

Федеральное государственное казенное  
образовательное учреждение высшего образования  
«СИБИРСКИЙ ЮРИДИЧЕСКИЙ ИНСТИТУТ  
МИНИСТЕРСТВА ВНУТРЕННИХ ДЕЛ РОССИЙСКОЙ ФЕДЕРАЦИИ»

*Галушин П.В.,  
Молоков В.В.,  
Шерстяных А.С.*

# **ПРАВОВАЯ СТАТИСТИКА**

**Учебно-практическое пособие**

КРАСНОЯРСК  
СИБЮИ МВД РОССИИ  
2017

УДК 519.254 : 311.21

ББК 22.172 : 32.97

Рецензенты: Н.В. Рычкова – доцент кафедры информатики и специальной техники Барнаульского юридического института МВД России, полковник полиции  
А.Б. Афанасьев – начальник ИЦ ГУ МВД России по Красноярскому краю, полковник внутренней службы

Учебно-практическое пособие подготовлено сотрудниками кафедры информационно-правовых дисциплин и специальной техники СибЮИ МВД России: старшим преподавателем кафедры кандидатом технических наук капитаном полиции Галушиным П.В., начальником кафедры кандидатом технических наук, доцентом полковником полиции Молоковым В.В., доцентом кафедры кандидатом технических наук, доцентом полковником полиции Шерстяных А.С.

### **Галушин, П.В.**

Правовая статистика : учебно-практическое пособие / П.В. Галушин, В.В. Молоков, А.С. Шерстяных. – Красноярск: СибЮИ МВД России, 2017. – 100 с.

В пособии рассматриваются основные вопросы учебной дисциплины «Правовая статистика». Изложение материала базируется на возможностях библиотеки статистических функций Microsoft Excel. Каждый раздел снабжен списком задач для самостоятельного решения.

Учебно-практическое пособие предназначено для курсантов, слушателей, адъюнктов образовательных организации МВД России, будет полезным лицам, ведущим научные исследования в сфере криминологии. Может быть использовано сотрудниками органов внутренних дел, занимающихся сбором и обработкой статистической информации.

© СибЮИ МВД России, 2017

© П.В. Галушин, В.В. Молоков, А.С. Шерстяных, 2017

## ОГЛАВЛЕНИЕ

Пояснительная записка.....	4
1. Основные понятия статистики.....	7
2. Статистическое наблюдение. Сводка и группировка материалов статистического наблюдения.....	9
3. Обработка сводных данных с помощью статистических показателей.....	23
4. Возможности автоматизации обработки статистических данных в MS Excel.....	52
5. Динамические ряды.....	58
6. Статистический анализ взаимосвязи социально-правовых явлений.....	72
7. Экспертные оценки.....	94
Список рекомендуемой литературы.....	99

## ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Практическое пособие подготовлено в соответствии с требованиями и содержанием курса «Правовая статистика», изучаемого в Сибирском юридическом институте МВД России. Основная цель пособия – оказать помощь обучающимся всех форм обучения в овладении (в том числе самостоятельном) методами правовой статистики, применяемыми при изучении различных социально-правовых явлений.

Основными задачами пособия являются:

- изучение основных понятий правовой статистики;
- развитие у обучающихся способностей к самостоятельному мышлению, логическим обобщениям и умозаключениям;
- совершенствование навыков работы с различными источниками статистической информации;
- формирование навыков работы с функциями статистической обработки данных в MS Excel;
- формирование умения использовать полученные навыки при принятии решений в оперативно-служебной деятельности.

Роль правовой статистики в совершенствовании деятельности правоохранительных органов (юстиции, суда, органов, внутренних дел, пенитенциарных и др.) очевидна, поскольку именно она позволяет установить эффективность их работы. Кроме того, к числу основных задач правовой статистики относится не только информирование об истинном состоянии дел в сфере борьбы с преступностью в стране, но и выявление взаимосвязей состояния преступности и правонарушений с политическим, социально-экономическим, демографическим положением в стране, уровнем развития других социальных явлений и процессов.

В пособии раскрывается содержание основных тем правовой статистики, иллюстрируемое практическими примерами. Каждый раздел снабжен списком задач для самостоятельного решения.

Эффективное применение аппарата правовой статистики невозможно без понимания её терминологии и методологии. Тема «Основные понятия статистики» раскрывает основные понятия правовой статистики и этапы статистических иссле-

дований в правовой сфере: статистическое наблюдение, статистическая сводка и анализ. Подробнее данные этапы рассматриваются в следующих темах.

Для получения в результате статистического исследования правовых явлений достоверных выводов необходима правильная организация сбора, фильтрации и первичной обработки статистических данных. Тема «Статистическое наблюдение. Сводка и группировка материалов статистического наблюдения» посвящена первым двум этапам статистического исследования: наблюдению и сводке. В ходе изучения данной темы читатели получают представление о том, как представить большие массивы статистических данных в форме, пригодной для непосредственного восприятия и анализа человеком, в том числе в виде различных графиков.

Использование результатов статистического исследования невозможно без глубокого понимания смысла статистических показателей. В теме «Обработка сводных данных с помощью статистических показателей» приводятся основные статистические показатели, используемые для описания состояния преступности, криминогенной среды, эффективности работы территориальных органов внутренних дел. Кроме того, в данной теме приводится методика определения размера выборки, обеспечивающего заданную точность выборочного статистического исследования.

Тема «Возможности автоматизации обработки статистических данных в MS Excel» посвящена изучению высокоуровневых возможностей Microsoft Excel, которые позволяют оперативно выполнять стандартные статистические процедуры без существенных затрат рабочего времени.

Изучения лишь одного состояния какого-либо явления недостаточно для получения полного представления об этом явлении. В теме «Динамические ряды» изучаются методы анализа развития социально-правовых процессов во времени, а также способы визуального представления динамики таких процессов.

Как правило, изучение отдельных социально-правовых явлений является лишь подготовительным этапом более крупного статистического исследования. Наибольший интерес представляет изучение взаимосвязей между такими явлениями, как, например, влияние изменения уровня жизни,

изменения уголовного законодательства и методов работы правоохранительных органов на уровень преступности. В теме «Статистический анализ взаимосвязи социально-правовых явлений» излагаются статистические методы, которые могут быть использованы, например, для изучения влияния деятельности правоохранительных органов на криминогенную ситуацию.

Зачастую помимо объективных данных и данных социологических исследований необходимо привлечение экспертных оценок. Тема «Экспертные оценки» посвящена обработке сведений, полученных в результате опроса экспертов, в ней излагаются методы, позволяющие обобщить мнения нескольких экспертов, а также оценить степень согласованности их мнений.

Для выполнения практических заданий необходим персональный компьютер с установленным табличным процессором Microsoft Office Excel (авторы используют версию 2007, но читатели могут использовать и более новые версии). Для работы Microsoft Office 2007 потребуется компьютер с операционной системой Windows XP (с установленным Service Pack 2) или более новой, 1 ГБ оперативной памяти, процессором с тактовой частотой от 500 МГц, экраном с разрешением не менее 1024x768 точек.

Некоторые задания требуют получения статистической информации с официального сайта Федеральной службы государственной статистики, для выполнения этих заданий необходим веб-браузер и подключение к сети Интернет.

# 1. ОСНОВНЫЕ ПОНЯТИЯ СТАТИСТИКИ

*Статистика – это отрасль науки, которая с помощью присущих ей приемов и методов изучает количественную сторону (в неразрывной связи с качественной стороной) массовых явлений и процессов и дает числовое выражение тенденций и закономерностей их развития.*

Статистикой называют также *различного рода числовые, или цифровые, данные*, характеризующие различные стороны жизни государства: экономику, население, преступность, т.е. с точки зрения сведений к ней относят только то, что получает количественное выражение.

Под статистикой в том числе понимают *процесс «ведения»*, т.е. процесс собирания и обработки сведений о фактах, необходимых для получения статистики в рассмотренном выше смысле. Кроме того, *статистика – это отрасль (форма) практической деятельности*, цель которой – результат указанного процесса «ведения».

*Статистическая совокупность* – множество элементов, обладающих массовостью, некоторыми общими, но не обязательно системными свойствами, существенными характеристиками – однородностью, определенной целостностью, взаимозависимостью состояний отдельных элементов и наличием вариации признаков, их характеризующих.

*Вариацией называют* количественное изменение значений признака при переходе от одной единицы совокупности к другой. Вариация возникает под воздействием случайных и прежде всего внешних причин и условий.

Признаки – свойства, характерные черты или особенности объектов (явлений), которые могут быть охарактеризованы рядом статистических величин.

В научной литературе *признаки делят на качественные и количественные*. Это деление условно, так как всегда существует неразрывная связь между качеством и количеством.

К первому виду относятся признаки, варианты которых отличаются друг от друга качественным содержанием. Так, профессии – дознаватель, следователь, оперативный сотрудник; преступления, связанные с незаконным оборотом наркотиков (НОН) – незаконное приобретение, хранение, пере-

возка и т.д.; способы сбыта наркотиков – контактный, бесконтактный и т.п. Это качественные признаки.

Во втором случае это признаки (количественные), варианты которых отличаются друг от друга определенной величиной: возраст человека (14, 16, 18 и т.д.), объем (вес) изъятых наркотических средств – 1 грамм, 2 килограмма, 1 тонна и т.д.

*Закономерностями (или регулярностями) обычно называют устойчивые отношения (связи).*

*Закономерность, проявившаяся лишь в большой массе явлений через преодоление свойственной ее единичным элементам случайности, называется статистической закономерностью.*

Методы статистики принято делить на две основные группы: методы статистического наблюдения и методы обработки и анализа статистических данных (т.е. результатов наблюдения).

Процесс статистического исследования проходит три основных этапа.

*Первый этап* – массовое статистическое наблюдение. Здесь в результате регистрации фактов по научной и тщательно разработанной программе получают объективные данные об изучаемых социально-экономических явлениях.

*Статистическое наблюдение – первый этап статистического исследования – планомерное, научно организованное и, как правило, систематическое получение данных (сбор сведений) о массовых явлениях и процессах социальной и экономической жизни путем регистрации существенных признаков каждой единицы их совокупности.*

*Второй этап* – сводка и обработка, группировка (классификация) и систематизация материалов, собранных в результате массового статистического наблюдения.

*Статистическая сводка* – второй этап статистического исследования, который представляет собой проверку, систематизацию, научную обработку материалов статистического наблюдения (подсчет первичного статистического материала, например карточек на лиц, совершивших преступления), подытоживание отдельных единиц и сведение их в массы или совокупности в целях получения обобщенной характеристики изучаемого явления по ряду существенных для него

признаков (например, число несовершеннолетних, совершивших преступления).

*Третий этап* – анализ показателей, полученных в результате сводки и обработки статистических материалов.

Как заключительный этап статистического исследования *анализ статистических данных* – это процесс изучения, сопоставления, сравнения полученных цифровых данных (между собой и с данными других отраслей социально-экономической статистики), их обобщения, истолкования и формулирования научных и практических выводов.

Все три этапа статистического исследования находятся между собой в тесной связи и требуют взаимосогласованности при их реализации.

## **2. СТАТИСТИЧЕСКОЕ НАБЛЮДЕНИЕ. СВОДКА И ГРУППИРОВКА МАТЕРИАЛОВ СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ**

### **2.1. МЕТОД ВЫБОРОЧНОГО НАБЛЮДЕНИЯ**

*Статистическое исследование* – это анализ и обработка, как правило, большого количества элементов, иногда достигающее до десятков и сотен миллионов. Например, чтобы установить среднее количество детей в российской семье, строго говоря, необходимо обследовать все семьи страны. Такая работа потребует больших временных и материальных затрат и будет неэффективной. Поэтому чаще всего обработке подвергаются не все элементы, а только небольшую их часть.

*Генеральная совокупность* – это полная совокупность элементов, подвергающихся статистическому исследованию.

*Выборка* – это часть генеральной совокупности (выборочная совокупность), которая подвергается непосредственному исследованию.

*Объемом совокупности* называется общее количество единиц наблюдения в совокупности. Объем генеральной совокупности ( $N$ ) всегда значительно превосходит объем выборки ( $n$ ).

Сущность выборочного обследования заключается в том, что исследованию подвергается только часть интересующей

нас генеральной совокупности, а полученные результаты служат характеристикой всех единиц совокупности.

Основная проблема в теории выборочных исследований – решение вопроса о правомерности распространения на всю генеральную совокупность тех выводов, которые будут получены при анализе выборочной совокупности. Правомерность такого распространения во многом зависит от объема выборки.

Одним из первых вопросов, который встает при проведении социологического исследования, является установление требуемого числа обследуемых объектов (например, правонарушителей), или *репрезентативного объёма выборки*. Такое число, с одной стороны, должно быть минимальным, но вместе с тем достаточным для того, чтобы исследование было показательным, т.е. обладающим достоверностью выводов об изучаемом явлении. Репрезентативность в качественном отношении означает достаточно полное приближение параметров выборки к характеристикам генеральной совокупности. Поэтому в процессе отбора необходимо стремиться к максимально возможному учёту особенностей изучаемого явления. Выборка, достаточно точно воспроизводящая генеральную совокупность, называется репрезентативной (представительной).

Для правильной организации выборочного обследования необходимо соблюдать следующие условия:

– число взятых в выборку единиц должно быть достаточно велико, поскольку закономерности могут быть выявлены только при массовом наблюдении;

– выбор отдельных единиц должен происходить таким образом, чтобы каждая из них имела совершенно одинаковые (равновероятные) шансы попасть в выборку;

– выбор должен быть произведён из всех частей изучаемой совокупности.

Применительно к сфере деятельности правоохранительных органов объектами (элементами) исследования могут быть, например: все сотрудники МВД России в определенном регионе; все лица, родившиеся в определенном году, выявленные в связи с совершением преступлений, связанных с НОН; все осужденные на территории региона, содержащиеся в колониях общего режима, и т.д. Поскольку всю ге-

неральную совокупность исследовать трудно, ее показатели изучают по выборке.

Часто при анализе данных возникает необходимость выборочного их просмотра. В MS Excel имеется несколько способов выборочного анализа участков таблиц. Самое простое – это установка **автоматических фильтров**.

**Упражнение 1.** Найдите в сетевой папке по адресу «\\infserv\install\disk\_D\ПС» файл *Группировка.xlsx*. Скопируйте его в каталог своей группы и откройте.

В файле дана выборка осужденных по статьям, связанным с незаконным оборотом наркотиков. Сделайте активной одну их ячеек списка. Щелкните по кнопке **Сортировка и фильтр**, расположенной на вкладке **Главная** в группе **Редактирование**. В открывшемся списке выберите пункт меню **Фильтр**.



Разверните список возможных значений столбца «Статья». Выберите значение 174. Просмотрите результат.

Разверните список столбца «Возраст». Выберите опцию **Числовые фильтры**. Ознакомьтесь с вариантами выбора условий отбора. Выберите **больше или равно**. Задайте условия «больше или равно» – 16 И «меньше или равно» – 22. Нажмите **ОК**. Просмотрите результат.

Отмените автофильтр, щелкнув по кнопке **Сортировка и фильтр**. В открывшемся списке выберите пункт меню **Очистить**.

**Упражнение 2.** На новом листе рабочей книги MS Excel по данным таблицы постройте группировку осужденных лиц по категории совершенного преступления (ст. 15 УК РФ): небольшой тяжести, средней тяжести, тяжкие и особой тяжести. Каждую группу охарактеризуйте числом лиц, совершивших преступление, из них: мужчин, женщин. Сделайте вывод.

**Упражнение 3.** По данным таблицы произведите группировку осужденных лиц по характеру совершенного преступления. Группировочным признаком возьмите разделы Особенной части УК РФ. При характеристике группы отдельно рассмотрите подгруппы, группировочным признаком которых служит глава Особенной части УК РФ. Каждую

группу охарактеризуйте числом лиц, совершивших преступление:

– из них мужчин (укажите число лиц с высшим образованием);

– женщин (укажите число лиц с высшим образованием).

На основе данных группировки охарактеризуйте преступления, совершаемые женщинами, мужчинами, с выделением лиц с высшим образованием.

Дайте социологический портрет «правонарушителя» отдельно взятого раздела Особенной части УК РФ.

Второй способ отбора данных предоставляется при задании **расширенных фильтров**, которые размещаются на листе рядом с таблицей. Расширенный фильтр, в отличие от обычного, позволяет отбирать данные по нескольким условиям.

Отбор данных по расширенному фильтру может производиться как непосредственно в месте расположения исходных данных путем скрывания строк, не удовлетворяющих условиям отбора, так и путем помещения отобранных данных в отдельное место таблицы. Извлеченные данные могут рассматриваться как новый список, к которому применимы все описанные выше приемы сортировки и фильтрации.

**Упражнение 4.** Найдите в сети по адресу «\\infserv\install\disk\_D\ПС» файл *Фильтры\_исходные\_данные.xlsx*. Скопируйте его в каталог своей группы и откройте.

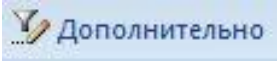
Анализировать данные, представленные в таком виде крайне неудобно. Гораздо удобнее, когда данные будут иметь структурированный вид, например, такой:

Возраст	228	228.1-228.4	229-229.1	230-233	234-234.1	Всего
16	2	6	2	4	2	16
18	3	7	3	2	4	19
...						
44	0	5	2	1	1	9

Для организации такого подсчета необходимо из исходной выборочной совокупности сформировать соответствующую

щие вспомогательные выборки. Для этого воспользуемся расширенным фильтром.

Справа от исходных данных (начиная со столбца **Е**) наберите шапку таблицы, как это показано на рисунке 1. Заполните столбцы **М** и **Н** в соответствии с данными рисунка. Отформатируйте таблицу соответствующим образом.

На вкладке **Данные** в группе **Сортировка и фильтр** нажмите кнопку . Установите переключатель в положение «Скопировать результат в другое место». Задайте исходный диапазон **В2:С502**. Диапазон условий **М2:М3**. Поместите результат в диапазон **Р2**. Нажмите **ОК**. Обратите внимание, что во всех случаях при задании диапазонов включаются и заголовки столбцов, содержащих критерии отбора и исходные данные.

Проделайте последовательно операции, аналогичные предыдущей, для диапазонов критериев **М5:М9**, **М11:М13**, **М15:М16**, **М19:М21**, размещая результаты в блоках, начиная с ячеек **Р2**, **Т2**, **У2**, **Х2** соответственно, смотрите рисунок 2. Обратите внимание, что максимальное число выбранных данных находится в столбце **Р** (до ячейки **Р167**).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Исходные данные				частоты								критерии	
2	№	Возраст	Статья	границы	228	228.1-228.4	229-229.1	230-233	234-234.1	общее			статья	
3	1	19	229	16									228	
4	2	17	228.3	18										
5	3	35	228.4	20									статья	
6	4	30	228.2	22									228.1	
7	5	18	228.2	24									228.2	
8	6	38	232	26									228.3	
9	7	19	228	28									228.4	
10	8	38	228.3	30										
11	9	31	234.1	32									статья	
12	10	38	229.1	34									229	
13	11	28	233	36									229.1	
14	12	17	234	38										
15	13	28	231	40									статья	статья
16	14	23	231	42									>229	<234
17	15	21	228.4	44										
18	16	30	232											
19	17	39	230										статья	
20	18	20	229										234	
21	19	19	232										234.1	

Рисунок 1. Критерии для расширенного фильтра

M	N	O	P	Q	R	S	T	U	V	W	X	Y
критерии												
			Возраст	Статья	Возраст	Статья	Возраст	Статья	Возраст	Статья	Возраст	Статья
статья	228		19	228	17	228.3	19	229	38	232	31	234.1
			17	228	35	228.4	38	229.1	28	233	17	234
статья	228.1		22	228	30	228.2	20	229	28	231	33	234
	228.2		33	228	18	228.2	22	229	23	231	19	234
	228.3		17	228	38	228.3	43	229.1	30	232	31	234
	228.4		16	228	21	228.4	31	229.1	39	230	30	234.1
			29	228	39	228.1	35	229	19	232	17	234
			28	228	22	228.4	32	229	18	231	19	234.1
статья	229		24	228	17	228.2	35	229	27	230	31	234.1
	229.1		25	228	28	228.2	32	229.1	41	231	23	234
			42	228	29	228.1	31	229.1	38	230	41	234

Рисунок 2. Выборочный частотный анализ

Введите в ячейки **E3:E17** границы возрастов для построения частотных рядов (в нашей выборке возраст меняется в интервале от 16 до 44 лет). В столбце **F** рассчитаем частоту появления возрастов осужденных по статье 228. Для этого выделите блок ячеек **F3:F17**. В строке формул введите формулу **=ЧАСТОТА(P3:P32; E3:E17)**. Нажмите **Ctrl + Shift+Enter**. Обратите внимание, что этот способ позволяет ввести формулу сразу в группу ячеек.

*Примечание.* Формула **ЧАСТОТА** является формулой массива, т.е. она возвращает ряд значений в указанные ячейки. Чтобы ввести формулу массива необходимо выделить нужное количество пустых ячеек, набрать формулу, затем нажать комбинацию клавиш **Ctrl + Shift+Enter**. Если формула не будет введена как формула массива, в ячейке **F3** отобразится только одно ее значение.

Аналогично введите формулу для подсчета остальных частот. Обратите внимание, что количество исходных данных для каждого критерия различно.

Посчитайте общее количество частот для каждого возраста (воспользуйтесь функцией автосуммирования). Убедитесь, что общее количество данных равно 500.

По данным блока **E3:J17** постройте диаграммы полученных распределений, задав тип **График** (для каждого столбца создайте свою диаграмму). Сделайте выводы.

## 2.2. ФОРМИРОВАНИЕ ВЫБОРОЧНОЙ СОВОКУПНОСТИ

**Упражнение 5.** Необходимо исследовать возраст осужденных по статье 228 УК РФ, отбывающих наказание в колониях города Красноярск.

Для этого необходимо сформировать выборку лиц, осужденных по статье 228 УК РФ, и определить их возраст. Объем выборки возьмем равным  $n=120$ .

Для начала работы в папке своей группы создайте файл **Статистические данные**. Переименуйте **Лист1** в **Вариация**. Для этого:

- щелкните правой кнопкой мыши по ярлыку **Лист1**;
- в появившемся контекстном меню выберите пункт **Переименовать**;
- наберите новое имя листа и нажмите **ОК**.

Для упрощения формирования выборки в учебных целях мы не будем брать реальные данные статистических наблюдений и смоделируем их случайным образом (по нормальному закону распределения). Для этого:

1. Занесем в ячейку **A1** формулу генератора случайных чисел:

**=ABS(ЦЕЛОЕ(НОРМОБР(СЛЧИС();45;9)))**

где *45* – математическое ожидание случайной величины (среднее значение), *9* – среднеквадратическое отклонение  $\sigma$  (разброс вокруг среднего значения в диапазоне  $\pm 3\sigma$ ).

2. С помощью маркера заполнения (в правом нижнем углу выделенной ячейки) скопируем содержимое ячейки **A1** в диапазон ячеек **A1:A120**.

3. Полученный ряд числовых данных изменяет свои значения (в этом можно убедиться, нажимая клавишу **F9**). Для фиксации числовых значений необходимо сделать следующее:

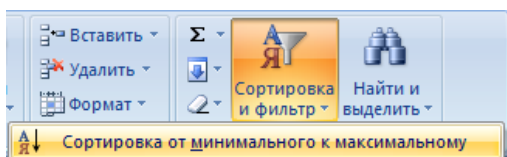
- выделите диапазон ячеек **A1:A120**;
- выберите пункт **Копировать** на вкладке **Главная**;
- выберите пункт **Специальная вставка** группа **Буфер обмена** вкладка **Главная**;
- в открывшемся окне выберите команду **значения** и нажмите **ОК**.

Теперь у нас есть варианты наблюдения случайной величины (возрастов осужденных) объемом 120 единиц.

### 2.3. ПОСТРОЕНИЕ НЕСГРУППИРОВАННОГО ВАРИАЦИОННОГО РЯДА

*Вариационным рядом* называется ранжированный (упорядоченный) в порядке возрастания или убывания ряд вариантов признака с соответствующими им весами (частотами или частостями).

Чтобы построить вариационный ряд возрастов нам необходимо произвести ранжирование (упорядочивание) данных ряда в порядке возрастания возраста осужденных. Для ранжирования вариационного ряда



выделим интервал значений признака **A1:A120** и воспользуемся кнопкой **Сортировка от минимального к максимальному** в группе **Редактирование** вкладки **Главная**. В таком виде изучать выборку удобнее.

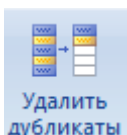
Для наглядности представления данных построим на этом же листе следующую таблицу:

Для наглядности представления данных построим на этом же листе следующую таблицу:

	В	С
1	Вариант (возраст в годах)	Частота
2	$x_1$	$f_1$
3	$x_2$	$f_2$
...	...	...
	$x_m$	$f_m$
	Итого:	$\sum f_i$

Оформите шапку таблицы. В первый столбец этой таблицы необходимо поместить по одному разу те значения возрастов, которые встречаются в полученной выборке.

Воспользуемся для этого функцией «Удалить дубликаты». Для этого скопируйте диапазон A1:A120 в столбец В, начиная с ячейки В2. На вкладке **Данные** в группе **Работа с данными** нажмите на кнопку **Удалить дубликаты**. В открывшемся диалоговом окне выберите пункт **«Сортировать в пределах выделенного диапазона»** и нажмите кнопку **«Удалить дубликаты...»**. В следующих диалоговых окнах нажмите кнопку **ОК**.



Второй столбец таблицы содержит частоту повторяемости каждого варианта из первого столбца вариационного ряда (то есть сколько раз тот или иной возраст встречается в выборке). Полученная в результате таблица должна выглядеть примерно так, как показано ниже:

	В	С
1	Вариант (возраст в годах)	Частота
2	16	2
3	17	1
4	19	3
...	...	...
	62	2
	Итого:	$\sum f_i$

После того как будут заполнены оба столбца вариационного ряда, необходимо подсчитать сумму частот  $\sum f_i$ . Для этого можно воспользоваться кнопкой **Автосумма** в группе **Редактирование** вкладки **Главная** (выделить диапазон данных, подлежащих суммированию, включая последнюю пустую ячейку и нажать кнопку). Полученная сумма частот должна совпадать с объемом  $n$  выборки и равняться 120.



## 2.4. ПОСТРОЕНИЕ СГРУППИРОВАННОГО ВАРИАЦИОННОГО РЯДА

Изучать выборку по несгруппированному вариационному ряду не совсем удобно из-за обилия числовых значений. Поэтому необходимо разбить варианты изменения признака на отдельные интервалы, т.е. провести группировку.

Число интервалов  $m$  следует брать не очень большим, чтобы после группировки ряд не был громоздким, но и не очень малым, чтобы не потерять особенности распределения признака.

Рекомендуемое число интервалов  **$m=6-8$** , а значение ширины интервалов определяют по формуле (округляя полученное значение до целого в ближайшую сторону):

$$k = \frac{x_{\max} - x_{\min}}{m}.$$

Для построения сгруппированного вариационного ряда необходимо воспользоваться вспомогательной таблицей 2, где в качестве примера уже рассмотрены сгруппированные по 6 интервалам значения вариантов возрастов (в нашем примере  $m=6$ ,  $k=10$ ).

Например, у нас получилось шесть подинтервалов по 10 лет в каждом: 16-25 лет, 26-35 лет, 36-45 лет, 46-55 лет, 56-65 лет, 66-75 лет:

	G	H	I	J	K	L	M
1	<b>Вариационный ряд распределения возрастов</b>						
2	Возраст в годах	16-25	26-35	36-45	46-55	56-65	66-75
3	Частоты	$f_{16} + \dots +$	$f_{26} + \dots +$	...	...	...	...
4	Относительные частоты	$\bar{f}_1$	$\bar{f}_2$	$\bar{f}_3$	...	...	$\bar{f}_6$
5	Накопленные частоты	$\bar{f}_1$	$\bar{f}_1 + \bar{f}_2$	$\bar{f}_1 + \bar{f}_2 + \bar{f}_3$	...	...	$\bar{f}_1 + \bar{f}_2 + \dots + \bar{f}_6$

На основе данных из таблицы 1 во второй строке рассчитаны суммы всех частот в конкретном возрастном интервале. В третьей строке рассчитываются относительные частоты  $\bar{f}_i$  по формуле:

$$\bar{f}_1 = \frac{f_{16} + f_{17} + f_{18} + \dots + f_{25}}{N}, \quad \text{и} \quad \bar{f}_2 = \frac{f_{26} + f_{27} + f_{28} + \dots + f_{35}}{N} \quad \text{т.д.},$$

где  $f_{16}$  – то, сколько раз возраст 16 встречается в выборке;  $f_{35}$  – то, сколько раз возраст 35 встречается в выборке и т.д.;  $N$  – объем выборки.

**Упражнение 6.** По вашей выборке на том же листе постройте сгруппированный вариационный ряд. Обратите внимание, что минимальное и максимальное значение возраста в вашей выборке может отличаться от приведенного в примере, соответственно, интервалы возрастов также могут иметь отличия. Для того чтобы построить сгруппированный вариационный ряд, выполните следующие действия:

1. Разбейте несгруппированный вариационный ряд на  $m=6-8$  интервалов и определите ширину интервала  $k$  по

формуле, указанной выше. Значение  $m$  выбираем таким образом, чтобы значение  $k$  получалось как можно ближе к целому числу ( $k$  округляем до целого). Данные по интервалам занесите в первую строку таблицы 2.

2. Используя таблицу 1, подсчитайте абсолютную частоту встречаемости признака попавшего в каждый из интервалов. Данные занесите во вторую строку таблицы 2.

3. На основе полученных значений абсолютной частоты, используя формулы, приведенные в таблице 2, вычислите значения относительных частот по каждому интервалу. Напомним, что объем выборки  $n = 120$ .

4. Вычислите накопленную частоту. Накопленная частота показывает (в долях), сколько наблюдалось вариантов со значением признака, меньшим  $x$ .

## 2.5. ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ ВАРИАЦИОННОГО РЯДА

Табличное и графическое представление вариационных рядов помогают выявить характерные особенности изучаемых социально-правовых явлений, найти максимальные и минимальные значения частот, дать им объяснение и определить тенденции развития соответствующих процессов.

Для графического изображения вариационных рядов наиболее часто используются гистограмма, полигон, кумулята.

**Гистограмма** служит только для отображения частот вариантов вариационного ряда и представляет собой ступенчатую фигуру из прямоугольников с высотами, равными абсолютным частотам (рис. 3).

**Полигон (или Полигон частот)**, как правило, служит для отображения относительных частот (долей, удельного веса) вариантов вариационного ряда и представляет собой ломаную, у которой концы отрезков прямой имеют координаты  $(x_i, f_i/n)$ ,  $i=1,2,\dots,m$ , где  $x_i$  – значение варианта,  $f_i/N$  – его относительная частота (рис. 4).

**Кумулята** (кумулятивная кривая) – кривая накопленных частот. Накопленные частоты определяются путем последовательного суммирования частот. Накопленные частоты показывают, сколько процентов единиц совокупности имеют значения признака не больше, чем рассматриваемое значение. Кумулята представляет собой ломаную, соединяющую

точки с координатами  $(x_i, f_i^{нак})$ ,  $i = 1, 2, \dots, m$ , где  $x_i$  – значение варианта,  $f_i^{нак}$  – его накопленная частота (рис. 5).

Обратите внимание, что на всех диаграммах по оси X откладываются интервалы возрастов вариационного ряда.

**Упражнение 7.** На основе сгруппированного вариационного ряда построить гистограмму, полигон и кумюляту.

Пример:

	G	H	I	J	K	L	M
1	<b>Вариационный ряд распределения возрастов</b>						
2	Возраст	16-25	26-35	36-45	46-55	56-65	66-75
3	Абсолютная частота	17	24	35	21	15	8
4	Относительная частота	0,1416667	0,2	0,2917	0,175	0,125	0,0667
5	Накопленная частота	0,1416667	0,34166667	0,6333	0,8083	0,9333	1

### Гистограмма

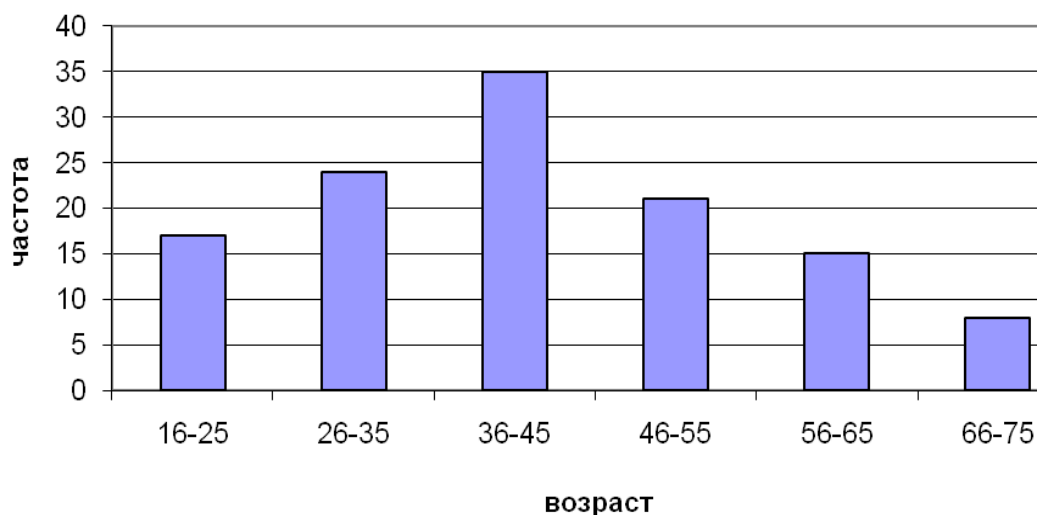


Рисунок 3. Гистограмма

### Полигон

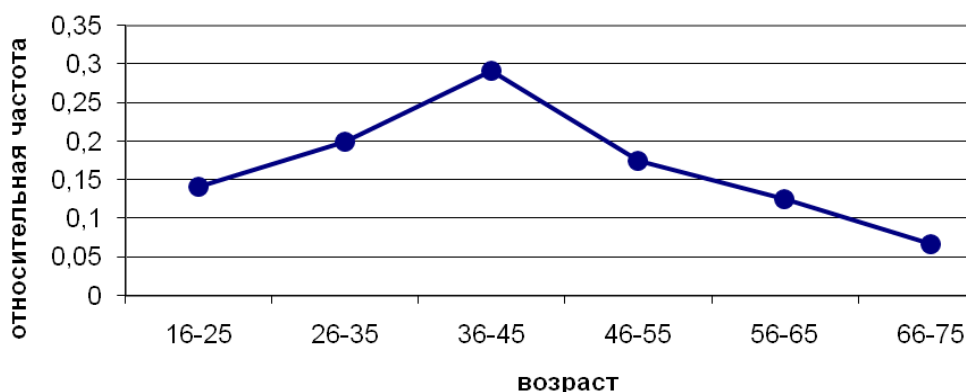


Рисунок 4. Полигон частот

### Кумулята

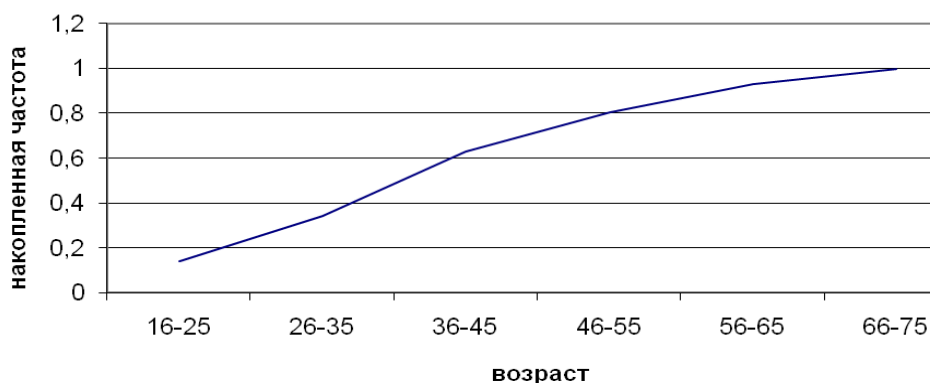


Рисунок 5. Кумулята

### ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

**Задание 1.** Органами полиции города  $N$  зарегистрированы преступления, совершенные несовершеннолетними в возрасте 13–17 лет в мае прошлого года. В результате получен ряд наблюдений возрастов малолетних преступников:  
16 13 15 13 15 17 17 15 15 14 16 14 14 14 14 14 13 16 17 16  
13 15 14 15

На основе этих данных:

- составьте сгруппированный вариационный ряд распределения и постройте полигон распределения;
- произведите сравнительный анализ (построив два полигона на одном графике), если известно, что за аналогичный период предыдущего года преступления, совершенные

несовершеннолетними, распределялись по возрасту следующим образом:

Возраст, лет	13	14	15	16	17
Относительная частота	0,12	0,22	0,30	0,28	0,08

**Задание 2.** Дана выборка количества зарегистрированных приводов в полицию в течение года для 20 несовершеннолетних, состоящих на учете в наркодиспансере. Уточнив у преподавателя номер Вашего варианта, по данным выборки составьте вариационный ряд и постройте полигон, гистограмму и кумуляту. Сделайте вывод.

Вариант	Выборка																			
	1	0	3	1	0	0	0	1	1	1	3	0	3	2	0	2	0	0	0	4
2	3	4	1	6	1	4	1	1	2	0	2	5	3	1	1	1	2	6	2	3
3	2	1	5	5	0	2	3	2	2	1	3	2	2	4	2	0	1	2	0	3
4	5	2	1	1	2	3	0	2	3	2	1	1	0	0	4	2	0	1	1	2
5	1	0	2	0	0	2	1	0	2	3	3	1	0	3	2	2	1	4	3	2
6	0	2	2	1	3	0	2	1	3	3	2	4	2	0	0	2	3	0	2	0
7	3	1	2	0	2	1	4	0	2	2	2	1	1	2	0	1	1	1	2	3
8	1	3	1	0	2	5	3	3	1	0	3	0	2	2	1	3	2	3	5	0
9	0	3	0	2	4	1	1	4	3	6	1	3	0	0	5	1	4	0	1	1
10	0	0	0	3	0	3	2	1	2	1	1	1	0	1	3	0	1	1	3	0
11	0	1	1	2	2	1	0	2	3	1	2	1	1	3	2	4	0	0	4	3
12	1	1	2	2	1	2	0	1	0	0	1	2	1	4	1	1	0	1	1	0
13	0	4	2	4	1	2	0	0	1	2	3	0	2	2	1	2	2	3	2	1
14	0	1	2	0	0	0	0	0	2	3	3	1	0	0	2	1	1	3	2	1
15	0	0	2	2	3	0	1	2	3	2	1	3	0	0	0	0	1	0	1	2

### 3. ОБРАБОТКА СВОДНЫХ ДАННЫХ С ПОМОЩЬЮ СТАТИСТИЧЕСКИХ ПОКАЗАТЕЛЕЙ

#### 3.1. АБСОЛЮТНЫЕ ВЕЛИЧИНЫ В ПРАВОВОЙ СТАТИСТИКЕ

В процессе статистического наблюдения, сводки и группировки получают данные о значениях тех или иных признаков исследуемой совокупности.

Конкретное количественное выражение каких-либо показателей (признаков) в статистике именуется *величиной*. В зависимости от метода исчисления *величины* делятся на *абсолютные* и *обобщающие*.

*Абсолютные величины* – исходная, первичная, самая общая форма выражения статистических показателей, характеризующая объем совокупности, т.е. число единиц, составляющих ее. Следовательно, абсолютные величины представляют собой суммарные числа, взятые из статистических таблиц без всякого преобразования.

Абсолютная величина – это всегда именованное число, связанное с единицей измерения (преступлений, потерпевших, уголовных дел, суммы ущерба и т.д.).

Однако, характеризуя сумму значений первичных признаков объекта наблюдения, абсолютные величины сами по себе недостаточно показательны, убедительны. Они не всегда могут быть подвергнуты непосредственному статистическому анализу. Зачастую они не дают возможности выявить реально существующие закономерности и взаимосвязи.

Более глубокий анализ требует измерить и характеризовать соотношение разных абсолютных величин, их изменения во времени, их взаимосвязи между собой и окружающей средой.

*Абсолютные величины, приведенные в сравнимый вид, называются в статистике обобщающими показателями. Обобщающие показатели* подразделяются на относительные и средние величины (показатели).

### 3.2. ОТНОСИТЕЛЬНЫЕ ВЕЛИЧИНЫ

*Величины, полученные путем сравнения, сопоставления двух абсолютных показателей, в статистике называют относительными величинами.* Относительная величина получается как частное от деления одной абсолютной величины на другую, которая служит основанием (базой) сопоставления.

Сопоставлять можно *одноименные абсолютные показатели*, относящиеся к различным периодам, различным объектам или разным территориям, а также *разноименные показатели*, относящиеся к одному объекту или явлению.

Все виды относительных величин принято делить на следующие группы:

1. Относительные величины, характеризующие долю. Они ставят своей задачей определить отношение части к целому (удельный вес) и отношение слагаемых к сумме (структуру совокупности). К этой группе относятся относительные величины интенсивности (ОВИ) и относительные величины структуры совокупности (ОВСС).

2. Относительные величины, характеризующие темпы изменения каких-либо явлений во времени. К ним относятся относительные величины динамики (ОВД).

3. Относительные величины степени, выражающие отношение разнородных величин, не связанных между собой как часть и целое или как слагаемое и сумма (например, сравнение фактического и планового выпуска продукции или числа жителей и площадь в квадратных километрах), а также отношение одной части совокупности к другой (например, отношение числа осужденных мужчин к числу осужденных женщин). К этой группе относятся отношения, характеризующие выполнение плана (ОВП), отношения степени и сравнения (ОВСр).

Все относительные показатели представляют собой результат сравнения различных величин, поэтому важное условие любого статистического исследования – соблюдение требования сравнимости показателей не только во времени (показатели должны быть взяты на одну и ту же дату или за один и тот же интервал) и пространстве (показатели должны

характеризовать один и тот же регион), но и сопоставимости данных с точки зрения правовых норм.

Рассмотрим отдельные виды относительных величин.

*Относительная величина интенсивности (ОВИ)* – отношение величины явления, характеризующегося определенным признаком, к размеру среды, которая является его базой, и выражаемое именованными числами, включающими наименование величин и измеряемого признака и среды его распространения:

$$\text{ОВИ} = \frac{\text{Показатель, характеризующий явление А}}{\text{Показатель, характеризующий среду распространения явления А}}$$

ОВИ показывает степень развития явления в данной среде: сколько единиц числителя приходится на 1, 100, 1000 и т.д. единиц знаменателя (например, при определении уровня рождаемости рассчитывается число родившихся на 1000 человек населения, при определении плотности населения рассчитывается число людей, приходящихся на 1 кв. км территории).

*Уровень преступности* – отражает количество преступлений и лиц, их совершивших, в относительных величинах. В уголовно-правовой статистике данный вид относительных величин в форме коэффициентов преступности занимает важное место. Они рассчитываются как отношение фактов преступлений или числа лиц, их совершивших, к численности населения, достигшего возраста, с которого наступает уголовная ответственность.

Уровень преступности измеряется в коэффициентах (индексах) преступности, к ним относятся следующие показатели: коэффициент преступности, коэффициент криминальной активности населения, коэффициент виктимности.

*Коэффициент преступности определяется* отношением числа совершенных преступлений к численности населения, достигшего возраста уголовной ответственности (с 14 лет), и рассчитывается по формуле:

$$K_{\Pi} = \frac{\Pi \times 100\ 000}{N}$$

где:  $K_{\Pi}$  – коэффициент преступности,  $\Pi$  – количество зарегистрированных преступлений,  $N$  – численность населения старше 14 лет, 100 000 – единая расчетная база, т.е. коэффициент преступности представляет собой не что иное, как количество совершенных преступлений приходящихся на 100 000 человек населения.

*Коэффициент криминальной активности населения*, рассчитывается по формуле:

$$K_{\Pi a} = \frac{m \times 100\,000}{N}$$

где:  $K_{\Pi a}$  – коэффициент криминальной активности населения,  $m$  – число лиц, совершивших преступление,  $N$  – численность активного населения старше 14 лет.

*Коэффициент виктимности* является показателем криминальной напряженности, он определяет относительное количество потерпевших от преступлений, рассчитывается по формуле:

$$K_{B} = \frac{K_{\text{ж}} \times 100\,000}{N}$$

где:  $K_{B}$  – коэффициент виктимности,  $K_{\text{ж}}$  – количество жертв от преступлений,  $N$  – численность населения.

*Относительные величины структуры совокупности (ОВСС)* – отношение части (частей) к целому, называемое удельным весом, или долей; выражается в процентах или дробях (как простых, так и десятичных). Иначе говоря, это отношение распределения внутри изучаемого явления, рассчитываемое по формуле:

$$\text{ОВСС} = \frac{\text{Показатель, характеризующий часть совокупности}}{\text{Показатель по всей совокупности в целом}}$$

Вся совокупность принимается за 100% (или за единицу), а ее составные части показывают степень распространенности интересующего нас признака. Примером может служить показатель доли городского населения в регионе, несовершеннолетних среди населения города и т.д.

*Частным случаем структуры преступности является ее география*, под которой понимается распределение преступности по регионам (территориям) страны – субъектам Российской Федерации.

Наряду с соотношением части и целого (показателями структуры) определяют соотношение между двумя частями одного целого. *Относительные величины, характеризующие соотношение между частями одного целого, называют относительными величинами координации (ОВК):*

$$\text{ОВК} = \frac{\text{Показатель, характеризующий } n\text{-ю часть совокупности}}{\text{Показатель, характеризующий часть совокупности, выбранной в качестве базы сравнения}}$$

ОВК показывают, в какой степени одна из величин больше или меньше другой.

К таким величинам относятся, например, показатели, характеризующие соотношение между численностью городского и сельского населения, между численностью рабочих и служащих и т.д.

Существенное значение имеют *отношения, характеризующие выполнение плана (ОВП)*. Технология вычисления их проста:

$$\text{ОВП} = \frac{\text{Показатель, достигнутый в } n\text{-м периоде}}{\text{Показатель, планируемый на } n\text{-й период}} \times 100$$

ОВП выражают уровень выполнения планового задания по данному показателю в виде отношения фактической (отчетной) величины показателя к запланированной на тот же период его величине.

*Относительные величины сравнения (ОВСр)* – отношение величины, характеризующей данный объект, к одноименной величине по другому аналогичному объекту:

$$\text{ОВСр} = \frac{\text{Показатель, характеризующий объект А}}{\text{Показатель, характеризующий объект Б}}$$

Обычно их исчисляют в процентах или кратных отношениях, показывающих, во сколько раз одна из сравниваемых величин больше (или меньше) другой. Это может быть отношение как пространственного, так и временного сравнения.

При помощи ОВСр сопоставляются показатели по разным странам, регионам, предприятиям и т.д.

При вычислении таких величин необходимо, как уже отмечалось, позаботиться, чтобы сравниваемые показатели определялись по единой методике построения, были сравнимы по единицам измерения и во всех других отношениях.

## ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

**Задание 1.** На основе представленных данных вычислите удельный вес количества зарегистрированных преступлений, связанных с НОН, в общем количестве преступлений, зарегистрированных правоохранительными органами России<sup>1</sup>:

Показатель	2012	2013	2014	2015
Всего зарегистрировано преступлений	2 302 168	2 206 249	2 190 578	2 388 476
Количество зарегистрированных преступлений, связанных с НОН	218 974	231 462	254 730	236 939

**Задание 2.** На основе следующих данных вычислите коэффициенты преступности и криминальной активности населения:

Показатель	2011	2012	2013	2014	2015
Численность постоянного населения старше 14 лет на 1 января (тыс. человек)	121 154	121 096	120 848	120 530	120 284
Количество преступлений, связанных с НОН	215 214	218 974	231 462	254 730	
Число лиц, совершивших преступления, связанные с НОН	109 144	115 214	117 912	123 300	

*Примечание.* Прежде всего, обратите внимание на сопоставимость представленных показателей. Например, численность постоянного населения дана на конкретную дату, такие показатели называются моментными, а количество зарегистрированных преступлений – за год, такие показатели называются интервальными. Очевидно, что использование таких данных без предварительной обработки может привести к получению результатов искаженных или лишенных смысла.

Поэтому необходимо перейти от моментных значений численности населения на 1 января к интервальным значениям за год в целом. Для этого используется формула:

<sup>1</sup> Данные официального сайта Единой межведомственной информационно-статистической системы. URL: <http://www.fedstat.ru>.

где  $S_{НГ}$  – численность населения на 1 января текущего года (на начало года);

$S_{КГ}$  – численность населения на 1 января года, следующего за текущим (на конец текущего года).

**Задание 3.** По данным, характеризующим распределение числа несовершеннолетних осужденных по возрасту, полу и социальному положению в городе N

Показатель	Число осужденных за преступления, связанные с НОН				
	2011	2012	2013	2014	2015
Всего	850	900	1050	1115	1120
из них:					
по полу					
мужчин	450	600	550	605	700
женщин	400	300	500	510	420
по возрасту:					
14—15 лет	200	250	200	300	250
16—17 лет	650	650	850	815	870
по социальному положению					
учащихся	450	600	500	605	500
работающих	200	150	250	210	220
неработающих и не учащихся	200	150	300	300	400

вычислите:

– ОВСС каждой группы, сделайте вывод.

– ОВК женщин по отношению к мужчинам, сделайте вывод.

– ОВК учащихся, по отношению к работающим, сделайте

вывод.

**Задание 4.** Охарактеризуйте состав обучающихся своей группы по полу, возрасту и рассчитайте относительные величины структуры.

**Задание 5.** По данным статистики МВД России, зарегистрированные в 2014 году преступления, связанные с незаконным оборотом наркотиков, на территории Российской Федерации характеризуются следующими показателями:

Количество преступлений, связанных с незаконным оборотом наркотиков	Удельный вес от всех пре- ступлений, %
Всего	100
Незаконное приобретение, хранение, перевозка, изготовление, переработка наркотиков (ст. 228 УК РФ)	44,9
Незаконный оборот сильнодействующих веществ (ст. 234 УК РФ)	1,5
Легализация денежных средств, полученных преступным путем (ст. 174, 174.1 УК РФ)	0,2
Незаконное производство, сбыт или пересылка наркотиков (ст. 228.1 УК РФ)	50,7
Контрабанда наркотиков (ст. 188, 226.1, 229.1 УК РФ)	0,8
Организация либо содержание притонов (ст. 232 УК РФ)	1,1
Иные наркопреступления	0,8

Рассчитайте количество преступлений по соответствующим категориям, если известно, что в 2014 г. зарегистрированы 254 730 преступлений, связанных с незаконным оборотом наркотиков.

**Задание 6.** На основании следующих данных о числе раскрытых преступлений и уровне раскрываемости преступлений, связанных с НОН, по городу N за истекший месяц определите средний уровень раскрываемости преступлений по городу:

Район	Число раскрытых преступлений	Раскрываемость преступлений, %
Центральный	97	46,1
Восточный	120	29,5
Южный	150	64,8
Западный	204	50,9

**Задание 7.** По информации, полученной вследствие реализации плана мероприятий региональной антинаркотической программы некоторого региона, рассчитайте процент выполнения плана мероприятий по целевым подпрограммам и плана мероприятий в целом по программе:

№ п/п	Наименование подпрограммы	Запланировано мероприятий	Фактически проведено мероприятий	% исполнения
1	Обеспечение правопорядка и профилактики правонарушений	15	12	
2	Комплексные меры профилактики немедического потребления наркотиков	22	18	
3	Профилактика безнадзорности и правонарушений несовершеннолетних	15	15	
4	Система мер по защищенности населения и территорий	15	16	
	Итого			

**Задание 8.** На основе представленных данных, сравните (определите ОВСр) количество преступлений, связанных с незаконным оборотом наркотических средств, в республиках Казахстан и Беларусь. Проанализируйте динамику полученного показателя, сделайте вывод.

Государства-члены ОДКБ	Год					
	2009	2010	2011	2012	2013	2014
Республика Беларусь	4522	4655	4532	4207	5008	7356
Республика Казахстан	9705	8795	4360	3886	3639	3548

### 3.3. СРЕДНИЕ ВЕЛИЧИНЫ

Вариационный ряд содержит достаточно полную информацию об изменчивости (вариации) признака. Однако обилие числовых данных, с помощью которых он задается, усложняет их использование. В то же время на практике часто оказывается достаточным знание лишь сводных характеристик вариационных рядов: средних или характеристик центральной тенденции; характеристик изменчивости (вариации) и др.

*Средняя величина* – это обобщающий показатель, выражающий типичные размеры количественно варьирующихся признаков (возраста, стажа работы, числа судимостей и т.д.) качественно однородных массовых общественных явлений и процессов. В средней величине выражается обобщенное, типичное для данной совокупности значение признака, погашаются случайные отклонения, присущие конкретным единицам совокупности, и таким образом проявляется действие закона больших чисел.

Средние величины широко используются при изучении социально-экономических и социально-правовых процессов, отражающих результаты деятельности государства, органов и учреждений, общественных структур (например, средние темпы роста и прироста объема преступности или раскрываемости, изменение структуры системы профилактики и другие).

Наиболее распространенным видом средних величин в статистике является средняя арифметическая, представляющая собой частное от деления суммы индивидуальных значений признаков на их количество.

Средняя арифметическая величина бывает *простой и взвешенной*.

**Пример 1.** Допустим, что количество правонарушений по 10 населенным пунктам региона за отчетный период составило: 3100, 3600, 4500, 4900, 5300, 5400, 5600, 5700, 5900, 6000. Набор цифр не позволяет получить представление об уровне преступности по всему региону. Для этого надо определить средний уровень преступности. Его в данном примере можно определить так: суммируем состояние преступности по всем населенным пунктам, и полученный итог разделим на число населенных пунктов в обследуемом регионе. Это составит следующее количество преступлений:

---

Среднегодовой уровень преступности по региону составляет 5000 преступлений. Это *простая* средняя арифметическая величина. Простой она называется потому, что ис-

числяется простым суммированием индивидуальных значений признака и делением этой суммы на число значений.

Отдельные значения признака единицы статистической совокупности, отличные от значений его у других единиц (в нашем примере уровень преступности по каждому населенному пункту – 3100, 3600, 4500 и т.д.), в статистике принято называть вариантами и обозначать символом  $x_i$  где  $i$  может принимать любые значения от 1 до  $n$ . При наличии в совокупности  $n$  вариантов последний из них обозначается  $x_n$ . В нашем примере  $n = 10$ , таким образом,  $x_1 = 3100$ ,  $x_2 = 3600$ , ...,  $x_{10} = 6000$ . Величина среднего варианта обозначается  $\bar{x}$ , а знак суммы вариантов – через  $\Sigma$ .

Расчет **средней арифметической простой** можно записать в виде следующей формулы:

$$\bar{x} = \frac{\sum x_i}{n}$$

где  $x_i$  – значение вариантов,  $\bar{x}$  – средняя арифметическая.

Средняя величина исчисляется по формуле средней арифметической простой тогда, когда значения вариантов встречаются по одному или по одинаковому числу раз, т.е. когда повторяемость всех вариантов одинакова.

Если же отдельные значения признака повторяются неодинаковое число раз, то средняя величина определяется по формуле средней арифметической взвешенной.

**Пример 2.** Имеются данные о состоянии преступности в различных населенных пунктах:

Количество совершенных преступлений	2500	4500	5000	6000	8000
Количество населенных пунктов	2	6	12	16	14

В данном примере только в двух населенных пунктах состояние преступности за рассматриваемый период составило 2500 фактов, в шести – по 4500 фактов и т.д., т.е. каждый вариант повторяется неодинаковое число раз. Поэтому при исчислении среднего уровня преступности нельзя пользоваться формулой средней арифметической простой. Чтобы определить уровень преступности по исследуемым населенным пунктам, сначала надо определить общее состояние преступности для всех населенных пунктов. Для этого ум-

ножают по каждой группе количество преступлений на число населенных пунктов и полученные произведения суммируют. Средний уровень преступности по всей совокупности населенных пунктов определяется делением полученной суммы на число населенных пунктов. В нашем примере:

---

Числа, которые показывают, сколько раз (или как часто) повторяются конкретные значения признака статистической совокупности, в статистике принято называть частотами. В нашем примере частотами являются 2, 6, 12, 16, 14. Они показывают, что количество преступлений, равное 2500 фактам, встречается в двух населенных пунктах, т.е. 2 раза, 4500 фактам – 6 раз и т.д. Частоты еще называют весами средней, отсюда и происходит название *средней взвешенной*.

Обозначив условно частоты буквой  $f$  расчет **средней арифметической взвешенной** можно выразить следующей формулой:

---

где  $x_i$  – значение вариантов,  $\bar{x}$  – средняя арифметическая,  $f_i$  – значение весов (частот).

Таким образом, при исчислении средней арифметической взвешенной величины все варианты надо умножить на их частоты, тогда каждая величина  $x_i$ , будет участвовать в образовании средней пропорционально своему значению.

На практике иногда встречается необходимость вычисления средней величины не из конкретных численных значений изучаемого признака, а из значений признака, сгруппированных в интервалы («от – до»). Тогда в качестве  $x_i$  выбирают СЕРЕДИНЫ соответствующих интервалов.

**Упражнение 1.** По имеющимся данным о возрасте лиц, совершивших преступления, связанные с НОН в городе N в прошлом году (данные условные), рассчитайте средний возраст указанных лиц (все вычисления проведите в MS Excel).

Так как исследованию подвергается сгруппированный вариационный ряд, то все показатели вариации будут взве-

шенными. В качестве весов будет выступать число выявленных лиц в соответствующей возрастной группе, совершивших преступление ( $f_i$ ).

Возрастная характеристика лиц, совершивших преступления, связанные с НОН в городе N в прошлом году (данные условные):

Группы по возрасту, лет	Середины интервалов ( $x_i$ )	Число выявленных лиц, совершивших преступление ( $f_i$ )
14-17	15,5	2 203
18-29	23,5	4 606
30 и старше <sup>1</sup>	35,5	1 298
Итого		8 107

Для удобства вычисления к исходной таблице справа добавьте столбец, в котором рассчитайте произведение  $x_i \times f_i$ :

Группы по возрасту, лет	Середины интервалов ( $x_i$ )	Число выявленных лиц, совершивших преступление ( $f_i$ )	$x_i \times f_i$
14-17	15,5	2 203	34 146,5
18-29	23,5	4 606	
30 и старше <sup>1</sup>	35,5	1 298	
Итого		8 107	188 467

В строке *Итого* рассчитайте сумму по третьему ( $\sum f_i$ ) и четвертому ( $\sum x_i \times f_i$ ) столбцам (для проверки правильности выполнения упражнения сравните полученные расчетные значения с представленными в таблице). В отдельной ячейке рассчитайте средний возраст.

---

<sup>1</sup> В тех случаях, когда ограничения интервала нет, его величина принимается равной предыдущему интервалу.

### 3.4. ПОКАЗАТЕЛИ ВАРИАЦИИ

Следующим этапом изучения вариации признака в совокупности является измерение характеристик силы, значения вариации, установления типичности или показательности средней, т.е. насколько точно характеризует средняя данную совокупность по определенному признаку. Другими словами, типичность средней должна показать, насколько однородна масса, которая характеризуется этой средней.

Простейшей из таких характеристик может служить *размах вариации, или амплитуда вариации* – абсолютная разность между максимальным и минимальным значением признака из имеющихся в изучаемой совокупности. Таким образом, размах вариации вычисляется по формуле:

$$R = x_{\max} - x_{\min}.$$

Размах вариации обладает и определенными недостатками, поскольку зависит от многих случайных причин и при повторных наблюдениях может резко менять свое значение. Главный недостаток размаха вариации заключается в том, что он не показывает, насколько велики отклонения от вариантов признака внутри него. Таким образом, размах вариации не может служить основным показателем меры вариации наблюдаемого признака.

Более точную характеристику вариации признака можно получить, если сравнить все имеющиеся значения с их средней величиной. Поскольку в этом случае отклонений столько же, сколько и вариантов, следует отыскивать их среднюю величину. Такими более точными показателями вариации статистической совокупности являются среднее линейное отклонение, среднее квадратическое отклонение и дисперсия.

**Дисперсией**  $\sigma^2$  вариационного ряда называется средняя арифметическая квадратов отклонений вариантов от их средней арифметической:

простая:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n};$$

взвешенная:

$$\sigma^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 f_i}{\sum f_i}.$$

Среднее квадратическое отклонение  $\sigma$  служит мерой рассеивания значений изучаемого признака около его среднего значения.

Среднее квадратическое отклонение находится как корень квадратный из дисперсии:

простой:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}};$$

взвешенной:

$$\sigma = \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{x})^2 f_i}{\sum f_i}}.$$

Аналогичные расчеты производят по интервальному ряду распределения. Вместо периодов времени можно использовать различные регионы, области и т.п., и дать на основе анализа вариации сравнительную характеристику признака.

**Упражнение 2.** По данным из упражнения 1 рассчитайте основные показатели вариации (*все вычисления проведите в MS Excel*):

Так как исследованию подвергается сгруппированный вариационный ряд, все показатели вариации будут взвешенными. В качестве весов будет выступать число выявленных лиц в соответствующей возрастной группе, совершивших преступление ( $f_i$ ).

Для удобства вычисления к исходной таблице, получившейся в результате выполнения упражнения 1, справа добавьте три столбца, в которых рассчитайте необходимые промежуточные значения (см. таблицу ниже):

– в 5-ом столбце рассчитайте отклонения середин интервалов от среднего (вспомните, что средний возраст мы рассчитали в упражнении 1, и он равен 23,3 года) как разность между серединами интервалов и средним значением;

– в 6-ом столбце возведите значение, полученное в предыдущем столбце в квадрат;

– в 7-ом столбце значение, полученное в предыдущем столбце, умножьте на соответствующую частоту (значение в третьем столбце).

Среднее квадратическое отклонение в нашем примере будет равно:

$$\sigma = \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{x})^2 f_i}{\sum f_i}} = \sqrt{\frac{327\,409,1}{8107}} = \sqrt{40,39} = 6,35 \text{ лет.}$$

Группы по возрасту, лет	Середины интервалов ( $x_i$ )	Число выявленных лиц, совершивших преступление ( $f_i$ )	Произведение середины возрастных интервалов и частоты	Отклонения от среднего значения	Квадрат отклонений от среднего	Произведение квадрата отклонений от среднего на частоту
14-17	15,5	2 203	34146,5	-7,74738	60,02188	132228,2
18-29	23,5	4 606	108241	0,252621		
30 и старше <sup>1</sup>	35,5	1 298	46079			
Итого		8 107	188466,5			327386,6

Рассчитаем размах вариации  $R = x_{\max} - x_{\min} = 41 - 14 = 27$  лет.

На основе проведенных расчетов можно сделать выводы, что средний возраст лиц, совершивших указанные преступления, составляет 23,3 года. Рассматриваемая совокупность лиц, совершивших преступление по возрастному критерию, весьма разрозненна: от 14 до 41 года, что характеризуется размахом вариации  $R=27$  лет. В среднем возраст преступников колеблется в пределах  $23,3 \pm 6,35$  лет.

**Упражнение 3.** Проиллюстрируем расчет показателей вариации по данным о числе зарегистрированных преступлений, связанных с НОН в городе N (данные условные) за 2011–2015 гг. (все вычисления повторите в MS Excel):

<sup>1</sup> В тех случаях, когда ограничения интервала нет, его величина принимается равной предыдущему интервалу.

Зарегистрировано преступлений, связанных с НОН в городе N:

Год	Зарегистрировано преступлений, связанных с НОН ( $x_i$ )		
2011	1 404	-185,4	34 372,2
2012	1 820	230,6	53 176,4
2013	1 702	112,6	12 678,8
2014	1 544	-45,4	2 061,2
2015	1 477	-112,4	12 633,8
Итого	7 947	0	114 923,4

В расчетах будем использовать показатели вариации невзвешенные, так как в таблице даны несгруппированные данные.

Рассчитаем среднее количество преступлений:

$$\bar{x} = \frac{1404 + 1820 + 1702 + 1544 + 1477}{5} = \frac{7947}{5} = 1589,4$$

Рассчитаем размах вариации:

$$R = x_{\max} - x_{\min} = 1820 - 1404 = 416 \text{ преступлений.}$$

Среднее квадратическое отклонение рассчитывается как корень из дисперсии:

$$\sigma = \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{x})^2}{n}} = \sqrt{\frac{34372,2 + 53176,4 + 12678,8 + 2061,2 + 12633,8}{5}} = \sqrt{\frac{114923,4}{5}} = \sqrt{22984,7} = 151,6 \text{ преступлений.}$$

По произведенным расчетам можно сделать следующие выводы. В среднем за год совершаются 1 589,4 преступлений, связанных с НОН. Диапазон колебания зарегистрированных преступлений варьирует в пределах от 1 404 до 1 820 преступлений и оставляет величину 416 преступлений. Так как наиболее точным показателем вариации признака является среднее квадратическое отклонение, то можем считать, что в среднем на 151,6 преступлений ежегодно отклоняется количество зарегистрированных преступлений, связанных с НОН, от средней величины за рассматриваемый период.

**Медианой** вариационного ряда называется значение признака, приходящееся на середину ранжированного (расположенного в возрастающем порядке) ряда наблюдений. Медиана делит упорядоченный ряд пополам. По обе стороны от нее находится одинаковое число единиц ряда. Медиана обозначается символом  $Me$ .

Для дискретного вариационного ряда с нечетным числом членов медиана равна срединному элементу, а для ряда с четным числом членов – полусумме двух срединных элементов (см. пример 3).

**Модой** ( $\tilde{Mo}$ ) вариационного ряда называется значение элемента ряда наблюдений, которому соответствует наибольшая частота. Мода применяется в тех случаях, когда нужно охарактеризовать наиболее часто встречающуюся величину признака (см. пример 3).

Могут встречаться ряды наблюдений, в которых все значения имеют одинаковую частоту, тогда говорят, что моды нет. В других случаях не одно, а два значения могут иметь одинаковую наибольшую частоту. Тогда будут две моды, такое распределение называется бимодальным. Бимодальные распределения могут указывать на качественную неоднородность совокупности по исследуемому признаку.

**Пример 3.** Рассмотрим вариационный ряд, представляющий собой зарегистрированную органами МВД города N подростковую преступность за май прошлого года в возрастах:

16 13 15 13 15 17 14 15 15 14 16 14 14 14 14 14 13 16 17 16  
13 15 14 15 17

Ранжированный (упорядоченный) ряд будет иметь следующий вид:

13 13 13 13 14 14 14 14 14 14 14 14 15 15 15 15 15 16 16 16  
16 17 17 17 17

Сгруппировав его, получим таблицу:

Возраст	13	14 $Mo$	15 $Me$	16	17
Частота	4	8	6	4	3

Очевидно, в этом примере модой будет возраст, равный 14 ( $Mo=14$ ), так как этому значению возраста соответствует наибольшая частота.

Чтобы найти медиану дискретного вариационного ряда, необходимо найти номера элементов, которые находятся в

середине вариационного ряда. Для этого нужно сумму частот разделить пополам и к полученному результату добавить  $\frac{1}{2}$ . Так как выборка содержит 25 значений возрастов подростков, совершивших преступления, медианой будет значение, имеющее номер

$$25/2 + \frac{1}{2} = 12,5 + \frac{1}{2} = 13,$$

т.е. 13-ый элемент, который делит упорядоченный ряд пополам. Каково же значение 13-го элемента? Для того чтобы это выяснить, нужно отсчитать 13-е значение в ранжированном ряду:

13 13 13 13 14 14 14 14 14 14 14 14 14 **15** 15 15 15 15 16 16 16  
16 17 17 17 17

В случае сгруппированного вариационного ряда для определения медианы нужно накапливать частоты, начиная от наименьшего значения. В нашем случае сумма частот 1-го и 2-го варианта равна 12 (см. таблицу ниже). Ясно, что накопленной частоты равной 13 нет. Если прибавить к 12 частоту 3-го варианта, то получим сумму, равную  $12 + 6 = 18$ . Следовательно, 13-ый элемент соответствует третьему варианту признака, и медианой будет возраст, равный 15.

Возраст	13	14 <i>Mo</i>	15 <i>Me</i>	16	17
Частота	4	8	6	4	3
Накопленная частота	4	12	18		

Если у нас была четная сумма частот (например, отбросив последнее значение, получим выборку из 24 элементов), то, применяя указанную выше формулу, получим номер медианной варианты, равный

$$24/2 + \frac{1}{2} = 12,5.$$

Поскольку элементов с дробным номером не существует, полученный результат указывает, что медиана находится посередине между 12 и 13 элементами. Поэтому в этом случае медиана будет равна:

————— .

В случае если дан интервальный вариационный ряд, находят модальный и медианный интервалы. Вычисление моды и медианы при этом осуществляется по более сложной формуле, которая в данном учебном пособии не приводится. Желающие могут ее найти в литературе по правовой статистике.<sup>1</sup>

<sup>1</sup> Напр.: Правовая статистика : учебник для студентов вузов, обучающихся по специальности 030501 «Юриспруденция», для курсантов и

**Упражнение 4.** По данным таблицы, содержащей распределение уголовных дел по срокам рассмотрения в судебных заседаниях, определите моду и медиану:

Сроки рассмотрения в судебном заседании, дни	Число уголовных дел
1	25
2	70
3	85
4	80
5	60
6	40
7	40

Для сравнения колебаний разнородных явлений, разных по своему характеру и размерам признаков, используется относительный показатель вариации, так называемый **коэффициент вариации**.

Коэффициент вариации обозначается буквой  $V$ . Коэффициент вариации как относительный показатель может быть представлен в коэффициентном или в процентном выражении. Наиболее часто в практических целях применяют коэффициент вариации, который представляет собой процентное отношение среднего квадратического отклонения к средней арифметической величине:

—

Коэффициент вариации используется для характеристики однородности совокупности по исследуемому признаку. Совокупность считается однородной, если коэффициент вариации не превышает 40%. Соответственно, надежность и типичность средней такой совокупности является достаточно высокой.

При помощи коэффициента вариации также можно сравнивать между собой вариацию производительности труда, эффективность деятельности и т.д. в одном и том же или разных учреждениях, органах и предприятиях.

## ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

**Задание 1.** Для сгруппированного вариационного ряда, характеризующего распределение осужденных по возрастным группам в городе N:

Возраст осужденных ( $x$ )	Число осужденных (частота $f$ )	Средина возрастных интервалов ( $x_i$ )	Произведение середины возрастных интервалов и частоты ( )	Отклонения от среднего значения ( )	Произведение квадрата отклонений от среднего на частоту
14-15	43 844				
16-17	88 382				
18-29	475 707				
30-49	1 411 842				
50-70	51 268				
Итого					

Рассчитайте:

– средний возраст осужденных как взвешенное среднее арифметическое осужденных в различных возрастных группах;

– среднее квадратическое отклонение;

– коэффициент вариации.

Сделайте выводы.

**Задание 2.** Имеется распределение территориальных органов МВД России по числу совершенных на их территории преступлений за определенный период (данные условные):

Количество преступлений ( $x_i$ )	150	250	350	450	550	650	750
Количество территориальных органов ( $f_i$ )	1	2	5	9	9	6	2

Рассчитайте:

– среднее число преступлений на территории;

– медиану;

– моду;

- коэффициент вариации;
- сделайте выводы.

**Задание 3.** В ходе исследования рецидивной преступности была организована случайная выборка осужденных за наркопреступления. В результате ее обработки было получено следующее распределение по числу повторной судимости осужденных:

Кол-во повторных судимостей ( $x_i$ )	0	1	2	3	4	5
Кол-во заключенных ( $f_i$ )	1 000	2 000	1 200	400	200	200

Определите:

- среднее число повторных судимостей;
- размах вариации;
- медиану;
- моду;
- сделайте выводы.

**Задание 4.** При изучении дисциплины «Правовая статистика» студент Иванов получил следующие оценки:

3, 2, 3, 5, 3, 5, 3, 2, 3, 4, 3, 3, 3, 5, 5, 4, 2, 4, 2, 4, 5, 5, 2, 3, 4, 2, 5, 4, 5, 4, 2, 4, 3, 3, 2, 3, 2, 4, 4, 3, 2.

Выполните следующие задания:

- постройте гистограмму и полигон частот распределения оценок;
- определите средний балл студента, найдите разброс оценок около среднего балла;
- сделайте вывод о качестве освоения дисциплины.

**Задание 5.** По данным о производительности труда сотрудников двух следственных подразделений МВД России (данные условные):

Расследовано уголовных дел за месяц			
Таб. №	Первое подразделение	Таб. №	Второе подразделение
1	2	6	8
2	3	7	9
3	12	8	10
4	15	9	11
5	18	10	12

Рассчитайте по каждому подразделению:

- размах вариации;
- среднюю производительность труда (кол-во расследованных дел на одного сотрудника в месяц);
- среднее квадратическое отклонение;
- коэффициент вариации.

Сделайте вывод.

**Задание 6.** Для вариационного ряда возрастов осужденных по статье 228, расположенного в файле *Статистические данные.xlsx* (созданного на прошлых занятиях), рассчитайте:

- средний возраст осужденных;
- размах вариации;
- моду и медиану;
- коэффициент вариации.

Сделайте выводы.

### **3.5. Точность, ошибки и методы контроля данных СТАТИСТИЧЕСКОГО НАБЛЮДЕНИЯ**

Точность данных – важнейшее требование, предъявляемое к статистическому наблюдению. Чем ближе значение показателей, полученных в результате статистического наблюдения, к фактическим их значениям, тем выше точность статистического наблюдения. Иными словами, *точность* характеризует степень достоверности информации, ее приближение к оригиналу, который она выражает. Отсюда *степень соответствия величины какого-либо показателя (значения какого-либо признака), определенной по материалам статистического наблюдения, действительной его величине называют точностью статистического наблюдения.*

*Расхождение между расчетным и действительным значением изучаемых величин, полученных при наблюдении, принято называть ошибками наблюдения.*

Чтобы избежать ошибок наблюдения, минимизировать их, необходимо тщательно составить инструментарий наблюдения, обеспечить качественное обучение и инструктаж персонала, который будет проводить наблюдение. Однако, учитывая массовый характер статистического наблюдения и

сложность его содержания, необходимо наряду с этим организовать специальные частные или сплошные контрольные проверки правильности заполнения статистических формуляров, провести логический и арифметический контроль полученных данных после окончания сбора информации.

В зависимости от причин возникновения ошибки наблюдения подразделяются на *три вида*: а) ошибки измерения; б) ошибки репрезентативности; в) ошибки регистрации.

*Ошибки измерения* ( $\varepsilon$ ) – разность между результатом измерения значения явления ( $x$ ) и истинным значением ( $X$ ):  $\varepsilon = x - X$ . Как и последующие ошибки, ошибки измерения объединяются в две группы: 1) систематические, которые при данных условиях проведения измерения имеют вполне определенное постоянное значение (например, ошибка измерительного прибора); 2) случайные, которые являются результатом взаимодействия большого числа незначительных в отдельности факторов и имеют в каждом случае различные значения.

*Ошибки репрезентативности* – расхождения между показателями выборочной и генеральной совокупности, возникшие вследствие того, что выборочная совокупность недостаточно полно воспроизводит (репрезентирует) совокупность генеральную, хотя установление и регистрация фактов были произведены точно. Они присущи только несплошному наблюдению (сплошным называется такое наблюдение, при котором обследованию подвергаются все без исключения единицы изучаемой совокупности; примером такого наблюдения являются переписи, при которых по основной программе обследованию подлежит все население страны без исключения. Несплошное – это такое наблюдение, при котором обследованию подвергаются не все единицы совокупности, а только часть их).

*Ошибки регистрации* представляют собой расхождение между сведениями, записанными в формуляры наблюдения (документы первичного учета и статистической отчетности), и фактическим положением в исследуемой совокупности. Они могут быть как в сплошном, так и в несплошном наблюдении. Ошибки регистрации могут быть случайными и систематическими. *Случайные* – это следствие невнимательности, небрежности при заполнении формуляра наблюдения или подсчете итогов.

*Систематические ошибки* подразделяются на преднамеренные и непреднамеренные.

*Преднамеренные* ошибки – сознательные, тенденциозные искажения. К ним относятся нарушения порядка представления государственной статистической отчетности, выразившиеся в непредставлении отчетов и других данных, необходимых для проведения статистических наблюдений, искажении отчетных данных или нарушении сроков представления отчетов, допускаемые некоторыми недобросовестными руководителями при составлении отчетности.

*Непреднамеренные* ошибки – неточность, возникающая в процессе статистического наблюдения в установлении или регистрации изучаемых фактов помимо воли и желания лица, производящего наблюдение или сообщаящего сведения (т.е. неумышленно, вследствие пропусков в записях отдельных фактов и т.п.).

Полученные в результате наблюдения данные до их обработки подвергаются самой тщательной проверке, контролю. Контроль достоверности осуществляется на всех этапах сбора и обработки статистических данных: при заполнении формуляров статистического наблюдения, составлении отчетов, разработке аналитических материалов. Прежде всего осуществляется так называемый внешний контроль. При этом проверяется правильность оформления документов, т.е. наличие всех необходимых записей (реквизитов), предусмотренных инструкцией, а также полнота материала и охвата всех отчетных единиц наблюдения.

Следующим этапом контроля точности информации является *счетный (арифметический) контроль*. Он основывается на использовании количественных связей между значениями различных показателей отчета. Задача такого контроля – исправление итогов и отдельных числовых показателей. Для этих целей во всех формах статистических отчетов правоохранительных органов и органов юстиции по многим показателям установлены контрольные арифметические равенства.

*Логический (смысловой) контроль*, так же как и арифметический, основывается на знании взаимосвязей между показателями, но не количественных, а логических, в процессе проверки ответов на вопросы программы наблюдения путем

их сопоставления между собой или сравнения полученных данных с другими источниками по этому же вопросу.

Сущность выборочного обследования заключается в том, что исследованию подвергается только часть интересующей нас генеральной совокупности, а полученные результаты служат характеристикой всех единиц совокупности.

Основная проблема в теории выборочных исследований – решение вопроса о правомерности распространения на всю генеральную совокупность тех выводов, которые будут получены при анализе выборочной совокупности. Правомерность такого распространения во многом зависит от объема выборки.

Одним из первых вопросов, который встает при проведении социологического исследования, является установление числа обследуемых объектов (например, правонарушителей), или *репрезентативного объёма выборки*. Такое число, с одной стороны, должно быть минимальным, но вместе с тем достаточным для того, чтобы исследование было показательным, т.е. обладающим достоверностью выводов об изучаемом явлении. Репрезентативность в качественном отношении означает достаточно полное приближение параметров выборки к характеристикам генеральной совокупности. Поэтому в процессе отбора необходимо стремиться к максимально возможному учёту особенностей изучаемого явления. Выборка, достаточно точно воспроизводящая генеральную совокупность, называется репрезентативной (представительной).

Статистикой выработаны определённые формулы вычисления репрезентативного объёма выборки.

Объем  $n$  случайной выборки для измерения среднего значения количественного признака, определяется по формуле:

$$n = \frac{\sigma^2 t^2}{\Delta^2},$$

где  $\sigma^2$  – дисперсия исследуемого признака в генеральной совокупности;

$t$  – коэффициент доверия;

$\Delta$  – предельная ошибка выборки.

Если же выборочно измеряют другую характеристику – долю изучаемого признака, то пользуются формулой:

$$n = \frac{P(1-P)t^2}{\Delta^2},$$

где  $P$  – доля признака.

Ниже приведены некоторые значения  $t$ :

Коэффициент доверия $t$	1,5	1,96	2,0	2,18	2,5	2,58	3,0	3,5
Уровень доверительной вероятности, %	86,6	95	95,4	97	98,8	99	99,7	99,9

Практика исследований даёт основание полагать, что в большинстве случаев достаточная достоверность выводов исследования обеспечивается при уровне доверительной вероятности 95%-99%.

Одним из основных вопросов выборочного исследования заключается в том, насколько выборочная средняя отличается от так называемой генеральной средней, т.е. как велика *ошибка репрезентативности*.

Для определения *предельной ошибки репрезентативности*  $\Delta$  рекомендуется пользоваться следующими двумя формулами:

1) при определении среднего размера изучаемого количественного признака:

$$\Delta = \sqrt{\frac{\sigma^2 t^2}{n}} = \frac{\sigma \cdot t}{\sqrt{n}};$$

2) при определении доли качественного признака:

$$\Delta = \sqrt{\frac{P(1-P)}{n}} \cdot t,$$

т.е. выборочная средняя будет лежать в пределах  $\bar{x} - \Delta \leq \bar{x} \leq \bar{x} + \Delta$ , где:

$\Delta$  – предельная ошибка репрезентативности;

$\sigma$  – среднее квадратическое отклонение;

$n$  – число единиц, попавших в выборку;

$t$  – коэффициент доверия;

$P$  – доля данного качественного признака в выборке;

$(1 - P)$  – доля противоположного признака.

**Упражнение 5.** В городе N из 11 000 студентов вузов добровольно сдали анализ на наркотики 550 человек. У 20 человек обнаружены следы употребления наркотических

средств. Определите с доверительной вероятностью 99,7% возможное количество студентов, употребляющих наркотические средства и психотропные вещества (НС и ПВ).

Создайте новый лист с именем **Доверительное оценивание**. Определите долю студентов, у которых выявлены следы НС : —.

Затем, выбрав соответствующее заданной вероятности достоверности значение  $t$ , рассчитайте значение  $\Delta$ .

Доверительный интервал определяет долю студентов, употребляющих НС и ПВ, от общего количества студентов:  $P-\Delta \leq P \leq P+\Delta$ . В результате вычислений у вас должно получиться следующее:  $0,012 \leq P \leq 0,06$ .

Возможное количество студентов  $K$ , употребляющих НС, можно рассчитать следующим образом:  $11\ 000 \times 0,012 \leq K \leq 11\ 000 \times 0,06$ .

### **ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ**

**Задание 1.** С целью определения средней фактической продолжительности рабочего дня в некотором территориальном органе МВД России с численностью сотрудников 480 человек была проведена 25-процентная случайная выборка. По результатам наблюдения оказалось, что у 10% обследованных потери времени достигали более 45 минут в день. С вероятностью 95,4% установите пределы, в которых находится генеральная доля сотрудников территориального органа МВД России с потерями времени более 45 минут в день.

*Указания:*

- 1) определите, сколько человек было выбрано для проведения статистического исследования;
- 2) определите предельную ошибку репрезентативности;
- 3) определите пределы, в которых находится интересующая доля сотрудников территориального органа.

**Задание 2.** При изучении 200 уголовных дел, связанных с НОН, отобранных из общего числа возбужденных в случайном порядке, оказалось, что 20% были необоснованно прекращены.

С вероятностью 0,954 определите пределы ( $P \pm \Delta\%$ ), в которых находится доля необоснованно прекращенных дел в общем числе возбужденных уголовных дел.

**Задание 3.** На основе обследования 200 осужденных по статье 234 (незаконный оборот сильнодействующих веществ) установлено, что 60% из них совершили преступление в состоянии наркотического опьянения. Определите предельную ошибку определения этого показателя с вероятностью 95%, 97%, 99,9%.

**Задание 4.** Определите, сколько нужно обследовать осужденных за нарушение правил дорожного движения и эксплуатации транспортных средств, чтобы выяснить среди них долю лиц, совершивших ДТП в состоянии наркотического опьянения. Предельная ошибка при исследовании должна быть не более 3% с доверительной вероятностью 0,97. На основе предварительного ознакомления с различными материалами (статистическими отчетами, судебной практикой, показателями прошлых обследований и т.д.) следует исходить из ожидаемой доли примерно 40%.

**Задание 5.** Для изучения общественного мнения о работе правоохранительных органов в порядке случайного отбора были опрошены 1500 человек, или 1% общей численности городского населения. Из числа опрошенных 340 человек положительно оценили работу правоохранительных органов. С вероятностью 0,997 определите пределы, в которых находится количество лиц, положительно оценивающих работу правоохранительных органов.

**Задание 6.** В результате выборочного обследования лиц, совершивших разбой в регионе, получены следующие данные о возрасте преступников:

Возраст преступников, лет	До 18	18-23	23-28	28-33	Итого
Число совершенных преступлений	11	54	42	23	130

Определите:

– с вероятностью 0,997 пределы среднего возраста преступников;

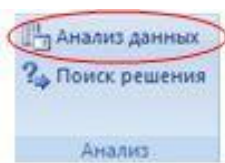
– с вероятностью 0,954 пределы доли преступлений, совершенных в возрасте до 18 лет.

## 4. ВОЗМОЖНОСТИ АВТОМАТИЗАЦИИ ОБРАБОТКИ СТАТИСТИЧЕСКИХ ДАННЫХ В MS EXCEL

### 4.1. ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Задача описательной статистики заключается в том, чтобы с использованием математических инструментов свести сотни значений выборки к нескольким итоговым показателям, которые дают представление о выборке. В качестве таких статистических показателей используются: среднее, медиана, мода, стандартное (среднеквадратическое) отклонение и др., рассмотренные в предыдущих разделах.


Возможности MS Excel позволяют автоматизировать процесс обработки статистических данных. Для этого в электронных таблицах предусмотрены специальные статистические функции и инструменты, которые сгруппированы в специализированный пакет анализа.



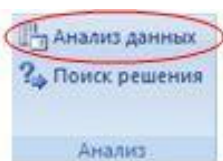
Проверьте, подключен ли **Пакет анализа** на вашем компьютере: перейдите на вкладку меню **Данные** и найдите кнопку **Анализ данных** в группе **Анализ**. Если такой кнопки нет, то ее нужно установить. Для этого щелкните по кнопке **Office** ⇒ **Параметры Excel** ⇒ **Надстройки** ⇒ **Управление надстройками Excel** (внизу окна) ⇒ **Перейти** ⇒ поставьте галочку слева от **Пакет анализа** ⇒ **ОК**.

Функции анализа данных можно применять только на **одном** листе. Если анализ данных проводится в группе, состоящей из нескольких листов, то результаты будут выведены на первом листе, на остальных листах будут выведены пустые диапазоны, содержащие только форматы. Чтобы провести анализ данных на всех листах, повторите процедуру для каждого листа в отдельности.

**Упражнение 1.** В качестве исходной статистической информации рассмотрим данные о количестве преступлений, связанных с незаконным оборотом наркотиков в Сибирском федеральном округе. Для этого на официальном сайте Федеральной службы государственной статистики ([www.gks.ru](http://www.gks.ru)) в разделе «Официальная статистика» найдите в разделе **Базы данных** ⇒ **ЦБСД** (Центральная база статистических данных)

данные по Сибирскому федеральному округу о количестве преступлений, связанных с незаконным оборотом наркотиков, начиная с 1993 года. С помощью кнопки **Выгрузить**  (в формате xls) сохраните выбранные данные в папку своей группы.

*Примечание.* Чтобы выбрать нужную территорию, в правой части окна щелкните по ссылке *Территории*, выберите Сибирский федеральный округ (можно задать в строке поиска, расположенной выше).



Запустите анализ данных: перейдите на вкладку меню *Данные* и щелкните по кнопке *Анализ данных* в группе *Анализ*. В окне *Анализ данных* выберите *Описательная статистика* и нажмите **ОК**. В поле *Входной интервал* укажите диапазон \$B\$2:\$B\$24. В ячейке B2 находится название федерального округа, поэтому установите флажок *Метки в первой строке*. В области *Параметры вывода* установите флажок *Выходной интервал* и в поле рядом укажите \$D\$3, а также установите флажок *Итоговая статистика* и нажмите **ОК**. При этом на лист Excel, начиная с ячейки D3, будут помещены данные итоговой статистики (рис.6):

	A	B	C	D	E	F
		Количество преступлений, связанных с незаконным оборотом наркотиков				
1		Сибирский федеральный округ				
2				Сибирский федеральный округ		
3	1993		11 073			
4	1994		12 673			
5	1995		14 148	Среднее	32784,63636	
6	1996		16 746	Стандартная ошибка	2144,912691	
7	1997		32 357	Медиана	36978,5	
8	1998		33 542	Мода	#Н/Д	
9	1999		36 879	Стандартное отклонение	10060,53229	
10	2000		42 070	Дисперсия выборки	101214310	
11	2001		42 771	Эксцесс	0,32020391	
12	2002		33 647	Асимметричность	-1,252549744	
13	2003		33 499	Интервал	31698	
14	2004		26 883	Минимум	11073	
15	2005		31 543	Максимум	42771	
16	2006		37 505	Сумма	721262	
17	2007		39 848	Счет	22	
18	2008		41 831			
19	2009		41 646			
20	2010		39 313			
21	2011		37 666			
22	2012		38 179			
23	2014		40 365			
24	2015		37 078			

Рисунок 6. Итоговая статистика

## 4.2. СТАТИСТИЧЕСКИЕ ФУНКЦИИ MS EXCEL

Функции категории *Статистические* предназначены в первую очередь для анализа диапазонов ячеек в MS Excel. С помощью данных функций можно вычислить наибольшее, наименьшее или среднее значение, подсчитать количество ячеек, содержащих заданную информацию, и т.д.

Данная категория содержит более 100 самых различных функций MS Excel, большая часть из которых предназначена исключительно для статистических расчетов и обычному рядовому пользователю чаще всего не нужна. Поэтому рассмотрим наиболее полезные и распространенные функции данной категории.

**Упражнение 2.** Для исходных данных, полученных в упражнении 1, рассчитаем основные показатели вариации. Для этого, начиная с ячейки G3, наберите таблицу, как на рисунке 7:

Количество значений		
Среднее		
Дисперсия		
Среднее квадратическое отклонение		
Максимум		
Минимум		
Размах вариации		
Мода		
Медиана		

Рисунок 7. Основные показатели вариации

В ячейках, начиная с H3, введите соответствующие формулы, в скобках функций укажите диапазон ячеек, в которых находятся значения количества преступлений по годам:

- в H3: =СЧЕТ( ) – подсчет количества значений;
- в H4: =СРЗНАЧ( ) – расчет среднего арифметического;
- в H5: =ДИСП( ) – расчет дисперсии;
- в H6: =СТАНДОТКЛОН( ) – расчет среднего квадратического отклонения;
- в H7: =МАКС( ) – расчет максимального значения;

- в Н8: =МИН( ) – расчет минимального значения;
- в Н9: =МАКС( ) – МИН( ) – расчет размаха вариации;
- в Н10: =МОДА( ) – расчет моды;
- в Н11: =МЕДИАНА( ) – расчет медианы.

### 4.3. ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ СГРУППИРОВАННОГО ВАРИАЦИОННОГО РЯДА

Для анализа дискретных и непрерывных величин используется их группировка. При построении сгруппированного вариационного ряда необходимо выбрать число интервалов и их длину. Для определения числа интервалов чаще всего используется формула Стержесса (Sturges):

$$k \approx 1 + 3,32Lg(n) = 1,44 Ln(n) + 1,$$

где  $n$  – количество исходных данных,  $k$  – количество интервалов. При вычислении величины  $k$  производится округление до большего целого.

После того как построен сгруппированный вариационный ряд, для его частотного анализа можно применить встроенный в MS Excel инструмент анализа данных – *Гистограмму*.

Данные для построения гистограммы должны быть представлены на рабочем листе в виде блока ячеек, содержащих исходные данные и блока ячеек, содержащих карманы (интервалы). Интервал карманов (отрезков) – блок ячеек, содержащих граничные значения интервалов для группировки исходных данных. Если интервал карманов не задается – он автоматически создается как набор равных отрезков в диапазоне от минимума до максимума исходных значений (по формуле Стержесса). Для каждого из интервалов, при проведении частотного анализа, определяется количество значений, попадающих в интервал. Оба блока могут располагаться как по столбцам, так и по строкам.

**Упражнение 3.** Для исходных данных, полученных в упражнении 1, проведите частотный анализ с использованием инструмента *Гистограмма*.

Выберите инструмент *Гистограмма (Анализ данных – Гистограмма)*. В появившемся окне в поле *Входной интер-*

вал укажем диапазон В3:В24. Поле **Интервал карманов** оставьте пустым – карманы будут созданы автоматически.

В области **Параметры вывода** установите вывод на текущем листе – в поле **Выходной интервал** укажите ячейку **К1** и установите флажки: **Интегральный процент** и **Вывод графика**. Нажмите кнопку **ОК** (если появится сообщение об ошибке, то пропустить – нажать на сообщении **ОК**).

На **Листе1**, начиная с ячейки **К1**, будет выведена таблица и гистограмма. Используя возможности: изменение ширины столбцов, изменение размеров диаграммы, следует отформатировать полученные данные для более удобного восприятия (рис. 8).

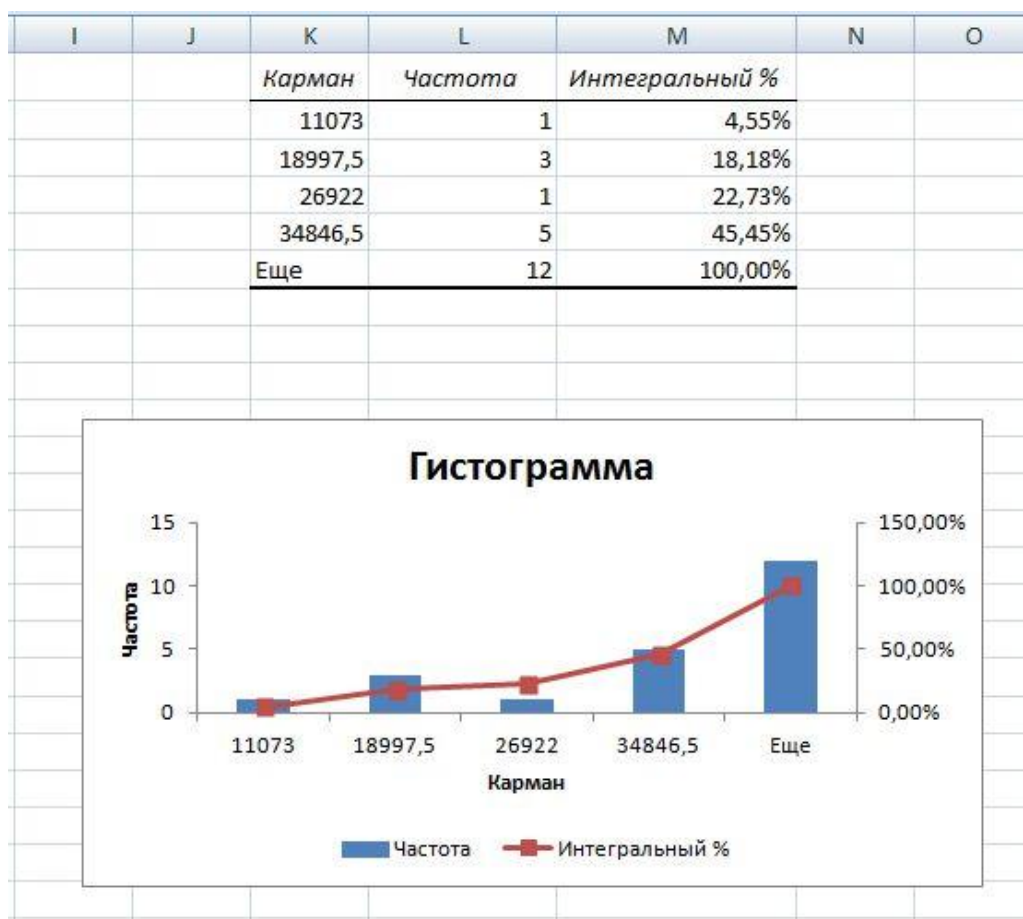


Рисунок 8. Результат работы инструмента Гистограмма

Рассмотрим полученные результаты. Таблица, выведенная на **Листе1** в ячейках **К1:М6**, содержит 3 столбца. Столбец **Карман** соответствует интервалам, на которые был автоматически разбит набор исходных значений. Столбец **Частота** содержит количество значений, попавших в интер-

вал. Столбец **Интегральный%** содержит значения накопленной частоты, выраженный в процентах.

График *Гистограмма* отражает информацию трех столбцов таблицы.

В случае если ваше представление о количестве и размере интервалов не совпадают с тем, как предлагает инструмент *Гистограмма*, можно самостоятельно разбить диапазон значения на интервалы и указать их в соответствующей опции инструмента *Гистограмма*.

### **ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ**

**Задание 1.** Скачайте с Центральной базы статистических данных (ЦБСД) данные по вашему родному федеральному округу о количестве преступлений, связанных с незаконным оборотом наркотиков, начиная с 1993 года (зайдите на официальный сайт Федеральной службы государственной статистики [www.gks.ru](http://www.gks.ru), в разделе «Официальная статистика» найдите в разделе *Базы данных* ⇒ *ЦБСД*). Если вы из Сибирского федерального округа, то уточните у преподавателя, какие данные вам необходимо скачать. С помощью статистических функций MS Excel и инструмента *Описательная статистика* рассчитайте основные показатели вариации (как в упражнении 1 и 2).

**Задание 2.** Для исходных данных, полученных в задании 1, проведите частотный анализ с использованием инструмента *Гистограмма*.

## 5. ДИНАМИЧЕСКИЕ РЯДЫ

### 5.1. ДИНАМИЧЕСКИЙ РЯД И ЕГО ГРАФИЧЕСКОЕ ПРЕДСТАВЛЕНИЕ

Анализ временных рядов представляет собой самостоятельную, весьма обширную и одну из наиболее интенсивно развивающихся областей математической статистики.

*Процесс развития, движения социальных явлений во времени в статистике принято называть динамикой.* Для отображения динамики строят ряды динамики (хронологические, временные), которые представляют собой ряды изменяющихся во времени значений статистического показателя, расположенных в хронологическом порядке. Их составными элементами являются цифровые значения данного показателя и периоды или моменты времени, к которым они относятся.

При изучении временных рядов ставятся следующие цели:

- дать представление об изменениях показателей за прошедший период;
- охарактеризовать интенсивность отдельных изменений в уровнях ряда от периода к периоду или от даты к дате;
- определить средние показатели временного ряда за тот или иной период;
- выявить основные тенденции и закономерности изменений изучаемого явления на отдельных этапах развития общества и в целом за рассматриваемый период;
- выявить факторы, обуславливающие изменение явления во времени;
- осуществить прогноз о возможном характере изменения явления в ближайшем и отдаленном будущем.

Важнейшая характеристика рядов динамики – их уровень. *Уровень в статистике – это размер (объем, величина) того или иного явления, достигнутый в определенный период или к определенному моменту.*

В тех случаях, когда уровни выражены абсолютными показателями, различают интервальные и моментные ряды.

*Интервальным* называется такой ряд, абсолютные уровни которого характеризуют величину изучаемого явления, полученную в итоге за определенный период времени.

*Моментным* называется ряд, абсолютные уровни которого характеризуют величину явления по состоянию на определенные моменты времени.

Таким образом, значения показателей интервального ряда можно складывать. Значения показателей же моментного ряда складывать нельзя.

Основным требованием, предъявляемым к построению любых динамических рядов, является сопоставимость их уровней. Это требование должно обеспечиваться выбором одинакового интервала, единством методологии учета и расчета показателей.

**Упражнение 1.** В таблице приведены данные по численности населения и числу зарегистрированных преступлений в России за период 2000–2014 гг. По данным таблицы оформите временной ряд и дайте его графическое представление.

Для этого следует:

1. В вашем рабочем файле *Статистические данные.xlsx* добавьте новый лист и переименуйте его в **Динамика**.
2. Постройте таблицу динамического ряда.

Год	Численность населения, тыс. чел.	Число зарегистрированных преступлений, связанных с незаконным оборотом наркотиков, тыс.
2000	146 890	243,6
2001	146 304	241,6
2002	145 649	189,6
2003	144 964	181,7
2004	144 168	150,1
2005	143 474	175,2
2006	143 236	212
2007	142 221	231,2
2008	142 009	232,6
2009	141 904	238,5
2010	141 914	222,6
2011	142 865	215,2
2012	143 056	219
2013	143 347	231,5
2014	143 667	254,7

3. Постройте диаграмму изменения численности населения и числа зарегистрированных преступлений. Для того

чтобы график численности населения откладывался по вспомогательной оси Y, необходимо:

- правой кнопкой мыши щелкнуть по графику изменения численности населения;

- в появившемся контекстном меню выбрать пункт **Формат рядов данных...**;

- в открывшемся окне выбрать раздел **Параметры ряда**, в котором отметить пункт **по вспомогательной оси**;

- нажать **ОК**.

Полученная диаграмма должна иметь вид, как на рисунке 9.

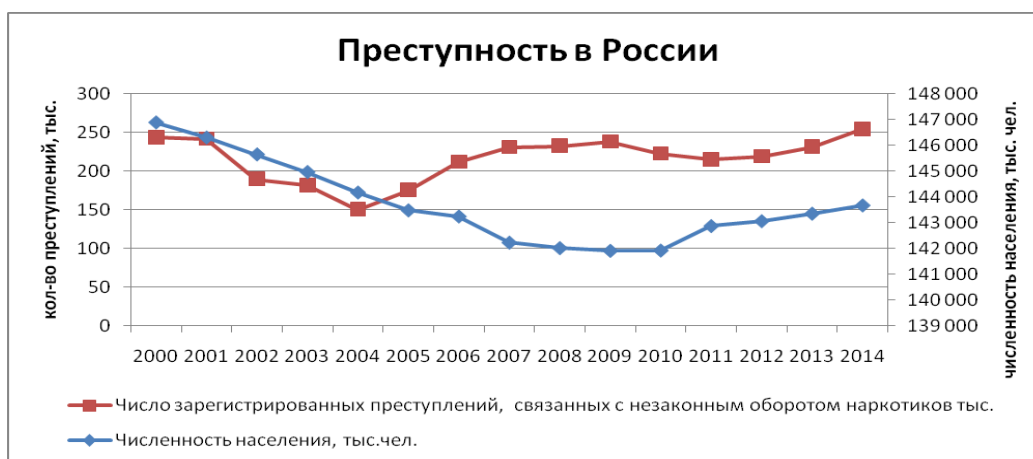


Рисунок 9. Диаграмма изменения численности населения и числа зарегистрированных преступлений в России

**Упражнение 2.** По данным таблицы оформите временной ряд изменения количество преступлений, связанных с незаконным оборотом наркотических средств, в Российской Федерации и одного из государств – членов ОДКБ (на свой выбор), и дайте их графическое представление (расположите на одном графике):

Государства-члены ОДКБ	Год					
	2009	2010	2011	2012	2013	2014
Республика Армения	1268	1524	1525	1139	1222	1046
Республика Беларусь	4522	4655	4532	4207	5008	7356
Республика Казахстан	9705	8795	4360	3886	3639	3548
Кыргызская Республика	1887	1543	1924	1933	1913	1955
Российская Федерация	238527	222564	215214	218974	231462	254730
Республика Таджикистан	796	816	927	895	936	993

## 5.2. ВЫЧИСЛЕНИЕ ХАРАКТЕРИСТИК ДИНАМИЧЕСКОГО РЯДА

*Относительные величины динамики (ОВД)* процесса, изменения явления во времени показывают, во сколько раз увеличился (или уменьшился) уровень показателя текущего периода по сравнению с каким-либо предшествующим (базисным) периодом:

$$\text{ОВД} = \frac{\text{Текущий показатель}}{\text{Предшествующий или базисный показатель}}$$

Этот вид относительных величин имеет самое широкое распространение в различных отраслях статистики. Например, в демографической статистике они применяются при рассмотрении сдвигов, происшедших в составе населения, – движении всего населения, в том числе по возрастному, половому и другим признакам и т.п. В правовой – при анализе изменений преступности – движении общего числа преступлений и их отдельных категорий, динамики нагрузки следователей, судей и т.д.

При анализе изменения интересующей нас совокупности исчисляются: абсолютный прирост (снижение), коэффициент роста (снижения), темп роста (снижения), темп прироста и значение абсолютной величины одного процента прироста. *Абсолютный прирост* – разность между данным уровнем ряда и уровнем ряда, принятым за базу сравнения; *коэффициент роста* – отношение одного уровня ряда к другому его уровню, принятому за базу; *темп роста* – *коэффициент роста*, выраженный в процентах; *темп прироста* – отношение абсолютного прироста к уровню, принятому за базу, выраженное в процентах; *абсолютная величина одного процента прироста* – отношение абсолютного прироста к темпу прироста.

При вычислении ОВД за базу сравнения может приниматься все время один уровень ряда динамики. Такие величины называются относительными величинами с *постоянной базой*, а способ вычисления – базисным. Они применяются в тех случаях, когда есть необходимость постоянного наблюдения за развитием изучаемых явлений по сравнению с периодом, имеющим почему-либо особо важное для подобного сравнения значение.

В случае необходимости выяснения темпов развития интересующего нас явления каждый показатель сравнивается

уже не с одной постоянной величиной, а со своей предыдущей, т.е. за базу принимается показатель предыдущего периода. Такие ОВД называются величинами, вычисленными *цепным способом*.

Для установления **средних показателей темпов роста** рядов динамики вычисляется средняя геометрическая, которая исчисляется путем извлечения корня степени  $n$  из произведений значений темпов роста  $x$ :

$$\bar{x}_g = \sqrt[n]{\prod x},$$

где  $\bar{x}_g$  – средняя геометрическая;  $\prod$  – знак перемножения.

### ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

**Задание 1.** На основе динамического ряда о количестве зарегистрированных преступлений в городе N заполните следующую статистическую таблицу:

Год	2010	2011	2012	2013	2014	2015
Кол-во зарегистрированных преступлений	2430	2673	2802	2851	2920	3231
Абсолютный прирост (цепной)						
Абсолютный прирост (базисный)						
Темп роста (цепной)						
Темп роста (базисный)						
Темп прироста (цепной)						
Темп прироста (базисный)						
Абсолютная величина одного процента прироста (цепной)						
Абсолютная величина одного процента прироста (базисный)						

Вычислите средний темп роста цепным и базисным способом.

**Задание 2.** На основе динамических рядов из упражнения 1 заполните следующую статистическую таблицу:

**Показатели наркопреступности в России за 2000–2014 гг.**

Год	Численность населения, тыс. чел.	Число зарегистрированных преступлений, тыс.	Абсолютный прирост (цепной)	Темп роста (цепной), %	Темп прироста, (цепной)%	Темп роста (базисный), %	
						к 2000 г.	к 2010 г.

Вычислите абсолютную величину одного процента прироста числа зарегистрированных преступлений в 2005 и 2014 годах.

**Задание 3.** На основе представленных данных рассчитайте цепным способом темпы роста количества преступлений, связанных с незаконным оборотом наркотических средств, по федеральным округам и Российской Федерации в целом, а также определите средний темп роста, сделайте вывод.

Федеральный округ	2010	2011	2012	2013	2014
Северо-Западный	22 389	24 318	24 620	26 066	26 557
Северо-Кавказский	10 192	10 134	10 693	10 396	10 836
Южный	21 642	20 800	20 931	19 509	19 237
Приволжский	40 414	37 961	37 573	39 607	45 627
Уральский	19 689	19 608	19 669	23 284	25 810
Сибирский	39 313	37 666	38 179	38 644	40 365
Центральный	53 505	51 230	53 648	58 779	67 524
Дальневосточный	13 410	11 486	11 649	13 164	16 760
Российская Федерация	220 554	213 203	216 962	229 449	252 716

**Задание 4.** Известны следующие данные о заболеваемости наркоманиями в РФ (данные условные). Рассчитайте отсутствующие значения, а также определите средний абсолютный прирост и средний темп роста за этот период.

Год	2007	2008	2009	2010	2011	2012	2013	2014
Число больных с диагнозом, установленным впервые в жизни тыс. чел.	2 483				3 146		3 787	3 719
Базисные показатели динамики								
Абсолютный прирост	–	122						
Коэффициент роста	–			1,19				
Темп прироста	–		13			32		
Абсолютное значение 1% прироста	–							

**Задание 5.** Известны следующие данные о количестве больных наркоманиями, прошедших реабилитацию в реабилитационных центрах субъектов Российской Федерации (данные условные). Рассчитайте отсутствующие значения, а также определите средний абсолютный прирост и средний темп роста за этот период.

Год	2007	2008	2009	2010	2011	2012	2013	2014
Число больных с диагнозом, установленным впервые в жизни тыс. чел.	2 227		2 246				2318	
Цепные показатели динамики								
Абсолютный прирост	–			– 72		– 50		
Коэффициент роста	–	0,9785						1,019
Темп прироста	–							
Абсолютное значение 1% прироста	–					22,28		

### 5.3 МЕТОДЫ ВЫЯВЛЕНИЯ ТЕНДЕНЦИИ ИЗМЕНЕНИЯ ЯВЛЕНИЯ И ПРОГНОЗИРОВАНИЯ

Самый простой способ выявления тенденции динамического ряда – *визуальное изучение его графического изображения*. Однако не всегда этот способ дает надежные результаты. Для выявления скрытых закономерностей, обнаружения за колебаниями определенных тенденций, применяются более тонкие *приемы обработки (преобразования) динамических рядов*. Выделяются эмпирические и аналитические способы преобразования рядов динамики. Из эмпирических приемов обработки рядов динамики наиболее распространены способ укрупнения интервалов и способ сглаживания путем скользящей средней, а из аналитических – способ средней геометрической.

*Способ укрупнения интервалов* – один из самых простых способов выявления тенденции развития изучаемого явления.

Укрупнение интервалов начинают с наименьшего возможного для укрупнения интервала (при интервале один год, например, берется двухгодичный интервал). Если такое укрупнение не проясняет картины, интервал укрупняется еще больше (берется, например, трехлетний интервал). Этот

процесс продолжают до тех пор, пока окончательно не выяснится тенденция изменения ряда динамики. Недостаток названного способа в том, что утрачивается картина изменения явления внутри укрупненного интервала.

*Сглаживание рядов динамики* – прием устранения колебаний значений показателей, образующих ряд динамики, в целях выявления основной тенденции развития исследуемого явления. Этот прием сводится к тому, что из показателей фактического ряда вычисляются средние, и фактический, колеблющийся ряд заменяется плавным, сглаженным рядом, характер и особенности которого будут четко выявлены.

**Пример.** Имеются следующие данные, характеризующие количество наркопреступлений в городе N за 8 месяцев:

Месяц	1-й	2-й	3-й	4-й	5-й	6-й	7-й	8-й
Количество наркопреступлений	470	443	718	644	574	907	775	894

Четкой тенденции в представленном ряду не наблюдается. Применим сглаживание, т.е. заменим фактические показатели так называемыми скользящими средними, вычисленными из трех рядом стоящих уровней с уступом каждый раз на один интервал (средняя из трех). Следовательно, первый показатель будет равен:

$$\frac{470+443+718}{3} = 543;$$

второй показатель

$$\frac{443+718+644}{3} = 601 \text{ и т.д.}$$

**Упражнение 3.** Рассчитайте сглаженный ряд динамики количества наркопреступлений в городе N. Постройте графики исходного ряда и сглаженного, сделайте вывод.

Месяц	1-й	2-й	3-й	4-й	5-й	6-й	7-й	8-й
Количество наркопреступлений	470	443	718	644	574	907	775	894
Сглаженный ряд динамики	–	543	601	...	...	...	...	–

**Упражнение 4.** На территории отдела полиции некоторого района г. N было совершено: в январе 35 преступлений; в феврале – 15; марте – 20; апреле – 25; мае – 30; июне – 20; июле – 35; августе – 55; сентябре – 45; октябре – 30; ноябре

– 20; декабре – 15. Преобразуйте динамический ряд методом укрупнения интервалов (ступенчатой средней). Охарактеризуйте проявившиеся тенденции.

Для расчета скользящей средней можно брать и сумму из пяти рядом стоящих уровней со смещением на один интервал. Однако чем больше интервал, за который исчисляется средняя (т.е. укрупненный интервал), тем усреднённый окажется полученный результат в сравнении с фактическим.

**Упражнение 5.** На основе приведенного динамического ряда постройте сглаженные ряды динамики (при вычислении среднего из трех и пяти уровней). Постройте графики исходного и сглаженных рядов динамики, сделайте вывод.

Год	Количество иногородних лиц, ежегодно прибывающих в город N на ПМЖ, тыс.чел.	Сглаженный ряд динамики 1	Сглаженный ряд динамики 2
2005	12,6	–	–
2006	10,8		–
2007	17,6		
2008	17,4		
2009	21,3		
2010	18,9		
2011	28,7		
2012	32,4		
2013	28,8		–
2014	36,7	–	–

При исследовании динамического ряда может сложиться ситуация, когда какие-то его показатели отсутствуют и их, основываясь на известных значениях уровней ряда, необходимо восстановить. Например, если в динамическом ряду показателей численности населения административно-территориальной единицы отсутствуют данные за некоторые годы, значит, нет возможности вычислить и коэффициент преступности за указанные периоды. Для того чтобы восстановить отсутствующие данные, применяется **интерполяция ряда динамики**, т.е. оценивание недостающих уровней ряда. Но может возникнуть необходимость «заглянуть» за пределы динамического ряда и отыскать, также на основе изучения известных уровней ряда, предполагаемые его уровни в будущем, т.е. осуществить прогноз. В этом случае применя-

ется **экстраполяция ряда динамики**, т.е. нахождение по динамическому ряду известных значений последующих значений, находящихся за пределами динамического ряда.

Благодаря относительной устойчивости закономерностей развития социальных явлений, экстраполяция вполне применима для их прогнозирования на сравнительно короткие сроки.

В общем виде при исследовании временных рядов чаще всего используют линию тренда. **Тренд** – плавно меняющаяся компонента, описывающая влияние долговременных факторов, т.е. длительную тенденцию изменения признака.

Линия тренда позволяет оценить тенденцию изменения исследуемого признака и развитие изучаемого процесса (например, уровень растет или убывает).

Продолжая линию тренда на несколько периодов вперед можно формировать прогноз исследуемого признака.

**Упражнение 6.** По таблице **Показатели преступности в России за 2000-2014 гг.** постройте диаграмму изменения зарегистрированной преступности в России за 2000-2014 гг. Добавьте линию тренда и дайте прогноз на год вперед.

**1.** Для построения линии тренда необходимо:

– правой кнопкой мыши щелкнуть по графику;

– в появившемся контекстном меню выбрать пункт **Добавить линию тренда...**;

– в открывшемся окне выбрать раздел **Тип**, в котором можно выбрать различные модели для аппроксимации графика (выберите в соответствии с рисунком, приведенным ниже);

– нажать **ОК**.

**2.** Изменяя тип линии тренда, подберите такую кривую, которая максимально близко подходит к исходным данным, для этого:

– правой кнопкой мыши щелкните по линии тренда;

– в появившемся контекстном меню выберите пункт

**Формат линии тренда...**;

– в разделе **Параметры** включите опции **показывать уравнение на диаграмме** и **поместить на диаграмму величину достоверности аппроксимации ( $R^2$ )**. Чем ближе к единице параметр достоверности аппроксимации  $R^2$ , тем более точно описывает данная модель наш график.

**3.** Постройте прогноз на один год вперед. Для этого:

– в разделе **Параметры** контекстного меню линии тренда задайте прогноз вперед на один период.

4. По линии тренда определите численные значения прогнозных показателей преступности. Полученная диаграмма должна иметь вид, как на рисунке 10.

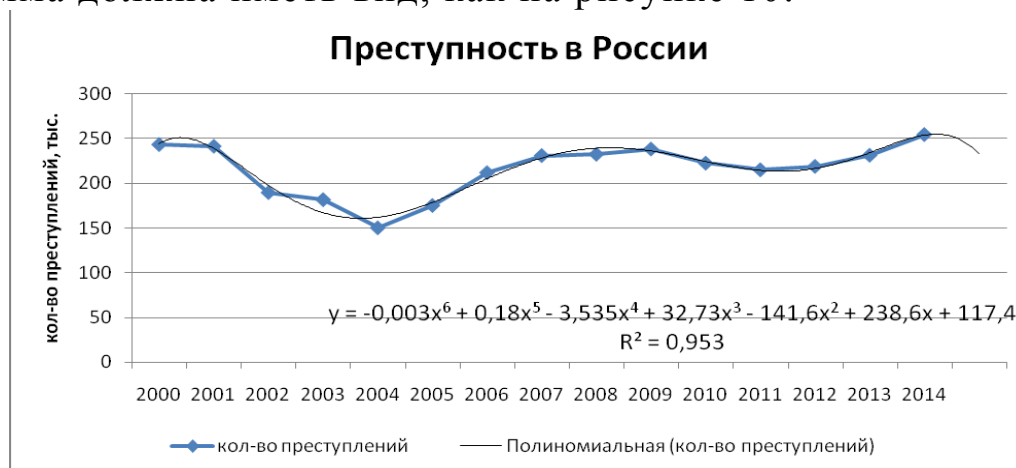


Рисунок 10. Диаграмма изменения зарегистрированной преступности в России за 2000-2014 гг. и линия тренда.

#### 5.4. ИНДЕКС СЕЗОННОЙ ПРЕСТУПНОСТИ

Абсолютному большинству юридически значимых явлений свойственны сезонные колебания. Изучать сезонность юридически значимых явлений необходимо в интересах организации адекватной организации управленческой деятельности. Наиболее простой метод выявления и измерения сезонных колебаний – это расчет среднего уровня (среднего арифметического) изучаемых явлений по месяцам за год и сопоставление месячных данных со средним уровнем. Это отношение уровней, выраженное в процентах, именуется *индексом сезонности*. Он рассчитывается по формуле:

$$ИС = \frac{Y_M}{Y_{CP}} \times 100\%$$

где ИС – индекс сезонности,  $Y_M$  – уровень (данные за месяц),  $Y_{CP}$  – средний уровень за год.

В случае если присутствуют данные за несколько лет, то индекс сезонности рассчитывается по формуле:

$$ИС = \frac{Y_{Mcp}}{Y_{CP}} \times 100\%$$

где ИС – индекс сезонности,  $Y_{Mcp}$  – среднее значение за соответствующий месяц, вычисленное по имеющимся данным

(например, среднее за январь, среднее за февраль и т.п.),  $U_{CP}$  – средний уровень, вычисленный как общее среднее по всем данным.

## ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

**Задание 1.** По данным о динамике преступности в России вычислите индексы сезонности. Изобразите на графике сезонную волну.

Месяц	Уровень преступности	Индекс сезонности
январь	202 013	
февраль	231 917	
март	236 772	
апрель	217 325	
май	216 308	
июнь	224 342	
июль	210 395	
август	214 087	
сентябрь	227 999	
октябрь	232 266	
ноябрь	206 925	
декабрь	204 732	
$U_{CP}$		

Выделите цветом только те строки, уровень преступности в которых превышает средний уровень преступности за год (воспользуйтесь **кнопкой Условное форматирование**, расположенной на вкладке меню **Главная** в группе **Стили**).

**Задание 2.** В таблице представлены данные о численности граждан, имеющих статус безработных, в одном из субъектов РФ (тыс. чел., данные условные). Выявите наличие сезонной волны в изменении численности безработных.

Месяц	2014	2015	2016
январь	10,9	8,9	8,4
февраль	10,8	9,6	8,8
март	10,4	9,4	8,5
апрель	9,6	9,0	8,0
май	8,5	8,3	7,0
июнь	8,1	7,9	6,6
июль	7,8	7,7	6,5

Месяц	2014	2015	2016
август	7,8	7,5	6,5
сентябрь	7,6	7,3	6,4
октябрь	7,7	7,2	6,6
ноябрь	8,2	7,7	7,5
декабрь	8,8	8,3	8,0

Изобразите на графике сезонную волну.

Выделите желтым цветом только те ячейки, уровень преступности в которых превышает средний уровень преступности (воспользуйтесь **кнопкой Условное форматирование**, расположенной на вкладке меню **Главная** в группе **Стили**).

**Задание 3.** Используя указанные данные, по следующей формуле рассчитайте *индекс тяжести преступлений (ИТП)*:

$$ИТП = \frac{\sum P_m \times B_m}{\sum P_{\text{б}} \times B_m}$$



где  $\sum P_m$  – сумма преступлений текущего периода,

$\sum P_{\text{б}}$  – сумма преступлений базового периода,

$B_m$  – баллы тяжести преступлений (рассчитываются как максимальный срок лишения свободы за указанное преступление), их значение возьмите из СПС КонсультантПлюс. Данные по количеству преступлений в Российской Федерации скачайте из ЕМИСС ([www.fedstat.ru](http://www.fedstat.ru) или [www.gks.ru](http://www.gks.ru)).

Вид преступления	Количество зарегистрированных преступлений в РФ		Максимальный срок лишения свободы (баллы)	Статья УК РФ
	2010	Данные за прошлый год		
Убийство				
Умышленное причинение вреда здоровью				
Изнасилование				
Грабёж				
Разбой				
Кража				
Вымогательство				
Хулиганство				
Преступления, связанные с незаконным оборотом наркотиков				228

*Указания.* Для скачивания информации зайдите на сайт ЕМИСС (Единой межведомственной информационной статистической системы) по адресу [www.fedstat.ru](http://www.fedstat.ru) или через сайт Федеральной службы государственной статистики [www.gks.ru](http://www.gks.ru) (раздел **Официальная статистика** ⇒ **Базы данных** ⇒ **ЕМИСС**). Перейдите в раздел **Ведомства**. Раскройте список показателей Министерства внутренних дел. Раскройте список «3.03. Информация о преступности». Щелкните по ссылке «Количество преступлений, зарегистрированных в отчетном периоде».

Щелкните по кнопке , расположенной в столбце **Справочник территорий**, очистите открывшийся список и выберите **Российская Федерация**, нажмите кнопку **Фильтровать**. Чтобы выбрать интересующий вас год нажмите на кнопку , расположенную справа от года, очистите фильтр, и в раскрывшемся списке выберите нужный год. Чтобы сохранить полученные данные щелкните по кнопке



В открывшемся списке выберите что позволяет сохранить данные в формате Excel.

Количество преступлений, связанное с незаконным оборотом наркотиков можно найти в разделе «3.05. Незаконный оборот, производство и использование наркотиков», списка показателей МВД России.

## 5. СТАТИСТИЧЕСКИЙ АНАЛИЗ ВЗАИМОСВЯЗИ СОЦИАЛЬНО-ПРАВОВЫХ ЯВЛЕНИЙ

### 6.1. ОЦЕНКА СВЯЗИ МЕЖДУ ПРИЗНАКАМИ, КОРРЕЛЯЦИОННАЯ ЗАВИСИМОСТЬ

Основная задача науки – вскрыть наиболее важные, наиболее существенные связи и взаимозависимости между признаками явлений и предметов, изучаемых данной наукой, чтобы установить управляющие ими законы. Поэтому *методы исследования, измерения связей составляют чрезвычайно важную часть методологии научного исследования, в том числе и статистического.*

В процессе статистического исследования объективно существующих связей вскрываются *причинно-следственные* отношения между явлениями, что позволяет выявлять факторы (признаки), оказывающие основное влияние на вариацию изучаемых явлений и процессов. Напомним, что признаки, обуславливающие изменение других, связанных с ними признаков, называются *факторными*, или просто *факторами*, признаки, изменяющиеся под действием факторных признаков, – *результативными*.

Различают два вида (типа) связей между различными явлениями и их признаками: *функциональную, или жестко детерминированную*, с одной стороны, и *статистическую, или стохастически детерминированную* – с другой.

**Функциональные связи** характеризуются полным соответствием между изменением факторного признака и изменением результативной величины, и каждому значению признака фактора соответствуют вполне определенные значения результативного признака. Так, изменение температуры внешней среды (фактор) адекватно отражается на шкале ртутного столбика (результативный признак) градусника. Функциональная связь двух величин возможна лишь при условии, что вторая из них зависит *только* от первой и ни от чего более. Понятно, что в общественных явлениях мы почти не встречаемся с такими связями.

Напротив, **стохастически детерминированные связи** не имеют ограничений и условий. При статистической связи

разным значениям одной переменной соответствуют разные распределения значений другой переменной.

Важнейший частный случай статистической связи – корреляционная связь. **Корреляционная зависимость** – *взаимосвязь между признаками, состоящая в том, что средняя величина значений одного признака меняется в зависимости от изменения другого признака* (например, зависимость между выработкой рабочих и стажем работы, между аварийностью и профессионализмом водителей автотранспорта, между числом судимостей преступника и временем его нахождения на свободе между ними и др.). С увеличением стажа средняя выработка изменяется (в общем, тоже возрастая), с повышением профессионализма, с увеличением срока вожения автомобиля изменяется средняя аварийность (в общем, снижается), по мере увеличения числа судимостей у преступника сокращается срок его пребывания на свободе и т.д. Но в отличие от функциональной зависимости в индивидуальных случаях при определенном значении одного признака могут быть разные значения другого. В последнем примере не исключаются случаи и при большом «наборе» судимостей становления преступника на праведный путь.

Основная **задача корреляции** заключается в том, чтобы на основе строго математических приемов установить количественное выражение этой зависимости, которая существует между исследуемыми признаками, абстрагируясь при этом от влияния всех других признаков, искажающих истинную связь. Такая связь определяется в корреляции на основе специальных уравнений, выражающих типичное соотношение между изучаемыми признаками.

Вместе с тем задача корреляционного анализа – не только подтверждение или обнаружение наличия связей между различными явлениями (например, зависимость между числом преступлений на одного работника полиции и раскрываемостью преступлений), но и установление формы и тесноты (силы) связей, количественная оценка влияния факторного признака на результат.

По форме корреляционные связи могут быть линейными и нелинейными, а по направлению – прямыми (положительными) и обратными (отрицательными). Прямая связь свидетельствует о том, что с увеличением значения одного при-

знака увеличивается и значение другого признака. Так, с увеличением числа пожаров увеличивается и число людей, пострадавших от них. При обратной связи увеличение значения одного признака ведет к уменьшению значения другого признака. Так, понижение уровня жизни ведет к росту преступности.

Главная задача корреляционного анализа решается путем вычисления коэффициентов корреляции:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

**Коэффициент корреляции**, характеризующий меру тесноты (силы) связи, тем больше, чем более жесткая зависимость. При коэффициенте корреляции, равном нулю, связь отсутствует, если он равен единице – связь не корреляционная, а функциональная. Коэффициент корреляции, равный +1 или –1, характеризует полную прямую (+) или обратную (–) связь между сопоставляемыми признаками. Следовательно, дроби, в пределах от 0 к 1, будут определять степень корреляции – большая дробь (например, 0,9) свидетельствует о большей корреляции, а меньшая (0,3) – о меньшей.

По степени тесноты связи различают следующие количественные критерии оценки тесноты связи.

**Количественные критерии оценки тесноты связи**

Коэффициент корреляции	Характер связи
До {0,3}	Практически отсутствует Слабая Умеренная Сильная
{±0,3} – {±0,5}	
{±0,5} – {±0,7}	
{±0,7} – {±1,0}	

**Упражнение 1.** Найдите в сетевой папке по адресу «\\infserv\install\disk\_D\ПС» файл *Корреляция.xlsx*. Скопируйте его в свой каталог и откройте. В файле представлены значения по ряду признаков по регионам Сибирского федерального округа за 2013 год. Затем на официальном сайте Федеральной службы государственной статистики ([www.gks.ru](http://www.gks.ru)) в разделе «Официальная статистика» найдите и скачайте сборник «Регионы России. Социально-экономи-

ческие показатели. 2014» (в сборнике за 2014 год представлены статистические показатели за 2013 г.) Из указанного сборника к имеющимся данным добавьте следующие:

- число зарегистрированных преступлений, связанных с незаконным оборотом наркотиков;
- среднедушевые денежные доходы населения (в месяц; рублей);
- удельный вес городского населения в общей численности населения (оценка на конец года; в процентах);
- численность безработных (тысяч человек);
- численность зрителей театров и число посещений музеев на 1000 человек населения;
- количество спортивных залов и плавательных бассейнов (на конец года).

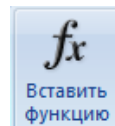
Определите, есть ли связь между каждым из признаков из получившегося набора данных и:

- количеством преступлений, связанных с НОН;
- количеством больных наркоманиями;
- сделайте выводы.

Для оценки степени связи между признаками вычислите коэффициент корреляции. Для этого следует:

- выделить ячейку, в которой будет вычисляться коэффициент корреляции;

– для запуска мастера функций воспользоваться кнопкой **Вставить функцию** на вкладке **Формулы**;



– на первом шаге в окне мастера функций в разделе **Категория** выбрать **Статистические**. Из этой категории выбрать функцию **КОРРЕЛ**. Нажать **ОК**;

– на втором шаге в окнах **Массив1** и **Массив2** задать диапазоны ячеек, которые содержат статистические данные. При задании массивов необходимо следить, чтобы их размерности совпадали. Нажать **ОК**.

**Упражнение 2.** Вычислите коэффициенты корреляции между количеством преступлений, связанных с НОН, и остальными признаками (как в упражнении 1) в вашем федеральном округе (данные возьмите из аналогичного сборника за прошлый год). Если вы из Сибирского федерального округа, то уточните у преподавателя, данные какого федерального округа вам следует выбрать

Еще одной компонентой, которую учитывают при анализе рядов динамики, является лаг или отставание одного явления от другого, связанного с ним.

**Лаг** – это смещение во времени изменений одних явлений по сравнению с другими (например, между регистрацией преступлений и осуждением виновных имеются заметные временные сдвиги, значительное число уголовных дел, завершенных в текущем году, рассматривается судами в следующем и т.д.). Учитывая возможность такого отставания, полезно, например, перед анализом связи двух динамических рядов сдвинуть один ряд относительно другого. Лаг и будет интервалом такого сдвига.

Например, можно предположить, что на признак *количество преступлений, связанных с НОН*, признак *численность лиц, отбывавших наказание за подобные преступления и освобожденных из мест лишения свободы*, окажет влияние только год спустя.

**Упражнение 3.** По имеющимся данным в городе N, проверьте гипотезу инерционности процесса влияния признака *численность лиц, отбывавших наказание за подобные преступления и освобожденных из мест лишения свободы* на *количество преступлений, связанных с НОН* с задержкой в 1 и 2 года.

Год	Количество преступлений, связанных с НОН	Численность освобожденных
2002	634	127
2003	536	96
2004	542	184
2005	689	102
2006	604	137
2007	678	110
2008	632	193
2009	701	190
2010	765	184
2011	789	189
2012	789	135
2013	865	175
2014	564	115

Для измерения связи альтернативных признаков применяются **коэффициент ассоциации** Дэвида Юла и **коэффициент контингенции** Карла Пирсона. Для расчета этих показателей применяется следующая матрица взаимного распределения частот:

	1 признак	ДА	НЕТ
2 признак			
ДА		a	b
НЕТ		c	d

где a, b, c, d – частоты взаимного распределения признаков.

При прямой связи частоты сконцентрированы по диагонали a-d, при обратной связи по диагонали b-c, при отсутствии связи частоты практически равномерно распределены по всему полю таблицы. Коэффициент ассоциации вычисляется по формуле:  $K_{ac} = \frac{ad - bc}{ad + bc}$

Значение коэффициента ассоциации, как и коэффициента корреляции, изменяется от -1 до 1.

**Упражнение 4.** Проанализируйте зависимость между полом и фактом совершения преступления, связанного с НОН.

	1 признак	М	Ж
2 признак			
Совершил		24	32
Не совершил		16	28

В случае если по каким-то причинам хотя бы один из показателей таблицы «четырех полей» отсутствует (равен нулю), величина коэффициента ассоциации будет равна  $\pm 1$ , что будет означать завышенную оценку степени тесноты связи между признаками. В этом случае применяется **коэффициент контингенции**, который рассчитывается по формуле:

$$K_{\text{конт}} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Коэффициент контингенции также изменяется в пределах от -1 до 1, но всегда его величина для тех же данных меньше коэффициента ассоциации.

**Упражнение 5.** Рассчитайте для предыдущего примера коэффициент контингенции. Сделайте вывод.

**Упражнение 6.** Проанализируйте зависимость между полом и фактом пробы наркотического вещества:

	1 признак	М	Ж
2 признак			
Пробовал		74	51
Не пробовал		1093	1619

Рассчитайте коэффициенты ассоциации и контингенции.

Если значения признака распределены более чем по 2 группам, то для определения тесноты связи применяют **коэффициенты взаимной сопряженности признаков Пирсона, Чупрова** и др.

**Коэффициент сопряженности признаков Пирсона** определяется по формуле:

$$C = \sqrt{\frac{J^2}{1+J^2}};$$

$$J^2 = L_1 + L_2 + L_3 + \dots - 1 = \sum L_i - 1;$$

$$\text{где } L_i = \frac{\left( \sum_j \frac{s_{ij}^2}{m_j} \right)}{n_i};$$

где  $J^2$  – показатель взаимной сопряженности признаков, который рассчитывается на основе матрицы взаимного распределения частот:

	1 гр.	2 гр.	3 гр.	...	j гр.	...	Итого
1 гр.	$s_{11}$	$s_{12}$	$s_{13}$	...	$s_{1j}$	...	$n_1$
2 гр.	$s_{21}$	$s_{22}$	$s_{23}$	...	$s_{2j}$	...	$n_2$
3 гр.	$s_{31}$	$s_{32}$	$s_{33}$	...	$s_{3j}$	...	$n_3$
...	...	...	...				
i гр.					$s_{ij}$		$n_i$
...	...	...	...				
Итого	$m_1$	$m_2$	$m_3$	...	$m_j$	...	

$s_{ij}$  – значения в соответствующих ячейках таблицы,

$m_j$  – сумма значений по столбцу;

$n_i$  – сумма значений по строке.

Более точным показателем тесноты связи является **коэффициент Чупрова**, который определяется по формуле:

$$K_{\text{чупр}} = \sqrt{\frac{J^2}{k_1 - 1 \cdot k_2 - 1}},$$

где  $k_1$  и  $k_2$  – соответственно число групп, выделенных по каждому признаку.

Коэффициент А.А. Чупрова в отличие от коэффициента ассоциации варьирует от 0 до 1. Если исходить из формулы, то его значение не может быть отрицательным. Но суть интерпретации та же. Связь считается существенной при величине  $K_{\text{чупр}} = 0,3$ . Чем ближе его значение к единице, тем сильнее связь.

**Упражнение 7.** Имеется информация о распределении сотрудников правоохранительных органов, совершивших коррупционные преступления в некотором федеральном округе, по полу и возрасту.

Для оценки тесноты связи между полом и возрастом лиц, совершивших данный вид преступления, рассчитайте коэффициенты взаимной сопряженности К. Пирсона и А. Чупрова. Указанные вычисления сделайте в MS Excel. Сделайте выводы.

Возраст	Удельный вес, %		Итого
	мужчины	женщины	
20-30	57	44	101
30-40	62	40	102
40 и старше	66	30	96
Итого	185	114	299

*Решение*

Вычислим показатель взаимной сопряженности признаков  $J^2$ :

$$J^2 = \frac{57^2}{185} + \frac{44^2}{114} + \frac{62^2}{185} + \frac{40^2}{114} + \frac{66^2}{185} + \frac{30^2}{114} - 1 = 0,342 + 0,341 + 0,328 - 1 = 1,011 - 1 = 0,011.$$

Коэффициент взаимной сопряженности Пирсона будет равен:

$$C = \sqrt{\frac{J^2}{1 + J^2}} = \sqrt{\frac{0,011}{1,011}} = 0,104.$$

Коэффициент взаимной сопряженности Чупрова:

$$K_{\text{чупр}} = \sqrt{\frac{J^2}{k_1 - 1} \cdot \frac{1}{k_2 - 1}} = \sqrt{\frac{0,011}{(2-1) \cdot (3-1)}} = \sqrt{\frac{0,011}{1 \cdot 2}} = 0,074.$$

**Упражнение 8.** Определите коэффициенты взаимной сопряженности К. Пирсона и А. Чупрова между двумя качественными признаками: число погибших и раненных в автомобильных авариях в зависимости от причины аварий: по вине пьяных водителей, по вине неисправности автомобиля и по вине пешеходов. Исходные данные представлены в таблице:

Причины аварий	Число погибших	Число раненных	Итого
Пьяные водители	100	300	
Неисправности автомобилей	200	800	
Вина пешеходов	300	1200	
Итого			

## 6.2. ПРОГНОЗИРОВАНИЕ

Задача прогнозирования критериальных показателей от будущего значения факторных дает возможность оценивать тенденции не только результативного признака, но и учитывать степень влияния независимых признаков (факторных). Как правило, в качестве факторных показателей выбирают только те, которые оказывают наибольший вклад (имеют высокий уровень корреляции с результативным признаком).

**Упражнение 9.** Откройте файл **Регрессия.xlsx**, в котором дана таблица статистических наблюдений, приведенная на рисунке 11. Для корректного использования функций критериальные и факторные показатели расположены в смежных ячейках. На основании оценки взаимосвязи признаков выдвинута гипотеза, что на значение  $Y_1$  наибольшее влияние оказывает факторный признак  $X_4$ , на  $Y_2$  признак  $X_3$ , на  $Y_3$  признаки  $X_1$  и  $X_4$ .

Следует осуществить прогноз критериальных признаков  $Y_1, Y_2, Y_3$  на следующий временной период – 2016 год.

	A	B	C	D	E	F	G	H	I
1		Количество преступлений, связанных с незаконным оборотом наркотиков (на 100 тыс. населения)	Количество преступлений ст.232 УК России (организация либо содержание притонов)	Количество фактов оптовых поставок НС и ПВ	Количество прибывшего населения	Учетная распространенность каннабиодной наркомании (на 100 тыс. населения)	Количество взятых под наблюдение лиц с синдромом зависимости от наркотических средств (наркомания) и потребителей наркотических средств (на 100 тыс. населения)	Учетная распространенность опийной наркомании (на 100 тыс. населения)	Употребление наркотических средств с вредными последствиями (на 100 тыс. населения)
2		Y1	Y2	Y3	X1	X4	X2	X3	X5
3	2007	136,0	3143	3178	8253	2,9	45,3	350,2	26,4
4	2008	161,1	4509	5590	7838	3,8	49,2	345,5	25,3
5	2009	193,0	5098	6175	7848	4,1	50,1	330,7	24,2
6	2010	206,1	5466	7099	7710	5,8	54,6	333,1	25,7
7	2011	216,8	6414	8193	7870	5,9	39,5	322,6	20,9
8	2012	216,0	6554	9287	8130	6,4	58,3	308,9	28,8
9	2013	204,0	7077	9830	10190	8,3	50,6	264,4	28,4
10	2014	195,6	6104	11990	19667	8,5	47,9	240,2	26,9
11	2015	198,1	6703	12139	24375	8,8	48,6	243,1	27,6
12	2016								

Рисунок 11. Статистические показатели региона N

Для начала работы создайте в папке своей группы файл **Регрессионный анализ.xlsx** и скопируйте на пустой лист рабочей книги статистическую таблицу (рисунок 11) из файла **регрессия.xlsx**. Лист 1 переименуйте в **Прогноз**.

Порядок вычисления прогнозных значений:

1. Введите в соответствующие ячейки на 2016 год предполагаемые в будущем значения факторных показателей X1, X3, X4 (их в дальнейшем можно менять, «играя» с моделью).

2. Сделайте активной ячейку **B12** (результат прогноза).

3. Перейдите на вкладку **Формулы** и в группе **Библиотека функций** выберите **Вставить функцию**. Затем в категории **Статистические** укажите **ТЕНДЕНЦИЯ**, как показано на рисунке 12.

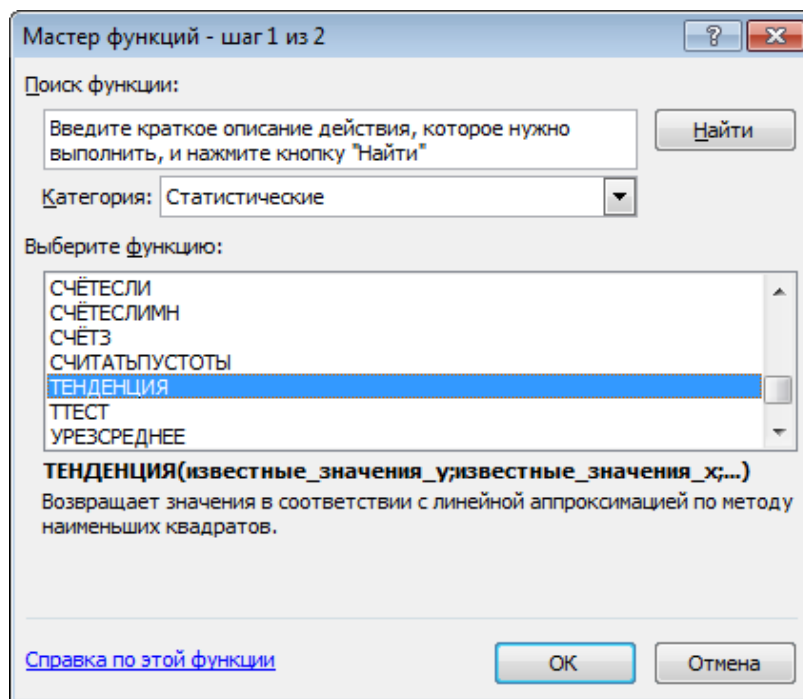


Рисунок 12. Мастер функций

4. Далее в открывшемся меню установите мышью курсор в позицию **Известные\_значения\_y** и укажите диапазон данных, соответствующий столбцу Y1 (**B3:B11**), как на рисунке 13.

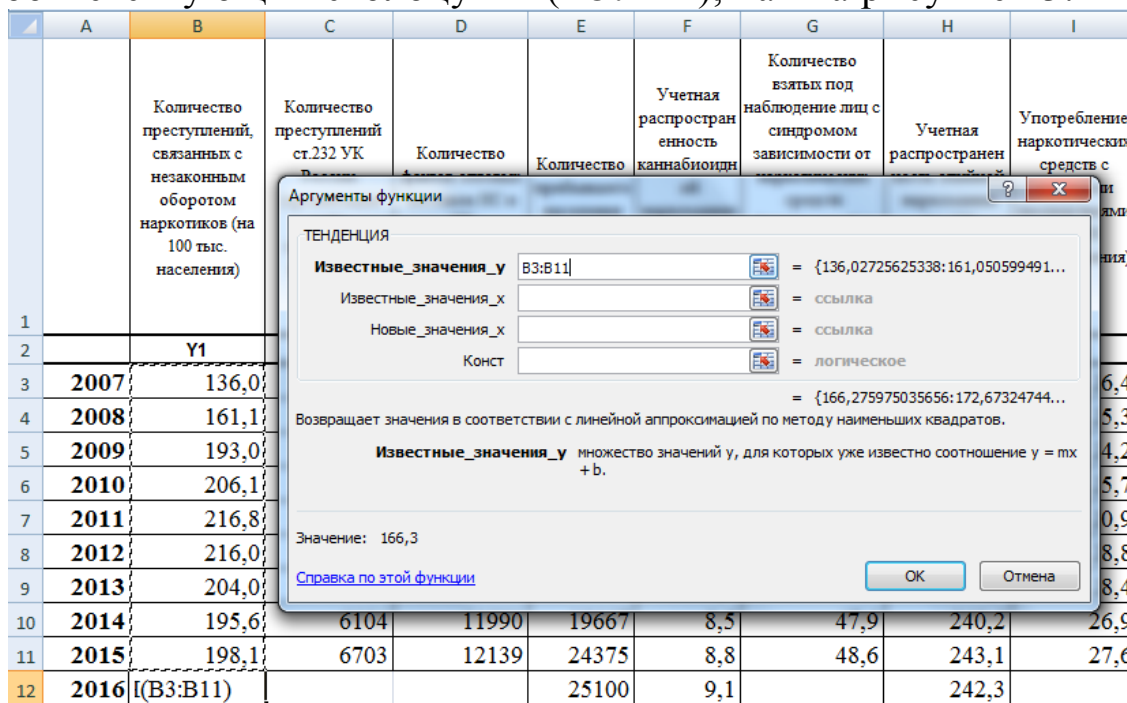


Рисунок 13. Аргументы функции ТЕНДЕНЦИЯ

5. Установите курсор в поле **Известные\_значения\_x** и выделите диапазон в соответствии с рисунком 14 (так как признак X4 оказывает наибольшее влияние на Y1).

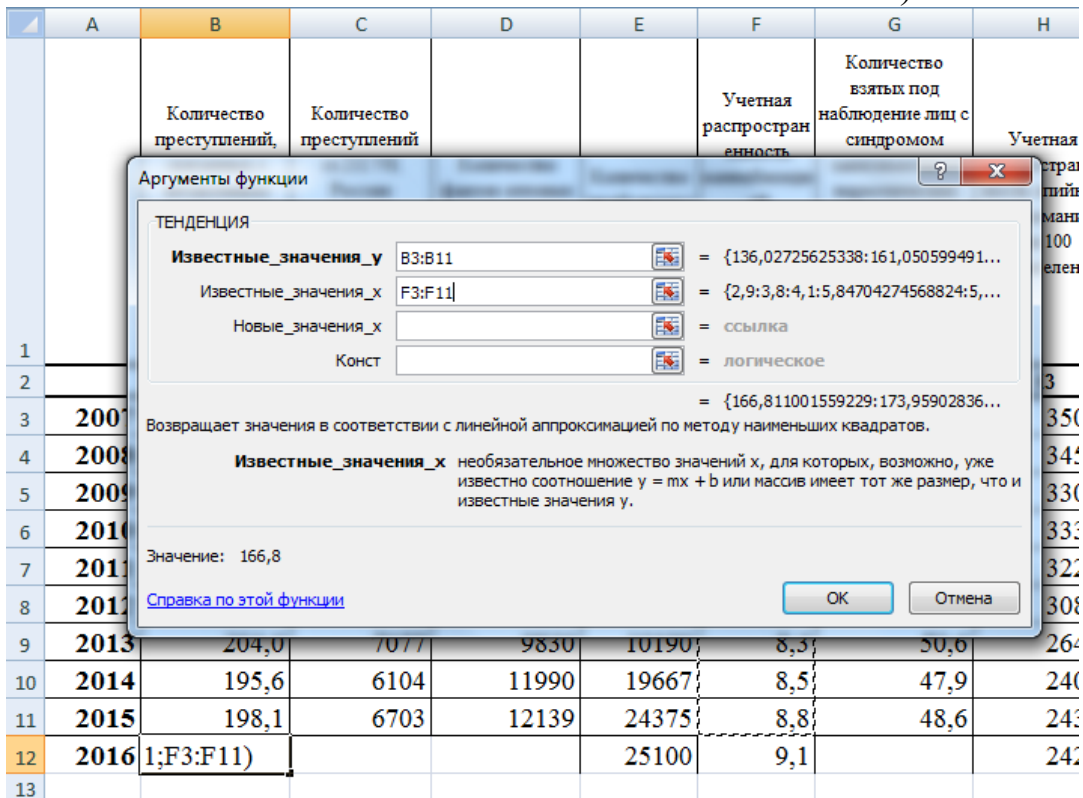


Рисунок 14. Аргументы функции ТЕНДЕНЦИЯ

6. В позиции **Новые\_значения\_x** установите диапазон предполагаемых значений факторного признака на прогнозный период (2016 год) в соответствии с рисунком 15. Поле **Конст** оставьте пустым.

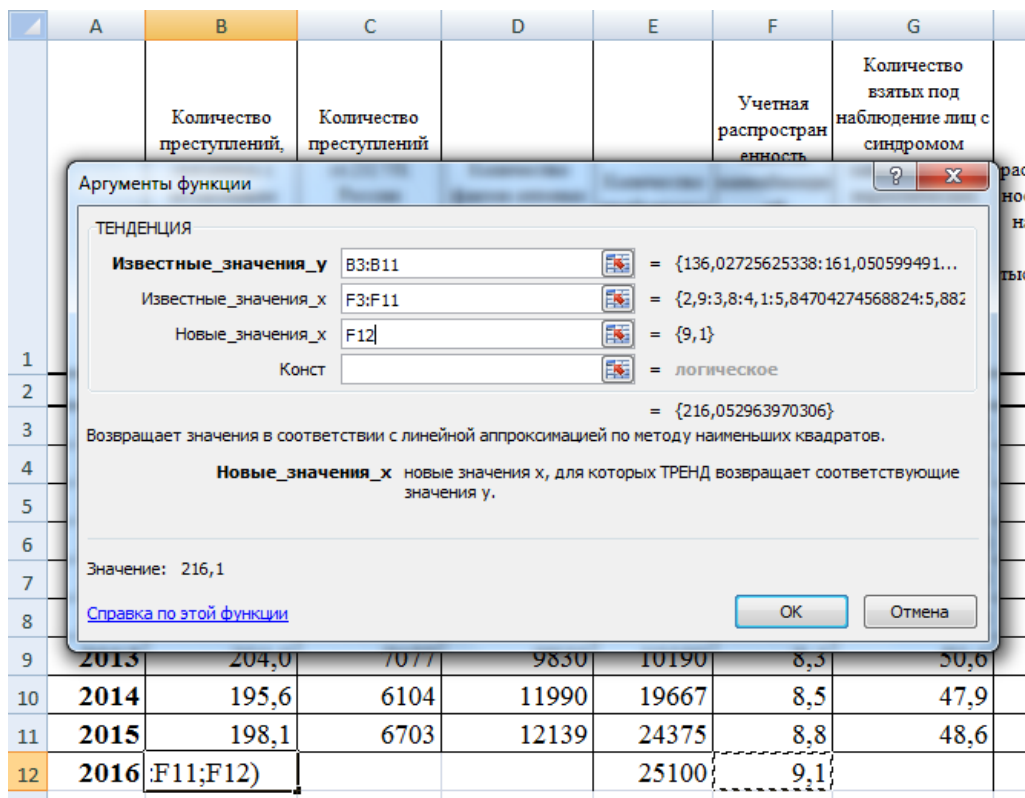


Рисунок 15. Аргументы функции ТЕНДЕНЦИЯ

7. Далее щелкнув по кнопке **ОК**, получим в ячейке B12 прогнозное значение количества преступлений, связанных с незаконным оборотом наркотиков (на 100 тыс. населения) в 2016 году.

8. Затем в соответствии с гипотезой о связи признаков вычислите остальные прогнозные значения показателей преступности на 2016 год. Значения аргументов функций приведены на рисунке 16.

		Y1	Y2	Y3	X1	X4
1						
2						
3	2007	136,0	3143	3178	8253	2,9
4	2008	161,1	4509	5590	7838	3,8
5	2009	193,0	5098	6175	7848	4,1
6	2010	206,1	5466	7099	7710	5,8
7	2011	216,8	6414	8193	7870	5,9
8	2012	216,0	6554	9287	8130	6,4
9	2013	=ТЕНДЕНЦИЯ(B3:B11;F3:F11;F12)				
10	2014	195,6	6104	11990	19667	8,5
11	2015	198,1	6703	12139	24375	8,8
12	2016	216,1	6962	12675	25100	9,1

Рисунок 16. Значения аргументов функций в 2016 году

Следует отметить, что в качестве известных и новых значений аргумента  $x$  можно указывать диапазон ячеек соответствующий нескольким факторным показателям, так например, в ячейке **D12 Известные\_значения\_x** диапазон **E3:F11, Новые\_значения\_x** – **E12:F12**. При этом их следует обязательно располагать в смежных столбцах данных, иначе функция выдаст ошибку.

Вместо функции **ТЕНДЕНЦИЯ** можно использовать функцию **РОСТ** которая прогнозирует экспоненциальную зависимость и идентична по своим параметрам.

### **ЗАДАНИЕ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ**

Получите прогнозные значения критериального показателя  $Y_1$  на 2016 год в зависимости от значений факторов  $X_2$ ,  $X_3$ ,  $X_5$ . Значения факторов на 2016 год выберите произвольно на основе тенденции их изменения.

### **МОДЕЛИ РЕГРЕССИОННОГО АНАЛИЗА**

Для установления связи между статистическими данными и получения прогнозных значений можно воспользоваться инструментами регрессионного анализа. Это средство предназначено для определения степени линейной зависимости столбца данных критериальных показателей ( $Y$ ) от нескольких линейно независимых (факторных) переменных ( $X$ ). Другими словами, регрессионный анализ сводится к нахождению коэффициентов  $\alpha_i$  в зависимости:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n.$$

В данном выражении  $Y$  и  $X_i$  – столбцы исходных данных. По сути, выводится уравнение зависимости переменной  $Y$  от переменных  $X_i$ .

Конечно, данный вид зависимости упрощает реально существующие связи в природе и обществе, но в ряде случаев позволяет определить наиболее значимые факторы для того или иного социального процесса.

Регрессионный анализ позволяет не только установить наличие линейной связи одной переменной от нескольких, но и определить коэффициенты этой зависимости. В дальнейшем на основе построенной модели можно определить

прогнозные значения результативных признаков (осуществить прогноз).

Рассмотрим два способа регрессионного анализа в MS Excel.

**Упражнение 10.** Для начала работы откройте, созданный вами ранее файл **Регрессионный анализ.xlsx** и перейдите на лист с именем **Прогноз**.

### 1 способ.

Согласно выдвинутым гипотезам определим коэффициенты регрессии для уравнений связей:

$$Y1 = \alpha_0 + \alpha_1 \times X4; \quad Y2 = \alpha_0 + \alpha_1 \times X3; \quad Y3 = \alpha_0 + \alpha_1 \times X1 + \alpha_2 \times X4.$$

Для этого:

1. Выделите диапазон ячеек **B26:C26**.

2. На вкладке **Формулы** в группе **Библиотека функций** выберите **Вставить функцию** и в категории **Статистические** укажите функцию **ЛИНЕЙН**.

3. В окне **Аргументы функции** для параметра **Известные\_значения\_у** укажите диапазон ячеек **B3:B11**.

4. Для параметра **Известные\_значения\_х** укажите диапазон ячеек **F3:F11**.

5. Параметру **Конст** присвойте значение 1 (ИСТИНА), а параметр **Статистика** оставьте пустым, как это изображено на рисунке 17. Далее нажмите клавишу **ОК**.

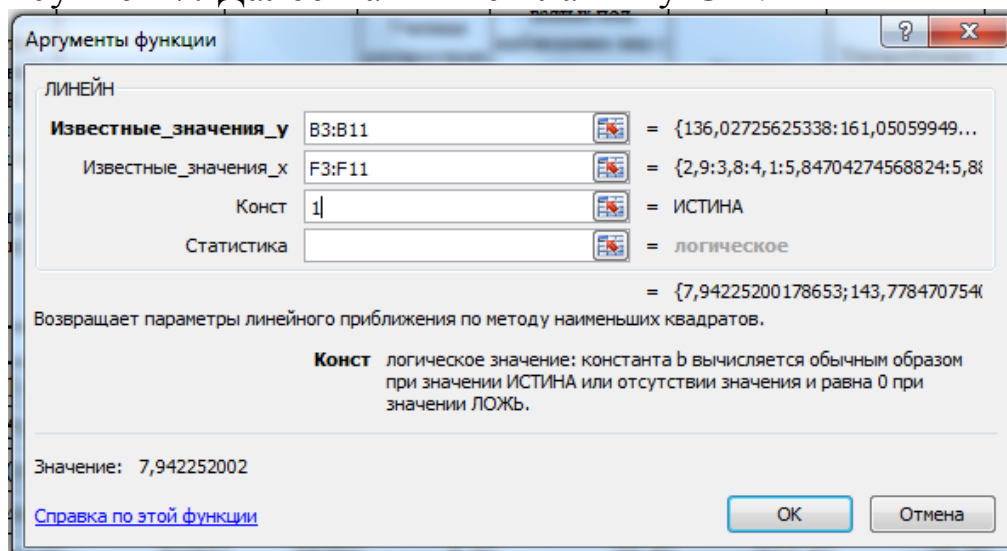


Рисунок 17. Аргументы функции ЛИНЕЙН

6. После ввода функции не снимая выделения с диапазона ячеек **B26:C26** нажмите клавишу **F2**, далее оставаясь в

режиме редактирования, нажмите комбинацию клавиш **Ctrl+Shift+Enter**. Важно, что данная формула должна вводиться только как формула массива и поэтому содержит по краям фигурные скобки



7. В ячейках **B26** и **C26** должны появиться соответственно значения коэффициентов  $\alpha_1$  и  $\alpha_0$  регрессии для первого уравнения.

Аналогично следует определить коэффициенты регрессии для второго и третьего уравнений связей:

- выделите диапазон ячеек **B27:C27**;
- в строке формул с клавиатуры введите **=ЛИНЕЙН(С3:С11;Н3:Н11;1)**;
- нажмите комбинацию клавиш **Ctrl+Shift+Enter**;
- выделите диапазон **B28:D28**;
- в строке формул введите **=ЛИНЕЙН(Д3:Д11;Е3:Е11;1)**;
- нажмите комбинацию клавиш **Ctrl+Shift+Enter**.

Значения коэффициентов регрессии для всех трех зависимостей изображены на рисунке 18.

The image shows a portion of an Excel spreadsheet with the following data:

23				
24				
25		$\alpha_1$	$\alpha_0$	
26	1	7,942	143,778	$\alpha_0$
27	2	-20,76827451	11993,91687	
28	3	1207,348719	0,060517734	169,5604214
29				
30		$\alpha_2$	$\alpha_1$	
31				

Рисунок 18. Коэффициенты уравнения регрессии

Функция **ЛИНЕЙН()** возвращает вначале коэффициенты перед независимыми переменными  $\alpha_i$  ( $\alpha_i, \alpha_2, \alpha_1$ ), а последним постоянную величину  $\alpha_0$ . Первым аргументом функции является блок значений зависимой величины, вторым – блок значений независимых переменных.

Важно, что при наличии нескольких независимых переменных, они должны быть расположены в соседних столбцах.

Используя полученные коэффициенты и общий вид уравнений регрессии, рассчитаем прогнозные значения за-

висимой переменной – количества зарегистрированных преступлений  $Y_1$ ,  $Y_2$ ,  $Y_3$ .

Для этого:

1. В ячейку **K3** введите формулу  $=\$C\$26+\$B\$26*F3$ , где к постоянному коэффициенту  $\alpha_0$  (ячейка C26) добавляется произведение коэффициента  $\alpha_1$  (ячейка B26) на значение признака  $X_4$  (ячейка F3).

2. Используя маркер заполнения, скопируйте содержимое ячейки **K3** в диапазон ячеек **K3:K12**.

3. В ячейку **L3** введите формулу  $=\$C\$27+\$B\$27*H3$ .

4. Скопируйте маркером заполнения ячейку **L3** в диапазон **L3:L12**.

5. В ячейку **M3** введите формулу  $=\$D\$28+\$C\$28*E3+\$B\$28*F3$ .

6. Скопируйте маркером заполнения ячейку **M3** в диапазон **M3:M12**.

Таблица полученных прогнозных значений признаков  $Y_1$ ,  $Y_2$ ,  $Y_3$  представлена на рисунке 19.

	K	L	M	N	O	P
	Прогнозные значения линейной аппроксимации					
	$=\$C\$26+\$B\$26*F3$		$=\$D\$28+\$C\$28*E3+\$B\$28*F3$			
	$Y_1$	$Y_2$	$Y_3$			
	166,8	4721	4170			
	174,0	4818	5232			
	$=\$C\$27+\$B\$27*H3$		5595			
	190,2	5077	7696			
	190,5	5294	7748			
	194,6	5580	8380			
	209,5	6502	10776			
	211,2	7006	11615			
	213,7	6945	12269			
	216,1	6962	12675			

Рисунок 19. Прогнозные значения признаков  $Y_1$ ,  $Y_2$ ,  $Y_3$

Постройте график прогнозных и действительных значений результативного признака  $Y_1$ , как это изображено на рисунке 20.

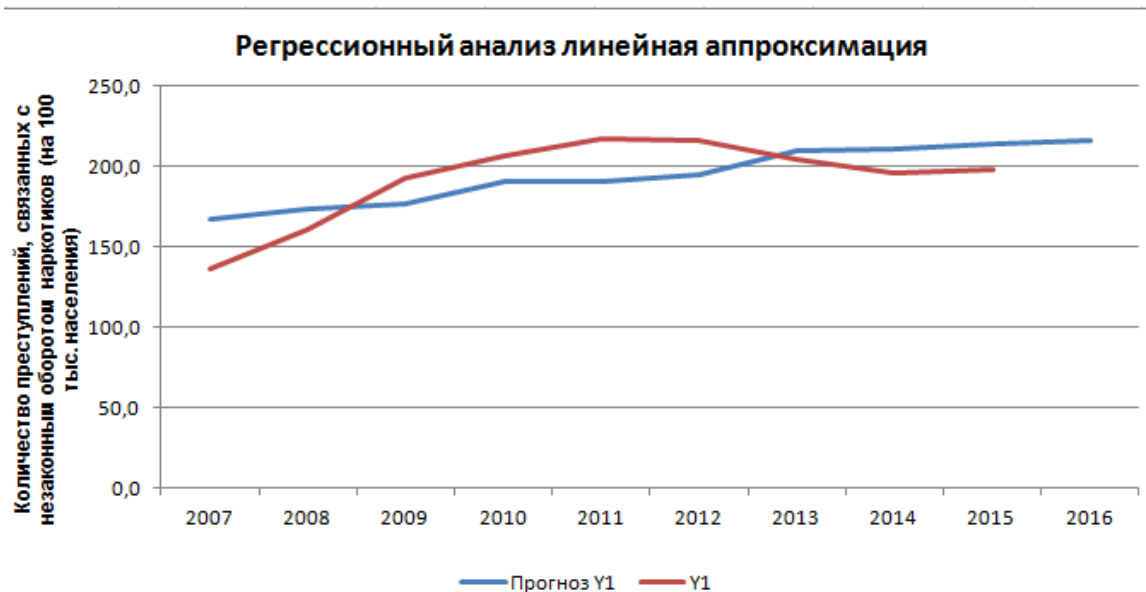


Рисунок 20. График прогнозных значений показателя Y1

Кроме линейной аппроксимации исходной зависимости (функция ЛИНЕЙН) используется и экспоненциальная кривая (функция ЛГРФПРИБЛ) которая позволяет лучше аппроксимировать исходные значения данных. При этом регрессионная модель (уравнение) изучаемых признаков будет выглядеть следующим образом (знак  $\wedge$  обозначает степень):

$$Y1 = \alpha_0 \times \alpha_1^X$$

$$Y2 = \alpha_0 \times \alpha_1^X$$

$$Y3 = \alpha_0 \times \alpha_1^X \times \alpha_2^X$$

Для определения коэффициентов уравнения регрессии для экспоненциальной кривой:

- выделите диапазон ячеек **B32:C32**;
- в строке формул введите **=ЛГРФПРИБЛ(B3:B11;F3:F11;1)**;
- нажмите комбинацию клавиш **Ctrl+Shift+Enter**;
- выделите диапазон ячеек **B33:C33**;
- в строке формул введите **=ЛГРФПРИБЛ(C3:C11;H3:H11;1)**;
- нажмите комбинацию клавиш **Ctrl+Shift+Enter**;
- выделите диапазон **B34:D34**;
- в строке формул введите **=ЛГРФПРИБЛ(D3:D11;E3:F11;1)**;
- нажмите комбинацию клавиш **Ctrl+Shift+Enter**.

Таблица значений коэффициентов уравнения регрессии для экспоненциального приближения изображена на рисунке 21.

	A	B	C	D
31				
32	1	1,047643058	143,3441209	
33	2	0,995959007	18947,15919	
34	3	1,218266914	0,999993507	2475,77023
35				

Рисунок 21. Коэффициенты уравнения регрессии

Далее постройте прогнозные значения результативных признаков:

1. В ячейку **O3** введите формулу  $=\$C\$32*\$B\$32^{\$F3}$ .
2. Используя маркер заполнения, скопируйте содержимое ячейки **O3** в диапазон ячеек **O3:O12**.
3. В ячейку **P3** введите формулу  $=\$C\$33*\$B\$33^{\$H3}$ .
4. Скопируйте маркером заполнения ячейку **P3** в диапазон **P3:P12**.
5. В ячейку **Q3** введите формулу  $=\$D\$34*\$C\$34^{\$E3}*\$B\$34^{\$F3}$ .
6. Скопируйте маркером заполнения ячейку **Q3** в диапазон **Q3:Q12**.

Постройте график прогнозных и действительных значений результативного признака **Y3**, как это изображено на рисунке 22.

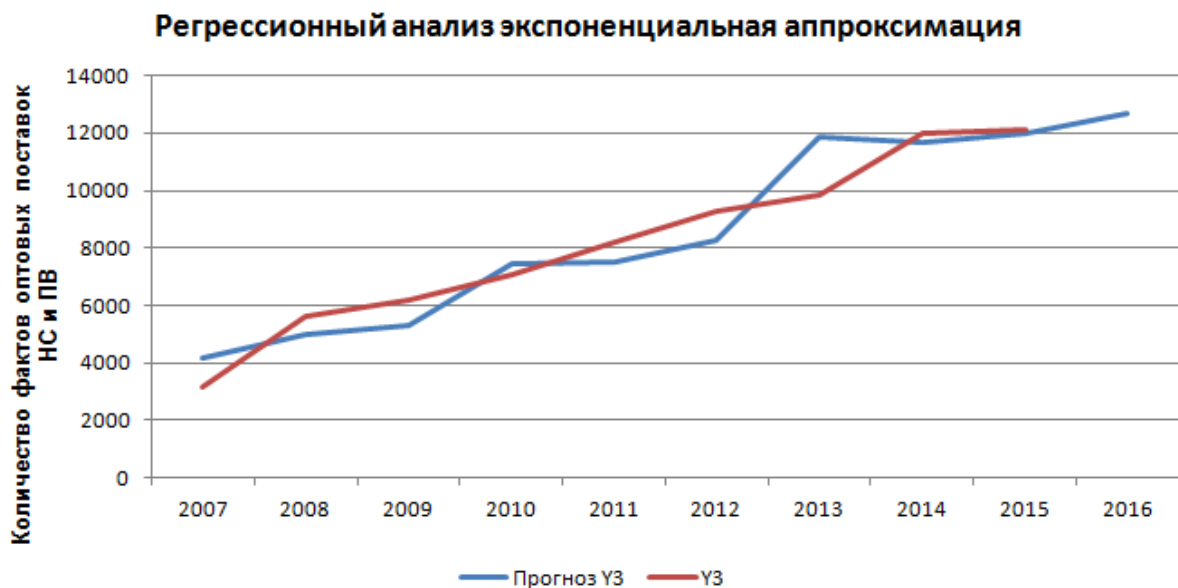


Рисунок 22. График прогнозных значений

## 2 способ.

Воспользуемся надстройкой **Регрессия** из **Пакета анализа MS Excel** для построения регрессионной линейной модели исследуемых зависимостей.

1. Скопируйте таблицу данных из листа **Прогноз** на новый лист с именем **Регрессия**.

2. Перейдите на вкладку **Данные** и в группе **Анализ** нажмите **Анализ данных**. В окне **Анализ данных** выберите инструмент **Регрессия**.

3. В открывшемся диалоговом окне следует задавать входные интервалы зависимой переменной  $Y$  и независимых переменных  $X$  аналогично, как это делалось ранее первым способом, а также выходной интервал (адрес верхней левой ячейки – в первом случае – **К3**, во втором – **К23**, в третьем – **К43**). Параметры инструмента **Регрессия** для первой модели представлены на рисунке 23.

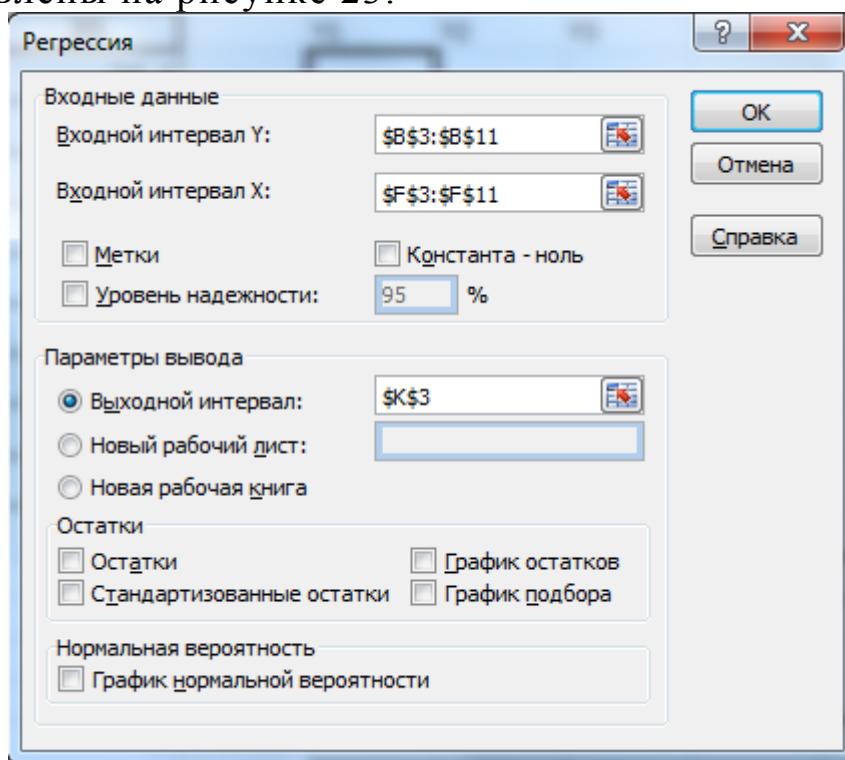


Рисунок 23. Параметры инструмента Регрессия

На рисунке 24 представлены результаты регрессионного анализа для признаков  $X_4$  и  $Y_1$ .

Вывод итогов								
Регрессионная статистика								
Множественный R	0,645166141							
R-квадрат	0,41623935							
Нормированный R-квадрат	0,332844971							
Стандартная ошибка	21,75737335							
Наблюдения	9							
Дисперсионный анализ								
	df	SS	MS	F	значимость F			
Регрессия	1	2362,758236	2362,758	4,991216	0,060603			
Остаток	7	3313,683065	473,3833					
Итого	8	5676,441301						
	Коэффициенты	Стандартная ошибка	t-статистика	P-значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	143,7784708	22,71290694	6,330254	0,000393	90,07098	197,486	90,07098	197,486
Переменная X 1	7,942252002	3,555007195	2,234103	0,060603	-0,464	16,34851	-0,464	16,34851

Рисунок 24. Регрессионный анализ признаков X4 и Y1

Степень качества прогнозных значений можно определить при помощи параметра **R-квадрат**, размещенного в блоке вывода. Принято считать данные хорошо согласованными, если это значение более 0,8.

Если значение параметра R-квадрат менее 0,8, то делается вывод, что для аппроксимации исходных данных необходимо учитывать влияние и других факторов.

Вычислите коэффициенты уравнений регрессии вторым способом для трёх линейных моделей. Убедитесь, что значения коэффициентов совпадают с приведёнными ниже:

$$Y1 = 143,78 + 7,94 \times X4; \quad Y2 = 11993,92 - 20,77 \times X3;$$

$$Y3 = 169,56 + 0,06 \times X1 + 1207,35 \times X4$$

### ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

**Задание 1.** По имеющимся статистическим данным по городу N рассчитайте коэффициент корреляции между признаками. Сделайте выводы о наличии линейной зависимости между исходными данными. С помощью функции ТЕНДЕНЦИЯ спрогнозируйте количество зарегистрированных больных наркоманиями в 2016 году. Определите коэффициенты

линейного уравнения регрессии для количества совершенных преступлений от количества зарегистрированных больных наркоманией. На основе полученных коэффициентов регрессии, рассчитайте прогнозные значения переменной количества зарегистрированных преступлений, связанных с НОН на 2016 год.

Год	2008	2009	2010	2011	2012	2013	2014	2015	2016
Количество преступлений, связанных с НОН, всего	3554	3855	3582	3209	2994	2628	2404	2302	
Зарегистрировано больных наркоманией, чел.	209	192	174	173	159	153	138	122	

**Задание 2.** По имеющимся статистическим данным по городу N рассчитайте коэффициент корреляции между признаками. Сделайте выводы о наличии линейной зависимости между исходными данными. С помощью функции ТЕНДЕНЦИЯ спрогнозируйте количество вынужденных переселенцев и беженцев в 2016 году. Определите коэффициенты линейного уравнения регрессии для количества совершенных преступлений от количества вынужденных переселенцев и беженцев. Используя полученные коэффициенты регрессии, рассчитайте прогнозные значения переменной количества зарегистрированных преступлений на следующий период.

Год	2008	2009	2010	2011	2012	2013	2014	2015	2016
Количество вынужденных переселенцев и беженцев, чел.	8914	7177	6802	4337	4092	2135	2096	2554	
Количество преступлений, связанных с НОН, всего	3554	3855	3582	3209	2994	2628	2404	2302	

**Задание 3.** Используя исходные данные из упражнений 1 и 2, определите коэффициенты линейного уравнения регрессии для количества совершенных преступлений, связанных с НОН от количества зарегистрированных больных наркоманией.

нией и количества вынужденных переселенцев и беженцев. На основе полученных результатов в заданиях 1-3 постройте графики сравнения прогнозных и исходных данных количества преступлений, связанных с НОН. Сделайте вывод о качестве аппроксимации в каждом случае.

## 7. ЭКСПЕРТНЫЕ ОЦЕНКИ

Кроме социологических опросов и анализа статистической информации, предоставляемой ведомствами – участниками мониторинга наркоситуации, при оценке наркоситуации в регионе используется мнение соответствующих экспертов по вопросам, касающимся оборота наркотиков, а также противодействия их незаконному обороту, профилактики немедицинского потребления наркотиков, лечения, медицинской и социальной реабилитации больных наркоманией. Для определения наиболее значимого мнения проводят математическую обработку данных опроса экспертов.

Анализ и обработка информации проводится по сформулированным ранее критериям (шкалам ценности). В тех случаях, когда требуется расположить объекты по степени приоритетности или значимости, используют *метод предпочтений* или *метод среднего*.

В методе предпочтений каждого из экспертов просят расположить объекты в порядке значимости. Наиболее значимому объекту эксперт присваивает номер 1, далее 2 и т.д. Если, по мнению эксперта, несколько объектов имеют одинаковую значимость, то им присваиваются равные номера.

**Упражнение 1.** В таблице приведены результаты экспертных оценок наркоситуации в семи регионах. Ранг 1 имеют регионы, в которых наркоситуация наименее напряженная, максимальный ранг эксперт присваивает территории, в которой наркоситуация наиболее неблагоприятная.

### Ранжирование регионов

Номера экспертов, <i>k</i>		Номера регионов, <i>j</i>						
		1	2	3	4	5	6	7
1	Иванов И.И.	3	1	2	4	3	3	5
2	Петров П.П.	4	1	2	3	3	4	4
3	Сидоров С.С.	3	2	1	3	4	5	3
4	Трофимов Т.Т.	2	2	1	3	3	4	5
5	Федоров Ф.Ф.	2	2	1	3	4	5	4

Результаты ранжирования приводятся к стандартизованному виду. Стандартизация заключается в перерасчете рангов. На первом этапе стандартизации регионы располагаются в порядке возрастания напряженности наркоситуации, определенном каждым экспертом. Каждому региону присваивается ранг, равный порядковому номеру. На втором этапе территориям, получившим одинаковые оценки у эксперта, присваивается их среднеарифметический номер. Например, первый и второй регионы у пятого эксперта будут стоять на втором и третьем местах и их стандартный ранг равен 2,5. Результаты стандартизации отражены в следующей таблице.

### Стандартизованные ранги объектов, $R_{kj}$

Номера экспертов, <i>k</i>		Номера объектов, <i>j</i>						
		1	2	3	4	5	6	7
1	Иванов И.И.	4	1	2	6	4	4	7
2	Петров П.П.	6	1	2	3,5	3,5	6	6
3	Сидоров С.С.	4	2	1	4	6	7	4
4	Трофимов Т.Т.	2,5	2,5	1	4,5	4,5	6	7
5	Федоров Ф.Ф.	2,5	2,5	1	4	5,5	7	5,5
$S_j$		<b>3,8</b>	<b>1,8</b>	<b>1,4</b>	<b>4,4</b>	<b>4,7</b>	<b>6,0</b>	<b>5,9</b>

После стандартизации рангов рассчитывается групповая оценка (среднее место) региона по формуле:

$$S_j = \frac{1}{n} \sum_{k=1}^n R_{kj}.$$

Здесь  $R_{kj}$  – стандартизованный ранг  $j$ -го региона, данный  $k$ -ым экспертом, а  $n$  – число экспертов.

При использовании метода средней оценки в качестве стандартизованных рангов  $R_{kj}$  выступают непосредственно числовые данные, приведенные экспертами.

Определение степени согласованности экспертов является важным этапом перед принятием окончательного решения. При использовании метода предпочтений согласованность мнений экспертов определяется коэффициентом конкордации  $W$ , рассчитываемым по формулам:

$$W = \frac{12 \cdot D}{n^2 - 1 - T/m}, \quad D = \frac{1}{m} \sum_{j=1}^m (S_j - \bar{S})^2, \quad \bar{S} = \frac{1}{m} \sum_{j=1}^m S_j.$$

В приведенных формулах используются  $S_j$ , рассчитанные без учета компетентности. В этом случае их сумма равна сумме номеров регионов. Учитывая, что в нашем случае число регионов  $m=7$ , среднее значение рангов будет равно 4.

Параметр  $T$  определяется количеством одинаковых оценок, данных регионам каждым экспертом. Пусть  $L_k$  – количество одинаковых групп данных у  $k$ -того эксперта, а  $t_l$  – количество одинаковых рангов в  $l$ -той группе. Тогда параметр  $T$  может быть вычислен по формулам:

$$T_k = \sum_{l=1}^{L_k} (S - t_l)^2, \quad T = \frac{1}{n} \sum_{k=1}^n T_k.$$

В рассматриваемом примере у второго эксперта две группы одинаковых оценок, в которых два и три элемента. Следовательно,  $T_2 = (3^3 - 2) + (3^3 - 3) = 30$ . Параметры расчета коэффициента конкордации приведены в таблице:

	1	2	3	4	5
$L_k$	1	2	1	2	2
$T_k$	24	30	24	12	12
$D=2,84$	$T=20,4$		$W=0,76$		

Для рассматриваемого случая значение коэффициента оказалось равным 0,76, что означает довольно высокую степень согласованности. Значение коэффициента конкордации может изменяться от нуля до единицы.

При использовании метода средней оценки мерой согласованности экспертных оценок выступает относительная вариация, рассчитываемая по формуле:

$$V_j = \frac{\sqrt{\frac{1}{n} \sum_{k=1}^n (R_{kj} - S_j)^2}}{S_j}.$$

Чем меньше вариация оценки, тем более согласованы мнения экспертов. Приемлемыми можно считать вариации на уровне до 0,3 – 0,5.

В том случае, когда согласованность оценок не велика, проводят выявление оценок тех экспертов, которые более всего повлияли на снижение коэффициента конкордации. Для этого из общего списка оценок исключают данные какого-либо эксперта и вновь рассчитывают коэффициент конкордации. Эту процедуру проводят для всех экспертов. Повторив описанные действия несколько раз, выявляют группу из нескольких человек, отрицательно повлиявших на степень согласованности. Для этой группы отдельно вычисляют свой коэффициент согласованности. Может оказаться так, что внутри отбракованной группы мнения согласованы. Причиной этому может быть либо существование двух противоположных подходов в группах на оценку регионов, либо различие в информации, которой располагали эксперты.

При выявлении расхождения в оценках регионов рабочая группа знакомит экспертов с предварительными результатами. Эксперты в форме дискуссии обмениваются мнениями, и опрос повторяется с последующей обработкой результатов.

### **ЗАДАНИЕ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ**

Проведите обработку результатов опроса экспертов по ранжированию семи территорий, приведенных в таблице. Определите уровень согласованности мнений экспертов. Рассчитайте уровни согласованности оценок каждой пары экспертов.

#### **Результаты ранжирования регионов экспертами**

<b>Вариант 1</b>							
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<b>Эксперт 1</b>	7	6	5	4	3	2	1
<b>Эксперт 2</b>	6	7	4	5	2	3	1
<b>Эксперт 3</b>	4	3	3	3	1	1	2
<b>Эксперт 4</b>	5	3	3	1	2	2	4
<b>Эксперт 5</b>	5	4	4	3	2	1	2

<b>Вариант 2</b>							
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<b>Эксперт 1</b>	5	5	5	4	3	2	1
<b>Эксперт 2</b>	6	7	4	5	2	3	1
<b>Эксперт 3</b>	3	4	4	3	1	1	2
<b>Эксперт 4</b>	5	3	3	1	2	2	4
<b>Эксперт 5</b>	5	4	4	3	2	1	2
<b>Вариант 3</b>							
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<b>Эксперт 1</b>	1	6	5	4	3	2	1
<b>Эксперт 2</b>	6	7	4	5	2	3	1
<b>Эксперт 3</b>	4	5	3	3	1	2	2
<b>Эксперт 4</b>	5	3	3	1	4	2	1
<b>Эксперт 5</b>	5	4	4	3	2	1	2
<b>Вариант 4</b>							
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<b>Эксперт 1</b>	5	6	5	4	2	2	1
<b>Эксперт 2</b>	6	7	4	5	2	3	1
<b>Эксперт 3</b>	4	3	5	3	1	1	2
<b>Эксперт 4</b>	5	3	3	1	2	2	4
<b>Эксперт 5</b>	5	6	4	3	2	1	2
<b>Вариант 5</b>							
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<b>Эксперт 1</b>	6	6	5	3	3	2	1
<b>Эксперт 2</b>	6	5	4	5	2	2	3
<b>Эксперт 3</b>	4	3	3	3	1	1	2
<b>Эксперт 4</b>	5	3	3	1	2	2	4
<b>Эксперт 5</b>	5	4	4	3	2	1	3

## СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Горбач, Н.А. Методология изучения и сохранения здоровья участников образовательного процесса в вузе / Н.А. Горбач и др. – Красноярск: СибЮИ ФСКН России, 2012. – 248 с.
2. Ефимов, С.Н. Правовая статистика : учебно-методическое пособие / С.Н. Ефимов, В.В. Молоков. – Красноярск : СибЮИ МВД России, 2007. – 47 с.
3. Ковалева, Т.Ю. Практикум по теории статистики : учебно-практическое пособие / Т.Ю. Ковалева. – М.: КНО-РУС, 2012. – 376 с.
4. Литвинов, В.А. Информатика и математика : практический курс / В.А. Литвинов. – Барнаул: БЮИ МВД России, 2004. – 212 с.
5. Литвинов, В.А. Информатика и информационные технологии в профессиональной деятельности : учебное пособие / В.А. Литвинов. – Барнаул : БЮИ МВД России, 2013. – 536 с.
6. Молоков, В.В. Методы автоматизированной обработки результатов мониторинга наркоситуации : практическое пособие / В.В. Молоков, А.С. Шерстяных. – Красноярск: СибЮИ МВД России, 2016. URL: <http://lib.sibli.ru/elib/0052595.pdf>.
7. Основы теории вероятностей и математической статистики : учебное пособие / Н.И. Косарев [и др.]. – Красноярск : СибЮИ МВД России, 2002. – 75 с.
8. Правовая статистика : учебник для студентов вузов, обучающихся по специальности «Юриспруденция», для курсантов и слушателей образовательных учреждений МВД / под ред. В.С. Лялина, А.В. Симоненко. – 2-е изд., перераб. и доп. – М.: ЮНИТИ-ДАНА, 2010. – 255 с.
9. Правовая статистика : учебник и практикум для академического бакалавриата / под общ. и науч. ред. Л.К. Савюка. – М.: Издательство Юрайт, 2017. – 409 с.

Павел Викторович Галушин,  
Вячеслав Витальевич Молоков,  
Александра Сергеевна Шерстяных

## **ПРАВОВАЯ СТАТИСТИКА**

Учебно-практическое пособие

Подготовлено к изданию Ю.В. Леонтьевой

Подписано в печать 20.06.2017

Формат Р 60x84. Бумага типографская. Гарнитура Times.

Печать офсетная. 3,36 уч.-изд. л. (6,25 усл.печ.л.).

Тираж 70 экз. Заказ 67.

Научно-исследовательский и редакционно-издательский отдел.  
Сибирский юридический институт МВД России.  
660131, г. Красноярск, ул. Рокоссовского, 20.

---

Отпечатано в типографии НИРИО СибЮИ МВД России.  
660050, г. Красноярск, ул. Кутузова, 6.