

Федеральное государственное образовательное бюджетное учреждение
высшего образования
**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»**
Новороссийский филиал
Кафедра «Информатика, математика и общегуманитарные науки»

И.Г. РЗУН

Методические рекомендации

АНАЛИЗ ДАННЫХ

Направление подготовки: 38.03.05 Бизнес-информатика

Направленность (профиль): IT- менеджмент в бизнесе.

Программа подготовки: академическая

Форма обучения: заочная, очная

Квалификация (степень) выпускника: бакалавр

Новороссийск 2017

1 ЦЕЛИ И ЗАДАЧИ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ.

1.1 Цель освоения дисциплины.

Дисциплина направлена на развитие теоретико-практической базы и формирование уровня математической подготовки, необходимых для понимания основных идей применения финансовых вычислений в экономике и финансах.

Анализ данных — один из важнейших разделов математической статистики и информатики, представляющий собой комплекс методов и средств, позволяющих получить из определенным образом организованных данных информацию для принятия решений. Целью изучения дисциплины «Анализ данных» является освоение теоретических основ и методов анализа данных, применяемых при решении прикладных (в том числе экономических) задач. Основным принципом, лежащим в основе данной дисциплины, состоит в повышении уровня фундаментальной экономико-математической и статистической подготовки студентов с усилением ее прикладной экономической направленности.

1.2 Задачи дисциплины.

Задачи изучения дисциплины вытекают из требований к результатам освоения и условиям реализации основной образовательной программы и компетенций, установленных Федеральным государственным образовательным стандартом высшего образования по направлению «Бизнес-информатика».

В ходе изучения дисциплины ставятся задачи:

1. Формирование навыков обработки, обобщения и анализа информации для оценки состояния, и выявления тенденций, закономерностей и конкретных особенностей развития социально-экономических и бизнес-процессов.

2. Овладение современными методиками статистического моделирования при решении задач.

3. Освоение компьютерных технологий, применяемых в анализе данных.

Дисциплина закладывает фундамент теоретических основ и методов анализа данных, применяемых при решении прикладных задач.

2. МЕТОДИЧЕСКИЙ МАТЕРИАЛ

Методические рекомендации по изучению дисциплины представлены в виде краткого описания некоторых разделов дисциплины. Приведён анализ решения типовых задач с применением MS Excel.

Технологию выполнения заданий, изложенных в методических рекомендациях, студенты выполняют на лабораторных занятиях во время учебного процесса или в качестве домашней контрольной в рамках КСР.

В рекомендациях подробно рассмотрены темы, соответствующие разделам рабочей программы дисциплины.

Задача самостоятельной работы при выполнении контрольной работы – выработать у студентов твердые навыки исследования и решения определенного круга задач, привить способность к самостоятельному аналитическому мышлению, умению работать со специальной и справочной литературой, таблицами.

При выполнении контрольной работы следует обратить внимание на следующие требования:

1. Задания к контрольной работе составлены в 100 вариантах. Каждый студент выполняет один вариант. Номер его варианта соответствует последним двум цифрам номера его зачетной книжки. Номер варианта указывается в самом начале работы.

2. Работы можно выполнять с помощью вычислительной техники и специального программного обеспечения (например, электронных таблиц MS Excel).

3. Нельзя ограничиваться приведением только готовых ответов. Расчеты должны быть представлены в развернутом виде, применяя где это необходимо табличные оформления исходной информации и расчетов, со всеми формулами, пояснениями и выводами, соблюдая достаточную точность вычислений. В пояснениях и

выводах показать, что именно и как характеризует исчисленный показатель.

4. Работа должна быть написана разборчиво, без помарок.

5. Контрольная работа может быть решена и защищена в рамках лабораторных работ.

Темы, разобранные в данных рекомендациях, расширяют знания, умения и навыки по разделам дисциплины.

Тема 1. ЛИНЕЙНЫЙ ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Одним из методов изучения стохастических связей между признаками является регрессионный анализ.

Регрессионный анализ представляет собой вывод уравнения регрессии, с помощью которого находится средняя величина случайной переменной (признака-результата), если величина другой (или других) переменных (признаков-факторов) известна. Он включает следующие этапы:

- 1) выбор формы связи (вида аналитического уравнения регрессии);
- 2) оценку параметров уравнения;
- 3) оценку качества аналитического уравнения регрессии.

Наиболее часто для описания статистической связи признаков используется линейная форма. Внимание к линейной связи объясняется четкой экономической интерпретацией ее параметров, ограниченной вариацией переменных и тем, что в большинстве случаев нелинейные формы связи для выполнения расчетов преобразуют (путем логарифмирования или замены переменных) в линейную форму.

В случае линейной парной связи уравнение регрессии примет вид: $y_i = a + b \cdot x_i + u_i$. Параметры данного уравнения a и b оцениваются по данным статистического наблюдения x и y . Результатом такой оценки является уравнение: $\tilde{y}_i = \tilde{a} + \tilde{b} \cdot x_i$, где \tilde{a} , \tilde{b} - оценки параметров a и b , \tilde{Y}_i - значение результативного признака (переменной), полученное по уравнению регрессии (расчетное значение).

Наиболее часто для оценки параметров используют **метод наименьших квадратов (МНК)**.

Метод наименьших квадратов дает наилучшие (состоятельные, эффективные и несмещенные) оценки параметров уравнения регрессии. Но только в том случае, если выполняются определенные предпосылки относительно случайного члена (u) и независимой переменной (x).

Задача оценивания параметров линейного парного уравнения методом наименьших квадратов состоит в следующем:

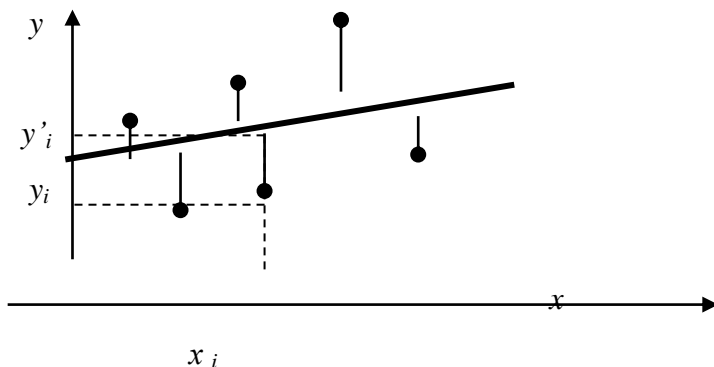
получить такие оценки параметров \tilde{a} , \tilde{b} , при которых сумма квадратов отклонений фактических

значений результативного признака - y_i от расчетных значений - \tilde{y}_i минимальна.

Формально критерий МНК можно записать так:

$$S = \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \rightarrow \min .$$

Проиллюстрируем суть данного метода графически. Для этого построим точечный график по данным наблюдений $(x_i, y_i, \quad i=1;n)$ в прямоугольной системе координат (такой точечный график называют корреляционным полем). Попытаемся подобрать прямую линию, которая ближе всего расположена к точкам корреляционного поля. Согласно методу наименьших квадратов линия выбирается так, чтобы сумма квадратов расстояний по вертикали между точками корреляционного поля и этой линией была бы минимальной.



Математическая запись данной задачи:

$$S = \sum_{i=1}^n (y_i - (\tilde{a} + \tilde{b} \cdot x_i))^2 \rightarrow \min .$$

Значения y_i и x_i $i=1;n$ нам известны, это данные наблюдений. В функции S они представляют собой константы. Переменными в данной функции являются искомые оценки параметров - \tilde{a} , \tilde{b} . Чтобы найти минимум функции 2-ух переменных необходимо вычислить частные производные данной функции по каждому из параметров и приравнять их нулю, т.е.

$$\frac{\partial S}{\partial \tilde{a}} = 0, \frac{\partial S}{\partial \tilde{b}} = 0 .$$

В результате получим систему из 2-ух нормальных

линейных уравнений:
$$\begin{cases} \sum_{i=1}^n y_i = \tilde{a} \cdot n + \tilde{b} \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i \cdot x_i = \tilde{a} \sum_{i=1}^n x_i + \tilde{b} \sum_{i=1}^n x_i^2 \end{cases}$$

Решая данную систему, найдем искомые оценки параметров:

$$\tilde{b}_1 = \frac{n \sum x_i y_i - \bar{x} \cdot \bar{y}}{n \sum x_i^2 - (\bar{x})^2} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x^2}$$

$$\tilde{a} = \bar{y} - \tilde{b} \cdot \bar{x}$$

Правильность расчета параметров уравнения регрессии может быть проверена сравнением сумм

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \tilde{y}_i \quad (\text{возможно некоторое расхождение из-за округления расчетов}).$$

Для расчета оценок параметров \tilde{a} , \tilde{b} можно построить таблицу 1.

Знак коэффициента регрессии b указывает направление связи (если $b > 0$, связь прямая, если $b < 0$, то связь обратная). Величина b показывает на сколько единиц изменится в среднем признак-результат -у при изменении признака-фактора - x на 1 единицу своего измерения.

Формально значение параметра a – среднее значение у при x равном нулю. Если признак-фактор не имеет и не может иметь нулевого значения, то вышеуказанная трактовка параметра a не имеет смысла.

Оценка тесноты связи между признаками осуществляется с помощью коэффициента линейной парной корреляции - $r_{x,y}$. Он может быть рассчитан по

формуле:
$$r_{x,y} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$$
. Кроме того, коэффициент

линейной парной корреляции может быть определен через

коэффициент регрессии b :
$$r_{x,y} = b \frac{\sigma_x}{\sigma_y}.$$

Область допустимых значений линейного коэффициента парной корреляции от -1 до $+1$. Знак коэффициента корреляции указывает направление связи. Если $r_{x,y} > 0$, то связь прямая; если $r_{x,y} < 0$, то связь обратная.

Если данный коэффициент по модулю близок к единице, то связь между признаками может быть интерпретирована как довольно тесная линейная. Если его модуль равен единице $|r_{x,y}| = 1$, то связь между признаками функциональная линейная. Если признаки x и y линейно независимы, то $r_{x,y}$ близок к 0 .

Для расчета $r_{x,y}$ можно использовать также таблицу

1.

Таблица 1

| N наблюдения | x_i | y_i | $x_i \cdot y_i$ | $(x_i - \bar{x})^2$ | $(y_i - \bar{y})^2$ |
|---------------------|------------|------------|--------------------|----------------------------|----------------------------|
| 1 | x_1 | y_1 | $x_1 \cdot y_1$ | $(x_1 - \bar{x})^2$ | $(y_1 - \bar{y})^2$ |
| 2 | x_2 | y_2 | $x_2 \cdot y_2$ | $(x_2 - \bar{x})^2$ | $(y_2 - \bar{y})^2$ |
| ... | | | | | |
| n | x_n | y_n | $x_n \cdot y_n$ | $(x_n - \bar{x})^2$ | $(y_n - \bar{y})^2$ |
| Сумма по столбцу | Σx | Σy | $\Sigma x \cdot y$ | $\Sigma (x_i - \bar{x})^2$ | $\Sigma (y_i - \bar{y})^2$ |

| | | | | | |
|------------------|--------------------------------|--------------------------------|---|---|---|
| Среднее значение | $\bar{x} = \frac{\sum x_i}{n}$ | $\bar{y} = \frac{\sum y_i}{n}$ | $\overline{x \cdot y} = \frac{\sum x_i \cdot y_i}{n}$ | $\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n}$ | $\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}$ |
|------------------|--------------------------------|--------------------------------|---|---|---|

Для оценки качества полученного уравнения регрессии рассчитывают теоретический коэффициент детерминации – R^2_{yx} :

$$R^2_{yx} = \frac{\delta^2}{\sigma_y^2} = \frac{\sum_{i=1}^n (\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\varepsilon^2}{\sigma_y^2} = 1 - \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

где δ^2 – объясненная уравнением регрессии дисперсия у;

ε^2 - остаточная (необъясненная уравнением регрессии) дисперсия у;

σ_y^2 - общая (полная) дисперсия у.

Коэффициент детерминации характеризует долю вариации (дисперсии) результативного признака у, объясняемую регрессией (а, следовательно, и фактором x), в общей вариации (дисперсии) у. Коэффициент детерминации R^2_{yx} принимает значения от 0 до 1. Соответственно величина $1-R^2_{yx}$ характеризует долю дисперсии у, вызванную влиянием прочих неучтенных в модели факторов и ошибками спецификации.

При парной линейной регрессии $R^2_{yx} = r^2_{yx}$.

Оценка статистической значимости параметров уравнения регрессии.

С помощью МНК мы получили лишь *оценки* параметров уравнения регрессии, которые характерны для конкретного статистического наблюдения (конкретного набора значений x и y). Если оценку параметров произвести по данным другого статистического наблюдения (другому набору значений x и y), то получим другие численные значения \tilde{a} , \tilde{b} . Мы предполагаем, что все эти наборы значений x и y извлечены из одной и той же генеральной совокупности. Чтобы проверить, значимы ли параметры, т.е. значимо ли они отличаются от нуля для генеральной совокупности используют статистические методы проверки гипотез.

В качестве основной (нулевой) гипотезы выдвигают гипотезу о незначимом отличии от нуля параметра или статистической характеристики в генеральной совокупности. Наряду с основной (проверяемой) гипотезой выдвигают альтернативную (конкурирующую) гипотезу о неравенстве нулю параметра или статистической характеристики в генеральной совокупности. В случае если основная гипотеза окажется неверной, мы принимаем

альтернативную. Для проверки этой гипотезы используется t -критерий Стьюдента.

Найденное по данным наблюдений значение t -критерия (его еще называют наблюдаемым или фактическим) сравнивается с табличным (критическим) значением, определяемым по таблицам распределения Стьюдента (которые обычно приводятся в конце учебников и практикумов по статистике или эконометрике). Табличное значение определяется в зависимости от уровня значимости (α) и числа степеней свободы, которое в случае линейной парной регрессии равно $(n-2)$, n -число наблюдений.

Если фактическое значение t -критерия больше табличного (по модулю), то основную гипотезу отвергают и считают, что с вероятностью $(1-\alpha)$ параметр или статистическая характеристика в генеральной совокупности значимо отличается от нуля.

Если фактическое значение t -критерия меньше табличного (по модулю), то нет оснований отвергать основную гипотезу, т.е. параметр или статистическая характеристика в генеральной совокупности незначимо отличается от нуля при уровне значимости α .

Для параметра b критерий проверки имеет вид:

$$t_{(b=0)} = \frac{\tilde{b}}{\mu_{\tilde{b}}},$$

где \tilde{b} - оценка коэффициента регрессии, полученная по наблюдаемым данным;

$\mu_{\tilde{b}}$ - стандартная ошибка коэффициента регрессии.

Для линейного парного уравнения регрессии стандартная ошибка коэффициента вычисляется по формуле:

$$\mu_{\tilde{b}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}}.$$

Числитель в этой формуле может быть рассчитан через коэффициент детерминации и общую дисперсию признака-результата:

$$\sum_{i=1}^n (y_i - \tilde{y}_i)^2 = n \cdot (1 - R_{yx}^2) \cdot \sigma_y^2.$$

Для параметра a критерий проверки гипотезы о незначимом отличии его от нуля имеет вид:

$$t_{(a=0)} = \frac{\tilde{a}}{\mu_{\tilde{a}}},$$

где \tilde{a} - оценка параметра регрессии, полученная по наблюдаемым данным;

$\mu_{\tilde{a}}$ – стандартная ошибка параметра a .

Для линейного парного уравнения регрессии:

$$\mu_{\tilde{a}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2 \sum_{i=1}^n x_i^2}{n(n-2) \sum_{i=1}^n (x_i - \bar{x})^2}}.$$

Для проверки гипотезы о незначимом отличии от нуля коэффициента линейной парной корреляции в генеральной совокупности используют следующий критерий:

$$t_{(r=0)} = \frac{r_{yx}}{\mu_r}, \quad \text{где } r_{yx} \text{ - оценка коэффициента}$$

корреляции, полученная по наблюдаемым данным; μ_r – стандартная ошибка коэффициента корреляции r_{yx} .

Для линейного парного уравнения регрессии:

$$\mu_r = \sqrt{\frac{(1 - r_{yx}^2)}{(n - 2)}}.$$

В парной линейной регрессии между наблюдаемыми значениями критериев существует взаимосвязь: $t_{(b=0)} = t_{(r=0)}$.

Прогноз ожидаемого значения результативного признака у по линейному парному уравнению регрессии.

Пусть требуется оценить значение признака-результата для заданного значения признака-фактора (x^p). Прогнозируемое значение признака-результата с доверительной вероятностью равной $(1-\alpha)$ принадлежит интервалу прогноза:

$$(\tilde{y}^p - t \cdot \mu_p; \tilde{y}^p + t \cdot \mu_p),$$

где \tilde{y}^p - точечный прогноз;

t – коэффициент доверия, определяемый по таблицам распределения Стьюдента в зависимости от уровня значимости α и числа степеней свободы $(n-2)$;

μ_p - средняя ошибка прогноза.

Точечный прогноз рассчитывается по линейному уравнению регрессии, как: $\tilde{y}^p = \tilde{a} + \tilde{b} \cdot x^p$.

Средняя ошибка прогноза определяется по формуле:

$$\mu_p = \sqrt{\frac{\sum (y_i - \tilde{y}_i)^2}{n-2} \left(1 + \frac{1}{n} + \frac{(x^p - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right)}.$$

Задание № 1

На основе данных, приведенных в Приложении 1 и соответствующих Вашему варианту (таблица 2), требуется:

1. Рассчитать коэффициент линейной парной корреляции и построить уравнение линейной парной регрессии одного признака от другого. Один из признаков, соответствующих Вашему варианту, будет играть роль факторного (x), другой – результативного (y). Причинно-следственные связи между признаками установить самим на основе экономического анализа. Пояснить смысл параметров уравнения.
2. Определить теоретический коэффициент детерминации и остаточную (необъясненную уравнением регрессии) дисперсию. Сделать вывод.
3. Оценить статистическую значимость уравнения регрессии в целом на пятипроцентном уровне с помощью F -критерия Фишера. Сделать вывод.
4. Выполнить прогноз ожидаемого значения признака-результата y при прогнозном значении признака-фактора x , составляющим 105% от среднего уровня x . Оценить точность прогноза, рассчитав ошибку

прогноза и его доверительный интервал с вероятностью 0,95.

Таблица 2

| Вариант | Номер начального наблюдения | Номер конечного наблюдения | Номер признака в из прил. 1 | Вариант | Номер начального наблюдения | Номер конечного наблюдения | Номер признака в из прил. 1 |
|-----------|-----------------------------|----------------------------|-----------------------------|-----------|-----------------------------|----------------------------|-----------------------------|
| <i>1</i> | <i>2</i> | <i>3</i> | <i>4</i> | <i>5</i> | <i>6</i> | <i>7</i> | <i>8</i> |
| 01 | 1 | 50 | 1,2 | 51 | 26 | 75 | 1,3 |
| 02 | 1 | 50 | 3,4 | 52 | 26 | 75 | 4,5 |
| 03 | 2 | 51 | 1,3 | 53 | 27 | 76 | 1,4 |
| 04 | 2 | 51 | 4,5 | 54 | 27 | 76 | 2,5 |
| 05 | 3 | 52 | 1,4 | 55 | 28 | 77 | 1,5 |
| 06 | 3 | 52 | 2,5 | 56 | 28 | 77 | 2,3 |
| 07 | 4 | 53 | 1,5 | 57 | 29 | 78 | 1,2 |
| 08 | 4 | 53 | 2,3 | 58 | 29 | 78 | 3,4 |
| 09 | 5 | 54 | 1,2 | 59 | 30 | 79 | 1,3 |
| 10 | 5 | 54 | 3,4 | 60 | 30 | 79 | 4,5 |
| 11 | 6 | 55 | 1,3 | 61 | 31 | 80 | 1,4 |
| 12 | 6 | 55 | 4,5 | 62 | 31 | 80 | 2,5 |
| 13 | 7 | 56 | 1,4 | 63 | 32 | 81 | 1,5 |
| 14 | 7 | 56 | 2,5 | 64 | 32 | 81 | 2,3 |
| 15 | 8 | 57 | 1,5 | 65 | 33 | 82 | 1,2 |
| 16 | 8 | 57 | 2,3 | 66 | 33 | 82 | 3,4 |
| 17 | 9 | 58 | 1,2 | 67 | 34 | 83 | 1,3 |
| 18 | 9 | 58 | 3,4 | 68 | 34 | 83 | 4,5 |
| 19 | 10 | 59 | 1,3 | 69 | 35 | 84 | 1,4 |
| 20 | 10 | 59 | 4,5 | 70 | 35 | 84 | 2,5 |

| | | | | | | | |
|-----------|----|----|-----|-----------|----|----|-----|
| 21 | 11 | 60 | 1,4 | 71 | 36 | 85 | 1,5 |
| 22 | 11 | 60 | 2,5 | 72 | 36 | 85 | 2,3 |
| 23 | 12 | 61 | 1,5 | 73 | 37 | 86 | 1,2 |
| 24 | 12 | 61 | 2,3 | 74 | 37 | 86 | 3,4 |
| 25 | 13 | 62 | 1,2 | 75 | 38 | 87 | 1,3 |
| 26 | 13 | 62 | 3,4 | 76 | 38 | 87 | 4,5 |
| 27 | 14 | 63 | 1,3 | 77 | 39 | 88 | 1,4 |
| 28 | 14 | 63 | 4,5 | 78 | 39 | 88 | 2,5 |
| 29 | 15 | 64 | 1,4 | 79 | 40 | 89 | 1,5 |
| 30 | 15 | 64 | 2,5 | 80 | 40 | 89 | 2,3 |
| 31 | 16 | 65 | 1,5 | 81 | 41 | 90 | 1,2 |
| 32 | 16 | 65 | 2,3 | 82 | 41 | 90 | 3,4 |
| 33 | 17 | 66 | 1,2 | 83 | 42 | 91 | 1,3 |
| 34 | 17 | 66 | 3,4 | 84 | 42 | 91 | 4,5 |
| 35 | 18 | 67 | 1,3 | 85 | 43 | 92 | 1,4 |
| 36 | 18 | 67 | 4,5 | 86 | 43 | 92 | 2,5 |
| 37 | 19 | 68 | 1,4 | 87 | 44 | 93 | 1,5 |
| 38 | 19 | 68 | 2,5 | 88 | 44 | 93 | 2,3 |
| 39 | 20 | 69 | 1,5 | 89 | 45 | 94 | 1,2 |
| 40 | 20 | 69 | 2,3 | 90 | 45 | 94 | 3,4 |
| 41 | 21 | 70 | 1,2 | 91 | 46 | 95 | 1,3 |
| 42 | 21 | 70 | 3,4 | 92 | 46 | 95 | 4,5 |
| 43 | 22 | 71 | 1,3 | 93 | 47 | 96 | 1,4 |
| 44 | 22 | 71 | 4,5 | 94 | 47 | 96 | 2,5 |
| 45 | 23 | 72 | 1,4 | 95 | 48 | 97 | 1,5 |
| 46 | 23 | 72 | 2,5 | 96 | 48 | 97 | 2,3 |
| 47 | 24 | 73 | 1,5 | 97 | 49 | 98 | 1,2 |
| 48 | 24 | 73 | 2,3 | 98 | 49 | 98 | 3,4 |
| 49 | 25 | 74 | 1,2 | 99 | 50 | 99 | 1,3 |
| 50 | 25 | 74 | 3,4 | 0 | 50 | 99 | 4,5 |

Тема 2. МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Построение уравнения множественной регрессии начинается с решения вопроса о *спецификации модели*, который в свою очередь включает 2 круга вопросов: отбор факторов и выбор уравнения регрессии.

Отбор факторов обычно осуществляется в два этапа:

1) теоретический анализ взаимосвязи результата и круга факторов, которые оказывают на него существенное влияние;

2) количественная оценка взаимосвязи факторов с результатом. При линейной форме связи между признаками данный этап сводится к анализу корреляционной матрицы (матрицы парных линейных коэффициентов корреляции):

$$r_{y,y} \quad r_{y,x1} \quad r_{yx2} \quad \dots \quad r_{y,xt}$$

$$r_{x1,y} \quad r_{x1,x2} \quad r_{x2x2} \quad \dots \quad r_{x2,xt}$$

.....

$$r_{xt,y} \quad r_{xt,x1} \quad r_{xt,x2} \quad \dots \quad r_{xt,xt}$$

где $r_{y,xj}$ – линейный парный коэффициент корреляции, измеряющий тесноту связи между признаками y и x_j $j=1;m$, m -число факторов.

r_{x_j, x_k} – линейный парный коэффициент корреляции, измеряющий тесноту связи между признаками x_j и x_k $j, k=1; m$.

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

1. Они должны быть количественно измеримы. Если необходимо включить в модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность (например, в модели урожайности качество почвы задается в виде баллов).

2. Каждый фактор должен быть достаточно тесно связан с результатом (т.е. коэффициент парной линейной корреляции между фактором и результатом должен быть существенным).

3. Факторы не должны быть сильно коррелированы друг с другом, тем более находиться в строгой функциональной связи (т.е. они не должны быть интеркоррелированы).

Разновидностью интеркоррелированности факторов является мультиколлинеарность - тесная линейная связь между факторами.

Мультиколлинеарность может привести к нежелательным последствиям:

1) оценки параметров становятся ненадежными. Они обнаруживают большие стандартные ошибки. С изменением объема наблюдений оценки меняются (не только по величине, но и по знаку), что делает модель непригодной для анализа и прогнозирования.

2) затрудняется интерпретация параметров множественной регрессии как характеристик действия факторов в «чистом» виде, ибо факторы коррелированы; параметры линейной регрессии теряют экономический смысл;

3) становится невозможным определить изолированное влияние факторов на результативный показатель.

Мультиколлинеарность имеет место, если определитель матрицы межфакторной корреляции близок к нулю:

$$\text{Det}|R| = \begin{vmatrix} r_{x1x1} & r_{x2x1} & r_{x3x1} \\ r_{x1x2} & r_{x2x2} & r_{x3x2} \\ r_{x1x3} & r_{x2x3} & r_{x3x3} \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

Если же определитель матрицы межфакторной корреляции близок к единице, то мультиколлинеарности нет. Существуют различные подходы преодоления сильной межфакторной корреляции. Простейший из них –

исключение из модели фактора (или факторов), в наибольшей степени ответственных за мультиколлинеарность при условии, что качество модели при этом пострадает несущественно (а именно, теоретический коэффициент детерминации $-R^2_{y(x1...xm)}$ снизится несущественно).

Определение факторов, ответственных за мультиколлинеарность, может быть основано на анализе матрицы межфакторной корреляции. При этом определяют пару признаков-факторов, которые сильнее всего связаны между собой (коэффициент линейной парной корреляции максимален по модулю). Из этой пары в наибольшей степени ответственным за мультиколлинеарность будет тот признак, который теснее связан с другими факторами модели (имеет более высокие по модулю значения коэффициентов парной линейной корреляции).

Еще один способ определения факторов, ответственных за мультиколлинеарность основан на вычислении коэффициентов множественной детерминации ($R^2_{xj(x1, \dots, xj-1, xj+1, \dots, xm)}$), показывающего зависимость фактора x_j от других факторов модели $x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_m$. Чем ближе значение коэффициента множественной детерминации к единице, тем больше ответственность за мультиколлинеарность фактора, выступающего в роли

зависимой переменной. Сравнивая между собой коэффициенты множественной детерминации для различных факторов можно проранжировать переменные по степени ответственности за мультиколлинеарность.

При выборе формы уравнения множественной регрессии предпочтение отдается линейной функции:

$$y_i = a + b_1 \cdot x1_i + b_2 \cdot x2_i + \dots + b_m \cdot x m_i + u_i$$

в виду четкой интерпретации параметров.

Данное уравнение регрессии называют уравнением регрессии в естественном (натуральном) масштабе. Коэффициент регрессии b_j при факторе x_j называют *условно-чистым коэффициентом регрессии*. Он измеряет среднее по совокупности отклонение признака-результата от его средней величины при отклонении признака-фактора x_j на единицу, при условии, что все прочие факторы модели не изменяются (зафиксированы на своих средних уровнях).

Если не делать предположения о значениях прочих факторов, входящих в модель, то это означало бы, что каждый из них при изменении x_j также изменялся бы (так как факторы связаны между собой), и своими изменениями оказывали бы влияние на признак-результат.

РАСЧЕТ ПАРАМЕТРОВ УРАВНЕНИЯ ЛИНЕЙНОЙ МНОЖЕСТВЕННОЙ РЕГРЕССИИ.

Параметры уравнения множественной регрессии можно оценить методом наименьших квадратов, составив и решив систему нормальных линейных уравнений.

Кроме того, для линейной множественной регрессии существует другой способ реализации МНК при оценке параметров - через β -коэффициенты (через параметры уравнения регрессии в стандартных масштабах).

Модель регрессии в стандартном масштабе предполагает, что все значения исследуемых признаков переводятся в стандарты (стандартизованные значения) по формулам:

$$t_{x_j} = \frac{x_j^i - \overline{x_j}}{\sigma_{x_j}}, \quad j=1;m,$$

где x_j^i - значение переменной x_j в i -ом наблюдении.

$$t_{y_i} = \frac{y_i - \overline{y}}{\sigma_y}.$$

Таким образом, начало отсчета каждой стандартизованной переменной совмещается с ее средним значением, а в качестве единицы изменения принимается

ее среднее квадратическое отклонение σ . Если связь между переменными в естественном масштабе линейная, то изменение начала отсчета и единицы измерения этого свойства не нарушат, так что и стандартизованные переменные будут связаны линейным соотношением:

$$\tilde{t}_y = \sum_{j=1}^m \beta_j \cdot t_{xj} .$$

Для оценки β -коэффициентов применим МНК. При этом система нормальных уравнений будет иметь вид:

$$r_{x1y} = \beta_1 + r_{x1x2} \cdot \beta_2 + \dots + r_{x1xm} \cdot \beta_m$$

$$r_{x2y} = r_{x2x1} \cdot \beta_1 + \beta_2 + \dots + r_{x2xm} \cdot \beta_m$$

...

$$r_{xmy} = r_{xmx1} \cdot \beta_1 + r_{xmx2} \cdot \beta_2 + \dots + \beta_m$$

Найденные из данной системы β -коэффициенты позволяют определить значения коэффициентов в регрессии в естественном масштабе по формулам:

$$\tilde{b}_j = \beta_j \cdot \frac{\sigma_y}{\sigma_x}, j=1; m; \quad \tilde{a} = \bar{y} - \sum_{j=1}^m \tilde{b}_j \cdot \bar{x}_j .$$

Показатели тесноты связи факторов с результатом.

Если факторные признаки различны по своей сущности и (или) имеют различные единицы измерения, то коэффициенты регрессии b_j при разных факторах являются несопоставимыми. Поэтому уравнение регрессии дополняют соизмеримыми показателями тесноты связи фактора с результатом, позволяющими ранжировать факторы по силе влияния на результат. К таким показателям тесноты связи относят: частные коэффициенты эластичности, β -коэффициенты, частные коэффициенты корреляции.

Частные коэффициенты эластичности \mathcal{E}_j рассчитываются по формуле:

$$\mathcal{E}_j = \frac{\partial y}{\partial x_j} \cdot \frac{\overline{x_j}}{\overline{y_{x_1, \dots, x_m}}}$$

Частный коэффициент эластичности показывают, на сколько процентов в среднем изменяется признак-результат y с изменением признака-фактора x_j на один процент от своего среднего уровня при фиксированном положении других факторов модели. В случае линейной зависимости \mathcal{E}_j рассчитываются по формуле:

$$\mathcal{E}_j = \tilde{b}_j \cdot \frac{\overline{x_j}}{\overline{y_{x_1, \dots, x_m}}}$$
, где \tilde{b}_j – оценка коэффициента регрессии при j -ом факторе.

Стандартизированные частные коэффициенты регрессии - β -коэффициенты (β_j) показывают, на какую часть своего среднего квадратического отклонения σ_y изменится признак-результат y с изменением соответствующего фактора x_j на величину своего среднего квадратического отклонения (σ_{x_j}) при неизменном влиянии прочих факторов (входящих в уравнение).

По коэффициентам эластичности и β -коэффициентам могут быть сделаны противоположные выводы. Причины этого: а) вариация одного фактора очень велика; б) разнонаправленное воздействие факторов на результат.

Коэффициент β_j может также интерпретироваться как показатель прямого (непосредственного) влияния j -ого фактора (x_j) на результат (y). Во множественной регрессии j -ый фактор оказывает не только прямое, но и косвенное (опосредованное) влияние на результат (т.е. влияние через другие факторы модели). Косвенное влияние измеряется

величиной: $\sum_{i=1, \dots, j-1, j+1, \dots, m} \beta_i \cdot r_{x_j, x_i}$, где m - число факторов в

модели. Полное влияние j -ого фактора на результат равно сумме прямого и косвенного влияний измеряет

коэффициент линейной парной корреляции данного фактора и результата – $r_{xj,y}$.

Коэффициент частной корреляции измеряет «чистое» влияние фактора на результат при устранении воздействия прочих факторов модели.

Для расчета частных коэффициентов корреляции могут быть использованы парные коэффициенты корреляции.

Для случая зависимости y от двух факторов можно вычислить 2 коэффициента частной корреляции:

$$r_{yx1/x2} = \frac{r_{x1y} - r_{x2y} \cdot r_{x1x2}}{\sqrt{(1 - r_{x1x2}^2)(1 - r_{x2y}^2)}},$$

(фактор x_2 фиксирован).

$$r_{yx2/x1} = \frac{r_{x2y} - r_{x1y} \cdot r_{x1x2}}{\sqrt{(1 - r_{x1x2}^2)(1 - r_{x1y}^2)}}$$

(фактор x_1 фиксирован).

Это коэффициенты частной корреляции 1-ого порядка (порядок определяется числом факторов, влияние которых устраняется).

Частные коэффициенты корреляции, рассчитанные по таким формулам изменяются от -1 до $+1$. Они используются не только для ранжирования факторов модели по степени влияния на результат, но и также для

отсева факторов. При малых значениях $r_{yxm/x1,x2,\dots,xm-1}$ нет смысла вводить в уравнение m -ый фактор, т.к. его чистое влияние на результат несущественно.

Коэффициенты множественной детерминации и корреляции характеризуют совместное влияние всех факторов на результат.

По аналогии с парной регрессией можно определить долю вариации результата, объясненной вариацией включенных в модель факторов (δ^2), в его общей вариации (σ_y^2). Ее количественная характеристика – теоретический множественный коэффициент детерминации ($R^2_{y(x1,\dots,xm)}$). Для линейного уравнения регрессии данный показатель может быть рассчитан через β -коэффициенты, как:

$$R^2_{y(x1,\dots,xm)} = \sum_{j=1}^m \beta_j \cdot r_{y \cdot xj} .$$

$$R_{y(x1,\dots,xm)} = \sqrt{R^2_{y(x1,\dots,xm)}} \quad - \quad \text{коэффициент}$$

множественной корреляции. Он принимает значения от 0 до 1 (в отличии от парного коэффициента корреляции, который может принимать отрицательные значения). Поэтому R не может быть использован для интерпретации направления связи. Чем плотнее фактические значения y_i располагаются относительно линии регрессии, тем меньше

остаточная дисперсия и, следовательно, больше величина $R_{y(x_1, \dots, x_m)}$. Таким образом, при значении R близком к 1, уравнение регрессии лучше описывает фактические данные и факторы сильнее влияют на результат. При значении R близком к 0 уравнение регрессии плохо описывает фактические данные и факторы оказывают слабое воздействие на результат.

Оценка значимости полученного уравнения множественной регрессии.

Оценка значимости уравнения множественной регрессии осуществляется путем проверки гипотезы о равенстве нулю коэффициент детерминации рассчитанного по данным генеральной совокупности: $R_{y(x_1, \dots, x_m)}^{2(\varepsilon)} = 0$ или $b_1 = b_2 = \dots = b_m = 0$ (гипотеза о незначимости уравнения регрессии, рассчитанного по данным генеральной совокупности).

Для ее проверки используют F -критерий Фишера.

При этом вычисляют фактическое (наблюдаемое) значение F -критерия, через коэффициент детерминации $R_{y(x_1, \dots, x_m)}^2$, рассчитанный по данным конкретного наблюдения:

$$F = \frac{R_{y(x_1, \dots, x_m)}^2}{1 - R_{y(x_1, \dots, x_m)}^2} \cdot \frac{n - h}{h - 1}, \text{ где } n - \text{число наблюдений;}$$

h – число оцениваемых параметров (в случае двухфакторной линейной регрессии $h=3$).

По таблицам распределения Фишера-Снедеккера находят критическое значение F -критерия ($F_{кр}$). Для этого задаются уровнем значимости α (обычно его берут равным 0,05) и двумя числами степеней свободы $k_1=h-1$ и $k_2=n-h$.

Сравнивают фактическое значение F -критерия ($F_{набл}$) с табличным $F_{кр}(\alpha; k_1; k_2)$. Если $F_{набл} < F_{кр}(\alpha; k_1; k_2)$, то гипотезу о незначимости уравнения регрессии не отвергают. Если $F_{набл} > F_{кр}(\alpha; k_1; k_2)$, то выдвинутую гипотезу отвергают и принимают альтернативную гипотезу о статистической значимости уравнения регрессии.

Задание № 2

На основе данных, приведенных в Приложении и соответствующих Вашему варианту (таблица 2), требуется:

1. Построить уравнение множественной регрессии. При этом признак-результат и один из факторов остаются теми же, что и в первом задании. Выберите дополнительно еще один фактор из приложения 1

(границы наблюдения должны совпадать с границами наблюдения признака-результата, соответствующего Вашему варианту). При выборе фактора нужно руководствоваться его экономическим содержанием или другими подходами. Пояснить смысл параметров уравнения.

2. Рассчитать частные коэффициенты эластичности. Сделать вывод.
3. Определить стандартизованные коэффициенты регрессии (β -коэффициенты). Сделать вывод.
4. Определить парные и частные коэффициенты корреляции, а также множественный коэффициент корреляции; сделать выводы.
5. Оценить значимость параметров уравнения регрессии с помощью t -критерия Стьюдента, а также значимость уравнения регрессии в целом с помощью общего F -критерия Фишера. Предложить окончательную модель (уравнение регрессии). Сделать выводы.

Тема 3. СИСТЕМЫ ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

Не всегда получается описать адекватно сложное социально-экономическое явление с помощью только одного соотношения (уравнения). Кроме того, некоторые переменные могут оказывать взаимные воздействия и трудно однозначно определить, какая из них является зависимой, а какая независимой переменной. Поэтому при построении эконометрической модели прибегают к системам уравнений.

В любой эконометрической модели в зависимости от конечных прикладных целей ее использования все участвующие в ней переменные подразделяются на:

- Экзогенные (независимые) – значения которых задаются «извне», автономно, в определенной степени они являются управляемыми (планируемыми) (X);
- Эндогенные (зависимые) - значения которых определяются внутри модели, или взаимозависимые (Y).
- Лаговые – экзогенные или эндогенные переменные эконометрической модели, датированные предыдущими моментами времени и находящиеся в уравнении с текущими переменными. Например:

y_t – текущая эндогенная переменная,

y_{t-1} – лаговая эндогенная переменная (отстоящая от текущей на 1 период назад),

y_{t-2} – тоже лаговая эндогенная переменная (отстоящая от текущей на 2 периода).

- Предопределенные переменные – переменные, определяемые вне модели. К ним относятся лаговые и текущие экзогенные переменные (x_t, x_{t-1}), а также лаговые эндогенные переменные (y_{t-1}).

Все эконометрические модели предназначены для объяснения текущих значений эндогенных переменных по значениям предопределенных переменных.

В дальнейшем для простоты будем рассматривать в качестве предопределенных переменных только текущие экзогенные переменные (x).

Система уравнений в эконометрических исследованиях может быть построена по-разному. Выделяют следующие 3 вида систем уравнений.

1. Система независимых уравнений, когда каждая зависимая переменная (y) рассматривается как функция только от предопределенных переменных (x):

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m + u_1 \\ y_2 = a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m + u_2 \\ \dots \\ y_k = a_{k1}x_1 + a_{k2}x_2 + \dots + a_{km}x_m + u_k \end{cases}$$

2. Система рекурсивных уравнений, когда в каждом последующем уравнении системы зависимая переменная представляет функцию от зависимых и предопределенных переменных предшествующих уравнений:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m + u_1 \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m + u_2 \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + \dots + a_{3m}x_m + u_3 \\ \dots \\ y_k = b_{k1}y_1 + b_{k2}y_2 + \dots + b_{kk-1}y_{k-1} + a_{k1}x_1 + a_{k2}x_2 + \dots + a_{km}x_m + u_k \end{cases}$$

В рассмотренных 2-ух видах систем каждое уравнение может рассматриваться самостоятельно, и его параметры можно определить с помощью метода наименьших квадратов (МНК).

3. Система взаимозависимых (совместных, одновременных) уравнений, когда зависимые переменные в одних уравнениях входят в левую часть (т.е. выступают в роли признаков-результатов), а в

других уравнениях – в правую часть системы (т.е. выступают в роли признаков-факторов) одновременно:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \dots + b_{1k}y_k + a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ y_2 = b_{21}y_1 + b_{23}y_3 + \dots + b_{2k}y_k + a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \dots \\ y_k = b_{k1}y_1 + b_{k2}y_2 + \dots + b_{kk-1}y_{k-1} + a_{k1}x_1 + a_{k2}x_2 + \dots + a_{kn}x_n \end{cases}$$

Название «система одновременных уравнений» подчеркивает тот факт, что в системе одни и те же переменные одновременно рассматриваются как зависимые в одних уравнениях и как независимые в других.

В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим, т.к. нарушаются предпосылки, лежащие в основе МНК. В результате оценки параметров получаются смещенными.

В эконометрике эта система уравнений также называется *структурной формой* модели.

Некоторые из уравнений системы могут быть представлены в виде тождеств, т.е. параметры этих уравнений являются константами.

От структурной формы легко перейти к так называемой *приведенной форме* модели. Число уравнений в приведенной форме равно числу эндогенных переменных модели. В каждом уравнении приведенной формы эндогенная переменная выражается через все предопределенные переменные модели:

$$\begin{cases} y_1 = A_{11}x_1 + \dots + A_{1m}x_m + U_1 \\ y_2 = A_{21}x_1 + \dots + A_{2m}x_m + U_2 \\ \dots\dots\dots \\ y_k = A_{k1}x_1 + \dots + A_{km}x_m + U_k \end{cases}$$

Так как правая часть каждого из уравнений приведенной формы содержит только предопределенные переменные и остатки, а левая часть только одну из эндогенных переменных, то такая система является системой независимых уравнений. Поэтому параметры каждого из уравнений системы в приведенной форме можно определить независимо обычным МНК.

Зная оценки этих приведенных коэффициентов можно определить параметры структурной формы модели. Но не всегда, а только если модель является идентифицируемой.

Проблема идентификации.

Модель считается точно идентифицированной, если все ее уравнения точно идентифицированы.

Если среди уравнений модели есть хотя бы одно сверхидентифицированное уравнение, то вся модель считается сверхидентифицированной.

Если среди всех уравнений модели есть хотя бы одно неидентифицированное, то вся модель считается неидентифицированной.

Уравнение называется точно идентифицированным, если оценки структурных параметров можно однозначно (единственным способом) найти по коэффициентам приведенной модели.

Уравнение сверхидентифицировано, если для некоторых структурных параметров можно получить более одного численного значения.

Уравнение называется неидентифицированным, если оценки его структурных параметров невозможно найти по коэффициентам приведенной модели.

Правила идентификации- необходимое и достаточное условия идентификации (применяются только к структурной форме модели).

Введем следующие обозначения:

M- число предопределенных переменных в модели;

m - число predeterminedных переменных в данном уравнении;

K – число эндогенных переменных в модели;

k – число эндогенных переменных в данном уравнении.

Необходимое (но недостаточное) условие идентификации уравнения модели:

Для того чтобы уравнение модели было идентифицируемо, необходимо, чтобы число predeterminedных переменных, не входящих в уравнение, было не меньше «числа эндогенных переменных, входящих в уравнение минус 1», т.е. : $M-m \geq k-1$;

Если $M-m = k-1$, уравнение точно идентифицировано.

Если $M-m > k-1$, уравнение сверхидентифицировано.

Эти правила следует применять в структурной форме модели.

Достаточное условие идентификации уравнения модели.

Введем обозначения: A – матрица коэффициентов при переменных не входящих в данное уравнение.

Достаточное условие идентификации заключается в том, что ранг матрицы A должен быть равен $(K-1)$. Ранг

матрицы – размер наибольшей ее квадратной подматрицы, определитель которой не равен нулю.

Сформулируем необходимое и достаточное условия идентификации уравнения модели:

1) Если $M-m > k-1$ и ранг матрицы A равен $K-1$, то уравнение сверхидентифицировано.

2) Если $M-m = k-1$ и ранг матрицы A равен $K-1$, то уравнение точно идентифицировано.

3) Если $M-m < k-1$ и ранг матрицы A меньше $K-1$, то уравнение неидентифицировано.

4) Если $M-m < k-1$, то уравнение неидентифицировано. В этом случае ранг матрицы A будет меньше $K-1$.

Оценка точно идентифицированного уравнения осуществляется с помощью *косвенного метода наименьших квадратов (КМНК)*.

Алгоритм КМНК включает 3 шага:

1) составление приведенной формы модели и выражение каждого коэффициента приведенной формы через структурные параметры;

2) применение обычного МНК к каждому уравнению приведенной формы и получение численных оценок приведенных параметров;

3) определение оценок параметров структурной формы по оценкам приведенных коэффициентов, используя соотношения, найденные на шаге 1.

Оценка свержидентифицированного уравнения осуществляется при помощи *двухшагового метода наименьших квадратов*.

Алгоритм двухшагового МНК включает следующие шаги:

1) составление приведенной формы модели;

2) применение обычного МНК к каждому уравнению приведенной формы и получение численных оценок приведенных параметров;

3) определение расчетных значений эндогенных переменных, которые фигурируют в качестве факторов в структурной форме модели;

4) определение структурных параметров каждого уравнения в отдельности обычным МНК, используя в качестве факторов входящие в это уравнение predetermined переменные и расчетные значения эндогенных переменных, полученные на шаге 1.

Рассмотрим пример.

Пусть имеется система:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2 + u_1 \\ y_2 = b_{21}y_1 + a_{21}x_1 + u_2 \\ y_3 = b_{32}y_2 + a_{31}x_1 + a_{33}x_3 + u_3 \end{cases}$$

Требуется составить приведенную форму модели, проверить каждое уравнение структурной модели на идентификацию, и предложить способ оценки параметров структурной формы модели.

Решение:

В этой системе y_1, y_2, y_3 - эндогенные переменные ($K=3$);

x_1, x_2, x_3 - предопределенные переменные ($M=3$).

$K-1=2; K+M=6$.

Составим приведенную форму модели:

$$\begin{cases} y_1 = A_{11}x_1 + A_{12}x_2 + A_{13}x_3 + U_1 \\ y_2 = A_{21}x_1 + A_{22}x_2 + A_{23}x_3 + U_2 \\ y_3 = A_{31}x_1 + A_{32}x_2 + A_{33}x_3 + U_3 \end{cases}$$

Проверим, как выполняется необходимое условие идентификации для каждого уравнения.

Для 1-ого уравнения имеем: $k_1=3; m_1=2$;

$M-m_1=1 < k_1-1=2$, следовательно, 1-ое уравнение неидентифицировано.

Для 2-ого уравнения имеем: $k_2=2; m_2=1$;

$M-m_2=2 > k_2-1=1$, следовательно, 2-ое уравнение сверхидентифицированно.

Для 3-его уравнения имеем: $k_3=2; m_3=2$;

$M-m_3=1 = k_3-1=1$, следовательно, 3-е уравнение точно идентифицированно.

Рассмотрим, как выполняется достаточное условие идентификации для каждого уравнения системы. Для того, чтобы оно выполнялось необходимо, чтобы определитель матрицы A (матрицы коэффициентов при переменных, не входящих в это уравнение) был равен $K-1=2$.

Составим матрицу A для 1-ого уравнения системы. В 1-ом уравнении отсутствует лишь одна переменная системы x_3 . Поэтому матрица A будет иметь вид:

$$\begin{array}{l} x_3 \\ 0 \quad - \text{ во 2-ом уравнении} \\ a_{33} \quad - \text{ в 3-ем уравнении} \end{array}$$

Ранг данной матрицы равен 1, что меньше $K-1=2$, следовательно, 1-ое уравнение модели неидентифицированно.

Составим матрицу A для 2-ого уравнения системы. Во 2-ом уравнении отсутствуют переменные u_3, x_2, x_3 :

$$\begin{array}{l} u_3 \quad x_2 \quad x_3 \\ b_{13} \quad a_{13} \quad 0 \quad - \text{ в 1-ом уравнении} \\ 1 \quad a_{32} \quad a_{33} \quad - \text{ в 3-ем уравнении} \end{array}$$

Ранг данной матрицы равен 2, что равно $K-1=2$, следовательно, 2-ое уравнение модели точно идентифицировано.

Составим матрицу A для 3-его уравнения системы. В 3-ем уравнении отсутствуют переменные y_1, x_2 :

$$\begin{array}{r} y_1 \quad x_2 \\ 1 \quad a_{12} \quad - \text{ в 1-ом уравнении} \\ b_{21} \quad 0 \quad - \text{ во 2-ом уравнении} \end{array}$$

Ранг данной матрицы равен 1, что меньше $K-1=2$, следовательно, 3-е уравнение модели неидентифицировано.

Сделаем выводы: 1-ое и 3-е уравнения системы неидентифицированы (т.к. не выполняются достаточные условия идентификации, а в случае 1-ого уравнения и необходимое условие также). 2-ое уравнение системы сверхидентифицировано. Следовательно, система в целом является неидентифицируемой.

Для оценки параметров 2-ого уравнения можно применить двухшаговый МНК. Параметры 1-ого и 3-его уравнений определить по коэффициентам приведенной формы нельзя. Поэтому модель должна быть модифицирована.

Задание № 3

На основе данных, приведенных в таблице 3 и соответствующих Вашему варианту (таблица 4) провести идентификацию модели и описать процедуру оценивания параметров уравнений структурной формы модели.

Таблица 3

| Уравнение | Вариант уравнения | Коэффициенты перед регрессорами | | | | |
|-----------|-------------------|---------------------------------|-----------------|-----------------|-----------------|-----------------|
| | | y2 | y3 | x1 | x2 | x3 |
| y1 | 1 | 0 | 0 | a ₁₁ | a ₂₁ | a ₃₁ |
| | 2 | 0 | b ₃₁ | 0 | a ₂₁ | a ₃₁ |
| | 3 | 0 | b ₃₁ | a ₁₁ | a ₂₁ | 0 |
| | 4 | 0 | b ₃₁ | a ₁₁ | 0 | a ₃₁ |
| | 5 | b ₂₁ | b ₃₁ | a ₁₁ | 0 | a ₃₁ |
| | | y1 | y3 | x1 | x2 | x3 |
| y2 | 1 | b ₁₂ | b ₃₂ | 0 | 0 | a ₃₂ |
| | 2 | b ₁₂ | 0 | a ₁₂ | a ₂₂ | 0 |
| | 3 | 0 | b ₃₂ | a ₁₂ | a ₂₂ | a ₃₂ |
| | 4 | b ₁₂ | b ₃₂ | a ₁₂ | a ₂₂ | 0 |
| | 5 | b ₁₂ | b ₃₂ | 0 | a ₂₂ | a ₃₂ |
| | | y1 | y2 | x1 | x2 | x3 |
| y3 | 1 | b ₁₃ | b ₂₃ | a ₁₃ | 0 | 0 |
| | 2 | b ₁₃ | 0 | 0 | a ₂₃ | a ₃₃ |
| | 3 | b ₁₃ | 0 | a ₁₃ | 0 | a ₃₃ |
| | 4 | b ₁₃ | 0 | a ₁₃ | a ₂₃ | a ₃₃ |

Таблица 4

| № варианта | Уравнение | | | № варианта | Уравнение | | |
|------------|-----------|----------|----------|------------|-----------|----------|----------|
| | y1 | y2 | y3 | | y1 | y2 | y3 |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 0 | y11 | y21 | y31 | 50 | y13 | y23 | y33 |
| 1 | y11 | y21 | y32 | 51 | y13 | y23 | y34 |
| 2 | y11 | y21 | y33 | 52 | y13 | y24 | y31 |
| 3 | y11 | y21 | y34 | 53 | y13 | y24 | y32 |

| | | | | | | | |
|----|-----|-----|-----|----|-----|-----|-----|
| 4 | y11 | y22 | y31 | 54 | y13 | y24 | y33 |
| 5 | y11 | y22 | y32 | 55 | y13 | y24 | y34 |
| 6 | y11 | y22 | y33 | 56 | y13 | y25 | y31 |
| 7 | y11 | y22 | y34 | 57 | y13 | y25 | y32 |
| 8 | y11 | y23 | y31 | 58 | y13 | y25 | y33 |
| 9 | y11 | y23 | y32 | 59 | y13 | y25 | y34 |
| 10 | y11 | y23 | y33 | 60 | y14 | y21 | y31 |
| 11 | y11 | y23 | y34 | 61 | y14 | y21 | y32 |
| 12 | y11 | y24 | y31 | 62 | y14 | y21 | y33 |
| 13 | y11 | y24 | y32 | 63 | y14 | y21 | y34 |
| 14 | y11 | y24 | y33 | 64 | y14 | y22 | y31 |
| 15 | y11 | y24 | y34 | 65 | y14 | y22 | y32 |
| 16 | y11 | y25 | y31 | 66 | y14 | y22 | y33 |
| 17 | y11 | y25 | y32 | 67 | y14 | y22 | y34 |
| 18 | y11 | y25 | y33 | 68 | y14 | y23 | y31 |
| 19 | y11 | y25 | y34 | 69 | y14 | y23 | y32 |
| 20 | y12 | y21 | y31 | 70 | y14 | y23 | y33 |
| 21 | y12 | y21 | y32 | 71 | y14 | y23 | y34 |
| 22 | y12 | y21 | y33 | 72 | y14 | y24 | y31 |
| 23 | y12 | y21 | y34 | 73 | y14 | y24 | y32 |
| 24 | y12 | y22 | y31 | 74 | y14 | y24 | y33 |
| 25 | y12 | y22 | y32 | 75 | y14 | y24 | y34 |
| 26 | y12 | y22 | y33 | 76 | y14 | y25 | y31 |
| 27 | y12 | y22 | y34 | 77 | y14 | y25 | y32 |
| 28 | y12 | y23 | y31 | 78 | y14 | y25 | y33 |
| 29 | y12 | y23 | y32 | 79 | y14 | y25 | y34 |
| 30 | y12 | y23 | y33 | 80 | y15 | y21 | y31 |
| 31 | y12 | y23 | y34 | 81 | y15 | y21 | y32 |
| 32 | y12 | y24 | y31 | 82 | y15 | y21 | y33 |
| 33 | y12 | y24 | y32 | 83 | y15 | y21 | y34 |
| 34 | y12 | y24 | y33 | 84 | y15 | y22 | y31 |
| 35 | y12 | y24 | y34 | 85 | y15 | y22 | y32 |
| 36 | y12 | y25 | y31 | 86 | y15 | y22 | y33 |

| | | | | | | | |
|----|-----|-----|-----|----|-----|-----|-----|
| 37 | y12 | y25 | y32 | 87 | y15 | y22 | y34 |
| 38 | y12 | y25 | y33 | 88 | y15 | y23 | y31 |
| 39 | y12 | y25 | y34 | 89 | y15 | y23 | y32 |
| 40 | y13 | y21 | y31 | 90 | y15 | y23 | y33 |
| 41 | y13 | y21 | y32 | 91 | y15 | y23 | y34 |
| 42 | y13 | y21 | y33 | 92 | y15 | y24 | y31 |
| 43 | y13 | y21 | y34 | 93 | y15 | y24 | y32 |
| 44 | y13 | y22 | y31 | 94 | y15 | y24 | y33 |
| 45 | y13 | y22 | y32 | 95 | y15 | y24 | y34 |
| 46 | y13 | y22 | y33 | 96 | y15 | y25 | y31 |
| 47 | y13 | y22 | y34 | 97 | y15 | y25 | y32 |
| 48 | y13 | y23 | y31 | 98 | y15 | y25 | y33 |
| 49 | y13 | y23 | y32 | 99 | y15 | y25 | y34 |

Эконометрическая модель содержит три уравнения.

Количество эндогенных переменных (y), экзогенных переменных (x) и вид уравнения определяются вариантом контрольной работы (таблицы 3 и 4).

Например, для варианта №1 (номер зачетной книжки заканчивается на 01) формируется система уравнений, содержащая уравнения y_{11} (1-ый вариант, соответствующий уравнению y_1), y_{21} (1-ый вариант, соответствующий уравнению y_2), y_{32} (2-ой вариант, соответствующий уравнению y_3) (см. таблицу 4). В результате из таблицы 3 формируем новую таблицу 5 коэффициентов при переменных, в соответствии с вариантом:

Таблица 5

| | | | | | |
|----------|----------|----------|----------|----------|----------|
| | y_2 | y_3 | x_1 | x_2 | x_3 |
| y_{11} | 0 | 0 | a_{11} | a_{21} | a_{31} |
| | y_1 | y_3 | x_1 | x_2 | x_3 |
| y_{21} | b_{12} | b_{32} | 0 | 0 | a_{32} |
| | y_1 | y_2 | x_1 | x_2 | x_3 |
| y_{32} | b_{13} | 0 | 0 | a_{23} | a_{33} |

Таким образом, окончательно система уравнений, соответствующая варианту 01, примет вид:

$$y_1 = a_{11} \cdot x_1 + a_{21} \cdot x_2 + a_{31} \cdot x_3$$

$$y_2 = b_{12} \cdot y_1 + b_{32} \cdot y_3 + a_{32} \cdot x_3$$

$$y_3 = b_{13} \cdot y_1 + a_{23} \cdot x_2 + a_{33} \cdot x_3$$

Тема 4. ВРЕМЕННЫЕ РЯДЫ В ЭКОНОМЕТРИЧЕСКИХ ИССЛЕДОВАНИЯХ

Большинство эконометрических моделей строится как динамические эконометрические модели. Это означает, что моделирование причинно-следственных связей между переменными осуществляется во времени, а исходные данные представлены в форме временных рядов.

Временной ряд x_t ($t=1;n$) – ряд значений какого-либо показателя за несколько последовательных промежутков времени.

Каждый временной ряд x_t складывается из следующих основных составляющих (компонентов):

1) Тенденции, характеризующей общее направление динамики изучаемого явления. Аналитически тенденция выражается некоторой функцией времени, называемой трендом (T).

2) Циклической или периодической составляющей, характеризующей циклические или периодические колебания изучаемого явления. Колебания представляют собой отклонения фактических уровней ряда от тренда. Объем продаж некоторых товаров подвержен сезонным колебаниям. Сезонные колебания (S) – периодические колебания, которые имеют определенный и постоянный

период равный годовому промежутку. Конъюнктурные колебания (K) связаны с большими экономическими циклами, период таких колебаний – несколько лет.

3) Случайной составляющей, которая является результатом воздействия множества случайных факторов (E).

Тогда уровень ряда можно представить как функцию от этих составляющих (компонентов): $\tilde{x} = f(T, K, S, E)$.

В зависимости от взаимосвязи между составляющими может быть построена либо аддитивная модель: $\tilde{x} = T + K + S + E$, либо мультипликативная модель: $\tilde{x} = T \cdot K \cdot S \cdot E$ ряда динамики.

Для **определения состава компонентов (структуры временного ряда)** в модели временного ряда строят автокорреляционную функцию.

Автокорреляция – корреляционная связь между последовательными уровнями одного и того же ряда динамики (сдвинутыми на определенный промежуток времени L - лаг). То есть, автокорреляция - это связь между рядом: x_1, x_2, \dots, x_{n-l} и рядом $x_{1+l}, x_{2+l}, \dots, x_n$, где L - положительное целое число. Автокорреляция может быть измерена коэффициентом автокорреляции:

$$r_{t,t-L} = \frac{\overline{x_t \cdot x_{t-L}} - \overline{x_t} \cdot \overline{x_{t-L}}}{\sigma_t \cdot \sigma_{t-L}},$$

$$\text{где } \overline{x_t \cdot x_{t-L}} = \frac{\sum_{i=1+L}^n x_i \cdot x_{i-L}}{n-L},$$

$$\overline{x_t} = \frac{\sum_{i=1+L}^n x_i}{n-L} \text{ — средний уровень ряда } (x_{1+L}, x_{2+L}, \dots, x_n),$$

$$\overline{x_{t-L}} = \frac{\sum_{i=1+L}^n x_{i-L}}{n-L} \text{ — средний уровень ряда } (x_1, x_2, \dots, x_{n-L}),$$

σ_t , σ_{t-L} — средние квадратические отклонения, для рядов $(x_{1+L}, x_{2+L}, \dots, x_n)$ и $(x_1, x_2, \dots, x_{n-L})$ соответственно.

Лаг (сдвиг во времени) определяет порядок коэффициента автокорреляции. Если $L=1$, то имеем коэффициент автокорреляции 1-ого порядка $r_{t,t-1}$, если $L=2$, то коэффициент автокорреляции 2-ого порядка $r_{t,t-2}$ и т.д. Следует учитывать, что с увеличением лага на единицу, число пар значений, по которым рассчитывается коэффициент автокорреляции уменьшается на 1. Поэтому обычно рекомендуют максимальный порядок коэффициента автокорреляции равный $n/4$.

Рассчитав несколько коэффициентов автокорреляции, можно определить лаг (L), при котором автокорреляция ($r_{t,t-L}$) наиболее высокая, выявив тем самым *структуру временного ряда*. Если наиболее высоким оказывается значение $r_{t,t-1}$, то исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался $r_{t,t-L}$, то ряд содержит колебания периодом L . Если ни один из $r_{t,t-L}$ не является значимым, можно сделать одно из двух предположений:

- либо ряд не содержит тенденции и циклических колебаний, а его уровень определяется только случайной компонентой;

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

Последовательность коэффициентов автокорреляции 1, 2 и т.д. порядков называют автокорреляционной функцией временного ряда. График зависимости значений коэффициентов автокорреляции от величины лага (порядка коэффициента автокорреляции) называют *коррелограммой*.

Для выявления закономерных колебаний внутри года при выполнении контрольной работы рекомендуется

рассчитывать не меньше 4-х уровней коэффициентов автокорреляции.

Рассмотрим на примере как построить коррелограмму, чтобы определяется структуру временного ряда.

Пусть нам даны поквартальные данные об объеме выпуска некоторого товара некоторой фирмой x (усл.ед.) за 3 года:

| 1993 | | | | 1994 | | | | 1995 | | | |
|------|-----|-----|-----|------|-----|-----|-----|------|-----|------|-----|
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| 410 | 560 | 715 | 500 | 520 | 740 | 975 | 670 | 705 | 950 | 1200 | 900 |

Чтобы построить коррелогорамму для нашего примера, исходный ряд динамики дополним рядами из уровней этого ряда, сдвинутыми во времени (таблица 6).

Таблица 6

| t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
|-----------|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|------|-------------------|
| x_t | - | 560 | 715 | 500 | 520 | 740 | 975 | 670 | 705 | 950 | 1200 | 900 | $r_{t,t}$ |
| x_{t-1} | - | 410 | 560 | 715 | 500 | 520 | 740 | 975 | 670 | 705 | 950 | 1200 | $r_{t,t-1}=0,537$ |
| x_t | - | - | 715 | 500 | 520 | 740 | 975 | 670 | 705 | 950 | 1200 | 900 | $r_{t,t}$ |
| x_{t-2} | - | - | 410 | 560 | 715 | 500 | 520 | 740 | 975 | 670 | 705 | 950 | $r_{t,t-2}=0,085$ |
| x_t | - | - | - | 500 | 520 | 740 | 975 | 670 | 705 | 950 | 1200 | 900 | $r_{t,t}$ |
| x_{t-3} | - | - | - | 410 | 560 | 715 | 500 | 520 | 740 | 975 | 670 | 705 | $r_{t,t-3}=0,445$ |
| x_t | - | - | - | - | 520 | 740 | 975 | 670 | 705 | 950 | 1200 | 900 | $r_{t,t}$ |
| x_{t-4} | - | - | - | - | 410 | 560 | 715 | 500 | 520 | 740 | 975 | 670 | $r_{t,t-4}=0,990$ |
| x_t | - | - | - | - | - | 740 | 975 | 670 | 705 | 950 | 1200 | 900 | $r_{t,t}$ |
| x_{t-5} | - | - | - | - | - | 410 | 560 | 715 | 500 | 520 | 740 | 975 | $r_{t,t-5}=0,294$ |

Рассчитаем коэффициенты корреляции:

1-ого порядка для рядов x_t и x_{t-1} ,

2-ого порядка для рядов x_t и x_{t-2} ,

3-его порядка для рядов x_t и x_{t-3} ,

4-ого порядка для рядов x_t и x_{t-4} ,

5-ого порядка для рядов x_t и x_{t-5}

Результаты расчетов представлены в таблице 7.

Таблица 7

| Лаг (порядок) – L | $r_{t,t-L}$ | Коррелограмма |
|------------------------|-------------|---------------|
| 1 | 0,537 | ***** |
| 2 | 0,085 | * |
| 3 | 0,445 | *** |
| 4 | 0,990 | ***** |
| 5 | 0,294 | ** |

Вывод: в данном ряду динамики имеется тенденция (т.к. $r_{t,t-1}=0,537 \rightarrow 1$) и периодические колебания с периодом (L) равным 4, т.е. имеют место сезонные колебания (т.к. $r_{t,t-4}=0,99 \rightarrow 1$).

Построение модели временного ряда с сезонными колебаниями (аддитивная модель).

Процесс построения модели временного ряда (x), содержащего n уровней некоторого показателя за Z лет, с L сезонными колебаниями включает следующие шаги:

1) *Выравнивание исходного ряда методом скользящей средней (x^c).* Произведем выравнивание исходного ряда взятого из примера, рассмотренного выше, методом скользящей средней с периодом усреднения равным 3. Результаты представлены в таблице 9 (столбец 4).

2) Расчет значений сезонной составляющей S_i , $i=1;L$, где L – число сезонов в году. Для нашего примера $L=4$ (сезоны - кварталы).

Расчет значений сезонных составляющих осуществляется после устранения тенденции из исходных уровней ряда: $x - x^c$ (столбец 5, таблица 9). Для дальнейшего расчета S_i построим отдельную таблицу. Строки данной таблицы соответствуют сезонам, столбцы - годам. В теле таблицы находятся значения: $x - x^c$. По этим данным рассчитываются средние оценки сезонных составляющих каждой строке (S_i^c). Если сумма всех средних оценок равна нулю ($\sum_{i=1}^L S_i^c = 0$), то данные

средние и будут окончательными значениями сезонных составляющих ($S_i=S_i^c$). Если их сумма не равна нулю, то рассчитываются скорректированные значения сезонных составляющих вычитанием из средней оценки величины равной отношению суммы средних оценок к их общему

числу ($S_i = S_i^c - \frac{\sum_{i=1}^L S_i^c}{L}$). Для нашего примера расчет

значений S_i представлен в таблице 8.

Таблица 8

| Номер сезона | Год 1 | Год 2 | Год 3 | Средняя оценка сезонной составляющей | Скорректированная оценка сезонной составляющей S_i |
|--------------|--------|---------|--------|--------------------------------------|--|
| 1 | - | -66,67 | -70,00 | -68,33 | -67,15 |
| 2 | -1,67 | -5,00 | -1,67 | -2,78 | -1,60 |
| 3 | 123,33 | 180,00 | 183,33 | 162,22 | 163,40 |
| 4 | -78,33 | -113,33 | - | -95,83 | -94,66 |
| Итого | | | | -4,72 | 0 |

3) Устранение влияния сезонной составляющей из исходного ряда динамики : $x^S = x - S_i$. Результаты расчета x^S для нашего примера представлены в столбце 6 таблицы 9.

4) Аналитическое выравнивание уровней x^S (построение тренда): $T = \tilde{a} + \tilde{b} \cdot t$.

Расчет параметров при аналитическом выравнивании чаще всего производится с помощью метода наименьших квадратов (МНК). При этом поиск параметров для линейного уравнения тренда можно упростить, если отсчет времени производить так, чтобы сумма показателей времени изучаемого ряда динамики была равна нулю. Для этого вводится новая условная переменная времени t^y , такая, что $\sum t^y = 0$. Уравнение тренда при этом будет следующим: $T = \tilde{a} + \tilde{b} \cdot t^y$.

При нечетном числе уровней ряда динамики для получения $\sum t^y = 0$ уровень, находящийся в середине ряда, принимается за условное начало отсчета времени (периоду

или моменту времени, соответствующему данному уровню присваивается нулевое значение). Даты времени, расположенные левее этого уровня, обозначаются натуральными числами со знаком минус (-1 -2 -3 ...), а даты времени, расположенные правее этого уровня – натуральными числами со знаком плюс (1 2 3 ...).

Если число уровней ряда четное, периоды времени левой половины ряда (до середины) нумеруются -1, -3, -5 и т.д. А периоды правой половины - +1, +3, +5 и т.д. При этом $\sum t^y$ будет равна 0.

Система нормальных уравнений (соответствующих МНК) преобразуется к виду:

$$\left\{ \begin{array}{l} \sum x^S = \tilde{a} \cdot n \\ \sum x^S \cdot t^y = \tilde{b} \sum (t^y)^2 \end{array} \right.$$

Отсюда параметры уравнения рассчитываются по формулам:

$$\tilde{b} = \frac{\sum x^S \cdot t^y}{\sum (t^y)^2}, \quad \tilde{a} = \frac{\sum x^S}{n}.$$

Интерпретация параметров линейного уравнения тренда $T = \tilde{a} + \tilde{b} \cdot t^y$:

\tilde{a} - уровень ряда за период времени $t^y=0$;

\tilde{b} - средний абсолютный прирост уровня ряда за единичный промежуток времени.

В нашем примере четное число уровней ряда: $n=12$. Следовательно, условная переменная времени для 6-ого элемента ряда будет равна -1 , а для 7-ого $+1$. Значения переменной t^y содержатся во 2-ом столбце таблицы 9.

Параметры линейного тренда будут: $\tilde{b} = 14257,5/572=24,93$; $\tilde{a} = 8845/12=737,08$. Это значит, что с каждым кварталом объем выпуска товара в среднем увеличивается на $2 \cdot 28,7$ усл.ед. А средний за период с 1993 по 1995гг объем выпуска составил $738,75$ усл.ед.

Рассчитаем значения трендовой компоненты по формуле $T = 737,08 + 24,93 \cdot t^y$ (столбец 7 таблицы 9).

5) Учет сезонной составляющей в выровненных уровнях ряда ($\tilde{x} = T+S$). Результаты расчета для нашего примера представлены в столбце 8 таблицы 9.

6) Расчет абсолютной ошибки временного ряда ($E=x-\tilde{x}$) осуществляется для оценки качества полученной

модели. Результаты расчета для нашего примера представлены в столбце 9 таблицы 9.

Таблица 9

| T | t^p | x | x^c | $x - x^c$ | x^s | T | \tilde{x} | E |
|------------|----------|----------|----------|-----------|----------|----------|-------------|----------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | -11 | 410 | - | - | 477,15 | 462,90 | 395,75 | 14,25 |
| 2 | -9 | 560 | 561,67 | -1,67 | 561,60 | 512,75 | 511,15 | 48,85 |
| 3 | -7 | 715 | 591,67 | 123,33 | 551,60 | 562,60 | 726,00 | -11,01 |
| 4 | -5 | 500 | 578,33 | -78,33 | 594,65 | 612,45 | 517,80 | -17,80 |
| 5 | -3 | 520 | 586,67 | -66,67 | 587,15 | 662,31 | 595,15 | -75,15 |
| 6 | -1 | 740 | 745,00 | -5,00 | 741,60 | 712,16 | 710,56 | 29,44 |
| 7 | 1 | 975 | 795,00 | 180,00 | 811,60 | 762,00 | 925,41 | 49,59 |
| 8 | 3 | 670 | 783,33 | -113,33 | 764,65 | 811,86 | 717,21 | -47,21 |
| 9 | 5 | 705 | 775,00 | -70,00 | 772,15 | 861,71 | 794,56 | -89,56 |
| 10 | 7 | 950 | 951,67 | -1,67 | 951,60 | 911,56 | 909,97 | 40,03 |
| 11 | 9 | 1200 | 1016,67 | 183,33 | 1036,60 | 961,41 | 1124,82 | 75,18 |
| 12 | 11 | 900 | - | - | 994,65 | 1011,27 | 916,61 | -16,61 |
| Итого 0 | | 8845 | | | 8845,00 | 8845,00 | 8845,00 | 16,61 |

Значимость параметров линейного уравнения тренда (T) определяется на основе t -критерия Стьюдента также как и в линейном парном регрессионном анализе.

Прогнозирование по аддитивной модели.

Пусть требуется дать прогноз уровня временного ряда на период $(n+1)$. Точечный прогноз значения уровня временного ряда x_{n+1} в аддитивной модели есть сумма трендовой компоненты и сезонной компоненты (соответствующей i -ому сезону прогноза): $\tilde{x}_{n+1}^p = T_{n+1} + S_i$.

Для построения доверительного интервала прогноза нужно рассчитать среднюю ошибку прогноза:

$$\mu^p = \sqrt{\left(\frac{\sum (x - \tilde{x})^2}{n - h} \left(1 + \frac{1}{n} + \frac{(t^{yp})^2}{\sum (t^y)^2} \right) \right)},$$

где h - число параметров в уравнении тренда;

t^{yp} – значение условной переменной времени для периода прогнозирования.

Затем рассчитаем предельную ошибку прогноза: $\Delta^p = t_a \cdot \mu_p$,

где t_a - коэффициент доверия, определяемый по таблицам Стьюдента по уровню значимости α и числу степеней свободы равным $(n-h)$.

Окончательно получим: $(\tilde{x}_{n+1}^p - \Delta^p; \tilde{x}_{n+1}^p + \Delta^p)$.

Задание № 4

На основе данных, приведенных в таблице 10 и соответствующих Вашему варианту (таблица 11), постройте модель временного ряда. Для этого требуется:

1. Построить коррелограмму и определить имеет ли ряд тенденцию и сезонные колебания.
2. Провести сглаживание ряда скользящей средней и рассчитать значения сезонной составляющей.

3. Построить уравнения тренда и сделать выводы.

4. На основе полученной модели сделать прогноз на следующие два квартала с учетом выявленной сезонности.

Таблица 10

Основные показатели развития производственной
фирмы
за период с 2002 по 2007 гг. (по сопоставимой
оценке)

| № набл юде ния | Год | Кварт ал | Объем произво дства продук ции, млн.руб . | Среднегод овая стоимость основных производс твенных фондов, млн.руб. | Средн е годов ая числе нность ь ППП, чел., | Дебито рская задолж енность , млн.ру б. | Среднего довая стоимост ь оборотны х средств, млн.руб. | Балан совая прибы ль, млн. руб. | Чиста я прибы ль, млн. руб. |
|----------------------|-------------|-------------|---|---|---|---|---|--|--|
| А | Б | В | Г | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 2002 | 1 | 1065 | 1062 | 713 | 25 | 837 | 94 | 36 |
| 2 | | 2 | 851 | 682 | 507 | 27 | 685 | 78 | 27 |
| 3 | | 3 | 531 | 726 | 361 | 34 | 837 | 87 | 22 |
| 4 | | 4 | 922 | 1153 | 557 | 44 | 1161 | 75 | 29 |
| 5 | 2003 | 1 | 1095 | 1213 | 607 | 42 | 1151 | 84 | 34 |
| 6 | | 2 | 986 | 898 | 598 | 39 | 822 | 63 | 28 |
| 7 | | 3 | 822 | 794 | 368 | 48 | 1383 | 86 | 30 |
| 8 | | 4 | 1137 | 1441 | 646 | 60 | 884 | 82 | 35 |
| 9 | 2004 | 1 | 1301 | 1600 | 693 | 63 | 1309 | 78 | 40 |
| 10 | | 2 | 1038 | 967 | 718 | 40 | 1028 | 72 | 33 |
| 11 | | 3 | 780 | 1246 | 363 | 48 | 1771 | 84 | 33 |
| 12 | | 4 | 1435 | 1458 | 639 | 71 | 1310 | 102 | 40 |
| 13 | 2005 | 1 | 1593 | 1412 | 708 | 87 | 1372 | 112 | 36 |
| 14 | | 2 | 1658 | 891 | 614 | 65 | 1272 | 92 | 27 |
| 15 | | 3 | 1363 | 1061 | 348 | 67 | 1821 | 99 | 30 |
| 16 | | 4 | 1737 | 1287 | 636 | 76 | 1571 | 113 | 36 |
| 17 | 2006 | 1 | 1719 | 1635 | 825 | 101 | 1758 | 95 | 36 |

| | | | | | | | | | |
|----|------|---|------|------|-----|----|------|-----|----|
| 18 | | 2 | 1521 | 1166 | 622 | 84 | 1505 | 79 | 28 |
| 19 | | 3 | 1049 | 1230 | 514 | 73 | 2109 | 112 | 28 |
| 20 | | 4 | 1790 | 1514 | 703 | 93 | 1787 | 116 | 28 |
| 21 | 2007 | 1 | 2016 | 1642 | 797 | 96 | 2197 | 90 | 39 |

Таблица 11

Номера наблюдений и показатель,
соответствующие варианту контрольной работы

| Номер варианта | Номер начального наблюдения | Номер конечного наблюдения | Номер показателя из табл. 4 | Номер варианта | Номер начального наблюдения | Номер конечного наблюдения | Номер показателя из табл.4 |
|----------------|-----------------------------|----------------------------|-----------------------------|----------------|-----------------------------|----------------------------|----------------------------|
| <i>1</i> | <i>2</i> | <i>3</i> | <i>4</i> | <i>5</i> | <i>6</i> | <i>7</i> | <i>8</i> |
| 1 | 1 | 12 | 1 | 51 | 1 | 12 | 6 |
| 2 | 2 | 13 | 1 | 52 | 2 | 13 | 6 |
| 3 | 3 | 14 | 1 | 53 | 3 | 14 | 6 |
| 4 | 4 | 15 | 1 | 54 | 4 | 15 | 6 |
| 5 | 5 | 16 | 1 | 55 | 5 | 16 | 6 |
| 6 | 6 | 17 | 1 | 56 | 6 | 17 | 6 |
| 7 | 7 | 18 | 1 | 57 | 7 | 18 | 6 |
| 8 | 8 | 19 | 1 | 58 | 8 | 19 | 6 |
| 9 | 9 | 20 | 1 | 59 | 9 | 20 | 6 |
| 10 | 10 | 21 | 1 | 60 | 10 | 21 | 6 |
| 11 | 1 | 12 | 2 | 61 | 1 | 12 | 7 |
| 12 | 2 | 13 | 2 | 62 | 2 | 13 | 7 |
| 13 | 3 | 14 | 2 | 63 | 3 | 14 | 7 |
| 14 | 4 | 15 | 2 | 64 | 4 | 15 | 7 |
| 15 | 5 | 16 | 2 | 65 | 5 | 16 | 7 |
| 16 | 6 | 17 | 2 | 66 | 6 | 17 | 7 |
| 17 | 7 | 18 | 2 | 67 | 7 | 18 | 7 |

| | | | | | | | |
|----|----|----|---|----|----|----|---|
| 18 | 8 | 19 | 2 | 68 | 8 | 19 | 7 |
| 19 | 9 | 20 | 2 | 69 | 9 | 20 | 7 |
| 20 | 10 | 21 | 2 | 70 | 10 | 21 | 7 |
| 21 | 1 | 12 | 3 | 71 | 1 | 12 | 1 |
| 22 | 2 | 13 | 3 | 72 | 2 | 13 | 2 |
| 23 | 3 | 14 | 3 | 73 | 3 | 14 | 3 |
| 24 | 4 | 15 | 3 | 74 | 4 | 15 | 4 |
| 25 | 5 | 16 | 3 | 75 | 5 | 16 | 5 |
| 26 | 6 | 17 | 3 | 76 | 6 | 17 | 6 |
| 27 | 7 | 18 | 3 | 77 | 7 | 18 | 7 |
| 28 | 8 | 19 | 3 | 78 | 8 | 19 | 1 |
| 29 | 9 | 20 | 3 | 79 | 9 | 20 | 2 |
| 30 | 10 | 21 | 3 | 80 | 10 | 21 | 3 |
| 31 | 1 | 12 | 4 | 81 | 1 | 12 | 4 |
| 32 | 2 | 13 | 4 | 82 | 2 | 13 | 5 |
| 33 | 3 | 14 | 4 | 83 | 3 | 14 | 6 |
| 34 | 4 | 15 | 4 | 84 | 4 | 15 | 7 |
| 35 | 5 | 16 | 4 | 85 | 5 | 16 | 1 |
| 36 | 6 | 17 | 4 | 86 | 6 | 17 | 2 |
| 37 | 7 | 18 | 4 | 87 | 7 | 18 | 3 |
| 38 | 8 | 19 | 4 | 88 | 8 | 19 | 4 |
| 39 | 9 | 20 | 4 | 89 | 9 | 20 | 5 |
| 40 | 10 | 21 | 4 | 90 | 10 | 21 | 6 |
| 41 | 1 | 12 | 5 | 91 | 1 | 12 | 7 |
| 42 | 2 | 13 | 5 | 92 | 2 | 13 | 1 |
| 43 | 3 | 14 | 5 | 93 | 3 | 14 | 2 |
| 44 | 4 | 15 | 5 | 94 | 4 | 15 | 3 |
| 45 | 5 | 16 | 5 | 95 | 5 | 16 | 4 |
| 46 | 6 | 17 | 5 | 96 | 6 | 17 | 5 |
| 47 | 7 | 18 | 5 | 97 | 7 | 18 | 6 |
| 48 | 8 | 19 | 5 | 98 | 8 | 19 | 7 |
| 49 | 9 | 20 | 5 | 99 | 9 | 20 | 1 |
| 50 | 10 | 21 | 5 | 0 | 10 | 21 | 2 |

4 ПЕРЕЧЕНЬ ОСНОВНОЙ И ДОПОЛНИТЕЛЬНОЙ УЧЕБНОЙ ЛИТЕРАТУРЫ, НЕОБХОДИМОЙ ДЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

4.1 Основная литература:

1. **Статистика** [Текст] : учебник и практикум для академического бакалавриата : учебник для студентов вузов, обучающихся по экономическим направлениям и специальностям : учебник для студентов вузов, обучающихся по направлению "Статистика" и другим экономическим специальностям / [В. С. Мхитарян и др.] ; под ред. В. С. Мхитаряна ; Нац. исслед. ун-т "Высшая школа экономики". - Москва : Юрайт, 2016. - 464 с. - (Бакалавр. Академический курс). - Библиогр.: с. 464. - ISBN 978-5-9916-5584-2
2. *Васильев, А. А.* Теория вероятностей и математическая статистика : учебник и практикум для академического бакалавриата / А. А. Васильев. — 2-е изд., испр. и доп. — М. : Издательство Юрайт, 2017. — 253 с. — (Серия : Бакалавр. Академический курс). ЭБС. URL: <https://biblio-online.ru/viewer/3F13A609-9D28-44A2-A070-1A025A293A4F#page/1>
3. Анализ данных : учебник для академического бакалавриата / В. С. Мхитарян [и др.] ; под ред. В. С. Мхитаряна. — М. : Издательство Юрайт, 2017. — 490 с. — (Серия : Бакалавр. Академический курс). — ISBN 978-5-534-00616-2. ЭБС. URL <https://biblio-online.ru/viewer/CC38E97A-CCE5-4470-90F1-3B6D35ACC0B4#page/1>
4. *Миркин, Б. Г.* Введение в анализ данных : учебник и практикум / Б. Г. Миркин. — М. : Издательство Юрайт, 2017. — 174 с. — (Серия : Авторский учебник). — ISBN 978-5-534-03762-3. ЭБС. URL <https://www.biblio-online.ru/viewer/46A41F93-BC46->

4.2 Дополнительная литература:

1. Болотова, Л. С. Системы поддержки принятия решений в 2 ч. Часть 1 : учебник и практикум для академического бакалавриата / Л. С. Болотова ; отв. ред. В. Н. Волкова, Э. С. Болотов. — М. : Издательство Юрайт, 2017. — 279 с.

2. **Статистика** [Текст] : учебник для студентов бакалавриата, обучающихся по направлению подготовки "Экономика" / [Л. И. Ниворожкина и др.] ; под общ. ред. Л. И. Ниворожкиной. - 2-е изд., доп. и перераб. - Москва : Дашков и К°, 2013.

3. Ивченко, Юлия Сергеевна. **Статистика** [Текст] : учебное пособие для студентов / Ю. С. Ивченко. - М. : РИОР [и др.], 2011.

4.3. Периодические издания:

1. "Алгебра и логика" / Институт математики им.Соболева СО РАН /Периодичность – 6 раз в год/ сайт: http://elibrary.ru/title_about.asp?id=7311/

4.4 Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля).

1. Российское образование, федеральный портал [Официальный сайт] — URL: <http://www.edu.ru>

2. Образовательный портал «Учеба» [Официальный сайт] URL: <http://www.uceba.com/>

3. Портал «Российское образование» [Официальный сайт] URL: <http://www.edu.ru/>

4. Единое окно доступа к образовательным ресурсам «Единое окно» [Официальный сайт] URL: <http://window.edu.ru/>

5. Федеральная университетская компьютерная сеть России [Официальный сайт] URL: <http://www.runnet.ru/>
6. Служба тематических толковых словарей [Официальный сайт] URL: <http://www.glossary.ru/>
7. Образовательный портал [Официальный сайт] URL: «Академик» <http://dic.academic.ru/>
8. Web of Sciense (архив с 2002 года) рефераты [Официальный сайт] URL: <http://webofknowledge.com>.
9. Лекториум “(Минобрнауки РФ) единая Интернет-библиотека лекций [Официальный сайт] URL <http://www.lektorium.tv/>
10. Электронный архив документов КубГУ полнотекстов [Официальный сайт] URL: <http://docspace.kubsu.ru>