

Предисловие

Сегодня в различных сферах общественной жизни к статистическим методам проявляется повышенный интерес как к одному из важнейших аналитических инструментов для поддержки процессов принятия решений. Статистикой пользуются все: от бизнесменов, стремящихся оптимизировать прибыль от инвестиций, до политиков, желающих предсказать исход выборов, или социологов, оценивающих доверие избирателей к этим политикам, не говоря уже о традиционных областях применения математической статистики — науке, технике, экономике. Очевидно, что, как правило, статистическими методами в своей деятельности пользуются не профессионалы-статистики (где набрать столько профессионалов!), а “обычные” профессионалы в своей области, которые, возможно, когда-то “проходили” в своих университетах курс математической статистики, но “это было так давно, что стало неправдой”.

Мой достаточно большой опыт применения статистических методов в совместной работе с биологами, медиками и в последние годы с экономистами показывает, что распространенное мнение о статистике как об одной из разновидностей лжи идет от неправомерного применения тех или иных статистических методов в конкретных ситуациях. Даже общеупотребительный и “безопасный” критерий Стьюдента, примененный бездумно, например, к выборкам из дискретных генеральных совокупностей, может в некоторых случаях показать удивительные результаты. С другой стороны, почти во всей литературе по математической статистике, включая практические руководства, материал излагается таким образом, что сначала идет “теория”, например основы метода максимального правдоподобия, а затем в качестве иллюстрации к “теории” предлагается несколько практических методов. В таком случае практику весьма сложно выбрать необходимые методы проведения статистического анализа, сравнить эти методы и тем более обосновать их применение. (Редким исключением в общем ряду такой статистической литературы является книга Дж. Полларда *Справочник по вычислительным методам статистики*, в которой представлены практические методы статистики и описание области применимости каждого из них.)

Эта книга задумывалась и написана как “сборник рецептов” статистических методов, которые часто используются на практике и сравнительно просто реализуются в электронной таблице Excel. Для каждого приведенного метода *четко описана статистическая модель*, в рамках которой его можно применять. Кроме того, методы сгруппированы по типу исходных данных, предъявляемых для статистического анализа. Таким образом, отдельно описаны методы для анализа одномерных выборок, отдельно — для зависимых наблюдений и т.д. Методы представлены в таком виде, чтобы их могли легко отобрать для своих потребностей и сравнительно просто реализовать практики (необязательно профессионалы-статистики), которым необходимо самостоятельно провести статистический анализ своих данных.

В этой связи необходимо отметить выбор электронной таблицы Excel как средства реализации методов статистического анализа. Существует множество специализированных программных средств для статистических расчетов: отечественные STADIA, СИГАМД, ОЛИМП:СтатЭксперт или зарубежные STATGRAPHICS,

STATISTICA, SPSS и общематематические пакеты (например, Mathcad, Matlab, Maple), которые также имеют встроенные статистические средства. Но наибольшее распространение как средство проведения различных расчетов, в том числе и статистических, в настоящее время получили электронные таблицы, среди которых безусловным лидером является Microsoft Excel. Эта электронная таблица входит в пакет Microsoft Office, который установлен практически на каждом компьютере. Microsoft Excel имеет достаточное количество встроенных статистических средств, включая надстройку Пакет анализа и порядка 80 статистических функций. Это обусловило выбор Excel в качестве основного средства для проведения статистического анализа. Несмотря на то что в книге все примеры реализованы в Excel 2002, они без существенных изменений могут быть перенесены на другие версии Excel, начиная с Excel 97 и заканчивая Excel 2003.

Хотя книга задумывалась только как сборник статистических методов, оказалось невозможным обойтись без вводной части, посвященной основам теории вероятностей и математической статистики, и специальной части, описывающей статистические возможности Excel. Поэтому книга состоит из четырех частей. В части I, *Основные понятия теории вероятностей и математической статистики*, приводятся основные понятия и сведения из теории вероятностей и математической статистики. Весь материал этой части представлен конспективно; здесь приведены все необходимые базовые понятия, определения, теоремы и статистические модели, которые позволят читателю вполне осознанно и продуктивно применять статистические методы, описанные в последующих частях книги. Конечно, эта часть совсем не предназначена для того, чтобы по ней изучать такую обширную и насыщенную область математики (хотя некоторые темы освещены достаточно подробно), как теория вероятностей и математическая статистика. Ее можно использовать как справочное пособие, к которому рано или поздно будет вынужден обратиться как практик-«нестатистик», который использует статистический анализ в своей работе, так и специалист-статистик (у любого специалиста рано или поздно возникает необходимость вернуться к «истокам» — базовым понятиям). Кроме того, материал этой части используется в части II для ссылок при описании статистических средств Excel.

В части II, *Средства Excel для статистического анализа*, описаны возможности Excel для проведения статистического анализа. Предполагается, что читатель знаком с основами работы в этой электронной таблице хотя бы в следующем объеме: он может вводить и редактировать данные, создавать формулы, использовать функции, строить диаграммы и графики, форматировать рабочий лист и т.п. Это базовые навыки работы с Excel, которые известны каждому, кто прослушал курс информатики и вычислительной техники (и при этом, конечно, усвоил необходимые знания) в любом вузе любого профиля. В этой части достаточно полно описаны статистические функции и средства, предоставляемые надстройкой Пакет анализа. К сожалению, справочная система Excel настолько неполно и невнятно (и даже с ошибками!) представляет эти функции и средства, что необходимость их полного описания очевидна. (Следует отметить, что в Excel 2003 справочная система написана более профессионально, при этом исправлены некоторые ошибки.) Кроме статистических функций и средств пакета анализа, в данной части рассмотрены общие средства и надстройки Excel, которые «не заявлены» как имеющие непосредственное отношение к статистическим методам, но которые также можно использовать в статистическом анализе. Это формулы массивов, специального вида

диаграммы и графики, а также надстройка Поиск решения. В конце части описаны способы моделирования случайных величин в Excel.

В части III, *Анализ одномерных выборок*, показана практическая реализация методов статистического анализа одномерных независимых выборок, рассмотрены вопросы предварительной обработки данных и подбора распределений по имеющимся выборочным значениям, а также приведены методы интервального оценивания параметров распределений и критерии проверки гипотез о значениях этих параметров. Последняя глава части посвящена сравнению распределений нескольких одномерных выборок.

В части IV, *Статистический анализ зависимостей*, описаны методы анализа статистических зависимостей, которые включают в себя широкий спектр статистических алгоритмов. Здесь рассмотрены методы корреляционного анализа, способы построения доверительных интервалов и критерии проверки гипотез о значениях коэффициента корреляции, а также показаны методы сравнения параметров распределений зависимых компонентов многомерных выборок. В последней главе описан ряд задач, связанных с построением регрессий, начиная с общей вычислительной схемы определения коэффициентов уравнений регрессии и заканчивая критериями проверки адекватности построенного уравнения регрессии. Хотя число рассмотренных в этой части методов достаточно велико и сами методы весьма громоздки, часть получилась на удивление небольшой. “Виной” этому Excel, в которой есть практически все средства, необходимые для реализации данных методов.

В конце книги приведен небольшой список литературы, на которую есть ссылки в тексте или которая может дополнить определенные темы, освещенные недостаточно полно.

Я буду признателен всем, кто поделится своими соображениями по улучшению содержания книги и стиля изложения материала, а также укажет на возможные ошибки (к сожалению, в книгах, содержащих более ста формул, вероятность ошибок всегда отлична от нуля). Мой адрес электронной почты — aminko@dialektika.com.

А.А. Минько

От издательства “Диалектика”

Вы, читатель этой книги, и есть главный ее критик. Мы ценим ваше мнение и хотим знать, что было сделано нами правильно, что можно было сделать лучше и что еще вы хотели бы увидеть изданным нами. Нам интересно услышать и любые другие замечания, которые вам хотелось бы высказать в наш адрес.

Мы ждем ваших комментариев и надеемся на них. Вы можете прислать нам бумажное или электронное письмо либо просто посетить наш Web-сервер и оставить свои замечания там. Одним словом, любым удобным для вас способом дайте нам знать, нравится ли вам эта книга, а также выскажите свое мнение о том, как сделать наши книги более интересными для вас.

Посылая письмо или сообщение, не забудьте указать название книги и ее авторов, а также ваш обратный адрес. Мы внимательно ознакомимся с вашим мнением и обязательно учтем его при отборе и подготовке к изданию последующих книг. Наши координаты:

E-mail: info@dialektika.com
WWW: <http://www.dialektika.com>

Информация для писем из:

России: 115419, Москва, а/я 783
Украины: 03150, Киев, а/я 152