

ФЕДЕРАЛЬНАЯ СЛУЖБА ИСПОЛНЕНИЯ НАКАЗАНИЙ
Псковский юридический институт

Ю.Н. Дятлов

**СТАТИСТИЧЕСКИЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ
СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ
В УГОЛОВНО-ИСПОЛНИТЕЛЬНОЙ СИСТЕМЕ**

Учебное пособие

Псков 2013

УДК 339.982
ББК 65.054
Д99

Рецензенты:

Т.А. Красева, кандидат экономических наук, доцент;
Г.В. Барина, кандидат экономических наук, доцент.

Дятлов Ю.Н.
Д99 Статистические методы прогнозирования социально-экономических процессов в уголовно-исполнительной системе: учебное пособие. – Псков: Псков. юрид. ин-т ФСИН России, 2013. – 65 с.

В учебном пособии рассмотрена сущность прогнозирования, классификация социально-экономических прогнозов в уголовно-исполнительной системе и статистических методов прогнозирования. Переставлены виды и примеры временных рядов социально-экономических показателей в уголовно-исполнительной системе. Изложены теоретические аспекты, рекомендации, примеры использования основных методов прогнозной экстраполяции и корреляционно-регрессионного анализа для определения перспективных значений социально-экономических показателей в уголовно-исполнительной системе.

Предназначено для курсантов, слушателей и адъюнктов образовательных учреждений ФСИН России, сотрудников уголовно-исполнительной системы.

УДК 343.81
ББК 67.409.021

© Дятлов Ю.Н.
© Псковский юридический институт
ФСИН России, 2013

СОДЕРЖАНИЕ

1. Концептуальные основы прогнозирования социально-экономических процессов в уголовно-исполнительной системе	4
1.1. Сущность социально-экономического прогнозирования.....	4
1.2. Эндогенные факторы и концептуальные положения, лежащие в основе прогнозирования развития пенитенциарного производства.....	8
1.3. Классификация социально-экономических прогнозов в уголовно-исполнительной системе и статистических методов прогнозирования.....	11
Контрольные вопросы	17
2. Временные ряды, используемые в прогнозировании социально-экономических процессов	18
2.1. Виды и примеры временных рядов социально-экономических показателей деятельности исправительных учреждений	18
2.2. Требования, предъявляемые к временным рядам	23
2.3. Компоненты временных рядов.....	26
Контрольные вопросы	27
3. Использование основных методов экстраполяции для прогнозирования социально-экономических процессов в уголовно-исполнительной системе	28
3.1. Простая экстраполяция	28
3.2. Оценка точности моделей и прогнозов, полученных методами экстраполяции	31
3.3. Метод скользящих средних	33
3.4. Метод экспоненциального сглаживания	40
Контрольные вопросы	46
4. Корреляционно-регрессионный анализ в социально-экономическом прогнозировании	47
4.1. Основные понятия корреляционно-регрессионного анализа.....	47
4.2. Классификация видов регрессии. Метод наименьших квадратов.....	49
4.3. Точность и надежность модели линейной регрессии. Доверительные интервалы прогноза	54
Контрольные вопросы	65

1. КОНЦЕПТУАЛЬНЫЕ ОСНОВЫ ПРОГНОЗИРОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ В УГОЛОВНО-ИСПОЛНИТЕЛЬНОЙ СИСТЕМЕ

1.1. Сущность социально-экономического прогнозирования

Уголовно-исполнительная система (УИС) является одним из важнейших государственных социально-экономических институтов, участвует в создании общественных, экономических благ в процессе реализации как правоохранительных, так и социально-экономических функций. В условиях последовательного реформирования УИС России в соответствии с требованиями международных и отечественных стандартов значительно возрастает роль социально-экономического прогнозирования как необходимой базы для выработки оптимальной стратегии развития производственного сектора УИС и связанных с ним социальных процессов.

Существуют различные подходы к определению категорий «прогноз» и «прогнозирование». Обычно **прогноз** рассматривается как научно-обоснованная гипотеза о вероятном, будущем состоянии социально-экономической системы, хозяйствующих объектов и характеризующих это состояние показателей. Соответственно процесс разработки, составления прогнозов называют **прогнозированием**¹.

Большинство представленных в научной и учебной литературе определений подчеркивает вероятностный аспект данных категорий. В частности, прогнозирование трактуется как «...вероятностная оценка возможного развития экономических процессов в фиксированный интервал времени»². В другом определении, которое также может быть отнесено к данному аспекту, прогнозирование представлено как научное выявление вероятных путей и результатов

¹ См. подр.: *Райзберг Б.А.* Современный экономический словарь / Б.А. Райзберг, Л.Ш. Лозовский, Е.Б. Стародубцева; Ред. Б.А. Райзберг. М., 2011. С. 265.

² См. подр.: *Федоров Н.В.* Прогнозирование социально-экономического развития регионов Российской Федерации. М., 1998. С. 20.

развития явлений и процессов, оценка показателей характеризующих эти явления и процессы для более или менее отдаленного будущего¹.

Приведенные выше определения в большей степени отражают качественные характеристики – надежность прогноза и верифицируемость (достоверность) прогнозирования, чем их содержательную сторону. С позиций **вероятностного подхода** сущность прогнозирования заключается, прежде всего, в минимизации неопределенности будущего, исследовании области допустимых траекторий развития процессов, в рамках которых целесообразно обоснование регулирующих воздействий.

Более полное определение категории «прогноз» дает И.В. Бестужев-Лада. По его мнению, «прогноз (от греч. prognosis - предвидение, предсказание) – это научно обоснованное суждение о возможных состояниях объектов в будущем и (или) об альтернативных путях и сроках их осуществления»². Здесь четко выражен функциональный аспект прогноза, предусматривающий его поисковую и нормативную (нормативно-целевую) направленность, и что не менее важно, отмечен научный подход к решению прогнозных задач.

С точки зрения **функционального подхода** вполне адекватным является определение экономического прогнозирования как научного предвидения о направлениях развития экономики или отдельных ее элементов в будущем или поиск оптимальных способов достижения поставленных целей³.

Определенный интерес представляет трактовка сущности прогноза как комплекса «аргументированных предположений, выраженных в качественной и количественной форме, относительно параметров экономической системы»⁴, выделяющая качественный (описательный) и количественный способы представления прогноза. Кроме того, заслуживает внимания определение сущности прогнозирования, в котором оно рассматривается как «...предопределение бу-

¹ См. подр.: *Гамбаров Г.М.* Статистическое моделирование и прогнозирование / Г.М. Гамбаров, Н.М. Журавель, Ю.Г. Королев. М., 1990.

² См. подр.: *Рабочая книга по прогнозированию* / Отв. ред. И.В. Бестужев-Лада. М., 1982.

³ См. подр.: *Прудников А.Г.* Прогнозирование и планирование в условиях рынка: учебное пособие для студентов экономических специальностей. Краснодар, 1999. С. 51.

⁴ См. подр.: *Прогнозирование и планирование в условиях рынка: учебное пособие для ВУЗов* / Под ред. Т.Г. Морозовой, А.Г. Пикулькина. М., 2003. С. 8.

дущего, построение активной модели желаемого и намечаемого будущего состояния экономики при одновременном установлении путей, способов, средств и сроков достижения этого состояния, конечных рубежей прогнозируемых действий»¹. Такой подход подчеркивает активную направленность прогнозирования, предполагающую необходимые воздействия на ход развития событий, что является характерной чертой функционирования современных социально-экономических систем.

В Федеральном законе «О государственном прогнозировании и программах социально-экономического развития Российской Федерации» № 115-ФЗ от 20.07.1995 отражены концептуальные подходы к государственному прогнозированию. Он регламентирует процесс разработки системы государственных прогнозов и программ социально-экономического развития Российской Федерации, устанавливает их содержание. В данном законе государственное прогнозирование рассматривается как «...система научно-обоснованных представлений о направлениях социально-экономического развития Российской Федерации, основанных на законах рыночного хозяйствования»².

Прогнозирование соотносится с более широким понятием - предвидением. Различают три формы предвидения: гипотезу, прогноз, план.

Гипотеза характеризует научное предвидение на уровне общей теории, то есть исходную базу построения гипотезы составляют теория и открытые на ее основе закономерности и причинно-следственные связи функционирования и развития исследуемых объектов. На уровне гипотезы дается качественная характеристика объектов.

Прогноз имеет значительно большую определенность, так как основан не только на качественных, но и на количественных параметрах. Прогноз выражает предвидение на уровне конкретно-прикладной теории. Он отличается от ги-

¹ См. подр.: *Райзберг Б.А.* Современный экономический словарь / Б.А. Райзберг, Л.Ш. Лозовский, Е.Б. Стародубцева; Ред. Б.А. Райзберг М., 2011. С. 265.

² См.: Федеральный закон от 20.07.1995 "О государственном прогнозировании и программах социально - экономического развития Российской Федерации" // Собрание законодательства Российской Федерации. 1995. № 30. Ст. 2871.

потезы меньшей степенью неопределенности и большей достоверностью. Прогноз носит вероятностный характер.

План представляет собой постановку точно определенной цели и предвидение конкретных событий исследуемого объекта. В нем фиксируются пути и средства развития в соответствии с поставленными задачами, обосновываются принятые управленческие решения. Отличительная черта плана – определенность. План основывается на результатах и достижениях конкретно прикладной теории.

Отечественный футуролог И.В. Бестужев-Лада дифференцировал понятия «прогнозирование» и «планирование» как «предсказание» и «предуказание». По его мнению, в состав понятия «**предсказание**» входят предчувствие и предугадывание. Если **предчувствие** представляет собой вероятное описание событий будущего на основе собственных знаний, эрудиции, работы подсознания, то **предугадывание** основывается на использовании житейского опыта и сравнительно подробном знании ситуации, в которой будут развиваться вероятные события. **Предуказание** включает в себя полностью планирование и основывается на принятии управленческих решений, учете наиболее вероятных критических аспектов будущего¹.

Процесс прогнозирования социально-экономических явлений включает в себя следующие этапы:

1. Постановка задачи и сбор необходимой информации.
2. Первичная обработка исходных данных.
3. Определение возможных моделей прогнозирования.
4. Оценка параметров моделей.
5. Исследование качества выбранных моделей, адекватности их реальному процессу.
6. Выбор лучшей из моделей.
7. Построение прогноза.

¹ См. подр.: Рабочая книга по прогнозированию / Отв. ред. И.В. Бестужев-Лада. М., 1982.

8. Содержательный анализ полученного прогноза.

Прогнозирование социально-экономических процессов является базой для разработки планов и программ развития производственного сектора УИС, совершенствования системы трудовой адаптации и профессиональной подготовки осужденных. Прогнозы постоянно используются не только для предплановых обоснований, но и на стадии реализации планов и программ.

1.2. Эндогенные факторы и концептуальные положения, лежащие в основе прогнозирования развития пенитенциарного производства

В процессе прогнозирования развития социально-экономических явлений в производственной сфере УИС, в отличие от других секторов национальной экономики, необходимо учитывать ряд **эндогенных факторов**. Такие факторы являются внутренними и обусловлены закономерностями развития как УИС в целом, так и отдельных ее элементов. К эндогенным факторам, прежде всего можно отнести следующие:

- необходимость привлечения осужденных в качестве основной части производственного персонала;
- доминирование цели исправления осужденных над целью максимизации прибыли в пенитенциарном производстве;
- обеспечение установленных законодательством требований по изоляции осужденных в связи с их опасностью для общества, в том числе в процессе оказания услуг по предоставлению рабочей силы из числа осужденных при организации взаимодействия со сторонними партнерами;
- низкий образовательный и профессиональный уровень осужденных, что требует их массового профессионального и трудового обучения;
- пространственное единство имущественного комплекса жилой и производственной зон исправительных учреждений;

- сочетание в составе персонала при осуществлении производственной деятельности осужденных, вольнонаемного персонала и аттестованных сотрудников, трудовая мотивация которых существенно дифференцирована;

- осуществление на производственных площадках надзора, соблюдение режимных требований, проведение профилактических мероприятий и выполнение других специальных функций, предусмотренных должностными обязанностями сотрудников исправительных учреждений;

- существенные ограничения, невозможность получения казенными учреждениями УИС кредитов (займов), использования внутренних финансовых источников для обновления основных фондов и пополнения оборотных средств;

- сложность реализации снабженческих и сбытовых функций, обусловленная исполнением режимных требований;

- определенные ограничения свободного доступа партнеров, с которыми реализуются совместные проекты, к складам, производственным площадкам, размещенным на территориях исправительных учреждений.

Рассмотрим ряд положений Концепции развития уголовно-исполнительной системы в Российской Федерации до 2020 г., которые следует учитывать в процессе разработки целевых прогнозов развития производственного сектора УИС.

Во-первых, планируется рост числа колоний-поселений до 210 к 2020 г., включая новые колонии-поселения с усиленным наблюдением. Такие колонии будут создаваться в основном на базе существующих исправительно-трудовых учреждений, располагающих развитой и более конкурентоспособной производственной базой. Дислокация колоний-поселений с обычным наблюдением будет осуществляться с учетом географии региональных инвестиционных, производственных проектов в целях задействования в них труда осужденных. Рас-

считается также возможность создания небольших камер-мастерских и внедрения индивидуальных форм занятости осужденных¹.

В-вторых, для развития и повышения эффективности пенитенциарного производства рекомендуется усилить совместную работу УИС с федеральными органами государственной власти Российской Федерации, связанную с решением вопроса о предоставлении исправительным учреждениям преференций, обязывающих государственных и муниципальных заказчиков размещать часть заказов на приобретение необходимой им продукции, выполнение работ, оказание услуг в исправительных учреждениях. Доля таких поставок учреждениями УИС должна составлять от 10 до 20% объема поставок товаров для государственных и муниципальных нужд по аналогии с субъектами малого предпринимательства.

Третьим перспективным направлением является установление специального налогового режима или полное освобождение исправительных учреждений от налогового бремени с целью направления доходов от труда осужденных на улучшение условий их содержания, создание дополнительных рабочих мест и развитие производственной и социальной сферы УИС.

Кроме того, предусмотрена активизация совместной работы учреждений ФСИН с органами власти субъектов Российской Федерации по установлению квот предприятиям и организациям для трудоустройства осужденных и льгот работодателям, предоставляющим для них рабочие места.

В отдельных случаях прогнозируется возможность оказания финансовой помощи за счёт региональных и муниципальных бюджетов в поддержании находящихся на балансе исправительных учреждений объектов, обеспечивающих жизнедеятельность населенных пунктов, для которых они являются градообразующими (объекты топливно-энергетического хозяйства, медицинского, социального, культурного назначения, автомобильные и железные дороги, мосты и т.д.). Предусматривается разработка мероприятий по привлечению бизнеса к

¹ См.: Концепция развития уголовно-исполнительной системы Российской Федерации до 2020 года: утв. распоряжением Правительства Российской Федерации от 14 октября 2010 г. № 1772-р // Собрание законодательства Российской Федерации. 2010. № 43. Ст. 5544.

созданию производственных участков в колониях-поселениях и к иным формам совместных производственных и инвестиционных проектов.

Для расширения производства сельскохозяйственной продукции рекомендуется принятие стимулирующих мер по созданию колоний-поселений с сельскохозяйственным производством в благоприятных природно-климатических зонах.

С целью разработки перспективных направлений производственной деятельности учреждений УИС с учетом приоритетного обеспечения нужд системы собственным производством и выпуска конкурентоспособной продукции, предусматривается систематическое проведение анализа состояния производственной и договорной базы УИС¹.

В целом указанные выше ориентиры целевого прогноза развития УИС до 2020 г. предполагают существенную трансформацию ее производственного сектора, приведение его структуры в соответствии с требованиями законодательства Российской Федерации и норм международного права.

1.3. Классификация социально-экономических прогнозов в уголовно-исполнительной системе и статистических методов прогнозирования

Основными социально-экономическими индикаторами, значения которых детально анализируют и прогнозируют при разработке программ и планов развития УИС, являются:

- количество трудоспособных осужденных;
- среднесписочная численность осужденных, подлежащих к обязательному привлечению к труду по объективным причинам;
- численность осужденных, занятых на оплачиваемых работах;
- средняя заработная плата на 1 отработанный чел.-день;
- вывод осужденных на оплачиваемые работы (в процентах);

¹См.: Концепция развития уголовно-исполнительной системы Российской Федерации до 2020 года: утв. распоряжением Правительства Российской Федерации от 14 октября 2010 г. № 1772-р // Собрание законодательства Российской Федерации. 2010. № 43. Ст. 5544.

- выполнение спецконтингентом норм выработки (в процентах) и др.

Различные социально-экономические показатели, характеризующие деятельность производственных подразделений УИС, могут быть сгруппированы в следующие объекты прогнозирования:

- с полным обеспечением количественной информацией, для которых имеется в наличии ретроспективная количественная информация в необходимом объеме;

- с неполным обеспечением количественной информацией, для которых имеющаяся в наличии ретроспективная информация допускает использование статистических методов;

- с наличием качественной ретроспективной информации и полным отсутствием либо чрезвычайной ограниченностью количественной информации;

- с полным отсутствием ретроспективной информации, что характерно для проектируемых объектов.

На практике для первого случая могут с уверенностью применяться **статистические методы прогнозирования**, которые позволяют выявлять закономерности на фоне случайностей, делать обоснованные прогнозы и оценивать вероятность их выполнения. Статистические методы также используются и для второго случая, но с некоторым уменьшением точности прогноза. Для двух последних случаев наиболее эффективно применение экспертных методов прогнозирования.

Классификация социально-экономических прогнозов в УИС может строиться по различным критериям в зависимости от целей, задач, объектов, проблем, периода упреждения и т.д.

В зависимости от **масштабности объекта прогнозирования** социально-экономические прогнозы в УИС могут охватывать два уровня: от микроуровня (рассматривающего прогнозы развития отдельных пенитенциарных производств: центров трудовой адаптации осужденных, производственных (трудовых) мастерских исправительных учреждений, федеральных государственных унитарных предприятий УИС, проектов, выполняемых на объектах организа-

ций любых организационно-правовых форм, расположенных на территориях учреждений УИС и вне их и т.п.) до макроуровня (анализирующего развитие социально-экономических процессов в УИС в масштабе всего государства).

Важной характеристикой является время упреждения прогноза - отрезок времени от момента, для которого имеются последние статистические данные об изучаемом объекте УИС, до момента, к которому относится прогноз. По **времени упреждения** прогнозы делятся на следующие виды:

- **оперативные** (с периодом упреждения до одного месяца). Такие прогнозы основаны на предположении о том, что в прогнозируемом периоде не произойдет существенных изменений в исследуемом объекте. В них преобладают детально количественные оценки ожидаемых событий;

- **краткосрочные** (период упреждения - от одного, нескольких месяцев до года) предполагают только количественные изменения. Оценка событий дается количественная;

- **среднесрочные** (период упреждения более 1 года, но не превышает 5 лет) проводится количественно-качественная оценка событий;

- **долгосрочные** (с периодом упреждения более 5 лет) предполагают преимущественно качественно-количественную или качественную (описательную) форму оценки прогнозируемых событий.

По **проблемно-целевому критерию** различают два типа прогнозов:

- **поисковый** (генетический, трендовый) прогноз определяет возможные состояния явления в будущем. Предполагается условное продолжение на перспективу тенденций развития изучаемого явления в прошлом и настоящем, абстрагирование от возможных событий, способных радикально изменить сложившиеся тенденции. Такой прогноз отвечает на вопрос: что вероятнее всего произойдет при условии сохранения существующих тенденций? В качестве примера поискового прогноза можно привести следующее суждение: «в связи с расширением правоприменительной практики судов по назначению наказаний, альтернативных лишению свободы, дальнейшей гуманизацией уголовных наказаний, предполагаемой передачей в уголовно-исполнительную систему полно-

мочий по осуществлению контроля за условно-досрочно освобожденными из мест лишения свободы прогнозируется увеличение численности лиц, состоящих на учете в уголовно-исполнительных инспекциях»¹;

- **нормативный** (целевой, программный) прогноз определяет пути и сроки достижения возможных состояний явления, принимаемых в качестве цели. Предусматривается прогнозирование достижения желаемых состояний на основе заданных норм, целей, стимулов в обратном порядке: от будущего состояния к настоящему. Данный прогноз отвечает на вопрос: какими путями достичь желаемого? Элементы нормативного прогноза содержатся в Концепции развития уголовно-исполнительной системы в Российской Федерации до 2020 г., например, следующее ее положение: «В сфере трудовой деятельности и профессиональной подготовки осужденных предполагается ... осуществление профессионального обучения и профессиональной подготовки осужденных с учетом результатов мониторинга прогнозных потребностей в рабочих кадрах учреждений уголовно-исполнительной системы и региональных рынков труда, в том числе по дефицитным рабочим специальностям, в целях создания высоких гарантий трудоустройства и возвращения в общество законопослушных граждан»².

В практике прогнозирования социально-экономических процессов преобладающими являются статистические методы. Это вызвано, главным образом, тем, что статистические методы опираются на аппарат анализа, развитие и практика применения которого имеют достаточно длительную историю. По общему принципу статистические методы относятся к группе формализованных методов прогнозирования.

С позиций общего подхода совокупность статистических методов прогнозирования, направленных на решение прикладных задач анализа состояния объекта и разработки прогноза его развития, можно дифференцировать на две

¹ См.: Концепция развития уголовно-исполнительной системы Российской Федерации до 2020 года: утв. распоряжением Правительства Российской Федерации от 14 октября 2010 г. № 1772-р // Собрание законодательства Российской Федерации. 2010. № 43. Ст. 5544.

² См.: Там же.

самостоятельные группы: методы прогнозной экстраполяции и статистическое моделирование (рис. 1). В состав первой группы входят методы простой экстраполяции, скользящих средних, экспоненциального сглаживания, аналитического выравнивания динамических рядов, прогнозирования на основе уравнений авторегрессии.



Рис. 1. Классификация статистических методов прогнозирования

Статистические модели, используемые в прогнозировании, могут строиться на основе одного уравнения регрессии или на основе системы уравнений регрессии. В отдельных случаях их называют производственными функциями. Лежащий в их основе корреляционно-регрессионный анализ позволяет определить количественную величину связей между изучаемыми социально-экономическими явлениями.

Описание основных критериев использования отдельных статистических методов прогнозирования представлено в табл. 1.

Процесс прогнозирования, опирающийся на статистические методы, включает на два этапа.

Таблица 1

Характеристика отдельных статистических методов прогнозирования

Метод прогнозирования	Количество статистических данных	Модель данных	Горизонт прогнозирования	Время, затрачиваемое на подготовку прогноза	Подготовка персонала
Простое скользящее среднее	От 3 до 5 наблюдений	Данные должны быть стационарными	Краткосрочный	Малое	Особой подготовки не требуется
Взвешенное скользящее среднее	От 3 до 5 наблюдений	Данные должны быть стационарными	Краткосрочный	Малое	Особой подготовки не требуется
Экспоненциальное сглаживание	От 5 до 10 наблюдений для установления весовых коэффициентов	Данные должны быть стационарными	Краткосрочный	Малое	Достаточно общей подготовки
Причинные регрессионные модели	8-10 и более наблюдений	Может обрабатывать сложные модели	Краткосрочный, среднесрочный и долгосрочный	Имеет длительный период разработки и малое время внедрения	Высокий уровень подготовки

Первый заключается в обобщении данных, собираемых за некоторый период времени, а также создании на основе этого обобщения модели процесса. Модель описывается в виде аналитически выраженной тенденции развития (экстраполяция тренда) или в виде функциональной зависимости от одного или нескольких факторов (аргументов) уравнения регрессии. Построение модели для прогнозирования, какой бы вид она ни имела, обязательно включает выбор формы уравнения, описывающего динамику и взаимосвязь явлений, и оценивание его параметров с помощью того или иного метода.

Второй этап - сам прогноз. На этом этапе на основе найденных закономерностей определяется ожидаемое значение прогнозируемого показателя, величины или признака. Безусловно, полученные результаты не могут рассматри-

ваться как нечто окончательное, так как при их оценке и использовании должны приниматься во внимание факторы, условия и ограничения, которые не участвовали в описании и построении модели. Корректировка результатов должна осуществляться в соответствии с ожидаемым изменением условий их формирования.

Контрольные вопросы

1. Дайте определение категорий «прогноз» и «прогнозирование».
2. Охарактеризуйте сущность прогнозирования с точки зрения вероятностного и функционального подхода.
3. Раскройте содержание понятий «гипотеза», «прогноз», «план».
4. Какие этапы включает процесс прогнозирования социально-экономических явлений?
5. Перечислите эндогенные факторы, которые необходимо учитывать при прогнозировании развития социально-экономических процессов в производственной сфере УИС.
6. Какие положения Концепции развития уголовно-исполнительной системы в Российской Федерации до 2020 г. должны найти отражение в процессе разработки целевых прогнозов развития производственного сектора УИС?
7. Назовите основные социально-экономические индикаторы деятельности исправительных учреждений.
8. Дайте классификацию социально-экономических прогнозов в УИС.
9. Причислите статистические методы прогнозирования.
10. Охарактеризуйте основные критерии использования статистических методов прогнозирования.

2. ВРЕМЕННЫЕ РЯДЫ, ИСПОЛЬЗУЕМЫЕ В ПРОГНОЗИРОВАНИИ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

2.1. Виды и примеры временных рядов социально-экономических показателей деятельности исправительных учреждений

Различные задачи прогнозирования требуют использования статистических данных, характеризующих исследуемые социально-экономические процессы в УИС, развернутых во времени в форме временных рядов. В них концентрировано отражаются изменения указанных процессов и явлений, позволяя провести необходимый анализ и разработать прогноз. При этом одни и те же временные ряды могут использоваться для решения разнообразных по постановке и содержанию проблем.

Под **временным (динамическим) рядом** понимается множество последовательных наблюдений за значениями некоторого показателя (признака), упорядоченных во времени, в порядке возрастания временного параметра. При этом отдельные наблюдения данного ряда называют **уровнями** временного ряда.

Таким образом, любой временной ряд состоит из двух совокупностей элементов - значений времени и соответствующих им значения уровней ряда.

В зависимости от характера временного параметра ряды делятся на моментные и интервальные¹.

В **моментных** временных рядах уровни характеризуют значения показателя по состоянию на определенные моменты времени. Например, моментными являются временные ряды цен на определенные виды товаров, закупаемых для нужд учреждений и органов УИС, уровни которых фиксируются на конкретную дату. К моментным также относятся ряды численности осужденных или стоимости основных фондов исправительных учреждений, поскольку значения уровней этих рядов определяются ежегодно на одно и то же число.

В **интервальных** рядах уровни характеризуют значение показателя за определенные периоды времени. Примерами рядов этого типа могут служить

¹ См. подр.: Дуброва Т.А. Статистические методы прогнозирования: учеб. пособие для вузов. М., 2003. С. 8.

временные ряды выпуска продукции производственными подразделениями УИС в натуральном или стоимостном выражении за месяц, квартал, год и т.д.

Моментный временной ряд представлен в табл. 2 и на рис. 2, интервальный - в табл. 3.

Таблица 2

Численность лиц, состоявших на учете в уголовно-исполнительных инспекциях (человек)¹

Дата	31.12.2006	31.12.2007	31.12.2008	31.12.2009	31.12.2010	31.12.2011	31.12.2012
Состоит на учете	574441	590703	558346	534409	491257	474775	465740

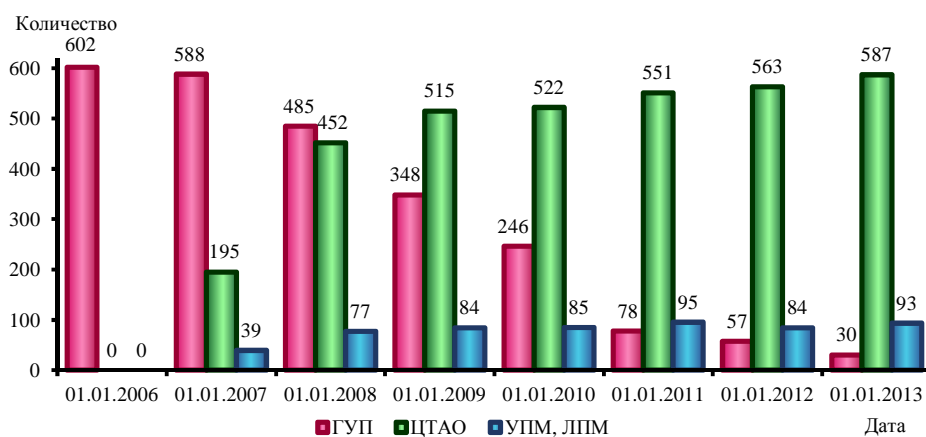


Рис. 2. Динамика реорганизации государственных унитарных предприятий (ГУП) УИС в центры трудовой адаптации осужденных (ЦТАО), учебные и лечебные производственные мастерские (УПМ, ЛПМ)²

В отдельных случаях уровни ряда представляют собой не непосредственно наблюдаемые значения, а производные величины: средние или относительные. Поэтому такие ряды называются **производными**. Уровни производных временных рядов определяются с помощью вычислений на основе непо-

¹ См.: Характеристика лиц, состоящих на учете в уголовно-исполнительных инспекциях <http://фсин.рф/structure/inspector/iao/statistika/Xar-ka%20v%20YII/>

² См.: Показатели, характеризующие производственный сектор УИС// <http://фсин.рф/structure/adaptation/soveshchanie-chelyabinsk-28-05-2013/index.php>

средственно наблюдаемых показателей. Примерами подобных рядов могут служить ряды средней заработной платы осужденного на 1 отработанный чел.-день, или ряды показателей вывода спецконтингента на оплачиваемые работы (в процентах) (табл. 4).

Таблица 3

Динамика объема выпуска товарной продукции
производственным сектором УИС в действующих ценах (млрд. руб.)¹

Год	2005	2006	2007	2008	2009	2010	2011	2012
Товарная продукция - всего, в том числе для внутри- системных нужд	18	17,8	22,6	26,2	25,2	29,4	30,3	30,7
	7,1	6,8	8,1	11,1	11,6	14,6	13,4	14,6

Таблица 4

Изменение удельного веса вывода осужденных на оплачиваемые работы
(без учета занятых на работах по хозяйственному обслуживанию)²

Год	Количество трудоспособных осужденных, тыс. чел.	Занято на оплачиваемых работах, тыс. чел.	Вывод на оплачиваемые работы, %
1	2	3	4
2005	618,4	182,9	29,6
2006	667,8	179,1	26,8
2007	686,8	182,5	26,6
2008	703,6	181,4	25,8
2009	693,5	149,4	21,5
2010	622,9	156,3	25,1
2011	575,1	158,4	27,5
2012	606,0	158,9	26,2

Данные графы 4 табл. 4 получаются с помощью деления данных графы 3 на графу 2 и умножения полученного значения на 100.

В качестве отличительной черты интервальных рядов динамики абсолютных величин следует отметить возможность определения суммы их уров-

¹ См.: Там же.

² См.: Там же.

ней, что позволяет получить накопленные итоги, которые имеют осмысленное содержание вследствие отсутствия повторного счета. Например, суммируя объем выпуска продукции производственными подразделениями УИС в натуральном или стоимостном выражении за первые три месяца и три последующих месяца, получаем, соответственно, объем производства продукции за первый и второй кварталы, а сумма этих квартальных данных дает объем выпуска продукции за полугодие.

Определение суммы уровней моментного временного ряда нецелесообразно, поскольку полученные таким образом значения не имеют смысла. Например, уровни моментного ряда «Численность лиц, состоявших на учете в уголовно-исполнительных инспекциях (на начало года)» (табл. 2) содержат элементы повторного счета. Второй уровень частично содержит лиц, состоявших на учете более года, учтенных ранее первым уровнем и т.д. Таким образом, моментные ряды динамики, в отличие от интервальных не обладают свойством аддитивности (добавления).

Однако для анализа моментного ряда динамики определенный смысл имеет расчет разности уровней, характеризующих изменение показателя за некоторый отрезок времени. Например, за 2012 г. количество ГУП УИС снизилось на 27, тогда как число ЦТАО возросло на 24, УПМ и ЛПМ - на 9 (рис. 2).

В аналитической деятельности учреждений и органов УИС часто применяются временные ряды с **нарастающими итогами**. При этом получают обобщающий результат развития показателя с начала отчетного периода (квартала, полугодия, года и т.д.). В качестве примера можно привести сбор данных о ежемесячных объемах производства продукции (выполнения работ, оказания услуг) производственными подразделениями исправительных учреждений, суммируя которые, получают ежеквартальные, полугодовые и годовые показатели.

Выделяют также **одномерные** временные ряды, полученные при фиксированной количественной характеристике, и **многомерные** временные ряды,

формируемые при наблюдении нескольких характеристик выделенного объекта.

Уровни ряда могут принимать детерминированные или случайные значения. Примером ряда с детерминированными значениями уровней служит ряд последовательных данных о количестве дней в месяцах. Естественно, анализу, а в дальнейшем и прогнозированию, подвергаются ряды со случайными значениями уровней. В таких рядах каждый уровень может рассматриваться как реализация случайной величины - дискретной или непрерывной¹.

Качество статистического анализа и прогнозирования в значительной степени зависит от соблюдения определенных правил построения временных рядов, среди которых важное значение имеет выбор оптимальных интервалов между соседними уровнями ряда. Равноотстоящие друг от друга уровни ряда наиболее удобны для исследования. Однако выбирая слишком большой интервал времени, можно упустить существенные закономерности в динамике показателя. Так, по квартальным данным невозможно анализировать месячные сезонные колебания. В то же время, малые интервалы между наблюдениями увеличивают объем вычислений и могут приводить к появлению ненужных деталей в динамике процесса.

Как правило, интервал времени между уровнями временного ряда социально-экономических показателей, характеризующих деятельность УИС, определяется лицами, осуществляющими аналитическую деятельность, исходя из целей каждого конкретного исследования, а также с учетом требований ведомственной статистической отчетности.

В дальнейшем, если не оговорено иначе, будем рассматривать одномерные, дискретные с равноотстоящими моментами наблюдений, случайные временные ряды.

¹ См. подр.: Дуброва Т.А. Статистические методы прогнозирования: учеб. пособие для вузов. М., 2003.

2.2. Требования, предъявляемые к временным рядам

Применяемые при обработке и прогнозировании временных рядов методы во многом опираются на математический аппарат, который предъявляет определенные требования к исходной информации. К таким требованиям относят сопоставимость, однородность данных, отсутствие пропусков наблюдений и др.

Сопоставимость уровней ряда является одной из основных характеристик, необходимых для адекватного отражения временным рядом развития реального социально-экономического процесса. Она достигается в результате одинакового подхода к наблюдениям на разных этапах формирования динамического ряда. Уровни во временных рядах должны иметь одинаковые:

- единицы измерения;
- шаг наблюдений;
- интервал времени;
- методику расчета;
- элементы, относящиеся к неизменной совокупности.

Для несопоставимых величин некорректно проводить исследование динамики. Например, уровни длительного временного ряда (за 15-20 лет), характеризующие объем товарной продукции производственного сектора УИС в стоимостном выражении, могут оказаться несопоставимыми из-за изменения цен в результате инфляции, наблюдавшейся в анализируемом периоде. Другой причиной несопоставимости являются структурные изменения, например, продолжающийся процесс преобразования государственных унитарных предприятий исправительных учреждений УИС в центры трудовой адаптации, создания лечебно-производственных и учебно-производственных мастерских, или укрупнение производства за счет слияния нескольких предприятий в одно объединение.

В большинстве случаев несопоставимость данных удастся устранить путем пересчета показателей с помощью формальных методов. Например, использовать в условиях инфляции индексы цен для перевода стоимостных пока-

зателей в цены базового года. Однако не всегда проведение такой обработки обеспечивает требуемую точность исходной информации, что может привести к затруднению дальнейшего анализа и прогнозирования.

Однородность данных означает отсутствие сильных изломов тенденций, а также аномальных, то есть резко выделяющихся данных, нетипичных для данного ряда наблюдений. Аномальные наблюдения проявляются в виде сильного изменения уровня – скачка или спада – с последующим приблизительным восстановлением предыдущего уровня. Часто появление таких значений может быть вызвано ошибками при сборе, записи и передаче информации. Возможными источниками появления ошибочных значений является сдвиг запятой при перенесении информации из документа, занесение данных в другую графу и т.п. Наличие аномалии резко искажает результаты прогнозирования. Поэтому аномальные наблюдения необходимо исключить из временного ряда, заменив их истинными или расчетными значениями.

Иногда аномальные значения могут отражать реальное развитие процессов, например, происходящее в условиях кризисных явлений в экономике. Как правило, эти значения также заменяются расчетными в процессе анализа и прогнозирования, но учитываются при расчете возможной величины отклонений фактических значений от полученных по модели.

Временные ряды **не должны иметь пропущенные наблюдения**. Пропуски могут объясняться как недостатками при сборе информации, так и происшедшими изменениями в системе отчетности или фиксирования данных. В этом случае для использования временного ряда в дальнейшем анализе необходимо восстановить пропущенные уровни одним из известных способов восстановления пропусков в зависимости от специфики конкретного временного ряда.

Если в систему показателей включен новый признак, учет которого не проводился ранее, то необходимо подождать, пока ряд достигнет требуемой

длины или попытаться восстановить прежние значения косвенными методами через другие показатели, если такой путь представляется возможным¹.

Для применения определенных экономико-математических методов и получения достоверных прогнозных расчетов необходимо, чтобы временной ряд имел достаточную длину, то есть информация была полной. Требование **полноты данных** обусловлено тем, что закономерность может обнаружиться лишь при наличии минимально допустимого объема наблюдений. Например, для использования регрессионного анализа требуется подготовить временные ряды, длина которых в несколько раз превосходит количество независимых переменных. При изучении сезонных колебаний на базе месячных или квартальных данных рекомендуется иметь информацию не менее чем за 3 года.

Устойчивость в изменении уровней ряда характеризуется преобладанием закономерности над случайностью. На графиках устойчивых временных рядов закономерность прослеживается визуально, на графиках неустойчивых рядов изменения последовательных уровней представляются хаотичными, и поэтому поиск закономерностей в формировании значений уровней таких рядов лишен смысла.

Следует иметь в виду, что при прогнозировании временных рядов социально-экономических показателей проверка выполнимости перечисленных требований проводится на этапах сбора необходимой информации и первичной обработки исходных данных. После этого переходят к расчету и анализу основных показателей динамики развития, построению моделей прогнозирования, получению прогнозных оценок.

¹ См. подр.: Статистические методы прогнозирования в экономике: учебное пособие, практикум, тесты, программа курса / Дуброва Т.А.; руководство по изучению дисциплины / Дуброва Т.А., Архипова М.Ю. М., 2004.

2.3. Компоненты временных рядов

При исследовании динамических рядов социально-экономических показателей выделяют следующие основные компоненты, образующие уровень временного ряда (Y):

1. **Тренд (T)** – устойчивая систематически изменяющаяся долговременная составляющая. Он показывает общее направление развития, основную длительную тенденцию временного ряда.

2. **Циклическая (конъюнктурная) составляющая (C)** проявляется на протяжении длительного времени и является результатом влияния факторов, обладающих большим последствием, либо циклически изменяющихся со временем. Характерным примером такой компоненты, влияющей на показатели деятельности производственных подразделений УИС, являются циклы деловой активности в стране. Циклическое изменение не всегда является периодическим.

3. **Сезонная составляющая (S)** имеет место, если период колебаний не превышает одного года. Часто причиной их возникновения являются природно-климатические факторы. Например, увеличение платежей за электроэнергию, используемую для нужд учреждений УИС в зимний период года, и др.

4. **Нерегулярная составляющая (U)** не поддается учету и регистрации, образована в результате действия большого числа внешних факторов, не участвующих в формировании детерминированной составляющей. Нерегулярная компонента может быть определена путем удаления из временного ряда тренда, циклической и сезонной составляющей. Среди факторов, влияющих на формирование нерегулярной компоненты, можно выделить факторы резкого, внезапного действия (стихийные бедствия, эпидемии и др.), которые вызывают большие отклонения по сравнению со случайными колебаниями (катастрофические колебания), а также текущие факторы, которые приводят к случайным колебаниям, являющимся результатом действия большого числа побочных причин.

Влияние каждого из текущих факторов незначительно, но ощущается их суммарное воздействие.

Предметом анализа временного ряда является выделение и изучение указанных компонент ряда. В тех или иных рядах некоторые составляющие могут отсутствовать.

Если временной ряд представляется в виде суммы соответствующих компонент ($Y=T+C+S+U$), то полученная модель называется **аддитивной**, если в виде произведения ($Y=T \cdot C \cdot S \cdot U$) - **мультипликативной** или смешанного типа ($Y=T \cdot C \cdot S+U$).

В результате анализа временного ряда необходимо определить, какие из случайных составляющих присутствуют в разложении ряда, построить для них адекватные оценки, подобрать модель, описывающую поведение остатков и оценить ее параметры.

Контрольные вопросы

1. Что понимается под временным рядом?
2. Назовите виды временных рядов.
3. В чем заключается отличие моментных временных рядов от интервальных?
4. Приведите примеры моментных, интервальных и производных временных рядов социально-экономических показателей деятельности исправительных учреждений.
5. Перечислите требования, предъявляемые к временным рядам.
6. Каким образом достигается сопоставимость уровней динамического ряда?
7. Назовите и охарактеризуйте компоненты, выделяемые при исследовании динамических рядов социально-экономических показателей?

3. ИСПОЛЬЗОВАНИЕ ОСНОВНЫХ МЕТОДОВ ЭКСТРАПОЛЯЦИИ ДЛЯ ПРОГНОЗИРОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ В УГОЛОВНО-ИСПОЛНИТЕЛЬНОЙ СИСТЕМЕ

3.1. Простая экстраполяция

Наиболее распространенными являются методы статистического прогнозирования, относящиеся к **экстраполяции**, сущность которой заключается в анализе изменений объектов в прошлом и продлении выявленных тенденций на будущее. Такие методы могут применяться при краткосрочном и среднесрочном прогнозировании.

Экстраполяция базируется на следующих допущениях:

- развитие явления может быть с достаточным основанием охарактеризовано плавной траекторией - **трендом**;

- общие условия, определяющие тенденцию изменения социально-экономических показателей в прошлом, не претерпят существенных изменений в будущем;

- отклонения фактических показателей от линии тренда носят случайный характер и распределяются по нормальному закону¹.

Простая экстраполяция может быть проведена на основе средних характеристик ряда: среднего уровня, среднего абсолютного прироста и среднего темпа роста.

Введем следующие обозначения:

$y_1, y_2, \dots, y_i, \dots, y_n$ - уровни временного ряда, зафиксированные в соответствующие моменты (периоды) времени $t=1, 2, \dots, i, \dots, n$;

n - длина временного ряда;

Если средний уровень ряда не имеет тенденции к изменению или, если это изменение незначительно, то рассчитывается простое среднеарифметическое значение показателя, которое закладывается в основу краткосрочного прогноза:

¹ См. подр.: *Личко К.П.* Прогнозирование и планирование развития агропромышленного комплекса: учебник. – М., 2007.

$$\hat{y}_{n+1} = \frac{\sum_{i=1}^n y_i}{n}, \quad (1)$$

где \hat{y}_{n+1} - прогнозная оценка значения уровня в точке $n+1$.

Абсолютный прирост (Δy) равен разности двух сравниваемых уровней.

Описание динамики ряда с помощью среднего абсолютного прироста соответствует его представлению в виде прямой, проведенной через две крайние точки. Величина среднего абсолютного прироста ($\bar{\Delta y}$) рассчитывается по формуле:

$$\bar{\Delta y} = \frac{y_n - y_1}{n - 1}. \quad (2)$$

В этом случае прогноз на один шаг вперед вычисляется с помощью арифметической прогрессии, то есть достаточно к последнему наблюдению добавить значение среднего абсолютного прироста:

$$\hat{y}_{n+1} = y_n + \bar{\Delta y}. \quad (3)$$

Указанный подход к получению прогнозного значения корректен, если характер развития близок к линейному.

Темп роста (T) характеризует отношение двух сравниваемых уровней ряда, выраженное в процентах. Применение среднего темпа роста для описания динамики ряда соответствует его представлению в виде показательной или экспоненциальной кривой, проведенной через две крайние точки. Поэтому использование данного показателя в качестве обобщающего целесообразно для тех процессов, изменение динамики которых происходит примерно с постоянным средним темпом роста (\bar{T}), определяемым по формуле:

$$\bar{T} = \sqrt[n-1]{\frac{y_n}{y_1}}. \quad (4)$$

В этом случае прогнозное значение на i шагов вперед может быть получено по формуле:

$$\hat{y}_{n+i} = y_n \cdot \bar{T}^i, \quad (5)$$

где \hat{y}_{n+i} - прогнозная оценка значения уровня в точке $n+i$.

В данном случае предполагается развитие по геометрической прогрессии или по экспоненте. Во всех случаях следует определять доверительный интервал, учитывающий неопределенность и погрешность используемых оценок.

К недостаткам среднего прироста и среднего темпа роста следует отнести то, что они учитывают лишь конечный и начальный уровни ряда, исключают влияния промежуточных уровней. Тем не менее, эти показатели имеют весьма широкую область применения, что объясняется чрезвычайной простотой их вычисления. Они могут быть использованы как приближенные, простейшие способы прогнозирования, предшествующие более глубокому количественному и качественному анализу.

Пример 1.

Рассчитать, используя простую экстраполяцию на основе среднего уровня, среднего абсолютного прироста и среднего темпа роста, прогноз среднесписочной численности осужденных, не подлежащих обязательному привлечению к труду по объективным причинам, используя следующие исходные данные (рис. 3).



Рис. 3. Динамика среднесписочной численности осужденных, не подлежащих обязательному привлечению к труду по объективным причинам, тыс. чел.

Решение.

1. Рассчитаем прогнозное значение показателя на следующий год (y_8) как простое среднее арифметическое трех последних уровней временного ряда, используя формулу (1):

$$y_8 = \frac{474,8 + 320,8 + 369,3}{3} = 388,3 \text{ тыс. чел.}$$

2. Вычислим прогноз на один шаг вперед, исходя из величины среднего абсолютного прироста, по формулам (2), (3):

$$\bar{\Delta y} = \frac{369,3 - 296,5}{7 - 1} = 12,1 ;$$

$$y_8 = 369,3 + 12,1 = 381,4 \text{ тыс. чел.}$$

3. Определим средний темп роста для прогнозной оценки рассматриваемого показателя, используя формулу (4):

$$\bar{T} = \sqrt[7-1]{\frac{369,3}{296,5}} = 1,037$$

В этом случае прогнозное значение на будущий год в соответствии с формулой (5) составит:

$$y_8 = 369,3 \cdot 1,037 = 383,0 \text{ тыс. чел.}$$

Таким образом, проведенные на основе простой экстраполяции расчеты показывают возможный рост среднесписочной численности осужденных, не подлежащих обязательному привлечению к труду по объективным причинам, в прогнозируемом периоде в интервале от 381,4 до 388,3 тыс. чел.

3.2. Оценка точности моделей и прогнозов, полученных методами экстраполяции

Как правило, точность прогноза характеризуется значением **ошибки прогноза**, под которой понимается величина, показывающая расхождения между прогнозным значением показателя и его фактическим значением.

Рассмотрим основные показатели, используемые для оценки точности прогнозов, полученных статистическими методами.

Абсолютная ошибка прогноза рассчитывается по формуле:

$$\Delta_t = \hat{y}_t - y_t, \quad (6)$$

где y_t - фактическое значение показателя;

\hat{y}_t - прогнозируемое значение.

Абсолютная ошибка измеряется в тех же единицах, что и прогнозируемый показатель и поэтому не очень удобна для оценки точности прогнозов.

Наиболее показательна **относительная ошибка** прогноза, исчисленная в процентах относительно фактического значения показателя:

$$s_t = \frac{\hat{y}_t - y_t}{y_t} \cdot 100. \quad (7)$$

На практике также широко используются средние ошибки, рассчитываемые по модулю. Так, **средняя абсолютная ошибка** вычисляется по формуле:

$$\bar{\Delta} = \frac{\sum_{t=1}^n |\hat{y}_t - y_t|}{n}, \quad (8)$$

где n - число уровней временного ряда, для которых определялось прогнозное значение.

Средняя относительная ошибка в процентах рассчитывается следующим образом:

$$\bar{s} = \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right| \cdot 100. \quad (9)$$

Чем ближе к нулю значения представленных выше показателей, тем выше точность полученного прогноза. Если абсолютная и относительная ошибка больше нуля, то это свидетельствует о завышенной прогнозной оценке, а если меньше нуля, то прогноз был занижен. Эти характеристики могут быть вычислены после того, как период упреждения уже закончился, и имеются данные о прогнозируемом показателе.

Значения средней относительной ошибки можно интерпретировать следующим образом:

$\bar{s} < 10\%$ - точность полученного прогноза высокая;

$10 \leq \bar{s} < 20\%$ - хорошая точность;

$20 \leq \bar{s} < 50\%$ - удовлетворительная точность;

$\bar{s} \geq 50\%$ - неудовлетворительная точность прогноза.

Ошибки прогнозов, полученные ретроспективно характеризуют точность применяемой модели. При проведении сравнительной оценки моделей прогнозирования применяются показатели **дисперсии** (δ^2) и **среднеквадратического отклонения** (δ):

$$\delta^2 = \frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}; \quad (10)$$

$$\delta = \sqrt{\delta^2}. \quad (11)$$

Дисперсия показывает меру рассеивания случайной величины, то есть ее отклонения от среднего значения и имеет единицы измерения исходных уровней ряда, возведенные в квадрат. Показатель среднеквадратического отклонения измеряется в тех же единицах, что и исходные данные и поэтому является более удобным для статистической оценки меры разброса величины от среднего уровня.

Следует иметь ввиду, что чем меньше значение дисперсии и среднеквадратического отклонения, тем выше точность модели.

О точности модели нельзя судить по одному значению ошибки прогноза, поскольку единичный хороший прогноз может быть получен и по плохой модели. Поэтому о качестве применяемых моделей можно судить только по совокупности сопоставлений прогнозных значений с фактическими.

3.3. Метод скользящих средних

Достаточно часто временные ряды характеризуются резкими колебаниями уровней. В таких рядах, как правило, сложно обнаружить четкие тенденции изменения. Поэтому применяется различные методы **сглаживания динамиче-**

ских рядов. Их суть заключается в замене фактических уровней временного ряда расчетными уровнями, в которых погашаются колебания, что позволяет более четко выявить тенденции развития. При этом сглаживание может применяться как предварительный этап обработки исходных данных перед использованием других методов выделения тенденции.

Наиболее простым и известным методом сглаживания является метод **скользящих средних**, осуществляющий механическое выравнивание временного ряда. Скользящие средние позволяют сгладить как случайные, так и периодические колебания, выявить имеющуюся тенденцию в развитии процесса, и поэтому, являются важным инструментом при фильтрации компонент временного ряда.

Используя скользящие средние можно выравнивать ряды динамики, имеющие тесную и слабую связь со временем. При этом появляется возможность определить среднее прогнозное значение для будущего периода.

Метод скользящих средних позволяет отвлечься от случайных колебаний временного ряда с помощью замены значений в рамках выбранного интервала средней арифметической величиной. Данный интервал последовательно сдвигается на одно наблюдение. При этом величина интервала сглаживания (g) может принимать любое значение - от минимального ($g=2$) до максимального ($g=n-1$). Сглаженный ряд короче исходного на $g-1$ наблюдение.

В случае если изменение уровней временного ряда носит линейный характер, возможно применение простой скользящей средней. Алгоритм ее использования включает следующие этапы¹:

1. Определяют длину интервала сглаживания g , включающего в себя g последовательных уровней ряда ($g < n$). При этом надо иметь в виду, что чем шире интервал сглаживания, тем в большей степени погашаются колебания, и тенденция развития носит более плавный, сглаженный характер. Поэтому при наличии сильных колебаний, интервал сглаживания должен быть шире.

¹ См. подр.: Статистические методы прогнозирования в экономике: учебное пособие, практикум, тесты, программа курса / Дуброва Т.А.; руководство по изучению дисциплины / Дуброва Т.А., Архипова М.Ю. М., 2004. С. 20.

2. Разбивают весь период наблюдений на участки, при этом интервал сглаживания скользит по ряду с шагом, равным 1.

3. Рассчитывают арифметические средние из уровней ряда, образующих каждый участок. При этом наблюдения, которые берутся для расчета среднего значения, называются **активным интервалом сглаживания**.

4. Заменяют фактические значения ряда, стоящие в центре каждого участка, на соответствующие средние значения.

Для любого интервала скользящая средняя рассчитывается по формуле:

$$\hat{y}_k^{(g)} = \frac{\sum_{i=k}^{g+k-1} y_i}{g}, \quad (12)$$

где y_i – фактическое значение i -го уровня ряда, $i=1, 2, \dots, n$;

$\hat{y}_k^{(g)}$ - k -я скользящая средняя при интервале g , $k=1, 2, \dots, m-(g-1)$.

Например, для $g=5$ первая скользящая средняя будет равна

$$\hat{y}_1^{(5)} = \frac{\sum_{i=1}^5 y_i}{g} = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}.$$

Чтобы полученные значения скользящей средней приходились на средний член интервала, рекомендуется брать длину интервала сглаживания (g) в виде нечетного числа: $g=2p+1$.

При нечетном значении g все уровни активного участка могут быть представлены в виде: $y_{t-p}, y_{t-p+1}, \dots, y_{t-1}, y_t, y_{t+1}, \dots, y_{t+p-1}, y_{t+p}$,

где y_t - центральный уровень активного участка;

$y_{t-p}, y_{t-p+1}, \dots, y_{t-1}$ - последовательность из p уровней активного участка, предшествующих центральному;

$y_{t+1}, \dots, y_{t+p-1}, y_{t+p}$ - последовательность из p уровней активного участка, следующих за центральным.

Для нечетного интервала скользящая средняя исчисляется по формуле:

$$\hat{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + \dots + y_{t+p-1} + y_{t+p}}{2p+1}, \quad (13)$$

где \hat{y}_t - значение скользящей средней в момент времени t ;

$2p+1$ – длина интервала сглаживания.

Для полного погашения периодических колебаний во временном ряду длина интервала сглаживания берется равной или кратной периоду колебаний. Соответственно для устранения сезонных колебаний целесообразно использовать четырехчленную и двенадцатичленную скользящие средние, но при этом не будет выполняться условие нечетности длины интервала сглаживания. Поэтому для расчета скользящего среднего с четным интервалом принято первое и последнее наблюдение на активном участке рекомендуется брать с половинными весами¹. Например, для расчета скользящих средних, обеспечивающих сглаживание сезонных колебаний при работе с временными рядами квартальной или месячной динамики рекомендуется применять следующие формулы:

$$\hat{y}_t = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}; \quad (14)$$

$$\hat{y}_t = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \dots + y_t + \dots + y_{t+5} + \frac{1}{2}y_{t+6}}{12}. \quad (15)$$

Метод скользящего среднего используется не только для сглаживания временного ряда, но и для краткосрочного прогнозирования. В последнем случае при $g=3$ используется следующая формула:

$$\hat{y}_{t+1} = \hat{y}_{t-1} + \frac{1}{m} \cdot (y_t - y_{t-1}), \quad (16)$$

где \hat{y}_{t-1} – скользящая средняя за два периода до прогнозного;

y_t – фактическое значение исследуемого явления за период, предшествующий прогнозному периоду;

y_{t-1} – фактическое значение исследуемого явления за два периода, предшествующих прогнозному.

Метод простой скользящей средней достаточно эффективен, если графическое изображение динамического ряда является близким по характеру к пря-

¹ См. подр.: Дуброва Т.А. Статистические методы прогнозирования: учеб. пособие для вузов. М., 2003.

мой линии. Для процессов с нелинейным развитием простая скользящая средняя может привести к существенным искажениям. В этих случаях более надежным является использование взвешенной скользящей средней

При определении простого скользящего среднего все уровни, как прошлых периодов, так и текущие, имеют одинаковый вес. Для целей прогнозирования более важными являются текущие уровни, поэтому им необходимо присваивать больший вес. Метод **взвешенного скользящего среднего** позволяет присваивать любой произвольный вес (w_i) каждому уровню ряда при условии, что сумма всех весов равна 1. При построении взвешенной скользящей средней на каждом участке сглаживания значение центрального уровня заменяется на расчетное, определяемое по формуле средней арифметической взвешенной:

$$\hat{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i \cdot w_i}{\sum_{i=t-p}^{t+p} w_i} \quad (17)$$

Взвешенная скользящая средняя приписывает каждому уровню вес, зависящий от его удаления до уровня, стоящего в середине активного участка. При этом схема присвоения веса может любой, например, некоторые уровни можно игнорировать (их вес принять равным нулю). Как правило, близкое прошлое служит наиболее важным индикатором будущего, а значит, этому периоду времени присваивают более высокий вес. Однако если данные имеют сезонные колебания, то их следует учитывать при установлении весовых коэффициентов. Преимущество метода взвешенного скользящего среднего перед простым заключается в возможности влияния на прогноз, изменяя результаты прошлых периодов.

Пример 2.

Имеются данные о численности осужденных, занятых на собственном производстве учреждений УИС и контрагентских производственных объектах (табл. 5). Рассчитайте скользящие средние значения с интервалом сглаживания $g=3$. Составьте прогноз на 2013 г. Постройте и проанализируйте график факти-

ческих и расчетных показателей с точки зрения применимости метода простого скользящего среднего. Определите ошибку полученного прогноза.

Таблица 5

Динамика численности осужденных, занятых на собственном производстве учреждений УИС и контрагентских производственных объектах¹

Год	2005	2006	2007	2008	2009	2010	2011	2012
Порядковый номер года, t	1	2	3	4	5	6	7	8
y_t , тыс. чел.	182,9	179,1	182,5	181,4	149,4	156,3	158,4	158,9

Решение.

1. Так как интервал сглаживания $g=3$, рассчитать первое значение скользящей средней можно только для второго временного периода. Она будет равна:

$$\hat{y}_2 = \frac{y_1 + y_2 + y_3}{3} = \frac{182,9 + 179,1 + 182,5}{3} = 181,5$$

Определим остальные значения скользящей средней:

$$\hat{y}_3 = \frac{y_2 + y_3 + y_4}{3} = \frac{179,1 + 182,5 + 181,4}{3} = 181,0;$$

$$\hat{y}_4 = \frac{y_3 + y_4 + y_5}{3} = \frac{182,5 + 181,4 + 149,4}{3} = 171,1;$$

$$\hat{y}_5 = \frac{y_4 + y_5 + y_6}{3} = \frac{181,4 + 149,4 + 156,3}{3} = 162,4;$$

$$\hat{y}_6 = \frac{y_5 + y_6 + y_7}{3} = \frac{149,4 + 156,3 + 158,4}{3} = 154,7;$$

$$\hat{y}_7 = \frac{y_6 + y_7 + y_8}{3} = \frac{156,3 + 158,4 + 158,9}{3} = 157,9.$$

2. Результаты расчетов оформим в табл. 6.

Таблица 6

Результаты расчётов для модели скользящего среднего

¹ См.: Показатели, характеризующие производственный сектор УИС// <http://фсин.рф/structure/adaptation/soveshchanie-chelyabinsk-28-05-2013/index.php>

Годы	t	Численность занятых осужденных, (y_t), тыс. чел.	Скользящая средняя \hat{y} , тыс. чел.	$(y - \hat{y})^2$	$\left \frac{y - \hat{y}}{y} \right $
2005	1	182,9	-	-	-
2006	2	179,1	181,5	5,76	0,013
2007	3	182,5	181,0	2,25	0,008
2008	4	181,4	171,1	106,09	0,057
2009	5	149,4	162,4	168,13	0,087
2010	6	156,3	154,7	2,56	0,010
2011	7	158,4	157,9	0,284	0,003
2012	8	158,9	-	-	-
Итого	-	-	-	285,08	0,179

3. Рассчитав скользящую среднюю для всех периодов, определим прогнозное значение на 2013 год по формуле (16):

$$y_9 = \hat{y}_7 + \frac{1}{g}(y_8 - y_7) = 157,9 + \frac{1}{3}(158,9 - 158,4) = 158,1 \text{ тыс. чел.}$$

4. Для оценки точности прогноза, на основе расчетных данных двух последних столбцов табл. 6, определим:

- среднеквадратическое отклонение:

$$\sigma = \sqrt{\frac{\sum (y - \hat{y})^2}{n}} = \frac{285,08}{6} = 6,89;$$

- среднюю относительную ошибку прогноза:

$$\bar{s} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% = \frac{1}{6} \cdot 0,179 \cdot 100 = 3,0\% < 10\%.$$

Значение средней относительную ошибку прогноза меньше 10%, что указывает на приемлемую точность прогноза. Поэтому в качестве прогнозной оценки численности осужденных, занятых на собственном производстве учреждений УИС и контрагентских производственных объектах на 2013 г. можно принять значение \hat{y}_9 , равное 158,1 тыс. чел.

5. Построим график по фактическим и выровненным с помощью скользящей средней уровням ряда (рис. 4).

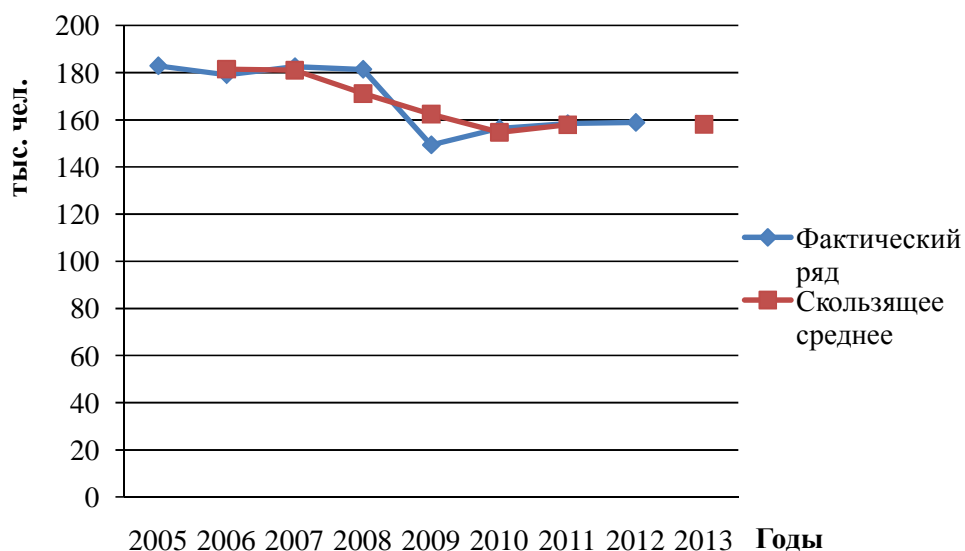


Рис. 4. Фактические и выровненные с помощью скользящей средней уровни ряда

Из графика видно, что в последние три года наблюдается явная тенденция к возрастанию значений временного ряда y_t , что указывает на невысокую точность прогноза, выполненного методом скользящего среднего (это следует из допущений метода), к подавлению этой тенденции.

3.4. Метод экспоненциального сглаживания

Часто на практике между достигнутым в текущем году значением уровня временного ряда и его значениями в предыдущие годы имеет место определенная взаимосвязь. При этом результаты более поздних наблюдений несут большую информацию об ожидаемом значении показателя. Следовательно, последним показателям временных рядов в процессе прогнозирования должно придаваться большее значение. В этом случае лучшей прогнозной оценкой на следующий шаг наблюдения выступает взвешенная средняя, причём веса наблюдения должны уменьшаться с убыванием наблюдений в прошлое, а сумма весов

должна быть равна единице. Такой подход лежит в основе метода **экспоненциального сглаживания** (модели Брауна), применяемого для выравнивания значительно колеблющихся временных рядов с целью последующего прогнозирования.

Р. Браун предложил задавать веса исходя из ряда бесконечной геометрической прогрессии. В результате подстановок и сокращений им была получена удобная формула для расчета S_t – сглаженного значения для t -го уровня ряда:

$$S_t = \alpha \cdot y_t + (1 - \alpha) \cdot S_{t-1}, \quad (18)$$

где S_t – значение экспоненциальной средней в момент времени t ;

α – коэффициент (параметр) сглаживания $0 \leq \alpha \leq 1$.

Для расчета прогноза временного ряда методом экспоненциального сглаживания используют формулу (18), представленную в виде:

$$S_{t+1} = \alpha \cdot y_t + (1 - \alpha) \cdot S_t. \quad (19)$$

Доказано, что дисперсия экспоненциальной средней $D[S_t]$ меньше дисперсии временного ряда (δ^2). При этом между ними существует следующее соотношение¹:

$$D[S_t] = \frac{\alpha}{2 - \alpha} \delta^2. \quad (20)$$

При высоком значении параметра сглаживания (α) дисперсия экспоненциальной средней незначительно отличается от дисперсии временного ряда, то есть в большей степени учитываются мгновенные текущие наблюдения отклика. С уменьшением данного параметра, дисперсия экспоненциальной средней уменьшается, и возрастает ее отличие от дисперсии временного ряда. Тем самым экспоненциальная средняя начинает выравнивать колебания временного ряда. Сглаженная величина в этом случае определяется в большей степени прошлой тенденцией развития, чем текущим состоянием отклика системы. Таким образом, с одной стороны следует увеличивать параметр сглаживания, а с другой стороны, для сглаживания случайных отклонений, его следует умень-

¹ См. подр.: Дуброва Т.А. Прогнозирование социально-экономических процессов: учебное пособие. М., 2010.

шать. Эти два требования находятся в противоречии, что требует поиска оптимального значения коэффициента сглаживания.

На практике выбор параметра сглаживания представляет собой субъективный процесс. При этом обычно ограничиваются интервалом от 0,2 до 0,5. Примерное значение коэффициента сглаживания можно также определить по формуле Р. Брауна:

$$\alpha = \frac{2}{n+1}, \quad (21)$$

где n – число уровней временного ряда, входящих в интервал сглаживания.

Достаточно часто поиск значения коэффициента сглаживания осуществляется путем перебора. Возможные значения параметра разбиваются сеткой с определенным шагом. Например, рассматривается сетка значений от $\alpha = 0,1$ до $\alpha = 0,9$, с шагом 0,1. Затем выбирается значение α , для которого сумма квадратов (или средних квадратов) остатков (наблюдаемые значения минус прогнозы на шаг вперед) является минимальной.

Процедура прогнозирования временного ряда по методу экспоненциального сглаживания состоит из следующих этапов:

1. Выбирается вид модели экспоненциального сглаживания, задается значение параметра адаптации;
2. Определяются начальное значение S_0 . В конкретных задачах его берут равным значению первого уровня ряда y_1 , или равным среднему арифметическому нескольких первых членов ряда. Вес, приписываемый этому значению, уменьшается по экспоненциальной зависимости по мере удаления от первого уровня, поэтому для длинных временных рядов влияние неудачного выбора начального значения S_0 погашается;
3. Находятся оценки коэффициента модели;
4. Осуществляется прогноз на одну точку вперед, определяются отклонения фактического значения временного ряда от прогнозируемых.

Пример 3.

Рассчитайте экспоненциальную среднюю и прогноз на один шаг вперед для временного ряда вакантных должностей сотрудников ФСИН (в процентах от общего числа должностей) с ежемесячными значениями уровня. В качестве начального значения экспоненциальной средней возьмите среднее трех первых уровней ряда. Значение первого коэффициента сглаживания примите равным 0,3, второго - 0,7. Оцените статистическую значимость полученного прогноза.

Таблица 7

Динамика удельного веса вакантных должностей сотрудников ФСИН, %

Номер месяца, t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	3,7	4	3,9	4,5	4,5	4,7	5	5,2	5,2	5,9	6	5,6

Решение.

1. Рассчитаем начальное значение экспоненциальной средней:

$$S_0 = \frac{1}{5} \sum_{t=1}^5 y_t = \frac{1}{5} \cdot (3,7 + 4 + 3,9 + 4,5 + 4,5) = 4,12$$

2. Вычислим значения экспоненциальной средней по формуле (18) при $\alpha = 0,3$:

$$S_1 = \alpha \cdot y_1 + (1 - \alpha) \cdot S_0; S_1 = 0,3 \cdot 3,7 + (1 - 0,3) \cdot 4,12 = 3,99;$$

$$S_2 = \alpha \cdot y_2 + (1 - \alpha) \cdot S_1; S_2 = 0,3 \cdot 4 + (1 - 0,3) \cdot 3,99 = 4;$$

$$S_3 = \alpha \cdot y_3 + (1 - \alpha) \cdot S_2; S_3 = 0,3 \cdot 3,9 + (1 - 0,3) \cdot 4 = 3,97;$$

$$S_4 = \alpha \cdot y_4 + (1 - \alpha) \cdot S_3; S_4 = 0,3 \cdot 4,5 + (1 - 0,3) \cdot 3,97 = 4,13 \text{ и т.д.}$$

Полученные результаты расчетов приведены в табл. 8.

Прогнозное значение удельного веса вакантных должностей сотрудников ФСИН на первый месяц следующего года вычислим по формуле (19):

$$S_{13} = 0,3 \cdot 5,6 + (1 - 0,3) \cdot 5,48 = 5,52\%$$

Для точности прогноза определим:

- дисперсию:

$$\sigma = \frac{\sum (y - \hat{y})^2}{12} = \frac{1,723}{12} = 0,14;$$

- среднюю относительную ошибку прогноза:

$$\bar{s} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% = \frac{1}{12} \cdot 0,775 \cdot 100 = 6,5\% < 10\%.$$

В целом точность полученного прогноза оценивается как высокая.

Таблица 8

Результаты расчётов для экспоненциальной модели при $\alpha = 0,3$

Номер месяца, (t)	Удельный вес вакантных должностей сотрудников ФСИН, % (y_t)	Экспоненциально сглаженные уровни ряда, (S_t)	$(y - \hat{y})^2$	$\left \frac{y - \hat{y}}{y} \right $
1	3,7	3,99	0,086	0,079
2	4	4,00	0,000	0,001
3	3,9	3,97	0,004	0,017
4	4,5	4,13	0,139	0,083
5	4,5	4,24	0,068	0,058
6	4,7	4,38	0,104	0,069
7	5	4,56	0,190	0,087
8	5,2	4,75	0,198	0,086
9	5,2	4,89	0,097	0,060
10	5,9	5,19	0,501	0,120
11	6	5,43	0,320	0,094
12	5,6	5,48	0,013	0,021
Итого	-	-	1,723	0,775

3. Аналогично проведем расчеты экспоненциальной средней при $\alpha = 0,7$:

$$S_1 = \alpha \cdot y_1 + (1 - \alpha) \cdot S_0; S_1 = 0,7 \cdot 3,7 + (1 - 0,7) \cdot 4,12 = 3,83;$$

$$S_2 = \alpha \cdot y_2 + (1 - \alpha) \cdot S_1; S_2 = 0,7 \cdot 4 + (1 - 0,7) \cdot 3,83 = 3,95 \text{ и т.д.}$$

Полученные значения представлены в табл. 9.

Прогнозируемый удельный вес вакантных должностей сотрудников ФСИН на первый месяц следующего года определим на основе формулы (19):

$$S_{13} = 0,7 \cdot 5,6 + (1 - 0,7) \cdot 5,69 = 5,63\%$$

Для оценки точности прогноза вычислим:

- дисперсию:

$$\sigma = \frac{\sum (y - \hat{y})^2}{12} = \frac{0,146}{12} = 0,012;$$

- среднюю ошибку аппроксимации:

$$\bar{s} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% = \frac{1}{12} \cdot 0,233 \cdot 100 = 1,9\% < 10\%.$$

Следовательно, точность прогноза, полученного с использованием коэффициента сглаживания $\alpha = 0,7$, также является высокой.

Таблица 9

Результаты расчётов для экспоненциальной модели при $\alpha = 0,7$

Номер месяца, (t)	Удельный вес вакантных должностей сотрудников ФСИН, % (y_t)	Экспоненциально сглаженные уровни ряда, (S_t)	$(y - \hat{y})^2$	$\left \frac{y - \hat{y}}{y} \right $
1	3,7	3,83	0,016	0,034
2	4	3,95	0,003	0,013
3	3,9	3,91	0,000	0,004
4	4,5	4,32	0,031	0,039
5	4,5	4,45	0,003	0,012
6	4,7	4,62	0,006	0,016
7	5	4,89	0,013	0,023
8	5,2	5,11	0,009	0,018
9	5,2	5,17	0,001	0,005
10	5,9	5,68	0,048	0,037
11	6	5,90	0,009	0,016
12	5,6	5,69	0,008	0,016
Итого	-	-	0,146	0,233

4. Представим влияние значения параметра адаптации α на характер сглаженного ряда (рис. 5).

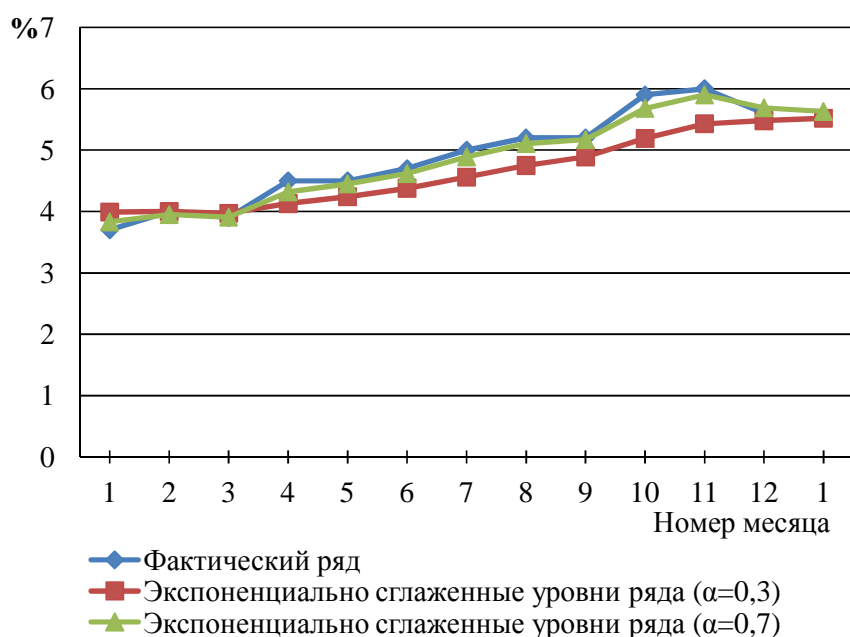


Рис. 5. Фактические и выровненные с помощью экспоненциального сглаживания уровни ряда удельного веса вакантных должностей сотрудников ФСИН

При $\alpha = 0,3$ экспоненциальная средняя носит более гладкий характер, так как в этом случае в наибольшей степени поглощаются случайные колебания временного ряда.

Контрольные вопросы

1. В чем заключается сущность прогнозной экстраполяции и допущения, на которых она базируется?
2. Охарактеризуйте метод простой экстраполяции.
3. Какие показатели используются для оценки точности моделей и прогнозов, полученных методами экстраполяции?
4. Как следует интерпретировать полученные значения средней относительной ошибки, дисперсии и среднеквадратического отклонения?
5. Раскройте сущность метода скользящих средних.
6. Опишите алгоритм использования простой скользящей средней для прогнозирования.
7. Укажите особенности построения взвешенного скользящего среднего.
8. Какой подход лежит в основе метода экспоненциального сглаживания?
9. В чем заключаются особенности выбора коэффициента (параметра) сглаживания для модели экспоненциального сглаживания?
10. Опишите процедуру прогнозирования временного ряда с использованием метода экспоненциального сглаживания.

4. КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ В СОЦИАЛЬНО-ЭКОНОМИЧЕСКОМ ПРОГНОЗИРОВАНИИ

4.1. Основные понятия корреляционно-регрессионного анализа

Для изучения форм связи, устанавливающих количественные соотношения между случайными величинами изучаемых социально-экономических процессов в УИС, целесообразно использовать возможности корреляционно-регрессионного анализа. В прогнозировании этот метод позволяет построить условные прогнозы и прогнозы, основанные на оценке устойчивых причинно-следственных связей, например разработать прогноз трудоустройства осужденных после их освобождения из исправительных учреждений исходя из изменения уровня затрат на профессиональную подготовку спецконтингента в производственных подразделениях УИС и профессиональных училищах.

Установление и оценка причинно-следственной связи предполагает определение зависимости изучаемой случайной величины Y от одной или нескольких других величин X с последующим прогнозированием значений Y . Параметр Y , значение которого нужно предсказать, является **результативным признаком** (зависимой или критериальной переменной). Параметр X , значения которого нам известны заранее и который влияет на значения Y , называется **факторным признаком** (независимой переменной или предиктором). Например, X – объем заказов на поставку товаров, выполнение работ оказание услуг, получаемый исправительными учреждениями, Y – численность осужденных, выводимых на оплачиваемые работы и т.д.

Регрессионный анализ – статистический метод, позволяющий определить аналитическое выражение связи между исследуемыми переменными, в котором изменение результативной переменной Y происходит под влиянием одной или нескольких факторных переменных: X_1, X_2, \dots, X_n . Модель регрессии или уравнение регрессии позволяет количественно оценить взаимосвязь между исследуемыми переменными Y и X и в самом простом виде может быть представлена следующим образом:

$$\bar{y}_x = f(x), \quad (22)$$

где \bar{y}_x – условное среднее всех возможных значений параметра Y , которые соответствуют значению $X = x$.

Функция $f(x)$ – называется **регрессией** Y на X , а ее график – **линией регрессии** Y на X .

Задача регрессионного анализа состоит в том, чтобы по данным наблюдений определить такую функцию $\bar{y}_x = f(x)$, которая наилучшим образом будет отражать исследуемую зависимость между переменными.

Основная задача **корреляционного анализа** – оценка тесноты (силы) корреляционной связи. Теснота корреляционной зависимости Y от X оценивается по величине рассеяния значений параметра Y вокруг условного среднего \bar{y}_x . Большое рассеяние говорит о слабой зависимости Y от X , либо об ее отсутствии и, наоборот, малое рассеяние указывает на наличие достаточно сильной связи.

Для проведения регрессионного анализа и прогнозирования необходимо выполнить следующие этапы:

- 1) выделить состав независимых переменных (факторных признаков);
- 2) обосновать характер возможных связей между результативным и факторными признаками. Для этого используют три основных приема:

а) **графический метод** или визуальная оценка характера связи. В этом случае на линейном графике по оси абсцисс откладываются значения факторной переменной (x), а по оси ординат – значения результативной переменной (y). Затем на пересечении соответствующих значений отмечаются точки. Полученный точечный график в системе координат (x, y) называется корреляционным полем. Линия, которая соединяет точки на графике, называется эмпирической линией. По её виду можно судить не только о наличии, но и о форме зависимости между изучаемыми переменными;

б) **теоретический и логический анализ** природы изучаемых явлений, их социально-экономической сущности;

в) определение **аналитической формы зависимости** между переменными экспериментальным путём;

3) выбрать форму уравнения регрессии, которая может описывать связь исходных данных;

4) определить численные коэффициенты функции регрессии;

5) оценить силу найденной регрессионной зависимости на основе коэффициентов корреляции и детерминации;

б) сделать прогноз или вывод о невозможности прогнозирования с помощью найденной регрессионной зависимости. Прогноз получают путем подстановки в регрессионное уравнение переменных. Результат представляет собой оценку среднего значения зависимой переменной при данных уровнях независимых переменных.

При использовании корреляционно-регрессионного анализа причинно-следственная связь между исследуемыми явлениями выступает в наглядной форме. Прогноз в этом случае лучше поддается содержательной интерпретации, чем простая экстраполяция тенденции. Во всяком случае, при проведении корреляционно-регрессионного анализа становится более ясным воздействие отдельных факторов и прогнозист лучше понимает природу исследуемого явления. Кроме того, регрессия создает базу для расчетного экспериментирования с целью получения ответов на вопросы типа «Что будет, если...?».

4.2. Классификация видов регрессии. Метод наименьших квадратов

В зависимости от количества переменных, учитываемых в регрессии, различаются:

1. **Простая** (парная) регрессия, которая предполагает исследование связи между двумя переменными. Следует отметить, что при наличии единственного аргумента времени (t) модель регрессии принимает вид трендовой зависимости;

2. **Многофакторная** (множественная) регрессия имеет место, когда уравнение регрессии включает два и более факторных признака.

Относительно формы зависимости между переменными регрессии, выделяются следующие ее виды:

1. **Линейная** регрессия с линейной зависимостью между переменными. В случае парной линейной регрессии уравнение имеет вид:

$$y = a_0 + a_1x + \varepsilon, \quad (23)$$

где x – факторный признак;

a_0, a_1 - коэффициенты регрессии;

ε - отклонение фактических значений результативного признака от выровненных.

При исследовании линейной зависимости одной переменной от нескольких объясняющих переменных x_1, x_2, \dots, x_n , уравнение регрессии принимает вид:

$$y = a_0 + a_1x + a_2x + \dots + a_mx + \varepsilon \quad (24)$$

2. **Нелинейная** регрессия с нелинейными зависимостями в уравнении регрессии. Различают два класса нелинейных регрессий. К первому классу относят регрессии, нелинейные относительно включенных в уравнение независимых переменных, но линейных по оцениваемым параметрам. Эти регрессии называются **существенно линейными** регрессиями. Преимущество таких уравнений в том, что для них остаются в силе все предпосылки классического линейного регрессионного анализа. Параметры оцениваются непосредственно обыкновенным методом наименьших квадратов.

Примером данного типа регрессий являются полиномы разных степеней ($y = a_0 + a_1x + a_2x^2 + \dots$), а также равносторонняя гипербола ($y = a_0 + a_1/x$).

Ко второму классу нелинейных регрессий относят регрессии, нелинейные по оцениваемым параметрам. Эти регрессии называются **существенно нелинейными** регрессиями. Оценить параметры обыкновенным методом наименьших квадратов невозможно, так как имеют место нелинейные уравнения относительно неизвестных параметров. Линеаризация таких уравнений осуществляется посредством логарифмирования.

Существенно нелинейными регрессиями, наиболее часто используемыми для описания социально-экономических процессов, являются следующие функции: степенная ($y = ax^b$), показательная ($y = ab^x$), экспоненциальная ($y = e^{a+bx}$).

Регрессионный анализ предполагает решение двух задач. Первая заключается в выборе независимых переменных, существенно влияющих на зависимую величину, и определения формы уравнения регрессии. Обычно этот этап в разработке регрессии называют спецификацией. Данная задача решается путем анализа изучаемой взаимосвязи по существу. Формальные средства могут служить здесь лишь некоторыми ориентирами. Вторая задача – расчет параметров регрессии – решается с помощью того или иного статистического метода обработки данных.

Рассмотрим простейшую модель парной регрессии – линейную регрессию, которая сводится к нахождению уравнения вида (23).

Построение линейной регрессии сводится к оценке ее параметров – a_0 и a_1 . При этом наиболее часто оценивание параметров регрессии достигается с помощью метода наименьших квадратов

Метод наименьших квадратов позволяет определить коэффициенты уравнения регрессии таким образом, чтобы точки, построенные по исходным данным (x_i, y_i) , лежали как можно ближе к точкам линии регрессии (25). Формально это записывается как минимизация суммы квадратов отклонений (ошибок) функции регрессии и исходных точек¹:

$$S = \sum_{i=1}^n (y_i^p - y_i)^2 \rightarrow \min, \quad (25)$$

где y_i^p – значение, вычисленное по уравнению регрессии;

$(y_i^p - y_i)$ – отклонение ε (ошибка, остаток) (рис. 6);

n – количество пар исходных данных.

¹ См. подр.: Алесинская Т.В. Экономико-математические методы и модели: учебное пособие по решению задач. Таганрог, 2002. С. 113-114.

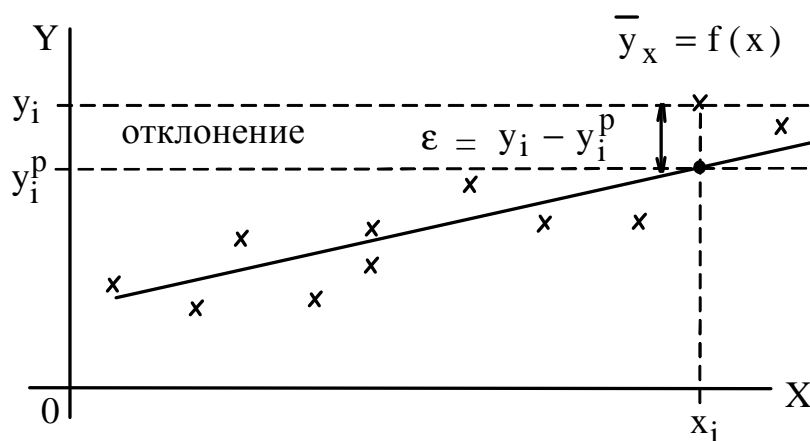


Рис. 6. Отклонение ε для случая линейной регрессии

В регрессионном анализе предполагается, что математическое ожидание случайной величины ε равно нулю и ее дисперсия одинакова для всех наблюдаемых значений Y . Отсюда следует, что рассеяние данных возле линии регрессии должно быть одинаково при всех значениях параметра X .

Исходя из метода наименьших квадратов, коэффициенты линейной регрессии вычисляются по следующим формулам (все суммы берутся по n парам исходных данных):

$$a_1 = \frac{n(\sum y_i x_i) - \sum y_i \sum x_i}{n(\sum x_i^2) - (\sum x_i)^2}; \quad (26)$$

$$a_0 = \frac{1}{n}(\sum y_i - a_1 \sum x_i). \quad (27)$$

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

Определение коэффициентов нелинейной регрессии основано на решении системы линейных уравнений. Для этого удобно применять универсальные пакеты численных методов или специализированные пакеты обработки статистических данных. В случае простых нелинейных регрессионных зависимостей используют прием их приведения к линейному виду

Рассмотрим наиболее простые случаи нелинейной регрессии: экспоненту, гиперболу и параболу. Для нахождения коэффициентов гиперболы и экспоненты будем использовать прием приведения к линейной зависимости. Это позво-

лит использовать для вычисления коэффициентов функций регрессии формулы (26), (27)¹.

С целью приведения к линейному виду экспоненциальной зависимости $y = a_0 e^{a_1 x}$, осуществляется ее логарифмирование:

$$\ln y = \ln(a_0 e^{a_1 x}); \quad (28)$$

$$\ln y = \ln a_0 + \ln(e^{a_1 x}); \quad (29)$$

$$\ln y = \ln a_0 + a_1 x. \quad (30)$$

Далее вводятся новые переменные $c_0 = \ln a_0$ и $c_1 = a_1$, что позволяет получить линейную зависимость $\ln y = c_0 + c_1 x$ и применить формулы (26), (27), в которых вместо значений y_i надо использовать $\ln y_i$

$$c_1 = \frac{n(\sum [\ln y_i] x_i) - \sum \ln y_i \sum x_i}{n(\sum x_i^2) - (\sum x_i)^2}; \quad (31)$$

$$c_0 = \frac{1}{n}(\sum \ln y_i - c_1 \sum x_i). \quad (32)$$

После указанных расчетов необходимо перейти от численных значений коэффициентов c_0 и c_1 к используемым в уравнении экспоненты коэффициентам a_0 и a_1 . Исходя из введенных обозначений и определения логарифма, получаем $a_0 = e^{c_0}$, $a_1 = c_1$.

При вычислении коэффициента детерминации экспоненты все значения параметра Y (исходные, регрессионные, среднее) необходимо заменить на их логарифмы, например, y_i^p – на $\ln(y_i^p)$ и т.д.

Для расчета параметров гиперболы $y = a_0 + \frac{a_1}{x}$ следует ввести новую переменную $q = \frac{1}{x}$. В этом случае уравнение гиперболы принимает линейный вид $y = a_0 + a_1 q$. Затем для нахождения коэффициентов линейной функции применяются формулы (26), (27) в которых вместо значений x_i используются значения

$$q_i = \frac{1}{x_i} :$$

¹ См.: Алесинская Т.В. Указ. соч. С. 120-121.

$$a_1 = \frac{n(\sum y_i q_i) - \sum y_i \sum q_i}{n(\sum q_i^2) - (\sum q_i)^2}; \quad (33)$$

$$a_0 = \frac{1}{n}(\sum y_i - a_1 \sum q_i). \quad (34)$$

Вычисление коэффициентов параболы $y = a_0 + a_1 x + a_2 x^2$ осуществляется на основе решения следующей линейной системы из трех уравнений:

$$\begin{cases} n \cdot a_0 + (\sum x_i) a_1 + (\sum x_i^2) a_2 = \sum y_i, \\ (\sum x_i) a_0 + (\sum x_i^2) a_1 + (\sum x_i^3) a_2 = \sum (y_i x_i), \\ (\sum x_i^2) a_0 + (\sum x_i^3) a_1 + (\sum x_i^4) a_2 = \sum (y_i x_i^2). \end{cases} \quad (35)$$

Используя формулы (36), (38) можно определить значение коэффициентов корреляции и детерминации непосредственно для гиперболы и параболы.

4.3. Точность и надежность модели линейной регрессии. Доверительные интервалы прогноза

Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение независимых переменных (одной или нескольких) для описания зависимой переменной.

В ходе построения регрессионной модели и ее использования в прогнозировании важны ответы на вопросы о том, велика ли теснота связи между:

- независимыми переменными;
- зависимой и указанной независимой переменной;
- зависимой переменной и совокупностью независимых переменных¹.

Желательно, чтобы все независимые переменные были сильно связаны с зависимой и слабо связаны друг с другом, что означает отсутствие мультиколлинеарности рядов факторных признаков. В связи с этим характеристика уравнения регрессии всегда дополняется показателем тесноты связи.

¹ См. подр.: Писарева О.М. Методы социально-экономического прогнозирования: учебник. М., 2003. С. 199.

Информацию о связи между параметрами можно получить при использовании выборочного **коэффициента корреляции** (коэффициента корреляции Пирсона), который можно рассчитать по следующей формуле:

$$r = \frac{\sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (36)$$

Фактическое значение коэффициента корреляции r лежит между -1 и 1 , причем можно выделить три предельных случая, а именно:

$r=1$ - присутствует сильная прямая (положительная) зависимость, изменение значений переменных однонаправленное, при росте значений одной переменной, значение другой также увеличивается;

$r=-1$ - имеет место сильная обратная (отрицательная) зависимость, изменение значений переменных разнонаправленное, при увеличении значения одной переменной, значение другой уменьшается;

$r=0$ - отсутствие связи между рассматриваемыми величинами.

В практике расчетов могут быть получены значения коэффициентов корреляции, близкие к одной из указанных выше величин. Весь спектр промежуточных значений данного коэффициента можно интерпретировать следующим образом:

$0 < r < 0,2$ ($0 < r < -0,2$) - связи практически нет;

$0,2 \leq r < 0,5$ ($-0,2 \leq r < -0,5$) - слабая прямая (обратная) зависимость;

$0,5 \leq r < 0,7$ ($-0,5 \leq r < -0,7$) - заметная прямая (обратная) зависимость;

$0,7 \leq r < 0,9$ ($-0,7 \leq r < -0,9$) - сильная прямая (обратная) зависимость;

$0,9 \leq r < 1$ ($-0,9 \leq r < -1$) - очень сильная прямая (обратная) зависимость.

Положительное значение коэффициента корреляции говорит о положительной связи между x и y , когда с ростом одной из переменных другая тоже растет. Отрицательное значение коэффициента корреляции означает, что с ростом одной из переменных другая убывает, с убыванием одной из переменной другая растет.

Считается, что если значение коэффициента корреляции $r > 0,7$, то можно строить регрессию, если $r < 0,5$, то в качестве независимых переменных для построения регрессии необходимо выбрать другие факторы или увеличить количество наблюдений.

Для анализа неслучайности обнаруженных связей рекомендуется осуществить проверку коэффициентов корреляции на значимость с помощью ***t*-критерия Стьюдента**:

$$t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad (37)$$

где $n-2$ – число степеней свободы, равное числу измерений за вычетом единицы для каждой выборки.

Вычисленное значение t_r обычно сравнивается с критическим (табличным) значением *t*-критерия Стьюдента $t_{табл}(\alpha, n-2)$ при уровне значимости $\alpha = 0,05$ (или $0,01$) (уровни доверительной вероятности $\beta = 1 - \alpha$ равны $0,95$ или $0,99$) и числе степеней свободы ($n-2$). Если $t_{факт} > t_{табл}$, то полученное значение r считается значимым и принимается гипотеза о наличии статистической связи между показателями, иначе принимается гипотеза об отсутствии связи между показателями и надо выбрать другой показатель X .

Для оценки качества подбора регрессии рассчитывается квадрат линейного коэффициента корреляции (r^2), называемый коэффициентом детерминации.

Коэффициент детерминации показывает, на сколько процентов ($r^2 \cdot 100\%$) найденная функция регрессии описывает связь между исходными значениями параметров X и Y :

$$r^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (38)$$

где $(\hat{y}_i - \bar{y})^2$ – объясненная вариация;

$(y_i - \bar{y})^2$ – общая вариация¹.

¹ См.: Алесинская Т.В. Указ. соч. С. 120-121. С. 114.

Соответственно, величина $(1-r^2) \cdot 100\%$ показывает, сколько процентов вариации параметра Y обусловлены факторами, не включенными в регрессионную модель.

Имеет место неравенство $0 < r^2 < 1$. Чем ближе значение коэффициента детерминации к 1, тем точнее модель линейной регрессии, если $r^2 > 0,8$, то модель линейной регрессии считается точной и ее можно использовать для прогнозирования, если $r^2 < 0,5$, то модель является неудовлетворительной, надо строить нелинейную регрессию или выбирать другой фактор X .

Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют **среднюю ошибку аппроксимации** (приближения), которая вычисляется по формуле:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% . \quad (39)$$

Средняя ошибка аппроксимации не должна превышать 8–10%.

Оценка статистической значимости параметров регрессии проводится с помощью ***t*-статистики Стьюдента**. При этом выдвигается и проверяется гипотеза о статистически незначимом отличии параметров от нуля.

Предварительно рассчитывается **дисперсия ошибки** по формуле:

$$D_{оcман} = \sqrt{\frac{\sum y^2 - a_0 \sum y - a_1 \sum xy}{n - 2}} . \quad (40)$$

Кроме того, по каждому из параметров линейной регрессии определяется его **стандартная ошибка** (m_{a0} и m_{a1}).

Стандартная ошибка коэффициента a_0 вычисляется по формуле:

$$m_{a0} = \frac{D_{оcман} \sqrt{\sum x^2}}{n \sqrt{x^2 - \bar{x}^2}} . \quad (41)$$

Величина стандартной ошибки совместно с *t*-распределением Стьюдента при $n - 2$ степенях свободы применяется для проверки существенности коэффициента регрессии и для расчета его доверительного интервала.

При оценке существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, то есть определяется фактическое значе-

ние t -критерия Стьюдента: $t_{a_0} = a_0 / m_{a_0}$, которое затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы $(n-2)$. Доверительный интервал для коэффициента регрессии определяется как $a_0 \pm t_{табл} \cdot m_{a_0}$. Границы доверительного интервала для коэффициента регрессии не должны содержать противоречивых результатов, например, $-1,2 \leq a_0 \leq 0,7$. Такого рода запись показывает, что истинное значение коэффициента регрессии одновременно содержит положительные и отрицательные величины и даже ноль, чего не может быть, и коэффициент признается незначимым.

Стандартная ошибка параметра регрессии a_1 определяется по формуле:

$$m_{a_1} = \frac{D_{оцман}}{\sqrt{n \cdot (\overline{x^2} - \bar{x}^2)}}. \quad (42)$$

После этого вычисляется t -критерий Стьюдента: $t_{a_1} = a_1 / m_{a_1}$, его величина сравнивается с табличным значением при $n-2$ степенях свободы.

Осуществление прогнозных расчетов по уравнению регрессии предполагает определение предсказываемого значения \hat{y}_p , которое является точечным прогнозом и рассчитывается на основе подстановки в уравнение регрессии соответствующего значения x_p . На практике обычно производится расчет стандартной ошибки для \hat{y}_p и, в дополнение к точечному прогнозу, определяются границы возможного изменения значения прогнозируемого показателя, то есть дается **интервальный прогноз**. При этом **стандартная ошибка** прогнозируемого значения может быть вычислена по формуле:

$$m_{\hat{y}_p} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} \cdot \left(1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}. \quad (43)$$

Далее производится интервальная оценка точечного прогнозируемого значения, то есть определяется **доверительный интервал прогноза**:

$$\hat{y}_p - \Delta_{\hat{y}_p} \leq \hat{y}_p \leq \hat{y}_p + \Delta_{\hat{y}_p}, \quad (44)$$

где $\Delta_{\hat{y}_p} = m_{\hat{y}_p} \cdot t_{табл}$, при этом $m_{\hat{y}_p}$ - стандартная ошибка прогнозируемого значения;

$t_{табл}$ - табличное значение t -критерия Стьюдента.

Расчет доверительных интервалов позволяет определить область, в которой следует ожидать значение прогнозируемой величины. Выход этой величины за границы интервала в силу случайных колебаний имеет незначительную вероятность – меньше, чем дополнение до единицы доверительной вероятности, то есть меньше уровня существенности.

Пример 4. Для прогнозирования объема производства товарной продукции УИС подготовлены следующие исходные данные (табл. 10).

Таблица 10

Исходные данные¹

Годы	2005	2006	2007	2008	2009	2010	2011	2012
Объем товарной продукции, произведенной УИС, млрд. руб. (Y)	18	17,8	22,6	26,2	25,2	29,4	30,3	30,7
Средняя заработная плата на 1 отработанный чел.-день, рублей (X)	50,16	64,56	84,54	110,66	141,34	146,03	165,1	171,98

Требуется:

1. Построить линейное уравнение парной регрессии, характеризующей зависимость объема товарной продукции, произведенной исправительными учреждениями от средней заработной платы осужденных на 1 отработанный человеко-день.

2. Рассчитать линейный коэффициент парной корреляции и коэффициент детерминации. Оценить статистическую значимость коэффициента корреляции.

3. Рассчитать среднюю ошибку аппроксимации.

4. Оценить статистическую значимость параметров регрессии.

¹ См.: Показатели, характеризующие производственный сектор УИС// <http://фсин.рф/structure/adaptation/soveshchanie-chelyabinsk-28-05-2013/index.php>

5. Сделать прогноз объема товарной продукции УИС в стоимостном выражении, если в качестве перспективного уровня оплаты труда осужденных принято значение 185 руб./чел.-день.

6. Рассчитать стандартную ошибку полученного точечного прогноза и провести его интервальную оценку.

Решение.

1. Построим точечный график по исходным данным в системе координат (x, y) (рис. 7).

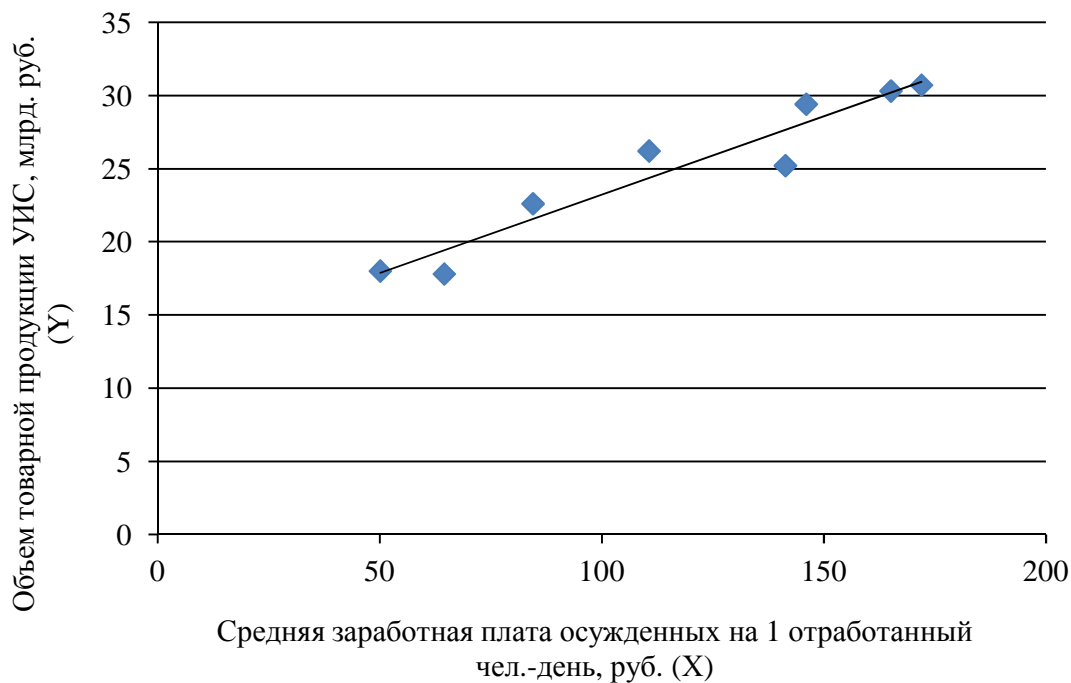


Рис. 7. График исходных данных задачи

Построенные точки не находятся точно на линии, поскольку кроме размера оплаты труда спецконтингента на объем пенитенциарного производства оказывает влияние большое число других микроэкономических и макроэкономических факторов. Но эти точки собраны вокруг прямой линии, поэтому можно предположить линейную связь между параметрами. Все исходные точки

почти равномерно распределены вдоль предполагаемой прямой линии, что позволяет применить метод наименьших квадратов.

Для расчета параметров уравнения линейной регрессии с помощью метода наименьших квадратов строим вспомогательную таблицу (табл. 11).

Таблица 11

Вспомогательная таблица

i	x_i	y_i	x_i^2	y_i^2	$x_i y_i$	\hat{y}_i	$(\hat{y}_i - \bar{y})^2$	$(y_i - \bar{y})^2$	$\left \frac{y - \hat{y}}{y} \right $
1	50,16	18,0	2516,03	324,0	902,88	17,9	51,03	49,35	0,01
2	64,56	17,8	4167,99	316,8	1149,17	19,4	31,35	52,20	0,09
3	84,54	22,6	7147,01	510,8	1910,60	21,6	11,95	5,88	0,04
4	110,66	26,2	12245,64	686,4	2899,29	24,4	0,43	1,38	0,07
5	141,34	25,2	19977	635,0	3561,77	27,7	6,93	0,03	0,10
6	146,03	29,4	21324,76	864,4	4293,28	28,2	9,83	19,14	0,04
7	165,01	30,3	27228,3	918,1	4999,80	30,2	26,74	27,83	0,00
8	171,98	30,7	29577,12	942,5	5279,79	30,9	35,02	32,21	0,01
Итого	934,28	200,2	124183,84	5198,02	24996,58	200,2	173,29	188,02	0,36
Сред. знач.	116,79	25,03	15522,98	649,75	3124,57	25,03	21,66	23,50	0,04

В соответствии с формулами (26), (27), используя данные вспомогательной таблицы, рассчитаем значения коэффициентов линейной регрессии:

$$a_1 = \frac{8 \cdot 24996,58 - 200,2 \cdot 934,28}{8 \cdot 124183,84 - 934,28^2} = 0,1072 ;$$

$$a_0 = \frac{1}{8} \cdot (200,2 - 0,1072 \cdot 934,28) = 12,503 .$$

Таким образом, искомое уравнение регрессии имеет следующий вид:

$$\hat{y} = 12,503 + 0,1072 x \quad (45)$$

Это значит, что при увеличении или уменьшении значения фактора на 1 руб. показатель увеличивается или уменьшается на 0,1072 млрд. руб., то есть между эконометрическими параметрами существует прямая пропорциональная или положительная зависимость.

Свободный член регрессии $a_0 = 12,503$ указывает значение показателя при нулевом значении фактора. Он имеет лишь расчетное значение, поскольку такой случай невозможен в реальной экономической ситуации.

2. Тесноту линейной связи оценим с помощью коэффициента корреляции, используя формулу (36):

$$r = \sqrt{\frac{\sum (\hat{y} - \bar{y})^2}{\sum (y - \bar{y})^2}} = \sqrt{\frac{173,29}{188,02}} = 0,96$$

Поскольку $0,9 \leq r \leq 1$, между факторным и результативным признаком имеется очень сильная прямая связь.

Коэффициент детерминации будет равен:

$$r^2 = 0,96^2 = 0,92 \text{ или } 92\%.$$

Это означает, что линейная модель объясняет 92% вариации объема производства. Не объясняется $100\% - 92\% = 8\%$ вариации данного показателя, которые обусловлены остальными факторами, не включенными в линейную модель регрессии.

Оценку статистической значимости коэффициента корреляции проведем с помощью t -статистики Стьюдента. Выдвигаем гипотезу о статистически незначимом отличии коэффициента от нуля.

Табличное значение t -статистики Стьюдента ($t_{табл}$) для числа степеней свободы $n-2=8-2=6$ и $\alpha = 0,05$ составит 2,45. При этом фактическое вычисленное значение будет равно:

$$t_r = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,96 \cdot \sqrt{8-2}}{\sqrt{1-0,92}} = 8,31$$

Фактическое значение t -статистики превосходит табличное значение, поэтому гипотеза отклоняется, то есть коэффициент корреляции не случайно отличается от нуля, а статистически значим.

3. Качество модели определим на основе средней ошибки аппроксимации по формуле (39):

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100 = \frac{1}{8} \cdot 0,36 \cdot 100 = 4,5\%$$

Качество построенной модели оценивается как хорошее, так как значение средней ошибки аппроксимации не превышает 10%.

4. Оценку статистической значимости параметров регрессии проведем с помощью t -статистики Стьюдента. Выдвигаем гипотезу о статистически незначимом отличии параметров от нуля.

$t_{табл}$ для числа степеней свободы $n-2=8-2=6$ и $\alpha = 0,05$ составит 2,45.

Рассчитаем дисперсию ошибки:

$$D_{осман} = \sqrt{\frac{\sum y^2 - a_0 \sum y - a_1 \sum xy}{n-2}} = \sqrt{\frac{5198,02 - 12,503 \cdot 200,2 - 0,1072 \cdot 24996,58}{6}} = 1,6$$

Определим случайные ошибки m_{a0} , m_{a1} :

$$m_{a0} = \frac{D_{осман} \sqrt{\sum x^2}}{n \sqrt{x^2 - \bar{x}^2}} = \frac{1,6 \cdot \sqrt{124183,84}}{8 \cdot \sqrt{15522,98 - 13639,9}} = 1,62$$

$$m_{a1} = \frac{D_{осман}}{\sqrt{n \cdot (x^2 - \bar{x}^2)}} = \frac{1,6}{\sqrt{8 \cdot (15522,98 - 13639,9)}} = 0,013$$

Тогда $t_{a0} = a_0 / m_{a0} = 12,503 / 1,62 = 7,72$

$t_{a1} = a_1 / m_{a1} = 0,1072 / 0,013 = 8,25$.

Фактические значения t -статистики превосходят табличное значение, поэтому гипотеза отклоняется, то есть параметры не случайно отличаются от нуля, а статистически значимы.

Рассчитаем доверительный интервал для коэффициентов регрессии a_0 и a_1 . Для этого определим предельную ошибку для каждого коэффициента:

$$\Delta_{a0} = t_{табл} \cdot m_{a0} = 2,45 \cdot 1,62 = 3,969 ;$$

$$\Delta_{a1} = t_{табл} \cdot m_{a1} = 2,45 \cdot 0,013 = 0,032 .$$

Доверительные интервалы имеют следующие значения:

$$\gamma_{a0} = a_0 \pm \Delta_{a0} = 12,503 \pm 3,969 ;$$

$$\gamma_{a0\min} = 12,503 - 3,969 = 8,534;$$

$$\gamma_{a0\max} = 12,503 + 3,969 = 16,472;$$

$$\gamma_{a1} = a_1 \pm \Delta_{a1} = 0,1072 \pm 0,032 ;$$

$$\gamma_{a1\min} = 0,1072 - 0,032 = 0,0752 ;$$

$$\gamma_{a1\max} = 0,1072 + 0,032 = 0,1392 .$$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью $p=1-\alpha =0,95$ параметры a_0 и a_1 , находясь в указанных границах, не принимают нулевых значений, то есть не являются статистически незначимыми и существенно отличны от нуля.

5. Поскольку результаты проведенной оценки качества регрессионной модели и ее параметров свидетельствуют о возможности использования полученного уравнения линейной регрессии для прогнозирования, определим прогноз объема товарной продукции УИС в стоимостном выражении. В данном случае прогнозирование заключается в подстановке в уравнение регрессии ожидаемого значения x , то есть перспективного уровня оплаты труда осужденных, равного 185 руб./чел.-день:

$$\hat{y} = 12,503 + 0,1072 x = 12,503 + 0,1072 \cdot 185 = 32,34 \text{ млрд. руб.}$$

Полученное прогнозное значение является точечным и не учитывает колебания уровней показателя. Поэтому произведем расчет стандартной ошибки прогноза ($m_{\hat{y}_p}$) и интервальной оценки точечного прогнозируемого значения \hat{y}_p .

Стандартная ошибка прогноза, вычисленная по формуле (43), будет равна:

$$m_{\hat{y}_p} = \sqrt{\frac{14,55}{6} \cdot \left(1 + \frac{1}{8} + \frac{(185 - 116,79)^2}{15073,95}\right)} = 1,86.$$

Табличное значение t -критерия Стьюдента $t_{табл}$ для числа степеней свободы $n-2=8-2=6$ и $\alpha = 0,05$ составит 2,45, следовательно

$$\Delta_{\hat{y}_p} = m_{\hat{y}_p} \cdot t_{табл} = 1,86 \cdot 2,45 = 4,56.$$

95-процентный доверительный интервал прогноза строится по формуле:

$$\gamma_{\hat{y}_p} = \hat{y}_p \pm \Delta_{\hat{y}_p} = 32,34 \pm 4,56.$$

Нижняя граница:

$$\gamma_{\hat{y}_p \min} = 32,34 - 4,56 = 27,78.$$

Верхняя граница:

$$\gamma_{\hat{y}_p \max} = 32,34 + 4,56 = 36,9.$$

При среднем уровне оплаты труда осужденных, равном 185 руб. за 1 чел.-день, значение объема товарной продукции УИС в стоимостном выражении с вероятностью 95% попадает в границы рассчитанного выше интервала.

Контрольные вопросы

1. Дайте определение следующих понятий: «результативный признак», «факторный признак», «регрессионный анализ», «регрессия».
2. Какова основная задача корреляционного анализа?
3. Перечислите этапы проведения регрессионного анализа и прогнозирования.
4. Дайте классификацию видов регрессии.
5. В чем заключается сущность метода наименьших квадратов?
6. Как следует интерпретировать полученное значение выборочного коэффициента корреляции?
7. Каким образом осуществляется проверка коэффициентов корреляции на значимость с помощью t -критерия Стьюдента?
8. О чем свидетельствует полученное значение коэффициента детерминации?
9. Как вычисляется и оценивается средняя ошибка аппроксимации?
10. Каким образом дается оценка статистической значимости параметров регрессии?
11. Опишите порядок вычисления доверительных интервала прогноза.