В. Т. Галочкин

ЭКОНОМЕТРИКА

УЧЕБНИК И ПРАКТИКУМ
ДЛЯ ПРИКЛАДНОГО БАКАЛАВРИАТА

Москва • Юрайт • 2017

УДК 330.43(075.8)
ББК 65в6я73
Г16

***Автор*:**

Галочкин Валерий Тимофеевич — доцент, кандидат физико-математических наук, доцент Департамента анализа данных, принятия решений и финансовых технологий Финансового университета при Правительстве Российской Федерации.

***Рецензенты*:**

Иванус А. И. — профессор, доктор экономических наук, профессор Департамента анализа данных, принятия решений и финансовых технологий Финансового университета при Правительстве Российской Федерации;

Бельчук А. И. — профессор, доктор экономических наук, профессор Всероссийской академии внешней торговли Министерства экономического развития Российской Федерации, профессор кафедры конституционного права имени Н. В. Витрука Российского государственного университета правосудия.

Галочкин, В. Т.

Г16 Эконометрика : учебник и практикум для прикладного бакалавриата / В. Т. Галочкин. — М. : Издательство Юрайт, 2017. — 288 с. — Серия : Бакалавр. Прикладной курс.

ISBN 978-5-9916-9201-4

В учебнике даны основные формулы эконометрики и подробно рассмотрены примеры решения задач с использованием основных приемов и методов эконометрического анализа. Учебник разбит на главы по основным разделам эконометрики. В начале каждого параграфа приведены основные формулы с объяснениями, необходимыми для решения задач. В конце каждой главы приведены типовые задачи для самостоятельного решения и вопросы для самоконтроля. В конце книги даны тестовые задания, ответы на задачи и тесты и математико-статистические таблицы.

Соответствует актуальным требованиям Федерального государственного образовательного стандарта высшего образования.

Для студентов-бакалавров экономических направлений и специальностей высших учебных заведений.

УДК 330.43(075.8)
ББК 65в6я73

ISBN 978-5-9916-9201-4

Оглавление

Предисловие

Эконометрика — дисциплина федерального компонента, она включена в основную образовательную программу подготовки экономистов, определяемую федеральными государственными образовательными стандартами высшего образования. Автором поставлена задача написать учебник, позволяющий выпускнику экономисту-бакалавру решать задачи в области практической экономики. Поэтому в учебнике основное внимание уделено решению задач по каждой теме и по материалу каждого параграфа.

При изложении материала предполагается, что читатель владеет основами теории вероятностей и математической статистики, а также линейной алгеброй. Везде, где требуется, даны ссылки на учебники, подробно излагающие теорию того или иного обсуждаемого вопроса.

Учебник состоит из основного учебного материала (гл. 1—11) и приложений.

В *главе 1* изложены предмет и задачи дисциплины «Эконометрика». Даны основные аспекты методологии построения эконометрических моделей. Показана необходимость сбора статистических данных для оценивания параметров эконометрической модели.

В *главе 2* представлен краткий обзор основных понятий теории вероятностей и математической статистики, которые понадобятся читателю при изучении последующих глав учебника. Подчеркнем, что этот обзор не заменяет перечисленные курсы, а вводит читателя в предмет «Эконометрика».

В *главах 3* и *4* рассмотрены модели линейной регрессии: в главе 3 — модель парной регрессии, в главе 4 — модель множественной регрессии. Даны основные понятия регрессионного анализа, предпосылки классической модели регрессии. В главе 4 показано применение аппарата матричной алгебры для нахождения коэффициентов уравнения регрессии. Изложена методика получения прогнозных характеристик с помощью уравнения регрессии. В обеих главах в качестве примера используется построение и исследование уравнения регрессии с помощью инструмента «Регрессия» *Microsoft Excel*[1].

Главы 5—7 посвящены проблемам, связанным с использованием регрессионных моделей, с невыполнением предпосылок метода наименьших квадратов. В главе 5 рассмотрены вопросы гетероскедастичности — непостоянство дисперсии случайных отклонений. Показаны пути преодоления

---

[1] Программа *Microsoft Excel* используется исключительно из-за ее доступности, простоты и широкого распространения. Существуют и другие программы решения подобных задач, например пакеты *Stata*, *R* и др.

гетероскедастичности. В главе 6 изложены вопросы, касающиеся проблемы автокорреляции случайных отклонений. Глава 7 освещает проблему мультиколлинеарности объясняющих переменных в модели множественной линейной регрессии. Во всех главах приводятся способы обнаружения и смягчения или устранения, если возможно, последствий нарушения перечисленных предпосылок.

В *главе 8* кратко изложены нелинейные модели и способы их линеаризации, сведения к линейным моделям регрессии. Рассмотрены признаки выбора формы модели, виды ошибок спецификации и другие вопросы, связанные с построением «хорошей» модели.

Глава 9 посвящена изучению моделей, содержащих вместе с количественными переменными и качественные (фиктивные) переменные. Рассмотрены качественные переменные сдвига и наклона.

В *главе 10* изучены эконометрические модели, касающиеся различных аспектов временны́х рядов. Рассмотрены способы сглаживания временно́го ряда, выделения трендовой составляющей аддитивной и мультипикативной моделей. Рассмотрено введение фиктивных переменных во временны́е ряды. Даны основы построения прогноза по временны́м рядам.

В *главе 11* изложены эконометрические модели, выраженные системой одновременных уравнений. Приведены необходимое и достаточное условия идентифицируемости системы уравнений. Рассмотрено нахождение коэффициентов уравнений косвенным и двухшаговым методами наименьших квадратов.

Во всех главах рассмотрены в основном линейные эконометрические модели как наиболее простые с точки зрения оценки коэффициентов и дающие наименьшие ошибки прогнозных значений. Рассмотрение временны́х рядов ограничено изучением стационарных рядов.

Изложение материала сопровождается подробным разбором примеров по каждому разделу, в том числе и использованием компьютерных средств. Приложение 1 посвящено работе с инструментом «Регрессия» *Microsoft Excel* с подробным разъяснением результатов работы программы. В конце каждой главы даны задачи для контроля читателем пройденного материала. Необходимые для решения примеров и задач математико-статистические таблицы даны в приложении 2. В приложении 3 приведен краткий перечень основных понятий, использованных при изложении материала.

В основу учебника положен материал проведения занятий по эконометрике со студентами-бакалаврами в Финансовом университете при Правительстве РФ.

Процесс изучения дисциплины направлен на формирование следующих компетенций. Студенты должны обладать способностью применять полученные знания на практике, анализировать, обобщать и систематизировать информацию, осуществлять постановку целей и задач исследований, выбор оптимальных путей и методов их достижения, а также:

знать

- основные понятия эконометрики;
- предпосылки применения эконометрических методов;
- основные типы задач эконометрического анализа и методы их решения;

уметь

- осуществлять сбор, анализ и обработку данных, необходимых для решения поставленных экономических задач;
- выбирать инструментальные средства для обработки экономических данных в соответствии с поставленной задачей, анализировать результаты эконометрического моделирования и обосновывать полученные выводы;
- на основе описания экономических процессов и явлений строить стандартные теоретические и эконометрические модели, анализировать и содержательно интерпретировать полученные результаты;
- анализировать и интерпретировать финансовую, бухгалтерскую и иную информацию, содержащуюся в отчетности предприятий различных форм собственности, организаций, ведомств и т.д., и использовать полученные сведения для принятия управленческих решений;
- анализировать и интерпретировать данные отечественной и зарубежной статистики о социально-экономических процессах и явлениях, выявлять тенденции изменения социально-экономических показателей;
- используя отечественные и зарубежные источники информации, собирать необходимые данные, анализировать их и готовить информационный обзор и (или) аналитический отчет;
- использовать для решения аналитических и исследовательских задач современные технические средства и информационные технологии;

владеть

- базовыми теоретическими знаниями, основными научными понятиями и категорийным аппаратом современной эконометрики;
- навыками применения методов эконометрического анализа, подготовки и представления аналитических обзоров и обоснований, помогающих сформировать профессиональное суждение при принятии управленческих решений на уровне экономических субъектов.

Глава 1
ПРЕДМЕТ И ЗАДАЧИ ДИСЦИПЛИНЫ «ЭКОНОМЕТРИКА»

После изучения данной главы студент должен:

знать

- основные понятия эконометрического подхода;

уметь

- осуществлять сбор, анализ и обработку данных, необходимых для решения поставленных экономических задач;

владеть

- базовыми теоретическими знаниями, основными научными понятиями и категорийным аппаратом современной эконометрики.

1.1. Определение эконометрики

Слово «эконометрика» представляет собой комбинацию двух слов: «экономика» и «метрика» (от греч. *metron* — мера, размер). Поэтому сам термин подчеркивает специфику и содержание эконометрики как науки: количественное выражение связей, которые обоснованы экономической теорией.

Эконометрика является наукой об измерении и анализе экономических явлений. Появление эконометрики как самостоятельного предмета является следствием междисциплинарного подхода к изучению различных экономических процессов. Эконометрика возникла в результате взаимодействия и объединения в единое целое трех компонент: экономической теории, математической статистики и других экономико-математических методов. В конце XX в. с развитием вычислительной техники эконометрика обогатилась различными новыми методами вычислений.

Эконометрика — это наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов.

Предметом эконометрики являются факты, формирующие развитие экономических процессов и явлений.

Закономерности в экономике выражаются в виде математических моделей зависимостей различных экономических показателей. В эконометрике модельные зависимости получают из реальных статистических данных.

Задачи эконометрики — построение математических моделей реальных экономических процессов, оценивание параметров полученных моделей, проверка гипотез о значимости полученных моделей.

Эконометрический анализ — основа экономического анализа: объяснение полученных экономических результатов и прогнозирование. Эконометрический анализ позволяет принять обоснованное экономическое решение.

1.1.1. Типы данных

Поскольку эконометрика оперирует преимущественно с процессами в экономике, дадим определение экономических данных.

Экономические данные — характеристика конкретного экономического объекта. Эти данные формируются под воздействием множества факторов. Не все экономические факторы возможно учесть, также экономическим данным присущ характер случайности (т.е. экономические данные имеют статистическую природу). Для обработки и анализа экономических данных необходимо применять методы теории вероятностей и математической статистики. Поэтому эконометрика оперирует не с функциональными, а со статистическими зависимостями.

При моделировании реальных экономических процессов, как правило, оперируют с двумя типами экономических данных: пространственными и временны́ми.

Пространственные данные обычно берутся по одному экономическому показателю от разных однотипных объектов (предприятий, губерний, краев и т.п.) и относятся к одному и тому же моменту времени. Например, данные о выпуске конкретного изделия, сборе урожая, численности работающих в один и тот же момент времени.

Временны́е данные характеризуют один и тот же объект в различные моменты времени. Например, ежедневные данные по выработке продукции, ежеквартальные данные о текучести кадров, статистические данные о национальном доходе за определенный период времени.

Временны́е данные всегда упорядочены во времени. Часто соседние и близкие данные зависимы.

1.1.2. Классы моделей

Любой экономический процесс можно описать математическими уравнениями, т.е. построить математическую модель (уравнение). В модели величины X выделяются как независимые (объясняющие), а величина Y — как зависимая (объясняемая). Например, совокупный рост дохода ведет к увеличению потребления; рост цены на некое благо ведет к снижению спроса на него; снижение банковской процентной ставки ведет к увеличению инвестиций.

Эконометрика оперирует множеством моделей. Выделим только три класса моделей: регрессионные модели с одним уравнением, регрессионные модели временны́х рядов и системы одновременных уравнений.

В *регрессионных моделях с одним уравнением* объясняемая переменная зависит от одной или нескольких объясняющих переменных (например, модель выпуска некоторого изделия в зависимости от капитала и количества работников).

Уравнение регрессии — это математическая модель статистической связи между переменными. По виду используемой функции регрессионные модели подразделяются на *линейные* и *нелинейные*. Если формула статистической связи линейна, имеем линейную регрессию. В общем случае модель может быть любого вида.

Разработаны эффективные методы оценки и анализа линейных и некоторых видов нелинейных регрессионных моделей. Отметим, что в прикладной эконометрике *базовым* является анализ линейных регрессионных моделей (парной и множественной). Только усвоив парную линейную регрессионную модель, можно переходить к изучению остальных типов моделей. Опираясь на анализ линейных моделей, строят анализ различных нелинейных моделей.

К *моделям временны́х рядов* относятся модели тренда и сезонности. *Тренд* — устойчивое изменение уровня показателя в течение длительного периода времени. *Сезонность* — регулярное устойчивое колебание уровня показателя (в течение недели, месяца, или года).

К этому классу относится множество и более сложных моделей: модель с распределенным лагом, модель авторегрессии k-го порядка, модель адаптивных ожиданий и др. Все эти модели объясняют поведение временно́го ряда исходя только из его предыдущих и настоящих значений.

Модели временны́х рядов подразделяются на стационарные и нестационарные.

Стационарные — ряды, имеющие постоянное среднее значение, колеблются вокруг среднего с постоянной дисперсией. Такой временной ряд не содержит трендовой или сезонной компоненты.

Нестационарные — ряды, распределение уровня которых зависит от времени, т.е. нестационарный ряд имеет трендовую и (или) сезонную компоненты.

Отметим, что область применения в экономике регрессионных моделей значительно шире, чем моделей временны́х рядов.

Системы одновременных уравнений — это модели, которые описываются системами нескольких уравнений, состоящими из регрессионных уравнений и тождеств. В общем случае в каждом уравнении и тождестве помимо объясняющих переменных содержатся объясняемые переменные из других уравнений системы.

При моделировании реальных экономических процессов используются все классы моделей. Факторы, явно не учтенные в экономической модели, но оказывающие на изучаемый объект некоторое результирующее воздействие, учитываются случайной компонентой — ε. Наличие случайной компоненты позволяет проводить проверку модели методами математической статистики.

1.1.3. Типы зависимостей

Задача экономических исследований — анализ зависимостей между переменными. Зависимости могут быть функциональными или статистическими.

Зависимость в виде точной формулы, в которой каждому значению одной переменной соответствует строго определенное значение другой, называется *функциональной*. Функциональная зависимость между переменными в экономике проявляется редко.

При *статистической* зависимости на учтенные в модели факторы накладывается воздействие случайных факторов. При этом изменение одной объясняющей переменной приводит к изменению математического ожидания зависимой (объясняемой) переменной.

Виды статистических зависимостей следующие:

а) корреляционные: при изменении одной из величин изменяется среднее значение другой (связь между переменными не носит направленного характера):

$$M(Y \mid X = x) = M_x(Y) = \varphi(x),\ M(X \mid Y = y) = M_y(X) = \psi(y),$$

где $M(Y \mid X = x)$ — математическое ожидание случайной величины Y, вычисленное при условии, что случайная величина X приняла значение x, $\varphi(x) \neq \text{const}$; $\psi(y) \neq \text{const}$;

б) регрессионные: односторонняя зависимость среднего значения случайной величины Y от одной X или нескольких $X_1, ..., X_m$ случайных величин.

1.2. Методологические вопросы построения эконометрических моделей

В зависимости от конечных прикладных целей эконометрической модели все участвующие в ней переменные подразделяются на следующие типы:

- *экзогенные переменные*, задаваемые извне, априори, в определенной степени управляемые (планируемые);
- *эндогенные переменные*, значения которых формируются в процессе функционирования анализируемой социально-экономической системы под воздействием экзогенных переменных и во взаимодействии друг с другом; эндогенные переменные являются целью объяснения в эконометрической модели;
- *предопределенные переменные*, которые выступают в роли аргументов или объясняющих переменных;
- *лаговые эндогенные переменные* — переменные, измеренные (определенные) в прошлые моменты, являющиеся известными, заданными.

Эконометрическая модель служит для объяснения поведения эндогенных переменных в зависимости от значений экзогенных и лаговых эндогенных переменных.

Весь процесс эконометрического моделирования можно разбить на семь основных этапов:

1-й этап (постановочный) — определение конечных целей моделирования, выбор набора участвующих в модели объясняющих факторов и показателей, выяснение их роли;

2-й этап (априорный) — предварительный анализ экономической сущности изучаемого явления, подбор и формализация априорной информации об экономическом объекте, исходных допущений, включая случайные остаточные составляющие в виде гипотез;

3-й этап (информационный) — подбор необходимой статистической информации для всех факторов, участвующих в модели;

4-й этап (спецификация модели) — обнаруженные на предыдущих этапах связи между эндогенными и экзогенными переменными выражаются в математической форме в виде одного или нескольких уравнений. Осмысливаются все данные по переменным, входящим в уравнение. Выбор общего вида модели: состава объясняющих переменных и формы связей между переменными Успех эконометрического моделирования зависит от того, насколько точно выполнена задача спецификации;

5-й этап (параметризация) — оценивание коэффициентов (параметров) выбранной модели на основе имеющейся статистической информации;

6-й этап (идентификация модели) — статистический анализ модели в целом и отдельных параметров модели. Этот этап непосредственно связан с проблемой идентификации модели, т.е. ответа на вопрос «Отвечает ли построенная эконометрическая модель существующим экономическим данным?»;

7-й этап (верификация модели) — сопоставление реальных статистических и модельных (рассчитанных) данных, проверка адекватности модели, оценка точности модельных данных, проверка статистических гипотез, проверка общего качества уравнения.

В ходе верификации модели решаются следующие вопросы:

- насколько хорошо удалось решить проблему спецификации;
- какова точность (абсолютная, относительная) расчетов экзогенных переменных, основанных на построенной модели;
- можно ли рассчитывать на то, что использование полученной модели в целях прогноза экзогенных переменных даст результаты, адекватные действительности.

Искусство построения математической модели состоит в том, чтобы совместить как можно бóльшую лаконичность в ее математическом описании с достаточной точностью модельного воспроизведения именно тех сторон анализируемой реальности, которые в наибольшей степени интересуют исследователя.

Перечисленные этапы построения эконометрической модели носят условный характер. Основное место в эконометрическом исследовании занимают этап оценки параметров модели и проверки их статистической значимости, а также проверка статистической значимости всего построенного уравнения регрессии. По результатам параметризации и идентификации модели конкретизируется модель (уравнение), уточняются связи параметров с выходными характеристиками модели.

Если модель удовлетворяет всем необходимым требованиям, то она может быть использована для прогнозирования эндогенных переменных или объяснения исследуемых экономических процессов. Такая модель позволяет предсказать среднее значение исследуемого экономического показателя на основе прогнозируемых значений входных (экзогенных) факторов, предвидеть допустимые отклонения конкретных значений изучаемой величины от предсказываемого по модели. Хорошая модель поможет определить, на какие факторы и в каком направлении и объеме следует воздействовать, чтобы значение исследуемого показателя лежало в определенных числовых границах. Отметим, что эконометрические модели, описывая взаимосвязи изучаемых процессов, не решают вопроса о причине этих взаимосвязей.

Схема, представленная на рис. 1.1, демонстрирует суть и последовательность эконометрических исследований экономических процессов.

Рис. 1.1. **Циклический характер эконометрического моделирования**

Приведенная схема отражает циклическую природу эконометрического исследования: от экономической теории к моделированию; от моделирования к совершенствованию экономической теории и более глубокому пониманию сути происходящих процессов. Дальнейшее развитие вычислительных систем и специальных прикладных программ, совершенствование методов анализа превратили эконометрику в мощнейший инструмент исследований экономических процессов.

1.3. Сбор статистических данных для оценивания параметров эконометрической модели

Важным этапом при проведении эконометрического исследования является сбор статистических данных об анализируемом объекте или процессе в виде конкретных значений эндогенных переменных и предопределенных переменных, входящих в спецификацию модели. Данная информация необходима для определения оценок неизвестных коэффициентов, входящих в эконометрическую модель.

В эконометрических исследованиях большое внимание уделяется проблеме данных, т.е. специальным методам работы при наличии данных с пропусками, влиянию агрегирования данных на эконометрические измерения. Зачастую по единицам исследуемой совокупности информация отсутствует, а в наличии имеются данные, характеризующие более крупные единицы (агрегаты).

Следует отметить, что при агрегировании временны́х данных опасность искажения результатов измерений гораздо больше, чем при агрегировании пространственных данных, потому что, с одной стороны, добавляется эффект автокорреляции, а с другой — происходит погашение случайной компоненты.

Таким образом, в итоге заключаем, что эконометрика — наука об измерении и анализе экономических явлений. Тщательно построенная и проверенная эконометрическая модель отвечает двум задачам: объяснение полученных экономических результатов и возможность построения прогноза на базе сформированной эконометрической модели.

Вопросы и задания для самоконтроля

1. Приведите определение эконометрики, отражающее современный взгляд на эту науку.
2. Каковы прикладные цели эконометрики?
3. Перечислите основные этапы эконометрического моделирования.
4. Что входит в спецификацию модели?
5. Что происходит на этапе идентификации модели?
6. Какие основные типы экономических данных вы знаете?
7. Перечислите основные типы эконометрических моделей.
8. Как происходит верификация модели?

Глава 2
ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ[1]

После изучения данной главы студент должен:

знать

- основные типы взаимосвязей признаков;
- способы определения тесноты взаимосвязи между признаками;

уметь

- проводить статистический анализ для различных типов экономических данных;
- интерпретировать результаты статистического анализа;

владеть

- методами статистического анализа.

2.1. Основные математические предпосылки эконометрического моделирования. Закон больших чисел, неравенство и теорема Чебышёва

Основными математическими предпосылками эконометрического моделирования являются теоремы Чебышёва, Бернулли и Ляпунова. Совокупность этих теорем носит общее название *закона больших чисел*.

На практике исследователи часто сталкиваются с таким комплексом условий, при осуществлении которого совокупное поведение достаточно большого количества случайных величин почти утрачивает случайный характер и приобретает определенные закономерности. Поэтому для решения подобных задач необходимо знать данный подобный комплекс условий. В этом случае опираются на закон больших чисел. Это целый комплекс теорем, венчает который центральная предельная теорема Ляпунова. Рассмотрим только некоторые из них.

Неравенство Чебышёва справедливо как для дискретных, так и для непрерывных случайных величин. Рассмотрим его на примере дискретных случайных величин.

Предположим, что случайная дискретная величина X подчиняется закону распределения вида

X	x_1	x_2	x_3	...	x_n
P	p_1	p_2	p_3	...	p_n

[1] В данной главе будут приведены основные сведения из теории вероятностей и математической статистики, необходимые для освоения основ эконометрики. Более подробную информацию можно найти в работах [5, 6].

Задача состоит в оценке вероятности того, что отклонение случайной величины X от ее математического ожидания $M(X)$ не превышает по абсолютной величине положительного числа ε. Если число ε достаточно мало, то задача будет состоять в оценке вероятности того, что случайная величина X примет значения, достаточно близкие к своему математическому ожиданию $M(X)$. Данная задача решается с применением неравенства Чебышёва.

Неравенство Чебышёва. Для любой случайной величины X, имеющей математическое ожидание a и дисперсию $D(X)$, справедливо неравенство

$$P(|X-a|>\varepsilon)\le\frac{D(X)}{\varepsilon^2}, \tag{2.1}$$

где $a = M(X)$, $a > 0$.

Неравенство Чебышёва можно записать в другой форме:

$$P(|X-a|\le\varepsilon)\ge 1-\frac{D(X)}{\varepsilon^2}. \tag{2.2}$$

Неравенство Чебышёва применимо для любых случайных величин. Первая запись неравенства (2.1) определяет верхнюю границу, а вторая (2.2) — нижнюю границу вероятности рассматриваемого события.

Теорема 2.1 (Чебышёва). *Если дисперсии n независимых случайных величин $X_1, X_2, ..., X_n$ ограничены одной и той же постоянной, то при неограниченном увеличении числа n средняя арифметическая случайных величин сходится по вероятности к средней арифметической их математических ожиданий $a_1, a_2, ..., a_n$:*

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+...+X_n}{n}-\frac{a_1+a_2+...+a_n}{n}\right|\le\varepsilon\right)=1. \tag{2.3}$$

Смысл теоремы Чебышёва: при большом числе случайных величин X_i ($i = 1, 2, ..., n$) их случайная средняя (далее в тексте — математическое ожидание $M(x)$)

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$$

сколь угодно мало отличается от неслучайной величины $\frac{1}{n}\sum_{i=1}^{n}a_i$ (т.е. величина $\bar{X}$ перестает быть случайной).

Теорема 2.2 (Бернулли). *Если в каждом из n независимых испытаний событие A имеет постоянную вероятность p, то как угодно близка к единице вероятность того, что отклонение относительной частоты m/n от вероятности p по абсолютной величине будет сколь угодно малым, если число испытаний достаточно велико, т.е. при соблюдении условий теоремы справедливо равенство*

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)=1. \tag{2.4}$$

Теорема Бернулли позволяет заменять неизвестную вероятность случайного события его относительной частотой.

Теоремы Чебышёва и Бернулли представляют собой факт приближения средней большого числа случайных испытаний к некоторым определенным постоянным.

При определенных условиях совокупное действие случайных величин приводит к нормальному закону распределения, который определяется центральной предельной теоремой, или теоремой Ляпунова. В качестве одной из математических предпосылок эконометрического моделирования выступает следствие из теоремы Ляпунова.

Следствие теоремы Ляпунова. Если случайная величина X является суммой очень большого числа попарно независимых случайных величин, влияние каждой из которых на всю сумму ничтожно мало, то случайная величина X подчиняется закону распределения, который близок к нормальному закону распределения вероятностей случайной величины.

Если суммарную ошибку наблюдений рассматривать как сумму очень большого числа попарно независимых частных ошибок, то можно сделать вывод о том, что суммарная ошибка подчиняется закону распределения, который близок к нормальному закону распределения вероятностей.

2.2. Случайные величины

Случайной величиной называется величина, которая под воздействием случайных факторов может с определенными вероятностями принимать разные значения из некоторого множества чисел.

Функцией распределения случайной величины X называется вероятность того, что X принимает значения, меньшее x, т.е.

$$F(x) = P(X < x);\ x \in \mathbb{R}.$$

Свойства функции распределения $F(x)$:

1) $0 \le F(x) \le 1$ для любых $x \in \mathbb{R}$;
2) $P(x_1 \le X \le x_2) = F(x_2) - F(x_1)$;
3) $F(x)$ — неубывающая функция;
4) $\lim\limits_{x\to-\infty} F(x) = 0$; $\lim\limits_{x\to+\infty} F(x) = 1$.

Пример 2.1

Функция распределения непрерывной случайной величины X имеет вид

$$F(x) = \begin{cases} 0 \text{ при } x \le 0, \\ \dfrac{x}{3} \text{ при } 0 < x \le 3, \\ 1 \text{ при } x > 3. \end{cases}$$

Найдем вероятность P того, что случайная величина X примет значение в полуоткрытом интервале[1] [1; 2).

[1] Интервал называется закрытым, если граничные точки входят в него. Обозначение закрытого интервала — квадратные скобки. Интервал называется открытым, если граничные точки не входят в него. Обозначение открытого интервала — круглые скобки. Полуоткрытый интервал — когда одна из границ входит в интервал, а другая — нет.

Решение. График функции распределения представлен на рис. 2.1.

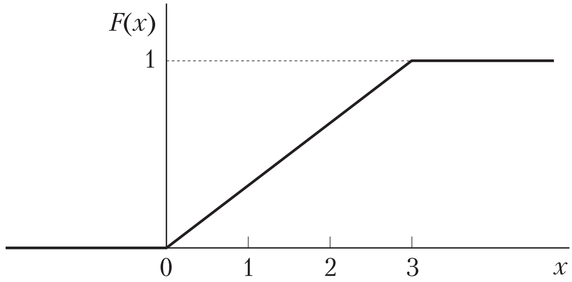

Рис. 2.1. **К примеру 2.1**

По формуле $P(x_1 \le X \le x_2) = F(x_2) - F(x_1)$ находим

$$P(1 \le X \le 2) = F(2) - F(1) = 2/3 - 1/3 = 1/3.$$

Пример 2.2

Дан ряд распределения случайной величины X:

x_i	1	2	4	5
p_i	0,2	0,3	0,4	0,1

Найдем и изобразим графически функцию распределения случайной величины X.

Решение. Будем задавать различные значения x и находить для них значение функции распределения $F(x) = P(X < x)$.

1. Если $x \le 1$, то, очевидно, $F(x) = 0$, в том числе и при $x = 1$, $F(1) = P(X < 1) = 0$.
2. Пусть $1 < x \le 2$; тогда $F(x) = P(X = 1) = 0{,}2$; $F(2) = P(X < 2) = 0{,}2$.
3. Пусть $2 < x \le 4$; тогда $F(x) = P(X < 4) = P(X = 1) + P(X = 2) = 0{,}2 + 0{,}3 = 0{,}5$; $F(4) = P(X < 4) = 0{,}5$.
4. Пусть $4 < x \le 5$; тогда $F(x) = P(X < 5) = P(X = 1) + P(X = 2) + P(X = 4) = 0{,}2 + 0{,}3 + 0{,}4 = 0{,}9$; $F(5) = P(X < 5) = 0{,}9$.
5. Пусть $x > 5$; тогда $F(x) = P(X = 1) + P(X = 2) + P(X = 4) + P(X = 5) = 0{,}2 + 0{,}3 + 0{,}4 + 0{,}1 = 1$.

Теперь построим график функции $F(x)$ (рис. 2.2), где

$$F(x) = \begin{cases} 0 & \text{при } x \le 1, \\ 0{,}2 & \text{при } 1 < x \le 2, \\ 0{,}5 & \text{при } 2 < x \le 4, \\ 0{,}9 & \text{при } 4 < x \le 5, \\ 1{,}0 & \text{при } x > 5. \end{cases}$$

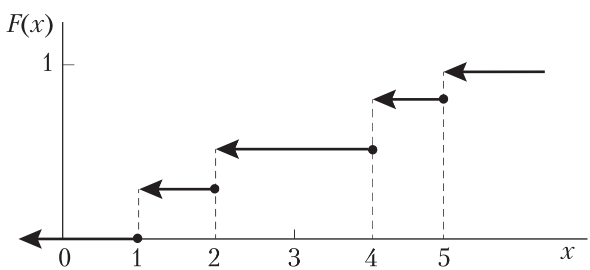

Рис. 2.2. **К примеру 2.2**

Замечание 2.1. При подходе к точкам разрыва слева функция $F(x)$ сохраняет свое значение (эти значения на графике $F(x)$ (см. рис. 2.2) выделены точками).

Случайные величины бывают дискретные и непрерывные.

Дискретной называется случайная величина, которая принимает отдельные, изолированные друг от друга значения. Число значений дискретной случайной величины конечно или счетно.

Дискретную случайную величину удобно задавать в виде ряда распределения

$$\begin{pmatrix} x_1 & x_2 & \dots & x_N \\ p_1 & p_2 & \dots & p_N \end{pmatrix},$$

где $p_i = P(X = x_i)$, $\sum_{i=1}^{N} p_i = 1$.

Непрерывной называется случайная величина, множество значений которой непрерывно заполняет некоторый числовой промежуток.

Непрерывную случайную величину задают плотностью распределения вероятности $f(x)$ или функцией распределения $F(x)$. Связь между ними: $f(x) = F'(x)$.

Вероятностный смысл $f(x)$:

$$f(x) = F'(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x \le X \le x + \Delta x)}{\Delta x}.$$

Свойства плотности распределения $f(x)$:

1) $f(x) \ge 0$ для всех $x \in \mathbb{R}$;

2) $P(a \le X \le b) = \int_a^b f(x)dx$;

3) $F(x) = \int_{-\infty}^{x} f(x)dx$;

4) $\int_{-\infty}^{+\infty} f(x)dx = 1$ — условие нормировки.

Генеральной совокупностью называется множество всех значений случайной величины, которые она может принять в процессе наблюдения (например, данные о численности населения всех городов страны).

Часть генеральной совокупности, отобранная для изучения, называется *выборочной совокупностью* (выборкой). Главное требование к выборочной совокупности — выборка должна быть репрезентативной, т.е. представительной, отражающей все характеристики генеральной совокупности.

2.3. Числовые характеристики распределения генеральной совокупности

Математическим ожиданием дискретной случайной величины x_i называется величина

$$m_x = M(X) = \sum_{i=1}^{N} x_i p_i, \tag{2.5}$$

где N — объем генеральной совокупности; p_i — вероятность значения случайной величины x_i.

Математическим ожиданием непрерывной случайной величины x называется величина

$$M(X)=\int_{-\infty}^{+\infty} xf(x)dx. \tag{2.6}$$

Пример 2.3

Найдем математическое ожидание для дискретной случайной величины из примера 2.2.

Решение. Имеем

$$m_x=M(X)=\sum_{i=1}^{N}x_i p_i=\sum_{i=1}^{4}x_i p_i=1\cdot 0,2+2\cdot 0,3+4\cdot 0,4+5\cdot 0,1=2,9.$$

Ответ. $m_x = 2,9$.

Пример 2.4

Найдем математическое ожидание для непрерывной случайной величины из примера 2.1.

Решение. Найдем плотность функции распределения $f(x) = F'(x)$:

$$f(x)=\begin{cases}0 \text{ при } x\le 0,\\ \dfrac{1}{3} \text{ при } 0<x\le 3,\\ 0 \text{ при } x>3.\end{cases}$$

Находим математическое ожидание по формуле (2.6):

$$M(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_0^3 x\frac{1}{3}dx=\left.\frac{x^2}{6}\right|_0^3=\frac{3}{2}.$$

Ответ. $M(X)=\dfrac{3}{2}$.

Пример 2.5

Известны законы распределения случайных величин X, Y — числа очков, выбиваемых 1-м и 2-м стрелками в тире (табл. 2.1, 2.2).

Таблица 2.1

Вероятности поражения мишени 1-м стрелком

x_i	0	1	2	3	4	5	6	7	8	9	10
p_i	0,15	0,11	0,04	0,05	0,04	0,1	0,1	0,04	0,05	0,12	0,2

Таблица 2.2

Вероятности поражения мишени 2-м стрелком

y_i	0	1	2	3	4	5	6	7	8	9	10
p_i	0,01	0,03	0,05	0,09	0,11	0,24	0,21	0,1	0,1	0,04	0,02

Выясним, какой из двух стрелков стреляет лучше.

Решение. Чтобы ответить на вопрос задачи, необходимо вычислить $M(X)$ и $M(Y)$:

$$M(X) = \bar{x} = 0 \cdot 0{,}15 + 1 \cdot 0{,}11 + \ldots + 9 \cdot 0{,}12 + 10 \cdot 0{,}2 = 5{,}36;$$

$$M(Y) = \bar{y} = 0 \cdot 0{,}01 + 1 \cdot 0{,}03 + \ldots + 9 \cdot 0{,}04 + 10 \cdot 0{,}02 = 5{,}36,$$

т.е. среднее число очков у двух стрелков одинаковое. Здесь $\bar{x}$ и $\bar{y}$ — выборочные средние случайных величин X и Y.

Ответ. По математическому ожиданию невозможно определить, какой из стрелков стреляет лучше.

Свойства математического ожидания.

Если a, b — постоянные числа; X, Y — случайные величины, то:

1) $M(a) = a$;
2) $M(bX) = bM(X)$;
3) $M(a + bX) = a + bM(X)$;
4) $M(X + Y) = M(X) + M(Y)$;
5) $M(X - m_x) = 0$.

Определение 2.1. Случайные величины X и Y называются *независимыми*, если $P(X = x; Y = y) = P(X = x)P(Y = y)$ для любых значений x и y.

Если случайные величины X и Y независимы, то $M(XY) = M(X)M(Y)$.

Разброс возможных значений случайной величины вокруг среднего значения (математического ожидания) определяется дисперсией.

Дисперсия случайной величины определяется как

$$D(X) = M(X - m_x)^2. \tag{2.7}$$

Для практических целей более удобна формула $D(X) = M(X^2) - m_x^2$.

Пример 2.6

Найдем дисперсию дискретной случайной величины из примера 2.2.

Решение. В примере 2.3 нашли математическое ожидание для этого дискретного распределения — $m_x = 2{,}9$. Дисперсию определим по формуле (2.7):

$$D(X) = M(X^2) - m_x^2 = 1^2 \cdot 0{,}2 + 2^2 \cdot 0{,}3 + 4^2 \cdot 0{,}4 + 5^2 \cdot 0{,}1 - 2{,}9^2 = \\ = 10{,}3 - 8{,}41 = 1{,}89.$$

Ответ. $D(X) = 1{,}89$.

Дисперсия непрерывной случайной величины определяется по формуле

$$D(X) = \int_{-\infty}^{+\infty} (x - m_x)^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - m_x^2. \tag{2.8}$$

Пример 2.7

Найдем дисперсию непрерывной случайной величины из примера 2.1.

Решение. В примере 2.4 найдены плотность функции распределения $f(x) = \dfrac{1}{3}$ и математическое ожидание этой непрерывной случайной величины $m_x = \dfrac{3}{2}$.

По формуле (2.8) имеем

$$D(X)=\int_{-\infty}^{+\infty} x^2 f(x)dx - m_x^2 = \int_0^3 x^2 \frac{1}{3} dx - \left(\frac{3}{2}\right)^2 = 3 - \frac{9}{4} = \frac{3}{4}.$$

Ответ. $D(X)=\frac{3}{4}$.

Средним квадратическим отклонением (СКО) случайной величины X называется величина $\sigma_x = \sqrt{D(X)}$. Отметим, что значение σ_x всегда берется с положительным знаком.

Пример 2.8

В примере о стрелках в тире (см. пример 2.5) вычислим дисперсию и среднее квадратическое отклонение числа выбитых очков для каждого стрелка.

Решение. В примере 2.5 было найдено $M(X) = M(Y) = 5{,}36$. Вычисляем дисперсии случайных величин по формуле (2.7), как в примере 2.6:

$$D_x = \sum_{i=1}^{n} (x_i - \overline{x})^2 p_i = (0-5{,}36)^2 \cdot 0{,}15 + (1-5{,}36)^2 \cdot 0{,}11 + \ldots + (10-5{,}36)^2 \cdot 0{,}2 = 13{,}61;$$

$$D_y = \sum_{i=1}^{n} (y_i - \overline{y})^2 p_i = (0-5{,}36)^2 \cdot 0{,}01 + (1-5{,}36)^2 \cdot 0{,}03 + \ldots + (10-5{,}36)^2 \cdot 0{,}02 = 4{,}1704;$$

$$\sigma_x = \sqrt{D_x} = 3{,}69; \sigma_y = \sqrt{D_y} = 2{,}04.$$

При равенстве средних значений числа выбиваемых очков $M(X) = M(Y)$ дисперсия (характеристика рассеяния относительно среднего значения) меньше у второго стрелка.

Свойства дисперсии.

Если a, b — постоянные числа, то:

1) $D(a) = 0$;
2) $D(bX) = b^2 D(X)$;
3) $D(a + bX) = b^2 D(X)$;
4) если случайные величины X и Y независимы, то $D(X \pm Y) = D(X) + D(Y)$.

Пример 2.9

Найдем дисперсию случайной величины $Z = 8X - 5Y + 7$, если известно, что случайные величины X и Y независимы и $D_x = 1{,}5$, $D_y = 1$.

Решение. Опираясь на свойства дисперсии, получим

$$D(Z) = 8^2 \cdot D(X) + 5^2 \cdot D(Y) + 0 = 64 \cdot 1{,}5 + 25 \cdot 1 = 121.$$

Обозначение нормального распределения: $X \sim N(m; \sigma^2)$.

Ковариацией между случайными величинами X и Y называется величина

$$\operatorname{cov}(x, y) = K(X; Y) = M((X - m_x)(Y - m_y)),$$

или

$$K(X; Y) = M(XY) - M(X)M(Y).$$

Ковариация $\operatorname{cov}(x, y)$ есть мера связи двух случайных величин. Ковариацию также называют корреляционным моментом.

Замечание 2.2. Если случайные величины X и Y независимы, то $K(X; Y) = 0$.

Коэффициентом корреляции r_{xy} между случайными величинами X и Y называется нормированный корреляционный момент:

$$r_{x,y} = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right] \cdot \left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} =$$

$$= \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\sqrt{[\overline{x^2} - (\overline{x})^2] \cdot [(\overline{y^2} - (\overline{y})^2]}}. \tag{2.9}$$

Коэффициент корреляции меняется в пределах $-1 \le r_{xy} \le +1$.

Коэффициент корреляции показывает тесноту линейной связи двух случайных величин. Если $r_{xy} > 0$, то между случайными величинами присутствует положительная линейная связь, если $r_{xy} < 0$ — отрицательная линейная связь, если $r_{xy} = 0$, то линейная связь отсутствует.

Случайные величины X и Y называются *некоррелированными*, если $r_{xy} = 0$, и *коррелированными*, если $r_{xy} \neq 0$.

Пример 2.10

Найдем ковариацию и коэффициент корреляции дискретных случайных величин X и Y по данным, приведенным в табл. 2.3.

Таблица 2.3

Данные для примера 2.10

x	2	6	10	14	18
y	1	2	4	11	12

Решение. Представим исходные данные (x, y) и расчетные величины в виде табл. 2.4.

Таблица 2.4

Расчетные величины для примера 2.10

№ п/п	x	y	x^2	xy	y^2
1	2	1	4	2	1
2	6	2	36	12	4
3	10	4	100	40	16
4	14	11	196	154	121
5	18	12	324	216	144
Сумма	50	30	660	424	286
Среднее	$\overline{x} = 10$	$\overline{y} = 6$	$\overline{x^2} = 132$	$\overline{xy} = 84{,}8$	$\overline{y^2} = 57{,}2$

Вычислим:

$$D(x) = \overline{x^2} - (\overline{x})^2 = 132 - 100 = 32;$$

$$D(y) = \overline{y^2} - (\overline{y})^2 = 57{,}2 - 36 = 21{,}2;$$

$$\operatorname{cov}(x, y) = \overline{xy} - \overline{x} \cdot \overline{y} = 84{,}8 - 60 = 24{,}8;$$

$$r_{xy} = \frac{\operatorname{cov}(x, y)}{\sigma_x \sigma_y} = \frac{24{,}8}{\sqrt{32 \cdot 21{,}2}} = 0{,}952.$$

Ответ. $\operatorname{cov}(x, y) = 24{,}8$, $r_{xy} = 0{,}952$.

Замечание 2.3. Если случайные величины X и Y независимы, то они не коррелированы, но из некоррелированности ($r_{xy} = 0$) не следует их независимость, т.е. $r_{xy} = 0$ указывает на отсутствие линейной связи, но не связи вообще.

2.4. Числовые характеристики выборочной совокупности

Пусть из генеральной совокупности случайной величины X извлекается выборка объема n: $x_1, x_2, ..., x_n$.

Выборочное среднее есть величина

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{2.10}$$

Если каждое значение x_i повторяется k_i раз, то выборочное среднее равно

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i k_i, \tag{2.11}$$

где k — число различных значений x_i.

Выборочной дисперсией (вариацией) называется величина

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2. \tag{2.12}$$

Справедлива формула

$$D_{\text{в}} = \overline{x^2} - (\overline{x})^2, \tag{2.13}$$

где $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$.

Свойства выборочной дисперсии.

Если a и b — постоянные числа, то:

1) $D_{\text{в}}(a) = 0$;
2) $D_{\text{в}}(bX) = b^2 D_{\text{в}}(X)$;
3) $D_{\text{в}}(a + bX) = b^2 D_{\text{в}}(X)$.

Пример 2.11

Для случайной величины X получен дискретный статистический ряд

x_i	1	3	7	12
n_i	8	16	6	10

Найдем выборочное среднее.

Решение. Выборочное среднее найдем по формуле (2.11):

$$\bar{x} = \frac{1}{40}(8 \cdot 1 + 16 \cdot 3 + 6 \cdot 7 + 10 \cdot 12) = 5{,}45.$$

Ответ. $\bar{x} = 5{,}45$.

Пример 2.12

Для случайной величины X получен дискретный статистический ряд

x_i	1	5	6	8
n_i	6	4	7	3

Найдем выборочную дисперсию.

Решение. Предварительно находим $\bar{x}$ и $\overline{x^2}$:

$$\bar{x} = \frac{1}{20}(6 \cdot 1 + 4 \cdot 5 + 7 \cdot 6 + 3 \cdot 8) = 4{,}6;$$

$$\overline{x^2} = \frac{1}{20}(6 \cdot 1^2 + 4 \cdot 5^2 + 7 \cdot 6^2 + 3 \cdot 8^2) = 27{,}5.$$

Выборочная дисперсия по формуле (2.13) равна

$$D_в = \overline{x^2} - (\bar{x})^2 = 27{,}5 - 4{,}6^2 = 6{,}34.$$

Ответ. $D_в = 6{,}34$.

Рассмотрим характеристики оценок, получаемых по выборочной совокупности.

Точечные оценки. Обычно характеристики генеральной совокупности неизвестны. Задача заключается в их оценке по характеристикам выборки.

Характеристики генеральной совокупности называются *параметрами*, а характеристики выборки — *оценками*.

Пусть искомый параметр генеральной совокупности равен θ_0, а на основе выборки объема n определяется его оценка θ.

Точечной оценкой θ параметра θ_0 называется числовое значение этого параметра, полученное по выборке, т.е. $\theta \approx \theta_0$.

Для того чтобы выборочная оценка давала хорошее приближение оцениваемого параметра, она должна удовлетворять определенным требованиям: несмещенности, эффективности и состоятельности.

Несмещенность оценок. Оценка θ называется несмещенной, если ее математическое ожидание равно оцениваемому параметру θ_0, т.е. $M(\theta) = \theta_0$. Если это не так, то оценка называется смещенной.

Требование несмещенности гарантирует отсутствие систематических ошибок при оценивании.

Выборочная средняя $\bar{x}$ является несмещенной оценкой генеральной средней, так как $M(\bar{x}) = m_x$.

Выборочная дисперсия $D_в$ является смещенной оценкой генеральной дисперсии D. В качестве несмещенной оценки генеральной дисперсии используется исправленная дисперсия

$$S_x^2 = \frac{n}{n-1} D_\text{в},$$

для которой $M(S_x^2) = D$.

Стандартным отклонением S_x случайной величины X в выборке называется величина

$$S_x = \sqrt{S_x^2}.$$

Стандартное отклонение всегда берется с положительным знаком.

Эффективность оценок. Несмещенная оценка θ называется эффективной, если она имеет минимальную дисперсию по сравнению с другими выборочными оценками.

Выборочная средняя $\overline{x}$ является эффективной оценкой генеральной средней m, т.е. имеет наименьшую дисперсию в классе несмещенных оценок.

Состоятельность оценок. Оценка θ называется состоятельной, если при $n \to \infty$ она стремится по вероятности к оцениваемому параметру θ_0, т.е.

$$\lim_{n\to\infty} P(|\theta - \theta_0| < \varepsilon) = 1.$$

Иначе говоря, состоятельной называется такая оценка, которая дает точное значение для большой выборки.

Теорема Чебышёва закона больших чисел утверждает, что

$$\lim_{n\to\infty} P(|\overline{x} - m| < \varepsilon) = 1,$$

т.е. выборочная средняя $\overline{x}$ является состоятельной оценкой генеральной средней m.

Интервальные оценки. Интервальной называется оценка θ, определяющая числовой интервал $(\theta - \varepsilon;\ \theta + \varepsilon)$, $\varepsilon > 0$, содержащий оцениваемый параметр θ_0, т.е. $\theta - \varepsilon < \theta_0 < \theta + \varepsilon$, или $|\theta - \theta_0| < \varepsilon$.

Доверительным интервалом называют интервал $|\theta - \theta_0| < \varepsilon$, в котором с заданной вероятностью γ заключен неизвестный параметр θ_0, а сама вероятность γ называется *доверительной вероятностью*, или надежностью, т.е. $P(|\theta - \theta_0| < \varepsilon) = \gamma$.

Уровнем значимости α называется вероятность того, что $P(|\theta - \theta_0| > \varepsilon) = \alpha$, причем $\alpha = 1 - \gamma$.

2.5. Проверка статистических гипотез

Статистической гипотезой H называется предположение относительно параметров или вида распределения случайной величины.

Нулевой (основной) называется выдвинутая гипотеза H_0, а конкурирующей (альтернативной) — гипотеза H_1, которая противоречит нулевой.

Проверку статистической гипотезы выполняют на основе результатов выборки. Так как выборка имеет ограниченный объем, то существует возможность принятия ошибочного решения.