ВЫСШАЯ ШКОЛА ЭКОНОМИКИ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

АНАЛИЗ ДАННЫХ

УЧЕБНИК ДЛЯ АКАДЕМИЧЕСКОГО БАКАЛАВРИАТА

Под редакцией профессора **В. С. Мхитаряна**

Рекомендовано Учебно–методическим отделом высшего образования в качестве учебника для студентов высших учебных заведений, обучающихся по экономическим направлениям и специальностям

Книга доступна в электронной библиотечной системе biblio-online.ru

Москва ▪ Юрайт ▪ 2016

УДК 519.2(075.8)
ББК 22.172я73
А64

***Ответственный редактор*:**

Мхитарян Владимир Сергеевич — доктор экономических наук, профессор, руководитель департамента статистики и анализа данных факультета экономических наук Национального исследовательского университета «Высшая школа экономики».

***Рецензенты*:**

Сажин Ю. В. — доктор экономических наук, профессор, заведующий кафедрой статистики, эконометрики и информационных технологий в управлении Национального исследовательского Мордовского государственного университета имени Н. П. Огарева;

Кремер Н. Ш. — профессор кафедры «Математика 1» Финансового университета при Правительстве Российской Федерации.

Анализ данных : учебник для академического бакалавриата / под ред. В. С. Мхитаряна. — М. : Издательство Юрайт, 2016. — 490 с. — Серия : Бакалавр. Академический курс.

А64

ISBN 978-5-9916-5591-0

Рассмотрены основные методы систематизации, обработки и анализа статистических данных, включающие описательные методы анализа данных, особенности и основные понятия вероятностно-статистического подхода к анализу данных. Для анализа многомерных данных представлены методы выявления и оценки степени зависимости между переменными, построения регрессионных моделей для определения вида статистической зависимости между переменными, методы снижения размерности признакового пространства и многомерной классификации данных. Также рассмотрены вопросы устойчивого, робастного оценивания параметров и непараметрического моделирования, анализа временны́х данных и прогнозирования.

В учебнике разбирается большое количество примеров анализа данных, приводятся задания для самостоятельной работы студентов.

Соответствует актуальным требованиям Федерального государственного образовательного стандарта высшего образования.

Для бакалавров экономических направлений вузов.

УДК 519.2(075.8)
ББК 22.172я73

ISBN 978-5-9916-5591-0

Оглавление

Авторский коллектив

Мхитарян Владимир Сергеевич — профессор, доктор экономических наук, руководитель департамента статистики и анализа данных факультета экономических наук Национального исследовательского университета «Высшая школа экономики» (предисловие, приложение, гл. 2, 4);

Архипова Марина Юрьевна — доктор экономических наук, профессор департамента статистики и анализа данных факультета экономических наук Национального исследовательского университета «Высшая школа экономики», ведущий научный сотрудник Института проблем управления имени В. А. Трапезникова Российской академии наук (гл. 3, 6);

Дуброва Татьяна Абрамовна — доктор экономических наук, профессор, заведующая кафедрой математической статистики и эконометрики Российского экономического университета имени Г. В. Плеханова (гл. 8);

Миронкина Юлия Николаевна — доцент, кандидат технических наук, доцент департамента статистики и анализа данных факультета экономических наук Национального исследовательского университета «Высшая школа экономики» (гл. 1);

Сиротин Вячеслав Павлович — кандидат технических наук, доцент, профессор департамента статистики и анализа данных факультета экономических наук Национального исследовательского университета «Высшая школа экономики» (гл. 5, 7).

Предисловие

Современный специалист, чтобы быть конкурентоспособным на рынке труда, должен владеть количественными методами анализа и прогнозирования. Отсюда и новые, повышенные, требования к его подготовке. Жизнь подтверждает справедливость утверждения о том, что кто владеет информацией, тот владеет миром. Однако чтобы информация превратилась в знания, нужно уметь правильно подготовить и обработать данные, владеть методами моделирования, анализа и интерпретации результатов их обработки. Анализ данных пронизывает все аспекты современной жизни, служит основой для многих решений в предпринимательской и общественной деятельности, информируют о тенденциях и факторах, которые влияют на нашу жизнь.

Обозреватель журнала *Business Week* Стивен Бейкер отмечает, что только с помощью математики и статистики современный бизнес сможет выжить во все возрастающих информационных потоках (№ 2, 2006 г.). Умение понимать и применять числовую аргументацию, сегодня востребовано везде. Специалист в области экономики, будь то менеджер, финансист, маркетолог или бухгалтер, должен хорошо владеть методами обработки и анализа данных для принятия эффективных управленческих решений.

Роль методов анализа данных в нашей жизни настолько значительна, что люди, часто не задумываясь и не осознавая, постоянно их используют в повседневной практике. Работая и отдыхая, делая покупки, знакомясь с другими людьми, принимая какие-то решения, человек определенным образом анализирует данные, для чего систематизирует и сопоставляет факты, делает необходимые для себя выводы и принимает определенные решения. Таким образом, в каждом человеке генетически заложены способности к анализу данных и синтезу информации об окружающем нас мире.

Анализ данных как научная дисциплина в системе прикладной статистики разрабатывает и систематизирует понятия, приемы, математические методы и модели, предназначенные для организации отбора из исследуемой совокупности подлежащих обследованию единиц, их стандартной записи, систематизации и обработке с целью их удобного представления и интерпретации, получения научных и практических выводов. Очевидно, что применение любого математического метода для изучения явления означает использование его формальной модели, т.е. определенной системы предпосылок и постулатов.

В настоящем учебнике анализ данных рассматривается как *дисциплина, основанная на статистических методах и вычислительных алгоритмах, позволяющих извлекать знания из результатов наблюдений.*

Анализ данных, опирающийся на множество подходов и алгоритмов, используется практически во всех областях науки и деятельности общества. Он осуществляется исследователем с целью формирования определенных представлений о характере анализируемого явления.

В процессе анализа данных исследователь чаще всего пытается их сжать, стремясь потерять при этом как можно меньше заложенной в них полезной информации. Делается это обычно с помощью статистических методов. Сокращение объема данных достигается за счет применения двух взаимно дополняющих принципов: выборочного метода и свертки информации. Первый из них декларирует отказ от всей совокупности данных (генеральной совокупности) в пользу специально организованной ее части — выборки, а второй заменяет всю выборку ее несколькими характеристиками, например средней арифметической, дисперсией и т.д., а также результатами применения методов исследования зависимостей, снижения размерностей и классификации.

Развитие теории и практики статистических методов обработки данных идет в двух параллельных направлениях. Одно из них представлено методами, предусматривающими возможность вероятностной интерпретации данных, использования вероятностных моделей для построения и выбора наилучших методов статистической обработки. Эти методы обычно называют вероятностно-статистическими. Они предполагают, что вероятностные модели адекватны явлениям, изучаемым с их помощью. В этом случае адекватность получаемых выводов основывается на строго доказанных математических результатах, дающих возможность также устанавливать точность получаемых выводов.

Другое направление представлено логико-алгебраическими методами анализа данных, которые не предполагают вероятностных моделей изучаемых явлений. Эти исходные данные, подлежащие статистической обработке, не могут интерпретироваться как выборка из генеральной совокупности. Отсюда следует неправомерность использования вероятностных моделей при выборе методов статистической обработки данных и вероятностной интерпретации полученных результатов. При этом процедуры свертки информации не всегда допускают формального алгоритмического подхода. Такое понимание термина «анализ данных» востребовано в социально-экономических приложениях и нашло отражение в работах многих современных статистиков и специалистов по обработке данных.

Эти методы обработки статистических данных не основываются на строго доказанных математических результатах и, как следствие, не позволяют оценивать точность получаемых с их помощью выводов. При решении таких задач наилучший метод обработки данных обычно выбирается с помощью оптимизации некоторого функционала качества, задаваемого из содержательных соображений. Естественно, что при этом проблема обоснованности выводов, получаемых с помощью методов анализа данных, требует дополнительного внимания, поэтому особую значимость приобретает вопрос согласования содержания задачи и используемых математических методов.

Однако даже в случаях применения вероятностно-статистических методов анализа данных, когда исследователь имеет возможности опираться на формальные критерии, проверка адекватности вероятностной модели изучаемому явлению также должна опираться и на содержательные соображения (критерии). При этом методы анализа данных в обоих рассмотренных случаях могут служить средством получения фундаментальных знаний, выявления неизвестных ранее закономерностей.

Учебник знакомит читателя с рядом статистических подходов и методов анализа данных, формирующих у него основы аналитического мышления для последующего их применения в экономике, социальной сфере и бизнесе.

Изучая дисциплину, студенты становятся компетентными при решении практических задач в профессиональном статистическом анализе данных, моделировании и прогнозировании социально-экономических явлений и процессов, в содержательной интерпретации результатов статистических расчетов.

В результате освоения дисциплины студент должен:

знать

- основные понятия и положения, связанные со сбором, систематизацией, обработкой и анализом статистической данных;
- основные подходы к анализу данных с использованием описательных и вероятностно-статистических методов;

уметь

- определять методы анализа, необходимые для оценки степени и вида зависимостей между переменными, снижения размерности признакового пространства и многомерной классификации данных;
- использовать методы устойчивого, робастного оценивания параметров и непараметрического моделирования;
- анализировать временны́е данные и прогнозировать;

владеть навыками

- описательной статистики, табличного и графического представления данных, их содержательной интерпретации;
- применения многомерных статистических методов исследования зависимостей, снижения размерностей и классификации при анализе социально-экономических данных с использованием аналитического программного обеспечения;
- анализа динамики временны́х данных и прогнозирования социально-экономических процессов с использованием статистических методов и аналитического программного обеспечения;

быть компетентным

- при решении практических задач в профессиональном статистическом анализе данных, моделировании и прогнозировании социально-экономических явлений и процессов;
- в содержательной интерпретации результатов статистических расчетов.

В учебнике рассмотрены основные методы систематизации, обработки и анализа статистической данных. Описательные методы анализа данных нашли отражение в гл. 1, особенности и основные понятия вероят-

ностно-статистического подхода к анализу данных изложены в гл. 2. Для анализа многомерных данных рассматриваются статистические методы выявления и оценки степени зависимости между переменными (гл. 3), построения регрессионных моделей для определения вида статистической зависимости между переменными (гл. 4), снижения размерности признакового пространства (гл. 5) и многомерной классификации данных (гл. 6). Вопросы устойчивого, робастного оценивания параметров и непараметрического моделирования совокупности данных рассмотрены в гл. 7. Анализу временны́х данных и прогнозированию посвящена гл. 8. В приложении учебника приводятся таблицы математической статистики, которые будут полезны при решении задач и упражнений по анализу данных.

Значительное внимание в учебнике уделяется логическому анализу исходной информации и экономической интерпретации получаемых результатов. Учебник снабжен достаточным количеством экономических примеров и задач для самостоятельного решения. Каждая глава завершается перечнем вопросов для повторения материала и тестами.

Учебник подготовлен авторским коллективом Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ) и Московского государственного университета экономики, статистики и информатики (ныне ставшего частью Российского экономического университета имени Г. В. Плеханова), под руководством доктора экономических наук, профессора В. С. Мхитаряна.

Авторы благодарны рецензентам: заведующему кафедрой статистики, эконометрики и информационных технологий в управлении Национального исследовательского Мордовского государственного университета имени Н. П. Огарева, доктору экономических наук, профессору Юрию Владимировичу Сажину и профессору кафедры «Математика 1» Финансового университета при Правительстве РФ Науму Шевелевичу Кремеру за ценные замечания, способствующие улучшению содержания учебника.

Авторы признательны Анне Артуровне Антонян за помощь в подготовке электронной версии учебника.

Глава 1
ПРЕДВАРИТЕЛЬНЫЙ АНАЛИЗ ДАННЫХ. ОПИСАТЕЛЬНАЯ СТАТИСТИКА

В результате изучения материала главы 1 обучающийся должен:

знать

- основные критерии классификации наборов данных и виды классификации;
- основные виды графического представления данных и методы их группировки;
- формулы расчета основных числовых характеристик количественных данных;

уметь

- определять тип шкалы измерения переменной и данных по упорядоченности во времени;
- таблично и графически изображать данные всех типов в наиболее удачной форме;
- строить таблицы частот и вариационные ряды — дискретные и интервальные;
- строить различные типы графиков и интерпретировать их;
- рассчитывать числовые характеристики количественных данных и интерпретировать их;
- находить основные показатели динамики временны́х рядов, строить на их основе прогнозы;

владеть

- категориями и понятиями современной классификации статистических данных;
- категориями, понятиями и методами современной описательной (дескриптивной) статистики и анализа временны́х рядов.

1.1. Классификация статистических данных

1.1.1. Критерии классификации данных

В процессе управления экономическими и техническими системами статистические методы позволяют выработать обоснованные решения, сочетающие интуицию и опыт специалиста с тщательным анализом имеющейся информации. И с каждым годом интерес к статистической обработке данных неуклонно возрастает, так как объемы окружающей нас информации угрожающе увеличиваются и без грамотной их обработки и представления, исследования закономерностей невозможно правильно принимать решения на их основе. При этом анализ данных может проводиться с целью:

- анализа и отображения конкретной собранной информации — в этом случае говорят о статистическом описании, *описательной* (*дескриптивной*) *статистике* (*descriptive statistics*);
- описания всего класса явлений по имеющимся выборочным данным, характеризующим только часть этого класса. Эти задачи относятся к *аналитической статистике*.

Как правило, любое статистическое исследование начинается с дескриптивной статистики, а потом уже при необходимости углубляется аналитической.

Под *данными* (*data*) в статистике понимают совокупность сведений, зафиксированных на определенном носителе в форме, пригодной для их постоянного хранения, передачи и обработки.

В статистике для характеристики изучаемых объектов используются различные типы данных, и к каждому типы применимы свои методы их обработки. Поэтому прежде всего необходимо определиться с их классификацией.

Основные критерии классификации наборов статистических данных [31, 34] следующие:

1) по числу переменных, характеризующих объект исследования, различают *одномерные*, *двумерные и многомерные данные*;

2) по наличию или отсутствию упорядочения во времени различают *пространственные*, *временны́е и пространственно-временны́е данные*;

3) по типу шкалы измерения каждого признака различают *количественные* (числовые) признаки, которые делятся на *дискретные* и *непрерывные*, и *качественные* (категориальные) признаки, которые делятся на *номинальные* и *порядковые*;

4) по способу получения данные делятся на *первичные* — если информация собиралась специально для данного анализа и *вторичные* — если используется информация из других источников, собранная для других целей.

Полная схема классификации данных по названным критериям представлена на рис. 1.1.

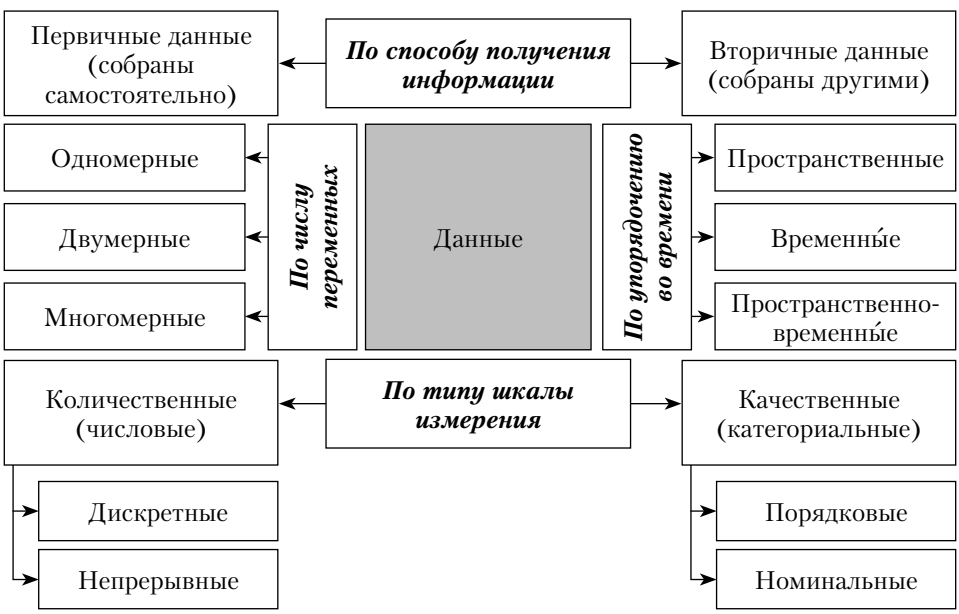


Рис. 1.1. **Классификация статистических данных по различным критериям**

1.1.2. Классификация данных по числу переменных

По числу переменных различают одномерный, двумерный и многомерный массивы данных (см. рис. 1.1).

В ***одномерных*** наборах данных у каждого наблюдения регистрируется только один признак.

В этом случае статистические методы используются для определения основных характеристик этого признака:

- расчет средних значений и показателей вариации, размаха признака;
- группировка данных и построение вариационных рядов (дискретных и интервальных);
- графическое представление данных с целью их визуализации и анализа;
- исследование различий наблюдений или групп наблюдений, требующих особого рассмотрения (задача классификации и выявления аномальных наблюдений).

Примеры одномерных данных

В качестве одномерных данных могут выступать:

- цена некоторого продукта питания в разных магазинах г. Пскова;
- динамика числа семей, нуждающихся в улучшении жилищных условий, в г. Самаре;

Пример 1.1

Индексы цен на первичном рынке жилья по Российской Федерации (на конец года; в % по отношению к концу предыдущего года). Данные представлены в табличной форме и в виде точечно-линейного графика (*line chart*, *time plot*), построенного в программе *MS Excel*.

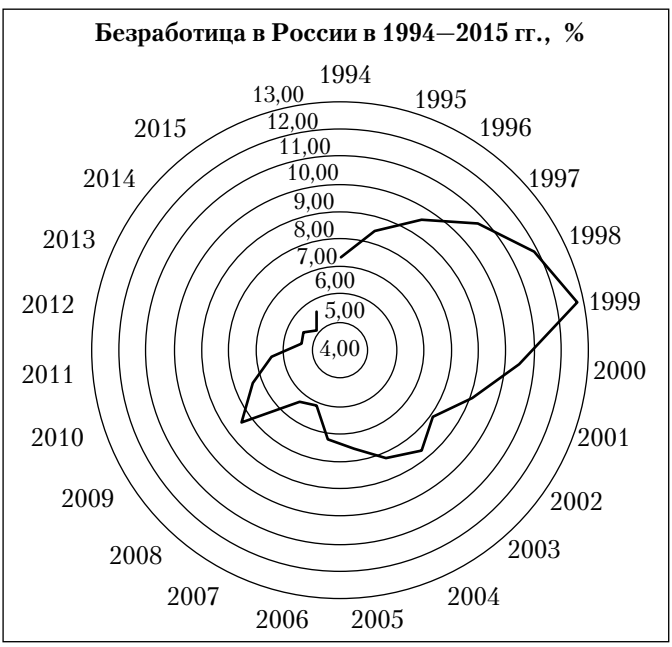
Год	Индекс цен, %	Год	Индекс цен, %	Год	Индекс цен, %	Год	Индекс цен, %
1998	156,9	2003	118,8	2008	110,3	2013	104,8
1999	146,3	2004	118,5	2009	92,4	2014	105,7
2000	113,1	2005	117,5	2010	100,3	2013	104,8
2001	125,1	2006	147,7	2011	106,7	2014	105,7
2002	122,5	2007	123,4	2012	110,7		

Источник: Росстат. URL: http://www.gks.ru/free_doc/new_site/prices/housing/tab9.htm.

Пример 1.2

Уровень безработицы (в %, на конец месяца, в среднем за год) в России в 1994—2015 гг. (в 2015 г. — среднее за первые шесть месяцев). Данные представлены в табличной форме и в виде лепестковой диаграммы (*MS Excel*).

Год	Уровень безработицы, %	Год	Уровень безработицы, %	Год	Уровень безработицы, %
1994	7,39	2002	8,06	2010	7,36
1995	8,53	2003	8,63	2011	6,51
1996	9,60	2004	8,16	2012	5,45
1997	10,81	2005	7,57	2013	5,50
1998	11,86	2006	7,17	2014	5,16
1999	12,74	2007	6,13	2015	5,65
2000	10,49	2008	6,36		
2001	9,03	2009	8,38		



Источник: URL: http://sophist.hse.ru/exes/tables/UNEMPL_M_SH.htm.

В ***многомерных*** (двумерных, трехмерных и т.д.) наборах данных у каждого наблюдения регистрируется несколько признаков.

Статистические методы в этом случае используются для решения задач:

- определения основных характеристик по каждому одномерному признаку;
- анализа наличия и степени зависимости между этими признаками;
- исследования вида зависимости одной переменной (результативной) от остальных (факторных);
- классификации наблюдений с целью получения однородных групп (кластеров) и выявления аномальных наблюдений;

• построения обобщающих, интегральных показателей с целью снижения размерности исходного признакового пространства;

• анализа рядов и прогнозирования (для временны́х данных).

Примеры многомерных данных

В качестве многомерных данных можно рассматривать следующие:

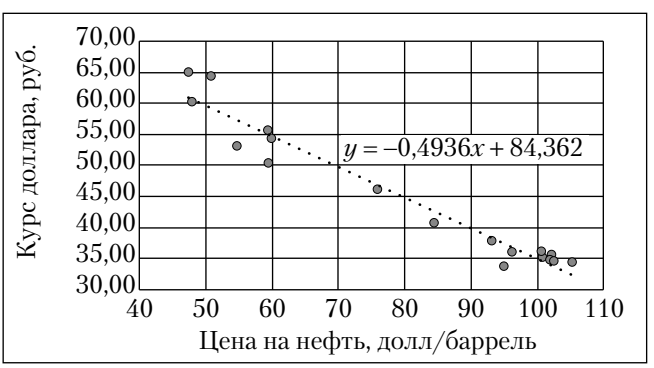
• характеристика работника некоторой фирмы с помощью нескольких показателей: заработная плата, пол, образование, стаж работы, категория работы и производительность труда;

• характеристика квартиры на рынке вторичного жилья в Москве с помощью нескольких показателей: стоимость квартиры, общая площадь, площадь кухни, удаленность от центра, этаж, материалы стен дома.

Пример 1.3

Среднемесячные данные мировых цен на нефть и курса доллара США в Российской Федерации в 2014–2015 гг. (двумерные данные). Данные представлены в табличной форме и в виде точечного графика (*scatter plot*, *point plot*) — диаграммы рассеяния с построенной линейной регрессионной зависимостью между переменными (см. гл. 4) в программе *MS Excel*.

Месяц, год	Цена на нефть, долл/баррель*	Курс доллара, руб.**	Месяц, год	Цена на нефть, долл/баррель*	Курс доллара, руб.**
Январь 2014	94,86	33,78	Октябрь 2014	84,34	40,80
Февраль 2014	100,68	35,24	Ноябрь 2014	75,81	46,22
Март 2014	100,51	36,20	Декабрь 2014	59,29	55,77
Апрель 2014	102,04	35,67	Январь 2015	47,33	65,15
Май 2014	101,80	34,83	Февраль 2015	50,73	64,52
Июнь 2014	105,15	34,45	Март 2015	47,85	60,36
Июль 2014	102,39	34,64	Апрель 2015	54,63	53,22
Август 2014	96,08	36,10	Май 2015	59,37	50,47
Сентябрь 2014	93,03	37,90	Июнь 2015	59,83	54,45

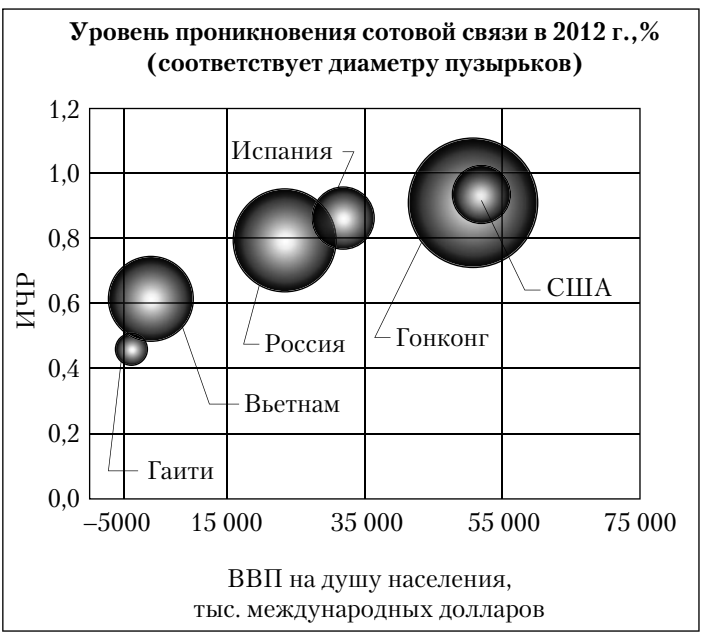


Источники: * — USA Energy Information Administration. URL: http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=rclc1&f=d; ** — Центральный банк РФ. URL: http://cbr.ru/currency_base/dynamics.aspx?VAL_NM_RQ=R01235&date_req1=01.07.1992&date_req2=22.07.1998&rt=1&mode=1.

Пример 1.4

Характеристика стран мира по уровню проникновения сотовой связи; ВВП на душу населения и индекса человеческого развития (ИЧР) (трехмерные данные). ИЧР — интегральный показатель, рассчитываемый ежегодно для межстранового сравнения и измерения уровня жизни, грамотности, образованности и долголетия как основных характеристик человеческого потенциала исследуемой территории. Данные представлены в табличной форме и в виде пузырькового графика (*bubble plot*) (*MS Excel*). Пузырьковая диаграмма позволяет наглядно представить на двумерной плоскости трехмерные данные — с помощью диаметра или площади поверхности пузырьков отобразить различия в каком-либо показателе у исследуемых объектов.

Страна	Уровень проникновения сотовой связи в 2012 г., %*	ВВП на душу населения, тыс. долл.	ИЧР (2012 г.)**
Гонконг	229,24	51 103	0,906
Россия	182,92	23 589	0,788
США	95,45	51 749	0,937
Гаити	59,91	1208	0,456
Испания	108,36	32 134	0,855
Вьетнам	147,66	3787	0,617

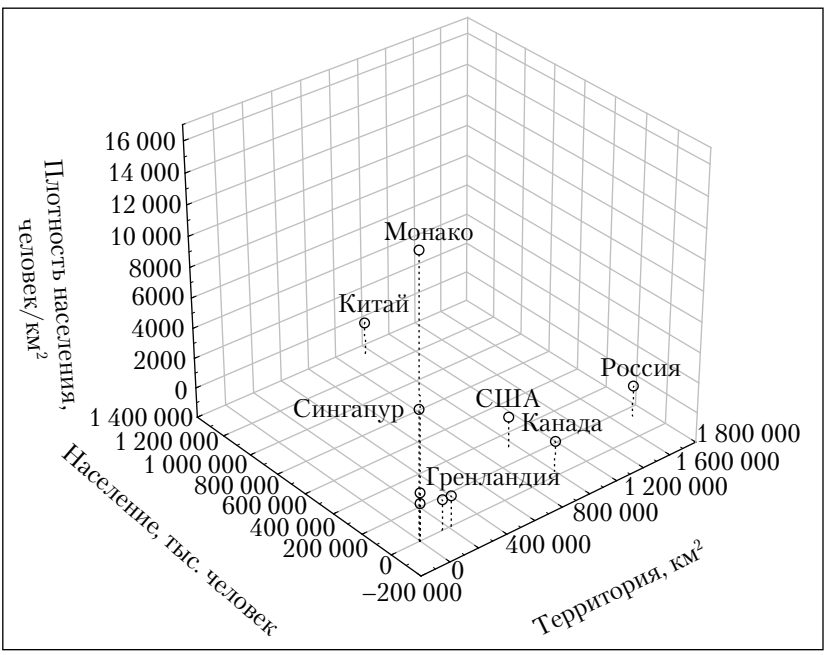


Источники: * — AC&M Consulting; ** — Human Development Report. URL: http://hdr.undp.org/en/2013-report.

Пример 1.5

Характеристика некоторых стран мира на 1 июля 2009 г. по показателям: площадь территории, численность населения и плотность населения на 1 км^2 (трехмерные данные). Данные представлены в табличной форме и в виде трехмерного точечного графика (3*D scatterplot*), построенного в пакете *STATISTICA* [9] (*StatSoft*).

Страна	Территория, км^2	Население, тыс. человек	Плотность населения, человек/км^2
Монако	2	33	16350
Сингапур	705	4615	6545
Бермудские острова	54	65	1190
Сан-Марино	61	31	512
Китай	9 596 961	1 337 411	139
США	9 629 091	311 666	32
Россия	17 098 240	141 394	8
Канада	9 984 670	33 259	3
Монголия	1 564 100	2641	2
Гренландия	2 166 086	57	0,03



Источник: URL: http://www.statistica.md/category.php?l=ru&idc=147.

Из приведенных примеров следует, что в статистике каждое наблюдение кроме количественных признаков содержит качественные характеристики, привязывающие это наблюдение ко времени и к месту (страна, город и т.д.).

1.1.3. Классификация данных по наличию или отсутствию упорядочения во времени

По рассматриваемому критерию различают пространственные, временны́е и пространственно-временны́е данные.

Пространственные данные — значения переменных, относящихся к однотипным объектам в один и тот же фиксированный момент времени. Они позволяют сравнить значение признака, измеренное на разных объектах исследования, сравнить эти объекты по степени проявления в них

того или иного свойства и разделить все объекты в соответствии с этим на группы и категории.

Примеры пространственных данных

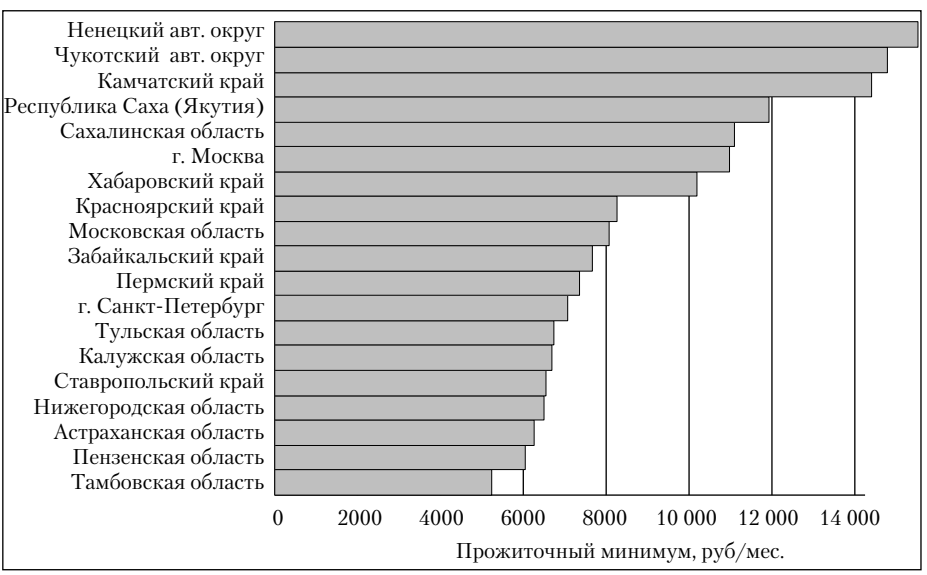
К пространственным данным можно отнести:

- данные по курсам покупки и продажи определенной валюты в разных обменных пунктах города на конкретный день;
- данные по курсам покупки и продажи валюты различных стран в определенном обменном пункте города на конкретный день;
- температуру воздуха в разных городах России в определенный день.

Пример 1.6

Величина прожиточного минимума (в среднем на душу населения), установленная в некоторых субъектах РФ за IV квартал 2013 г. Данные представлены в табличной форме и в виде ленточной (полосовой) диаграммы (*MS Excel*).

Регион	Рублей в месяц	Регион	Рублей в месяц
Тамбовская область	5230	Московская область	8072
Пензенская область	6057	Красноярский край	8249
Астраханская область	6271	Хабаровский край	10 182
Нижегородская область	6488	г. Москва	10 965
Ставропольский край	6543	Сахалинская область	11 083
Калужская область	6682	Республика Саха (Якутия)	11 923
Тульская область	6740	Камчатский край	14 384
г. Санкт-Петербург	7072	Чукотский авт. округ	14 766
Пермский край	7361	Ненецкий авт. округ	15 517
Забайкальский край	7670		



Источник: Росстат. «Регионы России. Социально-экономические показатели», 2014 г. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1138623506156.

Временны́е данные отражают динамику изменения переменных, характеризующих объект, на некотором промежутке времени. В зависимости от способа измерения значений признака временны́е данные делятся на *моментные* и *интервальные*.

Моментные временны́е данные представляют собой измерения (наблюдения) признака, сделанные в определенные моменты времени.

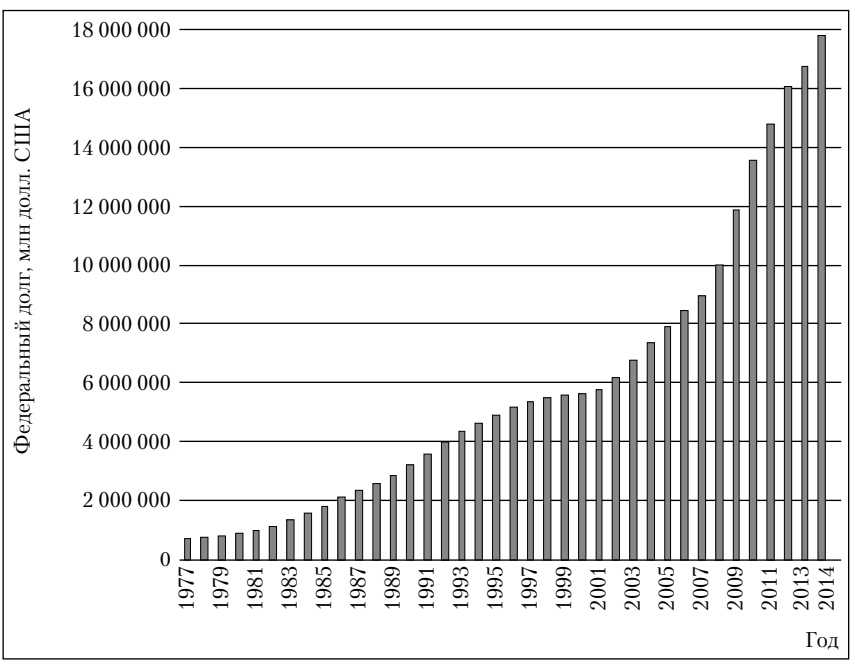
Примеры моментных временны́х данных

К таким данным относятся:

- данные по курсам покупки и продажи определенной валюты в определенном обменном пункте города на каждый день в течение месяца (года и т.д.);
- температура воздуха в некотором городе России, измеренная в определенное время каждый день в течение месяца (года и т.д.);
- величина прожиточного минимума в некотором регионе РФ на определенную дату в течение нескольких лет.

Пример 1.7

Федеральный долг США на конец года за 1977—2014 гг. (в млн долл. США). Данные представлены в виде столбиковой диаграммы (*MS Excel*).



Источник: The White House Office of Management and Budget. URL: http://www.whitehouse.gov/omb/budget/Historicals.

Интервальные временны́е данные характеризуют объект наблюдения за некоторый интервал времени. В отличие от моментных они измеряются не в какой-то момент времени, а в течение некоторого временно́го промежутка — недели, месяца, года и т.д.

Примеры интервальных временны́х данных

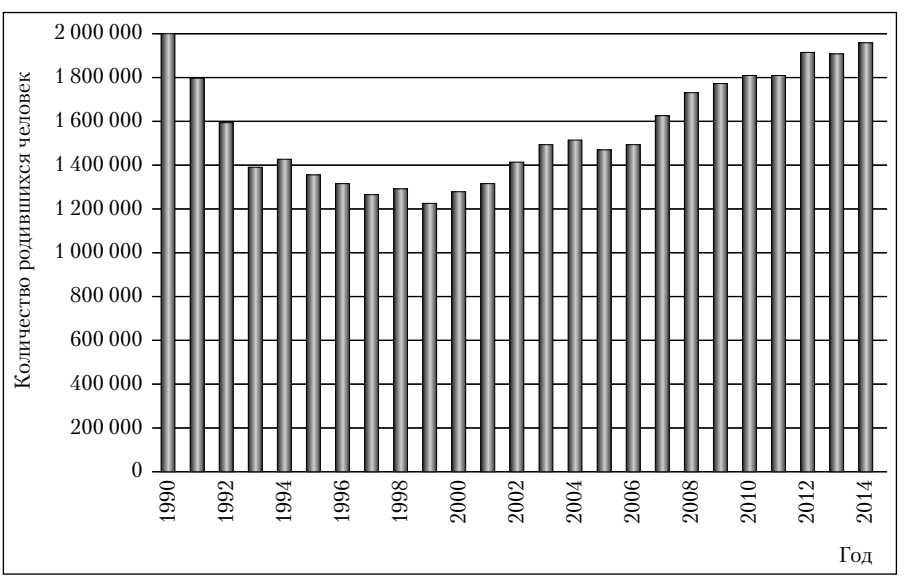
К таким данным относятся:

- сводки ГИБДД по количеству ДТП, произошедших в Москве за день, с 1 по 30 марта 2015 г.;
- годовой объем экспорта (импорта) России по годам за последние 20 лет;
- динамика ежегодных данных о количестве преступлений, совершенных в Российской Федерации за год, за последние 10 лет;
- число браков и разводов, зарегистрированных в Центральном федеральном округе по кварталам, за последние 10 лет.

Пример 1.8

Динамика числа родившихся (без мертворожденных) в Российской Федерации в течение года за 1990—2014 гг. (за 2014 г. — с учетом Крымского федерального округа). Данные представлены в табличной форме и в виде столбиковой диаграммы (*MS Excel*).

Год	Количество родившихся за год, человек	Год	Количество родившихся за год, человек	Год	Количество родившихся за год, человек
1990	1 988 858	1999	1 214 689	2008	1 713 947
1991	1 794 626	2000	1 266 800	2009	1 761 687
1992	1 587 644	2001	1 311 604	2010	1 788 948
1993	1 378 983	2002	1 396 967	2011	1 796 629
1994	1 408 159	2003	1 477 301	2012	1 902 084
1995	1 363 806	2004	1 502 477	2013	1 895 822
1996	1 304 638	2005	1 457 376	2014	1 942 683
1997	1 259 943	2006	1 479 637		
1998	1 283 292	2007	1 610 122		



Источник: Росстат. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/demography/#.

Пространственно-временны́е данные — значения переменных, относящихся к сходным объектам за несколько моментов времени.

Они могут быть также как моментными, так и интервальными.

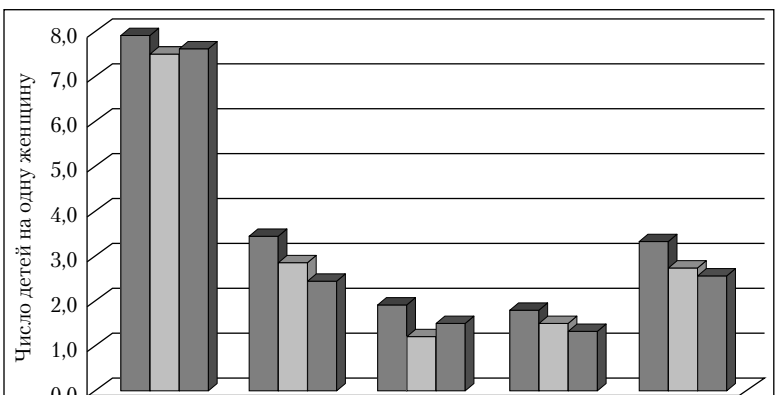
Примеры пространственно-временны́х данных

К таким данным относятся:

- сводки ГИБДД по количеству ДТП, произошедших в крупнейших российских городах за день, с учетом их видов, с 1 по 30 мая 2014 г.;
- годовой объем экспорта и импорта европейских стран по годам за последние 20 лет;
- динамика ежегодных данных о количестве преступлений по их видам, совершенных в азиатских странах, за последние 10 лет;
- число браков и разводов, зарегистрированных в федеральных округах Российской Федерации за последние 10 лет, по кварталам.

Пример 1.9

Данные Всемирной организации здравоохранения о среднем количестве детей у женщин некоторых стран мира за 1990—2013 гг. Данные представлены в табличной форме и в виде столбиковой диаграммы (*MS Excel*) по пяти объектам (четыре страны и весь мир в целом) по одному показателю — среднее число детей у женщин в динамике за три года.



	Нигер (max в 2013 г.)	Венесуэла (медиана)	Россия	Сингапур (min в 2013 г.)	В мире
1990	7,9	3,4	1,9	1,8	3,3
2000	7,5	2,8	1,2	1,5	2,7
2013	7,6	2,4	1,5	1,3	2,5

Источник: ВОЗ. World Health Statistics 2011, 2015. URL: http://www.who.int/gho/publications/world_health_statistics/en.

Пример 1.10

Данные Всемирной организации здравоохранения о продолжительности жизни мужчин и женщин некоторых стран мира за 1990—2013 гг. Данные представлены в виде структурной столбиковой диаграммы (*MS Excel*) по пяти объектам (четыре страны и весь мир в целом) по двум показателям — продолжительность жизни мужчин и женщин в динамике за три года.