позволяющих в условиях стохастической зависимости выходной переменной от входных параметров определить данную зависимость.

Результирующая переменная, или результативный признак — см. Односторонняя зависимость случайной переменной от одной (или нескольких) неслучайной независимой переменной.

Репрезентативность (представительность) выборки — достаточно полное отражение свойств генеральной совокупности. Должна удовлетворять следующим требованиям: 1) элементы генеральной совокупности выбираются случайным образом; 2) независимость результатов испытаний в выборке; 3) определенный (правильный) подбор объема выборки.

Распределение хи-квадрат с k степенями свободы — распределение суммы квадратов k независимых случайных величин, распределенных по стандартному нормальному закону.

Ряд динамики — числовые значения статистического показателя, представленные во временной последовательности; ряд, расположенный в хронологической последовательности значений статистических показателей.

Фиктивные переменные — переменные, полученные путем перевода качественных признаков переменных в количественные.

Экзогенные переменные — независимые переменные, которые определяются вне системы x .

Эндогенные переменные — взаимосвязанные переменные, которые определяются внутри модели y .

Учебное издание

Юрий Яковлевич Настин

ЭКОНОМЕТРИКА

Редактор И. В. Голубева

Подписано в печать 14.10.2020 г. Формат 60 × 90 1/16. Уч.-изд. л. 6,2.
Печ. л. 5,1. Тираж 30 экз. Заказ № 96

Издательство федерального государственного бюджетного
образовательного учреждения высшего образования
«Калининградский государственный технический университет».
236022, Калининград, Советский проспект, 1