МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«РОССИЙСКИЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ
ИМ.Г.В.ПЛЕХАНОВА»
Оренбургский филиал ФГБОУ ВПО « РЭУ им. Г.В.Плеханова»

Лаптева Е.В., Золотова Л.В.

Статистические методы оценки принятия управленческих решений

Учебное пособие

Оренбург
2015

ББК 311
УДК 60.6
Л 24

Одобрено и рекомендовано к изданию Советом Оренбургского филиала ФГБОУ ВПО «РЭУ им. Г.В.Плеханова». Протокол № 9 от 28.01.2015 г.

РЕЦЕНЗЕНТЫ:

Л.В.Портнова, кандидат экономических наук, доцент кафедры финансов и кредита Оренбургского филиала ФГБОУ ВПО «РЭУ им. Г.В.Плеханова»;
С.В.Хабарова, кандидат экономических наук, доцент кафедры статистики и экономического анализа ФГБОУ ВПО «Оренбургский государственный аграрный университет»

Лаптева Е.В., Золотова Л.В.

Л 24 Статистические методы оценки принятия управленческих решений / Е.В.Лаптева. – Оренбург: ООО ИПК «Университет», 2015. – 115с.
ISBN 978-5-9904198-1-0

В учебном пособии в систематизированном виде изложены и представлены такие вопросы как статистическое моделирование и прогнозирование как наука, априорный анализ компонент временного ряда, моделирование тенденции, моделирование периодической компоненты, моделирование случайной компоненты, статистические методы моделирования взаимосвязи, статистические методы прогнозирования динамики, статистические методы моделирования взаимосвязи, эвристические методы прогнозирования социально-экономических явлений.

Учебное пособие предназначено для бакалавров направления «Экономика» высших учебных заведений, преподавателей, аспирантов, научных работников, работников органов государственного и муниципального управления, руководителей организаций.

Работа издана в авторской редакции

ISBN 978-5-9904198-1-0

ОГЛАВЛЕНИЕ

1 Статистическое моделирование и прогнозирование как наука

1.1 Понятие о статистическом моделировании и прогнозировании

Исследование социально-экономических явлений вызывает необходимость практического использования различных статистических методов и приемов для получения адекватной картины развития того или иного процесса, для получения прогнозных значений на краткосрочную перспективу. Область применения статистических методов достаточно широка – начиная с экономики и заканчивая медициной.

Статистическое моделирование и прогнозирование базируется на системе статистических категорий, понятий и методов. Основными статистическими методами являются методы сводки и группировки, статистического наблюдения, обобщающих показателей, вариации, структуры, корреляционно-регрессионного анализа и т.д.

Рассмотрим более подробно понятия статистического моделирования и прогнозирования.

Моделирование - это воспроизведение или имитация поведения какой-либо реально существующей системы на специально построенном ее аналоге или модели.

Статистическая модель – это модель, построенная на том, что вначале делается предположение о характере связей между анализируемыми переменными, затем проверяется соответствие данных модели и в зависимости от степени этого соответствия делаются определенные выводы.

Простейшей формой статистической модели является линейная регрессия. При ее использовании делается предположение о том, что два показателя связаны друг с другом линейно и именно эта гипотеза и проверяется (кроме того делается предположение о том, что одна переменная зависит от другой). Таким образом, статистические модели базируются на двух типах допущений - как и методы тестирования статистических гипотез они предполагают, что данные распределены определенным образом (чаще всего по нормальному закону распределения), и в дополнение к этому делается предположение о характере связи. Поэтому сделать ошибку при использовании статистических моделей в два раза легче и они обычно рассматриваются как инструментарий, требующий дополнительной подготовки в области статистики[1].

[1] Плавинский С.Л. Биостатистика.Планирование, обработка и представление результатов биомедицинских исследований при помощи системы SAS. СПб: Издательский дом СПб МАПО.- 2005.

Прогнозирование – это научное, основанное на системе причинно-следственных связей и закономерностей, выявление состояния и вероятностных путей развития процессов.

В условиях рыночной экономики деятельность организаций в значительной степени зависит от того, насколько достоверно они могут предвидеть перспективы своего развития в будущем, т.е. от прогнозирования.

Прогноз – предположительная оценка будущего состояния организации. Организации используют прогнозы с целью предусмотрения возможных вариантов развития своего бизнеса, прогнозируют будущие события или условия их возникновения.

По оценкам ученых насчитывается более 150 методов прогнозирования. Базовых методов гораздо меньше, многие из «методов» скорее относятся к отдельным способам и процедурам прогнозирования, либо представляют собой набор отдельных приемов, отличающихся от базовых методов количеством частных приемов и последовательностью их применения.

Под **методом прогнозирования** понимается совокупность приемов и способов мышления, позволяющих на основе анализа ретроспективных данных, экзогенных (внешних) и эндогенных (внутренних) связей объекта прогнозирования, а также их измерения в рамках рассматриваемого явления или процесса вывести суждения определенной достоверности относительно будущего развития объект. Более краткое определение дано упоминавшемся Э.Янчем: «способ исследования объекта прогнозирования, направленный на разработку прогнозов».

1.2 Классификация методов статистического моделирования и прогнозирования

Проведение любого экономического исследования начинается с теоретических выкладок и определения основных понятий, но теория находиться в неразрывной взаимосвязи с практическим применением различных методов. В настоящее время в отечественной и зарубежной литературе предлагается более сотни различных приемов и методов моделирования для анализа социально - экономических явлений и процессов.

Метод (с греч. methodos - «путь к чему-либо») – это способ достижения цели, это деятельность, упорядоченная определенным образом.

В общем виде, обобщив и систематизировав накопленный опыт, предлагаем выделить следующие группы методов моделирования (рисунок 1.1).

Все из представленных на рисунке методов в конкретном экономическом исследовании могут быть использованы обособленно либо во взаимосвязи с другими методами. Для получения всестороннего аналитического материала и его дальнейшего практического применения целесообразно использовать методологический инструментарий комплексно и систематизировано.

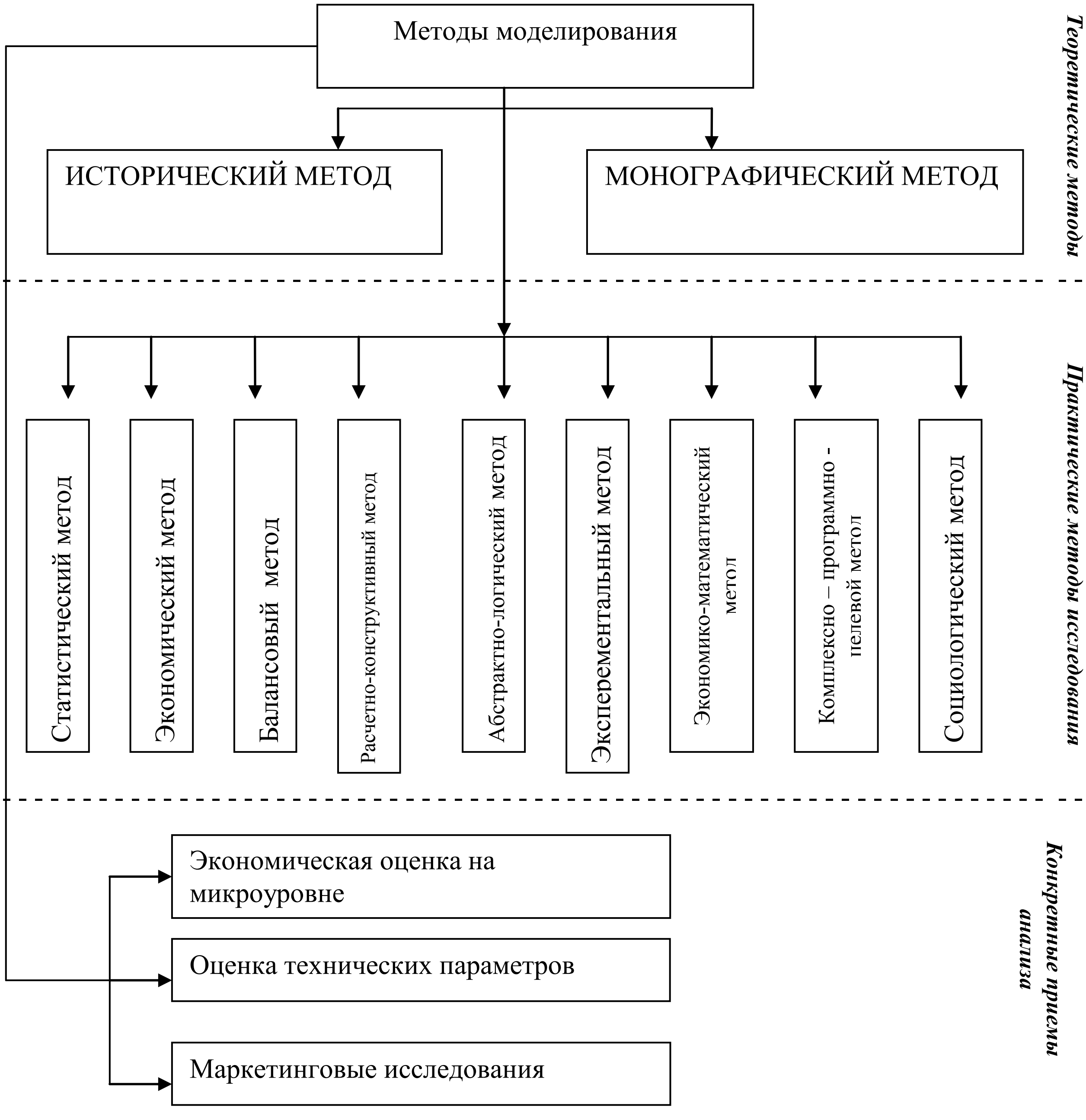

Рисунок 1.1 - Классификация методов моделирования

Методологические подходы построения прогнозов разрабатываются такой областью знаний как прогностика, одним из направлений которой является статистическое прогнозирование, в результате которого получают прогнозы развития того или иного явления.

Прогнозы можно подразделить на разные **типы** в зависимости от:

1) цели исследования:

- поисковый прогноз не ориентируется на заданную цель, а рассматривает возможные направления будущего развития прогнозируемого объекта, т.е. выявление того, как будет развиваться объект в будущем полностью зависит от сохранения существующих тенденций;

- нормативный прогноз – это прогноз, который предназначен для указания возможных путей и сроков достижения заданного конечного состояния прогнозируемого объекта, т.е. н.п. разрабатывается на базе заранее определенных целей и задач.

Таким образом, поисковый прогноз отталкивается при определении будущего от прошлого и настоящего объекта, а нормативный прогноз в обратной последовательности.

2) в зависимости от специфики области применения прогноза и от объекта:

- естественноведческие (биология, медицина);
- научно-технические (инженерия, техника, экономика).

3) в зависимости от масштабности объекта:

- глобальные (мировой масштаб);
- макроэкономические (страна);
- структурные (межотраслевые и межрегиональные);
- региональные (регион);
-отраслевые (отрасль);
- микроэкономические (организации, производства).

4) по сложности:

- сверх простые – прогноз на основе одномерного временного ряда, когда отсутствуют связи между признаками;
- простые – прогнозы, предполагающие учет оценки связей между факторными признаками;
- сложные – прогноз, оценка связей между признаками в котором определяется на основе системы уравнений или многофакторного динамического прогнозирования.

5) по времени упреждения: текущие (до 1 года); краткосрочные (1-3года); среднесрочные (3-5 лет); долгосрочные (5-10 лет); дальнесрочные (10 лет и более).

Вопросы для самопроверки:

1. *Чем моделирование отличается от прогнозирования?*
2. *Дайте определение моделированию.*
3. *Что такое статистическая модель?*
4. *Что такое прогноз? Дайте определение.*
5. *Какие основные виды прогноза вы знаете? В чем их смысл.*
6. *Перечислите типологию прогнозов в зависимости от различных признаков.*
7. *На какие группы можно разделить методы по набору прогнозируемых показателей?*
8. *На какие группы в зависимости от вида используемой модели можно подразделить методы прогнозирования?*

2 Априорный анализ компонент временного ряда

2.1 Ряды динамики. Классификация динамических рядов

Временной ряд (ряд динамики, хронологический ряд, динамический ряд) – это последовательность упорядоченных во времени числовых показателей, характеризующих уровень развития изучаемого явления. Всякий ряд динамики включает, следовательно, два обязательных элемента: во-первых, время и, во-вторых, конкретное значение показателя, или уровень ряда.

Ряды динамики различаются по следующим признакам.

1. **По времени** – моментные и интервальные ряды.

Интервальный ряд динамики – последовательность, в которой уровень явления относится к результату, накопленному или вновь произведенному за определенный интервал времени. Таковы, например, ряды показателей объема продукции по месяцам года, количества отработанных человеко-дней по отдельным периодам и т.д. Если же уровень ряда показывает фактическое наличие изучаемого явления в конкретный момент времени, то совокупность уровней образует *моментный ряд динамики.* Примерами моментных рядов могут быть последовательности показателей численности населения на начало года, величины запаса какого-либо материала на начало периода и т.д. Важное аналитическое отличие моментных рядов от интервальных состоит в том, что сумма уровней интервального ряда дает вполне реальный показатель – общий выпуск продукции за год, общие затраты рабочего времени, общий объем продаж акций и т.д., сумма же уровней моментного ряда, хотя иногда и подсчитывается, но реального содержания, как правило, не имеет.

2. **По форме представления уровней –** ряды абсолютных, относительных и средних величин.

3. **По расстоянию между датами или интервалам времени** выделяют полные и неполные хронологические ряды.

Полные ряды динамики имеют место, когда даты регистрации или окончания периодов следуют друг за другом с равными интервалами. Это равноотстоящие ряды динамики. *Неполные* – когда принцип равных интервалов не соблюдается.

Чтобы о развитии явления можно было получить представление при помощи числовых уровней, при составлении ряда динамики должны приводиться в сопоставительный вид.

Статистические данные должны быть сопоставимы по территории, кругу охватываемых объектов, единицам измерения, времени регистрации, ценам, методологии расчета. *Сопоставимость по территории* означает, что данные по странам и регионам, границы которых изменились, должны быть пересчитаны в старых пределах. *Сопоставимость по кругу охватываемых* *объектов* означает сравнение совокупностей с равным числом элементов. Территориальная и объемная сопоставимость обеспечивается смыканием рядов динамики, при этом либо абсолютные уровни заменяются относительными, либо делается пересчет в условные абсолютные уровни. Не возникает особых сложностей при обеспечении *сопоставимости* данных *по единицам измерения; стоимостная сравнимость* достигается системой сопоставимых цен.

Числовые уровни рядов динамики должны быть **упорядоченными во времени.** Не допускается анализ рядов с пропусками отдельных уровней, если же такие пропуски неизбежны, то их восполняют условными расчетными значениями.

2.2 Показатели анализа рядов динамики

При изучении явления во времени перед исследователем встает проблема описания интенсивности изменения и расчета средних показателей динамики. Решается она путем построения соответствующих показателей. Для характеристики интенсивности изменения во времени такими показателями будут:

1) абсолютный прирост;
2) коэффициент роста;
3) темпы роста;
4) темпы прироста,
5) абсолютное значение одного процента прироста.

Таблица 2.1 – Формулы расчета показателей динамики по цепной и базисной системе

Название показателя	Система расчета	
	Цепная (переменная)	Базисная (постоянная)
Абсолютный прирост (Δ)	$\Delta_{ö} = ó_i - y_{i-1}$	$\Delta_{á} = ó_i - y_0$
Коэффициент роста (Кр)	$K_{p_{ö}} = \frac{y_i}{y_{i-1}}$	$K_{p_{á}} = \frac{y_i}{y_0}$
Темп роста (Тр),%	$T_{p_{ö}} = K_{p_{ö}} \cdot 100$	$T_{p_{á}} = K_{p_{á}} \cdot 100$
Темп прироста (Тпр)	$T_{ïp_{ö}} = (K_{p_{ö}} - 1) \cdot 100$ $T_{ïp_{ö}} = \grave{O}_{p_{ö}} - 100$ $T_{ïp_{ö}} = \frac{\Delta_{ö}}{y_{i-1}} 100$	$T_{ïp_{á}} = (K_{p_{á}} - 1) \cdot 100$ $T_{ïp_{á}} = \grave{O}_{p_{á}} - 100$ $T_{ïp_{á}} = \frac{\Delta_{á}}{y_0} 100$
Абсолютное значение 1% прироста (А)	$A_{ö} = \frac{\Delta_{ö}}{T_{ïð_{ö}}}$ $A_{ö} = \frac{y_{i-1}}{100}$	$A_{á} = \frac{\Delta_{á}}{T_{ïð_{á}}}$ $A_{á} = \frac{y_0}{100}$

При расчете данных показателей приняты следующие условные обозначения:

ó_i - уровень текущего периода;

y_{i-1} - уровень периода, предшествующий текущему периоду;

ó_0 - уровень, принятый за постоянную базу сравнения (чаще всего начальный уровень временного ряда).

Для характеристики интенсивности развития явления за длительный период рассчитываются средние показатели динамики (таблица 2.2).

Система средних показателей динамики включает:

1) средний уровень ряда;
2) средний абсолютный прирост;
3) средний коэффициент роста;
4) средний темп роста;
5) средний темп прироста;
6) средняя величина одного процента прироста.

Таблица 2.2 – Формулы расчета средних показателей динамики

Название показателя	Метод расчета
Средний уровень ряда ($\bar{\text{ó}}$)	
- для интервального ряда	$\bar{\text{ó}} = \frac{\sum \text{ó}}{n}$
- для моментного ряда с равными интервалами	$\bar{y} = \frac{1/2 \cdot y_1 + y_2 + ... + y_{n-1} + 1/2 \cdot y_n}{n-1}$
- для моментного ряда с неравными интервалами	$\bar{\text{ó}} = \frac{\sum \text{ó} \cdot t}{\sum t}$
Средний абсолютный прирост ($\bar{\Delta}$)	$\bar{\Delta} = \frac{\sum \Delta}{n-1}$ $\bar{\Delta} = \frac{\sum y_n - y_1}{n-1}$
Средний коэффициент роста ($\bar{\text{Ē}}$)	$\bar{K}_p = \sqrt[n-1]{K_{p_1} \cdot K_{p_2} \cdot ... \cdot K_{n-1}}$ $\bar{K}_p = \sqrt[n-1]{\text{Ï}K_{\text{ð}}}$ $\bar{K}_p = \sqrt[n-1]{\frac{y_n}{y_1}}$
Средний темп роста ($\bar{\text{Ó}}_{\text{ç}}$)	$\bar{T}_{\text{ð}} = \bar{\text{Ē}}_{\text{ð}} \cdot 100$
Средний темп прироста ($\bar{\text{Ó}}_{\text{ïç}}$)	$\bar{T}_{\text{ïð}} = \bar{\text{Ó}}_{\text{ð}} - 100$ $\bar{T}_{\text{ïð}} = (\bar{\text{Ē}}_{\text{ð}} - 1) \cdot 100$
Средняя величина одного процента прироста ($\bar{A}$)	$\bar{A} = \frac{\bar{\Delta}}{\bar{\text{Ó}}_{\text{ïð}}}$

При расчете данных показателей приняты следующие условные обозначения:

$ó_1, ó_2, ..., ó_n$ - все уровни последовательных периодов (дат);

n - число уровней ряда;

t – продолжительность периода, в течение которого уровень не изменился.

2.3 Исследование временного ряда динамики

Опыт практического анализа рядов динамики свидетельствует о необходимости использования некоторых дополнительных показателей, представленных в таблице 2.3.

Таблица 2.3 – Дополнительные показатели, характеризующие изменения ряда динамики

Название показателя	Формула расчета	Характеристика показателя
Размах вариации	$R = x_{max} - x_{min}$	Разность между максимальным и минимальным уровнями динамического ряда.
Коэффициент выравненности	$\hat{E}_{\hat{a}} = \frac{ó_{min}}{y_{max}}$	Оптимальное значение составляет не ниже 0,1.
Среднее абсолютное отклонение	$\bar{d} = \frac{\sum \lvert x_i - \bar{x} \rvert}{n}$ (простое) $\bar{d} = \frac{\sum \lvert x_i - \bar{x} \rvert f_i}{\sum f_i}$ (взвешенное)	Если $\bar{d}$ имеет большое значение по сравнению с $\bar{x}$, т.е. большое отклонение от $\bar{x}$, то это свидетельствует о том, что данная совокупность в отношении данного признака неоднородна, а средняя нетипична.
Дисперсия	$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}$ (простая) $\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i}$ (взвешенная)	Дисперсия представляет собой средний квадрат отклонений индивидуальных значений признака от их средней величины

	$\sigma^2 = \overline{x^2} - (\bar{x})^2$	и вычисляется по формулам простой и взвешенной дисперсией (в зависимости от исходных данных).
Среднее квадратическое отклонение	$\sigma = \sqrt{\sigma^2}$	Среднее квадратическое отклонение равно корню квадратному из дисперсии.
Коэффициент вариации	$V_\sigma = \frac{\sigma}{\bar{x}} * 100\%$	Характеризует однородность совокупности. Вариация считается незначительной, если ее относительный уровень ниже 10%, средней – в пределах 10-30%, высокой – превышает 30% (больше 33% - совокупность неоднородна).
Устойчивость динамического ряда	$u = 1 - V_\sigma$	Если u составил 0,9, это означает, что среднее колебание составляет 10% среднего уровня. Однако вероятность того, что отдельное колебание (т.е. отклонение от тренда в отдельном периоде) не превзойдет средней величины колебаний, составляет лишь 0,68, если распределение колебаний по их величине близко к нормальному.
Коэффициент парной линейной корреляции	$r_{yt} = \frac{1/n\sum_{i=1}^{n}(y-\bar{y})(t-\bar{t})}{\sigma_y \cdot \sigma_t}$	Оценивает тесноту связи между уровнями (y) исследуемого показателя и текущим временем (t).
Коэффициент детерминации	$r_{yt}{}^2 = \sqrt{r_{yt}}$	Характеризует процент вариации в

		исследуемом динамическом ряду в зависимости от периода времени.
Теоретическое корреляционное отношение	$\eta = \sqrt{\frac{\delta^2}{\sigma^2}}$ δ^2 - дисперсия выровненных значений результативного признака; σ^2 - дисперсия фактических значений.	Изменяется в пределах от 0 до 1, показывает тесноту связи между признаками.
Коэффициент эластичности	$\acute{Y} = \frac{at}{\bar{y}}$ а - параметр уравнения зависимости динамики изучаемого явления (у) от времени (t) $\vec{ó}_t = a_0 + a_1 t$	Показывает на сколько процентов изменится значение результативного признак при измени факторного на 1%
Индекс или темп роста фактических уровней ряда (y) относительно теоретических значений ($\bar{y}$)	$I = \frac{y}{\bar{y}}$	Используется при оценке результатов производственной деятельности предприятий, например, по отдельным годам какого-то периода. Индекс лучше исчислять в процентах.
Мода	$\grave{I}\hat{i} = ó_0 + i * \frac{(f_{\grave{I}\hat{i}} - f_{\grave{I}\hat{i}-1})}{(f_{\grave{I}\hat{i}} - f_{\grave{I}\hat{i}-1}) + (f_{\grave{I}\hat{i}} - f_{\grave{I}\hat{i}+1})}$	Наиболее часто повторяющееся значение признака.
Медиана	$\grave{I}\mathring{a} = ó_0 + i * \frac{\frac{1}{2}\sum f_i - S_{Me-1}}{f_{\grave{I}e}}$	Значение признака, приходящееся на середину ранжированного ряда.
Асимметрия	$As = \frac{\bar{y} - Mo}{\sigma}$ $As = \frac{\bar{y} - Me}{\sigma}$	Характеристика распределения, с помощью которой оценивается симметрия расположения

		значений переменной относительно средней. Асимметрия выше 0,5 (независимо от знака) считается значительной; меньше 0,25 – незначительной.
Эксцесс	$Ex = ((\frac{\sum(y-\bar{y})^4}{n})/\sigma^4)-3$	Это показатель относительной крутости кривой ряда по сравнению с нормальным распределением.

На практике при изучении динамических рядов можно использовать большую совокупность различных аналитических характеристик, самое главное, чтобы они обеспечивали получении содержательной информации и реальных результатов. Часто исследователи ограничиваются расчетом 5-8 аналитических показателей, что не обеспечивает всестороннего изучения социально-экономических явлений.

Одной из основополагающих предпосылок проведения научно-обоснованного статистического анализа, адекватно отражающего причинно-следственные связи и зависимости, тенденции развития реальных явлений и процессов в динамике, является однородность статистической совокупности.

Анализ однородности статистической совокупности целесообразно проводить в следующей последовательности:

1) определение степени однородности всей совокупности по одному или нескольким существенным признакам;

2) определение и анализ аномальных наблюдений;

3) выбор оптимального варианта выделения однородных совокупностей.

В статистической теории и практике разработаны различные подходы к оценке степени однородности.

Проблемой оценки однородности совокупности занимались такие известные ученые, как Ю. Аболенцев, Г. Кильдишев, В. Овсиенко и другие.

Наиболее сложным и дискуссионным является вопрос о способах и критериях выделения однородных групп объектов в пределах исходной совокупности.[2]

Важной предпосылкой получения научно-обоснованных результатов статистического анализа и моделирования является проверка и выполнение гипотезы о близости распределения эмпирических данных нормальному закону. Для нормального закона распределения характерно:

[2] Садовникова Н.А., Шмойлова Р.А. Анализ временных рядов и прогнозирование. – М.: 2001.

$$X \approx Mo \approx Me\ ;\ As = 0;\ Ex = 0$$

Одним из недостатков данного подхода к оценке характера распределения является наличие субъективности в анализе достаточности величины отклонения X от Me и Mo от Me для подтверждения гипотезы.

Вопросы для самопроверки:

1. *Что такое временной ряд?*
2. *Перечислите основные виды рядов динамики в зависимости от классификационных признаков.*
3. *В чем заключается смысл сопоставимости динамических рядов?*
4. *Какие показатели характеризуют интенсивность изменения явления во времени?*
5. *Какие показатели входят в систему средних показателей динамики?*
6. *Охарактеризуйте дополнительные показатели, характеризующие изменения ряда динамики?*
7. *Что такое однородность совокупности? Перечислите основные этапы анализа однородности статистической совокупности?*

3 Моделирование тенденции

3.1 Основные элементы временного ряда

Статистическую модель можно построить, используя два типа данных:

1) данные, характеризующие совокупность различных объектов в определенный (момент) период времени;

2) данные, характеризующие один и тот же объект за ряд последовательных (моментов) периодов времени.

Модели, построенные по данным первого типа, называются *пространственными моделями*. Модели, построенные по данным второго типа – *моделями временных рядов*.

В практике исследования динамики явлений и прогнозирования принято считать, что значения уровней временных рядов могут содержать следующие компоненты (структурообразующие компоненты):

– Тренд (u_t);
– Сезонную компоненту (S_t);
– Циклическую компоненту (V_t);
– Случайную компоненту (ε_t).

Под *трендом* понимают изменение, определяющее общее направление развития, основную тенденцию временного ряда. Это систематическая составляющая долговременного действия.

Наряду с долговременными тенденциями во временном ряду часто возникают более или менее регулярные колебания – периодические составляющие рядов динамики.

Если период колебаний не превышает одного года, то их называют *сезонными* (например, колебания цен на сельскохозяйственную продукцию, увеличение закупок в предпраздничный период).

При большем периоде колебания считают, что во временных рядах имеет место *циклическая* составляющая (циклы деловой активности Кондратьева). В силу небольшой длины временного ряда экономических показателей, в экономических временных рядах редко выделяю циклические компоненты. Однако в естественных науках циклические составляющие хорошо изучены (цикличность солнечной активности примерно 11 лет).

Если из ВР удалить тренд и периодические составляющие, то останется *случайная компонента.*

Факторы, под действием которых формируется нерегулярная компонента, разделяют на два вида:

– Факторы резкого, внезапного действия - вызывают более значительные отклонения – катастрофические колебания (стихийные бедствия, эпидемии, война, кризис);

– Текущие факторы – вызывают случайные колебания и являются результатом действия большого числа побочных причин.

Если временной ряд представляется в виде суммы соответствующих компонент, то полученная модель носит название *аддитивной*:

$$Y_t = u_t + S_t + v_t + \varepsilon_t;$$

Если в виде произведения – *мультипликативной:*

$$Y_t = u_t \cdot S_t \cdot v_t \cdot \varepsilon_t$$

Также выделяют модели *смешанного типа*:

$$Y_t = u_t \cdot S_t \cdot v_t + \varepsilon_t$$

Отличительная особенность аддитивной модели заключается в том, что амплитуда сезонных колебаний, отражающих отклонения от тренда или среднего, остается примерно постоянной, неизменной во времени.

Не обязательно в процессе формирования значений уровней каждого ВР должны участвовать одновременно все компоненты.

Основные этапы анализа временного ряда:

– графическое представление и описание поведения временного ряда;

– выделение и удаление закономерных (неслучайных) составляющих временного ряда (тренда, сезонных и циклических составляющих);

– сглаживание и фильтрация (удаление низко- и высокочастотных составляющих временного ряда);

– исследование случайной составляющей временного ряда, построение и проверка адекватности математической модели для ее описания;

– прогнозирование развития изучаемого процесса на основе имеющегося временного ряда;

– исследование взаимосвязи между различными временными рядами.

3.2 Выявление наличия (отсутствия) тенденции во временном ряду

Для выявления факта наличия или отсутствия неслучайной составляющей $f(x)$, то есть для проверки гипотезы о существовании тренда - $H_0 : E_y(t) = a = const$ можно использовать различные критерии.

Например:

1) критерии серии, который имеет две модификации
   - критерий серий, основанный на медиане выборки;
   - критерий «восходящих - нисходящих» серий
2) метод Фостера – Стюарта и др.

Рассмотрим более подробно каждый из них.

Критерий серий, основанный на медиане выборки.

Для этого необходимо:

1. из исходного ряда с уровнями: $y_1, y_2, ..., y_n$ образовать ранжированный ряд: $y'_1, y'_2, ..., y'_n$, где y'_1- наименьшее значение уровней исходного ряда;

2. определить медиану Me этого вариационного ряда. В случае нечетного ряда $(n = 2m + 1) - Me = y'_{m+1}$, в противном случае $(n = 2m)$, $Me = \frac{y'_m + y'_{m+1}}{2}$.

3. образовать последовательность δ, из «+» и «-» по следующему правилу: $\delta_i = \begin{cases} "+", y_i > Me \\ "-", y_i < Me \end{cases}$ для всех t $\forall \in (1, n)$

4. необходимо подсчитать $\nu(n)$ совокупности δ_1, где под серией понимается последовательность подряд идущих «+» и «-». Один «+» или один « - « тоже считается серией. Определяем $\tau_{\max}(n)$- протяженность самой длинной серии.

5. проверка гипотезы основывается на том, что при условии случайности ряда (при отсутствии систематической составляющей) протяженность самой длинной серии не должно быть слишком большой, а общее число серии слишком маленьким. Поэтому, для того, чтобы не была отвергнута гипотеза о случайности исходного ряда, должны выполняться следующие неравенства:

$$\nu(n) > \left[\frac{1}{2}(n + 1 - 1.96\sqrt{n - 1})\right] \tag{3.1}$$

$$\tau_{\max}(n) < [1.43\ln(n + 1)] \tag{3.2}$$

Если хотя бы одно из неравенств нарушается, то гипотеза отвергается с вероятностью ошибки α. Следовательно, подтверждается наличие зависящей от времени неслучайной составляющей.

Критерий «восходящих - нисходящих» серий.

1. образуется последовательность знаков – плюсов и минусов по следующему принципу:

$$\delta_i = \begin{cases} +, \text{ åñëè } \acute{o}_{t+1} - y_t > 0 \text{ åñëè } t - 1, 2, ..., n - 1 \\ -, \text{ åñëè } y_{t+1} - y_t < 0 \text{ äëÿ } t = 1, 2, ..., n - 1 \end{cases}$$

В случае если последующее наблюдение окажется равным предыдущему, учитывается только одно наблюдение. Таким образом, элементы этой последовательности принимают значение «+», если последующее значение уровня ряда больше предыдущего, и «-« - если меньше. Общее число знаков «+» и «-« заранее не известно. Индекс *i* может принимать значение *1, 2, … , к,* где $к \le n$*-1*.

2. подсчитывается общее число серий $\nu(n)$ и протяженность самой длинной серии $\tau_{\max}(n)$ аналогично тому, как это делалось в предыдущем варианте критерия. Очевидно, что при этом каждая серия, состоящая из плюсов, соответствует возрастанию уровней ряда («восходящая» серия), а последовательность минусов – их убыванию («нисходящая» серия).

3. для того, чтобы не была отвергнута гипотеза, должны выполняться следующие неравенства (при уровне значимости α, заключенном между 0,05 и 0,0975):

$$\nu(n) > \left[\frac{1}{3}(2n+1) - 1.96\sqrt{\frac{16n-29}{90}}\right] \qquad (3.3)$$

$$\tau_{\max}(n) < \tau_0(n) \qquad (3.4)$$

где $\tau_0(n)$ - табличное значение, зависящее от длины временного ряда (таблица 3.1).

Таблица 3.1 – Значение $\tau_0(n)$ для критерия «восходящих – нисходящих» серий

n	n≤26	26 < n ≤ 153	153 < n ≤ 1170
$\tau_0(n)$	5	6	7

Если хотя бы одно неравенство нарушается, то нулевая гипотеза отвергается (следовательно, подтверждается наличие зависящей от времени неслучайной составляющей).

Метод Фостера – Стюарта.

Алгоритм метода выглядит следующим образом:

1. сравнивается каждый уровень ряда со всеми предыдущими, при этом

если $y_i > y_{i-1}$, U=1, e=0 и наоборот

2. вычисляются значения величин S, d:

$$d = \sum d_i \qquad S = \sum S_i \qquad (3.5)$$

где
$$S_i = U_i + e_i$$
$$d_i = U_i - e_i$$

Оба показателя асимптотически нормальны и имеют независимые распределения

3. проверка производится с использованием t-критерия Стьюдента гипотеза о том, можно ли считать случайным разности S-μ и d-0

$$t_S = \frac{S-\mu}{\sigma_1} \qquad t_d = \frac{d-0}{\sigma_2} \qquad (3.6)$$

значения μ и σ табулированы и приведены в специальных таблицах.

4. сравниваются расчетные значения t с табличными при заданном уровне значимости. Если t_s , t_d < t табличного, то гипотеза об отсутствии тренда с средней и дисперсии подтверждается.

Кроме рассмотренных подходов определения тенденции во временном ряду используется также критерий квадратов последовательных разностей (метод Аббе), метод проверки разностей средних уровней и др.

При наличии тенденции в ряду динамики его уровни можно рассматривать как функцию времени.

Существует несколько методов облегчающих процесс выбора формы кривой роста. Наиболее простой метод – визуальный, опирающийся на графическое изображение.

3.3 Методы анализа основной тенденции в рядах динамики

После установления наличия тенденции во временном ряду производится его описание с помощью методов сглаживания. Методы сглаживания можно разделить на две группы:

1) сглаживание или механическое выравнивание отдельных членов ряда динамики с использованием фактических значений соседних уровней;

2) выравнивание с применением кривой, проведенной между конкретными уровнями таким образом, чтобы она отображала тенденцию, присущую ряду, и одновременно освободила его от незначительных колебаний.

Рассмотрим каждый из методов более подробно[3].

Метод усреднения по левой и правой половине – простейший метод определения тренда, заключающийся в разделении ряда на две равные (примерно равные) части, определении для каждой из них среднего значения и отражении линии тренда на графике.

Пример.

Таблица 3.2 – Динамика объема экспорта

Год	2005	2006	2007	2008	2009	2010	2011
Объем экспорта продукции предприятия, тыс. долл. США	1200	1350	1400	1370	1350	1380	1310

Разделим ряд объема экспорта продукции предприятия (табл.3.1) на 2 части:

- 2005 – 2008 гг,
- 2009 – 2011 гг.

Средние значения для каждой части:

$$\bar{y}_1 = \frac{1200 + 1350 + 1400 + 1370}{4} = 1330$$

$$\bar{y} = \frac{1350 + 1380 + 1310}{3} = 1347$$

[3] Статистика. Контрольные задания и методические указания для студентов заочной формы обучения специальностей 060800, 061100, 311000/Составители: Гришакина Н.И., Фетисова Г.В. Учебное пособие (НовГУ им. Ярослава Мудрого: Великий Новгород, 2009.

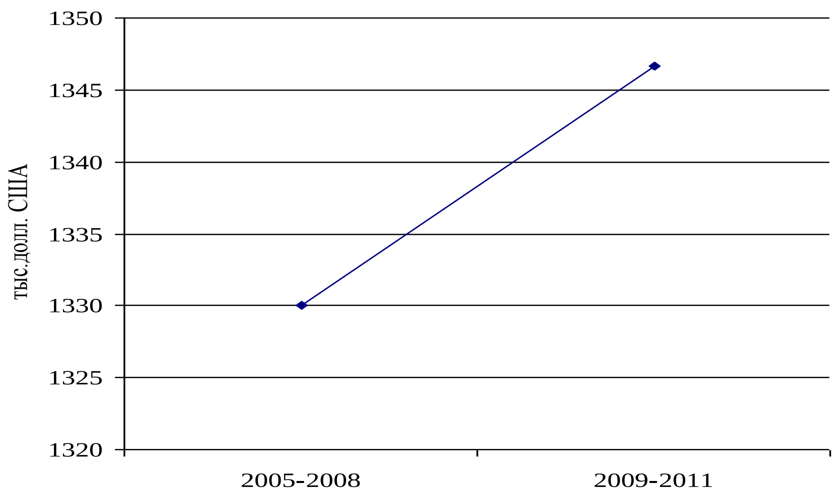

Рисунок 3.1 – Основная тенденция развития объема экспорта продукции предприятия

Метод **укрупнения интервалов** основан на укрупнении периодов времени, к которым относятся уровни ряда. В отдельных случаях (в зависимости от экономической сущности исследуемых показателей) целесообразно при укрупнении уровней рассчитывать средние значения.

Пример. Определим тенденцию развития средней заработной платы работников предприятия (табл.3.3). Месячные уровни заработной платы укрупним в квартальные, определив средние значения.

Таблица 3.3 – Динамика средней заработной платы работников предприятия

Месяц	Январь	Февраль	Март	Апрель	Май	Июнь
Средняя заработная плата работников предприятия, тыс. руб.	26,5	27,9	26,3	26,8	28,1	27,3
Месяц	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
Средняя заработная плата работников предприятия, тыс. руб.	27,9	28,1	27,3	27,6	28,5	28,9

Таблица 3.4 – Динамика среднемесячной заработной платы работников предприятия

Квартал	Итого за квартал	В среднем
1	80,7	26,9
2	82,2	27,4
3	83,3	27,8
4	85,0	28,3

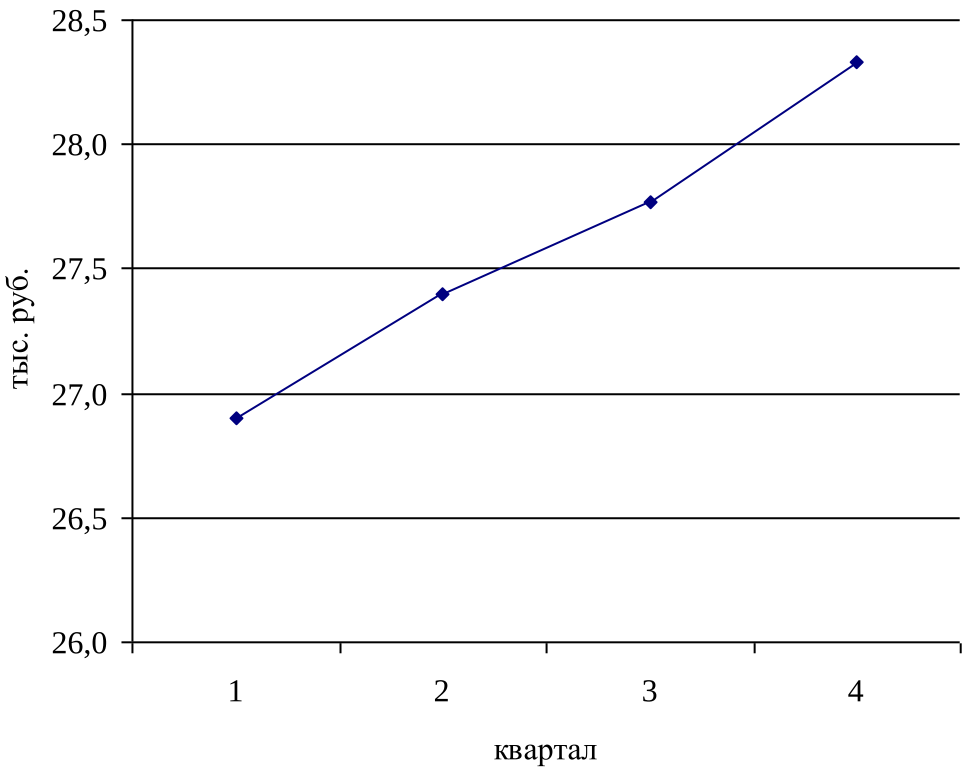

Рисунок 3.2 – Основная тенденция развития среднемесячной заработной платы работников предприятия

Сглаживание рядов динамики на основе **скользящих средних** основано на вычислении звеньев подвижной средней из такого числа уровней ряда, которая соответствует длительности наблюдаемых в ряду динамики циклов. То есть изначально выбирается период скольжения, равный двум, трем, четырем и т.д. периодам.

Например, трехчленная скользящая средняя исчисляется по следующей схеме:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3}{3} \qquad \text{(первая средняя),}$$

$$\bar{y}_2 = \frac{y_2 + y_3 + y_4}{3} \qquad \text{(вторая средняя),}$$

$$\bar{y}_3 = \frac{y_3 + y_4 + y_5}{3} \qquad \text{(третья средняя) и т.д.}$$

А для ряда внутригодовой динамики применяется чаще всего четырехчленные скользящие средние. Их расчет состоит в определении средних величин из четырех уровней ряда с отбрасыванием при вычислении каждой новой средней одного уровня ряда слева и присоединением одного уровня справа:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3 + y_4}{4} \qquad \text{(первая средняя),}$$

$$\bar{y}_2 = \frac{y_2 + y_3 + y_4 + y_5}{4} \qquad \text{(вторая средняя),}$$

$$\bar{y}_3 = \frac{y_3 + y_4 + y_5 + y_6}{4}$$ (третья средняя) и т.д.

Чтобы отнести скользящую среднюю к определенному периоду необходимо провести центрирование расчетных средних, определяемых как простая средняя арифметическая из 2-х рядом лежащих скользящих средних:

$\bar{y}_1^{ценр} = \frac{\bar{y}_1 + \bar{y}_2}{2}$ (1-й сглаженный средний уровень),

$\bar{y}_2^{ценр} = \frac{\bar{y}_2 + \bar{y}_3}{2}$ (2-й сглаженный средний уровень)

$\bar{y}_3^{ценр} = \frac{\bar{y}_3 + \bar{y}_4}{2}$ (3-й сглаженный средний уровень) и т.д.

Пример. Определим тенденцию развития средней заработной платы работников предприятия (табл.3.5).

Таблица 3.5 – Динамика средней заработной платы работников предприятия

Месяц	Средняя заработная плата, тыс. руб.	Четырехмесячная скользящая средняя	
		нецентрированная	центрированная
Январь	26,5		
Февраль	27,9	26,88	
Март	26,3	27,28	27,08
Апрель	26,8	27,13	27,20
Май	28,1	27,53	27,33
Июнь	27,3	27,85	27,69
Июль	27,9	27,65	27,75
Август	28,1	27,73	27,69
Сентябрь	27,3	27,88	27,80
Октябрь	27,6	28,08	27,98
Ноябрь	28,5		
Декабрь	28,9		

Центрированные средние наносят на график с эмпирическими данными.

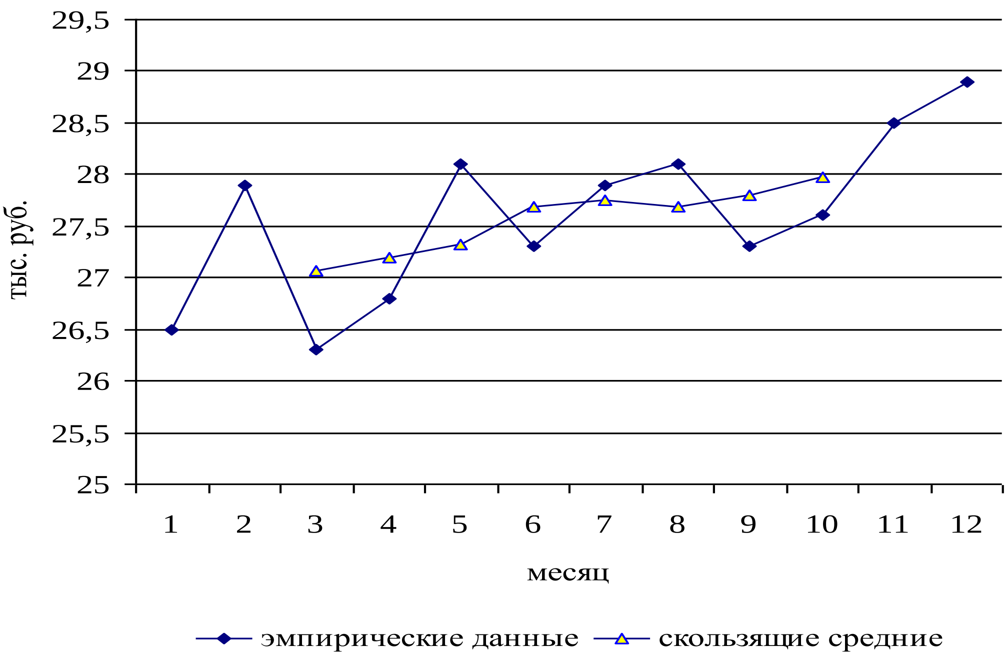

Рисунок 3.3 – Основная тенденция развития среднемесячной заработной платы работников предприятия

Особенность способа сглаживания рядов динамики на основе скользящих средних заключается в том, что полученные средние не дают теоретических рядов, в основе которых лежала бы определенная математическая закономерность.

Более совершенным приемом изучения общей тенденции в рядах динамики является **аналитическое выравнивание**. Оно основано на допущении, что изменения в рядах динамики могут быть выражены определенным математическим законом. На основе теоретического анализа выявляется характер явления во времени и на этой основе выбирается то или иное математическое выражение типа закономерности изменения явления:

- линейная функция $y_t = a + b \cdot t$
- полином второго порядка $y_t = a + b \cdot t + c \cdot t^2$
- полином третьего порядка $y_t = a + b \cdot t + c \cdot t^2 + d \cdot t^3$
- степенная функция $y_t = a \cdot t^b$
- показательная функция $y_t = a \cdot b^t$ и другие.

Данный прием сводится к следующему:

а) на основе экономического анализа явления за рассматриваемый период времени выявляется его характер;

б) исходя из характера явления выбирается то или иное математическое уравнение;

в) определяются параметры уравнения;

г) рассчитываются теоретические (выровненные) уровни ряда динамики, которые наносятся на график эмпирических значений;

д) прогнозируются уровни динамического ряда на основе аппроксимирующей модели на предстоящий период.

Для нахождения параметров уравнений используют метод наименьших квадратов (в случае линейной, параболической, гиперболической зависимостей) или линеаризации переменных (степенная, показательная и др.). Смысл метода наименьших квадратов состоит в том, что вычисленная линия теоретических уровней должна проходить в максимальной близости к фактическим уровням ряда, то есть

$$\sum(y - \hat{y}_t)^2 = \min$$

где у – исходные (эмпирические) уровни динамического ряда;

$\hat{y}_t$ - расчетные (теоретические) уровни ряда динамики.

Рассмотрим выравнивание ряда динамики **по прямой** (таблица 3.6):

$$\hat{y}_t = a + bt$$

где у – исходные (эмпирические) уровни ряда динамики

a и b – параметры уравнения,

t – время

Параметры уравнения находятся на основе системы уравнений:

$$\begin{cases} \sum y = a \cdot n + b\sum t, \\ \sum yt = a\sum t + b\sum t^2. \end{cases}$$

Расчет параметров заметно упрощается, если перенести начало отсчета времени в середину исходного ряда (что бы $\sum t = 0$). Причем, если число уровней ряда нечетное, нумерация **t** следующая: …-3, -2, -1, 0, +1, +2, +3…. ; а если число уровней ряда четное, нумерация **t** будет следующая: …-5, -3, -1, +1, +3, +5….

При условии, что Σt=0 (графа 2 таблицы 3.6) исходные нормальные уравнения принимают вид:

$$\begin{cases} \sum y = a \cdot n \\ \sum yt = b\sum t^2. \end{cases},$$

отсюда $\begin{cases} a = \dfrac{\sum y}{n} \\ b = \dfrac{\sum yt}{\sum t^2} \end{cases}$.

Необходимые величины рассчитаны в графах 3 и 4 таблицы 3.6.

Параметризованное уравнение, рассчитанное по данным таблицы 3.2, пример 1, имеет вид $\hat{y}_t = 1337{,}14 + 12{,}14 \cdot t$.

В полученное параметризованное уравнение подставляют значения **t** и получают расчетные значения результативного признака $\hat{y}_t$ (графа 5 таблицы 3.6), которые и являются тенденцией данного явления. Их наносят на график с эмпирическими данными.

Таблица 3.6 - Аналитическое выравнивание ряда динамики по прямой

Год	Объем экспорта, тыс. долл. США, (y)	Условные обозначения времени (t)	t^2	y*t	$\hat{y}_t$	$\frac{\lvert y-\hat{y}_t\rvert}{y}$
А	1	2	3	4	5	6
2005	1200	-3	9	-3600	1301	0,08393
2006	1350	-2	4	-2700	1313	0,02751
2007	1400	-1	1	-1400	1325	0,05357
2008	1370	0	0	0	1337	0,02399
2009	1350	1	1	1350	1349	0,00053
2010	1380	2	4	2760	1361	0,01346
2011	1310	3	9	3930	1374	0,04852
Всего	9360	0	28	340	9360	0,25152

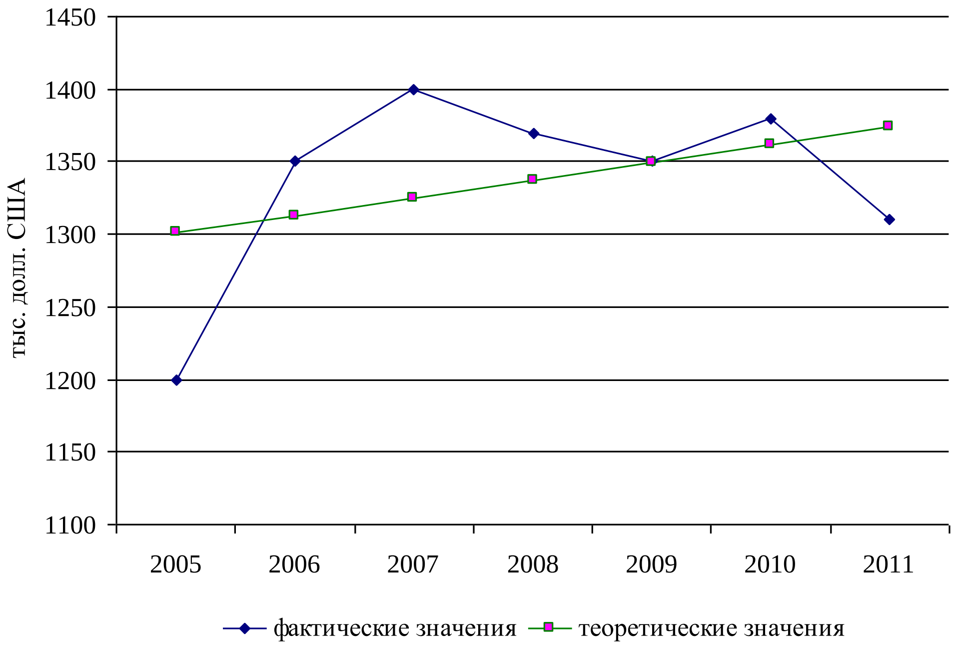

Рисунок 3.4 - Динамика эмпирических и теоретических уровней ряда динамики объема экспорта продукции предприятия

Рассмотрим выравнивание ряда динамики **по полиному второго порядка** (таблица 3.7):

$$\hat{y}_t = a + bt + ct^2$$

где y – исходные (эмпирические) уровни ряда динамики

a, b и c– параметры уравнения,

t – время

Параметры уравнения находятся на основе системы уравнений:

$$\begin{cases} \sum y = a \cdot n + b\sum t + c\sum t^2, \\ \sum yt = a\sum t + b\sum t^2 + c\sum t^3. \\ \sum yt^2 = a\sum t^2 + b\sum t^3 + c\sum t^4 \end{cases}$$

При условии, что Σt=0 (графа Всего таблицы 3.7) исходные нормальные уравнения принимают вид:

$$\begin{cases} \sum y = a \cdot n + c\sum t^2, \\ \sum yt = b\sum t^2. \\ \sum yt^2 = a\sum t^2 + c\sum t^4 \end{cases}$$

Необходимые величины рассчитаны в графах 3, 4, 5, 6 таблицы 3.7.

Параметризованное уравнение, рассчитанное по данным таблицы 3.2, , имеет вид $\hat{y}_t = 1393{,}34 + 12{,}14 \cdot t - 14.05 \cdot t^2$

В полученное параметризованное уравнение подставляют значения t и получают расчетные значения результативного признака $\hat{y}_t$ (графа 7 таблицы 3.7), которые и являются тенденцией данного явления. Их так же наносят на график с эмпирическими данными.

Таблица 3.7 - Аналитическое выравнивание ряда динамики по полиному второго порядка

Год	Объем экспорта, тыс. долл. США, (y)	Условные обозначения времени (t)	t^2	t^4	y*t	y*t^2	$\hat{y}_t$	$\frac{\lvert y-\hat{y}_t \rvert}{y}$
А	1	2	3	4	5	6	7	8
2005	1200	-3	9	81	-3600	10800	1230	0,0254
2006	1350	-2	4	16	-2700	5400	1313	0,0275
2007	1400	-1	1	1	-1400	1400	1367	0,0235
2008	1370	0	0	0	0	0	1393	0,0170
2009	1350	1	1	1	1350	1350	1391	0,0307
2010	1380	2	4	16	2760	5520	1361	0,0135
2011	1310	3	9	81	3930	11790	1303	0,0051
Всего	9360	0	28	196	340	36260	9360	0,1427

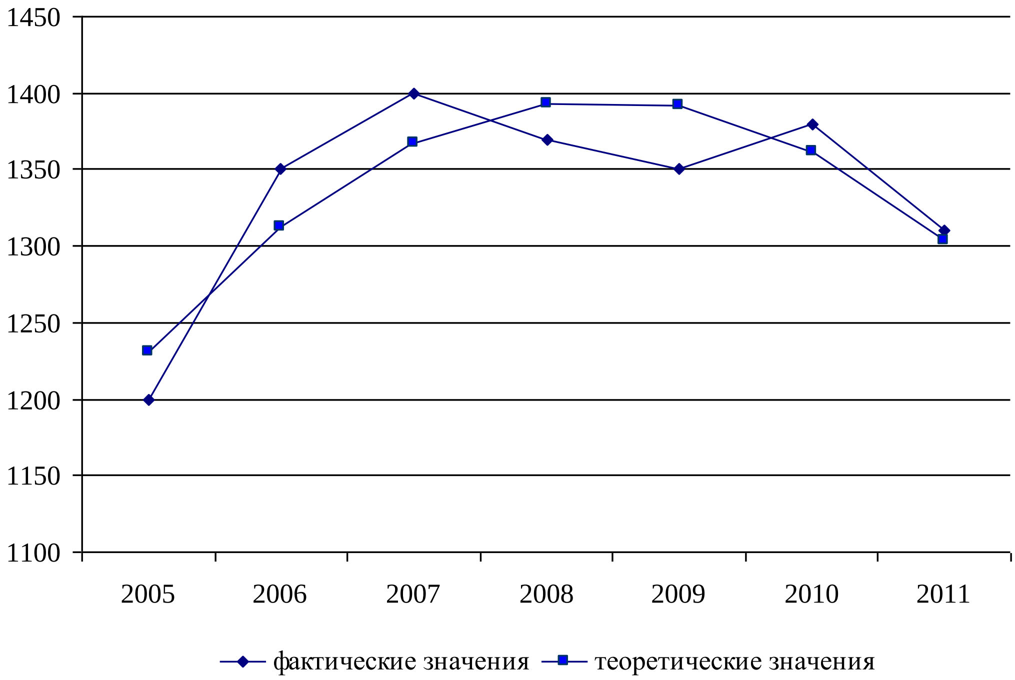

Рисунок 3.5 - Динамика эмпирических и теоретических уровней ряда динамики объема экспорта продукции предприятия

Рассмотрим выравнивание ряда динамики **по степенной функции** (таблица 3.8):

$$y_t = a \cdot t^b$$

где y – исходные (эмпирические) уровни ряда динамики

a, b – параметры уравнения,

t – время

В данном случае необходимо провести линеаризацию переменных, то есть привести степенную функцию к линейному виду путем логарифмирования обеих частей уравнения (для удобства используются логарифмы с постоянным основанием – натуральный или десятичный).

Воспользуемся натуральным логарифмом и получим линейную функцию, параметры которой определяются МНК, а в качестве расчетных данных используются не исходные уровни, а их натуральные логарифмы (графа 3 и 4 таблицы 3.8):

$\ln y = \ln a \cdot t^b$ или $\ln y = \ln a + b \cdot \ln t$

Согласно МНК, построим систему нормальных уравнений:

$$\begin{cases} \sum \ln y = \ln a \cdot n + b \sum \ln t, \\ \sum \ln y \cdot \ln t = \ln a \sum \ln t + b \sum (\ln t)^2. \end{cases}$$

Подставим необходимые величины, рассчитанные в графах 5, 6 таблицы 3.8, и получим:

$$\begin{cases} 50{,}38 = \ln a \cdot 7 + b \cdot 8{,}5252 \\ \ln a \cdot 8{,}5252 + b \cdot 13{,}196 = 61{,}4894 \\ . \end{cases}$$

Проведя преобразования, получим линейное уравнение: $\ln y = 7{,}14 + 0{,}0466 \cdot \ln t$

Для перехода обратно к степенной функции выполняется потенцирование, т.е.:

$$e^{\ln y} = e^{7{,}14+0{,}0466 \ln t}.$$

Таким образом, параметризованное уравнение, рассчитанное по данным таблицы 3.2 , имеет вид $y = 1234{,}243 \cdot t^{0{,}0466}$

В полученное параметризованное уравнение подставляют значения **t** и получают расчетные значения результативного признака $\hat{y}_t$ (графа 7 таблицы 3.8), которые и являются тенденцией данного явления. Их так же наносят на график с эмпирическими данными.

Таблица 3.8 - Расчетная таблица для аналитического выравнивания ряда динамики по степенной функции

Год	Объем экспорта, тыс. долл. США, (y)	Условные обозначения времени (t)	$\ln y$	$\ln t$	$(\ln t)^2$	$\ln y \cdot \ln t$	$\hat{y}_t$	$\frac{\lvert y - \hat{y}_t \rvert}{y}$
А	1	2	3	4	5	6	7	8
2005	1200	1	7,0901	0,0000	0,0000	0,0000	1234	0,028536
2006	1350	2	7,2079	0,6931	0,4805	4,9961	1275	0,055733
2007	1400	3	7,2442	1,0986	1,2069	7,9586	1299	0,072089
2008	1370	4	7,2226	1,3863	1,9218	10,0126	1317	0,038972
2009	1350	5	7,2079	1,6094	2,5903	11,6006	1330	0,01454
2010	1380	6	7,2298	1,7918	3,2104	12,9541	1342	0,027738
2011	1310	7	7,1778	1,9459	3,7866	13,9673	1351	0,031599
Всего	9360	28	50,38	8,5252	13,1965	61,4894	9148	0,269206

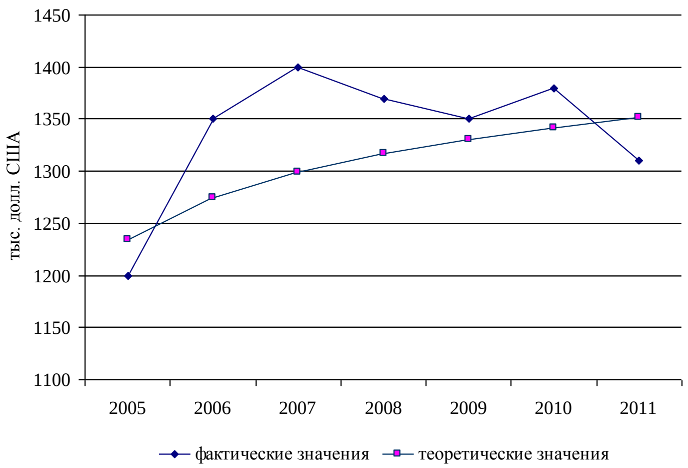

Рисунок 3.6 - Динамика эмпирических и теоретических уровней ряда динамики объема экспорта продукции предприятия

Критерием выбора параметризованного (лучшего для прогнозирования) уравнения является наименьшая ошибка аппроксимации:

$$E_a = \frac{1}{n} * \sum \frac{|y - \hat{y}_t|}{y} * 100$$

Средняя ошибка аппроксимации не должна превышать 10%.

Рассчитаем ошибки аппроксимации для анализируемых функций (графа 6 таблицы 10; графа 8 таблицы 11; графа 8 таблицы 12):

$E_a = \frac{1}{7} \cdot 0.25152 \cdot 100 = 3.59\%$ - для линейной функции

$E_a = \frac{1}{7} \cdot 0.1427 \cdot 100 = 2.04\%$ - для полинома второго порядка

$E_a = \frac{1}{7} \cdot 0.2692 \cdot 100 = 3.85\%$ - для степенной функции.

Так как все функции имеют ошибку аппроксимации в пределах средней, для прогнозирования можно выбрать любую. В данном случае отдадим приоритет линейной, как наиболее простой.

Вопросы для самопроверки:

1. *Какие типы данных используются для построения статистических моделей?*
2. *Какие компоненты формируют временной ряд?*
3. *Что такое тренд?*
4. *Опишите основные этапы анализа временного ряда?*
5. *Какие методы используются для определения наличия (отсутствия) тенденции во временных рядах?*
6. *В чем смысл критерия серий, основанного на медиане выборки?*
7. *В чем суть критерия «восходящих – нисходящих» серий?*
8. *В чем смысл метода Фостера - Стюарта?*
9. *В чем суть метода усреднения по левой и правой половине?*
10. *В чем суть метода укрупнения интервалов?*
11. *В чем суть метода скользящих средних?*
12. *Опишите основные этапы проведения аналитического выравнивания?*

4 Моделирование периодической компоненты

4.1 Выявление периодической компоненты

Для проверки предложения о существенности периодической компоненты ряда динамики целесообразно использовать такие критерии случайности, которые имеют наибольшую мощность относительно альтернативной гипотезы о цикличности ряда. Наиболее простым для применения и зрительно понятным является критерий «пиков» и «ям». В основе этого критерия лежит подсчет экстремальных точек ряда p, который осуществляется следующим способом.

$$\hat{p} = \sum_{t=2}^{n-1} p_t \tag{4.1}$$

где $p_t = \begin{cases} 1, \text{ åñëè} & y_{t-1} \langle\; y_t \;\rangle\; y_{t+1} \\ \text{èëè} & y_{t-1} \;\rangle\; y_t \;\langle\; y_{t+1} \\ 0 \text{ â ïðîòèâíîì} & \text{ñëó÷àå} \end{cases}$

t=1+n

n – число наблюдений в ряду динамики.

Для случайного ряда математическое ожидание числа экстремальных точек:

$$\bar{p} = \frac{2(n-2)}{3} \tag{4.2}$$

Проверка гипотезы сводится к сравнению $\bar{p}$ с расчетным значением $\hat{p}$. Если эти значения близки, то можно отказаться от дальнейшей проверки и признать ряд случайным. Если же $\hat{p}$ и $\bar{p}$ значительно отличаются друг от друга, тот производится дальнейшая проверка гипотезы, основанная на подсчете фаз различной длины.

Фазой называется интервал между двумя соседними уровнями, для которого $p_t = 1$. Для определения длины фазы е достаточно просто найти разности индексов двух соседних экстремальных точек. Затем подсчитывается число фаз N_1, N_2, N_3 длин $e_1=1$, $e_2=2$, $e_3=3$. Теоретическое значение числа фаз длины е для случайного ряда следующее:

$$\widehat{N}_e = \frac{2(n-e-2)(e^2-3e+1)}{(e+3)!} \tag{4.3}$$

Естественная процедура проверки случайности сводится к сравнению наблюдаемых значений N_1, N_2, N_3 с теоретическим значением N_e . Однако при небольшом числе наблюдений n критерий χ^2 здесь непосредственно использовать нельзя, так как в этом случае длины фаз e_i не являются независимыми. Доказательно, что при разбиении длины фазы на три группы: e=1, e=2, e=3 (две степени свободы) – статистика χ^2 может быть использована в обычной форме (v= 2,5) при χ^2 =6,3. Расчетные значения χ^2 в случае трех длинных фаз определяются по формуле:

$$\chi^2 = \sum_{e=1}^{3} \frac{(N_e - \widehat{N}_e)}{N_e} \tag{4.4}$$

Если $\chi^2 \geq 6{,}3$, то колебания исходного ряда нельзя считать чисто случайными и ряд содержит периодическую составляющую. Этот критерий весьма чувствителен к периодическим колебаниям и имеет практически нулевую эффективность относительно альтернативы наличия тренда, поэтому он может применяться непосредственно к исходному ряду динамики в отличие от других критериев, которые требуют, чтобы из ряда динамики предварительно была выделена систематическая составляющая. После того как установлена периодическая составляющая, производится ее анализ[4].

4.2 Анализ сезонной компоненты

При рассмотрении квартальных или месячных данных многих социально-экономических явлений часто обнаруживаются определенные, постоянно повторяющиеся колебания, которые существенно не изменяются за длительный период времени. Они являются результатом влияния природно - климатических условий, общих экономических условий, а также ряда многочисленных разнообразных факторов, которые частично являются регулируемыми. В статистике периодические колебания, которые имеют определенный и постоянный период, равный годовому промежутку, носят название «сезонны колебания» и «сезонные волны».

Под сезонностью понимается устойчивая закономерность внутригодовой динамики того или иного явления, которая проявляется во внутригодовых повышениях или понижениях уровней того или иного показателя на протяжении ряда лет.

Многие процессы имеет ярко выраженную зависимость от сезонных колебаний.

Изучение сезонности позволяет:

- определить степень влияния природно-климатических условий на формирование изучаемого потока;
- установить продолжительность сезона;

[4] Теория статистики / Под ред. Минашкин В.Г., Шмойлова Р.А. и др. - М.: ЕАОИ, 2008.

- раскрыть факторы, обусловливающие сезонность в; определить экономические последствия сезонности на уровне региона и фирмы;
- разработать комплекс мероприятий по снижению сезонной неравномерности.

Сезонность определяется целым рядом факторов:

- природно-климатических;
- экономических – структура потребления товаров и услуг, формирование платежеспособности спроса посредством предложения;
- социальных – наличие свободного времени;
- демографических – дифференцированный спрос по половозрастному составу и другие признакам;
- психологических – традиции, мода, подражание;
- материально-технических – развитие сети размещения, питания, транспорта, культурно-оздоровительного обслуживания;
- технологических – комплексный подход в предоставлении качественных услуг.

Все перечисленные выше факторы сезонных колебаний можно подразделить на первичные и вторичные. К первичным относятся факторы, формирующиеся под воздействием природно-климатических условий; ко вторичным - все остальные.

Следовательно, существует реальная возможность влияния на сезонную неравномерность спроса. Сезонность ведет к сезонному характеру занятости работников изучаемой индустрии. Это имеет свои положительные и отрицательные стороны.

С одной стороны, сезонность порождает неравномерное распределение рабочего времени (сверхурочные в туристский сезон и недостаточная загруженность рабочих в межсезонье) и, как следствие, значительный удельный вес не полностью занятых работников и текучесть кадров.

С другой стороны, сезонность стимулирует многопрофильный характер рабочих мест, когда один и тот же работник выполняет разные функции в зависимости от сезонных особенностей.

Кроме того, сезонная работа выгодна для многих категорий населения как источник дополнительного дохода.

Сезонность влияет на структуру занятости работников в туристской индустрии, особенностями которой являются:

- значительный удельный вес неполной занятости;
- сезонные колебания объема занятости и трудовой нагрузки;
- низкий удельный вес квалифицированного персонала;
- ограниченные возможности профессионального роста;
- значительный удельный вес женского труда.

Поэтому в процессе анализа и планирования объема реализованных услуг туристской фирмы необходимо учитывать закономерность отклонений показателей отдельных месяцев от среднегодовых показателей.

Эти расчеты производятся на основе коэффициентов сезонности, которые рассчитываются как процентное отношение средних месячных уровней за ряд лет к среднемесячному объему реализованных услуг за весь расчетный период по формуле:

$$k_c = \frac{\acute{o}_1}{\bar{\acute{o}}} \cdot 100\% \tag{4.5}$$

где k_c - коэффициент сезонности, %;

y_1 - среднее уровень объема реализованных услуг отдельного месяца, руб.;

$\bar{y}$ - среднемесячный объем реализованных услуг за расчетный период, руб.

В практике экономического анализа используются различные методы расчета коэффициента сезонности: простой средней, аналитического выравнивания, относительных чисел, скользящей (подвижной) средней, метод У. Пирсонса.

Наиболее легким из них является метод простой средней, который применяется для исчисления сезонных колебаний в тех случаях, когда внутригодовое изменение объема реализации услуг колеблется на протяжении года около определенного (постоянного) уровня.

Индексы сезонности показывают, во сколько раз фактический уровень ряда в момент или интервал времени t больше среднего уровня либо уровня , вычисляемого по уравнению тенденции f(t) . При анализе сезонности уровни временного ряда показывают развитие явления по месяцам (кварталам) одного или нескольких лет . Для каждого месяца (квартала) получают обобщенный индекс сезонности как среднюю арифметическую из одноименных индексов каждого года . Индексы сезонности – это , по либо уровень существу , относительные величины координации , когда за базу сравнения принят либо средний уровень ряда , либо уровень тенденции . Способы определения индексов сезонности зависят от наличия или отсутствия основной тенденции .

Если тренда нет или он незначителен, то для каждого месяца (квартала) индекс рассчитывается по формуле:

$$i_{t,\text{сез}} = \frac{y_t}{y_{\text{ср}}} \tag{4.6}$$

где $У_t$ -уровень показателя за месяц (квартал) t ;

$у_{\text{ср}}$ - общий уровень показателя.

Как отмечалось выше, для обеспечения устойчивости показателей можно взять больший промежуток времени. В этом случае расчет производится по формулам:

$$I_{t,сез} = \frac{\overline{У}_1}{У_{ср}} \text{ либо } I_{t,сез} = \frac{\Sigma i_{t,сез}}{T} \qquad (4.7)$$

где $\overline{У}_t$ - средний уровень показателя по одноименным месяцам за ряд лет ;

Т - число лет.

При наличии тренда индекс сезонности определяется на основе методов, исключающих влияние тенденции. Порядок расчета следующий:

1) для каждого уровня определяют выровненные значения по тренду f(t);

2) рассчитывают отношения $i_t = У_t / f(t)$;

3) при необходимости находят среднее из этих отношений для одноименных месяцев (кварталов) по формуле:

$$i_{t,сез} = \frac{i_t^1 + i_t^2 + \ldots + i_t^T}{T}, \text{(Т - число лет).} \qquad (4.7)$$

Слагаясь под совместным воздействием систематических и случайных факторов, уровень ряда динамики испытывает также воздействие причин, обусловленных периодичностью колебаний.

В рядах внутригодичной динамики, можно выделить три важнейшие составляющие колеблемости уровней временного ряда: тренд, сезонную и случайную компоненты.

Таким образом, при анализе колеблемости динамических рядов наряду с выделением случайных колебаний возникает и задача изучения периодических колебаний. Как правило, изучение периодических («сезонных») колебаний необходимо с целью исключения их влияния на общую динамику для выявления «чистой» (случайной) колеблемости.

В широком понимании к сезонным относят все явления, которые обнаруживают в своем развитии отчетливо выраженную закономерность внутригодичных изменений, т.е. более или менее устойчиво повторяющиеся из года в год колебания уровней. Часто эти колебания могут быть не связаны со сменой времен года. К сезонным явлениям относят, например, потребление электроэнергии; неравномерность производственной деятельности в отраслях пищевой промышленности, связанных с переработкой сельскохозяйственного сырья; перевозки пассажирским транспортом, спрос на многие виды продукции и услуг и т.д.

Как бы ни проявлялась сезонность, она наносит большой ущерб национальной экономике, связанный с неравномерным использованием оборудования и рабочей силы, с неравномерной загрузкой транспорта, необходимостью создания резервов мощностей и т.д. Комплексное регулирование сезонных изменений по отдельным отраслям должно основываться на исследовании сезонных отклонений.

Важнейшими задачами, решаемыми в ходе исследования сезонности, являются следующие:

1) определение наличия сезонности, численное выражение проявления сезонных колебаний и выявление их силы и характера в различных фазах годичного цикла;

2) характеристика факторов, вызывающих сезонные колебания;

3) оценка последствий, к которым приводит наличие сезонных колебаний;

4) математическое моделирование сезонности.

Для измерения **сезонных колебаний** статистикой предложены различные методы. Наиболее простые и часто употребляемые

из них:

а) метод абсолютных разностей;

б) метод относительных разностей;

в) построение индексов сезонности.

Первые два способа предполагают нахождение разностей фактических уровней и уровней, найденных при выявлении основной тенденции развития.

Применяя способ абсолютных разностей, оперируют непосредственно размерами этих разностей, а при использовании метода относительных разностей определяют отношение абсолютных размеров указанных разностей к выравненному уровню. При выявлении основной тенденции используют либо метод скользящей средней, либо аналитическое выравнивание. В некоторых случаях в стационарных рядах можно пользоваться разностью фактических уровней и средним месячным уровнем за год. Использование данных за несколько лет связано с тем обстоятельством, что в отклонениях по отдельным годам сезонные колебания смешиваются со случайными. Чтобы элиминировать случайные колебания, берут средние отклонения за несколько лет.

Для выделения сезонной волны надо определить средний уровень товарооборота за каждый месяц по пятилетним данным и общую среднюю за весь рассматриваемый период.

Общая средняя уд получается делением суммы уровней отпуска за все пять лет на 60 (общее число месяцев).

Затем определяется абсолютное отклонение средних месячных показателей от общей средней $\bar{y}_i - \bar{y}_0$.

Метод относительных разностей является развитием метода абсолютных разностей. Для нахождения относительных разностей абсолютные отклонения делят на общую среднюю и выражают в процентах.

Вместо относительных разностей за каждый месяц может быть вычислен индекс сезонности, который рассчитывается как отношение среднего уровня соответствующего месяца к общей средней, т.е.

$$y_{сез.} = \frac{\bar{y}_i}{\bar{y}_0} \times 100\% . \qquad (4.8)$$

Выделение сезонной волны можно выполнить на основе построения аналитической модели проявления сезонных колебаний.

Построение аналитической модели выявляет основной закон колеблемости данного временного ряда в связи с переходом от месяца к месяцу и дает среднюю характеристику внутригодичных колебаний.

При исследовании явлений периодического типа в качестве аналитической формы развития во времени принимается уравнение следующего типа (ряд Фурье):

$$\hat{y}_t = a_0 + \sum(a_k \cos kt + b_k \sin kt) \tag{4.9}$$

В этом уравнении величина k определяет гармонику ряда Фурье и может быть взята с разной степенью точности (чаще всего от 1 до 4). Для отыскания параметров уравнения используется метод наименьших квадратов, т.е.

$$\sum_1^n (y_i - \hat{y}_t)^2 = \min \tag{4.10}$$

Найдя частные производные этой функции и приравняв их к нулю, получим систему нормальных уравнений, решение которой дает следующие формулы для вычисления параметров:

$$a_0 = \frac{1}{n}\sum_1^n y_i\ ;$$

$$a_k = \frac{2}{n}\sum_1^n y_i \cos kt\,;$$

$$b_k = \frac{2}{n}\sum_1^n y_i \sin kt\,;$$

Параметры уравнения зависят от значений у и связанных с ними последовательных значений cos *kt* и sin *kt*.

Для изучения сезонных колебаний на протяжении года необходимо взять n=12 (по числу месяцев в году). Тогда, представляя периоды как части длины окружности, ряд динамики можно записать в виде таблицы, в первой строке которой будут записаны периоды, а во второй – соответствующие им уровни.

Применяя к этим же данным вторую гармонику ряда Фурье для выражения модели сезонности, получим коэффициенты a_2 и b_2. Подставляя их в уравнение ряда Фурье, будем иметь следующую модель сезонности данного ряда динамики:

$$\hat{y}_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t\,. \tag{4.11}$$

Рассмотрим выявление всех типов колебаний внутригодичной динамики уровней.

Для выравнивания уровней принимаем период сглаживания, равный четырем кварталам (m = 4).

Найденные скользящие средние будут отнесены не к конкретному кварталу, а попадут в промежуток между ними.

Для отнесения скользящей средней к определенному кварталу, находим средние из двух смежных скользящих средних, т.е. производим центрированные средних.

Для выявления сезонной составляющей в колеблемости уровней ряда динамики рассчитываем отношения фактических объемов товарооборота каждого квартала к соответствующей ему скользящей средней.

На основании полученных соотношений выполняется их группировка по кварталам путем занесения значений в таблицу.

Для расчета индекса сезонности на основании сравнений фактических квартальных значений за ряд лет с соответствующей скользящей средней можно воспользоваться следующими приемами:

1) рассчитать для каждого квартала среднюю арифметическую из полученных соотношений.

2) определить медиану из значений индексов сезонности за каждый квартал путем ранжирования.

Т. к. обычно сумма индексов сезонности хотя и незначительно, но отличается от 4 (для четырех кварталов сумма индексов должна быть равна 4, а их средняя равна 1,00), то для устранения этих расхождений определяется поправочный коэффициент как отношение теоретической суммы индексов (4,0) к фактической величине их суммы.

Для расчета индексов сезонности, скорректированных на поправочный коэффициент используются значения медиан.

Прежде чем анализировать основную тенденцию (тренд) или циклические колебания, необходимо исключить сезонную компоненту и проверить гипотезу о существовании тренда.

Для этого можно использовать метод проверки разностей средних уровней. Суть этого метода состоит в делении ряда на две части и нахождении их средних и дисперсий по формулам:

$$\bar{x} = \frac{x_1 + x \ + \ldots + x_x}{n}, \tag{4.12}$$

где n – число уровней ряда;

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n}; \tag{4.13}$$

Затем мы находим расчетное значение с помощью статистики Стьюдента:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}; \tag{4.14}$$

Затем полученное значение сравниваем с критическим табличным значением (число степеней свободы равно n1 + n2 - 2).

Сравнив критическое значение с расчетным, делаем вывод о наличии или отсутствии тренда в ряду динамики.

После ее исключения из колеблемости уровней временного ряда, рассчитаем уравнение тренда, воспользовавшись линейной функцией

$$\hat{y}_t = a + bt, \tag{4.15}$$

где $\sum t_i = 0$;

$$a = \frac{\sum_{i=1}^{n} y_i}{n};$$

$$b = \frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i^2};$$

С помощью полученного уравнения тренда выполним экстраполяцию на один год.

Найденные таким образом значения не учитывают сезонные колебания в объеме товарооборота. Для учета сезонной составляющей уровень, полученный в результате экстраполяции, умножают на индекс сезонности, т.е.

$$\hat{y}' = \hat{y}_t \times i_{\text{сез}} \qquad (4.16)$$

где $\hat{y}'$ - экстраполируемый уровень с учетом сезонных колебаний.

4.3 Анализ тренд – сезонной аддитивной модели

Известно несколько подходов к анализу структуры временного ряда, содержащих сезонные и циклические колебания (моделирование циклических колебаний в целом осуществляется аналогично моделированию сезонных колебаний, поэтому мы рассмотрим только методы моделирования последних).

Простейший подход – расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Построение аддитивной и мультипликативной моделей сводится к расчету значений u_t, s_t, ε_t.

Алгоритм построения тренд – сезонной аддитивной модели

1. сглаживание (выравнивание) ВР с помощью простой скользящей средней. Период скольжения должен быть равен (периодичности сезонных колебаний) 1 году.

Для этого:

1) просуммируем уровни ряда последовательно за каждые четыре квартала (12 месяцев) со сдвигом на один момент времени и определим условные годовые объемы;

2) разделив полученные суммы на 4 (12), найдем скользящие средние. Отметим, что полученные таким образом выровненные значения не содержат сезонной компоненты;

3) если период четный, то проводится центрирование скользящей средней, для чего определяется среднее значение из двух последовательных скользящих средних;

2. рассчитывают абсолютные показатели сезонности:

$$S_i = y_i - \tilde{y} \quad (4.17)$$

где $\tilde{y}$ - выровненные скользящие средние;

3. рассчитываются средние показатели сезонности для одноименных кварталов (месяцев):

$$\bar{S}_i = \frac{1}{n}\sum S_i\ ; \quad (4.18)$$

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

Если $\sum \bar{S}_i \neq 0$, проводится корректировка сезонной компоненты:

$$\hat{S}_i = \bar{S}_i - \frac{1}{n}\sum_{i=1}^{n}\bar{S}_i\ ; \quad (4.19)$$

4. проводим десезоналирование ВР или элиминируем влияние сезонной компоненты: из исходных уровней вычитаем скорректированную сезонную компоненту:

$$y_i - \hat{S}_i\ ; \quad (4.20)$$

5. по десезоналированному ВР проводим аналитическое выравнивание (определяем компоненту u_t);

6. рассчитываем тренд с учетом сезонности:

$$y_S = \hat{y}_t + \hat{S}_i\ . \quad (4.21)$$

4.4 Анализ построения тренд – сезонной мультипликативной модели

При мультипликативной модели уровень ВР можно представить в виде сомножителей:

$$y_i = \hat{y}_t \cdot K_S \cdot E \quad (4.22)$$

где K_S - коэффициент сезонности

E – коэффициент влияния случайности $\left(\frac{y_i}{y_S}\right)$.

Алгоритм построения тренд – сезонной мультипликативной модели:

1. сглаживание ВР с помощью скользящей средней
2. рассчитываем коэффициент сезонности

$$K_S = \frac{y_i}{\tilde{y}_i} \tag{4.23}$$

3. определяем средние показатели сезонности для одноименных кварталов (месяцев):

$$\overline{K}_j = \frac{1}{n}\sum K_{Si} \tag{4.24}$$

4. если при поквартальном наблюдении $\sum \overline{K} \neq 4$, а при помесячном $\sum \overline{K} \neq 12$, то выполняется корректировка коэффициента сезонности:

$$\hat{K}_j = \overline{K}_j \cdot \frac{4(12)}{\sum \overline{K}_j} \tag{4.25}$$

5. исключаем сезонность из уровней ряда:

$$\frac{y_i}{\hat{K}_j}; \tag{4.26}$$

6. проводится аналитическое выравнивание десезоналированного ряда;

7. рассчитываются уровни ВР, обусловленные влиянием тенденции и сезонности:

$$y_S = \hat{y}_t \cdot \hat{K}_j. \tag{4.27}$$

Аддитивная модель целесообразна, если размах сезонных колебаний изменяется слабо.

Вопросы для самопроверки:

1. *В чем заключается проверка предложения о существенности периодической компоненты?*
2. *Что такое сезонность?*
3. *Что позволяет выявить процесс изучения сезонности?*
4. *Что такое индекс сезонности? Для чего он рассчитывается?*
5. *Перечислите задачи, решаемые входе исследования сезонности?*
6. *Какие методы применяются для измерения сезонных колебаний?*
7. *В чем суть методов абсолютных разностей и относительных разностей?*
8. *Что такое ряд Фурье?*
9. *Опишите порядок расчета индекса сезонности?*
10. *Как осуществляется экстраполяция с учетом индекса сезонности?*
11. *Опишите алгоритм анализа тренд – сезонной аддитивной модели?*
12. *Опишите алгоритм анализа тренд – сезонной мультипликативной модели?*

5 Моделирование связных рядов динамики

5.1 Автокорреляция уровней временного ряда и выявление его структуры

Многомерные временные ряды, показывающие зависимость результативного признака от одного или нескольких факторных, называют **связными рядами динамики**. Применение метода наименьших квадратов для обработки рядов динамики не требует выдвижения никаких предположений о законах распределения исходных данных. Однако при использовании метода наименьших квадратов для обработки связных рядов динамики следует учитывать наличие автокорреляции (авторегрессии), которая не учитывалась при обработке одномерных временных рядов динамики, поскольку ее наличие способствовало более плотному и четкому выявлению тенденции развития рассматриваемого социально-экономического явления во времени.

В значительной части рядов динамики экономических процессов между уровнями, особенно близко расположенными, существует взаимосвязь. Ее удобно представить в виде корреляционной зависимости между рядами y_1, y_2, y_3, …, y_n и этим же рядом, сдвинутым относительно первоначального положения на h моментов времени y_{1+h}, y_{2+h}, y_{3+h}, …, y_{n+h}.

Временное смещение называется сдвигом, а само явление взаимосвязи – автокорреляцией.

Автокорреляционная зависимость особенно существенна между последующими и предшествующими уровнями ряда динамики. Поскольку классические методы математической статистики применимы лишь в случае независимости отдельных членов ряда между собой, то при анализе нескольких взаимосвязанных рядов динамики важно установить наличие и степень их автокорреляции.

Различают два вида автокорреляции:

1) автокорреляция в наблюдениях за одной или более переменными;
2) автокорреляция ошибок или автокорреляция в отклонениях от тренда.

При наличии во ВР тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих.

Степень тесноты связи между последовательностями наблюдений временного ряда: $y_1, y_2, ..., y_{n-\tau}$ и $y_{1+\tau}, y_{2+\tau}, ..., y_n$ (сдвинутых относительно друг друга на τ единиц, или, *с лагом* τ) может быть определена с помощью коэффициента корреляции:

$$r(\tau)=\frac{\overline{y_t \cdot y_{t+\tau}}-\bar{y}_t \cdot \bar{y}_{t+\tau}}{\sigma_t \cdot \sigma_{t+\tau}}, \qquad (5.1)$$

где $\overline{y_t \cdot y_{t-\tau}} = \dfrac{\sum_{i=1+\tau}^{n} y_i \cdot y_{i+\tau}}{n-\tau}$

$\bar{y}_t = \dfrac{\sum_{i=1+\tau}^{n} y_i}{n-\tau}$ - средний уровень ряда $y_{1+\tau}, y_{2+\tau},..., y_n$

$\bar{y}_{t+\tau} = \dfrac{\sum_{i=1+\tau}^{n} y_{i+\tau}}{n-\tau}$ - средний уровень ряда $y_1, y_2,..., y_{n-\tau}$

$\sigma_t, \sigma_{t+\tau}$ - средние квадратические отклонения для рядов $y_{1+\tau}, y_{2+\tau},..., y_n$ и $y_1, y_2,..., y_{n-\tau}$ соответственно.

Так как коэффициент $r(\tau)$ измеряет корреляцию между членами одного и того же ряда, его называют коэффициентом *автокорреляции*.

Замечания: 1) коэффициент $r(\tau)$ позволяет судить о наличии линейной (или близко к линейной) тенденции, если соседние уровни ряда тесно коррелируют;

2) по знаку $r(\tau)$ нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда.

Лаг определяет порядок коэффициента автокорреляции. Если $\tau = 1$, то имеем коэффициент автокорреляции 1-го порядка, если $\tau = 2$ - второго порядка и т.д. Следует учитывать, что с увеличением лага на единицу число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается на единицу. Поэтому обычно рекомендуют максимальный порядок коэффициента автокорреляции, равный n/4.

Рассчитав несколько коэффициентов автокорреляции, можно определить лаг τ, при котором автокорреляция $r(\tau)$ наиболее высокая, выявив тем самым структуру ВР. Если наиболее высоким оказывается значение $r(1)$, то исследуемый ряд содержит *только* тенденцию. Если наиболее высоким оказался $r(\tau)$, то ряд содержит помимо тенденции колебания периодом τ. Если ни один из коэффициентов не является статистически значимым, можно сделать одно из предположений:

– либо ряд не содержит тенденции и циклических колебаний

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

Последовательность коэффициентов автокорреляции 1-го, 2-го и т.д. порядков называют *автокорреляционной функцией*. График зависимости значений коэффициентов автокорреляции от величины лага – *коррелограммой.*

5.2 Способы исключения автокорреляции

Сущность всех методов исключения тенденции заключается в том, чтобы устранить или зафиксировать воздействие фактора времени на формирование уровней ряда. Основные методы исключения тенденции можно разделить на две группы.

Первая группа - это методы, основанные на преобразовании уровней исходного ряда в новые переменные, не содержащие тенденции. Полученные переменные используются далее для анализа взаимосвязи изучаемых временных рядов. Эти методы предполагают непосредственное устранение трендовой компоненты T из каждого уровня временного ряда. Два основных метода в данной группе – это метод последовательных разностей и метод отклонений от трендов.

Вторая группа - это методы, основанные на изучении взаимосвязи исходных уровней временных рядов при элиминировании воздействия фактора времени на зависимую и независимые переменные модели. В первую очередь, это метод включения в модель регрессии по временным рядам фактора времени. Рассмотрим методику применения, преимущества и недостатки каждого из перечисленных выше методов.

5.3 Методы, основанные на преобразовании уровней исходного ряда в новые переменные, не содержащие тенденции

Метод отклонений от тренда

Пусть имеются два временных ряда xt и yt, каждый из которых содержит трендовую компоненту T и случайную компоненту ε. Проведение аналитического выравнивания по каждому из этих рядов позволяет найти параметры соответствующих уравнений трендов и определить расчетные по тренду уровни $\tilde{o}_t$ и $\hat{y}_t$ соответственно. Эти расчетные значения можно принять за оценку трендовой компоненты T каждого ряда. Поэтому влияние тенденции можно устранить путем вычитания расчетных значений уровней ряда из фактических. Эту процедуру проделывают для каждого временного ряда в модели.

Дальнейший анализ взаимосвязи рядов проводят с использованием не исходных уровней, а отклонений от тренда $x_t - \tilde{o}_t$ и $y_t - \hat{y}_t$ при условии, что последние не содержат тенденции. Т. е. уравнение регрессии строится в виде

$$y_t - \hat{y}_t = a + b(x_t - \tilde{o}_t) \qquad (5.2)$$

Метод последовательных разностей

Если временной ряд содержит ярко выраженную полиномиальную тенденцию (в виде полинома от t), то с целью устранения тенденции можно применить метод последовательных разностей, заключающийся в замене исходных уровней ряда последовательными разностями соответствующих порядков (порядок разности равен порядку полинома).

Если временной ряд содержит ярко выраженную линейную тенденцию, ее можно устранить путем замены исходных уровней ряда цепными абсолютными приростами – первыми последовательными разностями

$$\Delta y_t = y_t - y_{t-1} \tag{5.3}$$

Если временной ряд содержит тенденцию в форме параболы второго порядка, то для ее устранения можно заменить исходные уровни ряда на вторые разности

$$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1} \tag{5.4}$$

Если тенденции временного ряда соответствует экспоненциальный или степенной тренд, метод последовательных разностей следует применять не к исходным уровням ряда, а к их логарифмам.

При всей своей простоте метод последовательных разностей имеет два существенных недостатка. Во-первых, его применение связано с сокращением числа пар наблюдений, по которым строится уравнение регрессии, и, следовательно, с потерей числа степеней свободы. Во-вторых, использование вместо исходных уровней временных рядов их приростов или ускорений приводит к потере информации, содержащейся в исходных данных.

5.4 Включение в модель регрессии фактора времени

В корреляционно-регрессионном анализе устранить воздействие какого-либо фактора можно, если зафиксировать воздействие этого фактора на результат и другие включенные в модель факторы. Этот прием широко используется в анализе временных рядов, когда тенденция фиксируется через включение фактора времени в модель в качестве независимой переменной. Модель вида:

$$y_t = a + b_1 x_t + b_2 t + \varepsilon_t\,, \tag{5.5}$$

относится к группе моделей, включающих фактор времени. Очевидно, что число независимых переменных в такой модели может быть больше единицы. Кроме того, это могут быть не только текущие, но и лаговые значения независимой переменной, а также лаговые значения результативной переменной.

Преимущество данной модели по сравнению с методами отклонений от трендов и последовательных разностей в том, что она позволяет учесть всю информацию, содержащуюся в исходных данных, поскольку значения xt и yt есть уровни исходных временных рядов. Кроме того, модель строится по всей совокупности данных за рассматриваемый период в отличие от метода последовательных разностей, который приводит к потере числа наблюдений. Параметры *a* и *b* модели с включением фактора времени определяются обычным МНК.

Интерпретация параметров уравнения регрессии – это параметр *b1* показывающий, насколько в среднем изменится значение результативного признака yt при увеличении фактора на единицу при неизменной величине других факторов. И параметр *b2* показывающий, насколько в среднем за год изменится значение результативного признака yt за счет воздействия всех факторов, кроме фактора xt.

5.5 Автокорреляция в остатках. Критерий Дарбина–Уотсона

Рассмотрим уравнение регрессии вида:

$$\hat{y}_t = a + \sum_j^k b_j \cdot x_{jt} + \varepsilon_t \qquad (5.6)$$

где *k* – число независимых переменных модели.

Для каждого момента (периода) времени *t* = 1,..., n значение компоненты ε_t определяется из соотношения

$$\varepsilon_t = y_t - \hat{y}_t = y_t - (a + \sum_j^k b_j \cdot x_{jt}) \qquad (5.7)$$

Рассматривая последовательность остатков как временной ряд, можно построить график их зависимости от времени. В соответствии с предпосылками МНК остатки ε_t должны быть случайными. Однако при моделировании временных рядов нередко встречается ситуация, когда остатки содержат тенденцию или циклические колебания. Что свидетельствует о том, что каждое следующее значение остатков зависит от предшествующих. В этом случае говорят о наличии автокорреляции остатков.

Автокорреляция остатков может быть вызвана несколькими причинами имеющими различную природу:

1) наличие ошибок измерения в значениях результативного признака;

2) модель может не включать фактор, окапывающий существенное воздействие на результат, влияние которого отражается в остатках, вследствие чего последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени *t*. Кроме того, в качестве таких существенных факторов могут выступать лаговые значения переменных, включенных в модель;

3) модель не учитывает несколько второстепенных факторов, совместное влияние которых на результат существенно ввиду совпадения тенденций их изменения или фаз циклических колебаний;

4) неправильная спецификация функциональной формы модели. В этом случае следует изменить форму связи факторных и результативного

признаков, а не использовать специальные методы расчета параметров уравнения регрессии при наличии автокорреляции остатков.

Существуют два наиболее распространенных метода определения автокорреляции остатков. Первый метод — это построение графика зависимости остатков от времени и визуальное определение наличия или отсутствия автокорреляции. Второй метод – использование критерия Дарбина — Уотсона и расчет величины:

$$d = \frac{\sum_{i=1}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{i=1}^{n}\varepsilon_t^2} \qquad (5.8)$$

Согласно (5.8) величина d есть отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии. Практически во всех статистических ППП значение критерия Дарбина – Уотсона указывается наряду с коэффициентом детерминации, значениями t- и F-критериев.

Коэффициент автокорреляции остатков первого порядка определяется как:

$$r_1^{\varepsilon} = \frac{\sum_{i=1}^{n}(\varepsilon_t - \bar{\varepsilon}_1)(\varepsilon_{t-1} - \bar{\varepsilon}_2)}{\sqrt{\sum_{i=2}^{n}(\varepsilon_t - \bar{\varepsilon}_1)^2 \cdot \sum_{i=2}^{n}(\varepsilon_{t-1} - \bar{\varepsilon}_2)^2}}$$

где

$$\bar{\varepsilon}_1 = \frac{\sum_{i=2}^{n}\varepsilon_t}{n-1}; \; \bar{\varepsilon}_{21} = \frac{\sum_{i=2}^{n}\varepsilon_{t-1}}{n-1}. \qquad (5.9)$$

Между критерием Дарбина–Уотсона и коэффициентом автокорреляции остатков первого порядка имеет место следующее соотношение: $d \approx 2 \cdot (1 - r_1^{\varepsilon})$.

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

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина–Уотсона следующий. Выдвигается гипотеза H0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H1 и H1* состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках. Далее по таблице (приложение 3) определяются критические значения критерия Дарбина–Уотсона dL и dU для заданного числа наблюдений n, числа независимых переменных модели k и уровня значимости a. По этим значениям числовой промежуток [0;4] разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью (1–a) рассматривается на рис. 5.1.

Есть положительная автокорреляция остатков. H_0 отклоняется. С вероятностью $P = (1-\alpha)$ принимается H_1.	Зона неопреде-ленности	Нет оснований отклонять H_0 (автокорреля-ция остатков от-сутствует)	Зона неопреде-ленности	Есть отрицательная автокорреляция остатков. H_0 отклоняется. С вероятностью $P = (1-\alpha)$ принимается H_1*

0 d_L d_U 2 $4-d_U$ $4-d_L$
4

Рисунок 5.1. Механизм проверки гипотезы о наличии автокорреляции остатков

Если фактическое значение критерия Дарбина – Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу H0.

5.6 Оценивание параметров уравнения регрессии при наличии автокорреляции в остатках

Обратимся к уравнению регрессии:

$$\acute{o}_t = a + b \cdot x_t + \varepsilon_t \tag{5.10}$$

Примем некоторые допущения относительно этого уравнения:

1. пусть yt и xt не содержат тенденции, например, представляют собой отклонения выравненных по трендам значений от исходных уровней временных рядов;

2. пусть оценки *a* и *b* параметров уравнения регрессии найдены обычным МНК;

3. пусть критерий Дарбина – Уотсона показал наличие автокорреляции в остатках первого порядка.

Основной подход к оценке параметров модели регрессии в случае, когда имеет место автокорреляция остатков, заключается в следующем: исходная модель регрессии (5.10) с помощью замены переменных приводится к виду:

$$\acute{o}_t' = a' + b \cdot x_t' + u_t \tag{5.11}$$

где

$$\begin{aligned} y_t' &= y_t - r_1^{\varepsilon} \cdot y_{t-1}; \quad x_t' = x_t - r_1^{\varepsilon} \cdot x_{t-1}; \\ u_t &= \varepsilon_t - r_1^{\varepsilon} \cdot \varepsilon_{t-1} \ ; \ a' = a(1 - r_1^{\varepsilon}). \end{aligned} \tag{5.12}$$

Здесь $r^{\varepsilon}{}_1$ – коэффициент автокорреляции первого порядка.

Поскольку ut, – случайная ошибка, то для оценки параметров преобразованного уравнения можно применять обычный МНК.

Итак, если остатки по исходному уравнению регрессии содержат автокорреляцию, то для оценки параметров уравнения используют обобщенный МНК.

Его реализация разбивается на следующие этапы:

1. Перейти от исходных переменных yt и xt к переменным y’t и x’t по формулам (5.12).

2. Применив обычный МНК к уравнению (5.11), определить оценки параметров *a’* и *b*.

3. Рассчитать параметр a исходного уравнения из соотношения (5.6) как

$$a = a'/(1 - r_1^{\varepsilon}) \tag{5.13}$$

4. Выписать исходное уравнение (5.10).

Обобщенный метод наименьших квадратов аналогичен методу последовательных разностей. Однако мы вычитаем из yt (или xt) не все значение предыдущего уровня yt-1 (или xt-1), а некоторую его долю $r_1^{\varepsilon} \cdot y_{t-1}$ (или $r_1^{\varepsilon} \cdot x_{t-1}$). Если $r_1^{\varepsilon} = 1$, то данный метод есть просто метод первых разностей, так как $y'_t = y_t - y_{t-1}$; $x'_t = x_t - x_{t-1}$.

Поэтому в случае, если значение критерия Дарбина – Уотсона близко к нулю, применение метода первых разностей вполне обоснованно.

Основная проблема, связанная с применением данного метода, заключается в том, как получить оценку r_1^{ε}. Основными способами являются оценка этого коэффициента непосредственно по остаткам, полученным по исходному уравнению регрессии, и получение его приближенного значения из соотношения между коэффициентом автокорреляции остатков первого порядка и критерием Дарбина–Уотсона $r_1^{\varepsilon} = 1 - d/2$.

Вопросы для самопроверки:

1. *Что такое связные ряды динамики?*
2. *Что такое автокорреляция? Какие виды выделяют?*
3. *С помощью какого коэффициента может быть определена степень тесноты связи между последовательными наблюдениями?*
4. *Какие выделяют группы методов исключения тенденции?*
5. *Опишите метод отклонения от тренда?*
6. *Опишите метод последовательных разностей?*
7. *В чем смысл включения в модель регрессии фактора времени?*
8. *Какие причины вызывают автокорреляцию остатков?*
9. *Для чего рассчитывается критерий Дарбина – Уотсона? Опишите методику его расчета?*
10. *В чем заключается суть оценивания параметров уравнения регрессии при наличии автокорреляции в остатках?*

6 Моделирование тенденции и колеблемости ряда динамики

6.1 Понятие колеблемости во временных рядах

Показатели силы и интенсивности колебаний аналогичны по построению, по форме показателям силы и интенсивности вариации признака в пространственной совокупности. По существу они отличаются тем, что показатели вариации вычисляются на основе отклонений от постоянной средней величины, а показатели, характеризующие колеблемость уровней временного ряда, - по отклонениям отдельных уровней от тренда, который можно считать «подвижной средней величиной».

Абсолютные показатели колеблемости:

1. Амплитуда (размах) колебаний: $A = Y_{\max} - Y_{\min}$ (6.1)

Это разность между наибольшим и наименьшим по алгебраической величине отклонений от тренда.

2. Средняя по модулю отклонений от тренда: $\alpha(t) = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$ (6.2)

3. Среднее квадратическое отклонение уровней ряда от тренда:

$$S_y(t) = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{\hat{y}}_i)^2}{n-p}} \qquad (6.3)$$

где y_i - фактический уровень;

$\hat{\hat{y}}_i$ - выровненный уровень;

n -число уровней;

p - число параметров тренда;

t - номера лет (знак отклонения от тренда).

Относительные показатели

$$V(t) = \frac{S(t)}{\bar{y}} * 100 \qquad (6.4)$$

где $\bar{y}$ - средний уровень ряда.

Этот показатель отражает величину колеблемости в сравнений со средним уровнем ряда. Он необходим для сравнения двух различных явлений и чаще всего выражаются в %.

Устойчивость временного ряда – это наличие необходимой тенденции изучаемого статистического показателя с минимальным влиянием на него неблагоприятных условий.

Из этого вытекают основные требования устойчивости:

- минимизация колебаний уровней временного ряда;

- наличие определенной, необходимой для общества тенденции изменения.

Устойчивость временного ряда можно оценить на различных явлениях. При этом в зависимости от явления будет меняться показатели, которые используются в качестве форм выражения существа исследуемого процесса, но содержание понятия устойчивости будет оставаться неизменным.

Наиболее простым, аналогичным процессу вариации при изменении устойчивости уровней временного ряда, является размах колеблемости средних уровней за благоприятный и неблагоприятные, в отношении к изучаемому явлению, периоды времени:

$$R_{\bar{y}} = \bar{Y}_{\text{áëàã}} - \bar{Y}_{\text{íåáëàã}} \tag{6.5}$$

К благоприятным периодам времени относятся все периоды с уровнями выше тренда, к неблагоприятным – ниже тренда (однако, например, при изучении динамики производительности труда если это трудоемкость, то все должно быть наоборот).

Отношение средних уровней за благоприятные периоды времени к средним уровням за неблагоприятные также может служить показателем устойчивости уровней. Чем ближе отношение к единице, тем меньше колеблемости и соответственно выше устойчивости. Назовем это отношение индексом устойчивости динамических рядов:

$$i_{\bar{y}} = \frac{\bar{y}_{благ}}{\bar{y}_{неблаг}} \tag{6.6}$$

При изменении колеблемости уровней исчисляются обобщающие показатели отклонений уровней от тренда за исследуемый период.

Основным абсолютным показателем является среднее линейное отклонение:

$$a(t) = \frac{\sum_{i=1}^{n} |y_i - \tilde{y}_i|}{n - p} \tag{6.7}$$

Он выражается в единицах измерения анализируемых уровней и не может, служить для сравнения колебаний различных динамических рядов.

Величину $K_y = (100 - V_y(t))$ (6.8)

называют <u>коэффициентом устойчивости</u>. Такое определение коэффициента устойчивости интерпретируется как обеспечение устойчивости уровней ряда относительно лишь в $(100 - V_y(t))$ случаях.

Наиболее простым показателем устойчивости тенденции временного ряда является <u>коэффициент критерия Спирмена</u> $Kp = 1 - \frac{6\sum_{i=1}^{n} d^2}{n^3 - n}$ (6.9)

где d - разность рангов уровней изучаемого ряда (P_y) и рангов номеров периодов или момент времени в ряду (P_t);

n – число таких периодов или моментов.

Для определения коэффициента Спирмена величины уровней изучаемого явления y_i нумеруются в порядке возрастания, а при наличии одинаковых уровней им присваивается определенный ранг, равный частному от деления суммы рангов, приходящихся на эти значения, на число этих равных значений.

Коэффициент рангов периодов времени и уровней динамического ряда может применить значения в пределах от 0 до ± 1.

Интерпретация этого коэффициента такова: если каждый уровень ряда исследуемого периода выше, чем предыдущего, то ранги уровней ряда и номера лет совпадают, $Kp = +1$. Это означает полную устойчивость самого факта роста уровней ряда, непрерывность роста.

Чем ближе Kp к +1, тем ближе рост уровней к непрерывному, выше устойчивость роста. При $Kp = 0$ рост совершенно неустойчив. При отрицательных значениях чем ближе Kp к -1, тем устойчивее снижение изучаемого показателя.[3]

В качестве характеристики устойчивости изменения можно применить индекс корреляции:

$$J_r = \sqrt{1 - \frac{\sum (y_i - \tilde{y}_i)^2}{\sum (y - \bar{y})^2}} \qquad (6.10)$$

где y_i - уровни динамического ряда;

$\bar{y}$ - средний уровень ряда;

$\tilde{y}_i$ - теоретические уровни ряда.

Индекс корреляции показывает степень сопряженности колебаний исследуемых показателей с совокупностью факторов, изменяющих во времени. Приближение индекса корреляции к 1 означает большую устойчивость изменения уровней динамического ряда. Сравнение индексов корреляции по разным показателям возможно лишь при условии равенства числа уровней.

6.2 Вероятностная оценка существенности параметров тренда и их колеблемости

Статистика лишь в виде редкого исключения может вести анализ, какого- то процесса от начала и до конца. Обычно исходный временной ряд – это лишь выборка во времени, отражающая некоторый этап или просто отрезок развития данного процесса и его показателей. Однако задача исследования может заключаться не только в получении характеристики процесса на ограниченном отрезке времени (показателей гипотетической генеральной совокупности).

Наличие тренда или его отсутствие на изучаемом отрезке времени может быть доказано лишь с некоторой вероятностью, для чего используется специальные критерии.

Вероятностная оценка любого выборочного показателя осуществляется путем сравнения его величины с величиной средней квадратической ошибки (среднего квадратического отклонения выборочных показателей при данном типе и объеме выборки от генерального показателя).

Надежность следует проверять для основного параметра тренда: при параболе 2-го порядка $(a\ )$, коэффициента роста при экспоненте (b).

Средняя ошибка репрезентативность параметра a параболического тренда определяется по формуле:

$$m_{a_2} = \frac{S(t)}{\sqrt{\overline{(t_i - \bar{t})^4} - \left[\overline{(t_i - \bar{t})^2}\right]^2}} \tag{6.11}$$

$\overline{(t_i - \bar{t})^4} - \left[\overline{(t_i - \bar{t})^2}\right]^2$ - рассчитывается при отсчете t_i от начала ряда или $\bar{t}_i^{\,4} - \left(\bar{t}_i^{\,2}\right)^2$ - при отсчете t_i от середины ряда.

Отношение среднегодового изменения к его средней ошибке – это t – критерий Стьюдента:

$$t = \left|\frac{a_t}{m_{a_2}}\right| \tag{6.12}$$

Величину критерия сравниваем с табличной величиной критерия Стьюдента для (n-p(число параметров тренда)) степеней свободы. Если фактические значения критерия намного больше табличного, значит, вероятность нулевой гипотезы (о равенстве параметра a_t нулю) чрезвычайно мала, т.е. наличие тренда достоверна.

Чтобы сравнить показатели колеблемости разных временных рядов и сделать вывод об изменении интенсивности и силы колебаний с течением времени, необходимо использовать методы вероятностной оценки среднего квадратического отклонения уровней от тренда и коэффициента колеблемости.

Средняя ошибка репрезентативности выборочной оценки генерального квадратического отклонения от тренда при нормальном распределении имеет вид:

$$m_{s(t)} = \frac{S(t)}{\sqrt{2n}} \tag{6.13}$$

где $S(t)$ - среднее квадратическое отклонение уровней от тренда;

n - число уровней ряда.

Критерий Стьюдента – отношение среднего квадратического отклонения уровней от тренда к его средней ошибке:

$$t = \frac{S(t)}{m_{s(t)}} = \sqrt{2n} \tag{6.14}$$

Используя готовую таблицу для оценки надежности отличия генерального среднего квадратического отклонения уровня от нуля, можно сказать, что если число уровней больше пяти, то считают надежно установленным отличием $S(t)$ от нуля.

Доверительные границы среднего квадратического отклонения уровней от тренда с заданной вероятностью определяется по формуле:

$$S(t) \pm t_a * m_{s(t)} \tag{6.15}$$

Средняя ошибка репрезентативности выборочной оценки генерального коэффициента колеблемости имеет вид:

$$m_{v(t)} = \frac{V(t)}{\sqrt{2n}} * \sqrt{1 + 2\left(\frac{V(t)}{100}\right)^2} \tag{6.16}$$

где $V(t)$ - коэффициент колеблемости, %.

Доверительные границы коэффициента колеблемости: $V(t) \pm t_a * m_{v(t)}$ (6.17)

При сравнении двух значений показателей колеблемости, которые требуют вероятностной оценки для того, чтобы определить надежно ли изменение величины $S(t)$ в сравнении с прошлым годом нужно проверить нулевую гипотезу о случайном различии величины $S(t)$ двух периодов. Для решения задачи о различии двух или более дисперсии применяется критерий Бартлетта М., он основан на том, что если сравниваемые величины равны, то их арифметическая средняя (взвешенная или простая) равна их геометрической средней, а если величины различаются, то чем больше они различаются, тем больше и различие между арифметической и геометрической средними. [4]

Взвешенная арифметическая средняя дисперсия равна:

$$\bar{S}(t)^2{}_{арифм} = \frac{\sum S(t)_i{}^2 * n}{\sum_{i=1}^{k} n_i} \tag{6.18}$$

где k - число дисперсии;

n_i - их веса, число уровней в подпериодах.

Взвешенная геометрическая средняя:

$$\bar{S}(t)^2{}_{геом} = \sqrt[\sum_{i=1}^{k} n_i]{\Pi_{i=1}^{k}\left[S(t)^2\right]^{n_i}} \tag{6.19}$$

Критерий Бартлетта М имеет вид:

$$M = \ln \frac{\bar{S}(t)^2{}_{арифм}}{\bar{S}(t)^2{}_{геом}} \cdot \sum_{i=1}^{k} n_i \tag{6.20}$$

Его средняя ошибка:

$$C = 1 + \frac{\sum_{i=1}^{k} \frac{1}{n_1} - \frac{1}{\sum n_1}}{3(k-1)} \tag{6.21}$$

При сравнении двух дисперсий и равном числе в каждом педпериоде (средние будет невзвешенной) формулы упрощаются:

$$M = 2\ln \frac{S(t)^2{}_{арифм}}{S(t)^2{}_{геом}} \cdot \sum_{i=1}^{k} n_i \tag{6.22}$$

$$C = 1 + \frac{\frac{1}{n} - \frac{1}{2n}}{3} \qquad (6.23)$$

Отношение $\frac{M}{C}$ имеет распределение χ^2 (хи-квадрат) с числом степеней свободы $k-1$. Табличное значение χ^2 сравнивают с расчетным, если фактические больше табличного, то различие в значениях $S(t)$ неслучайно.

Вопросы для самопроверки:

1. *Что такое колеблемость?*
2. *Какие показатели относятся к абсолютным показателям колеблемости?*
3. *Какие показатели относятся к относительным показателям колеблемости?*
4. *Что такое устойчивость временного ряда? Основные требования устойчивости?*
5. *Для чего рассчитывается коэффициент критерия Спирмена?*
6. *В чем суть вероятностной оценки существенности параметров тренда и их колеблемости?*
7. *Какой критерий используется для решения задачи о различии двух и более дисперсий? Опишите методику расчета?*

7 Статистические методы прогнозирования динамики

7.1 Понятие и классификация статистических методов прогнозирования

Статистический прогноз (conditional prediction) — прогноз, применяемый тогда, когда предусматривается, что лицо, принимающее решение, может осуществлять различные меры, которые способны воздействовать на прогнозируемые показатели.

Различают активный и пассивный статистический прогноз. Например, если наблюдается неблагоприятная тенденция к понижению фондоотдачи, то пассивный прогноз предскажет дальнейшее снижение этого показателя. Активный же прогноз ответит на вопрос, что будет, если окажется принятой та или иная программа действий по повышению эффективности фондов.

Для реализации статистического прогноза используются определенные методы. Выделяют следующие методы статистического прогнозирования (рисунок 7.1).

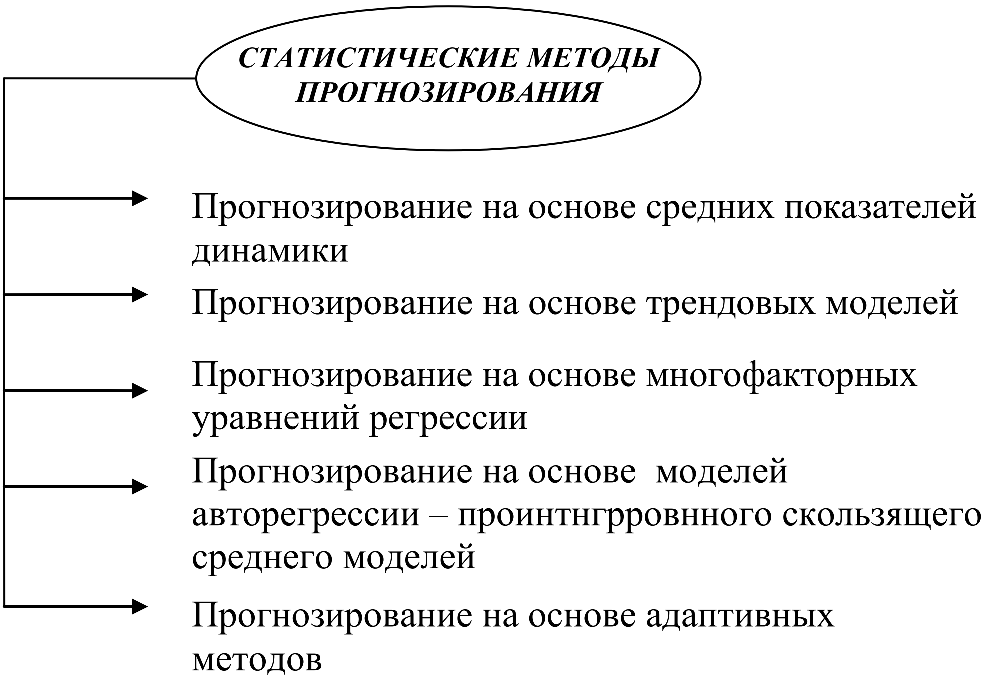

Рисунок 7.1 – Статистические методы прогнозирования, используемые в анализе финансового состояния организации

Статистические методы прогнозирования дополняются эконометрическими методами анализа, статистическими методами прогнозирования, приемами эконометрического моделирования и другими.

Рассмотрим более детально методы прогнозирования, представленные на рисунке 7.1.

7.2 Прогнозирование на основе средних показателей динамики

Простейшими методами прогнозирования динамики являются методы прогнозирования на основе среднего абсолютного прироста и среднего темпа роста.

Прогноз на основе абсолютного прироста (7.1).

$$S_t = S_0 + \overline{\Delta}_s \cdot t, \qquad (7.1)$$

где S_t - прогнозируемый уровень;

S_0 - конечный уровень исходного ряда;

$\overline{\Delta}_s$ - средний абсолютный прирост за ряд лет, примыкающий к началу прогнозного периода: $\overline{\Delta}_s = \dfrac{S_n - S_0}{n-1}$

t – номер периода, на которое делается прогноз.

Прогноз на основе среднего темпа роста (7.2).

$$S_t = S_0 \cdot \overline{T}_s^{\,t}\ , \qquad (7.2)$$

где S_t - прогнозируемый уровень;

S_0 - - конечный уровень исходного ряда;

$\overline{T}_s^{\,t}$ - средний темп роста численности за ряд лет, примыкающих к началу прогнозируемого периода: $\overline{T}_s^{\,t} = \sqrt[n-1]{\dfrac{S_n}{S_0}}$

t – номер периода, на которое делается прогноз.

7.3 Прогнозирование на основе трендовых моделей

Для описания тенденции временного ряда на практике используют модели кривых роста, описывающиеся формулой вида:

$$y = f(x) \qquad (7.3)$$

При таком научном подходе предполагается, что изучаемое явление связано с течением времени. Точно выбранная модель кривой роста должна соответствовать характеру изменений изучаемого явления. Прогнозирование на основе модели кривой роста основано на экстраполяции, т.е. на продлении в будущее тенденции, наблюдавшейся за предыдущие периоды времени.

Проблема выбора кривой роста – первоочередная при прогнозировании, т.к. от ее правильности зависит точность исследования в целом и, соответственно, точность прогноза.

В современной литературе описано несколько десятков кривых роста, к ним относятся: линейные (прямая) и нелинейные (гипербола, парабола, экспонента, логарифма, кривая Гомперца и т.д.)

Процедура разработки прогноза с использованием кривых роста включает несколько этапов:

Этап 1. определение наличия (отсутствия) тенденции в исходном динамическом ряду

↓

Этап 2. подбор кривых роста, описывающих имеющийся временной ряд

↓

Этап 3. оценка параметров кривых роста

↓

Этап 4. оценка точности и адекватности модели

↓

Этап 5. расчет точечного и интервального прогнозов

↓

Этап 6. принятие управленческих решений

Этап 1. Для того чтобы анализировать исходные данные и прогнозировать их дальнейшее развитие, в первую очередь необходимо выяснить существует ли тенденция вообще в изучаемом явлении. Это осуществляется путем проверки статистической гипотезы о случайности ряда. Для этого могут быль использованы критерий серий, основанный на медиане выборки, критерий «восходящих и нисходящих» серий или метод Фостара-Стюрта.

Этап 2. Если тенденция существует, то переходим к выбору кривой роста, наиболее точно описывающей процесс изменения в исследуемом явлении. После выявления тенденции приступаем к выбору лучшего уравнения тренда. Лучшим считается то уравнение тренда, в котором коэффициент апроксимациии очень близок к единице. Существует несколько методов облегчающих процесс выбора формы кривой роста. Наиболее простой метод – визуальный, опирающийся на графическое изображении.

Самым простым типом тренда является прямая линия, описываемая линейным уравнением тренда:

$$\tilde{y}_i = a + b \cdot t_i, \qquad (7.4)$$

где $\tilde{y}_i$ - выровненные, т.е. лишенные колебаний, уровни тренда для лет с номером i;

a - свободный член уравнения, численно равный среднему выровненному уровню для момента или периода времени, принятого за начало отсчета;

b - средняя величина изменения уровней ряда за единицу изменения времени.

Основные свойства тренда в форме прямой линии таковы:

- равные изменения за равные промежутки времени;
- если средний абсолютный прирост - положительная величина, то относительные приросты или темпы прироста постепенно уменьшаются;
- если тенденция к сокращению уровней, а изучаемая величина является по определению положительной, то среднее изменение b не может быть больше среднего уровня a;
- при линейном тренде ускорение, т.е. разность абсолютных приростов за последние периоды, равно 0.

Помимо прямолинейного тренда в статистике при анализе временных рядов также могут использовать параболический, экспоненциальный, логарифмический, гиперболический и логистический тренды.

Этап 3. Оценка параметров кривых роста осуществляется с помощью метода наименьших квадратов (МНК), суть которого сводится к минимизации квадрата отклонений фактических и теоретических значений временного ряда:

$$\sum (y_i - \tilde{y}_i)^2 \to \min \qquad (7.5)$$

В результате минимизации получается система нормальных уравнений, имеющая решение. Решение системы нормальных уравнений позволяет вычислить оценки искомых коэффициентов.

Этап 4. Проверка адекватности выбранных моделей реальному процессу строится при анализе случайных компонент. Ряд остатков как отклонение физических уровней от выровненных:

$$e_t = y_t - \hat{y}_t \qquad (7.6)$$

Принято считать, что модель адекватно описываемому процессу, если значение остаточной компоненты подчиняется случайному закону распределения.

Если вид функции выбран неудачно, то последовательные значения ряда остатков могут не обладать свойствами независимости, так как они могут коррелировать между собой, т.е. будет иметь место автокорреляции ошибок. Существует несколько методов обнаружения автокорреляции. Наиболее распространен критерий Дарбина – Уотсона:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^{\,2}} \approx 2(1 - r_t) \qquad (7.7)$$

где r_t - коэффициент автокорреляции первого порядка.

Если в ряду остатков имеется сильная положительная автокорреляция, то d=0. Если сильная отрицательная, то d=4. При отсутствии автокорреляции d=2. Применение на практике данного критерия основано на сравнении величины d с теоретическими табулированными значениями d_1 и d_2.

Если $d \langle d_1$, то нулевая гипотеза о независимости случайных отклонений отвергается (отсутствие автокорреляции).

Если $d > d_2$, то нулевая гипотеза не отвергается.

Если $d_1 \le d \le d_2$, нет достаточных оснований для принятия решений.

Когда расчетное значение d превышает 2, то с d_1 и d_2 сравнивается не сам коэффициент d, а $(4-d)$.

Важнейшими характеристиками качества моделей являются показатели ее точности:

Абсолютная ошибка прогноза: $$\Delta t = \hat{y}_t - y_t \qquad (7.8)$$

Относительная ошибка прогноза: $$\delta_t = \frac{\hat{y}_t - y_t}{y_t} * 100 \qquad (7.9)$$

Средняя абсолютная ошибка по модулю: $$S_t = \frac{\sum |\hat{y}_t - y_t|}{n} \qquad (7.10)$$

Средняя относительная ошибка по модулю: $$|\bar{\delta}_t| = \frac{1}{n} \sum \left| \frac{\hat{y}_t - y_t}{y_t} \right| * 100 \qquad (7.11)$$

Если $|\bar{\delta}_t| < 10\%$, это свидетельствует о высокой точности. Если $10\% \le |\bar{\delta}_t| \le 20\%$ - точность хорошей модели. Если $|\bar{\delta}_t| > 20\%$, но $|\bar{\delta}_t| < 50\%$ - удовлетворительная точность.

Этап 5. После проведенного предварительного анализа переходим к прогнозированию простой трендовой модели. Простая трендовая модель динамики – это уравнение тренда с указанием начала отсчета единиц времени. Прогноз по этой модели заключается в подстановке в уравнение тренда номера периода, который прогнозируется. Такой прогноз называется точечным. Точечный прогноз – «это скорее абстракция, чем реальность» [2], поэтому необходимо также делать интервальный прогноз. При таком прогнозе учитываются как вызванная колеблемостью ошибка репрезентативности выборочной оценки тренда, так и колебания уровней в отдельные периоды (моменты) относительно тренда.

Изучение закономерностей развития явлений, выявление и характеристика трендов создают базу для прогнозирования, т.е. для определения ориентированных размеров явлений в будущем.

Экстраполяция применяется в перспективном прогнозировании. Предполагает, что установленная тенденция в прошлом периоде будет сохраняться и в будущем. Для этих целей могут быть использованы выравнивание уровней динамического ряда по способу наименьших квадратов и подстановка в полученное уравнение соответствующих значений t. [6]

Прежде всего, вычисляют точечный прогноз – значение уравнения тренда, получаемое при подстановке в уравнение тренда номера прогнозируемого года t_m, однако параметры тренда, вычисленные по ограниченному периоду – это лишь выборочные оценки генеральных параметров. Прогноз должен иметь вероятностный характер, как любое суждение о будущем. Для этого вычисляется средняя ошибка прогноза положения тренда на год за номером t_m, обозначающая m_y, по формуле:

$$m_y = S(t) \cdot \sqrt{\frac{1}{N} + \frac{(t_m - \bar{t})^2}{\sum (t_i - \bar{t})^2}} \tag{7.12}$$

где N- число уровней исходного ряда;

t_m - номер прогнозируемого года;

S(t) – среднее квадратическое отклонение уровней от тренда.

$$S_{\bar{y}} = \sqrt{\frac{\sum_1^n (y_t - \hat{y}_t)^2}{n-p}} \tag{7.13}$$

где n – число уровней; p – число параметров тренда;

$y_t, \hat{y}_t$ - соответственно фактические и расчетные значения уровней динамического ряда.

Для вычисления доверительного интервала прогноза положения тренда среднюю ошибку необходимо умножить на величину t – критерия Стьюдента, при имеющимся числе степеней свободы колебаний и при выбранной вероятности (надежности прогноза). Следовательно, доверительный интервал прогноза положения тренда вычисляется по формуле:

$$y_{точ \pm t_{ст} \cdot m_y} \tag{7.14}$$

где $y_{точ}$ - точечный прогноз.

где t_α - доверительная величина (надежностью 95%) и (n-1)- степенями свободы.

Прогнозирование по тренду имеет качественное ограничение: оно допустимо в условиях сохранения основной тенденции.

<u>Этап 6.</u> После получения прогноза можно принимать управленческие решения в зависимости от полученных значений.

Вопросы для самопроверки:

1. *Что такое статистический прогноз?*
2. *Какие существуют методы прогнозирования?*
3. *Каким образом осуществляется процесс прогнозирования по средним показателям динамики?*
4. *Какие этапы включает в себя процедура разработки прогноза с использованием кривых роста?*
5. *Каковы свойства тренда в форме прямой?*
6. *Какие существуют показатели адекватности выбранной модели реальному процессу?*
7. *Как осуществляется экстраполяция по уравнению тренда?*

8 Методы моделирования взаимосвязи

8.1 Сущность, задачи и методы корреляционно-регрессивного анализа

Изучение зависимости вариации признака от окружающих условий составляет содержание теории корреляции («корреляция» – соотношение, соответствие).

В действительности нетрудно заметить, что каждое явление находится в тесной связи и взаимодействует с другими явлениями.

При изучении конкретных зависимостей выделяют следующие виды признаков. Признаки, выступающие в роли факторов, обуславливающих изменение других признаков, называются факторными, а признаки, испытывающие воздействие – результативными.

Рассматривая зависимости между признаками, выделяют следующие категории связей:

1) функциональные связи (взаимооднозначные), где каждому значению фактора соответствует одно или несколько определенных значений результативного признака;

2) корреляционные связи, где между изменением факторного и результативного признака нет полного соответствия, влияние отдельных факторов проявляется лишь в среднем при массовом наблюдении факторов.

Сравнивая между собой функциональные и корреляционные связи отметим, что при наличии функциональной зависимости между признаками, зная величину факторного признака можно точно определить величину результативного признака, при наличии корреляционной: с изменением величины факторного признака меняется средняя величина результативного признака.

В обосновании связей решающая роль принадлежит экономической теории. Статистика же дает количественную оценку этой зависимости.

Условия применения корреляционно-регрессивного анализа:

- требование наблюдений (в массе единиц происходят взаимопогашения действия случайных факторов);
- требования однородности единиц (однотипные предприятия, по которым изучаются технико-экономические показатели).

Проведение корреляционно-регрессивного анализа имеет две основные цели:

1) измерение параметров уравнения, выражающего связь средних значений зависимой переменной со значениями независимой переменной;
2) измерение тесноты связи признаков между собой.

Существуют следующие методы выявления наличия корреляционной связи:

1) сопоставление двух параллельных рядов – ряда значений факторного признака и соответствующих ему значений результативного признака.

Значения факторного признака располагают в возрастающем порядке и затем прослеживают направление изменения величины результативного признака.

Если с возрастанием величины факторного признака ведет к возрастанию результативного признака, то можно предположить прямую корреляционную связь;

Если с возрастанием факторного признака наблюдается убывание результативного признака, то можно предположить обратную связь между признаками.

2) построение аналитической группировки, где все наблюдения разбиваются на группы по величине факторного признака, и по каждой группе вычисляется среднее значение результативного признака.

Предполагается, что все прочие причины, если они носят случайный характер, при определении средней по группам взаимопогашаются, т.е. дают в каждой группе один и тот же результат. Недостаток: неоднозначность результатов, которые зависят как от числа выделенных групп, так и от установления границ интервалов.

3) графический метод

Для предварительного выявления наличия связи и раскрытия ее характера применяют графический метод. Используются данные о индивидуальных значения факторного признака и соответствующих ему значений результативного признака, можно построить точечный график, называемый полем корреляции. Точки корреляционного поля не лежат на одной линии, но вытянуты в определенном положении. Далее материал был сгруппирован, по каждому интервалу были определены значения средней. Соединив, получили эмпирическую линию связи, которая приближается к прямой, мы можем предположить наличие прямой корреляционной связи;

4) корреляционная таблица.

Задачи корреляционно-регрессионного анализа.

1. Выделение важных факторов, влияющих на результативный признак, на базе мер тесноты связи факторов с результативным признаком.
2. Оценка уравнения регрессии, где в качестве результативного признака выступает признак, являющийся следствием других признаков – причин.
3. Прогнозирование возможных значений результативного признака при задаваемых значениях факторных признаков. В уравнение подставляется планируемые значения факторных признаков и вычисляются ожидаемые значения результативного признака.

8.2 Измерение тесноты связи в случае парной корреляции

При рассмотрении количественных переменных основное внимание уделяется линейным связям.

Для изучения основных задач и особенностей корреляционного анализа удобно рассматривать генеральную совокупность 3-х признаков.

Рассмотрим случай трех признаков $X=(x_1, x_2, x_3)$, (p=3). Будем предполагать, что поведение многомерного вектора X описывается нормальным законом распределения, т.е. плотность совместного распределения одномерных случайных величин x_1, x_2, x_3 задается в виде:

$$p(x_1, x_2, x_3) = \frac{1}{\sqrt{(2\pi)^2}} \cdot \frac{1}{\sqrt{\sigma_{x_1}^2 \sigma_{x_2}^2 \sigma_{x_3}^2 \det(R)}} e^{-0,5z^T R^{-1} z}$$

где R — симметрическая положительно определенная матрица парных коэффициентов корреляции, а det(R) — определитель этой матрицы (обобщенная дисперсия случайной величины X), т.е.

$$R = \begin{bmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2x_1} & 1 & r_{x_2x_3} \\ r_{x_3x_1} & r_{x_3x_2} & 1 \end{bmatrix} \text{ и}$$

$$|R| = \det R = 1 + 2r_{12}r_{13}r_{23} - r_{13}^2 - r_{23}^2 - r_{12}^2 > 0$$

R^{-1} — матрица, обратная к R.

z - обозначен вектор значений нормированных случайных величин x_1, x_2, x_3.

$$z = \begin{bmatrix} \dfrac{x_1 - \mu_{x_1}}{\sigma_{x_1}} \\ \dfrac{x_2 - \mu_{x_2}}{\sigma_{x_2}} \\ \dfrac{x_3 - \mu_{x_3}}{\sigma_{x_3}} \end{bmatrix}$$

Таким образом, имеется трехмерная нормально распределенная случайная величина, которая определяется девятью параметрами: μ_{x1}, μ_{x2}, μ_{x3}, $\sigma_{x_1}^2$, $\sigma_{x_2}^2$, $\sigma_{x_3}^2$, $r_{x_1x_2}$, $r_{x_1x_3}$, $r_{x_2x_3}$.

Распределения одномерных x_1, x_2, x_3, двумерных (x_1, x_2), (x_1, x_3), (x_2, x_3), условные распределения при фиксировании одной из переменных $(x_1, x_2)|x_3$, $(x_1, x_3)|x_2$, $(x_2, x_3)|x_1$ и двух $x_1|x_2x_3$, $x_2|x_1x_3$, $x_3|x_1x_2$ являются нормальными.

Базовым инструментом измерения линейной связи между признаками является парный коэффициент корреляции.

Парный коэффициент корреляции характеризует тесноту линейной зависимости между двумя переменными на фоне действия всех остальных показателей, входящих в модель.

$$r_{jl}=\frac{1/n\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)(x_{il}-\bar{x}_l)}{\sigma_j\cdot\sigma_l},\ -1\le r_{jl}\le 1.$$

Если r_{jl} близок к ±1, то связь между переменными сильная, если r_{jl} =0 – линейная связь отсутствует.

r_{jl} > 0 – связь между переменными прямая;

r_{jl} < 0 – связь обратная.

Для многомерной корреляционной модели важную роль играют частные и множественные коэффициенты корреляции, детерминации (квадраты соответствующих коэффициентов корреляции).

<u>*Частный коэффициент корреляции*</u> между x_1 и x_2 *k-2-го порядка* при фиксированном воздействии переменной x_3 может быть определен по следующей формуле:

$$r_{x_1x_2}(x_3)=\frac{-R_{12}}{\sqrt{R_{11}R_{22}}}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

R_{ij} — алгебраическое дополнение матрицы R к элементу r_{ij},

Остальные частные коэффициенты определяются аналогично, путем замены соответствующих индексов.

$$r_{yx_j}(x_1,...,x_{j-1},x_{j+1},...,x_k)=r_{yx_j/1,...,j-1,j+1,...k}=\sqrt{1-\frac{1-R^2_{yx_1...x_k}}{R^2_{yx_1...x_{j-1},x_{j+1}...x_k}}},$$

$R^2_{yx_1...x_k}$ - множественный коэффициент детерминации всего комплекса *k* факторов с результатом;

$R^2_{yx_1...x_{j-1},x_{j+1,...}x_k}$ - тот же показатель детерминации, но без введения в модель фактора x_j.

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. В практических исследованиях предпочтение отдают показателям частной корреляции самого высокого порядка.

Частный коэффициент корреляции показывает тесноту линейной связи между двумя переменными независимо от влияния остальных случайных величин. Обладает всеми свойствами парного коэффициента корреляции.

Если частный коэффициент корреляции меньше парного, т.е. $r_{x_1x_2}(x_3)<r_{x_1x_2}$, то взаимодействие между x_1 и x_2 обусловлено частично (или полностью, если $r_{x_1x_2}(x_3)=0$) воздействием фиксируемых прочих переменных, т.е. - x_3. Если частный коэффициент корреляции $r_{x_1x_2}(x_3)>r_{x_1x_2}$, то фиксируемые прочие переменные ослабляют линейную связь.

Корреляционная матрица не в полной мере отражает связи признака с множеством переменных.

8.3 Измерение тесноты связи в случае множественной корреляции

<u>*Множественный коэффициент корреляции*</u> является мерой связи между одной переменной и другими (остальными), входящими в модель, и может быть рассчитан по формуле:

$$r_{x_3} = r_{x_3}(x_1 x_2) = \sqrt{1 - \frac{\det R}{R_{33}}} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{12}r_{31}r_{32}}{1 - r_{12}^2}};$$

$$r_y = r_y(x_1, x_2, ..., x_k) = \sqrt{1 - \frac{\det R}{R_{yy}}}.$$

Если $r_{X_3}=1$, то связь между x_3 и двумерной переменной (x_1, x_2) является функциональной, линейной. Если $r_{x_3} = 0$, то; линейной связи нет.

Из формулы r_{x_3} следует, что $r_3 >| r_{31}|$, $r_3 >| r_{32}|$, $r_3 > |r_{31|2}|$, $r_3 >| r_{32|1}|$.

Отсюда можно заметить, что коэффициент множественной корреляции может только увеличиваться, если в модель включать дополнительные признаки – случайные величины, и не увеличиться, если из имеющихся признаков производить исключение.

Если $r_3 =0$, то $r_{31}= r_{32}= r_{31|2}= r_{32|1}=0$.

Наибольшему множественному коэффициенту детерминации соответствуют большие частные коэффициенты детерминации (например, r_1^2 соответствуют $r_{12|3}^2$ и $r_{13|2}^2$)

Множественный коэффициент детерминации (квадрат соответствующего множественного коэффициента корреляции) показывает долю дисперсии, например, случайной величины x_3, обусловленную изменением случайных величин x_1 и x_2.

8.4 Проверка статистической значимости коэффициентов корреляции и их доверительные интервалы.

Проверка ***статистической значимости парного и частного коэффициентов корреляции.***

Назовем параметр связи в генеральной совокупности значимо отличающимся от нуля (значимым), если гипотеза о равенстве нулю этого параметра отвергается с заданным уровнем значимости α. Если же эта гипотеза принимается, генеральный параметр связи называется незначимым. В корреляционной модели соответствующая связь между величинами считается недоказанной или отсутствующей.

H_0: ρ=0, H_1: ρ≠0

<u>*t – критерий Стьюдента*</u>

$$t_{расч} = \frac{r}{\sqrt{1-r^2}}\sqrt{n-l-2} \quad ,$$

где *r* – частный или парный коэффициент корреляции;

l – порядок частного коэффициента корреляции;

Если $|t_{расч}| > t_{табл}(\alpha,\ v=n\text{-}l\text{-}2)$, то гипотеза H_0 с вероятностью ошибки 5% отвергается, проверяемый коэффициент корреляции считается значимым с вероятностью 95%.

Если же $|t_{расч}| < t_{табл}(\alpha,\ v=n\text{-}l\text{-}2)$, то гипотеза H_0 не отвергается.

<u>С помощью таблиц Фишера-Иейтса</u>.

По таблице Фишера–Иейтса определяется $r_{кр}(\alpha,\ v=n\text{-}l\text{-}2)$.

Если $|r_{расч}| > r_{кр}(\alpha,\ v=n\text{-}l\text{-}2)$, то гипотеза H_0 с вероятностью ошибки 5% отвергается, проверяемый коэффициент корреляции считается значимым с вероятностью 95%.

Если же $|r_{расч}| < r_{кр}(\alpha,\ v=n\text{-}l\text{-}2)$, то гипотеза H_0 не отвергается.

Проверка *статистической значимости множественного коэффициента корреляции.*

Нулевая гипотеза: отсутствует линейная связь между переменной x: и остальными переменными, образующими многомерный признак, H_0: $r_x=0$, H_1: $r_x \neq 0$

Рассчитываем статистику

$$F = \frac{r_x^2/2}{(1-r_x^2)/(n-3)}$$

Затем с заданным уровнем значимости α и числами степеней свободы $v_1=2$ и $v_2=n\text{-}3$ находят $F_{табл}$.

Если $F_{расч} > F_{табл}$ (2, n - 3), то гипотеза H_0 отвергается, т.е. r_x значимо отличается от нуля, следовательно, линейная связь есть и она статистически значима на α100% уровне значимости. Иначе связь между случайной величиной x и остальными случайными величинами отсутствует.

Конечно, проверку значимости коэффициентов связи начинать с частных коэффициентов корреляции не обязательно. Можно в некоторых случаях сократить такую проверку, например, если r_1 незначим, то коэффициенты $r_{12|3}$ и $r_{13|2}$ становятся незначимыми. Далее, если $r_{12|3}$ незначим, то $r_1=|r_{13}|$ (множественный коэффициент корреляции незначимо отличается от абсолютной величины парного коэффициента корреляции).

Для значимых параметров связи имеет смысл найти интервальные оценки.

Интервальная оценка для частного коэффициента корреляции. Для получения интервальной оценки коэффициента корреляции ρ используется *Z* статистика Фишера:

$$Z_r = Z(r) = \frac{1}{2}\ln\frac{1+r}{1-r} \text{ и } \Delta Z = t_\gamma \frac{1}{\sqrt{n-l-3}}$$

где t_γ является решением уравнения $\Phi(t_\gamma) = \gamma$ и находится по таблице интегральной функции Лапласа.

Значение Z_r определяется по таблице *Z* преобразований по найденному значению *r*. При этом $Z(-r) = -Z(r)$. Тогда ($Z_r \pm \Delta Z$) — границы интервальной оценки для M(Z). Доверительный интервал для коэффициента корреляции ρ получают по таблице обратного преобразования Фишера: $r_{\min} \le \rho \le r_{\max}$.

Интервальная оценка для множественного коэффициента корреляции находится по графикам и таблицам с помощью Z-преобразования Фишера, с дисперсией, приблизительно равной 1/n для достаточно больших значений n. Имеются графики и таблицы (Эзекиела и Фокса, К.Крамера) для получения интервальных оценок $\rho_м$ по значениям $r_м$.

8.5 Корреляция для нелинейной регрессии

Уравнение нелинейной регрессии, так же как и в линейной зависимости, дополняется показателем корреляции, а именно индексом корреляции (R):

$$R = \sqrt{1 - \frac{\sum (y - \tilde{y}_x)^2}{\sum (y - \bar{y})^2}}$$

Величина данного показателя находится в границах: $0 \le R \le 1$, чем ближе к единице, тем теснее связь рассматриваемых признаков, тем более надежно найденное уравнение регрессии.

Парабола второй степени, как и полином более высокого порядка, при линеаризации принимает вид уравнения множественной регрессии. Если же нелинейное относительно объясняемой переменной уравнение регрессии при линеаризации принимает форму линейного уравнения парной регрессии, то для оценки тесноты связи может быть использован линейный коэффициент корреляции, величина которого в этом случае совпадет с индексом корреляции $R_{yx} = r_{yz}$, где z - преобразованная величина признака-фактора, например $z = \frac{1}{x}$ или $z = \ln x$.

Иначе обстоит дело, когда преобразования уравнения в линейную форму связаны с зависимой переменной. В этом случае линейный коэффициент корреляции по преобразованным значениям признаков дает лишь приближенную оценку тесноты связи и численно не совпадает с индексом корреляции. Так, для степенной функции $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$ после перехода к логарифмически линейному уравнению $\ln y = \ln a + b \ln x + \ln \varepsilon$ может быть найден линейный коэффициент корреляции не для фактических значений переменных *x* и *y,* а для их логарифмов, т. е. $r_{\ln y \ln x}$. Соответственно квадрат его значения будет характеризовать отношение факторной суммы квадратов отклонений к общей, но не для *y,* а для его логарифмов:

$$r^2_{\ln y \ln x} = \frac{\sum (\ln \tilde{y} - \overline{\ln y})^2}{\sum (\ln y - \overline{\ln y})^2} = 1 - \frac{\sum (\ln y - \ln \tilde{y})^2}{\sum (\ln y - \overline{\ln y})^2}$$

Поскольку в расчете индекса корреляции используется соотношение факторной и общей суммы квадратов отклонений, то R^2 имеет тот же смысл, что и коэффициент детерминации. В специальных исследованиях величину R^2 для нелинейных связей называют индексом детерминации.

Оценка существенности индекса корреляции проводится, так же как и оценка надежности коэффициента корреляции.

Индекс детерминации используется для проверки существенности в целом уравнения нелинейной регрессии по F-критерию Фишера:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m}{m-1}$$

где R^2 - индекс детерминации;

n - число наблюдений;

m — число параметров.

Величина (m-1) характеризует число степеней свободы для факторной суммы квадратов, а (n - m) — число степеней свободы для остаточной суммы квадратов.

8.6 Ранговая корреляция

При *изучении неколичественных признаков или количественных* с непрерывными и неизвестными законами распределения классический подход корреляционного анализа оказывается не эффективен, и в этих случаях применяют методы непараметрической статистики и, в частности, метод ранговой корреляции.

Ранговая корреляция предназначена для изучения статистической связи между различными упорядочиваниями (ранжировками) объектов по степени проявления в них того или иного свойства.

Пусть $X^{(j)} = (x_1^{(j)}, x_2^{(j)}, \ldots, x_n^{(j)})$ некоторая ранжировка из n объектов по j-му свойству. Компонента $x_i^{(j)}$, $i = 1, \ldots ,n$ этой ранжировки определяет порядковое место (ранг), которое присвоено i-му объекту в общем ряду n анализируемых объектов, упорядоченных по убыванию; j-го свойства. Другая ранжировка $X^{(k)} = (x_1^{(k)}, x_2^{(k)}, \ldots, x_n^{(k)})$ может интерпретироваться либо как упорядочивание объектов по другому k-му свойству, либо как ранжировка по тому же свойству, но полученная, например, другим экспертом.

***Ранговый коэффициент корреляции Спирмена*.**

Ранговый коэффициент корреляции Спирмена измеряет степень согласованности двух различных ранжировок $X^{(j)}$ и $X^{(k)}$ одного и того же множества из n объектов и рассчитывается по формуле

$$r_c = 1 - \frac{6}{n^3 - n}\sum_{i=1}^{n}(x_i^{(j)} - x_i^{(k)})^2$$

Причем $-1 < r_c < 1$. Соответственно -1 означает, что ранжировки противоположны, 1 - они совпадают, r =0 - между ранжировками связь отсутствует.

Связанные ранги возникают в случае дележки мест в ранжировке. Объектам, которые делят места, приписывается ранг, равный среднему арифметическому соответствующих мест.

Например: объекты делят места с 3 по 5, тогда (3+4+5)/3 = 4 и объектам на 3, 4 и 5 местах приписывается ранг 4.

Для связанных рангов формула коэффициента корреляции Спирмена усложняется:

$$r_c = \frac{\frac{1}{6}(n^3 - n) - \sum_{i=1}^{n}(x_i^{(j)} - x_i^{(k)})^2 - T^{(j)} - T^{(k)}}{\left[\frac{1}{6}(n^3 - n) - 2T^{(j)}\right]^{\frac{1}{2}}\left[\frac{1}{6}(n^3 - n) - 2T^{(k)}\right]^{\frac{1}{2}}}$$

где $T^{(l)} = \frac{1}{12}\sum_{t=1}^{m}(n_t^{(l)})^3 - n_t^{(l)}$ здесь m — число групп, для которых имеются связанные ранги, n_t — число рангов, входящих в t-ю группу.

Проверка гипотезы H_0: $r_c = 0$.

Если $n > 10$, то статистика

$$\gamma_n = \frac{r_c\sqrt{n-2}}{\sqrt{1-r_c^2}}$$

распределена по закону Стьюдента с (n-2) степенями свободы.

При $\gamma_n > t_{\alpha/2}(n-2)$ гипотеза H_0 отклоняется, иначе - не отклоняется.

Если n < 10, то рассчитывают

$$r_c^{\max} = \frac{2S_c}{(n^3 - n)/3},$$

где S_c извлекается из таблицы для уровня значимости $\alpha/2$ и если $|r_c| > r_c^{\max}$, то H_0 отвергается.

Ранговый коэффициент корреляции Кендалла.

Расчетная формула имеет вид:

$$r_k = \frac{S}{\frac{1}{2}n(n-1)} = \frac{P-Q}{P+Q}, \quad S = P - Q$$

Ранжируем все элементы по признаку $x^{(1)}$, по ряду другого признака $x^{(2)}$ подсчитываем для каждого ранга число последующих рангов, превышающих данный (их сумму обозначим P) и число последующих рангов ниже данного (их сумму обозначим Q)

Значения коэффициента $-1 < r_k < 1$.

На практике при $n \geq 10$, $r_c = \frac{3}{2}r_k$

Для связанных рангов формула коэффициента корреляции Кендалла имеет вид:

$$r_k^* = \frac{r_k - \frac{2(u^{(1)} + u^{(2)})}{n(n-1)}}{\sqrt{(1 - \frac{2u^{(1)}}{n(n-1)})(1 - \frac{2u^{(2)}}{n(n-1)})}}, \quad u^{(l)} = \frac{1}{2}\sum_{t=1}^{m^{(l)}} n_t^{(l)}(n_t^{(l)} - 1), \ l=1,2.$$

Коэффициент конкордации (согласованности) Кендалла. Измеряет степень тесноты, статистической связи между m различными ранжировками:

$$W = \frac{12}{m^2(n^3 - n)} \sum_{i=1}^{n} \left(\sum_{l=1}^{m} x_i^{(l)} - \frac{m(n+1)}{2}\right)^2 .$$

Причем 0 < W < 1, если W= 1 — ранжировки совпадают, W= 0 — связь между ранжировками отсутствует.

Для связанных рангов

$$W = \frac{\sum_{i=1}^{n}\left(\sum_{l=1}^{m} x_i^{(l)} - \frac{m(n+1)}{2}\right)}{\frac{1}{12} m^2(n^3 - n) - m\sum_{l=1}^{m} T^{(l)}}$$

Проверка гипотезы H_0: W=0,

- при n>7 рассчитывают статистику γ_n = m(n-1)W и при $\gamma_n > \chi^2_\alpha(n-1)$ гипотеза отклоняется, иначе не отклоняется.

Вопросы для самопроверки:

1. *Какие виды признаков выделяют при изучении конкретных зависимостей?*
2. *Какие бывают категории связей?*
3. *Какие существуют условия применения корреляционно – регрессионного анализа?*
4. *Какие выделяют цели корреляционно – регрессионного анализа?*
5. *Какие существуют методы выявления наличия корреляционно – регрессионного анализа?*
6. *Какие задачи преследует процесс проведения корреляционно – регрессионного анализа?*
7. *Что показывает парный коэффициент корреляции?*
8. *Что показывает частный коэффициент корреляции?*
9. *Что показывает множественный коэффициент корреляции?*
10. *Что показывает коэффициент детерминации?*
11. *В чем суть проверки статистической значимости парного и частного коэффициентов корреляции?*
12. *В чем суть проверки статистической значимости множественного коэффициентов корреляции?*
13. *Каким образом осуществляется интервальная оценка для частного коэффициента корреляции?*
14. *Что показывает индекс корреляции?*
15. *Что показывает индекс детерминации ?*
16. *Для чего и как рассчитывается ранговый коэффициент корреляции Спирмена?*
17. *Для чего и как рассчитывается ранговый коэффициент корреляции Кенлалла?*
18. *Для чего и как рассчитывается коэффициент конкордации Кендалла?*

9 Методы прогнозирования взаимосвязи

9.1 Линейна и нелинейная парная регрессия: оценка параметров

Уравнение регрессии математически описывает теоретическую линию регрессии. Теоретической линией регрессии называется та линия, вокруг которой группируются точки корреляционного поля, указывающие основное направление связи.

Уравнение парной линейной регрессии описывается уравнением функции:

$$\tilde{y}_i = a + b \cdot x \tag{9.1}$$

a – свободный член уравнения, показывает усредненное влияние всех прочих факторов, не включенных в исследование, если x может принимать значение 0 (или не интерпретируется).

Интерпретировать можно лишь знак при a: если a > 0, то вариация результата меньше, чем вариация фактора.

b – коэффициент парной линейной регрессии, показывает, на сколько в среднем изменится величина результативного признака y при изменении факторного признака x на единицу собственного измерения.

b>0 - связь прямая, b<0 - связь обратная.

Построение линейной регрессии сводится к оценке ее параметров – a и b.

Классический подход к оцениванию параметров линейной регрессии основан на *методе наименьших квадратов* (МНК), разработанный Гауссом.

Метод наименьших квадратов позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений фактических значений результативного признака y от расчетных (теоретических) минимальна:

$$S = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \longrightarrow \min \quad \text{т.е} \quad \sum \varepsilon_i^2 \to \min \tag{9.2}$$

Другими словами, из множества линий линия регрессии выбирается так, чтобы сумма квадратов расстояний по вертикали между точками и этой линией была бы минимальной.

Для линейной парной зависимости:

$$S = \sum_{i=1}^{n} \left[y_i - (a + b \cdot x) \right]^2 \longrightarrow \min . \tag{9.3}$$

Для нахождения параметров a и b решают систему нормальных уравнений (методом последовательно исключения или методом определителей):

$$\begin{cases} \sum y_i = n \cdot a + b \cdot \sum x_i \\ \sum y_i \cdot x_i = a \cdot \sum x_i + b \cdot \sum x_i^2 \end{cases}$$

Откуда: $b = \dfrac{\sum xy - n\bar{x} \cdot \bar{y}}{\sum x^2 - n(\bar{x})^2}$ или

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}, \quad b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

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

Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций.

Различают *два класса нелинейных регрессий:*

• регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам;

• регрессии, нелинейные по оцениваемым параметрам.

Примером нелинейной регрессии по включаемым в нее объясняющим переменным могут служить следующие функции:

• *полиномы разных степеней*

$\tilde{y}_x = a + bx + cx^2 + \varepsilon$

• *равносторонняя гипербола*

$\tilde{y}_x = a + \dfrac{b}{x} + \varepsilon$

К нелинейным регрессиям по оцениваемым параметрам относятся функции:

• *степенная* - $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$

• *показательная* - $\tilde{y}_x = a \cdot b^x \cdot \varepsilon$

• *экспоненциальная* - $\tilde{y}_x = e^{a+bx} \cdot \varepsilon$

Нелинейная регрессия по включенным переменным не таит каких-либо сложностей в оценке ее параметров. Она определяется, как и в линейной регрессии, методом наименьших квадратов (МНК), ибо эти функции линейны по параметрам. Так, в параболе второй степени $\tilde{y}_x = a + bx + cx^2 + \varepsilon$ заменяя переменные $x = x_1$, $x^2 = x_2$, получим двухфакторное уравнение линейной регрессии: $\tilde{y}_x = a + bx_1 + cx_2 + \varepsilon$.

Применение МНК для оценки параметров параболы второй степени приводит к следующей системе нормальных уравнений:

$$\begin{cases} \sum y = na + b\sum x + c\sum x^2 \\ \sum yx = a\sum x + b\sum x^2 + c\sum x^3 \\ \sum yx^2 = a\sum x^2 + b\sum x^3 + c\sum x^4 \end{cases}$$

Решение ее возможно методом определителей:

$$a = \frac{\Delta a}{\Delta},\ b = \frac{\Delta b}{\Delta},\ \text{c} = \frac{\Delta c}{\Delta}$$

где Δ - определитель системы;

Δa, Δb, Δc – частные определители для каждого из параметров.

При $b > 0$ и $c < 0$ кривая симметрична относительно высшей точки, т.е. точки перелома кривой, изменяющей направление связи, а именно рост на падение.

При b < 0 и $c > 0$ парабола второго порядка симметрична относительно своей низшей точки, что позволяет определять минимум функции в точке, меняющей направление связи, т. е. снижение на рост.

Такого рода функцию можно наблюдать в экономике труда при изучении зависимости заработной платы работников физического труда от возраста – с увеличением возраста повышается заработная плата ввиду одновременного увеличения опыта и повышения квалификации работника. Однако с определенного возраста ввиду старения организма и снижения производительности труда дальнейшее повышение возраста может приводить к снижению заработной платы работника.

Парабола второй степени целесообразна к применению, если для определенного интервала значений фактора меняется характер связи рассматриваемых признаков: прямая связь меняется на обратную или обратная на прямую. В этом случае определяется значение фактора, при котором достигается максимальное (или минимальное) значение результативного признака.

Если же исходные данные не обнаруживают изменения направленности связи, то параметры параболы второго порядка становятся трудно интерпретируемыми, а форма связи часто заменяется другими нелинейными моделями.

Ввиду симметричности кривой парабола второй степени далеко не всегда пригодна в конкретных исследованиях. Чаще исследователь имеет дело лишь с отдельными сегментами параболы, а не с полной параболической формой. Кроме того, параметры параболической связи не всегда могут быть логически истолкованы. Поэтому если график зависимости не демонстрирует четко выраженной параболы второго порядка (нет смены направленности связи признаков), то она может быть заменена другой нелинейной функцией, например степенной.

Среди класса нелинейных функций, параметры которых без особых затруднений оцениваются МНК, следует назвать хорошо известную в эконометрике равностороннюю гиперболу: $\tilde{y}_x = a + \frac{b}{x} + \varepsilon$.

Классическим ее примером является **кривая Филлипса,** характеризующая нелинейное соотношение между нормой безработицы x и процентом прироста заработной платы *y.* Английский экономист А. В. Филлипс, анализируя данные более чем за 100-летний период, установил обратную зависимость процента прироста заработной платы от уровня безработицы.

Для равносторонней гиперболы вида $\tilde{y}_x = a + \frac{b}{x} + \varepsilon$ заменив $\frac{1}{x}$ на z, получим линейное уравнение регрессии $\tilde{y}_z = a + bz + \varepsilon$, оценка параметров которого может быть дана МНК. Система нормальных уравнений составит:

$$\begin{cases} \sum y = na + b\sum \frac{1}{x} \\ \sum \frac{y}{x} = a\sum \frac{1}{x} + b\sum \frac{1}{x^2} \end{cases}$$

При b>*0* имеем обратную зависимость, которая при $x \to \infty$ характеризуется минимальным предельным значением *y,* оценкой которого служит параметр *a.*

При b < *0* имеем медленно повышающуюся функцию.

В отдельных случаях может использоваться и нелинейная модель вида $\tilde{y}_x = \frac{1}{a + bx + \varepsilon}$, так называемая обратная модель, являющаяся разновидностью гиперболы. Но если в равносторонней гиперболе $\tilde{y}_x = a + \frac{b}{x} + \varepsilon$ преобразованию подвергаются объясняющие переменные $\frac{1}{x} = z$, то для получения линейной формы зависимости в обратной модели преобразовывается *y,* а именно: 1/*y* = z и $z = a + bx + \varepsilon$. В результате обратная модель оказывается внутренне нелинейной и требование МНК выполняется не для фактических значений признака *y,* а для их обратных величин 1/y , а именно: $\sum (z - \tilde{z}_x)^2 \to \min$

Соответственно получаем систему нормальных уравнений

$$\begin{cases} \sum \frac{1}{y} = na + b\sum x \\ \sum \frac{x}{y} = a\sum x + b\sum x^2 \end{cases}.$$

Иначе обстоит дело с регрессией, нелинейной по оцениваемым параметрам. Данный класс нелинейных моделей подразделяется на два типа: нелинейные модели внутренне линейные и нелинейные модели внутренне нелинейные. Если *нелинейная модель внутренне линейна,* то она с помощью соответствующих преобразований может быть приведена к линейному виду. Если же *нелинейная модель внутренне нелинейна,* то она не может быть сведена к линейной функции. Например, в эконометрических исследованиях при изучении эластичности спроса от цен широко используется степенная функция:

$$\tilde{y}_x = a \cdot x^b \cdot \varepsilon ,$$

где y - спрос (количество);

x - цена;

ε - случайная ошибка.

Данная модель нелинейна относительно оцениваемых параметров. Однако ее можно считать внутренне линейной, ибо логарифмирование данного уравнения по основанию ε приводит его к линейному виду:

$$\ln y = \ln a + b \ln x + \ln \varepsilon .$$

Соответственно оценки параметров a и b могут быть найдены МНК. Если же модель представить в виде $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$, то она становится внутренне нелинейной, ибо ее невозможно превратить в линейный вид.

В специальных исследованиях по регрессионному анализу часто к нелинейным относят модели, только внутренне нелинейные по оцениваемым параметрам, а все другие модели, которые внешне нелинейны, но путем преобразований параметров могут быть приведены к линейному виду, относятся к классу линейных моделей.

Модели внутренне нелинейные по параметрам могут иметь место в эконометрических исследованиях. Однако гораздо большее распространение получили модели, приводимые к линейному виду. Решение такого типа моделей реализовано в стандартных пакетах прикладных программ.

Среди нелинейных функций, которые могут быть приведены к линейному виду, в эконометрических исследованиях очень широко используется степенная функция $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$. Связано это с тем, что параметр b в ней имеет четкое экономическое истолкование, т. е. он является **коэффициентом эластичности.** Это значит, что величина коэффициента b показывает, на сколько процентов изменится в среднем результат, если фактор изменится на 1 %.

Для оценки параметров степенной функции $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$ применяется МНК к линеаризованному (логарифмическому) уравнению $\ln y = \ln a + b \ln x + \ln \varepsilon$, т.е. решается система нормальных уравнений:

$$\begin{cases} \sum \ln y = n \ln a + b \sum \ln x \\ \sum \ln y \ln x = \ln a \sum \ln x + b \sum (\ln x)^2 \end{cases}$$

Параметр b определяется непосредственно из системы, а параметр a — косвенным путем после потенцирования величины *ln*a.

В моделях, нелинейных по оцениваемым параметрам, но приводимых к линейному виду, МНК применяется к преобразованным уравнениям. Если в линейной модели и моделях, нелинейных по переменным, при оценке параметров исходят из критерия $\sum(y-\tilde{y}_x)^2 \to \min$, то в моделях, нелинейных по оцениваемым параметрам, требование МНК применяется не к исходным данным результативного признака, а к их преобразованным величинам, т. е. lny, 1/y.

Возьмем, например, показательную кривую: $\tilde{y}_x = a \cdot b^x \cdot \varepsilon$ или равносильную ей экспоненту $\tilde{y}_x = e^{a+bx} \cdot \varepsilon$. Прологарифмировав, имеем: $\ln y = \ln a + x \ln b + \ln \varepsilon$

Применяя МНК, минимизируем $\sum(\ln y - \ln \tilde{y}_x)^2 \to \min$. Система нормальных уравнений составит:

$$\begin{cases} \sum \ln y = n \ln a + \ln b \sum x \\ \sum x \ln y = \ln a \sum x + b \sum x^2 \end{cases}$$

Практическое применение экспоненты возможно, если результативный признак не имеет отрицательных значений. Поэтому если исследуется, например, финансовый результат деятельности предприятий, среди которых наряду с прибыльными есть и убыточные, то данная функция не может быть использована.

9.2 Коэффициент эластичности

Коэффициент эластичности является показателем силы связи. На его основе делают выбор наиболее значимого фактора. Рассмотрим **формулу расчета коэффициента эластичности**

$$\acute{Y} = f'(x)\frac{x}{y}, \tag{9.4}$$

где f'*(x)* - первая производная, характеризующая соотношение приростов результата и фактора для соответствующей формы связи.

Во всех функциях, кроме степенной, коэффициент эластичности зависит от значений фактора *x*.

В силу того, что коэффициент эластичности для линейной функции не является величиной постоянной, а зависит от соответствующего значения *x*, то обычно рассчитывается средний показатель эластичности по формуле

$$\bar{\acute{Y}} = b\frac{\bar{x}}{\bar{y}}. \tag{9.5}$$

Средний коэффициент эластичности $\bar{\acute{Y}}$ показывает, на сколько процентов в среднем по совокупности измениться результат y от своей средней величины при изменении фактора x на 1 % от своего среднего значения.

Поскольку коэффициенты эластичности представляют экономический интерес, а виды моделей не ограничиваются только степенной функцией, приведем формулы расчета коэффициентов эластичности для наиболее распространенных типов уравнений регрессии (табл. 9.1).

Таблица 9.1 – Коэффициенты эластичности для ряда математических функций

Вид функции, $\tilde{y}_x$	Первая производная, $f'(x)$	Коэффициент эластичности, $\acute{Y} = f'(x)\dfrac{x}{y}$
Линейная $\tilde{y}_x = a + bx + \varepsilon$	b	$\acute{Y} = \dfrac{bx}{a+bx}$
Параболическая $\tilde{y}_x = a + bx + cx^2 + \varepsilon$	$b + 2cx$	$\acute{Y} = \dfrac{(b+2cx)x}{a+bx+cx^2}$
Гипербола $\tilde{y}_x = a + \dfrac{b}{x} + \varepsilon$	$\dfrac{-b}{x^2}$	$\acute{Y} = \dfrac{-b}{ax+b}$
Степенная $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$	$a \cdot b \cdot x^{b-1}$	$\acute{Y} = a \cdot b \cdot x^{b-1}\dfrac{x}{a \cdot x^b} = b$
Показательная $\tilde{y}_x = a \cdot b^x \cdot \varepsilon$	$\ln b \cdot a \cdot b^x$	$\acute{Y} = \ln x \cdot b$
Обратная $\tilde{y}_x = \dfrac{1}{a+bx+\varepsilon}$	$\dfrac{-b}{(a+bx)^2}$	$\acute{Y} = \dfrac{-bx}{a+bx}$

Несмотря на широкое использование в эконометрике коэффициентов эластичности, возможны случаи, когда их расчет экономического смысла не имеет. Это происходит тогда, когда для рассматриваемых признаков бессмысленно определение изменения значений в процентах. Например, вряд ли кто будет определять, на сколько процентов может измениться заработная плата с ростом стажа работы на 1 %. Или, например, на сколько процентов изменится урожайность пшеницы, если качество почвы, измеряемое в баллах, изменится на 1 %. В такой ситуации степенная функция, даже если она оказывается наилучшей по формальным соображениям (исходя из наименьшего значения остаточной вариации), не может быть экономически интерпретирована.

9.3 Прогнозирование по уравнению регрессии

Уравнения регрессии используются для:

1. *Для оценки хозяйственной деятельности*

Сравнение фактических уровней результативного признака с расчетными позволяет установить эффективность использования фактора *x* (средств) в конкретном объекте (хозяйстве) по сравнению со средней эффективностью использования фактора *x* по совокупности объектов.

Линия регрессии отражает изменение среднего значения результативного признака. Объекты, чьи фактические значения результата превышают теоретические, наиболее эффективно используют ресурсы. А объекты, чьи фактические значения результата меньше теоретических, имеют неиспользованные ресурсы повышения результата.

Например, при изучении зависимости между надоем на одну корову (ц) и затратами на одну корову (тыс. руб.)

2. *Для прогнозирования возможных значений результативного признака*

- <u>авторегрессионное прогнозирование</u> по тренду и колеблемости;
- <u>факторное прогнозирование</u>, основанное на изучении и количественном измерении взаимосвязи между признаками.

Основным условием прогнозирования на основании регрессионного уравнения является стабильность или, по крайней мере, малая изменчивость других факторов и условий изучаемого процесса, не связанных с ними. Если резко изменится «внешняя среда» протекающего процесса, прежнее уравнение регрессии потеряет свое значение.

Прогнозирование по уравнению регрессии проводится в два этапа:

1. Вычисляется «точечный прогноз».
2. Определяется доверительный интервал с достаточно большой вероятностью (интервальный прогноз).

Точечный прогноз – это значение результативного показателя, получаемого при подстановке в уравнение регрессии ожидаемой величины факторного признака. Однако, нельзя подставлять значения факторного признака, значительно отличающихся от входящих в базисную информацию, по которой вычислено уравнение регрессии. При качественно иных уровнях фактора, если они даже возможны в принципе, были бы другими параметры уравнения.

Вероятность точной реализации такого прогноза крайне мала. Необходимо сопроводить прогноз *доверительным интервалом*.

Статистический прогноз с учетом доверительного интервала:

«Точечный прогноз» ± α,

где α – доверительный интервал,

$\alpha = m \cdot t$,

m – стандартная ошибка прогноза

$$m = S(y)\sqrt{1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}, \qquad (9.6)$$

m – стандартная ошибка положения линии регрессии в генеральной совокупности при $x=x_k$;

n – объем выборки;

x_k - ожидаемое значение фактора;

S(y) – среднее квадратическое отклонение результативного признака от линии регрессии в генеральной совокупности с учетом степеней свободы вариации.

$$S(y) = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}}\,. \qquad (9.7)$$

t критерий Стьюдента определяется по таблице, зависит от вероятности и числа степеней свободы $df = n - p$.

По уравнению множественной регрессии можно прогнозировать несколькими способами:

1) при подстановке в уравнение регрессии желаемых показателей;
2) с учетом прогнозных значений факторного признака.

Вопросы для самопроверки:

1. *Что такое парное уравнение регрессии? Для каких целей оно используется?*
2. *В чем заключается суть метода наименьших квадратов?*
3. *Какие выделяют классы нелинейных моделей? Какие к ним относятся функции?*
4. *Опишите алгоритм применения метода наименьших квадратов для параболы второго порядка?*
5. *Опишите алгоритм применения метода наименьших квадратов для равносторонней гиперболы?*
6. *Опишите алгоритм применения метода наименьших квадратов для обратной модели?*
7. *Опишите алгоритм применения метода наименьших квадратов для степенной функции?*
8. *Как осуществляется оценка параметров регрессии нелинейной по оцениваемым параметрам?*
9. *Что такое коэффициент эластичности? Что он показывает? Для чего рассчитывается?*
10. *Для каких целей используется уравнение регрессии?*
11. *Как осуществляется прогнозирование по уравнению регрессии?*

10 Эвристические методы прогнозирования социально-экономических явлений

10.1Общая характеристика метода экспертных оценок

В настоящее время в экономике, в частности при обработке результатов эконометрического исследования, широкое применение нашел метод экспертных оценок, применяемых в тех областях экономических знаний, где невозможно провести оценку характеристик объекта какими-либо физическими приборами.

Под методом экспертных оценок понимают комплекс логических и математических процедур, направленных на получение от специалистов информации, ее анализ и обобщение с целью подготовки и выбора рациональных решений. Экспертные коллективные оценки широко использовались в государственном масштабе для решения сложных проблем управления народным хозяйством уже в первые годы Советской власти. В 1918 году при Высшем совете народного хозяйства был создан Совет экспертов, задачей которого являлось решение наиболее сложных проблем реорганизации народного хозяйства страны. При составлении пятилетних планов развития народного хозяйства страны систематически использовались экспертные оценки широкого круга специалистов. В настоящее время в нашей стране и за рубежом метод экспертных оценок широко применяется для решения важных проблем различного характера. В различных отраслях, объединениях и на предприятиях действуют постоянные или временные экспертные комиссии, формирующие решения по различным сложным неформализуемым проблемам.

Сущность этого метода заключается в проведении высококвалифицированными специалистами интуитивно-логического анализа проблемы с качественной или количественной оценкой суждений и формальной обработкой результатов. Комплексное использование интуиции, логического мышления и соответствующего математического аппарата позволяет получить· эффективное решение поставленной проблемы. Метод экспертных оценок отличается от традиционной экспертизы наличием научной организации всех этапов и применением математического аппарата как при организации экспертизы, так и при обработке и анализе полученной информации.

Проведение экспертизы следует рассматривать как начальный цикл исследований, которому предшествует детальный анализ существа проблемы. Систематизация полученных при таком анализе сведений позволяет целенаправленно и планомерно организовывать экспертизу, ориентированную на получение новой информации в необходимой форме. Экспертизу чаще всего нужно рассматривать как последний, наиболее формальный шаг эконометрического анализа. Полученная и обработанная экспертная информация используется в дальнейшем в рамках выбранной процедуры выработки решений.

Целесообразно в общем виде выделить два класса проблем, решаемых экспертным путем.

К первому классу относятся проблемы, для решения которых имеется достаточная информационная база. Поэтому основная трудность их решения заключается в эффективном использовании этой базы, т.е. в правильном подборе экспертов, рациональной организации процедуры их опроса и применении оптимальных математических методов обработки результатов. При этом методы опроса и обработки основываются на том, что эксперт обобщает значительный объем переработанной им информации, а групповое мнение, рассчитанное как математическое ожидание мнений отдельных экспертов, близко к истинному значению. Эти допущения позволяют применять для обработки экспертных оценок специальные математические методы.

Ко второму классу относятся проблемы, для решения которых информационная база недостаточна. При анализе таких проблем эксперт уже не может количественно представлять итоги оценки, и поэтому здесь возможен в основном лишь качественный анализ результатов экспертизы.

Простейшим способом получения экспертной информации является учет мнения одного специалиста - ***индивидуальная экспертиза***. Индивидуальная экспертиза используется для решения уникальных проблем, относящихся ко второму классу. Однако при решении проблем первого класса целесообразно привлекать по возможности не одного, а нескольких специалистов, Т.е. организовывать групповую экспертизу.

В процессе проведения групповой экспертизы можно выделить следующие этапы:

1. формулирование цели и задач исследования;
2. постановка задачи;
3. разработка методов получения экспертной информации и способов ее обработки;
4. формирование экспертной группы;
5. проведение экспертизы;
6. сбор, обработка и анализ полученной информации;
7. интерпретация полученных результатов и формирование вариантов рекомендаций.

При решении важных практических задач организация экспертизы начинается с подготовки и издания руководящего документа. в котором формулируется цель работы, указываются сроки ее выполнения, задачи и состав группы управления, ее права и обязанности. устанавливается материальное и финансовое обеспечение работ. Для, подготовки такого документа и руководства всей работой назначается руководитель экспертизы. Группа управления организует проведение экспертизы, причем оформление результатов ее работы, осуществляется в виде отчета, который после обсуждения и одобрения представляется на утверждение в соответствующие инстанции.

На начальном этапе при постановке задачи исследования обосновывается целесообразность получения экспертной информации и намечаются пути использования ожидаемых результатов. Вопросы выбора методов получения и обработки экспертной информации решаются в соответствии со спецификой поставленной задачи и с учетом выделенных ресурсов и времени.

При составлении экспертных групп пользуются соображениями здравого смысла, учитывая цели экспертизы, ограниченность ресурсов и требования, обусловленные выбранными методами получения и обработки информации от экспертов. Вначале намечаются кандидатуры экспертов, которые по своим профессиональным качествам могут быть привлечены к работе в роли экспертов. При поиске таких кандидатур обычно используются общепринятые документальные показатели, отражающие профессиональный уровень специалиста (должность, ученая степень и звание, количество опубликованных научных работ и др.), а также его прежнее участие в экспертизах. Документальная характеристика часто дополняется взаимной оценкой экспертов. При этом в число «потенциальных» экспертов включаются специалисты, рекомендованные несколькими ранее выявленными экспертами, ранее активно проводившими экспертизу.

Важно, чтобы кандидат в экспертную группу имел широкий кругозор и эрудицию. Компетентность эксперта есть степень его квалификации в определенной области знаний, которая определяется на основе анализа его профессиональной деятельности, широты кругозора по перспективам развития рассматриваемой проблемы.

В зависимости от профессиональной подготовки эксперта (должность, ученое звание, степень) ему присваивают определенный балл. В табл. 10.1 приведены баллы оценок профессиональных качеств эксперта, работающего в научно-исследовательской организации крупной фирмы. Такая информация позволяет произвести предварительный отбор и ранжирование кандидатов по их компетентности.

Таблица 10.1 – Оценка степени компетентности эксперта

Занимаемая должность	Специалист без степени	Кандидат наук	Доктор наук	Академик, член-корреспондент
Инженер	1	-	-	-
Мл. научный сотрудник	1	1,5	-	-
Ст. научный сотрудник	-	2,25	3	-
Зав. лабораторией, сектором, руководитель группы	2	3	4	6
Зав. отделом, зам. зав. отделом	2	3,75	5	7,5

Ведущий руководитель комплекса, управления	3	4,5	6	9
Ген. директор, зам. ген. директора	4	6	8	12

На следующем этапе решается вопрос о численном составе экспертной группы. Количество привлекаемых к работе экспертов, как правило, обусловлено их квалификацией и не должно превышать предела, диктуемого ограничениями финансового, временного и организационного характера. Кроме того, в составе группы по возможности должно быть обеспечено равное представительство специалистов различных направлений, существующих в исследуемой области, и организаций, имеющих профессиональное отношение к рассматриваемой задаче.

Далее осуществляется формирование группы путем выделения тех экспертов, которые являются наиболее компетентными с точки зрения конкретной решаемой задачи. Для предварительной оценки компетентности экспертов рассчитывают коэффициент их компетентности, представляющий собой сумму баллов, приписываемых эксперту в зависимости от его документальных показателей, и данных самооценки (касающихся производственного опыта, знания экономической литературы по решаемому вопросу и др.). Эти баллы берутся из специальных таблиц. При другом подходе коэффициенты компетентности вычисляются на основе матрицы, составленной по результатам взаимной оценки экспертов.

В экспертную группу включаются те эксперты, для которых коэффициент компетентности не ниже некоторого порогового значения. Это значение устанавливается таким, чтобы обеспечить набор из числа тех потенциальных экспертов, которые действительно смогут участвовать в ее работе. При проведении экспертиз целесообразно использовать такую обобщенную характеристику эксперта, как достоверность его суждений. Количественно достоверность эксперта оценивают соотношением N_u/N, глд N_u - число случаев, когда эксперт дал решение, приемлемость которого подтвердилась практикой; N - общее число случаев участия эксперта в решении задачи.

К сожалению, даже самый квалифицированный эксперт может проявить себя некомпетентным в конкретной экспертизе как в силу стохастического характера, так и из-за отсутствия стимулов в данной группы. Поэтому возникает проблема оценки компетентности экспертов на основе анализа информации, полученной от него и других членов экспертной группы, участвующих в данной экспертизе.

Сбор и обработка экспертной информации осуществляются согласно разработанным (выбранным) для этих целей методам. Практически для сбора информации составляются соответствующие документы (например, специальные анкеты), а при ее обработке и анализе в случае необходимости может использоваться ПЭВМ. Содержательный анализ и интерпретация обработанных результатов осуществляются с учетом специфики решаемой проблемы в соответствии с задачами, которые ставились перед экспертизой.

Типичные ошибки, приводящие к некачественной экспертизе и негативному пониманию экспертного оценивания в эконометрике:

1. Преувеличение возможностей экспертного оценивания.
2. Нечеткая постановка задачи перед экспертами.
3. Излишнее стремление оставаться в рамках одной экспертной процедуры и увлечение количественными оценками и свертками, а также некорректная обработка и интерпретация результатов экспертизы.
4. Недостаточная информированность экспертов о конкретном объекте экспертизы и использование некомпетентных экспертов.
5. Противоречивость, несогласованность и неточность экспертных оценок при коллективной экспертизе.
6. Конформизм и конъюнктурность экспертов.

10.2 Классификация методов получения экспертной информации

Квалифицированное применение метода экспертных оценок в существенной степени зависит от выбранного способа сбора и обработки ответов целенаправленно сформированной группой специалистов. Существует большое количество методов получения экспертной информации. Классификация этих методов приведена на рис.10.1

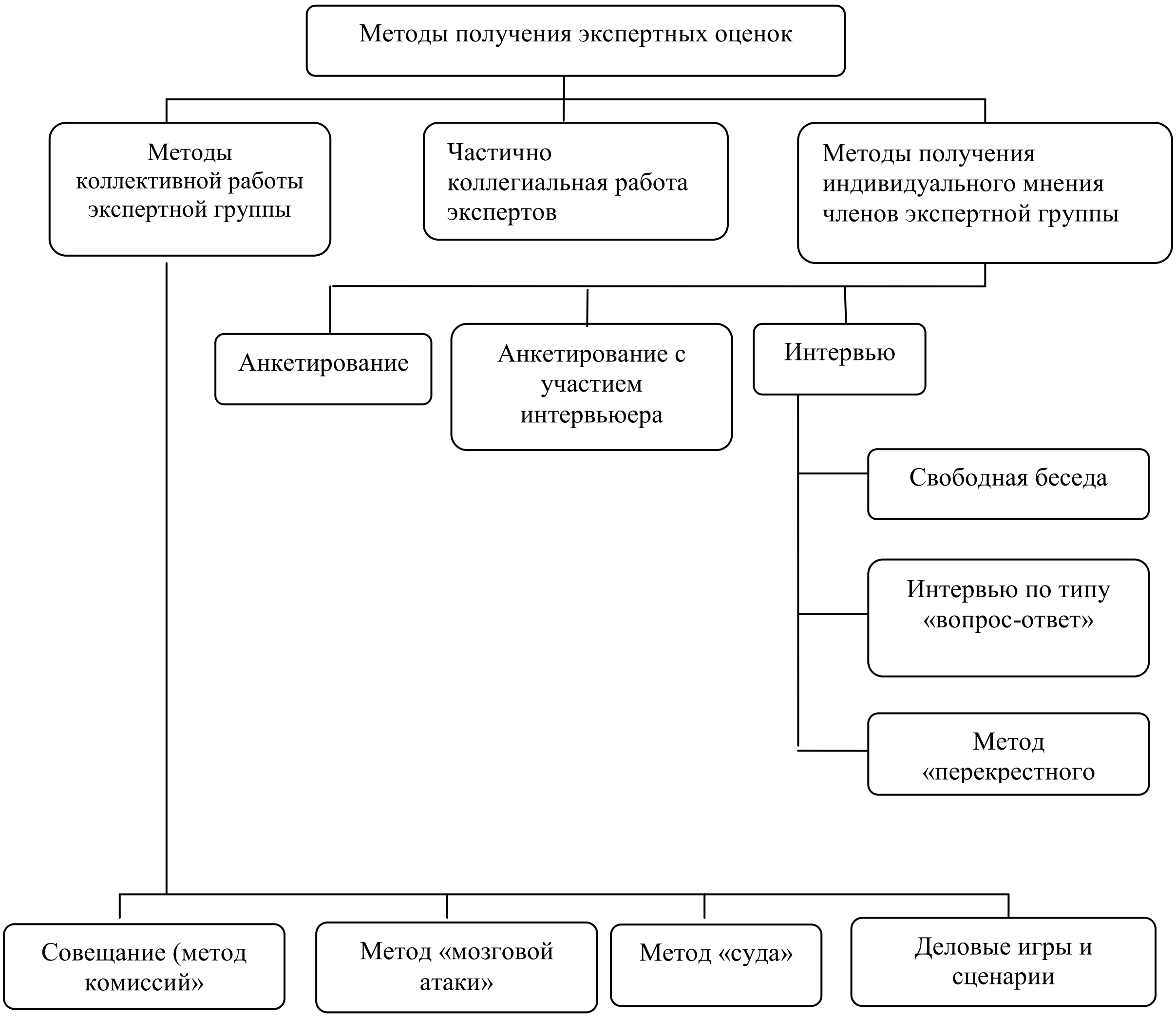

Рисунок 10.1 - Классификация методов экспертного оценивания

Методы коллективной работы экспертной группы предполагают получение обобщенного мнения в ходе совместного обсуждения и решения поставленной задачи всеми экспертами. Первые три метода - методы комиссий (совещания), «суда», и «мозговой атаки» основаны на коллективном обсуждении решаемой задачи, а деловые игры - на совместной работе коллектива специалистов, за каждым членом которого закреплены определенные обязанности в соответствии с заранее составленными правилами и программой. Деловые игры, широко используемые в экономике, предназначены в основном для выяснения поведения специалистов в исследуемой обстановке, а также для их обучения.

Метод комиссий. Самым простым и традиционным является метод комиссий. Он предполагает проведение совещания или же общей дискуссии с целью выработки единого коллективного мнения. Количество участников может быть от 3 до 10 и более. Достоинства этого метода состоят в том, что каждый член группы может не только высказать свое мнение, но и аргументировано его защитить, а также критиковать мнения других экспертов. Тщательное, обсуждение уменьшает возможность ошибок. Однако на совещании победить мнение какого-либо участника в силу его авторитетного положения, блестящих ораторских способностей, высокой активности и т.д.

Метод «суда». Этот недостаток присущ и методу «суда», При использовании которого работа экспертной группы осуществляется подобно ведению судебного процесса, где роли «судей («заседателей») исполняют члены группы управления, а в качестве «прокурора» и «защитников» выступают эксперты. «Свидетелями» являются различные факты, результаты экспериментов, доводы экспертов. Использование такого метода особенно целесообразно при наличии нескольких группировок экспертов, каждая из которых придерживается своей точки зрения.

Метод «мозговой атаки» («мозгового штурма», «коллективной генерации идей») основан на предположении, что если выдвигать большое число самых разнообразных идей и предложений по решению поставленной проблемы, то среди них обязательно окажется одна или даже несколько ценных. Сущность метода состоит в том, что задачи выдвижения новых идей и их анализа полностью разделены, поэтому образуются две группы - генераторов идей и аналитиков. В состав первой включаются эрудированные, с богатым воображением люди, специалисты из смежных областей. Для создания непринужденной обстановки лучше собирать людей примерно одинакового служебного и общественного положения. Заседанием группы генераторов идей руководит ведущий, основной задачей которого является всемерное поощрение инициативы и творчества, свободы высказывания предложений и полное недопущение критики. Все выступления фиксируются и впоследствии тщательно анализируются группой аналитиков - специалистов по профилю решаемой проблемы.

Методы опроса. Этот метод целесообразно использовать в критических ситуациях дефицита творческих решений. Однако реализация всех методов коллективного получения экспертной информации обычно связана с различными трудностями организационного характера, так как собрать вместе многих ведущих специалистов далеко не всегда возможно. Поэтому в настоящее время все чаще используются методы опроса, заключающиеся в получении коллективного мнения на основе совместной обработки индивидуальных мнений членов экспертной группы, опрашиваемых независимо друг от друга. Существуют два основных вида опроса - интервью и анкетирование. Третий является промежуточным: в нем сочетаются идеи первых двух. Специфической особенностью интервью является то, что исследователь и эксперт находятся непосредственном контакте. Исследователь ведет беседу, стараясь получить ответы на заранее сформулированные вопросы.

Условно выделяют три формы организации бесед. Первые две из них, очевидно, характеризуются своими названиями. Третья осуществляется с участием нескольких интервьюеров и проводится по типу перекрестного допроса в судебном расследовании. Интервьюеры, получая необходимую информацию, проверяют непротиворечивую последовательность рассуждений эксперта.

Прямой контакт интервьюера и эксперта создает определенный психологический климат, позволяющий получать сведения, малодоступные анкетному опросу. Однако при этом на результаты большое влияние могут оказать личность интервьюера, а также способность эксперта к контакту, быстрота его мышления и другие психологические факторы. Кроме того, при беседе эксперт обычно не в состоянии рассматривать и анализировать проблему в целом.

Метод анкетного опроса заключается в том, что эксперту предлагается для заполнения специальная анкета, содержащая заранее составленный набор вопросов. В зависимости от степени свободы, предоставляемой эксперту при выборе варианта ответа, вопросы бывают открытого и закрытого типа. При вопросе открытого типа эксперт свободен в выборе формулировки ответа. Вопросы закрытого типа содержат в своей формулировке варианты возможных ответов, один из которых и должен выбрать эксперт. С точки зрения возможности получения сопоставимых результатов анкетирования, наиболее предпочтительны вопросы закрытого типа.

В приведенной классификации выделена группа методов частично коллегиальной работы экспертов. Сущность их состоит в том, информация собирается методом опроса коллективных экспертов. Такие методы позволяют сочетать преимущества коллективной работы и опроса.

Метод Дельфы. В рассмотренную классификацию не вошли комплексные процедуры экспертизы, осуществляемые в несколько туров. Одним из наиболее известных методов анкетирования подобного рода является метод Дельфы, разработанный в начале 60-х годов прошлого столетия американской компанией РЭНД Корпорейшен, который предусматривает полный отказ от личных контактов экспертов и обеспечение их информацией, включая и обмен информацией между ними после каждого тура опроса, при сохранении анонимности оценок, аргументации и критики. Метод Дельфи (иногда дельфийский метод) появился в 1950-1960 годы для анализа планов атомной войны США (разработан корпорацией RAND, авторами считаются Olaf Helmer, Norman Dalkey, и Nicholas Rescher . Имя заимствовано от Дельфийского Оракула.

Суть этого метода в том, чтобы с помощью серии последовательных действий – опросов, интервью, мозговых штурмов – добиться максимального консенсуса при определении правильного решения. Анализ с помощью дельфийского метода проводится в несколько этапов, результаты обрабатываются статистическими методами.

Базовым принципом метода является то, что некоторое количество независимых экспертов (часто несвязанных и не знающих друг о друге) лучше оценивает и предсказывает результат, чем структурированная группа (коллектив) личностей. Позволяет избежать открытых столкновений между носителями противоположенных позиций т.к. исключает непосредственный контакт экспертов между собой и, следовательно, групповое влияние, возникающее при совместной работе и состоящее в приспособлении к мнению большинства. Даёт возможность проводить опрос экстерриториально, не собирая экспертов в одном месте (например, посредством электронной почты)

Предварительный: подбор группы экспертов - чем больше, тем дольше - до 20.

Основной:

постановка проблемы - экспертам рассылается вопрос и предлагается его разбить на подвопросы. Организационная группа отбирает наиболее часто встречающиеся. Появляется общий опросник.

Этот опросник рассылается экспертам. Их спрашивают - можно ли добавить ещё что-то; достаточно ли информации; есть ли дополнительная информация по вопросу? В итоге - 20 вариантов ответов, где есть дополнительные аспекты, запрос информации, предоставленная информация. На этой основе составляется следующий опросник.

Улучшенный опросник вновь рассылается экспертам, которым теперь надо дать свой вариант решения, а также рассмотреть наиболее крайние точки зрения, высказанные другими экспертами. Эксперты должны оценить проблему по аспектам: эффективность, обеспеченность ресурсами, в какой степени соответствует изначальной постановке задачи. Таким образом выявляются преобладающие суждения экспертов, сближаются их точки зрения. Всех экспертов знакомят с доводами тех, чьи суждения сильно выбиваются из общего русла. После этого все эксперты могут менять мнение, а процедура повторяется.

Итерации повторяются, пока не будет достигнута согласованность между экспертами, или не будет установлено отсутствие единого мнения по проблеме. Изучение причин расхождений в оценках экспертов позволяет выявить незамеченные ранее аспекты проблемы и зафиксировать внимание на вероятных последствиях развития анализируемой проблемы или ситуации. В соответствии с этим и вырабатывается окончательная оценка и практические рекомендации. Обычно проводится три этапа, но если мнения сильно разнятся - то больше.

Применяется в стратегическом планировании в технике, футурологии, бизнесе.

В России применяется мало, так как: долгое время аналитика была централизована - высок конформизм экспертов; нужны независимые аналитические структуры - в России их нет; стратегический анализ мало востребован - все авторы знают свои цели; нет традиции проведения таких исследований - нет профессионалов, все эксперты знают друг друга и их очень мало.)

10.3 Типы шкал и методы получения элементарных суждений

В группе задач анализа экспертных оценок можно выделить три основных типа задач, решаемых экспертами:

1) оценка имеющихся объектов (примером служит оценивание варианта решения по принятию критерия);

2) построение объектов (к задачам такого типа относятся формирование множества стратегий, выявление неопределенных факторов и установление областей их возможных значений);

3) построение объектов и их оценка (типичной задачей данного вида является составление перечня критериев, конструирование их шкал и оценка стратегий (исходов) по тем или иным критериям).

Рассмотрим наиболее употребительные в практике принятия решений типы шкал: номинальную (или классификационную), порядковую, интервалов (отношений, разностей, абсолютную).

Номинальная шкала используется для описания принадлежности элементов к определенным классам. Всем элементам одного и того же класса присваивается одно и то же число, а элементам разных классов - разные числа. Допустима любая замена чисел для обозначения классов, лишь бы это было взаимно-однозначное преобразование и каждый класс получил бы свое число. Это обстоятельство и определяет множество допустимых преобразований номинальной шкалы как множество всех взаимно-однозначных функций. Эта шкала наименее совершенная.

Порядковая шкала используется для измерения упорядочения элементов по одному или нескольким признакам. Она позволяет установить, что один элемент лучше, важнее, предпочтительнее другого или равноценен ему. Порядковая шкала отражает лишь порядок следования элементов и не дает возможности сказать, на сколько или во сколько раз один элемент предпочтительнее другого. Иными словами, в этой шкале нельзя определить меру степени предпочтительности.

Шкала интервалов применяется для отображения величины различия между характеристиками элементов. Она позволяет указать, на сколько один элемент отличается от другого в принятых единицах измерения. Интервальная шкала может иметь произвольные начало и масштаб. Множество допустимых преобразований данной составляют все линейные преобразования. Основным свойством шкалы интервалов является сохранение отношения длин интервалов. Частными случаями шкалы интервалов являются шкала отношений (нулевое начало отсчета) и шкала разностей (произвольное начало отсчета и единичный масштаб), а также абсолютная шкала (нулевое начало отсчета и единичный масштаб отсчета).,

Количественные и качественные шкалы. Номинальная и порядковая шкалы относятся к качественным шкалам. Шкалы интервалов, отношений, разностей и абсолютная относятся к количественным шкалам, которые позволяют установить количественные соотношения между элементами.

Для количественных шкал справедливы аксиомы арифметики. Шкала считается тем более совершенной, чем уже множество допустимых преобразований. С этой точки зрения самой совершенной является абсолютная шкала, наименее совершенной - номинальная.

В зависимости от существа или важности того или иного элемента и его характеристик могут быть использованы разные шкалы. Однако при выборе шкалы необходимо учитывать, какие действия в дальнейшем предполагается производить с оценками по выбранной шкале. Осмысленные арифметические действия можно производить над оценками, имеющими количественную шкалу.

В дальнейшем ограничимся рассмотрением лишь задач первого из трех указанных типов, так как именно на их решение ориентированы существующие математические методы анализа и обработки экспертной информации.

Экспертные оценки объектов, в соответствии с типом их шкал, делятся на количественные, балльные и качественные.

Качественные оценки, к которым относят оценки, полученные в порядковых и номинальных шкалах, называют также элементарными суждениями. К количественным оценкам относятся, например, достоверные эквиваленты и субъективные вероятности.

Методы получения элементарных суждений. Информация в форме элементарных суждений более доступна и надежна, чем количественная и балльная экспертные оценки, так как она наиболее и привычна экспертам. Поэтому именно ее предпочтительнее получать анкетированием. Существуют следующие методы получения элементарных суждений: группировка (сортировка); балльные оценки; ранжирование; попарные сравнения; множественные сравнения.

Группировка состоит в том, что эксперт последовательно относительно предлагаемые ему объекты к одному из заранее установленных классов. Иногда и сами классы возникают в процессе сортировки объектов. Получаемые при группировке оценки объектов имеют номинальную шкалу.

Для математического описания предпочтений широко используются бинарные отношения. Бинарными отношениями можно непосредственно представить результаты оценки предпочтений, выполненные по любому из трех разобранных выше способов. Вообще бинарные отношения могут быть использованы и практически применяются для описания связей самого различного характера между объектами произвольной природы.

Бинарным отношением p на множестве объектов A называется множество упорядоченных пар (a, b) объектов, т.е. формально подмножество совокупности $A \times A$ всех упорядоченных пар (a, b), где $a,b \in A$(последняя запись, как известно, означает, что и a,и b принадлежат А, n.e. являются элементами этого множества).

В дальнейшем мы ограничимся рассмотрением только бинарных отношений. Если объекты a и b связаны отношением p, то пишут $(a, b) \in p$. Если же a и b отношением p не связаны, то этот факт записывают так: $(a, b) \notin p$.

Поскольку отношение есть множество специального вида, то и задавать отношение можно теми же способами, что и множества перечислением всех его элементов или же указанием характеристического свойства, выделяющего его элементы из совокупности всех упорядоченных пар.

Существуют и специальные способы задания отношений - табличный и при помощи графов.

Способы задания отношений. При табличном способе отношение задается специальной матрицей - матрицей смежности, каждая строка и каждый столбец которой соответствуют некоторому объекту, а на пересечении a строки и столбца b ставится 1, если $(a, b) \notin p$, и 0 - в противном случае. Например, коммерческая фирма состоит из генерального директора (ГД), его заместителя (3ГД), а также менеджеров ($М_1$, $М_2$, ... , $М_n$), которые подчинены ГД и 3ГД. Для отношения подчинения p матрица смежности выглядит следующим образом:

$$\begin{Vmatrix} ГД & \text{ЗГД} & \text{М}_1 \ldots M_n \\ 0 & 1 & 1 \ldots 1 \\ 0 & 0 & 1 \ldots 1 \\ 0 & 0 & 0 \ldots 0 \\ \ldots & \ldots & \ldots \ldots \ldots \\ 0 & 0 & 0 \ldots 0 \end{Vmatrix}$$

10.4 Статистические методы оценки результатов

После получения ответов экспертов необходимо провести их оценку. Это позволяет:

1. Оценить согласованность мнений экспертов. При отсутствии значимой согласованности экспертов необходимо выявить причины несогласованности (наличие групп) и признать отсутствие согласованного мнения (ничтожные результаты)

2. Оценить ошибку исследования.

3. Построить модель свойств объекта на основе ответов экспертов

(для аналитической экспертизы).

4. Подготовить отчет с результатами экспертного оценивания. В отчете указывается цель исследования, состав экспертов, полученная оценка и статистический анализ результатов.

В настоящее время существует множество числовых показателей для измерения степени и характера взаимосвязи двух переменных – коэффициентов связи. Наиболее известный из них – коэффициент χ^2 (хи - квадрат), вычисляемый как показатель, фиксирующий степень расхождения реальных и ожидаемых частот, определяется по формуле:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{10.1}$$

где O_i - наблюдаемые частоты;

E_i – ожидаемые частоты;

n – число клеток в таблице.

На основе данного критерия можно хотя бы приблизительно определить наличие либо отсутствие связи, рассчитав коэффициент сопряженности Пирсона (C) и коэффициент сопряженности Крамера (V). Расчет производиться по формулам:

1. Коэффициент сопряженности Пирсона: $C = \sqrt{\dfrac{\chi^2}{\chi^2 + N}}$, (10.2)

где N – число опрошенных.

2. Коэффициент сопряженности Крамера: $V = \sqrt{\dfrac{\chi^2}{N(K-1)}}$, (10.3)

где N – число опрошенных; K – наименьше из чисел (r;c), где r – число строк; c – число столбцов.

Также для анализа взаимосвязи в таблицах сопряженности можно использовать ранговый коэффициент корреляции γ Гудмена – Краскэла:

$$\gamma = \frac{S - D}{S + D} \tag{10.4}$$

Смысл, которого заключается в вычислении количества пар, в которых значения первой переменной не меньше значений второй справа налево (S) и сравнении с количеством пар, в которых значения первой переменной не меньше значений второй слева направо (D).

Вопросы для самопроверки:

1. *Что понимается методом экспертных оценок ?*
2. *Какие проблемы решаются экспертным путем?*
3. *Что такое индивидуальная экспертиза?*
4. *Какие выделяют этапы проведения групповой экспертизы?*
5. *Какие существуют ошибки, приводящие к некачественной экспертизе?*
6. *Какие существуют методы экспертных оценок?*
7. *Какие бываю шкалы оценивания?*
8. *Какие выделяют методы получения элементарных суждений?*

9. С помощью, каких коэффициентов осуществляется оценка согласованности мнения экспертов?

Задания для практических занятий

1. Статистическое моделирование и прогнозирование как наука

Задача 1

На основании научной литературы по статистике, эконометрике и эконометрическому моделированию составить классификационную схему методов прогнозирования.

Задача 2

На основании научной литературы по моделированию составить классификационную схему методов моделирования.

Задача 3

Составьте схему последовательности методики комплексного анализа статистической информации и выявления причинно-следственных связей.

Задача 4

Составьте схему последовательности методики комплексного анализа и прогнозирования динамической информации.

2. Априорный анализ компонент временного ряда

Задача 5

По данным таблицы 5.1 решить задачу.

Таблица 5.1 – Динамика объема выручки торговой организации, тыс.руб.

Года	2008	2009	2010	2011	2012
Выручка, тыс. руб.	132	145	123	190	188

Для анализа динамики рассчитайте:

а) абсолютный прирост объема выручки (на цепной и базисной основе);

б) темпы роста и прироста объема выручки (на цепной и базисной основе);

в) абсолютное значение 1% прироста;

г) средний уровень ряда динамики объема выручки;

д) средний абсолютный прирост, средний темп роста и прироста объема выручки;

е) изобразить интенсивность развития ряда динамики графически. Сделайте выводы.

Задача 6

Заполните таблицу 6.1.

Таблица 6.1 – Динамика товарооборота, усл.ден.ед.

Год	Товарооборот, усл.ден.ед.	Абсолютное изменение, (+,-)		Темп роста, %		Темп прироста, %		Абсолютное значение 1%	
		цеп	баз	цеп	баз	цеп	баз	цеп	баз
2003	234	-	-	-	-	-	-	-	-
2004	254								
2005	236								
2006	267								
2007	256								
2008	301								
2009	298								
2010	287								
2011	291								
2012	300								

Сделайте вывод.

Задача 7

Таблица 7.1 – Динамика выполнения плана реализации продукции ООО «Север»

Год	Выполнение плана реализации продукции, %
2001	103,5
2002	97,6
2003	101,1
2004	84,6
2005	103
2006	100,2
2007	90,5
2008	102,8
2009	99,3
2010	100,1

2011	104
2012	100,8

По данным таблицы 7.1 рассчитайте все возможные дополнительные показатели, характеризующие изменения ряда динамики (для расчета используйте форму таблицы 2.3 лекционного материала).

Задача 8

По данным о динамике производства товара определить недостающие уровни и цепные показатели динамики.

Таблица 8.1 – Динамика производства товара, тыс.шт.

Год	Товар, тыс.шт.	Цепные показатели динамики			
		Абсолютное изменение, тыс.шт.	Темп роста, %	Темп прироста, %	Абсолютное значение 1 % прироста, тыс.шт.
2008	1,2	-	-	-	
2009		0,7			
2010			136,8		
2011					
2012		1,4			0,034

Задача 9

На основании данных таблицы 9.1 рассчитайте средние показатели прибыли организации за 2006-2012 гг.

Таблица 9.1 – Динамика прибыли торговой организации, тыс.руб.

Год	2006	2007	2008	2009	2010	2011	2012
Прибыль, тыс.руб.	134	124	189	123	145	167	190

Сделайте вывод.

Задача 10

Имеются следующие данные об объеме выручки торговой организации за 2012г. на первое число каждого месяца.

Таблица 10.1 – Динамика прибыли торговой организации, тыс.руб.

Январь	Февраль	Март	Апрель	Май	Июнь	Июль

456	458	510	356	425	478	509

Определите среднемесячные уровни прибыли торговой организации за первый, второй кварталы и за полугодие в целом.

3. Моделирование тенденции

Задача 11

По данным таблицы 11.1 определить наличие (отсутствие) тенденции на основании критерия серий, основанного на медиане выборки, критерия «восходящих - нисходящих» серий и методом Фостера-Стюарта.

Таблица 11.1 – Динамика прибыли торговой организации, тыс.руб.

Год	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Прибыль, тыс.руб.	213,6	314,7	246,5	220,2	225,7	248,7	259,2	223,7	214,0	206,1

Сделать вывод.

Задача 12

Имеются следующие данные о численности торговых организаций одного из регионов (таблица 12.1). Определить наличие (отсутствие) тенденции методом усреднения по левой и правой половине.

Таблица 21.1 – Динамика прибыли торговой организации, тыс.руб.

Год	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Число торговых организаций	107	109	108	109	110	109	110	108	108	109

Сделать вывод.

Задача 13

Имеются следующие данные об издержках обращения торговой организации одного из регионов за 2012г. (таблица 13.1). Определить наличие (отсутствие) тенденции методом укрупнения интервалов.

Таблица 13.1 – Динамика издержек обращения торговой организации, руб.

Месяц	Январь	Февраль	Март	Апрель	Май	Июнь
Издержки обращения, руб.	134	136	135	138	139	136
Месяц	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
Издержки обращения, руб.	137	139	134	135	140	136

Сделать вывод.

Задача 14

По данным задачи 13 проведите определение наличия (отсутствия) тенденции методом скользящих средних. Сделайте вывод.

Задача 15

Имеются следующие данные об уровне оптового товарооборота одного из регионов за 2012г. (таблица 15.1). Проведите аналитическое выравнивание по прямой, параболе и степенной функции. Определите наилучшую функцию.

Таблица 15.1 – Динамика оптового товарооборота за 2012 г., млн.руб.

Год	2001	2002	2003	2004	2005	2006
Оптовый товарооборот, млн.руб.	1868	1979	1985	1972	1863	1954
Год	2007	2008	2009	2010	2011	2012
Оптовый товарооборот, млн.руб.	1987	2002	1998	1872	2027	1966

Сделать вывод.

4. Моделирование периодической компоненты

Задача 16

По данным одного из регионов об уровне розничного товарооборота (таблица 16.1) выявить наличие сезонной составляющей и рассчитать величину сезонной волны, построить график.

Таблица 16.1- Динамика розничного товарооборота региона, млн.руб.

Месяц	2010	2011	2012
Январь	110,3	113,6	114,0
Февраль	111,1	114,3	114,7
Март	111,5	114,4	115,1
Апрель	112,0	114,6	115,6
Май	112,6	115,6	116,0
Июнь	116,0	117,1	117,4
Июль	115,9	116,9	118,2
Август	116,2	117,0	118,4
Сентябрь	116,4	116,5	117,5
Октябрь	115,2	116,0	117,0
Ноябрь	115,0	114,9	116,5
Декабрь	112,8	113,8	118,9

Сделать вывод.

Задача 17

Имеются данные о продаже обуви в магазинах города, тыс.руб.

Таблица 17.1 – Динамика продажи обуви в магазинах города, тыс.руб.

Январь	Февраль	Март	Апрель	Май	Июнь
39,88	41,77	52,33	55,23	55,59	56,75
Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
69,54	72,03	71,3	71,72	86,76	88,83

Необходимо провести выравнивание по ряду Фурье, построить график. Сделать вывод.

Задача 18

На основании имеющихся данных об уровне продажи овощей одного из магазинов региона (таблицы 18.1) необходимо построить тренд - сезонную аддитивную модель, построить на ее основе прогноз на 2013 г

Таблица 18.1 – Динамика продажи овощей, тыс.руб.

Год	Квартал	Объем продаж, тыс.руб.
2008	1	198,801
	2	174,857
	3	163,079
	4	171,612
2009	1	160,467
	2	134,185
	3	139,464
	4	158,426
2010	1	116,311
	2	110,264
	3	131,33
	4	136,581
2011	1	126,536
	2	126,043
	3	119,174
	4	113,666
2012	1	113,791
	2	104,694
	3	102,702
	4	123,426

Сделать вывод.

Задача 19

На основании имеющихся данных об уровне продаж тканей одного из магазинов региона (таблицы 18.1) необходимо построить тренд - сезонную мультипликативную модель, построить на ее основе прогноз на 2013 г

Таблица 18.1 – Динамика продажи тканей, тыс.руб.

Год	Квартал	Объем продаж, тыс.руб.
2008	1	345,92
	2	367,32

	3	379,67
	4	393,40
2009	1	358,97
	2	380,22
	3	192,86
	4	389,41
2010	1	365,32
	2	340,22
	3	333,66
	4	301,44
2011	1	305,23
	2	288,88
	3	262,13
	4	265,34
2012	1	237,85
	2	215,48
	3	227,31
	4	236,73

Сделать вывод.

5. Моделирование связных рядов динамики

Задача 20

На основании данных таблицы 20.1 об изменении уровня цен на условную единицу товара фирмы за 12 месяцев 2012 г. рассчитайте коэффициент автокорреляции.

Таблица 20.1 – Динамика уровня цен на продукцию фирмы за 12 месяцев 2012 г.

Месяц	Январь	Февраль	Март	Апрель	Май	Июнь
Цена за единицу товара, руб.	111,34	131,36	161,35	1231,38	1191,39	178,36
Месяц	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
Цена за единицу товара, руб.	171,37	201,39	191,34	171,35	201,40	198,36

Сделать вывод.

Задача 21

На основании таблицы 21.1 и с условием наличия автокорреляции устранить автокорреляцию, используя методы, основанные на преобразовании уровней исходного ряда в новые переменные, не содержащие тенденции.

Таблица 21.1 – Динамика уровня цен на продукцию фирмы за 12 месяцев 2012 г.

Месяц	Январь	Февраль	Март	Апрель	Май	Июнь
Индекс цен на продукцию	1,3	1,4	1,5	1,7	2,1	2,2
Месяц	Июль	Август	Сентябрь	Октябрь	Ноябрь	Декабрь
Индекс цен на продукцию	2,5	2,7	3,0	3,3	3,5	3,7

Сделать вывод.

Задача 22

На основании данных об уровне прибыли организации за 2000-2012 гг. проверить имеющийся динамический ряд на наличие (отсутствие) автокорреляции, используя критерий Дарбина - Уотсона.

Год	Прибыль, тыс.руб.
2000	139,89
2001	122,21
2002	119,25
2003	124,90
2004	117,59
2005	105,92
2006	111,86
2007	116,67
2008	111,45
2009	106,83
2010	109,62
2011	109,69
2012	105,06

Сделать вывод.

6. Моделирование тенденции и колеблемости ряда динамики

Задача 23

По имеющемуся динамическому ряду (таблица 23.1) рассчитать показатели колеблемости.

Таблица 23.1 – Динамика уровня цен за единицу товара фирмы

Год	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Цена за единицу товара, руб.	5,8	4,6	5,2	4,2	4,5	3,4	3,6	2,4	2,5	2,4	2,8

Сделать вывод.

Задача 24

По имеющемуся динамическому ряду (таблица 24.1) рассчитать показатели колеблемости, провести вероятностную оценку существенности параметров тренда и их колеблемости.

Таблица 23.1 – Динамика численности продавцов супермаркета, чел.

Год	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Численность продавцов супермаркета, чел.	29	31	35	35	45	46	45	44	38	37

Сделать вывод.

Задача 25

На основании данных таблицы 25.1 и 25.2 проинтерпретируйте значения полученных показателей (данные рассчитаны по динамическим рядам за 2005-2012 гг.)

Таблица 25.1 – Показатели силы и интенсивности колебаний

Показатели	Амплитуда (размах) колебаний	Среднее (по модулю) отклонение	Среднее квадратическое отклонение от	Коэффициент колеблемости, %

		от тренда	тренда	
Уровень прибыли организации, руб.	8594	275,32	387,63	31,89
Стоимость покупных товаров, руб.	3736	56,92	88,41	30,75

Таблица 25.2 – Показатели устойчивости уровней динамики

Показатели	Размах колеблемости, $R_{\bar{y}}$	Отношение средних уровней, $i_{\bar{y}}$	Среднее линейное отклонение, $a(t)$	Коэффициен устойчивости, K_y	Коэффициент Спирмена, K_p	Индекс корреляции, J_r
Уровень прибыли организации, руб.	0,56	1,0005	330,378	68,11	-0,23	0,66
Стоимость покупных товаров, руб.	-0,47	0,9984	68,303	69,25	-0,19	0,85

Сделать вывод.

7. Статистические методы прогнозирования динамики

Задача 26

По результатам аналитического выравнивания по прямой, параболе второго порядка и степенной функции динамического ряда уровня цен за выполнение строительных работ одной из организаций за 2006-2012гг.(руб.) были получены следующие уравнения и их основные характеристики (таблица 26.1). Проинтерпретируйте полученные результаты. Осуществите прогнозирование на основе уравнений тренда на следующие три года.

Таблица 26.1 – Характеристики трендов развития уровня цен выполнение строительных работ одной из организаций

Форма тренда	Модель	R^2	Стандартная ошибка

Прямая	$\tilde{y}_t = 442{,}15t + 20841$	0,5152	4,45
Парабола второго порядка	$\tilde{y}_t = 92{,}373t^2 - 1920t + 25029$	0,8486	3,44
Степенная	$\tilde{y}_t = 23018t^{-0{,}1597}$	0,6188	4,12

Сделать вывод.

Задача 27

Осуществите прогнозирование на основе средних показателей динамики на три года.

Таблица 27.1 – Динамика числа посетителей сайта объявлений, чел.

Год	2004	2005	2006	2007	2008	2009	2010	2011	2012
Число посетителей сайта объявлений, чел.	20	22	25	30	35	36	40	50	60

Сделать вывод.

Задача 28

Осуществите прогнозирование динамического ряда уровня розничного товарооборота на три года, дайте оценку адекватности полученной модели, проинтерпретируйте полученные результаты.

Таблица 28.1 – Динамика розничного товарооборота за 2012 г., млн.руб.

Год	2001	2002	2003	2004	2005	2006
Розничный товарооборот, млн.руб.	1868	1979	1985	1972	1863	1954
Год	2007	2008	2009	2010	2011	2012
Розничный товарооборот, млн.руб.	1987	2002	1998	1872	2027	1966

Сделать вывод.

Задача 29

Сделайте интервальный прогноз на 2013г. по линейному тренду по следующим данным об уровне продажи товара.

Таблица 29.1 – Динамика продажи товара, тыс.шт.

Год	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Продано, тыс.шт.	16	17	26	24	22	21	32	18	30	20

Сделать вывод.

Задача 30

Осуществите процесс прогнозирования на последующие три года по следующим данным (таблица 30.1) всеми известными вам методами прогнозирования.

Таблица 29.1 – Динамика продажи товара, тыс.шт.

Год	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Продано, тыс.шт.	60	60	53	35	25	44	30	49	68	48

Сделать вывод.

8. Статистические методы моделирования взаимосвязи

Задача 32

На основании имеющихся данных рассчитайте ранговый коэффициент корреляции Спирмена.

Таблица 32.1 – Динамика показателей

Инвестиции в основной капитал на душу населения (x)	5,1	4,2	3,8	11	6,9	7,5	5,5	5,8	4,9	6	10,4	8,8
Среднемесячная заработная плата (y)	3,8	4,1	3	6,3	4,8	5,2	3,7	3,5	4,2	4,5	6,6	6,7

Сделать вывод.

Задача 33

На основании имеющихся данных рассчитайте ранговый коэффициент корреляции Кендалла.

Изучается зависимость между ценой квартиры y и размером ее общей площади x.

Размер общей площади (x), $м^2$	35	35	33	34	38	40	40	39	37	36
Цена квартиры (y), тыс. долл.	29	31	35	35	45	46	45	44	38	37

Сделать вывод.

Задача 34

Необходимо определить степень согласованности мнения пяти экспертов, результаты ранжирования которыми семи объектов приведены в таблице 34.1. Для определения степени согласованности используйте коэффициент конкордации Кендалла.

Номер объекта экспертизы	Оценка эксперта				
	1	2	3	4	5
1	4	6	4	4	3
2	3	3	2	3	4
3	2	2	1	2	2
4	6	5	6	5	6
5	1	1	3	1	1
6	5	4	5	6	5
7	7	7	7	7	7

Сделать вывод.

9. Статистические методы прогнозирования взаимосвязи

Задача 35

По 25 странам изучается зависимость индекса человеческого развития y от переменных:

x_1 - суточная калорийность питания населения, ккал на душу населения;

x_2 – ожидаемая продолжительность жизни при рождении 2003 г., число лет.

Признак	Среднее значение	Среднее квадратическое отклонение	Парный коэффициент корреляции
y	0,85	0,1	$r_{yx_1} = 0.0{,}75$
x_1	3179,24	321,6	$r_{yx_2} = 0.96$
x_2	74,42	4,8	$r_{x_1x_2} = 0.7$

1. Построить уравнение множественной линейной регрессии в стандартизованном масштабе и в естественной форме.

2. Рассчитайте частные коэффициенты эластичности.

3. Рассчитать линейные коэффициенты частной корреляции и коэффициент множественной корреляции.

4. Оцените значимость уравнения регрессии в целом с помощью F – критерия Фишера.

Сделать вывод.

Задача 36

Имеются данные о выручке организации (у), млн.руб. и затратах на реализацию – (х), тыс.руб.

y	8	5	4,9	4	3,8	3,5	3,8	3,7	3,6	3,5	3,4	3	3
x	5	10	12	15	20	22	25	30	35	36	40	50	60

Задание:

1. Рассчитайте оценки параметров уравнения парной линейной регрессии.
2. Оцените тесноту связи между признаками с помощью выборочного коэффициента корреляции. Проверьте значимость коэффициента корреляции ($\alpha = 0{,}05$).
3. Рассчитайте выборочный коэффициент детерминации. Сделайте экономический вывод.
4. Проверьте значимость оценки коэффициента регрессии с помощью критерия Стьюдента при уровне значимости $\alpha = 0{,}05$.
5. Постройте доверительный интервал для коэффициента регрессии. Дайте экономическую интерпретацию.
6. Оцените с помощью F- критерия Фишера значимость уравнения линейной регрессии ($\alpha = 0{,}05$).
7. Рассчитайте выручку организации, если затраты составят 65 тыс.руб. Постройте доверительный интервал для прогнозного значения объясняемой переменной. Сделайте экономический вывод.
8. На поле корреляции постройте линию регрессии.

Задача 37

На основании данных предыдущего задания необходимо:

1. Рассчитать параметры следующих функций:

- степенной;
- равносторонней гиперболы;
- показательной;
- экспоненты;
- обратной.

2. Найти показатели тесноты связи по каждой модели.
3. Оценить каждую модель через показатель детерминации, F – критерий Фишера, ошибку аппроксимации и выбрать наилучшую из них.

10. Эвристические методы прогнозирования социально-экономических явлений

Задача 38

На основании данных таблиц 38.1, 38.2 и 38.3 заполните расчетную таблицу 38.4.

Таблица 38.1 – Зависимость решающих факторов при покупке молока и увеличением цены на продукт

Решающие факторы при покупке молока	Перейдут на более дешевое	Будут покупать то же	Откажутся от употребления	Итого по строке
Производитель	32	372	29	433
Марка	24	179	0	203
Жирность	96	716	2	814
Реклама	2	15	0	17
Цена	3	210	0	213
Итог в группах	157	1492	31	1680

Таблица 38.2 – Зависимость жирности молока и цены за 1 литр

Цена за 1 литр молока	1,5%	2,5%	3,2%	3,5%	Итого по строке
20-28 руб.	0	32	21	10	63
29-34 руб.	43	224	165	85	517
35-40 руб.	12	228	385	72	697
Более 40 руб.	30	123	168	82	403
Итог в группах	85	607	739	249	1680

Таблица 38.3 – Зависимость известных марок молока и цены за 1 литр

Известные марки молока	20-28	29-34	35-40	Более	Итого по

	руб.	руб.	руб.	40 руб.	строке
Кошкинское	1	87	69	41	198
Молоко с большой буквы «М»	22	214	342	274	852
Магистраль	10	144	115	144	413
Самарское	0	2	5	1	8
Давлеканово	30	68	78	33	209
Итог в группах	63	515	609	493	1680

Таблица 38.4 – Значения коэффициентов связи для таблиц зависимости

Взаимосвязи между показателями	Хи – квадрат (χ^2)	Коэффициент сопряженности Пирсона (C)	Коэффициент сопряженности Крамера (V)	Коэффициент корреляции γ Гудмена – Краскэла
1) решающие факторы при покупке молока и увеличение цены на продукт				
2) жирность молока и цена за 1 литр				
3) известные марки молока и цена за 1 литр				

Сделать вывод.

Список литературы

1. Афанасьев и др. Эконометрика: учебник (В.Н.Афанасьев, М.М. Юзбашев, Т.И. Гуляева); под ред. В.Н. Афанасьева. – М.: Финансы и статистика, 2005. – 256с.
2. Бабешко Л.О. Основы эконометрического моделирования: учебн. пособие для вузов / Л.О. Бабешко. – 2-е изд. перераб. и доп. – М.: Ком Книга, 2006. -432 с.
3. Богаткова Л.В., Пройдакова Е.В. Математические методы в исследовании экономического развития регионов Приволжского Федерального округа. // Вопросы статистики. – 2008. - №8. – с. 45-52.
4. Бородич С.А. Эконометрика: учебное пособие. – Минск: ООО «Новое знание», 2005 – 408с.
5. Валентинов В.А. Эконометрика: Учебник. - 2-е изд.- Дашков и ко, 2009. - 488 с.
6. Еремеева Н.С., Лебедева Т.В. Эконометрика: учебн. пособие для вузов. – Оренбург: ОАО «ИПК «Южный Урал», 2010. – 296 с.
7. Золотова Е.А., Чернышева Н.А., Планирование финансовых показателей деятельности филиала коммерческого банка на основе линейных регрессионных моделей. // Финансовый менеджмент. – 2007. - № 27. – с. 7-15.
8. Кремер Н.Ш. Эконометрика: учебник (Н.Ш. Кремер, Б.А. Путко). – М.: ЮНИТИ-ДАНА, 2006 – 311с.
9. Малюгин В.И., Демиденко М.В., Калечиц Д.Л., Миксюк А.Ю., Цукарев Т.В. Разработка и применение эконометрических моделей для прогнозирования и анализа вариантов денежно-кредитной политики. // Прикладная эконометрика. – 2009. - №2. – с. 24-39
10. Мицек С.А., Мицек Е.Б. Эконометрические и статистические оценки инвестиций в основной капитал в регионах России. // Прикладная эконометрика. – 2009. - №2. – с.39-47.
11. Первадчук В.П., Масенко И.Б. Математическая модель прогнозирования финансового состояния предприятия. // Вестник ОГУ. – 2007. - №77. – с. 181-190.

12. Петров А.Н. Эконометрические модели индекса валового регионального продукта. // Экономический анализ: теория и практика. – 2010. - №31. – с. 43-52.
13. Плавинский С.Л. Биостатистика.Планирование, обработка и представление результатов биомедицинских исследований при помощи системы SAS. СПб: Издательский дом СПб МАПО.- 2005.
14. Попова И.Н. Долгосрочный прогноз производства зерна в России на основе гипертренда. // Вопросы статистики. – 2009. - №12. – с. 51-55.
15. Прикладная статистика. Основы эконометрики: учебник для вузов в 2 т. Т. 1. Теория вероятностей и прикладная статистика / С.А.Айвазян, В.С. Мхитарян. – 2-е изд., испр. – М.: ЮНИТИ-ДАНА, 2001. – 656 с.
16. Прикладная статистика. Основы эконометрики: учебник для вузов в 2 т. Т. 2. Основы эконометрики / С.А.Айвазян, В.С. Мхитарян. – 2-е изд., испр. – М.: ЮНИТИ-ДАНА, 2001. –432 с.
17. Реннер А.Г., Седова Е.Н. Методы прогнозирования экономических показателей на основе временных рядов с учетом пространственной неоднородности данных и нелинейной взаимосвязи между факторами. // Вестник ОГУ. – 2007. - №4. – с. 105-111.
18. Садовникова Н.А., Шмойлова Р.А. Анализ временных рядов и прогнозирование. – М.: 2001.
19. Салин В.Н., Левит Б.Ю. Проверка значимости статистических показателей с помощью таблиц их критических значений. // Вопросы статистики. – 2009. - №9. – с. 69-77.
20. Статистика. Контрольные задания и методические указания для студентов заочной формы обучения специальностей 060800, 061100, 311000/Составители: Гришакина Н.И., Фетисова Г.В. Учебное пособие (НовГУ им. Ярослава Мудрого: Великий Новгород, 2009.
20. Шмойлова Р.А., Илюхина И.Е. Построение показателей прогноза связных рядов динамики основных показателей информационно-статистических услуг. //Вопросы статистики. – 2008. - №4. – с. 54-57.
21. Эконометрика: учебник / И.И. Елисеева, С.В. Курышева, Т.В. Костеева и др.; под ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2008. – 576 с.

ПРИЛОЖЕНИЕ 1

Таблица значений F-критерия Фишера при уровне значимости 0,05

d.f.2	d.f.1									
	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,5	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	307	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
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64

30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44

ПРИЛОЖЕНИЕ 2

Критические значения t–критерия Стьюдента при уровне значимости 0,10; 0,05; 0,01 (двухсторонний)

Число степеней свободы	α			Число степеней свободы	α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,495	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

ПРИЛОЖЕНИЕ 3

Значения статистики Дарбина–Уотсона при 5-%-ном уровне значимости

n	$k^1=1$		$k^1=2$		$k^1=3$		$k^1=4$		$k^1=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40	-	-	-	-	-	-	-	-
7	0,70	1,36	0,47	1,90	-	-	-	-	-	-
8	0,76	1,33	0,56	1,78	0,37	2,29	-	-	-	-
9	0,82	1,32	0,63	1,70	0,44	2,13	0,30	2,39	-	-
10	0,88	1,32	0,70	1,64	0,53	2,02	0,38	2,41	0,24	2,82
11	0,93	1,32	0,66	1,60	0,60	1,93	0,44	2,28	0,32	2,65
12	0,97	1,33	0,81	1,58	0,66	1,86	0,51	2,18	0,38	2,51
13	1,01	1,34	0,86	1,56	0,72	1,82	0,57	2,09	0,45	2,40
14	1,05	1,35	0,91	1,55	0,77	1,78	0,63	2,03	0,51	2,30
15	1,08	1,36	0,95	1,54	0,92	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	,186
28	1,33	1,48	126	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83