

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Тамбовский государственный технический университет»

Н.П. ПУЧКОВ, Т.В. ЖУКОВСКАЯ,
Е.А. МОЛОКАНОВА, И.А. ПАРФЁНОВА, А.И. ПОПОВ

**ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКИХ
ЗНАНИЙ В ПРОФЕССИОНАЛЬНОЙ
ДЕЯТЕЛЬНОСТИ**
ПОСОБИЕ ДЛЯ САМОРАЗВИТИЯ БАКАЛАВРА

Часть 2. Теория вероятностей и математическая статистика

Рекомендовано Учебно-методическим объединением
по университетскому образованию
в качестве учебного пособия для студентов высших учебных заведений,
обучающихся по направлению подготовки бакалавров «Инноватика»



Тамбов
• Издательство ФГБОУ ВПО «ТГТУ» •
2013

УДК 514.12:512.64(075.8)
ББК В11я73
П764

Рецензенты:

Доктор физико-математических наук,
профессор, заведующий кафедрой распределённых
вычислительных систем ФГБОУ ВПО «ТГТУ»
С.М. Дзюба

Доктор физико-математических наук, профессор, заведующий кафедрой
алгебры и геометрии ФГБОУ ВПО «ТГУ им. Г.Р. Державина»
А.И. Булгаков

П764 Применение математических знаний в профессиональной
деятельности. Пособие для саморазвития бакалавра. Ч. 2. Теория
вероятностей и математическая статистика : учебное пособие /
Н.П. Пучков, Т.В. Жуковская, Е.А. Молоканова, И.А. Парфёнова,
А.И. Попов. – Тамбов : Изд-во ФГБОУ ВПО «ТГТУ», 2013. – 64 с. –
100 экз. – ISBN 978-5-8265-1186-2.

Представлены базовые понятия теории вероятностей и математической статистики, изложены методы по использованию математических знаний при решении задач профессиональной деятельности, даны рекомендации по организации самостоятельной работы.

Предназначено для студентов высших учебных заведений, обучающихся по направлению подготовки бакалавров 222000 «Инноватика».

УДК 514.12:512.64(075.8)
ББК В11я73

ISBN 978-5-8265-1186-2 (Ч. 2)
ISBN 978-5-8265-1151-0 (Ч. 1)

© Федеральное государственное бюджетное
образовательное учреждение высшего
профессионального образования
«Тамбовский государственный технический
университет» (ФГБОУ ВПО «ТГТУ»), 2013

ВВЕДЕНИЕ

Цель настоящего пособия состоит в оказании помощи обучающимся в постижении методов теории вероятностей и математической статистики – наук, изучающих закономерности случайных явлений.

Методы теории вероятностей широко применяются в различных отраслях естествознания и техники: теории надёжности, теории массового обслуживания, теоретической физике, геодезии, астрономии, теории стрельбы, теории ошибок наблюдений, теории автоматического управления и во многих других теоретических и практических науках. Теория вероятностей служит также для обоснования математической статистики, которая, в свою очередь, используется при планировании и организации производства, при анализе технологических процессов, предупредительном и приёмочном контроле качества продукции и для многих других целей.

Теория вероятностей и математическая статистика рассматриваются в пособии как математические дисциплины, поэтому получение конкретных естественнонаучных или технических результатов в нём не является самоцелью. Более того, излагаемый учебный материал не претендует на математическую строгость изложения. При желании это можно получить в фундаментальных учебниках. Все примеры в тексте пособия имеют целью разъяснение общих положений теории и указание на связь этих положений с практикой, развивают умения применять теоретические положения в конкретных задачах.

Такое направление изучения даёт возможность выработать своеобразную теоретико-вероятностную интуицию, которая позволяет предвидеть в общих чертах выводы раньше применения формул. Главное, надо иметь в виду, что без систематического решения задач изучать теорию вероятностей и математическую статистику нельзя, поэтому основной упор в пособии делается на организацию самостоятельной работы студентов, на развитие умений применять знания на практике, т.е. на формирование компетенций.

Пособие содержит задания для самостоятельного решения. В каждом разделе они расположены по мере возрастания сложности их решения и, на наш взгляд, вполне доступны для основной массы обучающихся, познакомившихся с основами теории и технологией решения на примерах, приведённых авторами. Для желающих закрепить полученные навыки указаны координаты аналогичных заданий в задачниках, рекомендуемых в учебных программах.

Перечень контрольных вопросов имеется в каждом разделе и в обобщённом виде в конце пособия в форме вопросов к экзамену.

Мы считаем, что формированию наиболее глубоких знаний способствует практика самостоятельного составления (разработки) заданий, на-

пример, аналогичных тем, что имеются в пособии. Это эффективный путь становления компетентным специалистом.

Ещё раз обращаем внимание на то, что данное пособие, как и любое другое пособие, предполагает его использование совместно с классическим учебником, где даётся полное и глубокое изложение изучаемых вопросов.

Мы рекомендуем три учебника, указанных в списке литературы [1, 2, 5], предоставляя тем самым выбор учебного материала, стиль изложения в котором наиболее доступен, понятен для обучающегося. Также в списке литературы указаны два задачника [3, 6], используя которые можно закрепить навыки решения задач по математической статистике и теории вероятностей. В брошюре [4] приводятся решения вероятностных задач, взятых из «повседневной жизни».

1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ. ОСНОВНЫЕ ПОНЯТИЯ

I. Учебные цели. В результате изучения материала студенты должны получить представление о сущности теории вероятностей как математической дисциплины; иметь понятие о случайном событии, его вероятности; уметь классифицировать события, использовать основные понятия теории вероятностей и формулы комбинаторики для непосредственного вычисления вероятностей случайных событий.

II. Формирование компетенций. Формирование математической культуры, совершенствование общей культуры мышления, развитие мышления.

III. Введение в тему.

Задача любой науки, в том числе инноватики, состоит в выявлении и исследовании закономерностей, которым подчиняются реальные процессы. Найденные закономерности имеют не только теоретическую ценность, они широко применяются на практике – в планировании, управлении и прогнозировании.

Теория вероятностей – математическая наука, изучающая закономерности случайных явлений. Её изучение начинается с усвоения основных понятий, которые излагаются в данном разделе.

Для контроля качества усвоения изложенного материала необходимо сконцентрировать внимание на следующих вопросах:

1. Что такое случайное событие?

2. Какие события называются несовместными, элементарными, равновероятными?

3. В каком случае события образуют полную группу?

4. Классическое определение вероятности.

5. Свойства вероятности.

6. Статистическое определение вероятности.
7. Что такое перестановки и как подсчитывается их число?
8. Определение сочетания, формула для подсчёта числа сочетаний, свойства сочетаний.
9. Что такое размещения и чем они отличаются от других комбинаций?
10. Формулировка двух основных правил комбинаторики.

1.1. СЛУЧАЙНЫЕ СОБЫТИЯ И ИХ ВИДЫ

В основе теории вероятностей лежит понятие случайного события и случайной величины. Предмет теории вероятностей – изучение закономерных случайных событий (и случайных величин) при их массовом проявлении.

Определение 1.1. Случайным относительно комплекса условий S называется событие, которое при осуществлении указанного комплекса условий может либо произойти, либо не произойти.

Случайные события обозначают большими (прописными) буквами A, B, C, \dots

Случайное событие трактуется как результат испытания.¹

Например, экзамен – испытание, отличная оценка – событие; выстрел – испытание, попадание – событие.

Определение 1.2. События A и B называют несовместными, если в одном и том же испытании появление одного из них исключает появление другого.

Например, выпадение «орла» при подбрасывании монеты исключает появление «решки».

Определение 1.3. Несколько событий образуют полную группу, если в результате испытания появление хотя бы одного из них достоверно.

Например, при произведении выстрела по мишени (испытание) обязательно будет или промах или попадание. Эти два события образуют полную группу.

Каждый из возможных результатов испытания называют элементарным событием (или исходом). Те элементарные исходы, которые интересуют исследователя, называют благоприятными событиями.

1.2. ВЕРОЯТНОСТЬ СОБЫТИЯ И ЕЁ СВОЙСТВА

Определение 1.4 (классическое определение вероятности). Отношение числа M – благоприятствующих событию A элементарных исходов к числу N – равновозможных несовместных элементарных исходов, обра-

¹ Испытание рассматривается как организованное (человеком) осуществление комплекса условий (подбрасывание монеты, экзамен в вузе, стрельба по мишени), так и «организованное» природой (метеосостояния, аварии).

зующих полную группу, называется вероятностью события A и обозначается $p(A) = M/N$.

Свойства $p(A)$:

1. Вероятность достоверного события равна единице.
2. Вероятность невозможного события равна нулю.
3. Вероятность случайного события $0 < p(A) < 1$.

Это определение не единственное, так как содержит определённые ограничения, связанные с доказательством равновозможности событий, полноты их группы, конечным характером исходов.

Обозначим: n — число испытаний; m — число появлений события A в этих испытаниях (частота события A).

Определение 1.5. Отношение частоты события к числу испытаний называется относительной частотой события.

Определение 1.6 (статистическое определение вероятности).

Предельное значение относительной частоты появления события A при неограниченном возрастании числа испытаний называется вероятностью события A :

$$p(A) = \lim_{n \rightarrow \infty} \frac{m}{n}.$$

Обобщением классической схемы является пространство событий, элементарные исходы которых можно представить в виде точек, заполняющих некоторую область в пространстве R^1 — на прямой, R^2 — на плоскости или R^3 . Если при этом событию A благоприятствуют элементарные исходы, заполняющие некоторую подобласть Ω_1 из Ω , то геометрической вероятностью события A называют отношение меры области Ω_1 к мере области Ω .

1.3. ИСПОЛЬЗОВАНИЕ КОМБИНАТОРИКИ ДЛЯ НАХОЖДЕНИЯ ВЕРОЯТНОСТЕЙ

Комбинаторика изучает операции над конечными множествами. Пусть задано конечное множество элементов некоторой природы. Из них можно составить определённые комбинации, количества которых изучает комбинаторика. Некоторые её формулы используются в теории вероятностей.

Определение 1.7. Комбинации, состоящие из одной и той же совокупности n различных элементов и различающиеся только порядком их расположения, называются перестановками (операция — упорядочение множества). Их число определяется произведением чисел от 1 до n :

$$P_n = 1 \cdot 2 \cdot \dots \cdot n = n!$$

Пример 1.1. Множество $N_1 = 1, 2, 3$. $P_3 = 1 \cdot 2 \cdot 3 = 6$. Это: 123, 132, 213, 231, 312, 321.

Определение 1.8. Комбинации из m элементов, составленные из n различных элементов $m \leq n$, отличающиеся друг от друга либо самими элементами, либо их расположением, называются размещениями (образование упорядоченных подмножеств данного множества). Их число определяется по формуле

$$A_n^m = n(n-1)(n-2)\dots(n-m+1); \text{ очевидно, что } A_n^n = P_n = n!$$

$$A_3^2 = 3 \cdot 2 = 6. \text{ Для множества } N_1 \text{ (пример 1.1) это: } 13, 23, 12, 31, 32, 21.$$

Определение 1.9. Комбинации, содержащие по m элементов каждая, составленные из n различных элементов $m \leq n$ и различающиеся хотя бы одним элементом, называются сочетаниями (образование подмножеств данного множества). Число сочетаний определяется формулой:

$$C_n^m = \frac{n!}{m!(n-m)!}; \text{ очевидно, что } C_n^{n-m} = C_n^m; A_n^m = P_m \cdot C_n^m.$$

$$C_3^2 = \frac{3!}{2!1!} = 3. \text{ Для множества } N_1 \text{ (пример 1.1) это: } 12, 13, 23.$$

Последние комбинации (сочетания) участвуют в качестве коэффициентов в широко известной в математике формуле бинома Ньютона:

$$(p+q)^n = C_n^n p^n q^0 + C_n^{n-1} p^{n-1} q + C_n^{n-2} p^{n-2} q^2 + \dots + C_n^1 p q^{n-1} + C_n^0 p^0 q^n.$$

Два основных правила комбинаторики.

Правило суммы. Если a_1 из множества A можно выбрать n_1 способами, a_2 из $A - n_2$ способами (способ выбора a_1 не совпадает со способом выбора a_2), то a_1 или a_2 можно выбрать $n_1 + n_2$ способами.

Пример 1.2. В студенческой группе 8 человек – жители г. Тамбова, 4 – Тамбовского района, 2 – жители Моршанского района и т.д. Тогда можно выбрать жителя г. Тамбова – 8 способами, Тамбовского района – 4 способами, жителя Тамбова или Тамбовского района – 12 способами.

Правило произведения. Если a_1 из A можно выбрать n_1 способами, а элемент a_2 из A можно выбрать n_2 способами, то выбор a_1 и a_2 может быть осуществлен $n_1 \cdot n_2$ способами.

В примере 1.2 выбор жителя Тамбова и Тамбовского района может быть осуществлен $8 \cdot 4 = 32$ способами.

Пример 1.3. В коробке лежит 10 шаров: 6 белых и 4 чёрных. Найти вероятность того, что среди пяти наугад взятых шаров будут 3 белых и 2 чёрных.

$$\text{Общее число исходов } C_{10}^5 = \frac{10!}{5!5!} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 3 \cdot 2 \cdot 7 \cdot 6 = 252.$$

Число благоприятных исходов согласно правилу произведения

$$C_6^3 \cdot C_4^2 = \frac{6!}{3!3!} \cdot \frac{4!}{2!2!} = \frac{6 \cdot 5 \cdot 4}{1 \cdot 2 \cdot 3} \cdot \frac{4 \cdot 3}{1 \cdot 2} = 20 \cdot 6 = 120, \text{ а } p = \frac{120}{252} \approx 0,48.$$

1.4. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

1.4.1. Даны четыре числа: 2, 4, 6, 8. Сколько можно образовать:

- а) всевозможных четырёхзначных чисел?
- б) четырёхзначных чисел, начинающихся с 2?
- в) различных произведений двух заданных чисел?
- г) различных двузначных чисел?

1.4.2. Найти вероятность того, что из 10 книг, расположенных в случайном порядке, 3 определённые книги окажутся рядом?

1.4.3. В урне находятся 10 красных и 6 белых шаров. Наудачу вынимаются два шара. Какова вероятность, что они будут одноцветными?

1.4.4. Брошены две игральные кости. Найти вероятность следующих событий:

- а) сумма выпавших очков равна семи;
- б) сумма выпавших очков равна восьми, а разность четырём;
- в) число очков, выпавших на первой кости, окажется равным числу очков, выпавших на второй кости.

1.4.5. Собрание, на котором присутствуют 25 человек, в том числе 5 женщин, выбирает делегацию из трёх человек. Считая, что каждый присутствующий с одинаковой вероятностью может быть выбран, найти вероятность того, что в делегацию войдут две женщины и один мужчина.

1.4.6. Студенты A и B знают, соответственно, 40 и 45 экзаменационных вопросов из 50 заявленных преподавателем. Преподаватель случайным образом формулирует два вопроса. Какова вероятность, что на них сможет ответить только студент B ?

Кроме того, можно закрепить полученные навыки, решив задачи № 1.10; 1.12; 1.16а из [1], а также № 9; 22; 43 из [3].

2. ОСНОВНЫЕ ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

I. Учебные цели. В результате изучения материала студенты должны получить представление о математических операциях над вероятностями, что позволит им моделировать структуру сложных событий и уметь находить их вероятности. Изучение этой темы предполагает, что в результате студент должен знать формулы нахождения вероятностей суммы и произведения событий, понятие условной вероятности, понятие зависимых и независимых событий.

II. Формирование компетенций. Формирование математической культуры, развитие аналитического и логического мышления. Закрепление умений лаконично и точно формулировать определения и теоремы, грамотно записывать математические формулы.

III. Введение в тему.

Использование только одних определений при решении вероятностных задач трудоёмко и малопрактично. Поэтому необходимы новые дополнительные определения сложных событий, математических операций, правил, которые существенно расширяют круг решаемых задач, делают соответствующий процесс более алгоритмичным. Расчётная часть задач по теории вероятностей всегда сравнительно проста, наибольшую трудность составляет моделирование – перевод условия задачи на язык формул, необходима способность «угадать» нужную формулу. В развитии такого рода способностей и помогают теоремы теории вероятностей.

При изучении содержания постарайтесь сформулировать ответы на следующие вопросы:

1. Что такое условная вероятность события?
2. Что называется произведением двух событий? Какой пример Вы можете привести?
3. Какой формулой определяется вероятность произведения двух событий?
4. Какие события называются независимыми? Примеры таких событий.
5. Какой формулой определяется вероятность суммы двух событий? Различные виды этой формулы для совместных и несовместных, зависимых и независимых событий.
6. В каких задачах можно использовать формулу полной вероятности? Как она записывается?
7. Что такое переоценка гипотез и на основе какой формулы она осуществляется?
8. Если в серии повторных испытаний некоторое событие может наступить с одной и той же вероятностью, то как подсчитать вероятность наступления этого события заданное число раз?

2.1. УМНОЖЕНИЕ ВЕРОЯТНОСТЕЙ

Определение 2.1. Произведением двух событий A и B называется событие AB , означающее совместное появление этих событий.

Пример 2.1. Абитуриент закончил школу с золотой медалью – событие A ; абитуриент сдал первый экзамен на «отлично» – событие B . Событие AB – абитуриент подлежит зачислению в вуз.

Определение 2.2. Вероятность события B в предположении, что событие A произошло (B/A), называется условной вероятностью $p_A(B) = p(B/A)$.

Если $p_A(B) = p(B)$, то вероятность называется безусловной.

Пример 2.2. Из 25 экзаменационных билетов студент выгучил 20.

Если он первым берёт билет, то вероятность взять известный $p(B) = \frac{20}{25} = 0,8$. Студент выбирает билет вторым. Событие A – первый

студент выбрал билет, известный второму, тогда $p_A(B) = \frac{19}{24}$.

Теорема 2.1. Вероятность произведения двух событий определяется формулой $p(AB) = p(A)p_B(A) = p(B)p_A(B)$.

Пример 2.3. В коробке 6 белых и 4 чёрных шара. Последовательно, без возврата вынимаются два шара. Какова вероятность, что оба окажутся белыми?

$$p(AB) = \frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3}.$$

Теорема 2.1 допускает обобщение на случай произведения любого числа событий A_1, A_2, \dots, A_n .

$$p(A_1, A_2, \dots, A_n) = p(A_1)p_{A_1}(A_2)p_{A_1A_2}(A_3)\dots p_{A_1\dots A_{n-1}}(A_n).$$

Определение 2.3. Событие B называется независимым от события A , если условная вероятность события B равна его безусловной вероятности $p_A(B) = p(B)$.

Для n независимых событий

$$p(A_1A_2\dots A_n) = p(A_1)p(A_2)\dots p(A_n).$$

Пример 2.4. Стреляют три стрелка. Вероятность поражения мишени первым стрелком $p(A_1) = 0,7$; вторым – $p(A_2) = 0,75$; третьим – $p(A_3) = 0,8$. Какова вероятность что в результате залпа мишень будет поражена трижды?

$$P(A_1A_2A_3) = p(A_1)p(A_2)p(A_3) = 0,7 \cdot 0,75 \cdot 0,8 = 0,42.$$

Теорема 2.2. Вероятность появления хотя бы одного из независимых событий $A_1A_2 \dots A_n$, образующих полную группу, определяется формулой $p(A) = 1 - q_1q_2 \dots q_n$, где $q_i = 1 - p_i$ – вероятности соответствующих противоположных событий $\bar{A}_i, i = 1, n$.

Если $p(A_1) = p(A_2) = \dots = p(A_n) = p$, то $p(A) = 1 - q^n$.

В предыдущем примере вероятность, что хотя бы один стрелок поразит мишень, равна $p(A) = 1 - 0,3 \cdot 0,25 \cdot 0,2 = 1 - 0,015 = 0,985$.

2.2. СЛОЖЕНИЕ ВЕРОЯТНОСТЕЙ

Определение 2.4. События A и B называются совместными, если в одном и том же испытании появление одного из них не исключает появление другого.

Пример 2.5. Студент Иванов сдал экзамен – событие A , студент Петров сдал экзамен – событие B . A и B – совместные события.

Определение 2.5. Суммой двух событий A и B называют событие $C = A + B$, которое состоит в появлении либо события A , либо события B , либо A и B одновременно. Сумма нескольких событий $\sum_{i=1}^n A_i$ состоит в появлении хотя бы одного из них.

Пример 2.6. Подбрасывается игральная кость. Событие A – выпало число 2, событие B – выпало число 4, событие C – выпало число 6. Событие $A + B + C$ – выпало число очков, кратное 2.

Теорема 2.3. Вероятность суммы двух событий равна сумме их вероятностей без вероятности их произведения:

$$p(A + B) = p(A) + p(B) - p(AB).$$

Если A и B независимы, то $p(A + B) = p(A) + p(B) - p(A)p(B)$.

Если A и B зависимы, то $p(A + B) = p(A) + p(B) - p(A)p_A(B)$.

Если A и B несовместны, то $p(A + B) = p(A) + p(B)$.

Если при этом A_1, A_2 образуют полную группу, то

$$p(A + B) = p(A) + p(B) = 1.$$

Следствие. Сумма вероятностей противоположных событий равна 1:

$$p(A) + P(\bar{A}) = 1.$$

2.3. ФОРМУЛА ПОЛНОЙ ВЕРОЯТНОСТИ

Пусть события B_1, B_2, \dots, B_n попарно несовместны и образуют полную группу событий $\sum p(B_i) = 1$.

Пусть событие A может наступать при условии появления одного из событий B_i , причём известны как вероятности $p(B_i)$, так и условия вероятности $p_{B_i}(A)$, $i = \overline{1, n}$.

Теорема 2.4. Вероятность события A , появление которого возможно лишь при наступлении одного из несовместных событий B_i , образующих полную группу событий, равно сумме попарных произведений вероятности каждого из этих событий на соответствующую условную вероятность появления события A .

$$p(A) = p(B_1)p_{B_1}(A) + p(B_2)p_{B_2}(A) + \dots + p(B_n)p_{B_n}(A).$$

Пример 2.7. С остановки транспорта можно уехать на автобусе, троллейбусе или такси. В течение 5 минут через остановку проходят 1 троллейбус, 2 автобуса, 3 такси. Вероятность уехать на троллейбусе

равна 0,5 (подошедший троллейбус идёт в нужном для пассажира направлении); на автобусе – 0,8; на такси – 0,3. Какова вероятность, что пассажир уехал с данной остановки в течение ближайших пяти минут с первым подошедшим транспортным средством?

$$\begin{aligned} p(A) &= p(B_1)p_{B_1}(A) + p(B_2)p_{B_2}(A) + p(B_3)p_{B_3}(A) = \\ &= \frac{1}{6} \cdot 0,5 + \frac{2}{6} \cdot 0,8 + \frac{3}{6} \cdot 0,3 = \frac{3}{6} = 0,5. \end{aligned}$$

2.4. ФОРМУЛЫ БАЙЕСА

Пусть заданы исходные условия формулы полной вероятности. События B_i называют гипотезами, так как заранее неизвестно, какое из них наступит. Пусть произведено испытание и в результате появилось событие A . Тогда возможно определить условные вероятности гипотез B_i по следующим формулам:

$$p_A(B_i) = \frac{p(B_i)p_{B_i}(A)}{p(A)}; \quad i = \overline{1, n}.$$

Эти формулы называются формулами Байеса.

Возвращаясь к *примеру 2.7*, можно переоценить вероятности гипотез: $p_A(B_1)$ – пассажир уехал на троллейбусе; $p_A(B_2)$ – на автобусе; $p_A(B_3)$ – на такси:

$$p_A(B_1) = \frac{1/6 \cdot 0,5}{0,5} = \frac{1}{6}; \quad p_A(B_2) = \frac{2/6 \cdot 0,8}{0,5} = \frac{8}{15}; \quad p_A(B_3) = \frac{0,5 \cdot 0,3}{0,5} = 0,3.$$

Наиболее вероятным остаётся событие, что пассажир уехал на автобусе.

2.5. ФОРМУЛА БЕРНУЛЛИ

Определение 2.6. Несколько испытаний называются независимыми относительно события A , если вероятность события A в каждом испытании не зависит от исходов других испытаний.

Будем рассматривать такие независимые испытания, в которых событие A имеет одинаковую вероятность. Пусть производится n независимых испытаний, в каждом из которых событие A может появиться с вероятностью p : $p(A) = p$. Тогда $p(\bar{A}) = 1 - p = q$.

Теорема 2.5. Вероятность сложного события, состоящего в том, что в n испытаниях событие A наступит ровно k раз и не наступит $n - k$ раз, подсчитывается по формуле (Бернулли):

$$P_n(k) = C_n^k p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}.$$

Пример 2.8. Найти вероятность, что при пяти подбрасываниях монеты герб выпадет 3 раза.

$$P_5^3 = \frac{5!}{3!2!} \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^2 = 10 \cdot \frac{1}{32} = \frac{5}{16}.$$

2.6. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

2.6.1. По данным социологов, в городе A данный кандидат в депутаты будет поддержан на выборах большей частью населения с вероятностью 0,6; в городе B – с вероятностью 0,7. Какова вероятность, что на выборах кандидат одержит победу хотя бы в одном из городов A и B ?

2.6.2. Каждый из трёх независимо работающих сигнализаторов своевременно сообщает о нарушении заданного режима работы реактора с вероятностью, соответственно, $p_1 = 0,9$; $p_2 = 0,8$; $p_3 = 0,75$. Какова вероятность того, что при нарушении заданного режима работы сигнал не поступит?

2.6.3. В урне находятся 4 белых, 5 красных и 3 синих шара. Наугад извлекаются по одному шару, не возвращая его обратно. Найти вероятность того, что в первый раз появится белый шар, во второй – красный, в третий – синий.

2.6.4. В первой урне находятся 4 белых и 5 красных шаров, во второй – 7 белых и 3 красных. Из второй урны наугад взяли шар и переложили его в первую урну. Найти вероятность того, что взятый после этого из первой урны шар будет белым.

2.6.5. В ящике в равном количестве находятся внешне одинаковые электролампочки двух номиналов. Известно, что электролампочки с первым номиналом на 92% исправные, со вторым номиналом – на 98%. Случайно выбранная электролампочка оказалась бракованной. Какова вероятность, что она с первым номиналом?

2.6.6. Два человека загадывают по двухзначному числу. Рассматривается событие A : первый человек загадает число, содержащее цифру пять, а второй – число, кратное пяти. Найти вероятность события A .

Кроме того, можно закрепить полученные навыки, решив задачи № 55, 59, 64, 66, 69, 85, 90, 92, 102, 111 из [3].

3. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

I. Учебные цели. Познакомиться с одним из основных понятий теории вероятностей – понятием случайной величины.

В результате изучения материала студенты должны иметь представления о способах задания случайных величин, знать их основные числовые характеристики, уметь находить представление случайных величин в виде формул или графически.

II. Формирование компетенций. Формирование математической культуры, развитие способностей к обобщению, творческих способностей.

III. Введение в тему.

В практической жизни часто приходится сталкиваться с различными величинами. Значения многих из них можно найти непосредственно из опыта или с помощью вычислений, получив предварительные данные с помощью измерений, т.е. также из опыта. В результате повторения некоторых опытов можно получить одно и то же значение определённой величины, а в результате других значение величины изменяется, причём результат каждого отдельного опыта невозможно предугадать заранее. Величины, которые могут принять в результате опыта любое из возможных значений, неизвестно заранее – какое, заслуживают особого внимания и являются предметом изучения.

Вопросы для контроля усвоения материала.

1. Какая величина называется случайной и какие различают их виды?
2. Как можно задать дискретную случайную величину?
3. Запишите (в общем виде) закон распределения случайной величины X – число появления события A в n испытаниях.
4. Что такое математическое ожидание случайной величины?
5. Перечислите свойства математического ожидания.
6. Что такое дисперсия случайной величины?
7. Перечислите свойства дисперсии.
8. Как определяется функция распределения случайной величины?
9. Что такое плотность распределения вероятностей?
10. Какое распределение называется нормальным и почему?
11. Изобразите график нормального распределения и поясните, как влияют числовые характеристики на форму этой кривой.
12. Как на основе нормального закона распределения строятся распределения χ^2 Пирсона, Стьюдента и Фишера?

3.1. ДИСКРЕТНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ И ИХ РАСПРЕДЕЛЕНИЕ

Определение 3.1. Величина называется случайной, если в результате испытания она примет лишь одно возможное значение, заранее неизвестное и зависящее от случайных причин.

Обозначение случайных величин – прописными буквами X, Y, Z, \dots , значения случайных величин – строчными: $x_1, x_2, \dots, x_n; y_1, y_2, \dots, y_m; z_1, z_2, \dots, z_k; \dots$.

Различают два вида случайных величин.

Определение 3.2. Случайная величина, принимающая отдельные возможные значения с определёнными вероятностями, называется дискретной случайной величиной (ДСВ).

Пример 3.1. X – количество очков, выпадающих при подбрасывании игральной кости: $x_1 = 1, x_2 = 2, \dots, x_6 = 6$; Z – число попаданий в мишень при 5 выстрелах: $z_1 = 0, \dots, z_6 = 5$.

Определение 3.3. Непрерывной называется случайная величина, которая может принимать все значения из некоторого промежутка. Обозначение: НСВ.

Пример 3.2. X – дальность полета снаряда; Y – возможный вес яблока; Z – возможный рост человека.

Определение 3.4. Соответствие между отдельными возможными значениями ДСВ и их вероятностями называется законом распределения ДСВ.

Задать ДСВ – перечислить её возможные значения и указать их соответствующие вероятности, например, в виде таблицы:

X	x_1	x_2	...	x_n
p	p_1	p_2		p_n

Так как в одном испытании случайная величина принимает только одно возможное значение, то события $X = x_1, X = x_2, \dots, X = x_n$ образуют полную группу попарно несовместных событий, поэтому сумма их вероятностей равна 1: $p_1 + p_2 + \dots + p_n = 1$.

Пример 3.3. Из каждой сотни лотерейных билетов 50 – не имеют выигрыша, 25 имеют выигрыш 50 р., 15 – 150 р. и 10 – 250 р. Записать закон распределения ДСВ X – стоимости выигравшего билета.

X	0	50	150	250
p	0,5	0,25	0,15	0,1

Одним из наиболее часто встречающихся законов ДСВ является биномиальное распределение.

Пусть производится n независимых испытаний и в каждом из них событие A может появиться с одной и той же вероятностью p (не появиться с вероятностью $q = 1 - p$). В качестве ДСВ X рассмотрим число появлений события A в этих n испытаниях. Очевидно, что $x_1 = 0, x_2 = 1, \dots, x_{n+1} = n$.

Вероятности возможных (k) появлений в n испытаниях даются формулой Бернулли $P_n(k) = C_n^k p^k q^{n-k}$, а соответствующий закон распределения называется биномиальным, так как $C_n^k p^k q^{n-k}$ общий член бинома Ньютона $(p+q)^n$.

X	0	1	2	...	n
p	$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$	$C_n^2 p^2 q^{n-2}$		$C_n^n p^n q^0$

Пример 3.4. Монета брошена 5 раз. Написать закон распределения ДСВ X – числа появлений «герба».

Имеем:

$$P \sum_5^k = C_5^k \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{5-k} = C_5^k \cdot \left(\frac{1}{2}\right)^5 = C_5^k \cdot \frac{1}{32}.$$

$$C_5^0 = C_5^5 = \frac{5!}{0!5!} = 1; \quad C_5^1 = C_5^4 = \frac{5!}{1!4!} = 5; \quad C_5^2 = C_5^3 = \frac{5!}{2!3!} = 10, \text{ поэтому}$$

X	0	1	2	3	4	5
p	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

Если n достаточно велико, а p – достаточно мало, то вместо формулы Бернулли используют формулу Пуассона

$$P_n(k) = \lambda^k e^{-\lambda} / k!,$$

где $\lambda = np$ считается постоянной величиной. Формула Пуассона относится к числу приближённых.

3.2. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ДИСКРЕТНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Хотя закон распределения полностью характеризует ДСВ, на практике часто используют числовые характеристики случайной величины, которые дают её некоторое осреднённое описание, получаемое на основе закона её распределения.

Определение 3.5. Математическим ожиданием ДСВ называется сумма произведений всех её возможных значений на их вероятности:

$$M(X) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = \sum_{i=1}^n x_i p_i.$$

Из данного определения следует, что $M(X)$ есть некоторая постоянная неслучайная величина. Вероятностный смысл $M(X)$ – оно приближенно равно среднему арифметическому значению X (особенно для большого числа испытаний).

Для примера 3.3:

$$M(X) = 0 \cdot 0,5 + 50 \cdot 0,25 + 150 \cdot 0,15 + 250 \cdot 0,1 = 0 + 12,5 + 22,5 + 25 = 60.$$

Шестьдесят рублей – среднее значение выигрыша.

Для примера 3.4:

$$M(X) = 0 \cdot \frac{1}{32} + 1 \cdot \frac{5}{32} + 2 \cdot \frac{10}{32} + 3 \cdot \frac{10}{32} + 4 \cdot \frac{5}{32} + 5 \cdot \frac{1}{32} = \frac{80}{32} = 2,5.$$

«Герб», в среднем, появится 2,5 раза.

Для биномиального распределения $M(X) = np$.

Свойства $M(X)$:

1. $M(C) = C$, $C - \text{const}$.
2. $M(CX) = CM(X)$.
3. $M(X_1 + \dots + X_n) = M(X_1) + \dots + M(X_n)$.
4. Если X_1, X_2, \dots, X_n – независимые случайные величины, то $M(X_1 X_2 \dots X_n) = M(X_1)M(X_2) \dots M(X_n)$.

Определение 3.6. Разность между случайной величиной и её математическим ожиданием называется отклонением: $X - M(X)$.

Определение 3.7. Математическое ожидание квадрата отклонения (случайной величины от её математического ожидания) называется дисперсией или рассеянием:

$$D(X) = M[X - M(X)]^2;$$

$$D[X] = [x_1 - M(X)]^2 p_1 + [x_2 - M(X)]^2 p_2 + \dots + [x_n - M(X)]^2 p_n.$$

На практике чаще используют преобразованную формулу для дисперсии $D(X) = M(X^2) - M^2(X)$.

Для примера 3.3:

$$\begin{aligned} M(X^2) &= 0^2 \cdot 0,5 + 50^2 \cdot 0,25 + 150^2 \cdot 0,15 + 250^2 \cdot 0,1 = \\ &= 0 + 625 + 3375 + 6250 = 10250; \end{aligned}$$

$$M^2(X) = 60^2 = 3600; D(X) = 10250 - 3600 = 6650.$$

Свойства $D(X)$:

1. $D(C) = 0$.
2. $D(CX) = C^2 D(X)$.

3. Если X_1, X_2, \dots, X_n – независимые случайные величины, то

$$D\left(\sum X_i\right) = \sum D(X_i).$$

Существует доказательство, что для биномиального распределения $D(X) = np(1-p) = npq$.

Определение 3.8. Средним квадратическим отклонением случайной величины X называется квадратный корень из её дисперсии: $\sigma(X) = \sqrt{D(X)}$.

3.3. НЕПРЕРЫВНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

3.3.1. Функция распределения

Пусть все значения НСВ X сплошь заполняют отрезок $[a, b]$.

Определение 3.9. Функцией распределения X называется функция $F(x)$, определяющая вероятность того, что X принимает значение, меньшее x :

$$F(x) = p(X < x).$$

Свойства $F(x)$:

1. $0 \leq F(x) \leq 1$.

2. При $x_2 > x_1$ $F(x_2) \geq F(x_1)$.

3. $F(x) = \begin{cases} 0, & \text{если } x \leq a; \\ 1, & \text{если } x \geq b, \end{cases}$ если все значения случайной величины принадлежат отрезку $[a, b]$.

надлежат отрезку $[a, b]$.

Следствия:

1. $P(\alpha \leq X \leq \beta) = F(\beta) - F(\alpha)$.

2. $P(X = x_1) = 0$.

3. Если $-\infty < X < +\infty$, то $\lim_{x \rightarrow -\infty} F(x) = 0$; $\lim_{x \rightarrow +\infty} F(x) = 1$.

3.3.2. Плотность распределения вероятностей

Определение 3.10. Производная от $F(x)$ называется плотностью распределения вероятностей случайной величины X : $f(x) = F'(x)$.

Таким образом, $F(x)$ является первообразной для $f(x)$ и

$$P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f(x) dx = F(\beta) - F(\alpha). \quad \text{В то же время } F(x) = \int_{-\infty}^x f(t) dt.$$

Свойства $f(x)$:

1. $f(x) \geq 0$.

2. $\int_{-\infty}^{\infty} f(x) dx = 1$; $\int_a^b f(x) dx = 1$, если все значения X принадлежат отрезку $[a, b]$.

3.3.3. Числовые характеристики НСВ

$$M(X) = \int_a^b xf(x)dx; \quad D(X) = \int_a^b [x - M(X)]^2 f(x)dx \text{ или}$$

$$D(X) = \int_a^b x^2 f(x)dx - [M(X)]^2; \quad \sigma(X) = \sqrt{D(X)}.$$

Пример 3.5. Плотность распределения вероятностей случайной величины X (м) – рост взрослого жителя города N задана функцией

$$f(x) = \begin{cases} 0, & x < 1,5; \\ -24 \cdot (2x^2 - 7x + 6), & 1,5 \leq x \leq 2; \\ 0, & x > 2. \end{cases}$$

Найти $M(X)$.

$$\begin{aligned} M(X) &= \int_{1,5}^2 -24x(2x^2 - 7x + 6)dx = -24 \cdot \left(2 \frac{x^4}{4} - 7 \frac{x^3}{3} + \frac{6x^2}{2} \right) \Bigg|_{1,5}^2 = \\ &= -24 \cdot \left(0,5 \cdot (16 - 5,0625) - \frac{7}{3} \cdot (8 - 3,375) + 3 \cdot (4 - 2,25) \right) = 1,75. \end{aligned}$$

Таким образом, «средневзвешенный» рост взрослого жителя города N составляет 1,75 м.

$$\begin{aligned} D(X) &= \int_{1,5}^2 x^2 [-24 \cdot (2x^2 - 7x + 6)] dx - (1,75)^2 = \\ &= -24 \cdot \left(\frac{2x^5}{5} - \frac{7x^4}{4} + \frac{6x^3}{3} \right) \Bigg|_{1,5}^2 - 3,0625 = -24 \cdot (0,4 \cdot (32 - 7,59375) - \\ &- 1,75 \cdot (16 - 5,0625) + 2 \cdot (8 - 3,375)) - 3,0625 = 3,075 - 3,0625 = 0,0075. \\ \sigma(X) &= \sqrt{D(X)} \cong 0,087. \end{aligned}$$

Среднеквадратическое отклонение составляет около 9 см.

3.4. НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Определение 3.11. Общим нормальным распределением вероятностей НСВ X называется распределение с плотностью

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (3.1)$$

Отсюда становится ясным вероятностный смысл параметров a и σ нормального распределения. Эти параметры имеют простую геометрическую интерпретацию, для выяснения которой стандартными методами математического анализа исследуют поведение функции $f(x)$:

1. Областью определения функции $f(x)$ является вся числовая ось;
2. Функция $f(x)$ может принимать только положительные значения;
3. $\lim_{x \rightarrow \pm\infty} f(x) = 0$, т.е. ось OX является горизонтальной асимптотой

графика функции;

4. Функция $f(x)$ имеет в точке $x = a$ максимум, равный $\frac{1}{(\sigma \cdot \sqrt{2\pi})}$;

5. График функции $f(x)$ симметричен относительно прямой $x = a$;

6. График функции $f(x)$ в точках $x = a \pm \sigma$ имеет перегиб.

На основании этих свойств можно построить график плотности нормального распределения $f(x)$ (см., например, [1, с. 555]).

Зная $f(x)$, можно найти функцию распределения

$$F(x) = \int_{-\infty}^x f(t) dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt. \quad (3.2)$$

Использование формул (3.1) и (3.2) для практических расчетов затруднительно, так как функции $f(x)$ и $F(x)$ в данном случае являются

трансцендентными, а интеграл $\int e^{-\frac{(x-a)^2}{2\sigma^2}} dx$ относится к «неберущимся».

Решение задач по указанным формулам можно упростить, если от нормального распределения с произвольными параметрами a и σ перейти также к нормальному распределению с параметрами $a = 0$ и $\sigma = 1$.

Определение 3.12. Нормальное распределение с параметрами $a = 0$ и $\sigma = 1$ называется нормированным (стандартным);

его плотность $f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$, а функция распределения

$$F(u) = \int_{-\infty}^u f(t) dt.$$

График функции $f(x)$ такого вида называют стандартной кривой Гаусса; её можно рассматривать как кривую распределения случайной вели-

чины $U = \frac{X - a}{\sigma}$, где X распределена нормально с параметрами a и σ .

Выделяют три, наиболее часто встречающиеся задачи.

1. Вычисление вероятности попадания случайной величины, распределённой нормально, в заданный интервал (α, β) основано на использовании формулы:

$$P(\alpha < X < \beta) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{u_1}^{u_2} e^{-t^2/2} dt = \frac{1}{2} \Phi(u_2) - \frac{1}{2} \Phi(u_1), \quad (3.3)$$

где $u_1 = \frac{\alpha - a}{\sigma}$, $u_2 = \frac{\beta - a}{\sigma}$,

а $\Phi(u) = \frac{2}{\sqrt{2\pi}} \int_0^u e^{-t^2/2} dt$ – функция Лапласа, значения которой можно найти в специальных таблицах.

2. Вероятность того, что случайная величина X , распределённая нормально с параметрами a и σ , отклоняется от своего математического ожидания на малую величину ε , равна:

$$P(|X - a| < \varepsilon) = P(a - \varepsilon < X < a + \varepsilon) = \Phi(\varepsilon/\sigma), \quad (3.4)$$

так как в формуле (3.4):

$$u_1 = (a - \varepsilon - a)/\sigma = -\varepsilon/\sigma, \quad u_2 = (a + \varepsilon - a)/\sigma = \varepsilon/\sigma \quad \text{и}$$

$$\Phi(-\varepsilon/\sigma) = \Phi(\varepsilon/\sigma), \quad \text{то}$$

$$P(a - \varepsilon < X < a + \varepsilon) = \frac{1}{2} \Phi(\varepsilon/\sigma) - \frac{1}{2} \Phi(-\varepsilon/\sigma) = \Phi(\varepsilon/\sigma).$$

3. Обозначим в (3.4) $\varepsilon/\sigma = u$, тогда $\varepsilon = u \cdot \sigma$ и $P(|X - a| < u \cdot \sigma) = \Phi(u)$.

$$\text{При } u = 1: P(|X - a| < \sigma) = \Phi(1) = 0,6837;$$

$$u = 2: P(|X - a| < 2\sigma) = \Phi(2) = 0,9545;$$

$$u = 3: P(|X - a| < 3\sigma) = \Phi(3) = 0,9973.$$

Последнее равенство показывает, что вероятность того, что то или иное значение X попадает в интервал $(a - 3\sigma, a + 3\sigma)$, практически равна 1, т.е. соответствующее событие – достоверное.

В теории вероятностей это обстоятельство формулируется как правило «трёх сигм». Если случайная величина имеет нормальное распределение, то отклонение этой величины от её математического ожидания по абсолютной величине практически не превышает утроенного среднего квадратического отклонения.

Пример 3.6. Случайная величина X распределена нормально с параметрами $a = 8$, $\sigma = 3$.

Найти вероятность $P(12,5 < X < 14)$.

Решение. Используем формулу (3.4), где

$u_1 = (12,5 - 8)/3 = 1,5$; $u_2 = (14 - 8)/3 = 2$, тогда

$$P(12,5 < X < 14) = \frac{1}{2}\Phi(2) - \frac{1}{2}\Phi(1,5) = \frac{1}{2}(0,9545 - 0,8664) = 0,0441.$$

Пример 3.7. Случайная погрешность измерения подчинена нормальному закону распределения с параметрами $a = 0$, $\sigma = 9$ мм. Проводятся три независимых измерения. Найти вероятность того, что погрешность хотя бы одного измерения не превосходит по абсолютной величине 3 мм.

По формуле для $a = 0$, $\sigma = 9$, $\varepsilon = 3$ находим вероятность того, что погрешность измерения в одном испытании не превышает 3 мм. Имеем

$$P(|X| < 3) = \Phi(3/9) \approx \Phi(0,33) = 0,2586.$$

Вероятность того, что эта погрешность превышает 3 мм, равна

$$P(|X| > 3) = 1 - P(|X| < 3) = 0,7414.$$

Вероятность того, что во всех трёх испытаниях погрешность измерения превышает 3 мм, по теореме умножения вероятностей равна произведению вероятностей $[P(|X| > 3)]^3 \approx 0,4075$.

Искомая вероятность равна

$$1 - [P(|X| > 3)]^3 \approx 0,5925.$$

3.5. РАСПРЕДЕЛЕНИЯ χ^2 , СТЬЮДЕНТА И ФИШЕРА

Для того чтобы успешно реализовать методы математической статистики, необходимо иметь представление о распределении некоторых случайных величин, являющихся функциями нормальных величин (распределениях, построенных на основе нормального закона распределения).

3.5.1. Распределение χ^2 (хи-квадрат)

Определение 3.13. Пусть независимые случайные величины U_1, U_2, \dots, U_k являются стандартными нормально распределёнными величинами, т.е. $U_i = N(0, 1)$, $i = 1, 2, \dots, k$.

Распределение случайной величины

$$\chi^2(k) = U_1^2 + U_2^2 + \dots + U_k^2 = \sum_{i=1}^k U_i^2 \quad (3.5)$$

называется распределением хи-квадрат с k степенями свободы, а сама величина (3.5) – величиной хи-квадрат с k степенями свободы. Характерно, что $\chi^2(k) \geq 0$.

Подобно тому, как математическое ожидание a и среднее квадратическое σ являются параметрами нормального закона, так и число k является параметром $\chi^2(k)$ -распределения. Число k – число степеней свободы определяют как разность между числом суммируемых случайных величин и числом линейных связей, ограничивающих свободу этих величин. Так как в (3.5) слагаемые независимы, то число степеней свободы равно числу слагаемых.

Формула функции плотности $\chi^2(k)$ -распределения имеет сложный вид; на практике обычно пользуются специальными таблицами [1, с. 558].

3.5.2. Распределение Стьюдента (t -распределение)

Определение 3.14. Пусть U – стандартная нормально распределённая случайная величина, т.е. $U = N(0, 1)$, а $\chi^2(k)$ – случайная величина, имеющая хи-квадрат распределение с k степенями свободы, причём $\chi^2(k)$ и U – независимые величины. Тогда

$$t(k) = \frac{U}{\sqrt{\chi^2(k)/k}} \quad (3.6)$$

называется t -распределением с k степенями свободы или $t(k)$ -распределением, а сама величина (3.6) – t -величиной с k степенями свободы.

Графики функции плотности распределения $t(k)$ – кривые Стьюдента напоминают график стандартного нормального распределения, в частности, симметричны относительно оси ординат.

В математической статистике используют сведения о таких значениях t_γ величины $t(k)$, при которых вероятность $P(|t(k)| < t_\gamma) = \gamma$. Так как формула плотности $t(k)$ -распределения достаточно сложна, то значения $t(k)$ представлены специальными таблицами [1, с. 557].

С возрастанием числа степеней свободы распределение Стьюдента быстро приближается к нормальному.

3.5.3. Распределение Фишера (F -распределение)

Определение 3.15. Пусть $\chi^2(k_1)$ и $\chi^2(k_2)$ – независимые случайные величины, имеющие χ^2 -распределение соответственно с k_1 и k_2 степенями свободы. Распределение случайной величины

$$F(k_1, k_2) = \frac{\chi^2(k_1)/k_1}{\chi^2(k_2)/k_2} \quad (3.7)$$

называют F -распределением с k_1 и k_2 степенями свободы, а сама величина (3.7) $F(k_1, k_2)$ -величиной.

F -распределение определяется двумя параметрами – числами степеней свободы.

Очевидно, что $F(k_1, k_2) \geq 0$.

В математической статистике используют сведения о таких значениях f_γ величины $F(k_1, k_2)$, при которых вероятность $P(F(k_1, k_2) < f_\gamma) = \gamma$.

Значения $F(k_1, k_2)$ можно найти в специальных таблицах [1, с. 559].

3.6. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

3.6.1. Из пяти человек, среди которых четыре знают английский язык, наудачу отобраны три. Составить закон распределения дискретной случайной величины X – количества «англоговорящих» среди отобранных и найти $M(X)$.

3.6.2. Найти дисперсию случайной величины X , заданной законом распределения

X	–4	2	3	4
p	0,4	0,3	0,1	0,2

3.6.3. Случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0, & x \leq 2; \\ 0,5x - 1, & 2 < x \leq 4; \\ 1, & x > 4. \end{cases}$$

Найти вероятность того, что X примет значения: а) меньше 0,2;

б) менее 3; в) не менее 2,5 и не более 3,5.

3.6.4. Случайная величина X задана на интервале (0, 5) плотностью распределения $f(x) = 2x/25$; вне этого интервала $f(x) = 0$. Найти дисперсию X .

3.6.5. Задан закон распределения случайной величины X

X	14	16	18	20
p	0,1	0,2	0,3	0,4

Найти $M(X)$ и $D[X]$.

3.6.6. Дана плотность распределения вероятностей $f(x)$ случайной величины X

$$f(x) = \begin{cases} 0, & x \leq 0; \\ \alpha x^2, & 0 < x \leq 2; \\ 0, & x > 2. \end{cases}$$

Найти параметр α , математическое ожидание $M(X)$, дисперсию $D[X]$, функцию распределения $F(x)$, вероятность выполнения неравенства $\alpha < X < 2$.

Кроме того, можно закрепить полученные навыки, решив задачи № 171, 175, 180, 216, 260, 270, 287 из [3].

4. ДВУМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

I. Учебные цели. В результате изучения материала студенты должны получить представление о двумерных случайных величинах: способах задания закона их распределения, получения закона распределения компонентов двумерной случайной величины, условных распределений случайных величин, числовых характеристиках, методах оценки взаимосвязанных случайных величин – компонентов.

II. Формируемые компетенции. Формирование математической культуры, развитие способностей к обобщению, творческих способностей.

III. Введение в тему.

Кроме одномерных случайных величин, возможные значения которых определяются одним числом, существуют также величины, возможные значения которых определяются несколькими числами. Двумерную случайную величину обозначают через (X, Y) ; каждая из величин X и Y называется компонентой (составляющей). Обе величины X и Y рассматриваются одновременно, образуют систему двух случайных величин. Например, при одновременном подбрасывании двух игральных костей образующееся при этом двузначное число – двумерная случайная величина. Совокупность «рост и вес» случайно встретившегося человека – двумерная случайная величина.

4.1. ЗАКОН РАСПРЕДЕЛЕНИЯ ДВУМЕРНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ

Определение 4.1. Законом распределения двумерной случайной величины (X, Y) называется множество возможных пар чисел (x_i, y_j) и их вероятностей $p(x_i, y_j)$. Двумерную случайную величину можно трактовать как случайную точку $A(X, Y)$ на координатной плоскости.

Закон распределения двумерной (дискретной) случайной величины обычно задаётся в виде таблицы, в строках которой указаны возможные значения x_i случайной величины X , а в столбцах – возможные значения y_j случайной величины Y , на пересечениях строк и столбцов указаны соответствующие вероятности p_{ij} :

Таблица 4.1

X/Y	y_1	y_2	y_3	...	y_n	
x_1	p_{11}	p_{12}	p_{13}	...	p_{1n}	$p(x_1)$
x_2	p_{21}	p_{22}	p_{23}	...	p_{2n}	$p(x_2)$
...
x_m	p_{m1}	p_{m2}	p_{m3}	...	p_{mn}	$p(x_m)$
	$p(y_1)$	$p(y_2)$	$p(y_3)$...	$p(y_n)$	1

Из этой таблицы можно найти законы распределения каждой из случайных компонент. Например, вероятность того, что случайная величина X примет значение x_k , равна, согласно теореме сложения вероятностей независимых событий,

$$p(x_k) = p_{k1} + p_{k2} + \dots + p_{kn}, \quad k = 1, 2, \dots, m.$$

Таким образом, для нахождения вероятности $p(x_k)$ нужно просуммировать все n вероятностей по k -й строке. Аналогично вероятность того, что Y примет значение y_r , равна:

$$p(y_r) = p_{1r} + p_{2r} + \dots + p_{mr} = \sum_{i=1}^m p_{ir}, \quad r = 1, 2, \dots, n.$$

В таблице эти значения указаны в крайнем правом столбце $p(x_k)$ и крайней нижней строке $p(y_r)$.

Так как события $[(X = x_i)(Y = y_j)]$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$), состоящие в том, что случайная величина X примет значение x_i , а Y – значение y_j , несовместны и единственно возможны, т.е. образуют полную группу, то

$$\sum_{i=1}^m \sum_{j=1}^n p_{ij} = 1.$$

Если зафиксировать значение одного из аргументов, например, положить $Y = y_j$, то полученное распределение случайной величины X называется условным распределением X при условии $Y = y_j$. Вероятности $p_j(x_i)$ этого распределения будут условными вероятностями события $X = x_i$, найденными в предположении, что событие $Y = y_j$ произошло и $p_j(x_i) = p_{ij}/p_j$.

Аналогично условное распределение случайной величины Y при условии $X = x_i$ задается с помощью условной вероятности $p_i(y_j) = p_{ij}/p_i$.

Пример 4.1. Задано распределение двумерной случайной величины:

X/Y	1	2	3
1	0,1	0,15	0,12
2	0,2	0,22	0,21

Найти распределение X и Y .

Решение. Здесь возможные значения X : $x_1 = 1$ и $x_2 = 2$.

Тогда $p(x_1) = 0,1 + 0,15 + 0,12 = 0,37$.

$p(x_2) = 0,2 + 0,22 + 0,21 = 0,63$.

Имеем:

X	1	2
p_i	0,37	0,63

Аналогично находим:

$p(y_1 = 1) = 0,1 + 0,2 = 0,3$;

$p(y_2 = 2) = 0,15 + 0,22 = 0,37$;

$p(y_3 = 3) = 0,12 + 0,21 = 0,33$

Y	1	2	3
p_j	0,3	0,37	0,33

Используя известные для одномерной случайной величины формулы, можно найти числовые характеристики – математические ожидания $a_x = M(X)$, $a_y = M(Y)$ и дисперсии $\sigma_x^2 = D(X)$ и $\sigma_y^2 = D(Y)$.

$$a_x = M[X] = 1 \cdot 0,37 + 2 \cdot 0,63 = 1,63; \quad M[X^2] = 1 \cdot 0,37 + 4 \cdot 0,63 = 2,89;$$

$$D[X] = 2,89 - 1,63^2 = 0,233; \quad \sigma_x = 0,483;$$

$$a_y = M[Y] = 1 \cdot 0,3 + 2 \cdot 0,37 + 3 \cdot 0,33 = 2,03;$$

$$M[Y^2] = 1 \cdot 0,3 + 4 \cdot 0,37 + 9 \cdot 0,33 = 4,75;$$

$$D[Y] = 4,75 - 2,03^2 = 0,629; \quad \sigma_y = 0,793.$$

Пример 4.2. Закон распределения дискретной двумерной случайной величины (X, Y) задан в таблице:

Таблица 4.2

$X \backslash Y$	-1	0	1	2
1	0,10	0,25	0,30	0,15
2	0,10	0,05	0,00	0,05

Найти:

- а) законы распределения одномерных случайных величин X и Y ;
 б) условные законы распределения случайной величины X при условии $Y = 2$ и случайной величины Y при условии $X = 1$.

Решение.

а) Случайная величина X может принимать значения:

$X = 1$ с вероятностью $p_1 = 0,10 + 0,25 + 0,30 + 0,15 = 0,8$;

$X = 2$ с вероятностью $p_2 = 0,10 + 0,05 + 0,00 + 0,05 = 0,2$,

т.е. её закон распределения

X	1	2
p	0,8	0,2

Аналогично закон распределения

Y	-1	0	1	2
p	0,2	0,3	0,3	0,2

б) Условный закон распределения X при условии, что $Y = 2$, получим, если вероятности p_{ij} , стоящие в последнем столбце таблицы 4.2, разделим на их сумму, т.е. $p(Y = 2) = 0,2$. Получим:

$X_{Y=2}$	x_i	1	2
	$p_j(x_i)$	0,75	0,25

Аналогично для получения условного закона распределения Y при условии $X = 1$ вероятности p_{ij} , стоящие в первой строке таблицы 4.2, делим на их сумму, т.е. на $P(X = 1) = 0,8$. Получим:

$Y_{X=1}$	y_j	-1	0	1	2
	$p_i(y_j)$	0,125	0,3125	0,375	0,1875

4.2. КОВАРИАЦИЯ И КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Математические ожидания и дисперсии случайных величин X и Y недостаточно полно характеризуют двумерную случайную величину (X, Y) , так как не выражают степени зависимости её составляющих X и Y . Эту роль выполняют ковариация и коэффициент корреляции.

Определение 4.2. Ковариацией (или корреляционным моментом) K_{xy} случайных величин X и Y называется математическое ожидание произведения отклонений этих величин от своих математических ожиданий, т.е.

$$K_{xy} = M[(X - M(X))(Y - M(Y))] = M[(X - a_x)(Y - a_y)].$$

Из этого определения следует, что $K_{xy} = K_{yx}$, кроме того

$$K_{xx} = M[(X - M(X))^2],$$

т.е. ковариация случайной величины с самой собой есть её дисперсия.

$$\text{Для дискретных случайных величин } K_{xy} = \sum_{i=1}^m \sum_{j=1}^n (x_i - a_x)(y_j - a_y) p_{ij}.$$

Из свойств математического ожидания следует, что

$$K_{xy} = M(XY) - M(X) \cdot M(Y) = \sum_{i=1}^m \sum_{j=1}^n x_i y_j p_{ij} - a_x \cdot a_y.$$

Корреляционный момент двух независимых величин равен нулю, в противном случае эти величины являются зависимыми.

Из определения корреляционного момента следует, что его размерность равна произведению размерностей X и Y . Это обстоятельство затрудняет сравнение корреляционных моментов различных систем случайных величин. Для устранения этого недостатка вводят безразмерную числовую характеристику – коэффициент корреляции.

Определение 4.3. Коэффициентом корреляции случайных величин X и Y называют отношение их корреляционного момента к произведению средних квадратических отклонений этих величин.

$$r_{xy} = K_{xy} / (\sigma_x \cdot \sigma_y) = K_{xy} / \sqrt{D[X] \cdot D[Y]}.$$

Свойства коэффициента корреляции:

1. Абсолютная величина коэффициента корреляции не превосходит 1:

$$-1 \leq r_{xy} \leq 1.$$

2. Если случайные величины независимы, то их коэффициент корреляции равен нулю, а сами величины называются (являются) некоррелируемыми.

3. Если $|r_{xy}| = 1$, то между X и Y существует линейная функциональная зависимость.

Пример 4.3. Найти корреляционный момент и коэффициент корреляции двух случайных величин X и Y , распределения которых заданы в примере 4.1.

$$\begin{aligned} K_{xy} &= \sum_{i=1}^2 \sum_{j=1}^3 x_i y_j p_{ij} - M[X] \cdot M[Y] = 1(1 \cdot 0,1 + 2 \cdot 0,15 + 3 \cdot 0,12) + \\ &+ 2(1 \cdot 0,2 + 2 \cdot 0,22 + 3 \cdot 0,21) - 1,63 \cdot 2,03 = 0,1 + 0,3 + 0,36 + 0,4 + \\ &+ 0,88 + 1,26 - 3,3098 = -0,0098. \end{aligned}$$

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y} = \frac{-0,0098}{0,483 \cdot 0,793} = -0,0256.$$

В данном случае коэффициент корреляции близок к нулю; это означает, что случайные величины X и Y слабокоррелированы.

4.3. ЛИНЕЙНАЯ РЕГРЕССИЯ

Пусть зависимые случайные величины X и Y являются компонентами двумерной случайной величины (X, Y) . В этом случае возможно приближённое представление величины Y в виде линейной функции величины X :

$$Y \approx q(x) = a + bX, \quad (4.1)$$

где a и b – параметры, подлежащие определению.

Определение 4.4. Функция (4.1) называется наилучшим приближением в смысле метода наименьших квадратов, если математическое ожидание $M[Y - q(X)]^2$ принимает наименьшее возможное значение. Функция $q(x)$ называется среднеквадратической регрессией Y на X .

Существует доказательство, что линейная средняя квадратическая регрессия Y на X имеет вид:

$$q(X) = a_y + r_{xy} \frac{\sigma_y}{\sigma_x} (X - a_x),$$

где r_{xy} – коэффициент корреляции, $a_x = M(X)$, $a_y = M(Y)$.

Коэффициент $b = r_{xy} \cdot \frac{\sigma_y}{\sigma_x}$ этого уравнения называется коэффициентом регрессии Y на X , а прямая $y - a_y = b(x - a_x)$ – прямой среднеквадратической регрессии Y на X .

Поскольку зависимость (4.1) является приближённой, то существует оценка погрешности этого приближения, называемая остаточной дисперсией:

$$\varepsilon^2 = \sigma_y^2 (1 - r_{xy}^2).$$

Пример 4.4. Найти линейную среднюю квадратическую регрессию и остаточную дисперсию случайной величины Y на случайную величину X по данным примеров № 4.1 и 4.3.

Имеем:

$$y - 2,03 = (-0,0256) \cdot \frac{0,793}{0,483} (x - 1,63)$$

или $y - 2,03 = -0,042x + 0,0685$ – уравнение регрессии и $\varepsilon^2 = 0,629(1 - 0,000655) = 0,6286$; $\varepsilon = 0,793$.

4.4. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

4.4.1. Двумерная случайная величина (X, Y) определена следующим образом. Если при подбрасывании игральной кости выпадает чётное число очков, то $X = 1$, противном случае $X = 0$; $Y = 1$, когда число очков кратно трём, в противном случае $Y = 0$. Найти:

а) законы распределения двумерной случайной величины (X, Y) и её одномерных составляющих;

б) числовые характеристики случайных величин.

4.4.2. Дано распределение двумерной дискретной случайной величины (X, Y) .

$X \backslash Y$		2	4	6
1		0,2	0,1	0
3		0,1	0,2	0,1
5		0	0,1	0,2

Найти ковариационный момент K_{xy} и коэффициент корреляции X и Y .

4.4.3. Пусть двумерная случайная величина (X, Y) задана законом распределения

$X \backslash Y$		1	2	3	4
10		0,2	0,02	0,01	0
20		0,03	0,3	0,02	0
30		0,02	0,1	0,2	0,1

Найти:

а) законы распределения величин X и Y ;

б) условное распределение Y при условии $X = 30$;

в) коэффициент корреляции между величинами X и Y .

4.4.4. Двумерная дискретная случайная величина (X, Y) задана законом распределения

$X \backslash Y$		6	5	4	3
0,1		0,2	0,1	0,1	0
0,2		0,1	0,2	0,15	0
0,3		0	0	0,05	0,1

Найти условное распределение величины X при условии $Y = 4$. Являются ли величины X и Y независимыми?

4.4.5. Задан совместный закон распределения величин X и Y :

$X \backslash Y$	-2	-1	0	1	2
-1	1/8	1/16	1/8	1/16	1/8
1	1/8	1/16	1/8	1/16	1/8

Являются ли величины X и Y независимыми?

4.4.6. Двумерная случайная величина (X, Y) имеет закон распределения

$X \backslash Y$	15	20	25	30
-1	0,05	0,05	0,05	0
0	0,05	0,1	0,05	0
1	0	0,05	0,2	0,05
2	0	0	0,05	0,3

Найти условное распределение Y при условии $X > 0$.

Кроме того, можно закрепить полученные навыки, решив задачи № 421, 422 из [3], № 5.10 и 5.14 из [1].

5. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. ОСНОВНЫЕ ПОНЯТИЯ

I. Учебные цели. Познакомить студентов с основными понятиями математической статистики, задачами, которые решаются в изучаемом курсе.

В результате изучения материала студенты должны иметь представление о способах сбора статистических данных, о способах их представления в удобной для статистической обработки форме (вариационный ряд, статистическое распределение выборки, полигон, гистограмма, эмпирическая функция распределения), уметь осуществлять наглядное представление статистического распределения, находить числовые характеристики вариационных рядов.

II. Формирование компетенций. Развитие математической культуры, совершенствование общей культуры мышления, развитие способностей применять методы математической статистики в профессиональной деятельности, умение лаконично и точно формулировать определения, давать графическую интерпретацию математических зависимостей.

III. Введение в тему.

Математическая статистика является частью общей прикладной математической дисциплины «Теория вероятностей и математическая статистика», однако задачи, решаемые ею, носят специфический характер.

Если теория вероятностей исследует явления, полностью заданные их моделью, то в математической статистике вероятностная модель определена с точностью до неизвестных параметров. Отсутствие сведений о параметрах компенсируется «пробными» испытаниями, на основе которых и восстанавливается недостающая информация. Цель математической статистики состоит в создании методов сбора и обработки статистических данных для получения научных и практических выводов.

Вопросы для контроля усвоения излагаемого материала:

1. Что является предметом изучения математической статистики?
2. Что такое статистические данные?
3. Какие основные задачи решает математическая статистика?
4. Что такое генеральная и выборочная совокупности?
5. Какие существуют способы образования выборки?
6. Что такое вариационный ряд и статистическое распределение выборки?
7. Графики статистического распределения: полигон и гистограмма.
8. Как задаётся эмпирическая функция распределения?
9. Что такое выборочная средняя и какие у неё свойства?
10. Что такое выборочная дисперсия и какие у неё свойства?

5.1. СТАТИСТИЧЕСКИЕ ДАННЫЕ. ГЕНЕРАЛЬНАЯ И ВЫБОРОЧНАЯ СОВОКУПНОСТИ

Статистической совокупностью называется совокупность предметов или явлений, объединённых каким-либо признаком. Результатом наблюдений над статистической совокупностью являются статистические данные – данные о количестве элементов в какой-либо совокупности, обладающих определённым свойством.

Например:

- количество дождливых дней в году;
- количество жителей города в возрасте 20 лет;
- количество дубов на территории Тамбовской области.

Обработка статистических данных методами математической статистики приводит к установлению определённых закономерностей, присущих массовым явлениям.

Статистические данные, как правило, представляют собой ряд значений $\{x_1, x_2, \dots, x_n\}$ некоторой случайной величины X . Её исследование начинается с обработки этого ряда значений. Затем строятся функции, характеризующие случайную величину. Эти функции называются статистиками. Статистика – это функция, которая набору значений случайной величины сопоставляет по некоторому правилу действительное число.

Как правило, статистику T рассматривают как функцию от случайных величин X_1, X_2, \dots, X_n , имеющих такое же распределение, как и случайная величина X :

$$T = T(X_1, X_2, \dots, X_n).$$

В этом случае X рассматривается как набор одинаковых случайных величин $\{X_1, X_2, \dots, X_n\}$ и статистика становится случайной величиной; изучение её распределения сводится к изучению распределения самой случайной величины X .

Простейшей статистикой является, например, среднее значение одинаково распределённых случайных величин.

Можно выделить основные задачи математической статистики, которые решаются в изучаемом курсе:

1. Поиск способов сбора и группировки статистических данных, полученных в результате наблюдений или эксперимента.

2. Разработка методов анализа статистических данных в зависимости от целей исследования:

- оценка неизвестной вероятности события (по сути использование статистического определения вероятности);
- оценка неизвестной функции распределения;
- оценка параметров (известного) распределения;
- оценка степени зависимости случайных величин;
- проверка статистических гипотез о виде неизвестного распределения или о величине параметров распределения, вид которого известен.

Генеральной совокупностью называется совокупность объектов или наблюдений, все элементы которой подлежат изучению при статистическом анализе.

Генеральная совокупность может быть конечной или бесконечной. Число объектов в генеральной совокупности называется её объемом.

Изучение всего набора элементов генеральной совокупности не всегда бывает возможным (долговечность электролампочки, урожайность зерновых культур и т.п.); в этом случае рассматривают некоторую часть генеральной совокупности и называют выборочной совокупностью (или выборкой).

Задача математической статистики – по результатам изучения свойств выборки «спроектировать» свойства генеральной совокупности. Для того, чтобы по выборке можно было адекватно судить об изучаемой величине, она должна быть представительной (репрезентативной); это условие обеспечивается случайностью её элементов: все элементы генеральной совокупности должны иметь одинаковую вероятность попадания в выборку.

Здесь мы имеем дело с решением первой задачи математической статистики, сформулированной ранее – поиск способов сбора и группировки статистических данных.

Различают такие способы образования выборки, как:

1) повторная выборка, когда каждый элемент, случайно отобранный и исследованный, возвращается в генеральную совокупность и может быть отобран повторно;

2) бесповторная выборка, когда отобранный элемент не возвращается в генеральную совокупность.

Каждый из этих способов, в свою очередь, может осуществляться в виде чисто случайной, механической, типической и серийной выборки. Их подробное описание дано в учебнике [1].

5.2. ВАРИАЦИОННЫЙ РЯД И ЕГО ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ

Пусть из генеральной совокупности осуществлена выборка $\{x_1, x_2, \dots, x_n\}$ объёма n . Элементы этой выборки (варианты) представляют собой значения случайной величины X . Если они проранжированы по возрастанию, то такое представление называют рядом вариант или вариационным рядом.

Частотой варианты x_i называют число m_i , показывающее, сколько раз эта варианта встречается в выборке. Относительной частотой (долей варианты) называют число $w_i = \frac{m_i}{n}$.

Количество вариант m_x , значения которых меньше некоторого числа x , называют накопленной частотой $m_x = \sum_{x_i < x} m_i$.

Статистическим распределением выборки (в некоторых литературных источниках вариационным рядом) называют ряд вариант, расположенных в порядке возрастания их значений, с соответствующими им частотами (относительными частотами).

Вариационные ряды бывают дискретными и интервальными. Вариационный ряд называется дискретным, если он представляет собой выборку значений дискретной величины, и интервальным, если представляет собой выборку значений непрерывной величины.

Пример 5.1. Случайная величина X – экзаменационная оценка. Очевидно, что соответствующий вариационный ряд 2, 3, 4, 5.

Для построения интервального вариационного ряда множество значений вариант, заключённых на интервале $[a_j, a_{j+1}]$, разбивают на k полуинтервалов $[a_j, a_{j+1})$ ($j = \overline{1, k}$), последний из которых интервал $[a_k, a_{k+1}]$, т.е. производят их группировку (сгруппированные данные).

Если варианта находится на границе интервала, то её приравнивают к правому интервалу.

Пример 5.2. Случайная величина Y – рост взрослого человека; рекомендуемый диапазон значений от 154 до 195 см. Рекомендуемое количество интервалов k выбирается по формуле Стерджерса:

$$k = 1 + 1,4 \ln d,$$

где $d = a_{k+1} - a_1$.

Если мы считаем единицей измерения роста сантиметр (см), то $d = 196 - 154 = 42$ (см), а $k = 1 + 1,4 \ln 42 \cong 1 + 1,4 \cdot 3,7 \approx 6$.

Длина каждого малого интервала равна:

$$\Delta = \frac{d}{k} = \frac{196 - 154}{6} = 7 \text{ (см)}.$$

Интервальный вариационный ряд имеет вид:

[154, 161); [161, 168); [168, 175); [175, 182); [182, 189); [189, 196].

Зачастую «опорными точками» служат средние на интервалах значения вариант c_i , которые подсчитываются как среднеарифметические их граничных (конечных) значений: $c_i = \frac{1}{2}(a_i + a_{i+1})$, $i = \overline{1, k}$.

Пусть в *примере 5.1.* экзамен сдавало 30 студентов, из них получило оценку «2» – 3 человека, «3» – 13 человек, «4» – 8 человек и «5» – 6 человек, тогда соответствующее статистическое распределение выборки имеет вид:

Варианты (оценки)	2	3	4	5
Частоты (кол-во студентов)	3	13	8	6

Пусть в *примере 5.2* объём выборки составил 100 человек; из них (по росту) в первый интервал попало 8 человек, во второй – 14 человек, в третий – 26 человек, в четвертый – 22 человека, в пятый – 15 человек, в шестой – 9 человек и в седьмой – 6 человек, тогда соответствующее статистическое распределение выборки имеет вид:

Варианты (рост)	[154, 161)	[161, 168)	[168, 175)	[175, 182)	[182, 189)	[189, 196]
Средние значения	157,5	164,5	171,5	178,5	185,5	192,5
Частоты (кол-во чел.)	8	19	26	22	19	6

Для наглядности представления статистического распределения используются различного рода графики: полигон и гистограмму.

Полигон (частот, относительных частот) используется в случае дискретного вариационного ряда и представляет собой ломаную, соединяющую точки плоскости с координатами (x_i, m_i) (или (x_i, w_i) , $i = \overline{1, n}$; n – количество вариант).

Для интервального ряда также строится полигон, только его ломаная проходит через точки (c_i, m_i) , где c_i – средние на интервалах значения.

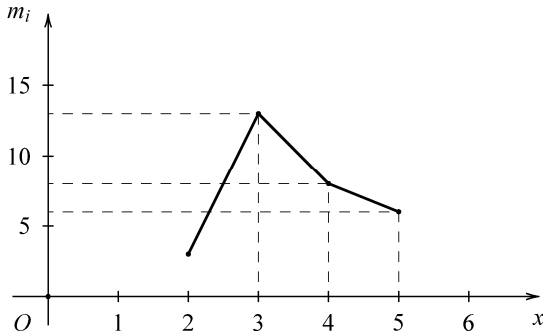


Рис. 5.1

Так, для *примера 5.1* полигон частот изображён на рис. 5.1.

Гистограмма служит для представления только интервальных вариационных рядов и имеет вид ступенчатой фигуры, состоящей из прямоугольников с основаниями, равными длине интервалов Δ и высотами, равными $W_i = \frac{M_i}{\Delta}$, $i = \overline{1, k}$, где M_i – сумма частот вариантов, попавших в i -й интервал, $\frac{M_i}{\Delta}$ – плотность частоты. Таким образом, площадь каждого прямоугольника равна $\Delta \cdot \frac{M_i}{\Delta} = M_i$ – сумме частот.

Для *примера 5.2* гистограмма имеет вид, представленный на рис. 5.2. Здесь $W_1 = 8/7$; $W_2 = 19/7$; $W_3 = 26/7$; $W_4 = 22/7$; $W_5 = 19/7$; $W_6 = 6/7$.

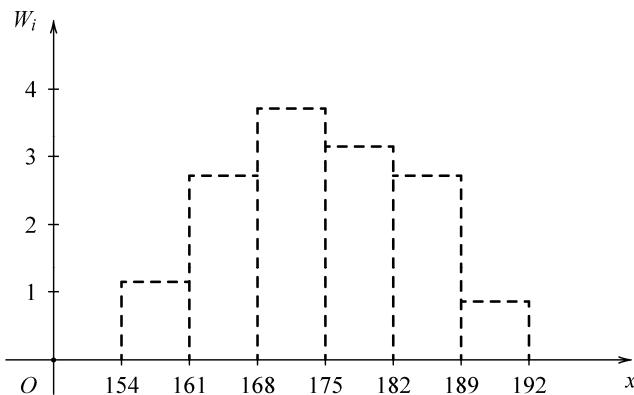


Рис. 5.2

5.3. ЭМПИРИЧЕСКАЯ ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

Эмпирической функцией распределения $F_n(x)$ называется функция, значение которой в точке x равно относительной накопленной частоте

$$F_n(x) = w_x = \frac{m_x}{n}, \quad n - \text{объём выборки, } m_x = \sum_{x_i < x} m_i.$$

Для *примера* 5.1 (дискретный вариационный ряд) эту функцию можно записать в виде:

$$F_n(x) = \begin{cases} 0, & x < 2, \\ 3/30, & 2 \leq x < 3, \\ 16/30, & 3 \leq x < 4, \\ 24/30, & 4 \leq x < 5, \\ 1, & x \geq 5 \end{cases}$$

и построить соответствующий график.

Для интервального ряда указываются не конкретные значения вариантов, а только их частоты на интервалах. В этом случае эмпирическая функция распределения определена только на концах интервалов; её можно изобразить ломаной, проходящей через точки $(a_i, F_n(a_i))$, $i = \overline{1, k}$.

Для *примера* 5.2 это будут точки:

(154; 0); (161; 0,08); (168; 0,27); (175; 0,53); (182; 0,75); (189; 0,94); (196; 1).

5.4. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ ВАРИАЦИОННЫХ РЯДОВ

Как следует из материала прошлого параграфа, вариационные ряды (выборки) можно охарактеризовать или с помощью статистического распределения, или с помощью эмпирической функции. На практике бывает достаточно иметь характеристики вариационных рядов в виде отдельных чисел, а именно: выборочной средней, выборочной дисперсии, выборочного среднеквадратического отклонения.

Пусть дискретный вариационный ряд задан статистическим распределением:

Варианты	x_1	x_2	...	x_k
Частоты	m_1	m_2	...	m_k

$$\sum_{i=1}^k m_i = n - \text{объём выборки, } k - \text{число вариантов.}$$

Выборочным средним называется величина

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i. \quad (5.1)$$

Если статистические данные не являются сгруппированными, т.е. $m_1 = m_2 = \dots = m_k = 1$, то выборочное среднее есть не что иное, как среднее арифметическое значений вариант $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

Формулой (5.1) можно пользоваться и для характеристики интервального вариационного ряда в виде:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k c_i m_i,$$

где c_i – середина i -го интервала; m_i – сумма частот вариант, попавших в i -й интервал; k – число интервалов.

Свойства выборочной средней аналогичны свойствам математического ожидания случайной величины (в теории вероятностей). Укажем одно из них, необходимое для дальнейшей работы: если вариационный ряд состоит из нескольких групп, то общая выборочная средняя равна:

$$\bar{x} = \sum_{i=1}^l \bar{x}_i \cdot \frac{n_i}{n}, \quad (5.2)$$

где \bar{x}_i – групповые средние; n_i – объёмы групп; l – число групп.

Пример 5.3. Дано распределение признака X

x_i	2	3	4	8	9	12
m_i	1	2	2	1	2	2

, $\sum m_i = 10$.

Общее выборочное среднее

$$\bar{x} = \frac{1}{10} \sum_{i=1}^6 x_i m_i = 0,1(2 + 6 + 8 + 8 + 18 + 24) = 6,6.$$

Выделим две группы вариант: чётных и нечётных

Группа 1					Группа 2				
	x_i	2	4	8	12		x_i	3	9
	m_i	1	2	1	2	;	m_i	2	2

для которых групповые выборочные средние: $\bar{x}_1 = 7$; $\bar{x}_2 = 6$.

По формуле (5.2)

$$\bar{x} = \frac{6}{10} \cdot 7 + \frac{4}{10} \cdot 6 = \frac{42 + 24}{10} = 6,6, \text{ что и требовалось получить.}$$

Выборочной дисперсией называется среднее арифметическое квадратов отклонений вариант от их выборочной средней:

$$S^2 = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^2 \cdot m_i.$$

Если $m_i = 1, i = \overline{1, n}$, то

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (5.3)$$

Для интервального вариационного ряда $S^2 = \frac{1}{n} \sum_{i=1}^k (c_i - \bar{x})^2$, где c_i – середина i -го интервала.

Для практических вычислений S^2 более удобной является формула $S^2 = \bar{x}^2 - (\bar{x})^2$, где \bar{x}^2 – выборочная средняя квадратов вариационного ряда.

Выборочное среднее квадратическое отклонение определяется как квадратный корень из дисперсии: $\sigma = \sqrt{S^2} = S$.

Для выборочной дисперсии справедливо свойство, которое лежит в основе раздела математики, называемого дисперсионный анализ, и гласит о том, что если вариационный ряд состоит из нескольких групп, то общая дисперсия равна сумме средней групповых дисперсий и межгрупповой дисперсии.

Пусть варианты выборки имеют обозначение x_{ij} , где $i = 1, 2, \dots, l$ – номер группы; $j = 1, 2, \dots, k_i$ – номер варианты в i -й группе; m_{ij} – соответствующая этой варианте частота; \bar{x}_i – групповые средние; \bar{x} – общая выборочная средняя; n_i – объём i -й группы; n – объём выборки.

Тогда

$$S_0^2 = \sum_{i=1}^l \sum_{j=1}^{k_i} (x_{ij} - \bar{x})^2 \frac{m_{ij}}{n} - \text{общая выборочная дисперсия;}$$

$$\delta^2 = \sum_{i=1}^l (\bar{x}_i - \bar{x})^2 \frac{n_i}{n} - \text{межгрупповая дисперсия;}$$

$$S_i^2 = \sum_{j=1}^{k_i} (x_{ij} - \bar{x}_i)^2 \frac{m_{ij}}{n_i} - \text{групповые дисперсии, а}$$

$$\bar{S}_{(i)}^2 = \sum_{i=1}^l S_i^2 \frac{n_i}{n} \text{ – их средняя и}$$

$$S_0^2 = \bar{S}_{(i)}^2 + \delta^2 \text{ – свойство дисперсии.} \quad (5.4)$$

Формула (5.4) может быть получена методом «разложения суммы квадратов» (см. [1, с. 286]).

Возвращаясь к примеру 5.3, найдём:

$$S_0^2 = \frac{1}{10} [(2-6,6)^2 \cdot 1 + (3-6,6)^2 \cdot 2 + (4-6,6)^2 \cdot 2 + (8-6,6)^2 \cdot 1 +$$

$$+ (9-6,6)^2 \cdot 2 + (12-6,6)^2 \cdot 2] = 13,24;$$

$$\bar{S}_{(i)}^2 = \frac{1}{10} [(2-7)^2 \cdot 1 + (4-7)^2 \cdot 2 + (8-7)^2 \cdot 1 + (12-7)^2 \cdot 2 +$$

$$+ (3-6)^2 \cdot 2 + (9-6)^2 \cdot 2] = 13;$$

$$\delta^2 = (7-6,6)^2 \cdot \frac{6}{10} + (6-6,6)^2 \cdot \frac{4}{10} = 0,24.$$

Действительно, $S_0^2 = \bar{S}_{(i)}^2 + \delta^2$.

Пример 5.4. Анализируется успеваемость по математике на учебном курсе. В выборку попали две студенческие группы (*A* и *B*) численностью 23 и 27 человек.

Экзамен был организован в форме тестирования по 100-балльной шкале. Результат – выборочные данные были сгруппированы в интервалы по 20 баллов с минимальным баллом 20 и представлены в таблице:

Баллы	Группа <i>A</i>	Группа <i>B</i>
[100 – 80)	4	3
[80 – 60)	8	10
[60 – 40)	9	11
[40 – 20]	2	3

Найти среднее выборочное значение количества баллов, набранных студентами групп *A* и *B*, а также дисперсии относительно средних.

Решение.

Обозначим: \bar{x}_1 – выборочное среднее набранных баллов в группе *A*; \bar{x}_2 – выборочное среднее набранных баллов в группе *B*; S_1^2 – выборочная дисперсия по группе *A*; S_2^2 – выборочная дисперсия по группе *B*.

По формулам (5.1) и (5.3) для интервального вариационного ряда

$$\bar{x}_1 = \frac{1}{23}(90 \cdot 4 + 70 \cdot 8 + 50 \cdot 9 + 30 \cdot 2) = \frac{1}{23} \cdot 1430 = 62,2;$$

$$\bar{x}_2 = \frac{1}{27}(90 \cdot 3 + 70 \cdot 10 + 50 \cdot 11 + 30 \cdot 3) = \frac{1}{27} \cdot 1610 = 59,6;$$

$$S_1^2 = \frac{1}{23}[(90 - 62,2)^2 \cdot 4 + (70 - 62,2)^2 \cdot 8 + (50 - 62,2)^2 \cdot 9 + (30 - 62,2)^2 \cdot 2] =$$

$$= \frac{1}{23}[3091,36 + 486,72 + 1339,56 + 2073,68] = 303,97;$$

$$S_2^2 = \frac{1}{27}[(90 - 59,6)^2 \cdot 3 + (70 - 59,6)^2 \cdot 10 + (50 - 59,6)^2 \cdot 11 + (30 - 59,6)^2 \cdot 3] =$$

$$= \frac{1}{27}[2772,48 + 1081,6 + 1013,76 + 2628,48] = 277,64.$$

5.5. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

5.5.1. В течение месяца (30 дней) государственная автоинспекция зарегистрировала 77 аварий. Распределение количества аварий по числу дней представлено следующей таблицей:

Кол-во аварий	0	1	2	3	4	5
Число дней	6	4	5	3	6	6

Найти выборочное среднее аварийности в день и среднее квадратическое отклонение от этого среднего.

Построить полигон частот и относительных частот.

5.5.2. В течение 1 часа (60 минут) в офис компании поступило 50 звонков. Хронология звонков выглядит следующим образом:

Временной интервал	[0, 10)	[10, 20)	[20, 30)	[30, 40)	[40, 50)	[50, 60]
Количество звонков	15	7	10	4	6	8

Построить гистограмму частот по данному распределению выборки.

Найти среднее выборочное времени звонков.

5.5.3. В результате пяти последовательных измерений нижнего кровяного давления получены следующие результаты (*mmHg*): 82; 84; 93; 95; 96.

Найти отношение среднего квадратического отклонения к выборочной средней (в %) – коэффициент вариации.

5.5.4. Вариационный ряд состоит из двух частей: четырех элементов x_i и четырех элементов x_j , которые вместе с соответствующими частотами образуют следующие таблицы:

x_i	1	2	3	4
m_i	2	3	3	2

x_j	2	4	6	8
m_j	2	4	3	1

Проверить свойство общей выборочной средней.

5.5.5. Ежедневные расходы на обслуживание и ремонт автомобилей в некотором автосалоне составляют в среднем 100 тыс. р., а число продаж X автомашин в течение дня подчиняется следующему закону распределения

X	0	1	2	3	4	5	6	7	8	9
p	0,25	0,2	0,1	0,1	0,1	0,1	0,05	0,05	0,025	0,025

Найти математическое ожидание ежедневной прибыли при цене на машину 400 000 р.

Найти дисперсию ежедневной продажи числа автомобилей.

5.5.6. В таблице приведено распределение $n = 50$ рабочих по производительности труда X (единиц за смену), распределённых на две группы. Найти общие и групповые средние и проверить «правило сложения дисперсии».

	Первая группа					Вторая группа					
x_{ij}	3	8	9	10	12	6	7	8	9	10	
m_{ij}	5	2	11	8	4	2	6	8	3	1	$\sum = 50$

Здесь индекс i соответствует номеру группы $i = 1, 2$; индекс j – номер варианты в группе $j = 1, 2, \dots, 5$.

Кроме того, закрепить полученные навыки можно, решив задачи № 446, 453, 463, 469 из [3], № 8.13 из [2].

6. ОЦЕНКА ПАРАМЕТРОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ

I. Учебные цели. Познакомить студентов с постановкой одной из основных задач математической статистики – задачей оценки неизвестных параметров известного распределения, сформировать умение осуществлять точечные и интервальные оценки неизвестных параметров.

II. Формирование компетенций. Формировать математическую культуру, развивать аналитическое и логическое мышление, развивать способность к обобщению.

III. Введение в тему.

Как было сформулировано в предыдущей главе, отличительной особенностью математической статистики от теории вероятностей является тот факт, что статистические модели определены с точностью до неизвестных параметров, которые надо оценить по результатам наблюдений.

Содержание данного раздела можно сформулировать как совокупность методов, позволяющих делать научно обоснованные выводы о числовых параметрах распределения генеральной совокупности по случайной выборке из неё. Если, например, нас интересует математическое ожидание генеральной совокупности, то задача статистической оценки параметров заключается в том, чтобы найти такую выборочную характеристику, которая позволила бы получить по возможности более точное и надёжное представление об интересующем нас параметре. Так как состав выборки случаен, то выводы, сделанные в этих условиях, могут быть ложными. С увеличением объёма выборки вероятность правильного вывода увеличивается. Поэтому всякому решению, принимаемому при статистической оценке параметров, стараются поставить в соответствие вероятность, характеризующую степень достоверности принятого решения.

При изучении материала обратите внимание на следующие вопросы для контроля качества усвоения:

1. Каким требованиям должна удовлетворять точечная оценка?
2. Как найти оценку параметра распределения Пуассона методом наибольшего правдоподобия?
3. Что является точечными оценками параметров нормального распределения?
4. Какова сущность интервальных оценок?
5. Как построить доверительные интервалы для параметров нормального распределения?

6.1. ПОСТАНОВКА ЗАДАЧИ ТОЧЕЧНОЙ ОЦЕНКИ

Одной из задач математической статистики является оценка неизвестных параметров известного распределения признака X в генеральной совокупности по конечной выборке $\{x_1, x_2, \dots, x_n\}$ объёма n из этой совокупности. Например, если распределение нормальное, то оценке подлежат математическое ожидание и среднее квадратическое отклонение.

Пусть Q – оцениваемый параметр, постоянное (неслучайное) число. Оценкой \bar{Q}_n параметра Q называется любая функция от значения выборки

$\bar{Q}_n = \bar{Q}_n(x_1, x_2, \dots, x_n)$, т.е. статистика (поэтому оценка называется статистической). Если принять утверждение, что x_i является реализацией случайной величины X_i , то статистику \bar{Q}_n можно рассматривать как функцию от случайных величин X_1, X_2, \dots, X_n .

Статистику \bar{Q}_n надо выбирать таким образом, чтобы её значения как можно точнее оценивали значение неизвестного параметра Q . Различают следующие требования к оценке \bar{Q}_n .

1. Состоятельность – при больших объёмах выборки \bar{Q}_n как угодно мало отличается от Q (\bar{Q}_n стремится к Q по вероятности):

$$\lim_{n \rightarrow \infty} P\{|\bar{Q}_n - Q| < \varepsilon\} = 1.$$

2. Несмещённость – её математическое ожидание должно быть равно оцениваемому параметру $M\bar{Q}_n = Q$.

Смещённость – наличие отклонения. Например: $Q = 2$; $M\bar{Q}_n = 2,05$.

3. Эффективность – при одном и том же объёме выборки её дисперсия минимальна среди всевозможных оценок:

$$D(\bar{Q}_n) = M(\bar{Q}_n - Q)^2 - \min.$$

При этом рассматриваются только несмещённые оценки.

Оценки \bar{Q}_n называют точечными, так как они оценивают численное значение параметра Q одним значением (точкой).

6.2. ТОЧЕЧНЫЕ ОЦЕНКИ ПАРАМЕТРОВ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

Пусть X – генеральная совокупность, $\{x_1, x_2, \dots, x_n\}$ – повторная выборка её значений; при этом случайные величины X_1, X_2, \dots, X_n – независимые. Пусть $MX = a$, $DX = \sigma^2$ – генеральная средняя и генеральная дисперсия.

В качестве оценок для a и σ рассматривают выборочную среднюю

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{и} \quad \text{выборочную дисперсию} \quad S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \bar{D}X, \quad \bar{D} - \text{выборочная дисперсия.}$$

Свойства этих оценок:

Свойства этих оценок:

$$M(\bar{x}) = M\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \frac{1}{n} \sum_{i=1}^n MX_i = \frac{1}{n} \sum_{i=1}^n a = \frac{1}{n} na = a.$$

Следовательно, \bar{x} является несмещённой оценкой для генеральной средней.

Также существуют доказательства, что эта оценка является состоятельной и эффективной. Если X распределена нормально, то $DX = \frac{\sigma^2}{n}$ или $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$.

Существует доказательство, что $MS^2 = \frac{n-1}{n}\sigma^2$, т.е. оценка S^2 является смещённой; в то же время оценка $\bar{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ является несмещённой; $\bar{S}^2 = \frac{n}{n-1} S^2$ – исправленная выборочная дисперсия. S^2 и \bar{S}^2 являются состоятельными оценками для σ^2 , а \bar{S}^2 и эффективной.

Для бесповторной выборки оценки \bar{x} и \bar{S}^2 также являются несмещёнными и состоятельными, а дисперсия $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)$, где N – объём генеральной совокупности.

6.3. ИНТЕРВАЛЬНЫЕ ОЦЕНКИ ПАРАМЕТРОВ

Точечная оценка \bar{Q}_n параметра Q даёт лишь его некоторое приближённое значение. Чтобы иметь представление о точности и надёжности оценки, используют интервальную оценку.

Интервальной оценкой параметра Q называется интервал (α, β) , который с заданной вероятностью γ покрывает неизвестное значение Q . Такой интервал (α, β) называется доверительным интервалом, а вероятность γ – доверительной вероятностью или уровнем надёжности.

На практике доверительный интервал симметричен относительно оценки \bar{Q}_n и определяется из формулы:

$$P\{|\bar{Q}_n - Q| < \Delta\} = \gamma \text{ и имеет вид } \bar{Q}_n - \Delta < Q < \bar{Q}_n + \Delta.$$

Последнее неравенство выполняется с вероятностью γ , а наибольшее отклонение Δ выборочного значения параметра \bar{Q}_n от его значения Q называется предельной ошибкой выборки.

6.3.1. Доверительные интервалы для генеральной средней

Пусть из генеральной совокупности объёма N произведена выборка $\{x_1, x_2, \dots, x_n\}$, \bar{x} – среднее выборочное значение, \bar{S}^2 – выборочная дисперсия, \bar{S} – выборочное среднее квадратическое отклонение.

Доверительный интервал уровня надёжности γ для генеральной средней a имеет вид:

$$\bar{x} - \Delta < a < \bar{x} + \Delta,$$

где Δ – предельная ошибка выборки, $\Delta = \Delta(\gamma)$.

При $n > 30$

$$\text{для повторной выборки } \Delta = t \frac{\bar{S}}{\sqrt{n}},$$

$$\text{для бесповторной} - \Delta = t \frac{\bar{S}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}},$$

где t определяется из уравнения $\Phi(t) = \frac{\gamma}{2}$, $\Phi(t)$ – функция Лапласа, значения которой даются в табличной форме [1, с. 556]. Задавая требуемый уровень надёжности γ , по таблицам можно найти соответствующее значение параметра t . Например, если $\gamma = 0,95$, то $t = 1,96$; если $\gamma = 0,97$, то $t = 2,17$; если $\gamma = 0,99$, то $t = 2,58$.

Таким образом, чем больше требуемый уровень надёжности, тем «шире» доверительный интервал, что естественно.

Если $n \leq 30$, то доверительный интервал для a строится только для нормальной генеральной совокупности и для повторной

выборки $\Delta = t_{n-1} \frac{\bar{S}}{\sqrt{n}}$, t_{n-1} определяется из условия $P\{|\xi| < t_{n-1}\} = \gamma$,

где случайная величина ξ имеет распределение Стьюдента с $v = n - 1$ степенью свободы и находится по заданному уровню надёжности γ из таблиц распределения Стьюдента [1, с. 557]. Например, если $\gamma = 0,95$ ($p = 0,05$) и $v = 29$, то $t = 2,04$; если $\gamma = 0,99$ ($p = 0,01$) и $v = 6$, то $t = 3,71$. Здесь наблюдается «двойная» зависимость: чем больше требуемый уровень надёжности, тем больше значение t ; чем больше объём выборки (число степеней свободы), тем меньше t , что вполне согласуется с интуитивными представлениями о величине доверительного интервала.

6.3.2. Доверительный интервал для генеральной дисперсии (среднего квадратического отклонения)

Если $\{x_1, x_2, \dots, x_n\}$ – выборка из нормальной совокупности, a и σ неизвестны, то статистика $Z = \frac{n\bar{S}^2}{\sigma^2}$ имеет распределение χ_{n-1}^2 .

Доверительный интервал для σ^2 находят из соотношения

$$P\left\{\frac{n\bar{S}^2}{z_2} < \sigma^2 < \frac{n\bar{S}^2}{z_1}\right\} = \gamma; \quad P\left\{z_1 < \frac{n\bar{S}^2}{\sigma^2} < z_2\right\} = \gamma.$$

В свою очередь z_1 и z_2 определяются из условия $P\{z_1 < \chi_{n-1}^2 < z_2\} = \gamma$, а на практике таким образом, чтобы $P\{\chi_{n-1}^2 < z_1\} = P\{\chi_{n-1}^2 > z_2\} = \frac{1-\gamma}{2} = \frac{1}{2} - \frac{\gamma}{2}$.

В этом случае $P\{\chi_{n-1}^2 > z_1\} = \frac{1+\gamma}{2}$; $P\{\chi_{n-1}^2 > z_2\} = \frac{1-\gamma}{2}$ – условия для выбора величин z_1 и z_2 , которые можно найти из таблиц распределения χ_{n-1}^2 [1, с. 558].

Например, если $\gamma = 0,90$, то $\frac{1+\gamma}{2} = 0,95$; $\frac{1-\gamma}{2} = 0,05$.

При $n = 30$ $P\{\chi_{29}^2 > z_1\} = 0,95$; $P\{\chi_{29}^2 > z_2\} = 0,05$. Из таблиц значений χ^2 -критерия Пирсона находим, что $z_1 = 17,7$; $z_2 = 42,6$.

Пример 6.1. На складе хранится партия товара в количестве 625 штук с различными сроками хранения (от 2 до 10 месяцев). Осуществлена бесповторная выборка товара в количестве 25 штук; при этом оказалось, что 8 единиц товара имеют срок хранения 2 месяца, 6 единиц – 5 месяцев; 4 единицы – 6 месяцев и 67 единиц – 10 месяцев.

Найти:

- выборочный средний срок хранения товара (\bar{x});
- выборочную дисперсию срока хранения (\bar{S}^2);
- доверительный интервал для оценки среднего срока хранения товара a с надёжностью 0,95.
- доверительный интервал для оценки среднего квадратического отклонения σ с надёжностью 0,95.

Решение. Имеем распределение выборки:

x_i	2	5	6	10
m_i	8	6	4	7

$$\text{а) } \bar{x} = \frac{1}{25}(2 \cdot 8 + 5 \cdot 6 + 6 \cdot 4 + 10 \cdot 7) = 5,6;$$

$$\text{б) } \overline{S^2} = \overline{x^2} - (\bar{x})^2 = \frac{1}{25}(4 \cdot 8 + 25 \cdot 6 + 36 \cdot 4 + 100 \cdot 7) - (5,6)^2 = 9,68;$$

$$\text{в) } \bar{x} - \Delta < a < \bar{x} + \Delta; \quad n = 25 < 30, \quad \Delta = t_{n-1} \frac{\bar{S}}{\sqrt{n}} \sqrt{1 - \frac{n}{N}},$$

здесь $\bar{S} = \sqrt{\overline{S^2}} = 3,11$; $n = 25$; $N = 625$.

$t_{n-1} = t_{24}$ из таблицы распределения Стьюдента при доверительной вероятности $\gamma = 0,95$: $t_{24} = 2,06$.

$$\Delta = 2,06 \frac{3,11}{\sqrt{25}} \left(\sqrt{1 - \frac{25}{625}} \right) \cong 1,25;$$

$$5,6 - 1,25 < a < 5,6 + 1,25; \quad 4,35 < a < 6,85;$$

$$\text{г) } \text{имеем: } P \left\{ \frac{n\overline{S^2}}{z_2} < \sigma^2 < \frac{n\overline{S^2}}{z_1} \right\} = \gamma.$$

$$P\{\chi_{24}^2 > z_1\} = \frac{1+0,95}{2} = 0,975, \text{ по таблицам } \chi^2 \quad z_1 = 12,4.$$

$$P\{\chi_{24}^2 > z_2\} = \frac{1-0,95}{2} = 0,025 \Rightarrow z_2 = 39,4.$$

$$\overline{S^2} = \frac{25}{25-1} \cdot 9,68 = 10,08.$$

$$\frac{n\overline{S^2}}{z_2} = \frac{25 \cdot 10,08}{39,4} = 6,4; \quad \frac{n\overline{S^2}}{z_1} = \frac{25 \cdot 10,08}{12,4} = 20,32.$$

$$6,4 < \sigma^2 < 20,32; \quad 2,53 < \sigma < 4,5.$$

6.4. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

6.4.1. В результате пяти измерений получены следующие результаты (верхнего) кровяного давления: 122, 124, 133, 135, 136. Найти выборочную среднюю, выборочную и исправленную дисперсию.

6.4.2. Некоторый товар в количестве 1000 кг решили разделить на три сорта в зависимости от качества и продавать по цене 60, 90 и 100 р. за килограмм. Для оценки возможных доходов от продажи сделали выборку в 30 кг. Оказалось, что она содержит товара низшего сорта – 5 кг, среднего – 10 кг и высшего – 15 кг. Выборка повторная.

Найти:

1. Выборочную среднюю стоимости товара;
2. Выборочную дисперсию стоимости;
3. Доверительный интервал для оценки средней стоимости товара с надёжностью 0,92;
4. Доверительный интервал для оценки среднего квадратического отклонения от средней стоимости с надёжностью 0,95.

6.4.3. В университете работает 625 преподавателей в возрасте от 25 до 65 лет. Осуществлена бесповторная выборка в количестве 30 человек, которая показала, что в возрасте от 25 до 35 лет – 5 человек; от 35 до 45 лет – 8 человек; от 45 до 55 лет – 7 человек и от 55 до 65 лет – 10 человек.

Найти:

1. Выборочный средний возраст преподавателей;
2. Выборочную дисперсию этого возраста;
3. Доверительный интервал для оценки среднего возраста с надёжностью 0,96;
4. Доверительный интервал для оценки среднего квадратического отклонения с надёжностью 0,9.

6.4.4. В следственных изоляторах Тамбовской области находится 480 человек. При выборке в 24 человека оказалось, что со сроком задержания до 2 месяцев было 8 человек, от 2 до 4 месяцев – 6 человек, от 4 до 6 месяцев – 8 человек, от 6 до 8 месяцев – 2 человека. Выборка бесповторная.

Найти:

1. Выборочный средний срок задержания;
2. Выборочную дисперсию срока задержания;
3. Доверительный интервал для оценки среднего срока задержания с надёжностью 0,97;
4. Доверительный интервал для оценки среднего квадратического отклонения срока задержания с надёжностью 0,96.

6.4.5. В ящике находится 1600 монет достоинством 1, 2, 5 и 10 р. Осуществлена повторная выборка 40 монет. При этом оказалось, что монета достоинством 1 р. была извлечена 8 раз; 2 р. – 4 раза; 5 р. – 12 раз; 10 р. – 16 раз.

Найти:

1. Выборочное среднее номинала монет;
2. Выборочную дисперсию отклонения номиналов монет от среднего;
3. Доверительный интервал для оценки среднего номинала монеты с надёжностью 0,95;
4. Доверительный интервал для оценки среднего квадратического отклонения номинала с надёжностью 0,98.

6.4.6. В саду имеется 900 яблонь. Для оценки предполагаемого урожая выбрали 40 яблонь (выборка бесповторная). Оказалось, что урожайность составила от 100 до 260 кг для различных яблонь и подчиняется следующему распределению:

Вес урожая, кг	[100, 140)	[140, 180)	[180, 220)	[220, 260]
Кол-во яблонь	20	6	8	6

Найти:

1. Выборочную среднюю урожайность одной яблони и доверительный интервал для оценки генеральной средней (a) с надёжностью 0,98;
2. Выборочное среднее квадратическое отклонение урожайности относительно среднего и доверительный интервал для оценки генерального среднего квадратического отклонения (σ) с надёжностью 0,90.

Кроме того, можно закрепить полученные навыки, решив задачи № 511 и 515 из [3].

7. ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИЙ

I. Учебные цели. Изучить некоторые варианты стохастической зависимости переменных величин.

В результате изучения материала студенты должны знать определения и свойства коэффициента корреляции, понятие корреляционной зависимости, линии регрессии, уметь оценивать тесноту и характер связи случайных величин.

II. Формирование компетенций. Формирование математической культуры, развитие способностей использовать законы естественнонаучных дисциплин в профессиональной деятельности, изложения сути предлагаемых решений, логически верно, аргументированно и ясно строить рассуждения.

III. Введение в тему.

Связи между различными явлениями в природе сложны и многообразны, однако, их можно определённым образом классифицировать. В технике и естествознании часто идёт речь о функциональной зависимости между переменными x и y , когда каждому возможному значению x поставлено в однозначное соответствие определённое значение y .

В реальном мире многие явления природы происходят в обстановке действия многочисленных факторов, влияние каждого из которых ничтожно, а число их велико. В этих случаях связь теряет свою строгую функциональность, и изучаемая система переходит не в определённое состояние, а в одно из возможных состояний. Поэтому речь идёт о так называемой стохастической связи, состоящей в том, что одна случайная

переменная реагирует на изменение другой изменением своего закона распределения. В практике статистических исследований рассматривают частный случай стохастической связи – статистическую связь, когда условное математическое ожидание одной случайной переменной является функцией значения, принимаемого другой случайной переменной, т.е. $M(Y/x) = f(x)$.

Знание статистической зависимости между случайными переменными имеет большое практическое значение: с её помощью можно прогнозировать значение зависимой случайной переменной в предположении, что независимая примет определённое значение. Однако такие прогнозы не могут быть безошибочными. Применяя вероятностные методы, можно вычислить вероятность того, что ошибка прогноза не выйдет за определённые границы.

Вопросы для контроля усвоения:

1. Что такое корреляционная зависимость?
2. Как найти коэффициент корреляции и что он характеризует?
3. Что такое линия регрессии?
4. Как найти уравнение линейной регрессии?
5. Что даёт статистический анализ уравнения регрессии?

7.1. КОРРЕЛЯЦИОННАЯ ЗАВИСИМОСТЬ И ПРЕДСТАВЛЕНИЕ ДАННЫХ В КОРРЕЛЯЦИОННОМ АНАЛИЗЕ

Пусть X – независимая переменная, Y – зависимая.

На практике одному значению X (фактору) может соответствовать ряд значений Y (определённое распределение Y), так как кроме фактора X имеет место влияние других факторов.

Например, X – срок курения; Y – заболевание лёгких (оценка в баллах).

Заболевание может быть инициировано другими факторами и неоднозначно зависеть от срока курения.

Такая зависимость называется стохастической (вероятностной) или корреляционной².

Корреляционной зависимостью между двумя переменными называется функциональная зависимость между одной из них и условным математическим ожиданием другой:

$\varphi(x) = M_x(Y)$ или $\psi(y) = M_y(X)$ – это уравнение регрессии³ (« Y » на « X » и « X » на « Y »), а их графики – линии регрессии.

² Корреляция – соотношение, взаимозависимость, связь между явлениями, если одно из них входит в число причин, определяющих другие.

³ Регресс – движение назад, возвращение, в противоположность тому, что прогресс – движение вперёд, развитие.

Если $\varphi(x) = \text{const}$, $\psi(x) = \text{const}$, то корреляционной связи нет.

Основная задача корреляционного анализа – выявление тесноты связи между переменными X и Y и количественная оценка этой связи.

Примеры корреляционной зависимости:

- X – уровень безработицы; Y – уровень преступности;
- X – образовательный уровень полицейских; Y – процент раскрываемости преступлений;
- X – уровень требовательности преподавателя; Y – успеваемость студентов.

В корреляционном анализе экспериментальные данные можно представлять в виде набора пар чисел (x_i, y_i) , $i = \overline{1, n}$, где (x_1, \dots, x_n) – выборка значений X ; (y_1, y_2, \dots, y_n) – выборка значений Y .

Пример 7.1. X – процент безработных в городе: {2, 4, 6, 8, 10}, Y – процент совершения краж имущества: {0,2; 0,25; 0,5; 0,65; 0,9}.

Данные для корреляционного анализа:

$$(2; 0,2); (4; 0,25); (6; 0,5); (8; 0,65); (10; 0,9)^4.$$

Насколько тесная связь: причины – безработицы и следствия – уровня преступности – задача корреляционного анализа.

Экспериментальные данные можно задавать и таблицей (корреляционная таблица).

Пример 7.2. Получены статистические данные по 10 сельскохозяйственным предприятиям с целью исследования зависимости объёма основных фондов X (стоимость техники и т.п.) и урожайности зерновых культур Y .

		x_i			n_i
		50	100	150	
y_i	y				
36 – 40	38	1			1
40 – 44	42		2		2
44 – 48	46		3	2	5
48 – 52	50			2	2
	n_i	1	5	4	10

⁴ Может быть и другое сочетание: (2; 0,5), {4; 0,2}, {6; 0,65}, {8; 0,25}, {10; 0,9}.

В первой строке таблицы записаны средние значения переменной X для интервалов $[25, 75)$, $[75, 125)$ и $[125, 175]$, в первом столбце – интервалы изменения переменной Y , во втором – середины этих интервалов. Центральную часть таблицы занимают частоты – частоты n_{ij} – число предприятий, соответствующих значениям переменных $X = x_i$, $Y = y_j$; $i = 1, 2, 3$; $j = 1, 2, 3, 4$. В последней строке записаны $n_i = \sum_{j=1}^4 n_{ij}$, в последнем

столбце – $n_j = \sum_{i=1}^3 n_{ij}$. Кроме того, $\sum_{i=1}^3 n_i = \sum_{j=1}^4 n_j = n$.

Далее будем рассматривать данные для корреляционного анализа, заданные таблицей:

X	x_1	x_2	...	x_n
Y	y_1	y_2	...	y_n

7.2. КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ

Основой оценки для тесноты связи между переменными X и Y служит выборочный коэффициент корреляции r (который мы уже упоминали в 4.2):

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x S_y}.$$

Если данные не сгруппированы, то

$$r = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{j=1}^n y_j \right)}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \sqrt{n \sum_{j=1}^n y_j^2 - \left(\sum_{j=1}^n y_j \right)^2}}. \quad (7.1)$$

Свойства выборочного коэффициента корреляции аналогичны свойствам коэффициента корреляции между случайными величинами X и Y .

1. $-1 \leq r \leq 1$; чем ближе $|r|$ к 1, тем теснее связь.

2. Если переменные X и Y умножить на одно и то же число, то r не изменится.

3. Если $r = \pm 1$, корреляционная связь между X и Y линейная.

Генеральный коэффициент корреляции ρ не является случайной величиной, а выборочный коэффициент корреляции r – величина случайная.

Если $r \neq 0$, то возникает вопрос о значимости найденного коэффициента, т.е. его отличие от нуля – факт или случайность.

Для проверки этого обстоятельства рассматривают статистику (экспериментальную величину) $t_3 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$, которая, в предположении, что

$\rho = 0$ (связь между X и Y отсутствует), имеет распределение Стьюдента с $k = n - 2$ степенями свободы. По соответствующим таблицам (распределения Стьюдента) при $k = n - 2$ и при заданном уровне доверия γ (или $\alpha = 1 - \gamma$) находят $t_{\text{табл.}}$. Если $|t_3| > t_{\text{табл.}}$, то считают, что найденный коэффициент корреляции значим.

Пример 7.3. Проведено исследование 6 хозяйств для изучения зависимости урожайности зерновых культур (Y , ц/га) от количества внесённых минеральных удобрений на 1 га пашни (X , ц/га). Получены следующие данные:

X	2,1	2,3	2,4	2,6	2,9	3,0
Y	18,0	21,0	22,1	25,3	28	28,5

Определить тесноту связи между величиной Y и величиной X , используя коэффициент корреляции, проверить на уровне $\alpha = 0,05$ его значимость.

Решение. Коэффициент корреляции найдём по формуле (7.1).

Для её реализации составим таблицу:

i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	2,1	18,0	37,8	4,41	324
2	2,3	21,0	48,3	5,29	441
3	2,4	22,1	53,04	5,76	488,41
4	2,6	25,3	65,78	6,76	690,09
5	2,9	28	81,2	8,41	784
6	3,0	28,5	85,5	9,00	812,25
$\sum_{i=1}^n$	15,3	142,9	371,62	39,63	3489,75

Тогда

$$r = \frac{6 \cdot 371,62 - 142,9 \cdot 15,3}{\sqrt{6 \cdot 39,63 - (15,3)^2} \sqrt{6 \cdot 3489,75 - (142,9)^2}} = 0,991.$$

Значимость r на уровне $\alpha = 0,05$.

$$\text{Статистика } t_3 = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,991 \cdot 2}{\sqrt{1-0,982}} = \frac{1,982}{0,134} = 14,79.$$

$t_{\text{табл}}$ (по Стьюденту) при $k = n - 2 = 4$: $t_{\text{табл}} = 2,78$.

Так как $t_3 = 14,79 > t_{\text{табл}}$, то r значим.

7.3. СТАТИСТИЧЕСКАЯ ЗАВИСИМОСТЬ. УРАВНЕНИЕ ЛИНЕЙНОЙ РЕГРЕССИИ

В статистическом анализе зависимость между входными параметрами (значениями неслучайной независимой переменной X) и выходной переменной Y рассматривается как статистическая, и представляет особый интерес установление вида зависимости Y от X_1, X_2, \dots, X_n , т.е. вида уравнения регрессии. Это связано в первую очередь с необходимостью прогнозирования исследуемых процессов.

Оценкой функции регрессии $\varphi(x) = M_x(Y)$ является функция

$$y_x = \varphi(x, b_0, b_1, \dots, b_n), \quad (7.2)$$

где x – значение величины X ; $y_x = M_x(Y)$; b_0, b_1, \dots, b_n – параметры функции регрессии.

Задача регрессионного анализа состоит в определении функции φ , её параметров и дальнейшем статистическом исследовании уравнения регрессии.

Ориентировочное определение вида функции φ можно осуществить эмпирически, построив корреляционное поле.

Если функция φ линейна по x , т.е. $y_x = a + bx$, то говорят, что имеет место линейная регрессия Y по X .

Используя метод наименьших квадратов, составим функцию неизвестных переменных a и b :

$$S(a, b) = \sum_{i=1}^n (y_i - (a + bx_i))^2 \rightarrow \min.$$

Из необходимого условия экстремума (равенства нулю частных производных) после преобразований получим систему:

$$\begin{cases} a + \bar{x}b = \bar{y}, \\ \bar{x}a + \bar{x}^2b = \overline{xy}, \end{cases}$$

$$\text{где } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i; \quad \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2; \quad \overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i.$$

Решение этой системы:

$$a = \frac{\overline{x^2 y} - \overline{x} \overline{xy}}{x^2 - (\overline{x})^2}; \quad b = \frac{\overline{xy} - \overline{x} \overline{y}}{x^2 - (\overline{x})^2}$$

определяет уравнение регрессии $y_x = a + bx$.

Коэффициент b в уравнении регрессии называется коэффициентом регрессии Y по X и обозначается b_{YX} .

Если учесть то обстоятельство, что $x^2 - (\overline{x})^2 = S_x^2$ – дисперсия случайной величины X , то

$$b_{yx} = \frac{\overline{xy} - \overline{x} \overline{y}}{S_x^2} = \frac{\overline{xy} - \overline{x} \overline{y}}{S_x S_y} \frac{S_y}{S_x} = r \frac{S_y}{S_x}, \quad (7.3)$$

где S_y^2 – дисперсия случайной величины Y .

Запишем (возьмём) математическое ожидание от правой и левой частей уравнения $y_x = a + bx$:

$$M(y_x) = M(a + bx) = M(a) + M(bx) = a + bM(x).$$

Тогда $\overline{y} = a + b\overline{x}$ и уравнение регрессии можно записать в обычно принятой в математической статистике форме:

$$y_x - \overline{y} = b_{YX}(x - \overline{x}) \quad (7.4)$$

Таким образом, уравнение линейной регрессии можно записать, если известны выборочные средние \overline{x} и \overline{y} и коэффициент регрессии.

Пример 7.4. Зависимость между стоимостью эксплуатации самолёта Y (млн. р.) и его возрастом X (лет) выражается следующей таблицей:

X	1	2	3	4	5	6	7	8
Y	3	3,5	3,5	4	4	6	9	10

Найти линейное уравнение регрессии Y по X , оценить его значимость на уровне $\alpha = 0,01$.

Решение.

Уравнение регрессии Y по X (7.4) с учётом (7.3) можно записать в виде:

$$y_x - \overline{y} = r \frac{S_y}{S_x}(x - \overline{x}).$$

Объединим результаты вычислений исходных данных в следующую таблицу.

i	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
1	1	3	3	1	9
2	2	3,5	7	4	12,25
3	3	3,5	10,5	9	12,25
4	4	4	16	16	16
5	5	4	20	25	16
6	6	6	36	36	36
7	7	9	63	49	81
8	8	10	80	64	100
Σ	36	43	235,5	204	283

Имеем: $\bar{x} = \frac{36}{8} = 4,5$; $\bar{y} = \frac{43}{8} = 5,38$; $\overline{xy} = \frac{235,5}{8} = 29,4$; $\overline{x^2} = \frac{204}{8} = 25,5$;
 $\overline{y^2} = \frac{1}{8} \cdot 283 = 35,4$.

По формуле (7.1) находим $r = 0,89$, тогда $r \frac{S_y}{S_x} = 0,99 \approx 1$,

а уравнение регрессии:

$$y_x - 5,38 = x - 4,5 \quad \text{или} \quad y_x = x + 0,88.$$

7.4. ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

7.4.1. Случайные величины X и Y связаны таблицей

X	2	3	4	5	6
Y	2	1,9	3,2	2,4	2,3

Определить тесноту связи X и Y , используя коэффициент корреляции, и проверить на уровне $\alpha = 0,05$ его значимость.

7.4.2. В автохозяйстве исследовали зависимость между сроком эксплуатации автомобиля (X , лет) и затратами на его обслуживание (ремонт, потребление топлива) (Y , % от первоначальной стоимости в год). Получены следующие данные:

X	[1 – 3)	[3 – 5)	[5 – 9)	[9 – 13)	[13 – 19]
Y	10	20	35	50	90

Определить тесноту связи X и Y , используя коэффициент корреляции, и проверить на уровне $\alpha = 0,05$ его значимость.

7.4.3. Определить тесноту связи общего веса X (г) некоторого растения и веса Y (г) его семян на основе следующих выборочных данных:

X	40	50	60	70	80	90	100
Y	20	25	28	30	35	40	45

Проверить значимость выборочного коэффициента корреляции при $\alpha = 0,05$.

7.4.4. Определить тесноту связи объема полива (X , л/м²) и урожайности (Y , кг/м²) некоторой сельскохозяйственной культуры на основе следующих выборочных данных:

X	26	35	36	40	41	45
Y	18	21	22,1	25,3	28	28,5

Проверить значимость выборочного коэффициента корреляции на уровне $\alpha = 0,05$.

7.4.5. Выборочные величины X и Y связаны таблицей:

X	8	12	16	20	24
Y	47	76	100	136	150

1. Определить с помощью выборочного коэффициента корреляции r тесноту связи между величинами X и Y .

2. Оценить значимость r на уровне 0,05 (используя распределение Стьюдента).

3. Для зависимости Y от X , заданной корреляционной таблицей, найти оценки параметров a и b уравнения линейной регрессии $y_x = a + bx$.

7.4.6. Выборочные величины X и Y связаны таблицей:

X	12	17	22	27	32
Y	20	22	24	25	27

1. Определить с помощью выборочного коэффициента корреляции r тесноту связи между величинами X и Y .

2. Оценить значимость r на уровне 0,05 (используя распределение Стьюдента).

3. Для зависимости Y от X , заданной корреляционной таблицей, найти оценки параметров a и b уравнения линейной регрессии $y_x = a + bx$.

Кроме того, закрепить полученные навыки можно, решив задачи № 12.5 и 12.15 из [1].

ВОПРОСЫ К ЭКЗАМЕНУ

1. Случайные события и их виды.
2. Вероятность события (классическое, статистическое и геометрическое определения) и её свойства.
3. Условная вероятность, теоремы об умножении вероятностей.
4. Теоремы о сложении вероятностей.
5. Формула полной вероятности. Формула Байеса.
6. Повторные испытания. Формула Бернулли.
7. Дискретные случайные величины и их распределение.
8. Числовые характеристики дискретных случайных величин.
9. Непрерывные случайные величины, их распределение и числовые характеристики.
10. Двумерная случайная величина и её распределение.
11. Оценки уровня зависимости компонентов двумерной случайной величины.
12. Генеральная и выборочная совокупности.
13. Вариационный ряд. Статистическое распределение выборки. Полигон и гистограмма.
14. Эмпирическая функция распределения и её свойства.
15. Выборочная средняя и выборочная дисперсия, их свойства.
16. Точечные оценки. Требования к оценкам.
17. Точечные оценки параметров нормального распределения.
18. Интервальные оценки. Построение доверительного интервала.
19. Корреляционная зависимость. Коэффициент корреляции.
20. Уравнение регрессии. Линейная регрессия.
21. Определение параметров уравнений регрессии методом наименьших квадратов.

ОТВЕТЫ К ЗАДАЧАМ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

1.3.1. а) 24; б) 6; в) 6; г) 12. **1.3.2.** 1/90. **1.3.3.** 0,5. **1.3.4.** а) 1/6; б) 1/18; в) 1/6. **1.3.5.** 6/23. **1.3.6.** 42/245. **2.6.1.** 0,88. **2.6.2.** 0,005. **2.6.3.** 1/22. **2.6.4.** 0,57. **2.6.5.** 0,8. **2.6.6.** 1/25. **3.6.1.** $M(X) = 2,4$. **3.6.2.** $D(X) = 6,29$. **3.6.3.** а) 0; б) 0,5; в) 0,5. **3.6.4.** $D(X) = 1,39$. **3.6.5.** $M(X) = 18$; $D(X) = 4$. **3.6.6.** $\alpha = 3/8$; $M(X) = 1,5$; $D(X) = 0,75$; $p = 0,95$. **4.4.1.** б) $M(X) = 1/2$; $M(Y) = 1/3$; $D(X) = 1/4$; $D(Y) = 2/9$. **4.4.2.** $K_{xy} = 1,6$; $r = 2/3$. **4.4.3.** в) $r = 0,7427$. **4.4.4.** Не являются. **4.4.5.** Являются. **4.4.6.**

y_k	20	25	30
$P_{(x>0)}(y_k)$	1/13	5/13	7/13

5.5.1. $\bar{x} = 2,57$; $S^2 = 3,31$. **5.5.2.** $\bar{x} = 25,6$. **5.5.3.** 6,4%. **5.5.5.** а) 975; б) 6,319. **5.5.6.** $\bar{x}_1 = 25/3$; $\bar{x}_2 = 7,75$; $\bar{x} = 8,1$. **6.4.1.** $\bar{x} = 130$; $S^2 = 34$; $\overline{S^2} = 42,5$. **6.4.2.** $\bar{x} = 90$; $S^2 = 200$; (85,62; 94,38); (11,46; 19,1). **6.4.3.** $\bar{x} = 47,34$; $S^2 = 119,52$; (43,24; 51,44); (9,33; 14,48). **6.4.4.** $\bar{x} = 3,33$; $S^2 = 3,89$; (2,35; 4,31); (1,55; 2,85). **6.4.5.** $\bar{x} = 5,9$; $S^2 = 13,29$; (5,79; 7,03); (2,93; 4,96). **6.4.6.** $\bar{x} = 160$; $S^2 = 2080$; $142,33 < \bar{x} < 177,67$; $(39,39 < \sigma < 55,7)$. **7.4.1.** $r = 0,339$; не значим. **7.4.2.** $r = 0,989$; значим. **7.4.3.** $r = 0,992$; значим. **7.4.4.** $r = 0,954$; значим. **7.4.5.** $r = 0,994$; значим; $a = 16$; $b = 101,8$. **7.4.6.** $r = 0,995$; значим; $a = 22$; $b = 23,6$.

СПИСОК ЛИТЕРАТУРЫ

1. Кремер, Н.Ш. Теория вероятностей и математическая статистика : учебник для вузов / Н.Ш. Кремер. – 2-е изд., перераб. и доп. – М. : ЮНИТИ-ДАНА, 2004. – 573 с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика : учеб. пособие для вузов / В.Е. Гмурман. – 11-е изд., стер. – М. : Высш. шк., 2005. – 479 с.
3. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике : учеб. пособие для студентов вузов / В.Е. Гмурман. – 9-е изд., стер. – М. : Высш. шк., 2004. – 404 с.
4. Пучков, Н.П. Математика случайного / Н.П. Пучков, Л.И. Ткач. – Тамбов : Изд-во ТГТУ, 2005.
5. Дворяткина, С.Н. Лекции по классической теории вероятностей / С.Н. Дворяткина, Л.Н. Ляхов. – М. : Книжный дом «ЛИБРОКОМ», 2010. – 176 с.
6. Фомин, В.И. Сборник заданий по теории вероятностей (типовые расчеты) : учеб. пособие / В.И. Фомин. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2002. – 80 с.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	3
1. ТЕОРИЯ ВЕРОЯТНОСТЕЙ. ОСНОВНЫЕ ПОНЯТИЯ	4
1.1. Случайные события и их виды	5
1.2. Вероятность события и её свойства	5
1.3. Использование комбинаторики для нахождения вероятностей	6
1.4. Задачи для самостоятельного решения	8
2. ОСНОВНЫЕ ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ	8
2.1. Умножение вероятностей	9
2.2. Сложение вероятностей	10
2.3. Формула полной вероятности	11
2.4. Формула Байеса	12
2.5. Формула Бернулли	12
2.6. Задачи для самостоятельного решения	13
3. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ	14
3.1. Дискретные случайные величины и их распределение	14
3.2. Числовые характеристики дискретных случайных величин ...	16
3.3. Непрерывные случайные величины	18
3.4. Нормальное распределение	19
3.5. Распределения χ^2 , Стьюдента и Фишера	22
3.6. Задачи для самостоятельного решения	24
4. ДВУМЕРНЫЕ СЛУЧАЙНЫЕ ВЕЛИЧИНЫ	25
4.1. Закон распределения двумерной случайной величины	25
4.2. Ковариация и коэффициент корреляции	28
4.3. Линейная регрессия	30
4.4. Задачи для самостоятельного решения	31
5. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. ОСНОВНЫЕ ПОНЯТИЯ	32
5.1. Статистические данные. Генеральная и выборочная совокупности	33
5.2. Вариационный ряд и его графическое изображение	35

5.3. Эмпирическая функция распределения	38
5.4. Числовые характеристики вариационных рядов	38
5.5. Задачи для самостоятельного решения	42
6. ОЦЕНКА ПАРАМЕТРОВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ...	43
6.1. Постановка задачи точечной оценки	44
6.2. Точечные оценки параметров нормального распределения	45
6.3. Интервальные оценки параметров	46
6.4. Задачи для самостоятельного решения	49
7. ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИЙ	51
7.1. Корреляционная зависимость и представление данных в корреляционном анализе	52
7.2. Коэффициент корреляции	54
7.3. Статистическая зависимость. Уравнение линейной регрессии	56
7.4. Задачи для самостоятельного решения	58
ВОПРОСЫ К ЭКЗАМЕНУ	60
ОТВЕТЫ К ЗАДАЧАМ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ	61
СПИСОК ЛИТЕРАТУРЫ	62