Министерство внутренних дел Российской Федерации
Академия управления МВД России

Д. В. Васильев, В. А. Ребрий

СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ ПРАВОВЫХ ЯВЛЕНИЙ И ПРОЦЕССОВ

Учебно-методическое пособие

Москва • 2018

УДК 31:343
ББК 67.5
В19

*Одобрено редакционно-издательским советом
Академии управления МВД России*

Рецензенты: Д. А. Артунян, начальник отдела УПЗиИП УПУ ДПД МВД России, канд. юрид. наук; А. В. Голубев, заместитель начальника Штаба ГУ МВД России по г. Москве – начальник Инспекции; М. П. Киреев, профессор кафедры управления органами внутренних дел центра командно-штабных учений Академии управления МВД России, д-р юрид. наук, проф.

Д. В. Васильев, В. А. Ребрий.
В19 Статистическое исследование правовых явлений и процессов : учебно-методическое пособие / Д. В. Васильев, В. А. Ребрий. – М. : Академия управления МВД России, 2017. – 84 с.

ISBN 978-5-906942-40-1

В учебно-методическом пособии рассмотрены базовые понятия, необходимые для овладения статистическими исследовательскими методами: статистическая закономерность, статистический признак, статистический показатель, вариация признака, шкала измерения и т. д. В пособии изложены типовые исследовательские задачи количественного анализа правовых явлений и процессов: изучение структурных параметров явления, динамики явления, исследование взаимосвязи явлений.

Издание предназначено для обучающихся по направлению подготовки 40.04.01 Юриспруденция и позволяет им освоить базовые методы статистического анализа, которые могут быть использованы в аналитической деятельности органов внутренних дел.

УДК 31:343
ББК 67.5

ISBN 978-5-906942-40-1

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Учебно-методическое пособие подготовлено в соответствии с рабочей программой учебной дисциплины «Статистическое исследование правовых явлений и процессов» для магистрантов, обучающихся в Академии управления МВД России по основной профессиональной образовательной программе по направлению 40.04.01 Юриспруденция, при изучении ими общих проблем статистики, а также при овладении принципами и методами статистического анализа на конкретном материале, относящемся к правовой тематике. В учебно-методическом пособии раскрываются понятия статистической совокупности, статического наблюдения, статистического признака, показателя и др., излагаются описательные и аналитические статистические методы, применяемые в статистических исследованиях правовых явлений и процессов. Данное учебно-методическое пособие вооружит обучающихся современным инструментарием аналитической обработки статистической информации, по уровню сложности вполне доступным для руководителя любого звена правоохранительных органов.

Кроме того, программа курса и соответствующее ей данное издание были подготовлены с учетом потребностей информационно-аналитической деятельности правоохранительных органов и, в частности, органов внутренних дел. При изучении программы курса обучающийся со школьным уровнем математических знаний, тем не менее, способен освоить многие важнейшие статистические принципы и методы, особенности их практического применения.

Необходимо отметить, что современный статистический анализ оперирует данными, имеющими как качественное, так и количественное измерение, поэтому в учебно-методическом пособии значительное место уделено разъяснению основ теории измерения, включая понятие измерительной шкалы и типов шкал: номинальной, ранговой, интервальной и абсолютной. Примеры шкал, приводимые в пособии, имеют прямое отношение к деятельности органов внутренних дел при проведении собственного исследования правовых проблем.

Для любой научной отрасли (научного исследования) важно не только определить предмет исследования, описать его, собрать эмпирические данные, но показать изучаемое явление как закономерное, обусловленное определенными факторами, объяснить рассматриваемое явление, т. е. выявить закономерности его становления, функционирования и развития. Только понимание закономерного характера исследуемого явления (процесса) позволяет ученому делать научный прогноз, а практику – осуществлять управление в соответствующей сфере. В этой связи в методическом пособии подробно рассмотрены проблемы количественного оценивания тесноты связи явлений для наиболее используемых в эмпирических исследованиях данных – номинальных и порядковых.

1. СТАТИСТИЧЕСКИЕ ПРИЗНАКИ И ПОКАЗАТЕЛИ В ОПИСАНИИ ПРАВОВЫХ ЯВЛЕНИЙ И ПРОЦЕССОВ. СРЕДНИЕ ВЕЛИЧИНЫ И ИЗУЧЕНИЕ ВАРИАЦИИ

1.1. Сущность и значение статистических показателей

Статистика изучает массовые явления и процессы в их количественных измерениях. Но числа, применяемые в статистике, это не абстрактные числа математики, характеризуемые только величиной, знаком, формой (целые/дробные; мнимые/действительные; рациональные/иррациональные и т. п.). Статистика применяет, иначе говоря, не числа, а показатели, точнее статистические показатели.

Не правильно уяснив содержание, форму, свойства того или иного статистического показателя, нельзя корректно применить его в анализе социально-экономических явлений и процессов и понять смысл статистической информации.

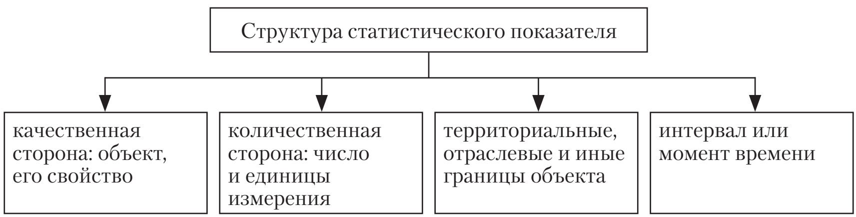
С философской точки зрения, статистический показатель – это мера, т. е. единство качественного и количественного отражения свойств объективных явлений и процессов в научном сознании. Поскольку статистика, как уже отмечалось, изучает массовые явления, статистический показатель – это обобщающая характеристика какого-то свойства совокупности, группы. Этим он отличается от индивидуальных значений, которые называются признаками. Например, средняя продолжительность ожидаемой жизни родившегося поколения людей в стране – статистический показатель, а продолжительность жизни конкретного человека – признак.

Рассмотрим содержание и форму статистического показателя на примере 20 065 совершенных преступлений на территории, например, выдуманной нами Петровской области в 2013 г. Показателем является не одно только число 20 065, а весь текст, поясняющий его содержание. Качественная сторона этого показателя – совершенное преступление. Далее статистический показатель имеет количественную сторону, которая выражается числом и единицей измерения – 20 065 преступлений.

Не всегда статистический показатель является именованным числом. Он может быть абстрактным и отвлеченным числом без наименования, может быть выражен в долях единицы: в процентах, промилле и т. п. Именованными числами являются абсолютные статистические показатели.

Статистический показатель имеет указание на территориальные границы (преступность на территории Петровской области) и границы во времени – 2013 г. Без указания территориальных, отраслевых или ведомственных границ объекта и привязки к определенному интервалу времени или моменту статистический показатель не существует. Структура статистического показателя представлена на схеме 1.

Схема 1



Являясь отображением каких-либо свойств изучаемых явлений и процессов, статистический показатель служит орудием их познания. Но всякое знание всегда ограничено, неполно соответствует изучаемому объекту. Ни один статистический показатель, ни целая их система не могут отразить все свойства, все особенности объекта и даже часть этих свойств с абсолютной точностью. Статистический показатель – приближенное, неточное и неполное отображение свойств изучаемого объекта, доступное при имеющемся уровне знаний и возможностях учета, измерения, сбора и передачи информации.

Остановимся на соотношении между признаком и статистическим показателем.

Признак – это свойство, присущее единице совокупности. Признак входит в качественное содержание показателя, он существует объективно независимо от того, отражает ли его наука с помощью тех или иных показателей. Например, возраст человека – это его признак, который можно измерять с разной степенью точности – в годах, месяцах, сутках или охарактеризовать датой рождения.

Показатель – характеристика группы единиц или совокупности в целом. Его построение зависит от цели исследования и изобретательности статистика. Средний возраст работников фирмы или жителей города – это статистические показатели, дающие возрастную характеристику определенных групп, совокупностей людей. Другим видом возрастных показателей может служить ряд распределения людей по возрасту и вычисленные на основе этого ряда системы показателей для характеристики структуры такого ряда и размеров вариации.

Объектами статистического исследования могут быть самые разнообразные явления и процессы. Поэтому чрезвычайно велико и разнообразие статистических показателей (табл 1.1).

Таблица 1.1

Классификация статистических показателей

По качественной стороне показателей	**По количественной стороне показателей**	**По отношению к характеризуемому свойству**
показатели конкретных свойств объектов	абсолютные	прямые
показатели статистических свойств любых массовых явлений	относительные	обратные

Показатели конкретных свойств изучаемого объекта – это, например, средний возраст сотрудника органа внутренних дел, количество совершенных преступлений и т. д. Особенностью этих показателей является то, что они формируются не только статистикой: в построении этих показателей качественное их содержание определяется конкретной предметной наукой: показатель рождаемости – демографией, показатель внутреннего валового продукта – теорией экономики, показатели эффективности правоохранительной деятельности – теорией управления в правоохранительной сфере. Статистика отвечает за методику учета или расчета количественной стороны этих показателей и их форму.

Совершенно иначе обстоит дело с показателями статистических свойств любых массовых явлений и процессов, не зависящих от конкретного содержания этих явлений. К таким статистическим показателям относятся: средние величины, показатели вариации, показатели связи признаков, показатели структуры и характера распределения, показатели скорости и темпов изменения, показатели колеблемости в динамике. К ним же относятся статистические оценки степени точности и надежности любых конкретных статистических показателей, полученных при выборочном изучении совокупности, а также оценки надежности и точности статистических прогнозов. За качественную, как и количественную сторону этих показателей, за их построение, интерпретацию и применение отвечает только статистика, а не какая-либо иная научная дисциплина. Система таких показателей создается и совершенствуется в ходе развития методов статистики.

Теоретическая статистика разрабатывает и изучает содержание, форму, методы расчета этих показателей в общем виде: что такое средняя арифметическая величина, коэффициент вариации, уравнение тренда ряда динамики. Если же любой из этих показателей рассчитан для определенного объекта, признака, периода времени, то он становится уже конкретным показателем.

Статистические показатели подразделяются на *абсолютные* и *относительные*.

Абсолютным показателем является показатель, который отражает либо суммарное число единиц, либо суммарное свойство объекта. Например, количество уголовных дел, направленных в суд территориальным органом внутренних дел в 2013 г., количество сотрудников органов внутренних дел, совершивших дисциплинарные проступки за конкретный месяц или год, и т. п.

Абсолютные показатели, как правило, выражаются именованными величинами в натуральных единицах измерения: тоннах, штуках, часах, амперах и т. п.; в условных единицах: условном топливе, нормо-сменах и т. д.; в стоимостных единицах: рублях, долларах, марках. Они характеризуют сумму значений первичных признаков объекта. Очевидно, что наука не может ограничиваться характеристиками только изолированных отдельных свойств объекта, поэтому статистика не ограничивается абсолютными показателями. Она измеряет и характеризует соотношение разных абсолютных величин, их изменения во времени, взаимосвязи между собой и окружающей средой. Статистика, как и все науки, широко пользуется общенаучными методами сравнения, обобщения, синтеза.

Относительным показателем является показатель, полученный путем сравнения, сопоставления абсолютных или относительных показателей в пространстве (между объектами), во времени (по одному и тому же объекту) или сравнения показателей разных свойств изучаемого объекта.

Относительные статистические показатели, получаемые при сопоставлении абсолютных показателей, могут быть названы относительными показателями первого порядка, а полученные при сопоставлении относительных же показателей – показателями высших (второго, третьего и т. д.) порядков. Показатели выше четвертого порядка ввиду сложности интерпретации почти никогда не применяются. Относительные статистические показатели выражают связь между абсолютными показателями: уровень преступности – отношение количества совершенных преступлений к общей численности населения; доля сотрудников органов внутренних дел с высшим юридическим образованием – отношение количества таких сотрудников к общему количеству сотрудников органов внутренних дел.

Основные виды относительных величин чаще выражаются отвлеченными числами, но могут быть также именованными относительными показателями. Построение их связано с применением различных методов статистики.

Относительные показатели можно подразделить на следующие группы:

1. Относительные показатели, характеризующие структуру объекта. Это доля (удельный вес) – отношение части к целому. Например, отношение количества преступлений против личности к общему количеству преступлений; числа женщин-полицейских к общему количеству сотрудников органов внутренних дел. В эту же группу входят характеристики отношения между отдельными частями объекта; показатели, характеризующие степень сложности структуры, степень неравномерности (вариации) долей и др. Доли выражаются нередко в процентах или промилле (тысячных долях).

2. Относительные показатели, характеризующие динамику процесса, изменение во времени. Это отношения показателей, характеризующих объект в более позднее время (текущий период), к аналогичным показателям того же объекта в более ранний (базисный) период. Такие показатели называют темпами роста. Темп роста может быть выражен в разах или в процентах. Темп роста говорит о том, во сколько раз больше показатель текущего периода в сравнении с базисным или сколько процентов он составляет по отношению к показателю базисного периода. К относительным показателям динамики принадлежат также темпы прироста, параметры уравнений трендов, коэффициенты колеблемости и устойчивости в динамике, индексные показатели динамики.

3. Относительные показатели, характеризующие взаимосвязь признаков в совокупности явлений, а также взаимосвязь результативных признаков-следствий с факторными признаками-причинами, например, количества профилактических мероприятий с количеством совершенных преступлений; уровня квалификации сотрудников органа внутренних дел с эффективностью деятельности органа и т. п. К таким показателям относятся коэффициенты корреляции, эластичности, детерминации, а также аналитические индексы. Относительные показатели взаимосвязи могут быть как отвлеченными, так и именованными числами.

4. Относительные показатели, характеризующие соотношение разных признаков того же объекта между собой (иногда их называют показателями интенсивности). Эти показатели обобщают вторичные признаки объектов. Показатели соотношения признаков могут быть прямыми и обратными. И прямые, и обратные показатели выражаются именованными числами с двойными единицами измерения обоих сравниваемых признаков. К данной группе показателей относится уровень преступности – соотношение количества совершенных преступлений и численности населения.

5. Особым видом относительных статистических показателей являются отношения фактически наблюдаемых величин признака к его нормативным, плановым, оптимальным или максимально возможным величинам. Это широко распространенные показатели выполнения норм выработки, норм расхода материалов и других ресурсов.

Отношения наблюдаемых величин признака к оптимальным или плановым характеризуют приближение изучаемого процесса к идеалу.

Отношение фактических значений признака к максимально возможным значениям часто характеризует качество процесса, агрегата, машины. Таковы, например, показатели целевой эффективности.

1.2. Средние величины как выражение закономерности

Статистика изучает массовые явления и процессы. Каждое из таких явлений обладает как общими для всей совокупности, так и особенными, индивидуальными свойствами. Различия между индивидуальными явлениями называют вариацией. Важно отметить, что массовым явлениям присуще такое свойство, как близость характеристик отдельных явлений. Если в сосуд с горячей водой добавить холодную, то температура воды во всем сосуде станет одинаковой (осреднится). Поведение сотрудников конкретного подразделения полиции тоже по прошествии времени приобретает в определенной степени общие, усредненные черты, после чего правомерно воспринимать каждого из большинства сотрудников как типичного представителя этого подразделения. При сохранении и различий у таких сотрудников появляется все больше общего, причем не только в поведении, но и в образе мыслей, ценностных ориентирах, моральных нормах. Как во всякой нивелировке, в этом есть и свои положительные, и свои отрицательные стороны, но здесь важно то, что это реальная закономерность, с которой всем (и в том числе руководителям) необходимо считаться. Опираясь на знание средних величин по важнейшим характеристикам, показателям, статистика решает много задач.

Итак, взаимодействие элементов совокупности приводит к ограничению вариации хотя бы части их свойств. Эта тенденция существует объективно. Именно в ее объективности заключена причина широчайшего применения средних величин на практике и в теории.

Главное значение средних величин состоит в их обобщающей функции, т. е. замене великого множества различных индивидуальных значений признака некоею единой средней величиной, характеризующей уже всю совокупность явлений. Известно, что руководитель всей совокупностью своих действий (включая и личный пример) способен положительно воздействовать на подчиненных, хотя и в разной мере. Но как измерить это явление, не является ли подобное наше мнение самообманом, тешащим наше самолюбие, но далекое от реальной жизни? По отношению к разным сотрудникам эффект влияния оказывается также различным, более того, в отношении одних сотрудников он будет несущественен, а некоторых – даже отрицателен вопреки всем ожиданиям. Но если, используя имеющиеся показатели деятельности сотрудников, измерить среднюю величину, на которую изменилось поведение подчиненных в процессе указанного воздействия руководителя, то можно установить и факт таких положительных сдвигов в их поведении, а также рассчитать и типичную среднюю величину такого сдвига для определенного времени и, разумеется, для определенных условий (характер коллектива, режим работы, возможности материального стимулирования и т. п.).

Если средняя величина обобщает качественно однородные значения признака, то она является типической характеристикой признака в данной совокупности. Так, можно говорить об измерении типичного уровня выпускников, определенного ведомственного вуза или об уровне этих же выпускников по прошествии, например, трех лет (предполагая, что за это время большинство из них уже в полной мере освоит профессию и сможет проявить ожидаемый положительный эффект).

Однако неправильно сводить роль средних величин только к характеристике типичных значений признаков в однородных по данному признаку совокупностях. На практике значительно чаще современная статистика использует средние величины, обобщающие явно неоднородные явления, как, например, число раскрытий преступлений в расчете на одного сотрудника органов внутренних дел Российской Федерации, включая сотрудников различных подразделений. Или рассмотрим такую среднюю, как средний ущерб от хищения золота в соответствующей добывающей отрасли народного хозяйства на одного в ней работающего: ведь среди работающих в этой отрасли большинство, полагаем, люди, честно выполняющие свои обязанности в трудных условиях. Тем не менее, в определенных аналитических исследованиях знание такой величины весьма необходимо, например, когда оценивается латентный (прямо ненаблюдаемый) ущерб на ранее не изучавшемся объекте той же отрасли в том же или сходном регионе страны.

Средняя величина национального дохода на душу, средняя криминализация региона или страны – это характеристики различных сторон существования страны или субъекта Федерации как единой народнохозяйственной системы, это так называемые *системные средние*.

Системные средние могут характеризовать как пространственные или объектные системы, существующие одномоментно (государство, регион, отрасль, правоохранительное силовое ведомство и т. п.), так и динамические системы, протяженные во времени (год, десятилетие, сезон и т. д.). Примером системной средней, характеризующей период времени, может служить средний по стране конкурс для молодежи, поступающей в высшие (или средние) учебные заведения системы МВД.

Итак, типическая средняя может обобщать системные средние для однородной совокупности, или системная средняя может обобщать типические средние для единой, хотя и неоднородной, системы. При этом даже типическая средняя не является раз и навсегда данной неизменной характеристикой. Так, можно ожидать, что с отменой обязательного призыва в армию конкурс среди поступающих в вузы системы МВД несколько снизится и, соответственно, снизится доля абитуриентов мужского пола.

Понятие средней арифметической

Виды средних величин различаются прежде всего тем, какое свойство, какой параметр исходной варьирующей массы индивидуальных значений признака должен быть сохранен неизменным.

Средней арифметической величиной называется такое среднее значение признака, при вычислении которого общий объем признака в совокупности сохраняется неизменным.

Иначе можно сказать, что средняя арифметическая величина – среднее слагаемое. При ее вычислении общий объем признака мысленно распределяется поровну между всеми единицами совокупности. Например, среднее денежное довольствие сотрудников подразделения – это такая сумма денег, которая приходилась бы на каждого сотрудника, если бы все деньги по статье соответствующих расходов были бы распределены между работниками поровну.

Исходя из определения, формула средней арифметической величины имеет вид:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum x}{n}, \tag{1.1}$$

где $\bar{x}$ – средняя величина; n – численность выборочной совокупности.

По формуле (1.1) вычисляются средние величины первичных (объемных) признаков, если известны индивидуальные значения признака. Если изучаемая совокупность велика, исходная информация чаще представляет собой ряд распределения или группировку, как, например, табл. 1.2.

Таблица 1.2

Распределение полицейских операций УМВД России по N-ской области по числу задержанных лиц, находящихся в розыске (за период с 2010 г. по 2013 г.)

Число задержанных разыскиваемых лиц, x	0	1	2	3	4	5	6	7	8	9	Всего
Число проведенных полицейских операций, n_i	5	8	11	24	30	29	45	41	23	1	187

Среднее число лиц, задерживаемых за одну полицейскую операцию определенного вида, должно представлять результат равномерного распределения общего числа задержанных разыскиваемых лиц по всем 187 операциям данного вида, проведенным в области за учитываемый период времени. Общее число задержанных разыскиваемых лиц согласно исходной информации табл. 1.2 можно получить как сумму произведений значения признака в каждой группе на число операций с таким количеством задержанных разыскиваемых лиц n_i (частоты). Получим следующую формулу (1.2):

$$\bar{x} = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \ldots + x_n n_n}{n_1 + n_2 + n_3 + \ldots + n_n}$$

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i}. \tag{1.2}$$

Такую форму средней арифметической величины называют взвешенной арифметической средней в отличие от простой средней, рассчитанной по формуле (1.1). В качестве весов выступают здесь числа единиц совокупности в разных группах. Название «вес» выражает тот факт, что разные значения признака имеют неодинаковую важность при расчете средней величины. Важнее, весомее число задержанных разыскиваемых лиц, которое встречалось чаще: 1, 2 и 3, а такие значения, как 7 или 9 задержанных разыскиваемых лиц при расчете средней, ввиду своей нетипичности, не играют особой роли – их вес мал.

Средняя арифметическая величина может быть дробным числом, если даже индивидуальные (первичные) значения целые. Первичный признак (число задержанных в ходе конкретной операции) может принимать, естественно, только целые значения (такой признак называется дискретным, т. е. прерывистым, в отличие от непрерывного, каким уже является получаемая на его основе средняя арифметическая величина). Ничего негативного для метода средних в этом нет; из сущности средней вовсе не вытекает, что эта величина обязана быть реальным значением признака, которое могло бы встретиться у какой-либо конкретной единицы совокупности (в данном примере единица – конкретная операция по задержанию). Но средняя может, конечно, иногда принимать и целые значения, которые, в свою очередь, могут совпадать со значениями отдельных конкретных операций, но могут и не совпадать ни с одним из таких первичных значений. Например, если бы ровно в половине случаев было по одному задержанию, а в оставшейся половине – тоже по одинаковому числу задержания, допустим, по три, тогда очевидно, что средняя оказывается равной двум, при том, что ни в одной из учитывавшихся конкретных операций по задержанию такого количества задержанных не было. Итак, хотя конкретная величина (число задержанных в ходе операции) всегда целое число, его средняя, выражающая уже абстракцию (т. е. отвлеченное значение) уже не обязана быть целой, а может быть и дробной, как в табличном примере.

В случае, когда имеем дело с группировкой данных и значения осредняемого признака заданы интервалами, при расчете средней арифметической величины в качестве значения признака в группах принимают середины этих интервалов. Делается так, исходя из наиболее вероятного (в этих условиях неполной нашей информированности) предположения, гипотезы о равномерном распределении единиц совокупности по интервалу значений признака – тогда середина интервала – их естественный компромисс. Для открытых интервалов в первой и последней из групп, если таковые есть, значение признака надо определить экспертным путем исходя из сущности, свойств признака и совокупности. Например, по табл. 1.3 можно минимальный возраст сотрудников ОВД (включая и курсантов, и слушателей) считать 17 лет. Тогда первый интервал будет от 17 до 20 лет, а максимальный возраст – 62 года, тогда последний интервал – от 50 до 62 лет.

Таблица 1.3

Распределение сотрудников УМВД России по N-ской области по возрасту

Возрастная группа сотрудников (интервал), лет	Число сотрудников такого возраста, n_i	Середина возрастного интервала, x_i (лет)	Общее количество лет сотрудников, считая, что все они в возрасте середины интервала, $x_i n_i$
До 20	48	18,5	888
20–30	120	25	3000
30–40	75	35	2625
40–50	62	45	2790
Старше 50	54	57,5	3105
Итого	359	34,56	12408

Средний возраст сотрудников, рассчитанный по формуле (1.2) с заменой точных значений признака в группах серединами интервалов, составил:

$$\overline{x} = \frac{\sum x_i n_i}{\sum n_i} = \frac{12408}{359} = 34{,}56 \text{ года,}$$

что и записано в итоговую строку. В то время, как итог абсолютного («объемного» показателя) – это сумма, итог же по графе относительных показателей или средних групповых величин – это средняя.

Числитель дроби – это общая сумма лет, прожитых сотрудниками УМВД. Эта рассчитанная сумма тем ближе к истинной, чем справедливее сделанные нами предположения относительно возможности замены истинного возраста средним значением соответствующего возрастного интервала, а также относительно назначенных нами нижней границы для первой и верхней границы для последней из групп. Разделив указанную сумму на число работников, получаем средний возраст сотрудника УМВД в годах.

Перейдем к рассмотрению средних вторичных (относительных) признаков. Сумма таких показателей сама по себе реальной величиной какого-либо признака в совокупности не является. Однако общее определение арифметической средней сохраняет силу и в этом случае. При вычислении таких средних величин необходимо, чтобы сохранялась сумма величины объемного признака, который является числителем при построении осредняемого относительного показателя.

Так, например, необходимо по N-ской области вычислить уровень преступности (на 100 тыс. чел. населения) исходя из значений уровня преступности по каждому из районов области. Этот показатель, иначе именуемый коэффициентом пораженности населения преступностью, представляющий типичную среднюю величину, должен быть таким, чтобы общее количество выявленных преступлений осталось бы неизменным при замене значений аналогичных коэффициентов по каждому из районов этим единым для всей области коэффициентом, т. е. средней по области величиной. Нельзя менять реальную величину объемного признака (в примере – это общее число выявленных преступлений) – она является базой расчета средней. Чтобы выполнить указанное условие, в качестве весов при расчете средней величины относительного показателя необходимо принять значения того признака, который является знаменателем при определении относительного показателя. Так, при вычислении уровня преступности (на 100 тыс. чел. населения) по совокупности районов весами должны служить численность населения каждого из районов.

Применение простой и взвешенной средней

Простая и взвешенная средние величины различаются не только по величине (не всегда), способу вычисления, но и по своей роли в решении различных задач статистического анализа. Рассмотрим, например, такую среднюю величину, как уровень преступности (на 100 тыс. чел. населения) для области. Если эта средняя при решении поставленной задачи входит в систему показателей численности населения, количества выявляемых преступлений, затрат ресурсов (сил и средств) и других характеристик деятельности территориального органа внутренних дел, то следует применять взвешенную среднюю, т. к. произведение невзвешенной средней на общую численность населения не даст общего количества выявленных преступлений.

Если же нас интересуют такие задачи, как измерение вариации (степени различий) указанного выше уровня преступности между хозяйствами или связь этого уровня с выделением определенного вида сил и средств в известных количествах, то следует применять простую среднюю – уровень преступности по каждому из районов, полностью абстрагируясь от численности населения районов. Иначе на полученный результат повлияют различия таких численностей, совершенно не касающиеся данного признака. Точно также, если необходимо изучить колебания раскрываемости преступлений определенного вида за несколько лет и выявить их связь с экономической ситуацией, единой для всех районов области, нужно применять простую среднюю – раскрываемость преступлений данного вида за несколько лет, абстрагируясь от различия в количестве выявленных преступлений в разные годы.

Чтобы правильно применять средние величины, следует знать, от каких причин зависит различие между простой и взвешенной средними.

Понятие средней в широком смысле слова сближается с такой философской категорией, как закон («закон есть общее в явлениях»), закономерность. Это далеко не случайное родство. Рассмотрим сущность процесса осреднения на примере арифметической средней согласно формуле (1.1). Среднюю считаем типической, определенной по однородной совокупности. Однородность индивидуальных значений признака – это проявление их общих свойств, обусловленных основными условиями и закономерностями массового процесса, порождающего данную совокупность. Однако кроме общих условий, кроме закономерности, на каждую единицу совокупности влияют индивидуальные, особенные условия, случайные события, не связанные причинно с общей закономерностью. Поэтому можно индивидуальные значения признака x_i представить как состоящие из единиц совокупности (обозначим этот элемент c), так и элемента Δ_i индивидуального для каждой единицы совокупности.

Итак, $x_i = c + \Delta_i$, где Δ_i может быть как положительной, так и отрицательной величиной, как малой, так и большой в сравнении с c.

Теперь вычислим среднее значение признака для совокупности из n единиц:

$$\begin{gathered}\bar{x} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n} = \frac{\sum x}{n} = \\ \frac{\sum(c + \Delta_i)}{n} = \frac{c + \sum \Delta_i}{n} = c + \Delta.\end{gathered} \tag{1.3}$$

Итак, **средняя величина** признака слагается из элемента, выражающего ***закономерность***, общую для всей совокупности, и из средней величины элементов, отражающих ***индивидуальные условия*** отдельных единиц этой совокупности. Элементы Δ_i могут иметь положительные и отрицательные, большие и малые значения. При осреднении они согласно закону больших чисел взаимопогашаются в зависимости от объема совокупности: тем в большей мере, чем больше объем совокупности n. Об этом говорит «закон больших чисел», данный великим русским математиком П. Л. Чебышевым (1821–1894). Чем больше объем однородной совокупности, тем полнее взаимопогашение случайных (по отношению к совокупности в целом и ее законам) элементов признака x; полнее и надежнее, с большей вероятностью среднее значение признака измеряет действие общих для совокупности закономерностей.

Однако случайная вариация индивидуальных величин признаков – это не только некоторая помеха, туман, «шум» в информационном смысле, затрудняющий познание закономерности. Вариация – неотъемлемая, необходимая черта, свойство массовых явлений, имеющая громадное самостоятельное значение в развитии природы и общества.

Создатель учения о средних величинах бельгийский статистик А. Кетле писал следующее: «В мире существует общий закон, предназначенный как бы для того, чтобы разливать жизнь во Вселенной; в силу этого закона все живущее подлежит бесконечному разнообразию. Каждый предмет подвержен флюктуациям»[1].

[1] Кетле А. Социальная система и законы ею управляющие / пер. Л. Н. Шаховской. СПб., 1866. С. 16.

2. РЯДЫ ДИНАМИКИ ПРИ ИЗУЧЕНИИ ПРАВОВЫХ ПРОЦЕССОВ. РОЛЬ ЭКСТРАПОЛЯЦИИ В КРИМИНОЛОГИЧЕСКОМ ПРОГНОЗИРОВАНИИ

2.1. Понятие и классификация динамических рядов

Одно из основных положений методологии научного познания – необходимость изучать все явления в развитии. Изучение протекания процессов (динамики явлений), изменений явлений во времени, осуществляемое с использованием статистических средств, позволяет исследователю получить ответы на многие важные вопросы. Например, как по годам изменяется в количественном отношении преступность в регионе, каковы тенденции этих изменений, как возрастает (или снижается) уровень нагрузки на сотрудников полиции, велики ли колебания уровня преступности, каковы ее сезонные изменения.

На эти и аналогичные вопросы исчерпывающие ответы может дать только специализированная система статистических методов, каждый из которых предназначен для изучения развития, изменения во времени (или, как принято говорить, для изучения динамики) тех или иных параметров меняющегося явления. Освоение этих методов, достаточно понятных даже при ограниченной математической подготовке аналитика, имеет большое значение для становления и профессиональной деятельности, как научной, так и практической, юристов различных специализаций.

Для отображения динамики строят так называемые ряды (хронологические, временные, динамические), которые представляют собой последовательности меняющихся во времени и расположенных в хронологическом порядке значений какого-то статистического показателя, характеризующего рассматриваемое явление с определенной стороны через признак этого явления. Примером явления могут служить квартирные кражи на определенной территории, а примером признака – их количество за соответствующий год.

Составными элементами динамических (временных) рядов являются количественные значения соответствующего показателя и периоды (моменты) времени, к которым они относятся.

Важнейшая статистическая характеристика ряда динамики – это его уровень члена ряда. Под уровнем (членом ряда) понимается показатель – размер (объем, величина), отражающий соответствующую сторону изучаемого явления, относящегося к определенному моменту времени или к целому временному периоду. Ввиду важности для последующих алгоритмов анализа данных из общего числа уровней выделяют *начальный* и *конечный* уровни динамического ряда, соответствующие крайним членам такого ряда.

Важная характеристика динамического ряда – его длина, измеряемая численностью членов ряда, что далеко не всегда соответствует длине рассматриваемого временного диапазона, поскольку даже на малом диапазоне может быть много отдельных наблюдений, каждому из которых соответствует член ряда, в то время как на обширном диапазоне их будет совсем немного. Далее длительность ряда будет рассматриваться именно как число членов ряда с соответствующими уровнями.

С точки зрения качества результатов статистического анализа длинные динамические ряды предпочтительнее коротких в том смысле, что чем длиннее ряд, тем, как правило, более надежны результаты. Однако с социально-правовой, криминологической точки зрения, использование очень длинных динамических рядов будет оправданным только при отсутствии резких социально-экономических, политических, правовых изменений. В ином же случае может создаться ситуация, когда начальные и конечные наблюдения будут принадлежать к качественно различным социально-правовым явлениям, лишь формально объединяемым общим названием. Необходим, следовательно, разумный компромисс между статистическими и криминологическими требованиями к длине динамического ряда, определяемый целями исследования.

Поскольку временной ряд включает два обязательных элемента: время и соответствующее ему конкретное значение уровня ряда, то в табличной форме динамический ряд задается с помощью двух столбцов или строк (табл. 2.1).

В качественном отношении временные ряды различаются по следующим признакам.

1. *По времени* – моментные и интервальные. Если уровень ряда характеризует изучаемое явление в конкретный момент времени, то совокупность уровней ряда образует *моментный ряд*. Примером

моментного ряда может быть последовательность количества преступников, находящихся в розыске, на 1-е число месяца. Ясно, что в предыдущий день, равно как и в последующий, их может быть больше или меньше, что никак не связано с их численностью именно в этот учетный день месяца. В отличие от этого, *интервальный ряд* – последовательность членов ряда, в которой уровень изучаемого признака явления характеризует накопление за весь протекший интервал времени. Примером интервального ряда может служить ряд, уровнями которого выступают количества преступлений, совершенных в регионе за месяц.

Таблица 2.1

Количество убийств и покушений на убийство, совершенных в Уральском федеральном округе (УФО) с 2012 г. по 2016 г.[1]

Год	2012	2013	2014	2015	2016
Количество убийств и покушений на убийство в УФО	1805	1629	1469	1358	1299

Вот почему интервальный ряд отличается от моментного и тем, что сумма уровней интервального ряда дает реальный показатель, наделенный определенным смыслом и содержанием, в то время как сумма уровней моментного ряда реального содержания не имеет. Так, сумма двенадцати показателей количества преступлений, совершенных на определенной территории за месяц, дает годовой аналогичный показатель. С другой стороны, сумма двенадцати показателей количества разыскиваемых преступников никакой смысловой интерпретацией не обладает.

2. *По форме представления уровней* – ряды абсолютных (табл. 2.1), относительных (табл. 2.2) и средних величин (табл. 2.3).

Таблица 2.2

Удельный вес тяжких и особо тяжких преступлений, совершенных в Центральном федеральном округе (ЦФО) с 2012 г. по 2016 г.[2]

Год	2012	2013	2014	2015	2016
Удельный вес тяжких и особо тяжких преступлений, %	28,5	28,7	27,8	27,5	27,5

[1] По данным ФКУ «ГИАЦ МВД России».

[2] Справочные материалы ФКУ «ГИАЦ МВД России».

Таблица 2.3

Средний стаж службы сотрудников УМВД России по N-ской области с 2012 г. по 2016 г.[1]

Год	2012	2013	2014	2015	2016
Средний стаж службы сотрудников, лет	15,4	14,8	13,9	14,1	14,5

3. *По расстоянию между датами или интервалами времени* – полные или неполные временные ряды. *Полные* ряды имеют место, когда даты регистрации показателей (для моментных рядов) или окончания периодов следуют друг за другом с равными интервалами; для *неполных* рядов принцип равных интервалов не соблюдается.

4. *По содержанию показателей* – ряды частных и агрегированных показателей. *Частные* показатели характеризуют изучаемое явление односторонне, изолированно от прочих признаков. *Агрегированные* показатели основаны на увязанных между собой частных показателях и потому характеризуют изучаемый процесс комплексно. Пример временного ряда частных показателей – ряд уровней, отражающих меру доверия населения в отношении деятельности органов внутренних дел (по данным ежегодных опросов общественного мнения). Другой пример ряда агрегированных показателей – динамика интегрального индекса безопасности личности от противоправных посягательств, также объединяющая разные признаки.

Важнейшее условие правильного построения и исследования рядов динамики – обеспечение сопоставимости уровней этих рядов, относящихся к различным периодам, представленным в ряду. Сопоставимость показателей – это соответствие условий и методов расчета ее показателей, обеспечивающих правильность получаемых при их сравнении выводов о различиях между изучаемыми явлениями (например, преступностью). Данное условие решается либо в процессе сбора и обработки данных, либо путем их пересчета.

Уровни ряда, подлежащие изучению, должны быть однородны по содержанию, учитывать существо изучаемого явления, а также цель исследования. При построении временного ряда, во-первых, необходимо соблюдать требование сопоставимости показателей ряда во времени и пространстве (по территории). Вполне понятно, что показатели, характеризующие, например, данные о числе осужденных за месяц и за год, непосредственно не сопоставимы друг

[1] Пример условный.

с другом ввиду неравенства интервалов или несовпадения моментов уровней. То же самое происходит и в случае изменения административно-территориального деления – образование новых районов, областей и, наоборот, их слияния. В данном случае при построении динамических рядов необходимо устранять несопоставимость данных по разным территориям путем пересчета этих данных за прежнее время применительно к новой территории.

Иногда для того чтобы привести уровни ряда динамики к сопоставимому виду, приходится прибегать к приему, который называется «смыкание рядов динамики». Под смыканием понимают объединение в один ряд (более длинный) двух или нескольких рядов динамики, уровни которых исчислены по разной методологии или разным территориальным границам. Так, при изменении границ региона (территориальные преобразования) сопоставимые данные ряда динамики можно получить, применяя смыкание рядов, при котором показатели за год (или квартал), в котором произошло изменение границ, принимаются за 100 %, и, исходя уже из этого, рассчитываются все другие показатели.

Необходимо обеспечить также полноту охвата различных частей явления, которое также важнейшее условие сопоставимости уровней ряда. Требование одинаковой полноты охвата разных частей изучаемого объекта означает, что уровни ряда за отдельные периоды должны характеризовать размер того или иного явления по одному и тому же кругу входящих в его состав частей.

Конечно, необходимо иметь в виду, что из-за существования латентной (скрытой от регистрации) преступности динамические ряды и отражают лишь число зарегистрированных, а не число реально совершенных (фактических) преступлений. Еще менее точны динамические ряды, отражающие информацию о преступниках. В этих рядах отсутствуют не только сведения о лицах, совершивших преступления, оставшихся латентными, но и сведения о лицах, совершивших преступления, которые были зарегистрированы, но не были раскрыты.

Формируя динамические ряды, необходимо следить за однокачественностью их уровней на протяжении всего временного периода. Вследствие многих обстоятельств однородность величин, составляющих динамический ряд, может нарушиться, и таким образом нарушается сопоставимость уровней динамического ряда. Если, например, анализируется уровень преступности, то необходимо учесть изменения уголовного законодательства и взять данные о тех видах преступлений, которые предусматривались уголовным законом на протяжении всего исследуемого периода. В противном случае будут иметь место дефектные ряды, и их показатели несопоставимы друг с другом по своему содержанию. Особенно такая дефектность заметна на длинных динамических рядах.

На сопоставимость уровней ряда динамики непосредственно влияет методология учета или расчета показателей. Особенно часто эта проблема возникает при международных сопоставлениях. В некоторых странах, например, раскрытым считается преступление только на основании решения, принятого судом. Такая статистическая практика значительно отличается от российских подходов.

Несопоставимость показателей, возникающая в силу неодинаковости применения единиц измерения, очевидна. Например, при анализе уровней преступности могут сравниваться показатели, рассчитанные лишь на одинаковое количество населения: на 1 тыс., на 10 тыс. или на 100 тыс. населения. Это и все вышеперечисленные обстоятельства следует учитывать при подготовке информации для анализа изменений явлений во времени (динамики явлений).

2.2. Показатели, характеризующие тенденцию динамики. Особенности вычисления среднего уровня динамического для моментных рядов

При статистическом изучении динамики необходимо четко разделить ее два основных элемента – тенденцию и колеблемость, чтобы дать каждому из них количественную характеристику с помощью специальных показателей.

Тенденция и колебания наглядно показаны на рис. 2.1. По оси абсцисс всегда отражается время, по оси ординат – уровни. По обеим осям строго соблюдается масштаб, иначе характер динамики будет искажен.

Сравнивая уровни разных лет, можно заметить, что в целом количество дорожно-транспортных происшествий возрастает. Однако нередко уровень рассматриваемого временного ряда следующего года оказывается ниже предыдущего. Иногда рост по сравнению с предыдущим годом велик, как в 2007 г., а иногда мал. Следовательно, рост дорожно-транспортных происшествий наблюдается лишь в среднем как тенденция. В отдельные же годы уровни временного ряда испытывают колебания, отклоняясь от основной тенденции.

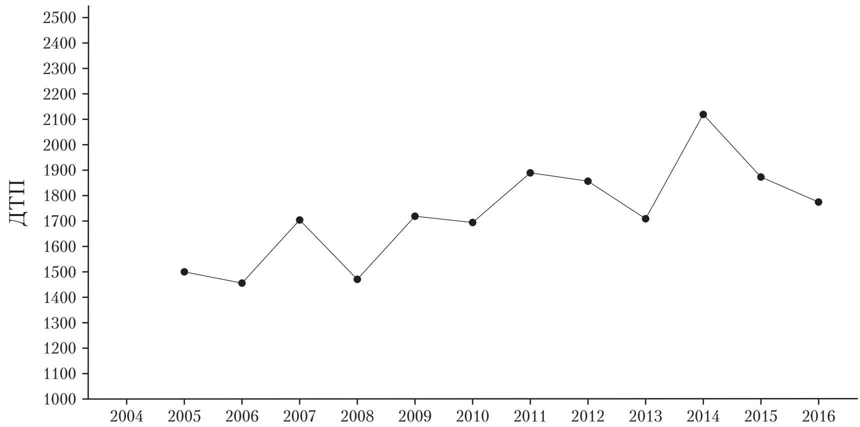


Рис. 2.1. Количество дорожно-транспортных происшествий, зарегистрированных в Петровской области по годам[1]

Если рассматривать динамические ряды показателей преступности по месяцам, выявятся регулярно повторяющиеся из года в год сезонные колебания уровней. Сезонные «пики» преступности в целом (отдельные виды преступных деяний могут иметь свои, не совпадающие с другими видами «пики») чаще всего падают на весну и осень, а точнее, на март и октябрь, а «провалы» регистрируются зимой (декабрь – январь), т. е. в конце и в начале отчетного периода. В летние месяцы (июль) уровни преступности ниже, чем весной и осенью.

Тенденция динамики связана с действием долговременно существующих причин и условий развития, хотя, конечно, после какого-то периода эти причины и условия тоже могут измениться и породить уже другую тенденцию развития изучаемого объекта. Колебания же, напротив, связаны с действием краткосрочных или циклических факторов, влияющих на отдельные уровни динамического ряда и отклоняющих уровни от тенденции то в одном, то в другом направлении. Например, тенденция динамики дорожно-транспортных происшествий связана с увеличением количества автолюбителей. Колеблемость количества преступлений, по мнению некоторых криминологов, обусловлена тем, что уровень преступной деятельности коррелирует с регистрационной активностью правоохранительной системы, на которую определенное влияние оказывает сезонность.

[1] Пример условный.

Таким образом, при статистическом изучении динамики исследователю необходимо в первую очередь выделить во временном ряду два основных ее элемента – тенденцию и колеблемость, определяющих закономерное проявление изменений анализируемого процесса. Смешение тенденции и колеблемости ведет к неверным выводам о динамике. Если из временного ряда дорожно-транспортных происшествий (табл. 2.4) произвольно взять данные за отдельные годы и сравнить их друг с другом, можно получить выводы, прямо противоположные истине. Например, если сравнить уровни ряда в 2015 г. и в 2007 г., то получим, что за 8 лет количество дорожно-транспортных происшествий возросло на 660 происшествий, т. е. более чем на 80 происшествий за год. Если же сравнить уровни ряда, относящиеся к 2016 г. и к 2008 г., то получим, что за 8 лет количество дорожно-транспортных происшествий возросло лишь на 26 происшествий.

Таблица 2.4

Количество дорожно-транспортных происшествий, зарегистрированных в N-ской области с 2006 г. по 2016 г.[1]

Год	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Кол-во ДТП	1491	1453	1686	1465	1772	1764	1908	1869	1768	2113	1895

Чтобы построить систему показателей, характеризующих тенденцию динамики, нужно ответить на вопрос: какие черты, свойства этой тенденции нужно измерить и выразить в статистических показателях? Очевидно, аналитика должна интересовать величина изменений уровня как в абсолютном, так и в относительном выражении (на какую соответственно долю или процент уровня, принятого за базу, произошло изменение). Далее аналитика будет интересовать, является ли изменение равномерным или неравномерным, ускоренным (замедленным). Наконец, интерес представит и выражение тенденции в форме некоторого уравнения, наилучшим образом аппроксимирующего (приближенно отображающего) фактическую тенденцию динамики. Уравнение тенденции динамики в статистике называется уравнением тренда.

Для того чтобы нагляднее представить показатели, характеризующие тенденцию, возьмем условный пример, в котором динами-

[1] Пример условный.

ческий ряд представлен в виде «чистого» тренда и колебания отсутствуют. Пример такого ряда представлен в табл. 2.5, где единица измерения данных – «преступление на 100 тыс. человек».

Таблица 2.5

Абсолютные и относительные показатели тенденции

Номера периодов или моментов времени	Уровни ряда, прест. / 100 тыс.	Абсолютный прирост уровней, прест. / 100 тыс. в год	Темп роста уровня, в % к предыдущему	Темп роста уровня, в % к начальному
0	100	–	–	100
1	112	12	112	112
2	128	16	114,3	128
3	148	20	115,6	148
4	172	24	116,2	172
5	200	28	116,3	200
6	232	32	116,0	232

Абсолютный прирост – это разность между сравниваемым уровнем и уровнем более раннего периода, принятым за базу сравнения. Если эта база – непосредственно предыдущий уровень, то показатель называют цепным; если за базу взят, например, начальный уровень, показатель называют базисным. Формулы абсолютного изменения уровня:

базисное: $\Delta_i = y_i - y_0$, (2.1)

цепное (с переменной базой): $\Delta_i = y_i - y_{i-1}$. (2.2)

Абсолютный прирост имеет ту же единицу измерения, что и уровни ряда с добавлением единицы времени, за которую определено изменение, а именно: 28 преступлений на 100 тыс. человек в год.

Лучше усвоить рассмотренные показатели поможет следующая аналогия с механическим движением: уровень – это аналог пройденного пути, причем начало его отсчета не в нулевой точке. Абсолютный прирост – аналог скорости движения тела, а ускорение абсолютного прироста – аналог ускорения движения в физике. Пройденный путь, считая и тот, который уже был пройден до начала отсчета времени (и представленный поэтому первым уровнем) в данной задаче, равен:

$$S = s_0 + v_0 t + \frac{at^2}{2}, \quad (2.3)$$

где s_0 – путь, пройденный до начала отсчета времени; v_0– начальная скорость; a – ускорение; t – время, прошедшее от начала его отсчета в задаче.

Видим, что s_0 – аналог свободного члена a; v_0 – аналог абсолютного прироста b; $a/2$ – аналог ускорения прироста c.

Система частных показателей, характеризующих явление, должна содержать не только абсолютные, но и относительные статистические показатели. Так, *относительные* показатели динамики необходимы для сравнения развития разных объектов из сферы информатики, особенно если их абсолютные (объемные) характеристики различны. Предположим, имеется временной ряд за аналогичный период с тенденцией, выраженной уравнением тренда:

$$y_t = 20 + 4t + 0{,}5t^2.$$

Коэффициент роста (или *темп роста* – этот же показатель, но выраженный в процентах) – это отношение сравниваемого уровня (более позднего) к уровню, принятому за базу сравнения (более раннему). Коэффициент роста (темп роста) исчисляется в цепном варианте – к уровню предыдущего года и в базисном варианте – к одному и тому же, обычно начальному уровню. Он говорит о том, сколько процентов составляет сравниваемый уровень по отношению к уровню, принятому за базу, или то, во сколько раз сравниваемый уровень больше уровня, принятого за базу. При этом если уровни снижаются со временем, то сказать, что последующий уровень больше в 0,33 раза или составляет 33,3 % базового уровня, это, разумеется, означает, что уровень уменьшился в 3 раза.

Таблица 2.6

Абсолютные и относительные показатели тенденции

Номера периодов или моментов времени	Уровни ряда, прест. / 100 тыс.	Абсолютное изменение уровней, прест. / 100 тыс. в год	Темп роста уровня, в % к предыдущему	Темп роста уровня, в % к начальному
0	20	–	–	100
1	24,5	4,5	122	122
2	30	5,5	122	150
3	36,5	6,5	122	182
4	44	7,5	120	220
5	52,5	8,5	119	262
6	62	9,5	118	310

Можно сказать, что относительная характеристика роста показателей для временного ряда в предыдущей табл. 2.5 за год близка к 115 % (рост приблизительно на 15 % за год) и за шесть лет уровни ряда увеличились в 2,32 раза; а для временного ряда (табл. 2.6), вычислив шесть уровней параболического тренда, можно убедиться, что в среднем за год уровень ряда возрастал примерно на 20 %, а за шесть лет фактический рост составил 310 %. Следовательно, в относительном выражении второй ряд возрастал быстрее. Таким образом, только в сочетании абсолютных и относительных характеристик динамики можно правильно отразить процесс развития всей совокупности объектов.

Рассмотрим связь абсолютных и относительных показателей динамики. Коэффициент роста обозначается $k_{рост}$ (или иногда, когда понятно, о чем речь идет, просто k) и измеряется в отвлеченных (безразмерных) числах, которые могут быть и больше единицы (реальный рост), и меньше ее (реальное снижение), и равным ей (стабильность). В отличие от этого темп роста, который обозначается $T_{рост}$, измеряется в процентах (что также является безразмерной величиной, т. к. «процент» – это просто «одна сотая», а «сто сотых» ничто иное, как единица, не имеющая размерности):

базисный темп роста:

$$k_{i/0} = \frac{y_i}{y_0},$$

цепной темп роста:

$$k_{i/i-1} = \frac{y_i}{y_{i-1}}.$$

Если сравниваемый уровень Y выразить через уровень предыдущего года плюс прирост или через уровень базисного года плюс базисное абсолютное изменение, получим:

$$k_{i/0} = \frac{y_0 + \Delta_{0i}}{y_0} = 1 + \frac{\Delta_{0i}}{y_0} \text{ или } T_{рост} = 100\% + \frac{\Delta_{0i}}{y_0},$$

$$k_{i/i-1} = \frac{y_{i-1} + \Delta_i}{y_{i-1}} = 1 + \frac{\Delta_i}{y_{i-1}} \text{ или } T_{рост} = 100\% + \frac{\Delta_i}{y_{i-1}}.$$

Величину $\frac{\Delta_i}{y_{i-1}}$ или $\frac{\Delta_{0i}}{y_0}$, т. е. отношение абсолютного изменения к предыдущему или базисному уровню, часто называют *относитель-*

ным приростом (относительным изменением) или же темпом прироста. Он равен (k–1). Темп прироста (относительное изменение) может иметь как положительные значения, так и отрицательные.

Рассмотрим соотношения между цепными и базисными показателями.

1. Сумма цепных абсолютных изменений равна базисному абсолютному изменению:

$$\sum \Delta_{i(\text{цепн.})} = \Delta_{i(\text{базис.})}.$$

2. Произведение цепных темпов изменения равно базисному темпу изменения:

$$\prod_i k_{i(\text{цепн.})} = k_{i(\text{базис.})},$$

где $\prod$ – знак произведения (перемножения).

Сумма цепных темпов прироста не равна базисному темпу прироста. Значения цепных темпов прироста, рассчитанных каждый к своей базе, различаются не только числом процентов, но и величиной абсолютного прироста, составляющей каждый процент. Поэтому складывать или вычитать цепные темпы прироста недопустимо.

При анализе временных рядов важная группа показателей, характеризующих динамику, – средние значения в динамическом ряду («средние показатели»): средний уровень ряда, средние абсолютные приросты и ускорения, средние коэффициенты роста (темпы роста). Они необходимы при обобщении характеристик тенденции за длительный период, по различным периодам и при сравнении развития за неодинаковые по длительности отрезки времени при выборе аналитического выражения тренда. При наличии в динамическом ряду существенных колебаний уровней определение средних показателей тенденции требует применения специальных методов статистики, а здесь рассматривается только форма, математические свойства средних показателей динамики и простейшие приемы их вычисления, применимые на практике к рядам со слабой колеблемостью.

Средний уровень интервального ряда динамики определяется как простая арифметическая средняя из уровней ряда за равные промежутки времени:

$$\bar{Y} = \frac{\sum_1^n y_1}{n},$$

или как взвешенная арифметическая средняя из уровней за неравные промежутки времени, длительность которых и является весами (весовыми коэффициентами).

В моментном ряду смысл среднего уровня в том, что он характеризует уже не состояние объекта в отдельные моменты, а его среднее, *обобщенное* состояние между начальным и конечным моментом учета. Из этого следует, что роль уровней, относящихся к начальному и конечному моменту, существенно иная, отличная от уровней для моментов внутри изучаемого отрезка времени. Начальный и конечный уровни находятся на границе изучаемого интервала, они (хоть и представлены точками) наполовину относятся один к предыдущему, а другой – к последующему интервалам, и лишь наполовину к изучаемому. Уровни, относящиеся к моментам внутри интересующего осредняемого интервала, целиком относятся только к нему. Отсюда получаем особую форму средней арифметической величины, называемой *хронологической средней*:

$$\bar{Y}_{хронол.} = \frac{\left[\frac{y_1}{2} + \sum_2^{n-1} y_1 + \frac{y_n}{2}\right]}{n-1}.$$

Если известны точные даты изменения уровней моментного ряда, то средний уровень определяется как:

$$\bar{Y} = \frac{\sum t_i y_i}{\sum t_i},$$

где t_i – время, в течение которого сохранялся уровень.

Средний абсолютный прирост (абсолютное изменение) определяется как простая арифметическая средняя из абсолютных изменений за равные промежутки времени (цепных абсолютных изменений) или как частное от деления базисного абсолютного изменения на число осредняемых отрезков времени от базисного до сравниваемого периода:

$$\bar{\Delta} = \frac{\sum \Delta_i}{n} = \frac{y_n - y_0}{n}.$$

Как уже было отмечено, при наличии существенной колеблемости уровней средний абсолютный прирост (изменение), как и средний темп, следует вычислять, отделив сначала тренд от колебаний (соответствующая методика будет изложена ниже). Прямой непосредственный расчет среднего абсолютного прироста по крайним (начальному и конечному) уровням ряда допустим, если нет существенных колебаний уровней.

Для правильной интерпретации показатель среднего абсолютного изменения должен сопровождаться указанием двух единиц времени:

– время, за которое он вычислен, к которому относится и которое он характеризует;

– время, на которое показатель рассчитан, время, входящее в его единицу измерения.

Можно рассчитать среднемесячный прирост за пятилетие, среднесуточное изменение за год, месяц, квартал. Очевидно, что среднемесячный прирост будет в 12 раз меньше среднегодового прироста за рассматриваемый период.

Средний коэффициент роста (как и темп роста) определяется наиболее точно при аналитическом выравнивании динамического ряда по экспоненте (показательной функции). Если можно пренебречь колеблемостью, то средний темп определяют как *геометрическую среднюю* из цепных темпов роста за n лет или из общего (базисного) темпа роста за n лет:

$$\bar{k} = \sqrt[n]{\prod_{1}^{n} k_i} = \sqrt[n]{\frac{y_n}{y_0}}.$$

Средний коэффициент роста (как и темп роста), так же как средний прирост, следует сопровождать указанием двух единиц времени:

– периода, который им характеризуется;

– периода, на который рассчитан темп.

Например, среднегодовой темп за последнее десятилетие; среднемесячный темп за полугодие и т. п.

2.3. Прогнозирование коротких временных рядов в форме экстраполяции

В качестве конкретного численного примера исследуем динамику показателя «Количество совершенных преступлений в Каменском районе» (табл. 2.7). Для этого района области требуется вычислить показатели динамики совершаемых преступлений по годам и указать соответствующие средние за весь анализируемый период.

Таблица 2.7

Количество преступлений, совершенных в Каменском районе N-ской области, по годам

Наименование показателя	Год, i				
	0-й	1-й	2-й	3-й	4-й
Количество преступлений, Y_i	31	25	24	28	34

Рассматриваемый здесь временной (динамический) ряд является рядом с равными промежутками (каждый из них равен одному году). Кроме того, этот конкретный ряд является интервальным, т. е. его значения относятся ко всему интервалу, служат итогом всего прошедшего года. В это значение включены и соответствующие преступления, относящиеся к 1 января, если таковые имели место в районе, и, скажем, к 31 мая, и к 31 декабря того же года и т. д. В течение года значение формирующегося показателя способно только возрастать или, в лучшем случае (учитывая, что речь идет о совершении преступлений) оставаться на том же уровне, но никак не снижаться. Здесь имеет место только эффект накопительный, кумулятивный. Иная картина для моментного временного ряда, примером которого служат ряды, образуемые такими показателями, как «Количество лиц, находящихся в розыске», «Количество лиц, находящихся в изоляторе временного содержания» или «Численность личного состава», «Количество сотрудников, временно отсутствующих по уважительным причинам» и т. п. Ясно, что значения по каждому из этих показателей в каждый момент могут меняться и в большую, и в меньшую сторону. Так, для первого показателя: кого-то из разыскиваемых задержали – показатель убыл, кто-то вновь объявлен в розыск – возрос (аналогично и для второго показателя: кто-то вновь принят – показатель возрос, кто-то уволен – убыл). Вид временного ряда – интервальный он или моментный – надо обязательно учитывать, поскольку формулы для расчета средних показателей динамики для них различны, соответственно и разные получаются на их основе результаты расчетов.

Каждое из значений временного ряда (и интервального, и моментного) носит наименование «уровень ряда» и имеет стандартное обозначение «y» с указанием внизу индекса – номера уровня ряда. В данном случае это порядковый номер года: 0, 1, 2 и т. д., но можно и 2000, 2011, 2012 и т. д. – сути дела при анализе динамики это не меняет, нумеруют как удобнее и проще для восприятия, а это чаще бывает именно 1, 2 и т. д. Здесь уровни y_1, y_2, y_3, y_4, y_5 – по числу учитываемых лет. Число лет обозначается символом n.

Таблица 2.8

Расчет показателей динамики

Статистический показатель	Тип	0-й год	1-й год	2-й год	3-й год	4-й год
Абсолютный прирост, Δ_i, Кол-во преступлений: $\Delta_{пос.i} = y_i - y_0$ $\Delta_{пер.i} = y_i - y_{i-1}$	с пер. базой	–	$\Delta_{пер.1} = 25 - 31 = -6$	$\Delta_{пос.2} = 24 - 31 = -7$	$\Delta_{пос.3} = 28 - 31 = -3$	$\Delta_{пос.4} = 34 - 31 = +3$
	с пост. базой	$\Delta_{пос.0} = 31 - 31 = 0$	$\Delta_{пер.2} = 24 - 25 = -1$	$\Delta_{пер.3} = 28 - 24 = +4$	$\Delta_{пер.4} = 34 - 28 = +6$	$\Delta_{пос.1} = 25 - 31 = -6$
Коэффициент роста, K_i (безразмерная величина, в «разах») $K_{пос.i} = \frac{y_1}{y_0}$ $K_{пер.i} = \frac{y_1}{y_{i-1}}$	с пер. базой	–	$K_{пер.1} = \frac{25}{31} = 0{,}81$	$K_{пер.2} = \frac{24}{25} = 0{,}96$	$K_{пер.3} = \frac{28}{24} = 1{,}17$	$K_{пер.4} = \frac{34}{28} = 1{,}21$
	с пост. базой	$K_{пос.0} = \frac{31}{31} = 1$	$K_{пос.1} = \frac{25}{31} = 0{,}81$	$K_{пос.2} = \frac{24}{31} = 0{,}77$	$K_{пос.3} = \frac{28}{31} = 0{,}90$	$K_{пос.4} = \frac{34}{31} = 1{,}10$
Темп роста, T_i, % $T_i = K_i \times 100\%$	с пер. базой	–	$T_{пер.1} = 81$	$T_{пер.2} = 96$	$T_{пер.3} = 117$	$T_{пер.4} = 121$
	с пост. базой	$T_{пос.0} = 100$	$T_{пос.1} = 81$	$T_{пос.2} = 77$	$T_{пос.3} = 90$	$T_{пос.0} = 110$
Темп прироста, $T_{пi}$, %, $T_{пi} = (K_i - 1) \times 100\%$ или $T_{пi} = T_i - 100\%$	с пер. базой	–	$T_{п1} = -19$	$T_{п2} = -4$	$T_{п3} = 17$	$T_{п4} = 21$
	с пост базой	$T_{п0} = 0$	$T_{п1} = -19$	$T_{п2} = -23$	$T_{п3} = -10$	$T_{п4} = 10$
Абсолютное значение одного процента прироста, A_i, кол-во прест. $A_i = \frac{\Delta_i}{T_i}$	с пер. базой	–	$A_1 = 0{,}32$	$A_2 = 0{,}25$	$A_3 = 0{,}24$	$A_4 = 0{,}29$
	с пост. базой	–	$A_1 = 0{,}32$	$A_2 = 0{,}30$	$A_3 = 0{,}30$	$A_4 = 0{,}30$

Средний уровень интервального временного ряда

$$Y_{ср.} = \sum \frac{y_i}{n} = \frac{(31 + 25 + 24 + 28 + 32)}{5} = 28{,}4.$$

Итак, в среднем в районе ежегодно совершается 28,4 преступлений.

Примечание: если бы временной ряд был не интервальным, а моментным (например, в вышеприведенной таблице данных в нижней строке стоял бы показатель «Количество лиц, имеющих непогашенную судимость»), то средний уровень моментного ряда с равными интервалами рассчитывается по иной формуле. А именно:

$$Y_{хр.} = \frac{\frac{1}{2}y_1 + \sum y_i + \frac{1}{2}y_{n-1}}{n-1},$$

для рассматриваемого примера:

$$Y_{хр.} = \frac{\frac{1}{2}y_1 + y_2 + y_3 + y_4 + \frac{1}{2}y_5}{4}.$$

Чтобы понять, почему формула иная и именно такая, как приведена выше, следует еще раз вспомнить, в чем суть моментного ряда. Здесь значение уровня (Y_i) не является итогом всего временного интервала (т. е. всех моментов времени, оказавшихся на этом интервале, как это имеет место для интервального ряда), напротив, здесь каждое значение относится только к какому-то одному моменту времени. При n уровнях у моментного и интервального рядов разное количество интервалов, поскольку y_1 стоит в начале первого интервала, в конце него стоит уже Y_2, а для интервального ряда Y_1, как итоговое значение, стоит лишь в конце первого интервала. Соответственно, в начале и в конце второго интервала для моментного ряда стоят Y_2 и Y_3, а для интервального ряда – Y_3 и Y_4 и т. д. Таким образом, номер последнего интервала, правой границе которого соответствует последний уровень ряда (Y_n), в случае интервального временного ряда оказывается имеющим номер n, а в случае моментного – (n–1). Внутренние в данном примере уровни (Y_2, Y_3 и Y_4) учитываются в полном объеме, а вот конечные уровни (Y_1 и Y_2) только отчасти. Так, граничный для всего данного ряда момент № 1 (начальный) только наполовину относится к настоящему (учитываемому) времени, а наполовину он относится к прошлому, не имеющему прямого отношения к данному ряду. Аналогично, граничный момент № 5 (здесь – конечный) – также только наполовину относится к настоящему (учитываемому) времени, а наполовину – к будущему, также не имеющему непосредственного отношения к данному ряду. Мы как бы разрезаем граничный момент на две половинки, после чего и уровень, соответствующий такому моменту, разделяется пополам (каждому «полумоменту» достается по «полууровню»), поэтому в формулу для данного ряда включаются лишь половины от уровней каждого из двух граничных моментов.

Все рассчитанные показатели динамики представлены в таблице попарно в зависимости от базы (основы) сравнения – идет ли речь о сравнении с предыдущим годом или с каким-то фиксированным, выбранным по каким-то определенным соображениям годом. В первом случае база не просто переменна, ею каждый раз становится предыдущий период, т. е. образуется своего рода цепь, когда прошедшее сравнение значение тут же переходит в основу для последующего сравнения уже другого значения, относящегося к последующему году. Поэтому показатели динамики с переменной базой именуются *цепными*. Для цепных показателей, характеризующих последовательные изменения от года к году, строятся показатели, усредняющие такие изменения. Они и будут в дальнейшем рассмотрены.

Средний абсолютный прирост ($\Delta_{средн}$)

Следует иметь в виду, что термины «рост» и «прирост» означают просто *динамику* без связи с тем, имеется ли рост в реальности или имеет место стабильность либо даже снижение (аналогично алгебраической сумме – понятию, которое включает и арифметические суммы, и разности; вспомним широко известное понятие экономистов «нулевого» или даже «отрицательного экономического роста», что при своей парадоксальности наименования привычно и никого не удивляет).

Вначале рассчитаем эту величину «в лоб»: просуммировав отдельные приросты и поделив затем полученную сумму на общее число таких приростов (т. е. на число переходов от года к году при сравнении) – в данном случае на 4.

$$\Delta_{средн} = \frac{\sum \Delta_i}{n-1} = 0{,}75.$$

Но, учитывая, что сумму в числителе можно преобразовать следующим простым образом:

$$\Delta_{средн} = \frac{\left(\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4\right)}{4} = \left(Y_2 - Y_1\right) +$$
$$\left(Y_3 - Y_2\right) + \left(Y_4 - Y_3\right) + \left(Y_5 - Y_4\right) = Y_5 - Y_1,$$

получаем второй более простой способ расчета этой же величины:

$$\Delta_{средн} = \frac{\left(y_{n-1} - y_0\right)}{\left(n-1\right)}.$$

Для наших данных имеем: $\frac{34-31}{4} = 0{,}75$, т. е. результат, разумеется, тот же самый.

Итак, при всех изменениях в сторону возрастания либо убывания в среднем ежегодно количество преступлений возрастает на 0,75 преступления. Следует обратить внимание на то, что прирост имеет единицу измерения – «преступление», а точнее «преступление/год».

Средний коэффициент роста

Здесь также счет можно вести по одной из двух ведущих к одному результату формулам. Первая формула:

$$K_{средн} = \sqrt[n-1]{K_1 \times \ldots \times K_{n-1}},$$

что для конкретных анализируемых данных означает:

$$K_{средн} = \sqrt[4]{K_1 K_2 K_3 K_4} = \sqrt[4]{0{,}81 \times 0{,}96 \times 1{,}17 \times 1{,}2} = 1{,}02.$$

Другой способ также основан на упрощении предыдущей формулы:

$$K_{средн} = \sqrt[n-1]{K_1 \times \ldots \times K_{n-1}} = \sqrt[n-1]{\frac{y_1}{y_0}\frac{y_2}{y_1} \times \ldots \times \frac{y_{n-2}}{y_{n-3}}\frac{y_{n-1}}{y_{n-2}}} = \sqrt[n-1]{\frac{y_{n-1}}{y_0}}.$$

Для рассматриваемых данных:

$$K_{средн} = \sqrt[4]{\frac{34}{31}} = 1{,}02.$$

Из-за округления возможны некоторые весьма небольшие расхождения в промежуточных расчетных данных. Желающие убедиться в полном совпадении обоих методов могут сделать это непосредственным счетом, увеличив точность расчета.

Итак, сопоставляя методики расчета показателей роста (изменения), с одной стороны, методику расчета $\Delta_{средн}$, а с другой – методику расчета $K_{средн}$ (что фактически то же $T_{средн}$, тот же показатель, но в процентах), можно сделать следующие весьма полезные для практики статистических расчетов выводы.

Когда в графе, соответствующей году № i, разность Δ_i отрицательна, тогда и коэффициент K_i обязан быть меньше единицы (соответственно T_i – меньше 100 %), но когда Δ_i больше нуля, то K_i обязан быть больше единицы (соответственно T_i – больше 100 %). Наконец, при равенстве нулю Δ_i (нет изменений в значениях показателя) K_i обязан быть равен единице (соответственно T_i равен 100 %). Верно и обратное: по тому, превосходит ли K_i единицу или нет, можно судить о знаке при Δ_i. Обнаружение в графе нарушения этого нехитрого правила свидетельствует о наличии ошибки в расчетах, но где именно (в расчете Δ_i или K_i), до пересчета сказать невозможно.

Другим важным выводом служит то, что средние показатели $\Delta_{средн}$ и $K_{средн}$ никаким образом от промежуточных уровней временного ряда не зависят. Так, значения этих уровней могут использоваться в расчете (первые способы в обоих средних показателях), но могут и не использоваться (вторые способы) – на итоговом результате это не сказывается. Это заставляет четко осознавать факт исключительной роли крайних уровней для усредненных (и только для них) показателей динамики.

Благодаря знанию среднего абсолютного прироста можно, начав с первого уровня (Y_1) и переходя от года к году на величину среднего прироста, прийти через ($n-1$) переход (здесь – на четвертом шагу) к конечному для временного ряда уровню, причем совершенно точно в него попадем, если считать с достаточной точностью. Можно продолжить движение и дальше, сделав, например, еще шаг № n, выйдя за пределы ряда. Этот выход, носящий наименование «экстраполяция», означает построение прогноза, когда получается значение в области будущего – по отношению к имеющимся данным и моментам времени, т. е. если мы смотрим данные, например, до 2010 г. включительно, то год 2011, хотя он уже и прошел, будет будущим по отношению к указанным данным. Переход от года к году в этом случае (для среднего абсолютного прироста) представляет собой известную из средней школы *арифметическую прогрессию*, изменения (в сторону повышения или понижения) происходят по лесенке с равными ступеньками – не быстрее и не медленнее. Кстати, это напрямую связано с тем, как получался указанный средний показатель: путем нахождения среднего арифметического значения.

Средний темп роста, в отличие от предыдущего показателя, определялся, что видно из формулы для него, как среднее геометрическое значение по отношению к усредняемым годовым темпам роста. Благодаря этой средней характеристике также можно, начав с первого уровня, через четыре (в данном случае) шага дойти в итоге до значения пятого уровня. Промежуточные значения в этом случае будут отличаться (и, как правило, заметно) от промежуточных значений для среднего абсолютного прироста, но заданные реальные граничные будут полностью совпадать с рассчитанными, в отличие от промежуточных, которые, как правило, не совпадают. Кстати, в этом случае переход от года к году на основе умножения значения, соответствующего предыдущему году, на средний коэффициент роста представляет собой геометрическую прогрессию. Выход за пределы, экстраполяция на будущее и здесь возможна и достаточно информативна. Однако следует иметь в виду, что прогноз на основе среднего коэффициента роста не совпадет с прогнозом на основе среднего абсолютного прироста подобно тому, как обе прогрессии не совпали и внутри ряда – между крайними уровнями. Построение графиков для обеих прогрессий позволяет лучше судить об использовании каждого вида усреднений.

Может быть задан вопрос: а где прогноз более правильный? Такая постановка вопроса не вполне правильна. Каждая из моделей прогнозирования опирается на свои отличающиеся допущения. При одном из них фактически считается, что все идет ровно «по ступенькам». Если это действительно в целом выполняется, то следует полагаться на прогноз на основе арифметической прогрессии (средний абсолютный прирост). Если же для ситуации следующего года существенным обстоятельством служит не фиксированная добавка (слагаемое) к имеющемуся, а доля, также фиксированная от достигнутого, которая в абсолютном выражении тем больше, чем больше уровень, на основе которого она определяется, от которого она отталкивается, то правильнее брать геометрическую прогрессию. Но когда в отношении указанных соображений имеются свои «за» и «против», то целесообразно построить оба прогноза, считая, что истина где-то вблизи обоих. Такой подход, как правило, обеспечивает более надежные выводы.

Когда значение конечного уровня резко отличается от предшествующих уровней, для экстраполяции может оказаться полезным прибегнуть к построению так называемой «скользящей средней» – пересчету значений с учетом уровней-соседей, что позволит скорректировать резкое отклонение конечного уровня от сложившейся на тот момент тенденции, снизить риск оказаться жертвой ошибки или нетипичной случайности, проявившейся в значении конечного уровня.

3. СТАТИСТИЧЕСКИЕ МЕТОДЫ ИЗУЧЕНИЯ ВЗАИМОСВЯЗИ ПРАВОВЫХ ЯВЛЕНИЙ И ПРОЦЕССОВ

3.1. Понятие о функциональной и статистической связях

Современная наука исходит из принципа взаимосвязи всех явлений и процессов природы и человеческого общества, хотя и весьма различающихся по силе своего проявления. Поскольку все такие взаимосвязи учесть невозможно теоретически, в человеческой практике появляется необходимость в большей или меньшей степени учета одних связей, возможность пренебрежения при определенных условиях другими и полное игнорирование третьих, наличие которых в обыденной жизни не ощущается ввиду ничтожности значений такой связи. Невозможно управлять явлениями и процессами, в том числе и социальными, без изучения и учета особенностей связей. Поэтому методы выявления и измерения связей, включая их силу и направленность, составляют чрезвычайно важную часть любого научного исследования.

Следует пояснить смысл, вкладываемый в такие понятия, как «явление» и «процесс». *Явление* – это статическое состояние, которое относится к определенному моменту времени, своего рода «стоп-кадр». В этот «стоп-кадр» попадает множество объектов, которые в жизни как-то изменяются, развиваются, причем каждый по своим законам. Однако, попав в «стоп-кадр», все эти объекты на миг замерли, образовав в совокупности некую конфигурацию, констелляцию, определенную статическую картину. Такая картина позволяет аналитику изучать в рассматриваемый момент структуры из объектов, а также соотношения таких структур. При переходе от одного объекта к другому внутри достаточно однородной совокупности можно отмечать и изучать особенности такого перехода, как изменение в пространстве. Примером явления может быть взятое состояние сил и средств правоохранительных органов в конкретный момент времени.

Для парной взаимосвязи (например, причины со следствием) возможны следующие варианты пар: между двумя явлениями, между двумя процессами, между явлением (причиной) и процессом (следствием) и между процессом (причиной) и явлением (следствием). Но поскольку процессы предстают в виде своих результатов-явлений, математический аппарат исследования, включающий различные приводимые ниже методы, будет одним и тем же для всех четырех указанных вариантов. Вот почему в последующем для краткости изложения чаще будет говориться о взаимосвязи лишь явлений, но подразумеваться при этом будут и процессы.

Связи между явлениями в зависимости от выбранного основания могут быть различным образом классифицированы. Вначале разделим связи в зависимости от степени категоричности выводов, делаемых в их отношении. В соответствии с этим различают два типа связей: *функциональную* (или *жестко детерминированную*) связь и *статистическую* (или *стохастическую*, или *вероятностную*) связь.

В пособии для простоты изложения речь будет идти о связи именно двух (а не трех или большего числа) явлений, которые здесь представлены своими признаками. Связь признаков явлений математически отображается в форме уравнения связи двух *переменных* (показателей)[1].

Если с изменением одной из переменных (X) вторая переменная (Y) изменяется строго определенным образом, если значению одной переменной обязательно соответствует хотя бы одно точно заданное значение (или сразу несколько) другой переменной, то связь между ними является *функциональной*. Пример: офицер на плацу командует строем, строй выполняет его приказы. Здесь зависимой переменной является поведение строя как единого объекта, но если говорить об отдельных военнослужащих как различающихся объектах, то это будет примером упомянутых выше нескольких значений (предполагается, однако, что в строю не новобранцы, иначе требуемая строгость в выполнении ими приказов может отсутствовать).

Функциональная связь двух величин возможна лишь при условии, что вторая из величин зависит от первой и только от нее

[1] Здесь термины *признак* и *показатель* будут синонимами, когда речь идет о свойстве объекта совокупности, а для меры силы и направленности связи будет использоваться термин *показатель*.

(ни от чего другого). В природе (тем более в обществе) таких связей нет; они являются лишь абстракциями, полезными и необходимыми при анализе явлений, но неизбежно упрощающими действительность, что далеко не всегда и не везде может считаться допустимым.

Однако некоторые разделы криминалистики, а также такие науки, как механика, электротехника, политическая экономия и ряд других успешно используют представление реальных связей как функциональных. Это возможно потому, что в сравнительно простых системах интересующая нас переменная величина зависит в основном (скажем, на 99 % или даже на 99,99 %) от какой-то другой, но также одной величины или лишь от нескольких, в то время как иное внешнее влияние минимально и, что важно, для имеющихся потребностей пренебрежимо мало, т. е. связь явлений в такой относительно несложной системе, хотя она и не является строго функциональной, но практически близка к таковой. Технически такая «простота» в приборе достигается ценой огромного труда создателей – ученых и инженеров, старающихся исключить любое постороннее влияние. Но простые системы демонстрирует и природа. Так, например, продолжительность земного года (период обращения Земли вокруг Солнца) почти функциональным образом зависит только от массы Солнца и расстояния Земли от него. На самом деле она зависит, пусть и в очень слабой степени, от масс других планет и их расстояний от Земли. Однако *нарушение функциональной связи*, вносимое планетами-соседями (удаленными звездами), для всех практических целей и задач человечества, включая и околоземную космонавтику, не имеет значения.

Пример из социальной сферы. Суд выносит приговор, когда, по его мнению, исследованы все относящиеся к делу существенные обстоятельства, хотя многие несущественные обстоятельства не учитываются да и вряд ли могут быть учтены, с одной стороны, ввиду чрезвычайно больших требуемых трудозатрат, а с другой – ввиду уверенности, убежденности суда, что и взятые все вместе такие неучтенные обстоятельства, будучи приняты во внимание, уже не способны заметным образом изменить сложившуюся в его представлении картину, а следовательно, и не способны повлиять на окончательное решение. Важная и нужная работа по совершенствованию деятельности суда направлена на повышение доли судебной правоты, но это доля никогда не будет доведена до 100 %, поскольку это невозможно в принципе. Это лишь один из примеров того, когда реальная взаимосвязь, ввиду практической близости к функциональной связи, вполне обоснованно рассматривается в практической деятельности.

Статистическая связь не имеет ограничений и условий, присущих функциональной связи. Если с изменением значения одной из переменных вторая переменная может в определенных пределах принимать любые значения с соответствующими вероятностями, но при этом ее статистическая (массовая) характеристика (например, среднее значение) изменяется по определенному закону, то говорят, что связь является статистической. Статистическая связь, которая не столь категоричная, как функциональная, принципиально не стопроцентная, приобретает законное право на описание реальной ситуации в тех случаях, когда без нее уже не обойтись и попытки описать эту ситуацию более простым способом порождают недопустимую для целей данного аналитического исследования неточность, и, кроме того, не позволяет охватить ситуацию целиком. Так, разброс попаданий по мишени при выстрелах в тире нельзя охарактеризовать какой-то одной точкой, поскольку результаты стрельб при этом оказались бы чрезмерно упрощенными.

Если говорить о направленности связи между явлениями, то они подразделяются на *причинно-следственные* связи, *сопутствующие* связи и связи, где каждое явление в паре есть одновременно и причина, и следствие для другого. Рассмотрим каждый из указанных видов направленности связи, представляющих собой различные пути возникновения связей.

1. Причинно-следственные связи. Это важнейший путь возникновения связей – причинная зависимость результативного признака (его вариации, изменения значений) от вариации, изменения факторного признака. Например, признак X – это относительный уровень преступности (на 100 тыс. чел.), а Y – это балльное значение оценки сложившемуся уровню преступности какой-то территории (например, качественный уровень опасности). Здесь логика ясно указывает, какой признак выступает как независимая переменная (X), а какой – как зависимая переменная (Y).

2. Сопутствующие связи. Если в качестве признака X взять численность сотрудников подразделения экономической безопасности и противодействия коррупции (ЭБиПК) в городе, а за признак Y – сумму убытков (ущерба) за год в городе от экономических преступлений, то между признаками X и Y в совокупности городов России окажется существенная прямая зависимость. Прямая зависимость в данном случае означает, что чем больше таких сотрудников в городе, тем и больше ущерб от экономических преступлений. Отсюда может создаться впечатление, что чуть ли не численность сотрудников по борьбе с экономическими преступлениями и кор-

рупцией определяет размер материального ущерба, что представляет очевидную несуразность. В чем же здесь кроется противоречие? Оказывается, данную пару явлений (сотрудники и ущерб) нельзя интерпретировать в виде причины и следствия, поскольку оба признака (численность подразделения и сумма ущерба) являются следствиями общей причины: такого явления, как экономическая активность города. Вполне логично, что в тех городах, где больше экономическая активность (а следовательно, и экономических преступлений), будет больше и численность сотрудников подразделений ЭБиПК, но ведь по той же самой причине там будет и больше ущерб от экономических преступлений. Два представленных следствия одной общей, но не рассматриваемой здесь причины, сопутствуют друг другу. Более того, по данным об одном из них можно достаточно определенно судить и о данных по другому (и наоборот). Однако, ограничиваясь рассмотрением только этой пары, не выходя за ее пределы, нельзя предложить никакую разумную причинно-следственную связь, что и заставляет аналитика, интерпретируя статистическую связь, искать иную, уже истинную причину. Здесь приведен несколько видоизмененный реальный пример, касающийся деятельности пожарных команд в городах дореволюционной России и ущерба от пожаров в них, рассмотренный известным русским статистиком А. А. Чупровым.

Другой пример. Прошедшие полтора десятка лет XXI в. отмечаются многократным ростом мобильных телефонов, причем как в среде сотрудников правоохранительных органов, так и в среде преступников. Соответствующий высокий показатель связи между ростом числа аппаратов у одних и у других свидетельствует о наличии сильной статистической связи. Это типичный пример сопутствующей связи, поскольку оба роста – это лишь следствия единой причины: общего бурного развития средств мобильной связи во всех без исключения слоях общества, включая и названные выше, при всем их различии.

3. Связи, где каждое из явлений – одновременно и причина, и следствие. В этом случае каждое из пары явлений поочередно или непрерывно выступает то причиной, то следствием. Такова, например, зависимость между уровнем производительности труда сотрудников ОВД и уровнем оплаты труда. С одной стороны, уровень оплаты труда устанавливается в зависимости от круга должностных обязанностей, проявляющихся в итоговой производительности труда: чем выше производительность, тем выше и оплата. Но, с другой стороны, при правильной системе оплаты материальные стимулы

выступают в качестве фактора, от которого зависит производительность труда. В такой системе признаков допустимы обе постановки задачи: каждый признак выступает и в роли независимой переменной X, и в качестве зависимой переменной Y.

Другой пример связи «одновременно и причина, и следствие» – «гонка» профессионализма правоохранительных органов и преступного мира – рост одного с неизбежностью должен вести и к росту другого (чтобы выстоять, выжить), одно явление «подстегивает» другое, и наоборот. В качестве отдельного примера можно указать на меры населения по усилению защиты своего имущества от проникновения преступников (установка железных дверей и решеток и т. п.) и одновременный рост квалификации преступников, заинтересованных в указанном проникновении. Модель такой взаимосвязи представляет собой раскручивающуюся спираль и носит заимствованное из кибернетики наименование «положительной обратной связи». Как правило, раскручивание на некотором этапе все же приостанавливается, поскольку начинают сказываться и другие сдерживающие факторы, до того в модели не учитываемые, поскольку были еще пренебрежимо малы.

Итак, специалисту в соответствующей отрасли требуется установить, наличествует ли факт взаимосвязи между двумя интересующими его явлениями и квалифицировать выявленную связь с точки зрения причинно-следственных отношений (имеет ли место тот или иной рассмотренный выше тип связи). Что касается подобной квалификации, то решение этого вопроса лежит сугубо на специалисте той отрасли, к какой относятся анализируемые явления и в какой он, следовательно, обязан быть компетентным. Именно он должен, начертав литеры *А* и *Б* как обозначение пары явлений, соединить их затем соответствующей причинно-следственной стрелкой, если определит связь в качестве таковой, или же, напротив, чертить уже не стрелку, а нечто иное, отражающее иные типы связи, если причинной зависимости между *А* и *Б* не может быть на основе содержательной теории. По сути дела, специалист в соответствующей предметной отрасли (криминолог, юрист-цивилист, политолог и т. п.) строит качественную модель, в которой отражена причинно-следственная логика, именно в этом заключается роль отраслевой статистики и ее представителей. В отличие от этого роль общей теории статистики заключается в том, чтобы рассчитать количественное значение показателя связи между явлениями. При *наличии пороговых значений* для рассчитанного показателя появляется возможность охарактеризовать его и в качественном отношении (сильная ли выявленная связь), а также указать, как с ростом значения признака (показателя) одного явления будет вести себя значение признака (показатель) другого. Но то, каков тип установленной статистической связи, на этот вопрос общая теория статистики ответа не дает, это вне ее компетенции. Касаемо проблематики органов внутренних дел, это компетенция криминалиста, криминолога, специалиста по работе с персоналом (кадровика) и т. д. – в зависимости от конкретной области проблемы.

Теперь следует отметить такой важный для рассматриваемой темы факт, как невозможность использования во всех случаях какого-то одного единственного (универсального) метода для измерения взаимосвязи явлений (процессов) и указать на те фундаментальные информационные обстоятельства, которые это обуславливают. На первый взгляд представляется, что если уж имеется несколько методов, находящих применение в исследовании взаимосвязей явлений (процессов), то достаточно выбрать из них лучший в каком-то смысле, чтобы только к нему в последующем и обращаться. Главная причина невозможности отбора единственного метода заключается в том, что различного рода данные, представляющие явление, и которые предстоит обрабатывать, при возможной внешней их схожести на письме (вплоть до полного совпадения символов) могут отличаться друг от друга принципиальным образом. Так, методы обработки информации вполне возможно уподобить оборудованию по переработке имеющегося информационного сырья в желаемую готовую информационную продукцию (в качестве таковой служат на практике промежуточные и окончательные аналитические выводы). Тогда понятно, что оборудование должно быть приспособлено к конкретному виду сырья, а с другим видом в лучшем случае будет работать весьма неэффективно, с большими потерями, а в худшем – вообще работать не сможет. Для методов обработки информации дело усугубляется еще и тем обстоятельством, что аналитик, неправомерно использующий какой-то метод (как правило, даже и не осознающий своей ошибки), получив результат, не поймет и того, что он бессмысленный (бессмыслица будет выдана компьютером в столь же красивом виде, что и прочие вполне осмысленные результаты, и затеряется в общей массе результатов). В итоге ложный результат ляжет в основу управленческого решения (и, возможно, даже и не одного), которое станет тоже ложным.

Поясним на простом сквозном примере, как может различаться содержание во внешне схожих данных. Так, возьмем символ

«5» и подумаем, что о нем, на основе одного только его вида, можно сказать применительно к сфере деятельности органов внутренних дел. Много это или мало и что вообще стоит за этим символом? Лучше или хуже «5», чем «4» или «7»? И даже проще: больше ли 5, чем символы 4 или 7? И если «да, больше», то на сколько именно? И во сколько раз больше? Казалось бы, ответы, по крайней мере на три последние вопроса, вполне очевидны даже младшему школьнику: вычитай и дели. Но дело будет выглядеть далеко не столь очевидным, если выяснится, что «5» – это номер отделения полиции. Но предпочтительнее ли 5 как номер отделения иным номерам? Разумеется, нет, он не лучше и не хуже. Без каких-либо дополнительных данных, без иной необходимой информации оценивать значения символов бессмысленно. Сами по себе символы 5, 4, 7 (как и любые другие) смыслом не обладают, смысл в них должен быть привнесен в виде дополнительной информации, которая становится для каждого из них после этого характеристикой. Таким образом, до попытки дать какой бы то ни было ответ на задаваемый вопрос, должно последовать уточнение относительно смысла затрагиваемых в вопросе символов. Итак, когда символ 5 служит номером для некоего органа внутренних дел, т. е. играет роль имени, его значение не хуже, но и не лучше любого другого имени (№ 4, или № 7, «Зареченский» или «Заводской» и т. д.).

Но в другой раз тот же символ «5» может означать уже номер места, которое занял соответствующий территориальный орган внутренних дел районного уровня по итогам года в своей области. Если лучшим местом является № 1, тогда про ОВД, занявший место № 5, справедливо будет сказать, что этот орган потрудился лучше, чем тот, что занял место № 7, но хуже, чем занявший № 4.

Но вот то, на сколько именно и во сколько именно ОВД с пятым местом сработал лучше (или хуже) каждого из прочих ОВД области, сказать на основе только информации о месте невозможно. Другими словами, сказать можно лишь то, на сколько единиц номер одного места больше (либо меньше) другого номера, но не насколько больше (меньше) результат работы, определивший место. Итак, «5» уже обретает некоторые признаки привычной в повседневном быту пятерки, которая больше четверки и меньше семерки. На этом сходство данного символа «5» с нашей «бытовой» пятеркой кончаются.

Для развития нашего примера с символом «5» предположим, что в некой области до сих пор еще сохраняется порочная «палочная» система оценки деятельности сотрудников ОВД с установленной для них месячной «нормой» на число «палок» (допустим, задержаний правонарушителей), ведется учет деятельности в форме учета отклонения каждого из сотрудников от назначенной «нормы» – в большую либо меньшую сторону. При таком учете «5» означает, что имело место «перевыполнение нормы» на это самое число единиц, то же самое относится и к «4», и к «7». Зато «–5», ввиду минуса, уже означало бы «недовыполнение нормы» с соответствующими неутешительными для сотрудника последствиями. Исходя из такого смысла всех рассматриваемых символов, можно обоснованно сказать, насколько именно лучше или хуже (с точки зрения руководства органа с его порочной системой оценки) сработал сотрудник с результатом «5» в сравнении с любым другим сотрудником. Так, очевидно, что результат «5» на единицу больше результата «4» (лучше) и на две единицы меньше результата «7» (хуже). В то же время в принципе невозможно сказать, во сколько раз сотрудник с показателем «перевыполнения» «5» сработал лучше (или хуже) тех, у кого этот же показатель «4» и «7», поскольку вывод основывается лишь на имеющихся данных, в состав которых информация о значении «нормы» не входит.

Но вот когда значение указанной «нормы» становится известным, входит в состав анализируемых данных, то появляется уже возможность ответить и на вопрос о том, во сколько раз результат одного сотрудника превосходит результат другого: это отношение значения («норма» плюс 5) к значениям («норма» плюс 4) и («норма» плюс 7) соответственно. Если же предписываемой сотрудникам «нормы» вообще нет (как то и положено), то символы 5, 4 и 7, будучи значениями непосредственно результатов какой-либо деятельности сотрудников (количество задержаний правонарушителей и т. п.), являются уже вполне привычными нам числами, а потому могут на законном основании (с точки зрения информационно-аналитических действий) вычитаться одно из другого и делиться одно на другое (кроме, разумеется, деления на нулевые значения).

Все вышеприведенные различающиеся между собой смыслы одного и того же символа «5» вполне определенно указывают на то, что и методы аналитической обработки для символьных данных с разным смыслом также должны быть различными, т. е. адекватными смыслу обрабатываемых данных. Проблемой смысла данных и, соответственно, допустимостью методов обработки каждого из типов данных занимается *теория измерения*[1], устанав-

[1] *Яковлев А. М. и др.* Рабочая книга социолога: монография / под ред. Г. В. Осипова. 5-е изд. М., 2009. С. 142–150.

ливающая для различных данных свои уровни измерения, определяемые измерительными шкалами. Говоря об объектах социальной природы, гуманитарной сферы, можно констатировать, что измерение находит широкое применение при решении самых разнообразных задач статистики в таких областях практического знания, как юриспруденция, криминология, социология, экономика, психология и т. п. Именно теория измерения ясно указывает, какие методы и предполагаемые ими математические действия допустимы, разрешены по отношению к подлежащим обработке информационным данным соответствующего уровня измерения. Нарушение же предписаний теории измерения с неизбежностью приводит к получению неправомерных, а потому и ложных выводов, что ведет в итоге к ошибочным управленческим решениям. При этом у управленца останется твердая уверенность, что все проделано верно и его решения базируются на выводах и рекомендациях науки, предложившей ему метод обработки данных. И это притом, что подобные неудачные решения чаще всего будут гораздо хуже тех, которые управленец выработал бы сам, не прибегая ни к каким рекомендациям и действуя лишь на основе собственной интуиции, по здравому смыслу.

Согласно теории измерения можно выделить четыре основных уровня измерения. Качественный уровень измерения представляют шкалы *номинальная* (*классификационная*) и *порядковая* (*ранговая*), а количественный уровень – шкалы *интервальная* и *абсолютная* (*метрическая*). Когда идет речь об измерении объекта (явления, процесса), то подразумевают, что измеряется какой-то конкретный признак, именно об уровне измерения этого признака и идет речь в этом контексте. Так, например, если в объекте аналитика интересуют два признака, то по одному признаку объект считается измеримым по шкале номинальной, в то время как по другому признаку – по порядковой шкале. Другими словами, нет понятия какого-то абстрактного измерения объекта, не зависящего от его конкретных признаков. Рассмотрим названные уровни измерения.

Номинальная шкала – это низший уровень измерения, при котором все измерение сводится к определению родового *имени* объекта или, что то же самое, к определению *класса* объекта. Но и это простейшее в сравнении с прочими измерение очевидным и беспроблемным оказывается далеко не всегда.

В подтверждение сказанного достаточно указать на такие примеры измерения по номинальной шкале, как квалификация совершенного уголовного преступления (имел ли место грабеж или разбой; было ли убийство или только тяжкое телесное повреждение, повлекшее смерть; имела ли место необходимая оборона или нет и т. п.), что никак не отнести к разряду фиксации очевидного и банального и способно повлечь за собой далеко идущие последствия и для конкретных людей, и для обеспечения правопорядка в обществе в целом. Ведь и вердикт – приговор суда присяжных («виновен» или «невиновен») – это также измерение по номинальной (классификационной) шкале. При сопоставлении данных, измеримых по номинальной шкале, допускается взаимное сравнение только по принципу «равно/неравно» («эквивалентно/неэквивалентно»), на основе чего и производится классификация объектов в классы (группы, типы и т. п.). Объекты из одного класса – эквивалентны друг другу, но только в том смысле, по которому выбиралось основание классификации. Соответственно, для номинальной шкалы не существует такого соотношения между объектами, как «больше/меньше». Характеристика сотрудника как полицейского – это также измерение по номинальной шкале. Другие из множества примеров измерения по этой шкале – имя и фамилия человека, его пол, адрес, профессия, национальность, наличие судимости, темперамент, наличие навыков владения оружием, а также: вид оружия, вид службы в ОВД, наименование города, края, страны, группы стран, вид структуры управления, способ совершения преступления и т. д. В тех информационных таблицах, с которыми постоянно имеют дело в штабах и иных информационно-аналитических подразделениях, большинство объектов измеряется именно по номинальной шкале. Каждый объект принадлежит к своему классу, причем, как правило, и не одному, но только в разных системах классификации. Так, плотницкий топор в одной классификации – орудие для взлома двери, а в другой – орудие убийства (разумеется, возможны и иные для него измерения, например, строительный инструмент и т. д.).

Столь привычное большинству аналитиков среднее арифметическое значение для номинальной шкалы не существует, поскольку здесь вообще ничего нельзя ни складывать, ни делить – операции сложения и деления попросту отсутствуют. Кому этого недостаточно, можем порекомендовать попробовать к имени Иван прибавить имя Павел, чтобы потом, усредняя, все это поделить пополам, и тогда этот вопрос в отношении шкалы имен, надеемся, отпадет сам собой. Но среднее, понимаемое как нечто типичное (в смысле характерное, образцовое, шаблонное, центральное) для всей совокупности объектов, существует и для номинальной шкалы. Таким средним для номинальной шкалы служит мода (обозначается *Mo*), являющаяся наиболее распространенным вариантом значений признака в совокупности объектов. Иными словами, это вариант, представляющий собой в совокупности относительное большинство.

У моды как вида средних характеристик совокупности, отличающей ее от всех других существующих видов, имеются две особенности.

Во-первых, это то, что мода может быть определена как в отношении качественных шкал (пример тому номинальная шкала), так и для количественных (если таков уровень измерения признака объекта). Так, качественный признак «вид подразделений полиции» имеет различные значения, какими являются наименования подразделений. Для конкретной совокупности полицейских, например, указанных в телефонном справочнике, естественно считать типичным, характерным именно то подразделение, которое представлено в наибольшей мере (причем именно в справочнике, а не в органах внутренних дел региона или в целом по стране). Другими словами, средней (в данном случае модальной) в рассматриваемой совокупности считается то подразделение, сотрудники которого представляют в совокупности относительное большинство; шанс встретить представителя этого подразделения наибольший. Если интересует количественный признак – возраст, то, исходя из числа прожитых лет, можно определить модальный для данной совокупности возраст.

Во-вторых, мода может быть не единственной – одновременно может быть и две, и больше мод, например, если сразу два варианта окажутся наиболее (и, разумеется, в равной мере) распространенными.

При относительной бедности информации об объектах, измеряемых по номинальной шкале, общая теория статистики смогла предложить определенные показатели, измеряющие взаимосвязи. Отметим, что результатом измерения взаимосвязи между номинальными данными является, тем не менее, количественная характеристика. Это обстоятельство позволяет аналитику делать вывод в отношении взаимосвязи, пусть и не слишком глубокий (в сравнении с другими шкалами), но вполне *объективный*. Альтернативой же этому объективному, научно обоснованному выводу могут служить лишь *субъективные* выводы (часто у аналитика свои выводы и даже диаметральные выводам коллег). И только научный анализ данных в силу объективного характера способен выступить в роли «общего знаменателя» в решении вопроса о наличии закономерностей и правильностей в данных анализируемых таблиц.

В порядковой (ранговой) шкале объекты могут быть упорядочены по значениям рассматриваемого признака, выстроены в цепочку, именуемую *рядом*. Каждый объект после этого получает свой номер (ранг) в ряду, который и является исчерпывающей характеристикой для измерения по порядковой шкале. Упорядочивать объекты можно по принципу «больше/меньше», «хуже/лучше», «активнее/пассивнее», «выше/ниже» и т. д. Примерами могут быть виды преступлений, если для любых двух видов можно указать, какой из них более опасный, а какой менее (или они в этом смысле равны). Но те же самые виды преступлений можно упорядочить и по иному признаку, например, по тому, какой пол преобладает среди лиц, совершающих преступления: мужчины или женщины. Это, как правило, будет уже совершенно иной ряд, поскольку его упорядоченность основывается на совершенно ином признаке. Ряд, упорядоченный по возрастанию значения признака, именуется в статистике *вариационным* рядом.

Порядковая (ранговая) шкала, являясь одновременно и номинальной (классификационной), обладает и всеми свойствами последней, но как частный ее случай, и некоторыми дополнительными свойствами, отсутствующими у номинальной. Это обстоятельство позволяет существенно расширить информационные возможности в аналитической работе, получая более глубокие выводы, в том числе и в отношении взаимосвязи явлений. Дополнительные свойства обусловлены тем, что к существующему у номинальной шкалы сравнению объектов по принципу «равно/неравно», добавляется и сравнение по принципу «больше/меньше». Это позволяет не только образовать классы эквивалентных (равных) объектов по значениям рассматриваемого признака, но и упорядочить эти классы по возрастанию. Однако же, казалось бы, на вопрос о том, *на сколько* (а также и *во сколько*) именно одно значение признака больше (или меньше) другого значения, порядковая шкала ответа не дает и дать не может – этот количественный вопрос не для качественной шкалы. Вот почему ни понятия разности, ни понятия отношения (деления) значений признака для порядкового уровня измерения попросту не существует. Хотя порядковая шкала, как и номинальная, относится к *качественным* шкалам, но уже более высокого уровня, а потому присущее ей сравнение по принципу «больше/меньше» позволяет делать выводы более глубокими, чем для номинальной, что заметно приближает ее к шкалам количественным.

Что касается среднего значения для порядковой шкалы, то кроме моды, воспринятой от номинальной шкалы, она обладает и другим, своим собственным средним значением, адекватным именно этому уровню измерения, именуемым *медианой* (*Me*). Медиана представляет собой значение признака объекта, располагающегося точно посередине вариационного ряда. Вот почему для медианного объекта верно то, что количество других объектов со значением рассматриваемого признака меньшим, чем у него, ровно такое же, как и объектов со значением рассматриваемого признака большим, чем у него.

Приведем некоторые примеры порядковых шкал. *Ранг*, т. е. место, занимаемое конкретным органом внутренних дел среди прочих (по итогам года, в спортивных соревнованиях и т. п.), – это типичный пример измерения по порядковой шкале: места, присвоенные ранжированным районам области в зависимости от сложности криминальной обстановки в них; места видов преступлений, если такие виды удалось расположить по средней общественной опасности преступлений соответствующего вида. Но виды преступлений можно расставить и по-иному: в соответствии со средней сложностью раскрытия преступлений соответствующего вида. В этом случае каждый вид преступлений получает свой ранг (номер места), отличный от ранга предыдущего упорядочения, поскольку здесь уже иной учитываемый признак. Также примерами измерения по указанной шкале служат уровни управления ОВД и перечень служебных должностей, упорядоченные по статусу.

Если про два значения признака можно сказать не только, что один из них больше другого, но и на сколько именно, то такая количественная шкала именуется интервальной. Интервальная шкала, относясь к количественным шкалам, одновременно является и номинальной, и порядковой, однако при этом имеющей и дополнительные (в сравнении с порядковой шкалой) свойства. Это еще больше расширяет ее возможности и, соответственно, позволяет сделать более глубокие выводы. Однако в отношении данных, измеримых по этой шкале, по-прежнему бессмысленно спрашивать, во сколько раз одно больше другого, что связано с такой особенностью интервальной шкалы как условность нуля. Хотя у этой шкалы и имеется нулевая точка отсчета, но этот нуль не истинный, поскольку он может быть выбран или изменен аналитиком, исходя из соображения его удобства. Никакого ущерба для выводов, сделанных ранее (при прежнем нуле) от смены нуля в интервальной шкале не происходит, хотя пересчет данных (их общий сдвиг по оси) все же необходим. Например, шкала Цельсия на градуснике – это интервальная шкала, нуль ее условен, поскольку выбрали в качестве него точку замерзания воды, но ведь могли же выбрать и что-то другое; скажем, медики могли бы предложить в качестве удобного для них нуля нормальную температуру человеческого тела. Именно поэтому, если с утра было 2 градуса тепла, а к полудню стало 6 градусов, то верным будет сказать, что полуденная температура на 4 градуса выше, но неверным – что выше она в 3 раза. Абсурдность последнего высказывания особо видна, если считать, что утром было 0 градусов, а в полдень похолодало до –2 градуса (тут и на нуль делить нельзя, и непонятно, что с минусом делать).

Примеров интервальной шкалы в сфере правоохранительной деятельности (как и вообще в социальной сфере) не так уж и много, и один из них – это временные даты, т. е. точки на оси времени (например, год рождения человека, год основания города, год упразднения организации и т. п.). Если говорить о дате, то, зная года рождения двух людей, можно сказать, на сколько лет один из них старше другого. Но, во сколько раз он старше, сказать только на основе имеющихся данных (без указания, когда, в какой именно момент задается вопрос) невозможно, т. к. в разные моменты и ответы были бы разными. Примером интервальной шкалы будет и рассмотренная выше ситуация с заданием незаконной «палочной нормы», тут значения отклонения от «нормы», взятые как самостоятельные показатели работы, представляют собой именно такие измерения.

В интервальной шкале к уже имеющимся средним значениям (моде и медиане) добавляется и знакомое всем *среднее арифметическое значение*, которое адекватно этой шкале.

Если интервальная шкала обладает истинным нулем, то это уже будет абсолютная (или метрическая) шкала. Данные, измеряемые по этой шкале, уже можно складывать и вычитать, умножать и делить (в последнем случае – только бы не на нуль). Если даты рождения людей измеримы по интервальной шкале, то их возрасты уже измеримы и по абсолютной. Примеры измерений по абсолютной шкале весьма многочисленны: численность сотрудников, количество управленческих уровней, доход предпринимателя, количество преступлений по району, продолжительность трудовой смены, относительный уровень преступности (на 100 тыс. чел.) по области и т. д. К данным, измеряемым по абсолютной шкале, применимы уже все без исключения методы математической обработки.

К трем видам средних значений, имеющихся у интервальной шкалы (среднему арифметическому, моде и медиане), добавляется еще одно, характерное для абсолютной шкалы, – *среднее геометрическое значение*, используемое, например, в анализе динамических (временных) рядов при расчете среднего коэффициента и темпа роста[1]. Если, как уже отмечалось, температурная шкала Цельсия – пример интервальной шкалы, то температурная шкала Кельвина, ввиду наличия у нее абсолютного нуля, уже является абсолютной шкалой. Для шкалы Кельвина становится верным то, что не было верно для шкалы Цельсия: замеренная по шкале Кельвина температура, равная 6 градусам, в три раза выше, чем температура, равная 2 градусам.

Таким образом, измерительные шкалы последовательно вложены одна в другую (по принципу «матрешки»), причем из любых двух более внутренней шкале соответствует более высокий уровень измерения и, соответственно, более глубокие выводы. Шкала более высокого уровня измерения обладает одновременно и всеми свойствами шкалы более низкого уровня, а потому к данным, измеряемым по шкале более высокого уровня, применимы и все методы, адекватные шкалам более низкого уровня.

3.2. Понятие о типе показателей коэффициентов корреляции

Ознакомившись с существующими уровнями измерения данных, можно для шкалы каждого уровня указать соответствующие ей методы измерения взаимосвязей, привести соответствующие им показатели взаимосвязи. К данным, измеримым по шкале более высокого уровня измерения, применимы методы, адекватные более низкому уровню измерения, однако при этом происходит определенная потеря информации, что нередко допустимо, но всякий раз должно четко осознаваться. Каждый из методов состоит в производстве расчета соответствующего показателя, значение которого и дает ответы на вопросы относительно связи. Рассмотрим *взаимосвязи пары* показателей как наиболее часто встречающегося и потому наиболее важного случая (хотя в статистике имеются и показатели взаимосвязи между множеством явлений). Показатели парных взаимосвязей специально конструировались их создателями таким образом, чтобы удовлетворять ряду стандартных требований, весьма упрощающими их применение и делающими более удобной интерпретацию результирующих значений. Два из трех приводимых ниже показателей взаимосвязи именуются *показателями*

[1] *Елисеева И. И., Юзбашев М. М.* Общая теория статистики. М., 2001. С. 86–92.

корреляции (с соответствующими уточняющими расширениями в наименовании), третий же показатель, хотя и носит иное наименование, ввиду единства конструкции примыкает к ним и потому также может быть отнесен к типу показателей коэффициентов корреляции. *Корреляция* – статистический термин, о котором в рамках изучаемой темы важно знать лишь то, что он означает наличие соответствия в изменениях, развитии соответствующих двух явлений. Эти изменения у явлений могут быть либо сходны в одновременном росте и убыли, либо, напротив, зеркально противоположны (росту соответствует убыль и наоборот), но в обоих случаях изменения одного явления способны достаточно точно описывать изменения другого.

Следует отметить, что независимо от того, по какой шкале измеряются явления, взаимосвязь между которыми исследуется, показатель, относящийся к типу показателей коэффициентов корреляции, всегда измеряется по высшей – *абсолютной* – шкале, что позволяет применять к его значениям при дальнейшей аналитической обработке арсенал математической статистики уже в полном объеме.

Итак, пусть K – показатель, относящийся к типу показателей коэффициентов корреляции. Отметим, в чем именно заключаются свойства таких показателей, и поясним их смысл.

1. $-1K \leq +1$ – значения показателя связи *нормированы*, т. е. никогда, ни при каких обстоятельствах они не могут выходить за указанные верхний и нижний пределы. Встречающиеся порой высказывания о том, что «коэффициент корреляции равен плюс 10» или «минус 2, 71» и т. п., свидетельствуют лишь о полном непонимании сути показателей данного вида).

2. $K > 0$ – связь между двумя явлениями прямая (положительная), это означает, что у их изменений одинаковая направленность. Если рассуждать по-иному, то при *положительной корреляции* с ростом значений показателя, относящегося к одному явлению, значения показателя, относящиеся к другому явлению, также возрастают (и наоборот – с убыванием одних убывают и другие).

3. $K < 0$ – связь между двумя явлениями обратная (отрицательная), это означает, что у их изменений противоположная направленность. Рассуждая по-другому, при *отрицательной корреляции* с ростом значений показателя, относящегося к одному явлению, значения показателя, относящиеся к другому явлению, напротив, убывают (и наоборот – с убыванием одних другие возрастают).

4. $K = 0$ – взаимосвязь между двумя явлениями отсутствует; это проявляется в том, что их изменения никак не согласуются между

собой. Другими словами, при *нулевой корреляции* рост или убывание значений показателя, относящегося к одному явлению, никаким образом не отражается на росте или убывании значений показателя, относящегося к другому явлению.

5. $K=+1$ – между двумя явлениями связь *прямая*, причем из статистической перешедшая уже в разряд *функциональных*. В данном случае по изменениям значений показателя, относящегося к одному явлению, можно со стопроцентной точностью описать и изменения значений показателя, относящегося к другому явлению, причем фазы таких изменений *полностью совпадают*.

6. $K=-1$ – между двумя явлениями связь *обратная*, причем из статистической перешедшая уже в разряд *функциональных*. Хотя, как и в предыдущем случае, по изменениям значений показателя, относящегося к одному явлению, можно со стопроцентной точностью описать и изменения значений показателя, относящегося к другому явлению, однако в отличие от предыдущего случая такие изменения *прямо противоположны* (изменения в противофазе, наблюдается *инверсия* одних относительно других).

В дополнение к парному коэффициенту корреляции целесообразно рассматривать и такой статистический показатель взаимосвязи явлений, как *коэффициент детерминации*, обозначаемый символом η^2 и представляющий собой *квадрат коэффициента корреляции* ($\eta^2=K^2$). Поскольку квадрат любого числа всегда положительный (точнее, неотрицательный, ведь K может равняться и нулю), значения коэффициента детерминации заключены в пределах от нуля до единицы. Смысл этого показателя взаимосвязи пары явлений в том, чтобы указать, на сколько велика относительная доля изменений значений признака (показателя), относящихся к одному явлению, которая может быть объяснена изменениями значений признака (показателя), относящихся к другому явлению. Другими словами, коэффициент детерминации говорит о том, в какой мере одно явление, рассматриваемое в качестве причины, влияет на другое явление, рассматриваемое в качестве следствия (если связь причинно-следственная). Дело в том, что на результат-следствие могут влиять и другие причины, и условия, другие явления (что происходит в большинстве случаев). Значения этого показателя, характеризующего долю влияния одного из рассматриваемых показателей на другой, можно указывать как привычнее аналитику: либо в процентах – от 0 % до 100 %, либо в виде дробных чисел: от 0 до 1.

Далее рассмотрим конкретные показатели взаимосвязи явлений, относящихся к типу *коэффициентов корреляции* (хотя и не все в наименовании содержат термин «корреляция» – соответствие), адекватных той или иной из приведенных шкал измерения.

3.3. Измерение взаимосвязи для данных, измеримых по номинальной (классификационной) шкале

Роль номинальных данных в социологии вообще и отраслевой социологии, связанной с деятельностью ОВД, огромна. Объяснить это можно следующими, во многом взаимосвязанными причинами.

Во-первых, именно номинальные данные чаще всего используются социологами. Это во многом объясняется объективной сложностью измерения объектов социальной природы, а также сравнительной простотой их получения, естественностью интерпретации, интуитивной уверенностью в состоятельности последней.

Во-вторых, номинальные данные в качестве исходных часто являются более надежными, чем данные, специальным образом измеренные на основе шкал более высокого типа, не всегда в реальности соответствуют такому более высокому уровню измерения.

В-третьих, в методах, используемых для анализа номинальных данных, обычно подразумеваются модели, не вызывающие сомнения, вполне отвечающие естественной логике социолога или иного исследователя, обрабатывающего собранную информацию наиболее простыми методами, когда использование компьютерных программ служит лишь ускорению расчета.

Таблица с исходными либо аналитическими целочисленными данными – типичная для управленческой практики форма представления данных для последующей работы руководителя, принимающего на основе нее свои решения. Это могут быть данные по районам области в отношении совершенных видов преступлений, т. е. сопрягается информация о месте совершении преступления с видом преступления. Но могут быть сопряжены и данные о видах преступлений с характеристиками лиц, их совершивших (например, такими, как пол, сведения о судимости и т. п.), а также иные варианты сопряжения.

Сформированная прямоугольная таблица, имеющая в своем составе m строк и n столбцов, именуется в статистике «таблицей сопряженности $m \times n$» (читается «m на n»). В таблице сопрягаются (взаимодействуют) два признака: значения одного из них составляют «боковины» таблицы, а значения другого признака определяют состав «шапки», что и определило такое наименование таблицы. Каждое табличное значение (элемент таблицы) отражает частоту сопряжения конкретных значений обоих сопрягаемых признаков.

Представляется естественным использовать для оценки связей между признаками частотные таблицы (как еще иногда называют таблицы сопряженности, выделяя не сопряжение признаков в таблице, а представленность в ней статистических частот). Заметим, что последний термин обязан своим происхождением именно тому обстоятельству, что на основе анализа подобных таблиц можно судить о сопряженности (сочетаемости) каких-то значений одного признака с какими-то значениями другого признака. Как будет показано ниже, связь значений номинально измеренных признаков адекватно выражается подобным образом.

Предположим, что мы имеем два признака X и Y, первый из которых принимает m, а второй – n значений. Таблицей сопряженности (частотной таблицей) именуют матрицу (прямоугольную таблицу чисел), на пересечении i-ой строки и j-го столбца которой стоит число a_{ij}, означающее количество объектов, обладающих i-ым значением первого признака и j-ым значением второго. Таблица сопряженности приведена ниже (табл. 3.1).

Таблица 3.1

Таблица сопряженности
(матрица абсолютных частот)

a_{11}	a_{12}	...	a_{1n}
a_{21}	a_{22}	...	a_{2n}
...	...	...	...
a_{m1}	a_{m2}	...	a_{mn}

Таблицей сопряженности является любая таблица с исходными либо с аналитическими данными, используемая в практической аналитической работе. Единственным отличием может служить то, что у такой таблицы из практики всегда имеется итоговая строка и/или итоговый столбец, а, кроме того, может присутствовать строка и/или столбец типа «всего» (до которого или после которого следует «из них», «в том числе» и т. п.). Когда общая итоговая строка и общий итоговый столбец будут отброшены, а внутренний итог заменен на «прочие», механическое дублирование в такой таблице будет исключено, после чего она может уже рассматриваться в качестве таблицы сопряженности со всеми вытекающими из этого аналитическими последствиями.

С учетом того, что в ряде аналитических расчетов используются и определенные суммы частот, ее обычно представляют в ином виде: с явно обозначенными наименованиями признаков и их значений и выписанными маргинальными (краевыми) суммарными частотами[1]: крайний правый столбец представляет собой суммы по строкам, а крайняя нижняя строка – суммы по столбцам (это так называемые «маргинальные», т. е. *краевые* частоты). В ячейке правового нижнего угла указана величина объема выборки, она равна сумме сумм (маргинальных частот) – хоть по вертикали, хоть по горизонтали, это значения не имеет, результат тот же. Нередко используется расширительное понимание таблицы сопряженности с тем, чтобы в ней были представлены текстовые наименования обоих признаков и их значений, как в табл. 3.2.

Таблица 3.2

Таблица сопряженности
(с маргинальными частотами)

Признак *X* (наименование)	**Наименования значений признака *X***	**Признак *Y* (наименование)**				**Всего:**
		Наименования значений признака *Y*				
		1	2	...	$\boldsymbol{n}$	
	1	a_{11}	a_{12}	...	a_{1c}	$\sum_i n_{1i}$
	2	a_{21}	a_{22}	...	a_{2c}	$\sum_i n_{2i}$
	...	...	...	...	...	...
	$\boldsymbol{m}$	a_{r1}	a_{r2}	...	a_{rc}	$\sum_i a_{2i}$
	Всего:	$\sum_j a_{j1}$	$\sum_j a_{j2}$	...	$\sum_j a_{j1}$	$\sum_{i,j} a_{ij}$

[1] Ранее отброшенные строка и столбец играют лишь вспомогательную роль – упрощают расчеты, но элементами таблицы сопряженности не становятся.

Таблицу сопряженности значительно расширяют. Так, например, предполагается, что в качестве ее элементов могут фигурировать не только частоты (как в классике), но и многие другие величины. Так, например, в клетках половозрастной таблицы могут стоять в качестве дополнительных сведений средние значения зарплаты тех людей, которые характеризуются отвечающим клетке значениям признаков пола и возраста. Таким же образом в клетки таблицы могут быть помещены средние другого рода (мода, медиана), показатели неоднородности данных – дисперсии, величины отклонений от средних величин по строке (столбцу), разница между наблюдаемой (эмпирической) и теоретической частотой и т. д. Такое расширительное понимание таблицы сопряженности используется в пакете прикладных программ SPSS.

Элементами таблицы сопряженности служат целые числа, представляющие собой абсолютные частоты встречаемости сочетаний значений сопоставляемых (сопрягаемых) признаков, т. е. измеримых по номинальной шкале, низшей по уровню измерения. Но даже и для этой шкалы статистика предлагает целый ряд показателей, способных обнаружить в таблице факт наличия закономерности, скрытой от непосредственного видения ее человеком. В числе этих показателей есть так называемый «критерий *ХИ*-квадрат» (X^2), который с определенной высокой вероятностью правильного ответа отметит наличие в таблице сопряженности $m \times n$ некой закономерности, хотя и не укажет, в чем именно она заключается. Но даже и одно лишь указание критерия *ХИ*-квадрат на то, что имеющееся сочетание данных в таблице никак не может быть случайным, а потому требует выявления скрытых связей, дает мощный толчок к последующим исследованиям данных с целью выявления этих закономерностей на основе уже иных показателей связи. Обнаружение факта имеющейся, но пока и скрытой для нас закономерности внутри таблицы сопряженности представляет собой новую информацию, особенно ценную на фоне объективной относительной бедности информации, характерной для обработки данных, измеримых по номинальной шкале.

К упомянутому выше ряду показателей, применимых для измерения взаимосвязей в таблице сопряженности $m \times n$, принадлежат такие, как рассматриваемый ниже *коэффициент контингенции Ф* («*фи*»), а также *коэффициент ассоциации Q* (он же коэффициент Юла), *показатель Чупрова* (*T*) и др. Заранее сказать, какой из них более пригоден для предстоящего содержательного анализа, нельзя. В основе каждого показателя лежат свои особые предположения (требования) идеального характера относительно используемых данных (это требования математической модели) и то, к какой из таких моделей реалии окажутся ближе, покажет только проведенный статистический анализ.

Для анализа таблиц сопряженности применяются также коэффициенты связи, основанные на моделях прогноза: *коэффициент лямбда* (λ); *коэффициент тау* Гудмана-Крускала (τ); *коэффициент неопределенности*; коэффициенты, основанные на понятии «энтропии».

Статистическая практика предлагает простую и вместе с тем весьма плодотворную альтернативу *непосредственному* рассмотрению таблицы сопряженности $m \times n$. Альтернатива заключается в предварительном сведении (на определенных условиях) каждой такой таблицы (вне зависимости от количества имеющихся у нее строк и столбцов) к иной таблице сопряженности, уже предельно простой и компактной: к таблице сопряженности 2×2, которая и будет в последующем предметом анализа вместо исходной (такое сведение будем именовать редукцией). Важно отметить, что редукция возможна для любой таблицы сопряженности и, хотя при таком преобразовании происходит некоторая потеря информации, она, как правило, не принципиальная для рассматриваемой проблемы и, кроме того, не безвозвратная (к исходным данным всегда можно вернуться). Но зато для анализа такой редуцированной таблицы, предельно упрощенной по форме, статистическая наука предоставляет математический аппарат, обеспечивающий получение достаточно глубоких и научно обоснованных выводов, хорошо интерпретируемых на практике.

Таблица сопряженности 2×2 важна не только как результат редуцирования таблиц иного формата, она и крайне востребована в практике статистического анализа взаимосвязей, поскольку многие сопрягаемые признаки, именуемые *дихотомичными*, уже изначально имеют лишь по два значения, а потому и образуют именно таблицу 2×2. Сюда в первую очередь относятся признаки с *естественными* дихотомическими парами значений, такими как «да/нет», «мужской/женский», «городской/сельский», «удовлетворительный/неудовлетворительный», «начальник/подчиненный», «ответственный/безответственный», «плюс/минус», «виновный/невиновный» и т. д.

Таблицу сопряженности 2×2 и работу с ней продемонстрируем на конкретном условном примере, касающемся выявления по итогам года взаимосвязи проведенной профилактической работы с несовершеннолетними, состоящими на учете в ОВД, и реальным положением дел, касающихся возможного совершения ими правонарушений уже после завершения профилактической работы.

Здесь сопрягаются два признака – «Проводимость профилактической работы до начала отчетного года» с двумя значениями «Проводилась» и «Не проводилась» и «Совершаемость правонарушений несовершеннолетним в течение отчетного года» с двумя значениями – «Не совершил» и «Совершил».

Таблица 3.3

Совершаемость правонарушений в отчетном году лицами из числа несовершеннолетних, состоящих на учете в ОВД, в зависимости от проведения с ними профилактической работы

Несовершеннолетний правонарушение:	**В отношении несовершеннолетнего профилактическая работа:**		
	проводилась	**не проводилась**	Всего
не совершил	311	134	445
совершил	23	44	67
Всего	334	178	512

Приведем формулы для расчета каждого коэффициента. В общем виде решаемый пример в отношении состоящих на учете несовершеннолетних имеет вид, представленный в табл. 3.4.

Таблица 3.4

Таблица сопряженности для изучения зависимости совершаемости правонарушений лицами из числа несовершеннолетних, состоящих на учете в ОВД, от проведения с ними профилактической работы

Несовершеннолетний правонарушение:	**В отношении несовершеннолетнего профилактическая работа**		
	проводилась	**не проводилась**	Всего
не совершил	a	b	$a + b$
совершил	c	d	$c + d$
Всего	$a + c$	$b + d$	$a + b + c + d$

Целочисленные значения таблицы a, b, c и d именуются *абсолютными частотами*, представляют таблицу сопряженности, ее ядро, именно они заключают в себе суть таблицы, поскольку содержат новую, но пока не выявленную информацию о связи признаков. Суммарные значения, добавленные к таблице сопряженности $(a + c)$, $(b + d)$, $(a + b)$ и $(c + d)$ – *маргинальные* абсолютные частоты. Маргинальные частоты используются здесь только в техническом отношении (для упрощения расчета), но их присутствие не добавляет никакой новой информации.

В качестве показателя взаимосвязи между признаками используется коэффициент контингенции Φ, рассчитываемый по формуле:

$$\Phi = \frac{ad - bc}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}.$$

Коэффициент контингенции Φ относится к типу коэффициентов корреляции, а потому для этого показателя связи справедливо все ранее приведенное в отношении этого типа. Подставляя конкретные значения условного примера и учитывая, что значения $(a + c)$, $(b + d)$, $(a + b)$, $(c + d)$ специально подсчитывать не требуется, т. к. они уже имеются в виде состава маргинальных (краевых) частот, получим:

$$\Phi = \frac{311 \times 44 - 23 \times 134}{\sqrt{331 \times 178 \times 445 \times 67}},$$

$\Phi = 0{,}14012194$. Округленно $\Phi = 0{,}14$ (или 14 %).

Итак, первый вывод из полученного значения для Φ: связь между профилактикой и отказом несовершеннолетних от совершения (несовершением) правонарушений оказалась прямой. Этот же самый вывод можно сформулировать иначе: связь между профилактикой и совершением преступлений оказывается обратной – на что и рассчитывали, затевая данное мероприятие. То, что профилактика дает положительный эффект, хорошо, однако остается вопрос, достаточен ли подобный эффект, чтобы счесть оправданными затраченные для его достижения силы и средства, т. е. достаточен ли уровень эффективности, означающей эффект, рассматриваемый в неразрывной связи с затратами всех видов использованных ресурсов. Итак, как правильно оценивать полученную величину в отношении того, большая она или малая. Общая теория статистики универсальных пороговых значений для показателя связи не может указать в принципе, поскольку таковых не существует для рассматриваемого уровня измерений. Этот вопрос должен решаться в рамках конкретных областей деятельности (такой является, например, криминология), для чего требуется уже учет специфики каждой из областей. При решении такого вопроса в рамках отраслевой (предметной) статистики специалист в соответствующей области на основе своего опыта и своей интуиции определяет для конкретного круга явлений, относящихся к конкретной, исследуемой им области, характерные для нее пороговые значения показателя связи между показателями явлений. Нередко при решении этого вопроса аналитики обращаются к результатам специалистов, представляющих смежные области знания, полагаясь на аналогию. Вот почему, например, аналитик-криминолог, если он не имел возможности определить пороги в своей области самостоятельно, вынужден руководствоваться порогами, используемыми в практике работы социологов, психологов и иных социальных исследователей. Там принято считать, что связь считается *слабой* (фактически отсутствующей), если она без учета знака (плюса либо минуса) меньше 0,3 (т. е. 30 %), *средней* – если от 0,3 до 0,7 (от 30 % до 70 %), *сильной* – если больше 0,7 (70 %).

Существует ряд показателей, относящихся к типу коэффициентов корреляции, и при всех различиях между ними они одинаково отвечают на вопрос, прямая ли связь между показателями явлений или обратная. Однако при решении вопроса о силе связи (о ее наличии, если значение одного из показателей связи окажется вблизи нуля, в точности с ним не совпадая) разные показатели связи могут по-разному оценивать одну и ту же ситуацию. Это указывает не на возможную неправильность некоторых из них, а на присущую каждому из них индивидуальность, обусловленную теми особенностями в допущениях, гипотезах, которые лежат в основе каждого из показателей, отличая его от прочих. Например, рассматриваемый показатель контингенции Φ относится к числу весьма строгих показателей (почему и был рекомендован для использования на практике), в то время как другой – коэффициент ассоциации Q:

$$Q = \frac{ad - bc}{ad + bc} = \frac{311 \times 44 - 23 \times 134}{311 \times 44 + 23 \times 134} = 0{,}632$$

при тех же данных оценивает силу связи гораздо выше (он равен 0,63). Поэтому считать, что 0,14 находится в нулевой зоне, в данном случае неправомерно, и для этого показателя в конкретной области применения (например, для криминологии и регионального уровня) надо нарабатывать присущие именно этому показателю пороговые значения.

На установление пороговых значений влияют и цели решаемой задачи: для целей профилактической работы в отношении нескольких сотен подростков пороги будут одни, а для строительства атомной электростанции – уже совсем иные, гораздо более строгие. В частности, в случае АЭС значение 0,1 явно не может быть проигнорировано, приравнено к нулевому.

Если в таблице рассмотренного примера переставить между собой строки (как на приведенной ниже табл. 3.5), то абсолютное значение Φ (без учета знака) останется неизменным, а сам знак поменяется на противоположный, т. е. связь станет обратной (отрицательной).

Таблица 3.5

Совершаемость правонарушений в отчетном году лицами из числа несовершеннолетних, состоящих на учете в ОВД, в зависимости от проведения с ними профилактической работы

Несовершеннолетний правонарушение:	**В отношении несовершеннолетнего профилактическая работа:**		
	проводилась	**не проводилась**	Всего
совершил	23	44	67
не совершил	311	134	445
Всего	334	178	512

Вернемся к вопросу сведения (редуцирования) произвольной таблицы сопряженности $m \times n$ к таблице 2×2. Возможность редуцирования основана на том, что любой качественный показатель с любым числом значений (именуемых в этом случае «атрибутами» или «категориями») может быть сведен к тому же показателю, но уже лишь с двумя значениями (атрибутами), т. е. к дихотомическому. Другими словами, редукция таблицы сопряженности основана на редукции сопрягаемых в ней значений каждого из сопрягаемых показателей. Дихотомия – это всегда принципиальное противопоставление внутри пары значений показателя: «да/нет», «белый/черный», «положительный/отрицательный», «виновный/невиновный», «активный/пассивный» и т. п. Чтобы обеспечить это, необходимо вначале в подлежащем редуцированию показателе из всех имеющихся у него значений выбрать лишь одно, но притом первостепенное, главное в данном исследовании, интересующее аналитика, в котором и заключена суть решаемой проблемы. Этому первому значению противопоставляются все прочие значения, являющиеся в этот момент уже лишь второстепенными, третьестепенными и т. д. Вот под этим именем «Прочие» («Иные», «Другие» и т. п.) и выступает второе значение в дихотомизированном показателе. Таким образом, в качестве первого значения отбирается готовое значение «*А*» из числа уже имеющихся как представляющее первостепенный интерес, а дихотомической парой ему будет служить значение «*не-А*», дополнительное по отношению к «*А*». В жесткие границы дихотомии «втискивается» исходный подлежащий редуцированию показатель исследуемого явления с неизбежными, пусть и с существенными, но все же не принципиальными, второстепенными потерями. После этого к преобразованному (дихотомизированному) показателю уже можно применять весь арсенал методов, разработанных специально для таблиц сопряженности 2 × 2. Но можно предположить, что после этого аналитика заинтересует и какое-то еще значение исходного показателя в качестве первостепенного. В этом случае действия аналитика будут полностью аналогичными выше описанным: он изберет заинтересовавшее его значение в качестве первого из двух возможных значений редуцированного показателя (в качестве «*А*»), а все остальные в совокупности (включая и ранее выделявшееся значение) – в виде второго дихотомического значения – «*не-А*».

Влияние цели проводимого анализа на дихотомизацию многозначного показателя, вычленения основного для проводимого анализа значения покажем на следующем примере. Пусть в ходе пресечения массовых беспорядков были задержаны несколько десятков человек, в числе которых были и мужчины, и женщины. В ходе разбирательства (например, в суде) эти ранее задержанные лица были подразделены на три класса: совершившие уголовные преступления, совершившие административные правонарушения и правонарушений не совершившие. При анализе взаимосвязи между степенью общественной опасности совершенных деяний и полом задержанных будет составлена таблица сопряженности 3 × 2. В этой исходной таблице один из показателей (пол) дихотомический уже по своей природе, а потому принудительной дихотомизации должен быть подвергнут другой качественный показатель – противоправность совершаемых деяний. Исходными значениями для него как раз и будут три вышеназванные значения, характеризующие степень противоправности их деяний. Исходя из цели исследования, определяется первостепенное значение. Если целью является исследование криминальной части среди задержанных, то понятно, что первостепенное значение приобретает именно совершение уголовных деяний, которому в качестве второго значения противопоставляется совершение всех прочих деяний. Поэтому в контексте проводимого исследования уже не будет важным то обстоятельство, что одна часть объединяемых во второе значение показателя вела себя противоправно, а другая часть была законопослушной. Важно для второго значения лишь то, что тех и других объединяет факт, что они не отнесены к преступникам. В этом смысле непреступные деяния (в число которых вошли и административные правонарушения) противостоят деяниям преступным. Однако если целью станет исследование любого противоправного поведения как антипода поведению законопослушному, то законопослушное поведение образует одно значение редуцированного показателя, а противоправное – другое. Наконец, если целью исследования являются именно лица, совершившие административные правонарушения, тогда противостоять этой категории лиц будут все прочие, что в данном случае означает объединение (разумеется, умозрительное) преступников и законопослушных граждан, хотя такое чисто логическое объединение и представляется на первый взгляд внутренне противоречивым.

3.4. Измерение взаимосвязи для данных, измеримых по порядковой (ранговой) шкале

Начнем рассмотрение вопроса с простого примера. В определенной оперативной службе отделы каким-то образом распределились по результатам боевой и физической подготовки, результатам своей профильной деятельности. Должен ли номер места отдела в первом распределении как-то сказываться на номере его места во втором? Должна ли существовать между ними взаимосвязь или это разные сущности, друг на друга не влияющие?

Видимо, факт введения подобного рода подготовки уже предполагает, что чем выше уровень такой подготовки, тем выше должны быть и конечные результаты профильной деятельности. Рассматривая данный вопрос с другой стороны, можно сказать, что в идеале должно быть так, что отдел, занимающий N-ое место по боевой и физической (профессиональной, психологической и т. п.) подготовке, имеет то же самое (N-е) место и по конечным результатам. С точки зрения практики было бы вполне достаточно, чтобы подобное соответствие номеров мест соблюдалось, пусть не во всех случаях, но, по крайней мере, в большинстве. Статистика располагает показателями, измеряющими взаимосвязь (соответствие) номеров мест в указанных распределениях, что дает ответ на поставленный вопрос.

Итак, если не видеть за боевой и физической подготовкой связи с достижением требуемых конечных результатов, то зачем отрывать людей от того, ради чего их общество и призвало в ряды полиции? К чему проведение лыжных кроссов, стрельб в тире, овладение криминалистикой и т. д.? Рассматриваемые в отрыве от конечных результатов все названные виды подготовок сотрудников полиции особого смысла не имеют. Логика учредителей указанной подготовки здесь очевидна: лучше стреляет – поразит противника, лучше бегает – догонит правонарушителя. Но эта логика основывается лишь на предположении, которое в реальности может оказаться верным вполне или верным отчасти, а то и вовсе неверным. Вводимая из самых благих побуждений подготовка сотрудников по какому-то иностранному языку (например, венгерскому) в связи с проведением в будущем международного спортивного мероприятия способна в определенных случаях оказать пагубное влияние на занятия иными видами подготовки.

Общество от полиции и входящих в нее подразделений в большинстве своем ждет только тех результатов, которые полицией именуются «конечными» (в отличие от других результатов – «промежуточных», отражающих только сугубо внутренние цели и задачи полиции). В статистике такие промежуточные результаты, достижение которых лишь предположительно способствует в той или иной мере достижению конечных результатов, именуются *факторами* (т. е. воздействующими обстоятельствами), а «конечные» результаты так и именуются просто *результатами*. Таким образом, под факторами понимаются причины и условия (объективного и субъективного характера), которые согласно теоретическим предположениям (подлежащим последующей проверке) влияют или способны порой влиять на результат.

Но как проверить такое предположение о влиянии каждого из факторов в реальной служебной практике? Ответить на этот вопрос, как правило, затруднительно, поскольку какие-то номера мест у отделов в сопоставляемых распределениях и в самом деле будут совпадать (подтверждая предположение, находящееся в основе введения указанной подготовки), в то время как другие будут резко различаться. Сделать же вывод в целом, полагаясь на свое внутреннее чутье, суметь уловить тенденцию, не впадая при этом в субъективизм, успешно действовать без привлечения положений и методов статистической науки чаще всего оказывается невозможным. Как же в условиях реальной противоречивости данных о совпадении и несовпадении занимаемых мест сделать итоговый общий вывод относительно пользы (или, наоборот, бесполезности, а то и об ощутимом ущербе) от наличия соответствующего вида подготовки? Номера мест – признаки, измеряемые по порядковой (ранговой) шкале, поэтому и ответ на подобные вопросы следует искать, обращаясь к тому или иному показателю взаимосвязи, использующему данные, измеримые именно по этой шкале.

К показателям тесноты (силы) парной связи между показателями, измеряемыми по такой качественной шкале, как порядковая, относится предложенный в начале XX в. Спирменом *коэффициент корреляции рангов* (*ранговый коэффициент* корреляции).

Следует напомнить, что *ранг* – это показатель, означающий порядковый номер единицы совокупности (объекта) в ранжированном по возрастанию ряду. Если проранжировать совокупность отдельно по каждому из двух признаков, связь между которыми изучается, то полное совпадение у признаков рангов означает максимально тесную (функциональную) прямую связь, а полная противоположность рангов (их инверсия) – максимально тесную (функциональную) обратную связь. Но такие предельные, максималистские случаи в реальной практике, если и встречаются, то крайне редко.

Чаще же сильной связью считается только приближение к максимальному значению, но не достижение его, поскольку всегда находится какое-нибудь нарушение в соответствии рангов. Аналогичным образом на практике крайне редко встречается и другой предельный случай – полное отсутствие связи, которому соответствует нулевое значение показателя взаимосвязи. Но даже и при отсутствии какой-либо связи, присутствующая в рассчитанном значении показателя связи случайная погрешность делает это значение уже ненулевым, хотя и весьма малым. Именно поэтому к идеальному нулевому значению приравниваются и все реальные иные (положительные и отрицательные) значения, мало отличающиеся от нуля. «Малость» такого отличия в данном случае определяется в итоге уже аналитиком, исходящим из цели и условий решаемой им задачи.

Ранжирование объектов совокупности по соответствующему признаку в статистике производят, как правило, по возрастанию его значений, именно такой ряд значений объектов носит наименование *вариационного ряда*. Вариационный ряд, ранжированный по одному признаку, вовсе не обязан быть ранжированным и по другому. Можно провести два независимых ранжирования по двум признакам и рассматривать далее в каждом из полученных вариационных рядов уже не значения признаков объектов, а лишь их ранги. Сопоставив между собой ранги двух признаков одного и того же объекта и учтя затем таким же образом все имеющиеся объекты совокупности, можно определить значение показателя взаимосвязи. На основе этого значения делаются все выводы о характере интересующих явлений.

Итак, один из пары сопоставляемых признаков – это фактор, т. е. показатель того явления, которое не является предметом ведения руководителя. Руководитель вообще не должен нести за факторный показатель какой-либо ответственности (либо потому, что это дело вне его компетенции, либо потому, что это и в принципе невозможно, если речь идет, например, о непредсказуемых природных явлениях). Однако, исследуя сугубо свою проблему, руководитель, тем не менее, наталкивается на тот или иной фактор как влияющий на результат, за который спрашивают с руководителя. Вот почему он вынужден включать такой фактор (один или несколько) в круг рассматриваемых им показателей, если предварительная теоретическая качественная модель приводит его к выводу, что данный фактор может оказывать воздействие на результат. Второй из пары сопоставляемых показателей это уже результат, за который руководитель несет ответственность и который поэтому представляет для него непосредственный интерес.

Сопоставив между собой ранги фактора и ранги результата для одних и тех же объектов, можно установить, имеется ли взаимосвязь между признаками, а если имеется, то какова она по направленности и силе.

Если в совокупности у объекта i-ый ранг по признаку X обозначить как P_x^i, а i-ый ранг по признаку Y обозначить как P_y^i, то коэффициент корреляции рангов, рассчитанный по *формуле Спирмена*, имеет вид:

$$R_s = 1 - \frac{6 \times \sum \left(p_x^i - p_y^i\right)^2}{(n-1)n(n+1)} = 1 - \frac{6 \times \sum \left(p_x^i - p_y^i\right)^2}{n\left(n^2 - 1\right)},$$

где n – число объектов.

Преимущественная особенность ранговых коэффициентов корреляции (включая и такой их вид, как коэффициент Спирмена) заключается в том, что они могут рассчитываться для большого числа ситуаций, встречающихся на практике. Так только ранговый коэффициент может быть рассчитан, когда упорядочением признаков будет только их ранжирование (по принципу «больше/меньше», «хуже/лучше», «активнее/пассивнее» и т. п.). Но для сравниваемых значений нельзя (основываясь на имеющейся информации) сказать, на сколько же именно больше, (лучше, активнее и т. п.) одно значение другого. Так, комиссией могут быть проранжированы кандидаты на занятие определенной должности по их профессиональному уровню, хотя на вопрос о том, насколько один профессиональнее другого, комиссия не будет в состоянии ответить. В отношении такого рода данных уместны вычисления рангового коэффициента корреляции, а вот другие методы сопоставления данных здесь не применимы.

Среди значений одного признака (что X, что Y) не часто, но порой встречаются два, три и т. д. одинаковых значения (в том числе и с точностью до округления данных). Например, равенство ОВД по нагрузке на сотрудника, степени удовлетворения работой, активности на учебных занятиях, гибкости организационных структур и т. п. В таком случае естественно присвоить равным значениям и некий единый, равный для них всех ранг. В качестве такого ранга принимается усреднение из тех рангов, которые этим значениям были бы присвоены, отличайся они один от другого хоть на малость, после чего они становились бы формально неравными. Такой усредненный ранг не обязательно будет целым числом, как все прочие ранги. Например, двум одинаковым значениям признака, претендующим в соответствии с порядком расположения на третий и четвертый ранги, присваивается усредненное для них значение 3,5, т. е. сумма этих двух рангов, равная 7, сохранится. В результате в конечном наборе рангов будут одновременно присутствовать сразу два ранга, равных 3,5 (и при этом рангов 3 и 4 уже не будет). Если имеются не два, а три совпадающих значения признака (претендующие, предположим, на ранги 6, 7 и 8), то общий (усредненный) ранг для всех равных значений признака будет равен 7 (и при этом будут отсутствовать уже ранги 6 и 8) и т. д. Далее эти ранги, наряду с прочими, используются в приведенной выше формуле для расчета коэффициента ранговой корреляции.

На примере с условными данными рассмотрим использование коэффициента корреляции Спирмена (табл. 3.6).

Таблица 3.6

Сведения о числе неполных семей и количестве правонарушений несовершеннолетних (по микрорайонам города)

№ микрорайона	1	2	3	4	5	6	7	8	9	10
Число неполных семей	213	111	98	137	241	222	276	189	217	195
Количество правонарушений несовершеннолетних	35	23	23	31	38	29	34	19	17	53

В действительности для органов внутренних дел непосредственный интерес должен представлять только второй показатель – «Количество правонарушений несовершеннолетних», потому такой показатель именуется *результативным* (или *результатом*), обозначается обычно буквой Y. Но в процессе содержательного анализа, предшествовавшего статистической обработке данных, аналитик ОВД предполагает, что указанное количество в определенной мере зависит (или хотя бы может зависеть) в числе прочих и от такого показателя, как «Число неполных семей». Тогда в силу этого обстоятельства аналитика ОВД уже начинает интересовать и этот показатель причины или условия. Показатели причин и показатели условий объединяются обобщающим понятием *факторного показателя* (фактора), обозначаемого обычно буквой X. Располагается факторный показатель в таблице раньше результативного в качестве подчеркивания предполагаемой (возможной) причинно-следственной связи.

В то время, как результат (Y) в исследуемой теоретической модели всегда один, факторов, напротив, может быть сразу несколько. В этом случае для факторов, кроме X, используются и другие буквенные обозначения (например, Z, U, V и т. п.). Тем не менее статистическое исследование взаимосвязи осуществляется пошагово: в таблице с результатом всегда сопоставляется только один фактор. Затем поочередно перебираются по одному и все прочие факторы (рассматриваются пары X-Y, Z-Y, U-Y, V-Y и т. д.), после чего на основе ряда частных выводов, определяемых соответствующим фактором, делается общий вывод, относящийся к теоретической модели в целом.

Если говорить не о технической информации (о порядковом номере объекта), а о содержательной информации, то факторный ряд и результативный ряд данных располагаются в таблице параллельно, и по этой причине анализ взаимосвязи между этими рядами получил в статистике наименование *анализа параллельных рядов*.

Рассчитаем ранговый коэффициент корреляции Спирмена для приведенного выше числового примера. Для этого составим на основе данных исходной таблицы (табл. 3.6) расчетную таблицу (табл. 3.7).

Таблица 3.7

Последовательность расчета рангового коэффициента корреляции

***i*-й номер объекта (здесь: микрорайона)**	**Ранг объекта по фактору (по *X*) (Px_i)**	**Ранг объекта по результату (по *Y*) (Py_i)**	**Разность рангов ($Px_i - Py_i$)**	**Квадрат разности рангов $(Px_i - Py_i)^2$**
1	6	8	–2	4
2	2	3,5	–1,5	2,25
3	1	3.5	–2,5	6,25
4	3	6	–3	9
5	9	9	0	0
6	8	5	3	9
7	10	7	3	9
8	4	2	2	4
9	7	1	6	36
10	5	10	–5	25
Сумма:	55	55	0	104,5

Итак, $R_s = \dfrac{1-6\times\left(4+2{,}25+6{,}25+9+0+9+9+4+36+25\right)}{\left(10-1\right)\times 10\times\left(10+1\right)} = 0{,}3667.$

Таким образом, полученное значение коэффициента указывает на то, что в отношении факторного и результирующего показателей имеется:

– прямая статистическая связь;

– связь достаточно сильная (особенно учитывая, что это всего лишь один из множества факторов, существенных для такого явления, как совершение правонарушений несовершеннолетними).

На основе вычисления коэффициента детерминации $\eta^2\left(=R_s^2\right)$ выясняется, в какой мере изменения (вариация) факторного показателя определяют изменения (вариацию) результативного показателя: $\eta^2 = (0{,}3667)^2 = 0{,}1345$, т. е. результат данным фактором определяется на 13,5 %, что достаточно много.

Общий вывод состоит в том, что наличие прямой связи между количеством неполных семей в микрорайоне и количеством правонарушений несовершеннолетних можно считать установленным; притом эта связь считается достаточно сильной, особенно с учетом того, что такой фактор неединственный. То, что такая связь причинно-следственная, это, как уже отмечалось, вытекает не из общей теории статистики (математической статистики), а правовой статистики как статистики предметной, рассматривающей уже и содержательную сторону модели связи признаков.

Если речь идет об особо важном явлении, предположим, об исследовании организованной преступности на этнической основе (описываемой соответствующим результативным показателем) в зависимости от численности соответствующей этнической диаспоры (что описывается факторным показателем), то уже и малые, ранее относимые к нулю значения должны приниматься во внимание. Исследование в этом случае необходимо продолжить до получения более определенных выводов.

В отношении показателя η^2 как меры вклада конкретного фактора в изменения результата верным будет утверждение, что на долю прочих факторов, не зависящих от рассмотренного, приходится остальной вклад – $(1 - \eta^2)$. Для рассмотренного примера на долю прочих независящих факторов должны приходиться остающиеся 86,5 % воздействия на результат. Однако просто суммировать подобные доли нельзя, предварительно не убедившись в том, что эти факторы статистически независимы между собой (корреляция близка к нулю), а подобрать такие факторы, как правило, непросто. Так, если в рассмотренном примере выбрать вторым фактором «Число неблагополучных семей», то этот новый фактор, несомненно, отличаясь по смыслу от первого, будет, тем не менее, достаточно сильно с ним статистически связан, коррелирован. Действительно, с одной стороны, многие неблагополучные семьи являются одновременно и неполными, а, с другой стороны, многие неполные семьи являются одновременно и неблагополучными. Итак, совпадения показателей нет, а значительная коррелированность очевидна. Неблагополучие в семье тогда следует выражать другим показателем, по возможности резко отличающимся от фактора неполной семьи.

Чем же корреляция между факторами может помешать, почему требуется их статистическая независимость? Дело в том, что можно найти степень влияния и второго фактора точно так же, как она определялась для первого фактора, но суммировать значения коэффициентов детерминации для них с целью рассчитать их совместное влияние на результат уже нельзя, поскольку общее влияние зависимых показателей гораздо меньше суммы значений коэффициента детерминации для каждого в отдельности. Из этого следует, что подбор аналитиком независимых факторов для достаточно полного объяснения изменений результата – это самостоятельная и достаточно сложная проблема, в данном случае – это проблема правовой (отраслевой, предметной) статистики. Однако общая статистика окажет отраслевой такую помощь, которая способна отбраковать предлагаемые к рассмотрению факторы по мере обнаружения зависимости между ними и уже используемыми факторами.

Ранговые коэффициенты корреляции (и Спирмена, и иные возможные, например, Кендалла) имеют то достоинство, что могут работать с признаками качественными, допускающими упорядочивание (по принципу «больше/меньше»). Если же данные соответствуют количественной шкале (как в рассмотренном выше числовом примере), то хотя использование коэффициента ранговой корреляции в этом случае достаточно продуктивно, однако надо ясно осознавать, что такой переход к порядковой шкале с ее рангами означает определенную потерю информации, иногда значимой для целей исследования. В рассмотренном примере данные таблицы измеряются по абсолютной шкале, и потому к ним могут применяться также и те методы, которые ранее нельзя было бы применить к данным, измеримым по порядковой (ранговой) шкале.

В программном средстве SPSS применяются такие меры связи для порядковых переменных, как коэффициенты: гамма, d Сомерса, τ_b Кендалла, τ_c Кендалла. Все эти коэффициенты основаны на учете количества нарушений согласованности в рангах по изучаемым признакам для выбранной пары объектов (наблюдений). Например, согласованность в табл. 3.8, где представлены данные как о доле сотрудников территориальных органов внутренних дел со стажем работы в полиции свыше 15 лет, так и об экспертной оценке деятельности данного органа.

Инверсией для пары изучаемых объектов (ОВД) называется нарушение согласованности (порядка) следования рангов по двум показателям. Так, для пары объектов (1 и 2) имеет место инверсия, поскольку для первого показателя ранг первого объекта больше ранга второго, а для второго показателя, наоборот, ранг первого объекта меньше ранга второго. Если такая согласованность сохраняется в паре, говорят о *проверсии*.

Таблица 3.8

Зависимость экспертной оценки деятельности органа внутренних дел от доли сотрудников с большим стажем службы

№ ОВД	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Экспертная оценка деятельности ОВД, баллы	6	2	1	3	9	8	10	4	7	5
Ранг по оценке	6	2	1	3	9	8	10	4	7	5
Доля сотрудников ОВД со стажем свыше 15 лет, %	32	37	43	29	17	18	33	35	28	38
Ранг по стажу	5	8	10	3	1	2	6	7	4	9

Например, в паре (1 и 4) порядок следования рангов сохраняется для двух показателей, следовательно, имеет место проверсия.

Коэффициент гамма (γ) вычисляется по следующей формуле:

$$\gamma = \frac{P - I}{P + I},$$

где P – общее количество проверсий во всех парах, I – общее количество инверсий во всех парах.

Если инверсий не наблюдается (т. е. $I = 0$, имеем полную прямую связь ($\gamma = 1$)). Если же вовсе не встречается проверсий (т. е. $P = 0$), то говорят о полной обратной связи ($\gamma = -1$). Если количество инверсий и проверсий равно, то в этом случае ($\gamma = 0$) нет выраженной связи признаков – ни прямой, ни обратной. Заметим, что $P = I$ равно количеству всех возможных пар, составленных из n объектов (сочетаний из n по 2), т. е.

$$P + I = \frac{n(n+1)}{2}.$$

В случае, если ранжируемые объекты по определенному признаку одинаковы, утверждают, что такие объекты являются *связанными*. Например, если исследователь связывает третий и четвертый объекты в ранжируемом ряду, то каждому объекту приписывается ранг, равный 3,5, но если бы это были третий, четвертый и пятый, каждый из трех объектов получил бы ранг, равный 4. Данный метод основывается на следующем рассуждении: если объекты по определенному рассматриваемому в данном случае признаку неразличимы, следовательно, они должны рассматриваться как один объект с одним и тем же рангом, при этом сумма рангов не должна меняться.

Наличие связанных рангов имеет определенное влияние на расчет коэффициента гамма. Произвольной паре объектов при расчете коэффициента присваивается значение 1 или –1 в зависимости от согласованности следования рангов по двум признакам. В случае связанных объектов следует данной паре приписать нулевое значение.

Коэффициент d Сомера учитывает связанные ранги. Для его вычисления используют формулу для γ с учетом количества связанных рангов:

$$d = \frac{P - I}{P + I - N_1},$$

где N_1 – количество всех пар, образуемых связанными объектами по первому признаку.

Аналогично рассчитывается мера связанности (коэффициент) d Сомера, где используются корректирующий член, соответствующий количеству пар, образуемых связанными объектами по второму признаку. Указанные коэффициенты являются ассиметричными мерами связанности. Симметричный коэффициент d Сомера в знаменателе имеет среднее значение знаменателей ассиметричных коэффициентов.

Коэффициент τ_b Кендалла одновременно учитывает связанные ранги по первому и второму признаку:

$$\tau_b = \frac{P - I}{\sqrt{P + I - N_1}\sqrt{P + I - N_2}}.$$

Если порядковые признаки принимают ограниченное количество значений, для оценки связи между ними можно воспользоваться таблицей сопряженности, разместив значения первого признака и второго признака одновременно в порядке убывания или возрастания (табл. 3.9).

Для данных, представленных в таблице сопряженности, применяют коэффициент τ_c Кендалла:

$$\tau_c = \frac{2m(P - I)}{N^2(m - 1)},$$

где N – общая сумма частот, m – наименьшее из числа строк и столбцов.

Этот коэффициент может достигать значений ±1 в любых таблицах.

Таблица 3.9

Стаж сотрудника органа внутренних дел и его квалификационное звание

	Без звания	3	2	1	Мастер	Наставник	Σ
0–5 лет	10	2	0	0	0	0	12
5–10 лет	8	12	2	0	0	0	22
10–15 лет	1	10	6	0	0	0	17
15–20 лет	0	3	4	2	0	0	9
свыше 20 лет	1	5	1	0	0	0	7
Σ	20	32	13	2	0	0	67

ЗАКЛЮЧЕНИЕ

В учебно-методическом пособии рассмотрен категориальный аппарат статистического анализа, необходимый для понимания и интерпретации использования математического аппарата в сфере познания правовых явлений. Особое внимание уделено описанию предметного поля рассматриваемой дисциплины, показаны особенности объекта статистического анализа в правовой сфере, изложены основные методы дескриптивной и аналитической статистики.

Необходимо отметить, что современный статистический анализ оперирует данными, имеющими как качественное, так и количественное измерение, поэтому в данном пособии разъяснены основы теории измерения, включая понятие измерительной шкалы и типов шкал: номинальной, ранговой, интервальной и абсолютной. Примеры шкал, приводимые в пособии, имеют прямое отношение к деятельности органов внутренних дел, что позволяет читателю ориентироваться при проведении собственного исследования правовых проблем.

Для любой научной отрасли (научного исследования) важно не только определить предмет исследования, описать его, собрать эмпирические данные, но показать изучаемое явление как закономерное, обусловленное определенными факторами, объяснить рассматриваемое явление, т. е. вывить закономерности его становления, функционирования и развития. Только понимание закономерного характера исследуемого явления (процесса) позволяет ученому делать научный прогноз, а практику осуществлять управление в соответствующей сфере. В этой связи в пособии подробно рассмотрены проблемы количественного оценивания тесноты связи явлений для наиболее используемых в эмпирических исследованиях данных – номинальных и порядковых.

СПИСОК ЛИТЕРАТУРЫ

Конституция Российской Федерации. Герб. Гимн. Флаг. – М. : Эксмо, 2016. – 64 с.

О полиции : федер. закон от 7 февраля 2011 г. № 3-ФЗ // Собр. законодательства Российской Федерации. 2011. № 7. Ст. 900.

Балдин К. В., Рукосуев А. В. Общая теория статистики : учеб. пособие / К. В. Балдин, А. В. Рукосуев. – М. : Дашков и К, 2015. – 312 с.

Забрянский Г. И. Методика статистического изучения преступности (введение в криминологическую статистику) : учеб. пособие / Г. И. Забрянский. – М. : Граница, Российская Академия адвокатуры и нотариата, 2010. – 128 с.

Кетле А. Социальная система и законы ею управляющие / пер. Л. Н. Шаховской. – М. : Н. Поляков и К, 1866. – 313 с.

Коробов В. Б., Ребрий В. А., Васильев Д. В. Проведение прикладных социальных эмпирических исследований, связанных с деятельностью органов внутренних дел : учеб.-метод. пособие / В. Б. Коробов [и др.]. – М. : Академия управления МВД России, 2013. – 75 с.

Плотинский Ю. М. Модели социальных процессов : учеб. пособие / Ю. М. Плотинский. – М. : Логос, 2009. – 296 с.

Рабочая книга социолога : монография / [А. М. Яковлев и др.] ; под ред. Г. В. Осипова. – 5-е изд. – М. : УРСС, 2009. – 477 с.

Ребрий В. А., Васильев Д. В. Основы сбора и аналитической обработки статистических данных, относящихся к деятельности ОВД (на основе SPSS) : метод. пособие / В. А. Ребрий, Д. В. Васильев. – М. : Академия управления МВД России, 2014. – 108 с.

Тарасова Т. Н., Давыдова Н. Ю. Правовая статистика : учеб. пособие / Т. Н. Тарасова, Н. Ю. Давыдова. – Оренбург : Оренбургский государственный университет, ЭБС АСВ, 2016. – 144 с.

Теоретические основы исследования и анализа латентной преступности : монография / [С. М. Иншаков и др.] ; под ред. С. М. Иншакова. – М. : ЮНИТИ-ДАНА, 2011. – 839 с.

Тюрин Ю. Н., Макаров А. А. Анализ данных на компьютере [Электронный ресурс] // Математическая книга. – М. : МЦНМО, 2014. – 367 с. – Режим доступа: http://маткнига.рф/wp-content/uploads/2017/02/978-5-4439-3011-4-Tyurin-Makarov-Analiz-dannyh.pdf

Для заметок

Учебное издание

Васильев Дмитрий Владимирович
Ребрий Владимир Александрович

Статистическое исследование правовых явлений и процессов

Учебно-методическое пособие

Редактор: *К. В. Громова*
Верстка: *Д. А. Бурцев*

Подписано в печать 29.03.2018. Формат 60 × 84 $^1/_{16}$.
Усл. печ. л. 4,9. Уч.-изд. л. 4,1. Тираж 68 экз. Заказ № ______

Отделение полиграфической и оперативной печати РИО
Академии управления МВД России.
125993, Москва, ул. Зои и Александра Космодемьянских, д. 8

ISBN 978-5-906942-40-1

9 785906 942401