МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Государственное образовательное учреждение высшего профессионального образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

УТВЕРЖДАЮ
Директор ИДО
_______________ С.И. Качин

«____»_____________2011 г.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания и индивидуальные задания
для студентов ИДО, обучающихся по направлениям
080100 «Экономика»,
080200 «Менеджмент»,
080400 «Управление персоналом»,
100700 «Торговое дело»

Составители
Л.И. Константинова
Е.А. Молдованова

Семестр	**3**
Кредиты	3
Лекции, часов	6
Практические занятия, часов	10
Индивидуальные задания	№1, № 2
Самостоятельная работа, часов	92
Формы контроля	зачёт

Издательство
Томского политехнического университета
2011

УДК 519.21(075.8)

Теория вероятностей и математическая статистика: метод. указ. и индивид. задания для студентов ИДО, обучающихся по напр. 080100 «Экономика», 080200 «Менеджмент», 080400 «Управление персоналом», 100700 «Торговое дело» / сост. Л.И. Константинова, Е.А. Молдованова; Томский политехнический университет.– Томск: Изд-во Томского политехнического университета, 2011.– 75 с.

Методические указания и индивидуальные задания рассмотрены и рекомендованы к изданию методическим семинаром кафедры прикладной математики ИК «____» ______ 2011 года, протокол № ___.

Зав. кафедрой ПМ
профессор, доктор физ.-мат. наук ________________ В.П. Григорьев

Методические указания и индивидуальные задания рассмотрены и рекомендованы к изданию методическим семинаром кафедры высшей математики ФТИ «____» ________ 2011 года, протокол № ___.

Зав. кафедрой ВМ,
профессор, доктор физ.-мат. наук _________________ К.П. Арефьев

Аннотация

Методические указания и индивидуальные задания по дисциплине «Теория вероятностей и математическая статистика» предназначены для студентов ИДО, обучающихся по направлениям 080100 «Экономика», 080200 «Менеджмент», 080400 «Управление персоналом», 100700 «Торговое дело». Данная дисциплина изучается в одном семестре.

Приводится содержание основных тем дисциплины, темы практических занятий, варианты заданий для индивидуальных домашних заданий и список рекомендуемой литературы. Даны основные требования и методические указания по выполнению индивидуальных домашних заданий.

1. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОСНОВНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Теория вероятностей и математическая статистика» изучается в третьем семестре студентами ИДО, обучающимися по направлениям 080100 «Экономика», 080200 «Менеджмент», 080400 «Управление персоналом», 100700 «Торговое дело».

Задачами дисциплины являются:

- развитие математической интуиции;
- воспитание математической культуры;
- формирование навыков, необходимых для использования знаний при изучении специальных дисциплин и дальнейшей практической деятельности;
- овладение студентами необходимым математическим аппаратом, дающим возможность анализировать, моделировать и решать экономические задачи;
- воспитание у студентов отношения к математике как к инструменту исследования и решения экономических задач, необходимому в их дальнейшей работе.

В результате изучения дисциплины студент должен:

знать основные формулы для определения вероятности события; основные законы распределения случайных величин и их числовые характеристики; вероятностный смысл числовых характеристик случайных величин; предельные теоремы теории вероятностей; способы представления результатов наблюдений; методы оценивания генеральных параметров по выборке; общий алгоритм решения задач по проверке гипотез; способы оценивания стохастической связи и определения зависимости между переменными;

уметь применять изученные методы для решения экономических задач, устанавливать границы применимости методов, уметь анализировать найденные решения;

владеть навыками применения современного математического инструментария для решения экономических задач, методиками для построения, анализа и применения математических моделей оценки состояния, прогноза и развития различных экономических процессов;

иметь опыт применения математической символики для выражения количественных и качественных отношений объектов исследования, аналитического и численного решения задач.

Дисциплина «Теория вероятностей и математическая статистика» является базовой дисциплиной естественно-научного цикла (Б2).

Пререквизитами являются дисциплины «Математический анализ 1», «Математический анализ 2».

Кореквизитом является дисциплина «Линейная алгебра».

2. СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОГО РАЗДЕЛА ДИСЦИПЛИНЫ

Тема 1. Случайные события

Основные понятия теории вероятностей. Алгебра событий. Способы задания вероятностей. Условная вероятность. Формула полной вероятности. Формула Байеса. Формула Бернулли.

Рекомендуемая литература: [1, раздел 1, тема 1, 1.1.1 – 1.1.15].

Методические указания

Изучение дисциплины и данной темы начните с выбора учебника и пособия по решению задач. Основной трудностью при работе с литературой по теории вероятностей является отсутствие единой терминологии и системы обозначений. Первая тема курса имеет особое значение, так как в ней излагаются основы теории вероятностей, без понимания и усвоения которых дальнейшее изучение вызовет значительные затруднения. Непосредственно подсчитать вероятность события возможно только в задачах, соответствующих экспериментам с конечным числом равновозможных несовместных исходов. При этом подсчёт числа элементов различных подмножеств пространства элементарных событий осуществляется по формулам и правилам комбинаторики. Чаще всего вероятность события вычисляется по формулам сложения и умножения вероятностей. Для этого сложное событие надо при помощи операций над событиями выразить через простейшие события, вероятности которых уже известны. При использовании формулы Байеса следует следить за тем, чтобы сформулированные гипотезы образовывали полную группу несовместных событий. «Теория вероятностей есть, в сущности, не что иное, как здравый смысл, сведённый к исчислению: она заставляет оценить с точностью то, что справедливые умы чувствуют как бы инстинктом, часто не умея отдать себе в этом отчёта» писал великий французский математик П. Лаплас.

Ниже приведены для самоконтроля усвоения темы. Если вопросы не вызвали затруднений, можно переходить к разбору задач, решённых в методических указаниях и в пособии [1]. В противном случае обратитесь к рекомендованной литературе и попробуйте найти ответы на вопросы.

Вопросы для самопроверки

1. Что изучает теория вероятностей?
2. Что называется элементарным событием (элементарным исходом)?
3. Что такое пространство элементарных событий?
4. Какое событие называется достоверным?
5. Какое событие называется невозможным?
6. Что называется суммой двух событий?
7. Что называется произведением двух событий?
8. Может ли сумма двух событий совпадать с их произведением?
9. Какие события называются несовместными?
10. Какие события называются совместными?
11. Какое событие называется противоположным для данного события?
12. Какими способами можно задать вероятность события?
13. Какие значения может принимать вероятность события?
14. Чему равна вероятность невозможного события?
15. Чему равна вероятность достоверного события?
16. Какое событие называется практически достоверным?
17. Какое событие называется практически невозможным?
18. Какие события образуют полную группу?
19. Какие события называются равновозможными?
20. В каком случае вероятность события вычисляется по формуле классической вероятности?
21. Как найти вероятность суммы двух несовместных событий?
22. Как найти вероятность суммы двух совместных событий?
23. Чему равна сумма вероятностей противоположных событий?
24. Как определяется условная вероятность события?
25. Какие события называются независимыми?
26. Как найти вероятность произведения двух событий?
27. Чему равна вероятность произведения двух независимых событий?
28. Чему равна сумма вероятностей гипотез в формуле полной вероятности?
29. Как пересчитать вероятности гипотез после опыта с учётом наблюденного результата?
30. В каком случае опыты называются независимыми?
31. Какая вероятность вычисляется по формуле Бернулли?
32. Как найти наиболее вероятное число появлений события в данной серии опытов?

Тема 2. Случайные величины

Понятие случайной величины. Закон распределения случайной величины. Ряд распределения. Функция распределения и её свойства.

Плотность распределения. Числовые характеристики случайной величины. Нормальное распределение. Распределение Бернулли. Распределение Парето. Показательный (экспоненциальный) закон.

Рекомендуемая литература: [1, раздел 1, тема 2].

Методические указания

Понятие случайной величины относится к важнейшим понятиям теории вероятностей. При рассмотрении случайных величин главную роль играет не пространство элементарных событий, на котором определена случайная величина как числовая функция, а закон распределения этой случайной величины. Случайные величины условно разделяют на два вида: дискретные и непрерывные. Универсальным способом задания закона распределения случайной величины является функция распределения. Однако во многих случаях необязательно знать закон распределения случайной величины, полностью её описывающий. Достаточно знать числовые характеристики случайной величины, т.е. константы, характеризующие наиболее важные черты её закона распределения. Особенно важно уяснить вероятностный смысл основных числовых характеристик (мода, медиана, математическое ожидание, дисперсия, среднее квадратическое отклонение).

Обратить внимание надо также на нормальный закон распределения, так как он имеет важнейшее практическое значение.

Ниже сформулированы теоретические вопросы, которые помогут подготовиться к выполнению индивидуальных заданий. Приступать к решению задач можно только в том случае, если ответы на вопросы не вызвали затруднений. В противном случае необходимо снова обратиться к рекомендованной литературе.

Вопросы для самопроверки

1. Что такое случайная величина?
2. Какие случайные величины являются дискретными, непрерывными?
3. Что такое закон распределения случайной величины?
4. Что такое ряд распределения случайной величины?
5. Что такое (интегральная) функция распределения случайной величины?
6. Какими свойствами обладает (интегральная) функция распределения случайной величины?
7. Что такое плотность распределения случайной величины?
8. Какими свойствами обладает плотность распределения случайной величины?
9. Что называется кривой распределения случайной величины?

10. Какими способами может быть задан закон распределения для дискретной случайной величины?
11. Какими способами может быть задан закон распределения для непрерывной случайной величины?
12. Как связаны функция распределения и плотность распределения непрерывной случайной величины?
13. Что такое математическое ожидание случайной величины?
14. Какой вероятностный смысл имеет математическое ожидание случайной величины?
15. Как определяется математическое ожидание дискретной случайной величины?
16. Как определяется математическое ожидание непрерывной случайной величины?
17. Что такое мода случайной величины?
18. Как определяется мода дискретной случайной величины?
19. Как определяется мода непрерывной случайной величины?
20. Что такое медиана непрерывной случайной величины?
21. Что такое дисперсия случайной величины?
22. Какой вероятностный смысл имеет дисперсия?
23. Что такое среднее квадратическое отклонение?
24. Как определяется дисперсия дискретной случайной величины?
25. Как определяется дисперсия непрерывной случайной величины?
26. Что такое начальные и центральные моменты?
27. Как центральные моменты выражаются через начальные моменты?
28. Что является начальным моментом первого порядка?
29. Что является центральным моментом второго порядка?
30. Какие числовые характеристики являются характеристиками расположения?
31. Какие числовые характеристики являются характеристиками рассеивания?
32. Что такое нормальный закон распределения?
33. Какие параметры имеет нормальный закон распределения?
34. Как определяется функция распределения стандартизованного нормального закона распределения?
35. Как связаны функция стандартизованного нормального закона распределения и функция Лапласа?
36. В чем состоит правило трёх сигм для нормального закона распределения?

Тема 3. Функции случайных величин и предельные теоремы теории вероятностей

Системы случайных величин. Закон распределения системы двух случайных величин. Числовые характеристики системы двух случайных величин. Функции случайных величин. Теоремы о числовых характеристиках случайных величин. Закон больших чисел. Теорема Бернулли. Центральная предельная теорема. Локальная и интегральная теоремы Лапласа.

Рекомендуемая литература: [2, главы 6-8].

Методические указания

При изучении этой темы не возникнет трудностей, если предыдущие темы хорошо усвоены. Для двумерных случайных величин вводятся понятия, аналогичные соответствующим понятиям для одномерных случайных величин. Однако появляются и новые понятия, такие как условные распределения, зависимость случайных величин, коррелированность, регрессия, корреляция и др.

Предельные теоремы теории вероятностей выражают связь теории вероятностей с практикой и являются основой для статистических исследований. Особое значение имеет теорема Бернулли, которая позволяет устанавливать вероятность события.

Ниже сформулированы теоретические вопросы, которые помогут подготовиться к выполнению индивидуальных заданий. Приступать к решению задач можно только в том случае, если ответы на вопросы не вызвали затруднений. В противном случае необходимо снова обратиться к рекомендованной литературе.

Вопросы для самопроверки

1. Что такое двумерная случайная величина?
2. Что такое компонента двумерной случайной величины?
3. Какими способами может быть задан закон распределения двумерной случайной величины?
4. Как определить законы распределения компонент, если известен закон распределения двумерной случайной величины?
5. Что такое функция распределения двумерной случайной величины?
6. Какими свойствами обладает функция распределения двумерной случайной величины?
7. Какие числовые характеристики определяются для двумерной случайной величины?
8. Что такое центр рассеивания двумерной случайной величины?

9. Как найти корреляционный момент двумерной случайной величины?
10. Какой вероятностный смысл имеет корреляционный момент?
11. Что такое условный закон распределения случайной величины?
12. Какие значения может принимать коэффициент корреляции?
13. Какими свойствами обладает математическое ожидание случайной величины?
14. Какими свойствами обладает дисперсия случайной величины?
15. В каком случае коэффициент корреляции по абсолютной величине равен единице?
16. В каком случае коэффициент корреляции равен нулю?
17. Что следует из равенства нулю коэффициента корреляции?
18. Чему равно математическое ожидание суммы двух случайных величин?
19. Чему равна дисперсия суммы двух случайных величин?
20. Чему равно математическое ожидание произведения двух случайных величин?
21. Чему равна дисперсия произведения двух независимых случайных величин?
22. Какие теоремы называются законом больших чисел?
23. Какие теоремы называются центральной предельной теоремой?

Тема 4. Определение законов случайных величин на основе опытных данных

Выборка и способы её записи. Графическое представление выборки. Точечные оценки параметров распределения. Интервальные оценки параметров распределения. Доверительные интервалы для параметров нормального распределения. Вычисление и статистический анализ параметров линейной модели.

Рекомендуемая литература: [1, раздел 2, тема 1, тема 2, 2.2.1 – 2.2.12].

Методические указания

При изучении этой темы необходимо уяснить разницу между случайной величиной и генеральной совокупностью, между генеральной совокупностью и выборкой. Отметим, что основной целью изучения статистических данных является установление статистических закономерностей, присущих массовым случайным явлениям. Поэтому особое внимание следует обратить на методы формирования выборки. Различные способы записи выборки служат для удобства анализа имеющихся данных. При изучении доверительных интервалов особое внимание следует уделить математико-статистическим таблицам.

Ниже сформулированы теоретические вопросы, которые помогут подготовиться к выполнению индивидуальных заданий. Приступать к решению задач можно только в том случае, если ответы на вопросы не вызвали затруднений. В противном случае необходимо снова обратиться к рекомендованной литературе.

Вопросы для самопроверки

1. Что такое выборка, объем выборки?
2. Что такое генеральная совокупность?
3. Какого типа могут быть результаты наблюдений?
4. Какие наблюдения называются непрерывными?
5. Какие наблюдения называются дискретными?
6. Что такое вариационный ряд?
7. Что такое статистический ряд для: непрерывных наблюдений; дискретных наблюдений?
8. Как определяется объем выборки по сгруппированному ряду?
9. Как определяется число классов для интервального ряда?
10. Как представляется графически интервальный ряд?
11. Как представляется графически сгруппированный ряд?
12. Как определяется эмпирическая функция распределения?
13. В каком интервале может принимать значения эмпирическая функция распределения?
14. Чему равна площадь гистограммы, построенной в координатах (x, m), где m – частота.
15. Чему равна площадь гистограммы, построенная в координатах (x, m/n), где m/n – частость?
16. Как определяется среднее арифметическое выборки?
17. Как определяется среднее арифметическое сгруппированного ряда, интервального ряда?
18. Как определяется выборочная дисперсия: для выборки, для сгруппированного ряда, для интервального ряда?
19. Как определяется выборочное среднее квадратическое отклонение?
20. С помощью каких числовых характеристик можно установить симметричность распределения?
21. Какие числовые характеристики относятся к структурным средним?
22. Как определяется мода?
23. Как определяется медиана?
24. Как определяется коэффициент вариации?

25. Влияют ли крайние члены вариационного ряда на медиану, среднее арифметическое?
26. Что такое n-процентиль?
27. Как определяются квартили?
28. Какие типы оценок используются в математической статистике?
29. Что является точечной оценкой для математического ожидания?
30. Что является точечной оценкой для дисперсии?
31. Что такое доверительный интервал?
32. Что такое уровень значимости?
33. Что такое доверительная вероятность?
34. Как доверительная вероятность связана с уровнем значимости?
35. Как задается уровень значимости?
36. Какие условия влияют на выбор формулы для определения доверительной оценки для математического ожидания?
37. Какой вид имеют формулы для интервального оценивания математического ожидания нормального распределения?

Раздел 5. Проверка статистических гипотез

Основные понятия. Критерии значимости для проверки гипотез о параметрах нормально распределённой генеральной совокупности. Проверка гипотез о коэффициенте корреляции. Проверка гипотезы о виде распределения генеральной совокупности. Проверка гипотезы о независимости двух случайных величин.

Рекомендуемая литература: [1, раздел 2, тема 2, 2.2.13 – 2.2.19, тема 3].

Методические указания

Изучение этой темы начните с чтения учебного пособия, выучите все новые понятия и определения. Обратите внимание на то, что не всякое предположение может быть статистической гипотезой. Например, высказывание типа – на Марсе есть жизнь – не является статистической гипотезой, так как в ней не идет речь ни о законе распределения, ни о параметрах распределения случайной величины.

Ниже сформулированы теоретические вопросы, которые помогут подготовиться к выполнению индивидуальных заданий. Приступать к решению задач можно только в том случае, если ответы на вопросы не вызвали затруднений. В противном случае необходимо снова обратиться к рекомендованной литературе.

Вопросы для самопроверки

1. Что такое статистическая гипотеза?

2. Какие гипотезы выдвигаются в задачах проверки гипотез?
3. Какая гипотеза называется нулевой гипотезой?
4. Какая гипотеза называется альтернативной гипотезой?
5. Что такое ошибка 1-го рода?
6. Что такое ошибка 2-го рода?
7. Как определяются вероятности ошибок 1-го и 2-го рода?
8. Если вероятность ошибки 1-го рода уменьшается, то что происходит с вероятностью ошибки 2-го рода?
9. Что такое критическая точка?
10. Что такое критическая область?
11. Что такое критерий для проверки гипотез?
12. В каком случае отвергается нулевая гипотеза?
13. В каком случае принимается нулевая гипотеза?
14. Критерий принятий нулевой гипотезы при сравнении двух средних.
15. Критерий принятий нулевой гипотезы при сравнении двух дисперсий
16. Что такое стохастическая связь между случайными переменными?
17. Как можно оценить стохастическую связь?
18. Что оценивает выборочный коэффициент корреляции?
19. Что оценивает ранговый коэффициент корреляции?
20. Как определяется значимость коэффициента корреляции?
21. Что называется регрессией **y** на **x**?
22. Как задается парная линейная регрессия?
23. Какой метод используется для оценки коэффициентов парной линейной регрессии?
24. Как определяются коэффициенты парной линейной регрессии?
25. Как оценивается качество аппроксимации результатов наблюдений регрессионной моделью?

3. СОДЕРЖАНИЕ ПРАКТИЧЕСКОГО РАЗДЕЛА ДИСЦИПЛИНЫ

3.1. Тематика практических занятий

1. Вычисление вероятностей событий (2 часа).
2. Случайные величины дискретного типа (2 часа).
3. Случайные величины непрерывного типа (2 часа).
4. Оценки параметров распределений (2 часа).
5. Проверка статистических гипотез (2 часа).

4. ИНДИВИДУАЛЬНЫЕ ДОМАШНИЕ ЗАДАНИЯ

4.1. Общие методические указания

Основной формой обучения студента заочного отделения является его самостоятельная работа над учебным материалом, которая состоит из следующих элементов: изучение материала по учебникам, решение задач, ответы на вопросы для самопроверки, выполнение домашних индивидуальных заданий (ИДЗ).

В соответствии с учебным графиком для студентов, обучающихся по направлениям 080100 «Экономика», 080200 «Менеджмент», 080400 «Управление персоналом», 100700 «Торговое дело», предусмотрено выполнение двух индивидуальных домашних заданий. Выполнение этих заданий необходимо для закрепления теоретических знаний и приобретения практических навыков решения типовых задач. Индивидуальное задание №1 соответствует темам 1-3 раздела 2 «Содержание теоретического раздела дисциплины». Индивидуальное задание №2 соответствует темам 4-5 раздела 2.

Студент выполняет вариант индивидуального домашнего задания, номер которого совпадает с последней цифрой шифра его зачётной книжки. Например, если номер зачетной книжки З-3В11/14, но студент выбирает вариант индивидуального домашнего задания под номером 4, если шифр З-3В11/20, то выбирать надо вариант под номером 0.

Индивидуальные задания выполняются в соответствии с графиком изучения дисциплины и высылаются на проверку преподавателю. Работы следует выполнять в течение семестра, чтобы к моменту сессии они уже были прорецензированы.

При оформлении индивидуального домашнего задания необходимо соблюдать следующие требования:

1. Обязательно должен быть титульный лист. На титульном листе указываются номер индивидуального задания, номер варианта, название дисциплины; фамилия, имя, отчество студента; номер группы, шифр.

2. Все страницы работы должны иметь сквозную нумерацию.

3. Обязательно прилагается список использованной литературы. В этот список необходимо включить рабочую программу и методические указания, в соответствии с которыми выполнены задания.

4. Решения задач следует располагать в той же последовательности, что и задания. Перед решением следует записать текст условия задачи.

5. Решения всех задач должны быть подробными, со всеми промежуточными расчётами, с указанием использованных формул и т.п. При необходимости студент должен давать пояснения по всем или некоторым задачам индивидуального задания.

6. В случае не соответствия работы требованиям к оформлению студент получает оценку «незачтено». В этом случае работа должна быть исправлена и повторно предоставлена на проверку преподавателю.

7. Студент, не получивший положительной аттестации хотя бы по одному индивидуальному заданию, не допускается к сдаче экзамена по данной дисциплине.

Перед выполнением индивидуального домашнего занятия студент должен ознакомиться с литературой, рекомендуемой по каждой теме, включенной в теоретический раздел дисциплины.

Студент может обращаться к преподавателю с вопросами для получения устной или письменной консультации. Указания студенту по текущей работе даются также в процессе рецензирования контрольных работ. Кроме того, студентам читаются обзорные лекции по наиболее важным и трудным разделам курса, проводятся практические занятия.

Студенты, обучающиеся по классической заочной форме (КЗФ) каждое индивидуальное задание оформляют в отдельной тетради.

Студенты, обучающиеся с использованием дистанционных образовательных технологий (ДОТ) каждое индивидуальное задание оформляют в отдельном файле. Студенты, обучающиеся с использованием ДОТ, в обязательном порядке получают письменную рецензию на каждое индивидуальное задание. Работы студентам не возвращаются.

4.2.2. Методические указания к решению индивидуального задания № 1

Классическое определение вероятности

Случайным событием называется событие, которое в результате эксперимента может произойти или не произойти. При этом рассматриваются только такие эксперименты, которые можно повторять, при неизменном комплексе условий, произвольное число раз.

Результатом эксперимента является **случайный исход** или **элементарное событие** $\mathbf{e_i}$. Все возможные элементарные исходы образуют выборочное пространство $\mathbf{E = \{e_i\}}$, где $\mathbf{i = 1, \ldots, n}$. Случайное событие может совпадать со случайным исходом эксперимента или являться какой-либо комбинацией этих исходов. Если случайное событие обозначить **A**, а все возможные исходы эксперимента $\mathbf{e_i}$, то каждому случайному событию **A** будет благоприятствовать какое-то число элементарных событий (исходов) $\mathbf{e_i}$ (ими будут $\mathbf{e_i}$, через которые можно определить случайное событие **A**).

Классическое определение вероятности случайного события **A** используется для экспериментов со случайными исходами $\mathbf{e_i}$ (элементарными событиями), которые можно повторять (воспроизводить) при неизменном комплексе условий произвольное число раз и при этом каждый раз любой исход $\mathbf{e_i}$ имеет один и тот же шанс появиться. В этом случае вероятность любого события **A**, являющегося комбинацией $\mathbf{e_i}$, определяется по следующей формуле:

$$\mathbf{P(A) = \frac{m}{n}},$$

где **m** число элементарных исходов $\mathbf{e_i}$, которые благоприятствуют появлению **A**, и **n** – число всевозможных исходов.

Пример. Эксперимент состоит в подбрасывании один раз правильной шестигранной игральной кости. Пусть **x** – число очков, выпавших на верхней грани кости. Описать множество всех элементарных событий, и указать состав подмножеств, соответствующих событиям **A = {x кратно 3}, B = {x нечетно}, C** = {**x > 3**}. Определить вероятности случайных событий **A, B, C.**

Решение. Множество всех возможных случайных исходов составляют элементарные события, состоящие в выпадении $\mathbf{e_1 - 1}$, $\mathbf{e_2 - 2}$, $\mathbf{e_3 - 3}$, $\mathbf{e_4 - 4}$, $\mathbf{e_5 - 5}$, $\mathbf{e_6 - 6}$. Событиям **A, B, C** соответствуют подмножества из $\{\mathbf{e_i}\}$, в которые входят следующие исходы:

$\mathbf{A = \{e_3, e_6\}}$, $\mathbf{B = \{e_1, e_3, e_5\}}$, $\mathbf{C = \{e_4, e_5, e_6\}}$.

$\mathbf{P(A) = 2/6}$, $\mathbf{P(B) = 3/6 = 1/2}$, $\mathbf{P(C) = 3/6 = 1/2}$.

Комбинаторный метод вычисления вероятностей в классической схеме

Предполагается, что производится выбор **m** элементов из **n** различных элементов всех возможных исходов $\mathbf{E = \{e_1, e_2, ..., e_n\}}$. При этом предполагается, что выбор осуществляется без возвращения. Это предполагает, что выбираются сразу все **m** элементов из **n** , или по одному, причем каждый выбранный исключается. В результате имеем различные схемы по выбору **m** элементов из общего числа исходов **E**.

1. Схема выбора, приводящая к сочетаниям

Если опыт состоит в выборе **m** элементов без возвращения и упорядочивания, то различными исходами будут **сочетания из n элементов по m**. Их общее число определяется по формуле

$$\mathbf{C_n^m = \frac{n!}{m!(n-m)!} = \frac{1 \cdot 2 \cdot 3 \cdot ... \cdot n}{1 \cdot 2 \cdot ... \cdot m \cdot 1 \cdot 2 \cdot ... \cdot (n-m)}}.$$

Пример. В урне **10** белых и **5** черных шаров. Из урны вынимают **2** шара. Определить вероятность того, что оба шара будут белыми.

Решение. Событие **A** – оба шара будут белыми.

Общее число случаев

$$\mathbf{C_{10+5}^2 = \frac{15!}{2!13!} = \frac{1 \cdot 2 \cdot 3 \cdot ... \cdot 14 \cdot 15}{1 \cdot 2 \cdot 1 \cdot 2 \cdot ... \cdot 12 \cdot 13} = 105}.$$

Число благоприятных случаев

$$C_{10}^2 = \frac{1 \cdot 2 \cdot ... \cdot 9 \cdot 10}{1 \cdot 2 \cdot 1 \cdot 2 \cdot ... \cdot 8} = 45.$$

Тогда по формуле классической вероятности получаем

$P(A) = 45/105 = 3/7$.

Пример. Ящик содержит **6** деталей, среди которых **2** детали неисправны, **4** детали исправны. **Три** из **шести** деталей выбираются из ящика.

а. Какое количество элементарных исходов мы имеем в данном случае?

б. Какова вероятность, что одна из 3-х деталей будет неисправна?

Решение.

а. $$\mathbf{n = C_6^3 = \frac{6!}{3!3!} = \frac{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 1 \cdot 2 \cdot 3} = 20}.$$

б. $$\mathbf{P(A) = \frac{C_2^1 \cdot C_4^2}{C_6^3} = \frac{2 \cdot \frac{4!}{2!2!}}{20} = \frac{2 \cdot 6}{20} = 0.6}.$$

Число благоприятных исходов в данном случае определяется как произведение числа комбинаций для выбора **одной** неисправной детали из **двух** и числа комбинаций для выбора **остальных двух** деталей из **трех** исправных.

2. Схема выбора, приводящая к размещениям

Если опыт состоит в выборе **m** элементов без возвращения, но с упорядочиванием их по мере выбора, в последовательную цепочку, то различными исходами будут **m** элементарных подмножеств множества **E,** отличающиеся либо набором элементов, либо порядком их следования. Получаемые при этом комбинации называются **размещениями из n элементов по m**, а их общее число определяется формулой

$$A_n^m = \frac{n!}{(n-m)!} = \frac{1\cdot 2\cdot 3\cdot \ldots \cdot n}{1\cdot 2\cdot \ldots \cdot (n-m)}.$$

Пример. На пяти карточках написаны цифры от **1** до **5**. Опыт состоит в случайном выборе трех карточек и раскладывании их в порядке поступления в ряд слева направо. Найти вероятность события **А = {появится число 123}.**

Решение. Общее число случаев $n = A_n^m = A_5^3 = \frac{5!}{2!} = 3\cdot 4\cdot 5 = 60$.

Благоприятных исходов 1, т.е. **Р(А)=1/60.**

В частном случае упорядочивания, когда ***n*** *=* ***m*** *имеем* $A_n^m = n!$ или ***число перестановок*** P_n ***из n элементов****, где* $P_n = 1\cdot 2\cdot \ldots \cdot n = n!$

Пример. Из урны, содержащей **n** перенумерованных шаров, наугад вынимаются один за другим все находящиеся в ней шары. Найти вероятность того, что номера вынутых шаров будут идти по порядку **1, 2, …, n.**

Решение. Общее число случаев $n = n!$. Благоприятных случаев **m** = **1**, $P(A) = 1/n!$.

Примечание: **0! = 1**.

Классификация событий

Сложным событием называется событие, выражаемое через другие наблюдаемые события в том же эксперименте с помощью допустимых алгебраических операций. Допустимыми алгебраическими операциями над случайными событиями являются сложение, умножение, дополнение, разность.

Суммой двух событий А и В является событие **С**, которое состоит в появлении хотя бы одного из **А** и **В**.

Произведением событий **А** и **В** является событие **С**, состоящее в совместном осуществлении **А** и **В.**

Дополнение $\overline{\mathbf{A}} = \mathbf{\Omega} - \mathbf{A}$ - противоположное событие, где $\mathbf{\Omega}$ - событие, которое происходит всегда, т.е. $\mathbf{P(\Omega) = 1}$**,** событие $\mathbf{\Omega}$ называется **достоверным** событием.

Разность А – В - событие **С**, состоящее в том, что событие **А** происходит, а событие **В** не происходит.

Событие $\varnothing$ называется **невозможным** и его вероятность $\mathbf{P(\varnothing) = 0}$**.**

События **А** и **В** называются **несовместными**, если появление события **А** исключает появление события **В**.

Операции над событиями помогают упростить вычисление вероятностей событий.

Формула сложения вероятностей **P(A+B)=P(A)+P(B)-P(AB)**, если **А** и **В** совместны, **P(A+B)=P(A)+P(B)**, если **А** и **В** несовместны.

Пример. В урне **12 белых** и **7 черных** шаров. Из урны вынимается **2** шара. Какова вероятность получить **один черный** шар и **один белый** шар?

Решение. Пусть **А** событие – вытащили **один белый** шар и **один черный** шар. Событие **А** произойдет, если произойдут следующие события:

В – в **первый** раз извлечен **белый** шар, во **второй черный** или

С - в **первый** раз извлечен **черный** шар, во **второй белый**.

Событие **А** можно представить как сумму событий **В** и **С**, т.е. **А=В+С**. События **В** и **С** несовместны, значит

$\mathbf{P(A) = P(B) + P(C);}$

$\mathbf{P(B) = 12/19 \cdot 7/18;}$

$\mathbf{P(C) = 12/19 \cdot 7/18;}$

$$\mathbf{P(A) = 12/19 \cdot 7/18 + 12/19 \cdot 7/18 = \frac{28}{57}.}$$

Условная вероятность. Независимость событий

Условная вероятность это вероятность события **А** при условии, что появилось событие **В.** Условная вероятность события **А** при условии, что произошло событие **В** обозначается как **P(A/B).**

Условная вероятность определяется как отношение вероятности совместного появления событий **А** и **В** к вероятности события **В**: **P(A/B)=P(AB)/P(B).** Аналогично определяется условная вероятность события **В** относительно **А**. Определение условной вероятности выражает принцип умножения вероятностей: вероятность совместного появления двух событий равна вероятности одного из них, умноженной на условную вероятность другого относительно первого:

P(AB)=P(A)P(B/A)=P(B)P(A/B).

Событие **A** зависит от события **B, если $P(A/B) \neq P(A)$,** т.е, вероятность события **A** изменяется при появлении события **B**. Если **A** не зависит от **B,** то **$P(A/B)=P(A)$.**

Пример 1. В урне имеется **5 черных и 7 белых** шаров. Опыт состоит в том, что из урны вынимается **один** шар и регистрируется его цвет, после чего шар **возвращается** в урну. После перемешивания снова вынимается **один** шар. Найти вероятность того, что оба раза будет вынут **белый** шар.

Решение. Пусть событие **A** появление **белого** шара в **первый** раз, событие **B** появление **белого** шара во **второй** раз. Как в первый так и во второй раз благоприятных возможностей вынуть белый шар **7**, а общее число возможностей **12.** Следовательно, в данном случае **$P(A) = P(B) = 7/12$**;
$P(A/B) = P(B/A) = 7/12$. Поэтому, события **A** и **B** в данном случае независимы.

Пример 2. В условиях предыдущего примера найти вероятность появления **двух белых** шаров, если **белый** вынутый шар в урну **не возвращается.**

Решение. В этом случае при вынимании **второго** шара имеется **11** случаев, из которых **6** благоприятных при условии, что **первый** шар **- белый.** Следовательно, **$P(B/A)=6/11$**, если **первый** шар был **черным.** то **$P(B/A)=7/12$**. Таким образом, появление белого шара во второй раз зависит от того, какой шар был вынут в первый раз, т.е. **$P(B/A) \neq P(B)$,** события **A** и **B** в данном случае зависимы.

$P(AB)=7/12 \cdot 6/11=7/22$, если первый шар белый.

$P(AB) =7/12 \cdot 7/11=49/132$, если первый шар черный.

Формула полной вероятности и формула Байеса

Пусть случайное событие **A** может произойти только с одним из событий **$H_1, \ldots, H_n$**, попарно несовместных. Тогда вероятность события **A** может быть определена по формуле

$$P(A)=P(H_1)P(A/H_1)+\ldots+P(H_n)P(A/H_n).$$

События **$H_1, \ldots, H_n$** называются **гипотезами.** Гипотезы образуют полную группу несовместных событий, поэтому сумма их вероятностей равна единице.

Пример. Имеются три одинаковые урны. В первой находятся **a** белых шаров и **b** черных, во второй – **c** белых и **d** черных; в третьей – только белые шары. Некто подходит наугад к одной из урн и вынимает из нее один шар. Найти вероятность того, что этот шар белый.

Решение. Рассмотрим событие **A** – появление белого шара. Сформулируем гипотезы:

H_1 - выбор первой урны;
H_2 - выбор второй урны;
H_3-выбор третьей урны.
$P(H_1)=P(H_2)=P(H_3)=1/3$.
Находим условные вероятности:

$$P(A/H_1)=\frac{a}{a+b};$$

$$P(A/H_2)=\frac{c}{c+b};$$

$$P(A/H_3)=1;$$

Найденные значения подставляем в формулу полной вероятности
$P(A)=P(H_1)P(A/H_1)+P(H_2)P(A/H_2)+P(H_3)P(A/H_3)$.
Получаем

$$\mathbf{P(A)=\frac{1}{3}\cdot\frac{a}{a+b}+\frac{1}{3}\cdot\frac{c}{c+d}+\frac{1}{3}\cdot 1=\frac{1}{3}\cdot\left(\frac{a}{u+b}+\frac{c}{c+d}+1\right)}.$$

Формула Байеса имеет вид:

$$P(H_i\,|\,A)=\frac{P(H_i)P(A|H_i)}{P(H_i)P(A|H_i)+...+P(H_n)P(A|H_n)},\qquad i=1,2,\ldots,n.$$

Эта формула используется для вычисления условной вероятности ***$P(H_i/A)$*** гипотезы ***H_i*** после испытания, при котором произошло событие ***A***. Формула Байеса позволяет переоценить вероятности гипотез, вычисленные до испытания (априорные), по результатам уже проведенного опыта.

Пример. Пусть в предыдущем примере некто вынул шар, и он оказался белым. Определить вероятность того, что он вынул его из первой урны.

Решение. По формуле Байеса пересчитаем вероятность первой гипотезы при условии, что событие **A** – появление белого шара уже произошло.

$$\mathbf{P(H_1/A)=\frac{\frac{1}{3}P(A/H_1)}{\frac{1}{3}\sum_{i=1}^{3}P(A/H_i)}=\frac{\frac{a}{a+b}}{\frac{a}{a+b}+\frac{c}{c+d}+1}}.$$

Отметим, что при использовании формулы полной вероятности рекомендуется придерживаться следующей последовательности действий:

1. Записать событие **A**, вероятность наступления которого требуется найти;

2. Сформулировать гипотезы;

3. Найти вероятности гипотез;

4. Сделать проверку (сумма вероятностей всех гипотез должна быть равна единице);

5. Найти условные вероятности события **А** при условии, что произошло событие $\boldsymbol{H_i}$ (если они не даны в условии);

6. По формуле полной вероятности найти вероятность события **А**.

Если из условия задачи известно, что событие **А** уже произошло, то по формуле Байеса необходимо найти условные вероятности гипотез.

Случайные величины

Случайной величиной (СВ) называется величина **X,** которая в результате опыта может принимать то или иное значение, причем заранее (до опыта) неизвестное. В теоретико-вероятностной схеме случайной величиной называют числовую функцию от элементарных исходов $\mathbf{X = X(e_i)}$. Следовательно, можно говорить о числовой функции $\mathbf{X(e_i)}$, определенной на пространстве элементарных событий. Значения случайных величин обозначаются **x, y, z, ...**

Случайные величины могут быть дискретными и непрерывными.

Дискретной случайной величиной называется величина, которая принимает конечное или бесконечное счетное множество значений.

Непрерывные случайные величины – это величины, значения которых непрерывно заполняют некоторый числовой отрезок.

В качестве примеров можно рассмотреть следующие.

1. Предположим эксперимент состоит в том, что необходимо определить число автомобилей, проезжающих по магистрали за **30** сек. Возможные значения случайной величины **0, 1, 2, …, n**.

2. Предположим, что нужно измерить время завершения двух операций на конвейере. Возможные значения, которые принимает случайная величина от **0** секунд до **n** секунд.

В первом случае имеем дискретную случайную величину, во втором - непрерывную.

Случайная величина считается заданной, если задан ее закон распределения. **Законом распределения случайной величины** называется всякое соотношение между значениями, принимаемыми случайной величиной и вероятностями, с которыми она принимает эти значения. Закон распределения позволяет полностью описать случайную величину с вероятностной точки зрения.

Для **дискретных** случайных величин закон распределения задается в виде **ряда распределения** и **функции распределения**.

Для **непрерывных** случайных величин закон распределения задается в виде **функции распределения** и **плотности распределения**, которые будут рассмотрены ниже.

Для **дискретной** СВ закон распределения может быть задан рядом распределения, который представляется таблицей соответствия значений СВ и соответствующих им вероятностей. Ряд распределения может быть также задан графиком, когда по оси **x** откладываются значения случайной величины, а по **y** значения вероятностей; ломаная линия, соединяющая точки $(\mathbf{x_i}, \mathbf{p_i})$ называется **многоугольником распределения**. Кроме того закон распределения СВ может быть задан аналитически в виде функции распределения.

Функцией распределения называется вероятность того, что случайная величина **X** примет значение, меньшее произвольно выбранного значения **x**, то есть $\mathbf{X \in [-\infty; x]}$;

$$\mathbf{F(x) = P(X < x) = P}.$$

Для дискретных случайных величин функция распределения **F(x)** вычисляется по формуле $\mathbf{F(x) = \sum_{x_i < x} P(X < x_i)}$, где суммирование ведется по тем значениям **i**, для которых $\mathbf{x_i < x, i = 1,2,\ldots}$ Свойства функции распределения приводятся далее (см. функция распределения для непрерывных случайных величин).

Наиболее часто встречающиеся законы распределения дискретных случайных величин

1. Гипергеометрическое распределение $\mathbf{P(x = k) = C_M^k C_{N-M}^{n-k} / C_N^n}$, где **N, M, n** – натуральные числа: $\mathbf{M \le N, n \le N, k = 0,1,\ldots,\min(n, M)}$.

Пример. Из партии, содержащей **M** белых и **N-M** черных, наудачу извлекаются **n** шаров. Какова вероятность того, что среди выбранных **n** шаров окажется ровно **k** белых?

В этом примере СВ, равная числу шаров, имеет гипергеометрическое распределение $\mathbf{P_n(k) = C_M^k C_{N-M}^{n-k} / C_N^n}$, где $\mathbf{k = k_0, k_0 + 1, \ldots, k_1}$, $\mathbf{k_0 = \max(0, M - N + n)}$, $\mathbf{k_1 = \min(n, M)}$.

2. Распределение СВ, равное числу успехов в **n** испытаниях, определяется по биноминальному закону распределения:

$$\mathbf{P(x = k) = P_n(k) = C_n^k p^k q^{n-k}}, \qquad (1)$$

где **n** – число испытаний, **k** – число успехов, **p** – вероятность успеха в одном испытании, $\mathbf{q = 1 - p}$ - вероятность неудачи.

Формула (1) называется **формулой Бернулли**.

3. Для бесконечной последовательности испытаний Бернулли СВ **X**, равная числу испытаний до первого успеха включительно, имеет геометрическое распределение

$$\mathbf{P(x = k) = (1 - p)^{k-1} p},$$

где **p** – вероятность успеха **k = 1, 2, …** .

4. Если вероятность появления события в последовательности испытаний Бернулли мала, а **n** велико, то применение формулы Бернулли затруднительно, в этом случае пользуются ее предельным значением $\mathbf{P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}}$, где $\mathbf{\lambda = np}$. Это формула Пуассона и распределение называется Пуассоновским.

Пример 1. Из партии, содержащей **100** изделий, среди которых **10** дефектных, выбраны случайным образом **5** изделий для проверки их качества. Построить ряд распределения случайного числа X дефектных изделий, содержащихся в выборке.

Решение. Случайная величина **X** распределена по гипергеометрическому закону, то есть $\mathbf{P(X = k) = \frac{C_{10}^k C_9^{5-k}}{C_{100}^5}}$, **k** – число бракованных изделий, **k = 0, 1, 2, 3, 4, 5**. Ряд распределения представляется следующей таблицей:

x_i	0	1	2	3	4	5
p_i	0.583	0.340	0.07	0.007	0	0

Пример 2. Вероятность того, что некоторая деталь окажется дефектной, равна **p**. В случае обнаружения дефекта линию останавливают и делают переналадку. Составить ряд распределения для числа **X** годных деталей между двумя переналадками.

Решение. Случайная величина **X** принимает значения **0, 1, 2, …, k, …** . **X = k**, когда **(k+1)**-ая деталь окажется дефектной. Следовательно,

$$\mathbf{P_k = P(X = k) = q^k \cdot p \quad (k = 0,1,2,...)}$$

и ряд распределения представится в виде

x_i	0	1	2	…	k	…
p_i	p	qp	q^2p	…	q^kp	…

Плотность распределения. Свойства функции распределения

Для непрерывной СВ закон распределения может быть задан не только функцией распределения **F(x)**, но и ещё одной функцией – плотностью распределения

$\mathbf{f(x)=F'(x)}$.

График плотности распределения называется кривой распределения. Функция распределения выражается через плотность распределения формулой

$$\mathbf{F(x)=\int_{-\infty}^{x} f(x)dx}.$$

Основные свойства функции распределения **F(x):**

1. $\mathbf{F(-\infty)=0;}$
2. $\mathbf{F(+\infty)=1;}$
3. **F(x) – неубывающая;**
4. $\mathbf{P(a \le x \le b)=F(b)-F(a)}$.

Основные свойства плотности распределения **f(x):**

1. $\mathbf{f(x)\ge 0;}$
2. $\mathbf{\int_{-\infty}^{+\infty} f(x)dx=1;}$
3. $\mathbf{P(a\le x\le b)=\int_{a}^{b} f(x)dx}$.

Прежде, чем ввести наиболее встречающиеся законы распределения непрерывных СВ, рассмотрим основные числовые характеристики СВ.

Числовые характеристики случайных величин

Математическое ожидание СВ **X** есть число, обозначаемое **M(x)**, а также $\mathbf{m_x}$ и μ, и определяемое по формуле

$$\mathbf{M(x)}=\begin{cases}\sum_i \mathbf{x_i p_i}, & \text{для дискретных } \mathbf{CB\ X},\\ \int_{-\infty}^{+\infty} \mathbf{xf(x)dx}, & \text{для непрерывных } \mathbf{CB\ X},\end{cases}$$

где $\mathbf{p_i=P(X=x_i)}$, **f(x)** – плотность распределения.

Предполагается, что $\sum_i \mathbf{x_i p_i}$ и $\int_{-\infty}^{+\infty}\mathbf{xf(x)dx}$ абсолютно сходятся, в противном случае считают, что **M(x)** не существует. **M(x)** является характеристикой положения СВ **X**.

Дисперсией D(x) СВ X называют математическое ожидание квадрата отклонения СВ **X** от ее математического ожидания

$$\mathbf{D(x) = M(x - m_x)^2},$$

$$\mathbf{D(x)} = \begin{cases} \sum\limits_{\mathbf{i}} \mathbf{(x - m_x)^2 p_i}, \text{ для дискретных } \mathbf{СВ\ X}, \\ \int\limits_{-\infty}^{+\infty} \mathbf{(x - m_x)^2 f(x) dx}, \text{ для непрерывных } \mathbf{СВ\ X}. \end{cases}$$

Дисперсия является характеристикой рассеяния СВ X. Кроме нее характеристикой рассеяния является $\boldsymbol{\sigma = \sqrt{D(x)}}$, называемая **средне-квадратическим отклонением**.

Пример. Производится ряд независимых испытаний, в каждом из которых может появиться некоторое событие **A**. Вероятность **A** в каждом опыте равна **p**. Опыты производятся до первого появления **A**, после чего они прекращаются. Случайная величина **X** – число произвольных опытов. Построить ряд распределений этой случайной величины и найти математическое ожидание и дисперсию.

Решение. СВ **X** имеет геометрическое распределение $\mathbf{P(x = k) = (1 - p)^{k-1} p}$, **q = 1-p, k** – число испытаний до первого успеха.

$\mathbf{x_i}$	**0**	**1**	**2**	**3**	**4**	**…**	**i**	**…**
$\mathbf{p_i}$	**p**	**qp**	**q2p**	**q3p**	**q4p**	**…**	**qip**	**…**

$$\mathbf{M(x) = \sum_{i=1}^{\infty} x_i p_i = 0 \cdot p + 1qp + 2q^2p + 3q^3p + ... + iq^ip + nq^np = p\sum_{i=1}^{\infty} iq^{i-1}},$$

$$\mathbf{D(x) = \sum_{i=1}^{\infty} (x_i - m_x)^2 p_i}.$$

На практике удобней бывает использование формулы $\mathbf{D(x) = M[x^2] - M^2(x)}$, в нашем случае

$$\mathbf{M[x^2] = \sum_{i=1}^{\infty} x_i^2 p_i = \sum_{i=1}^{\infty} i^2 q^{i-1} p_i = p\sum_{i=1}^{\infty} i^2 q^{i-1}},$$

$$\mathbf{D[x] = p\sum_{i=1}^{\infty} i^2 q^{i-1} - p(\sum_{i=1}^{\infty} iq^{i-1})^2}.$$

Основные законы распределения непрерывных СВ

Среди различных распределений особое место занимает нормальное или гауссовское распределение, которое задается плотностью распределения вида

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-m_x)^2}{2\sigma^2}} \qquad (-\infty < x < +\infty)$$

и функцией распределения вида

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{(x-m_x)^2}{2\sigma^2}} dx \qquad (-\infty < x < +\infty),$$

где m_x – математическое ожидание,

σ - среднеквадратическое отклонение.

Величины m_x и σ являются параметрами нормального закона распределения, которое обозначается $N(m_x,\sigma)$. Для каждого конкретного значения $M(x)$ и σ имеем свое нормальное распределение.

Кривая нормального распределения (кривая Гаусса) симметрична относительно прямой $x = m_x$ и имеет единственный максимум в точке $x = m_x$. Кривая нормального распределения приведена на рис. 1. Ось абсцисс является асимптотой для этой кривой.

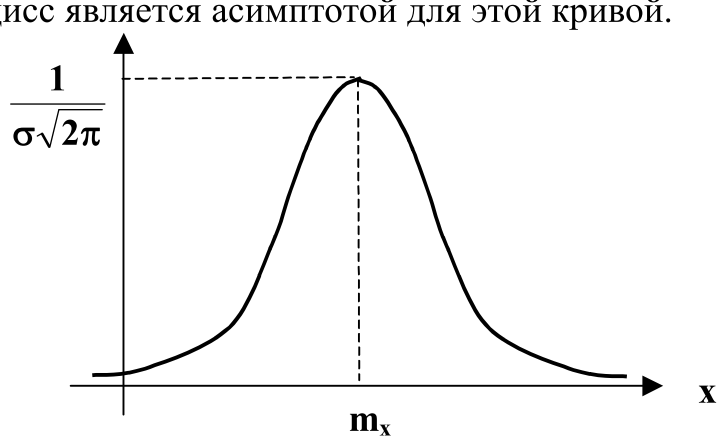


Рис. 1. График нормального распределения $N(m_x,\sigma)$

Нормальное распределение с параметрами $m_x = 0$ и $\sigma = 1$ называется стандартным или нормированным и обозначается $N(0,1)$, его плотность распределения определяется формулой

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (-\infty < x < +\infty).$$

Функция распределения стандартного (нормированного) нормального закона распределения имеет вид:

$$\Phi_{N(0,1)}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx.$$

Значения этой функции протабулированы и приводятся в таблицах (см. приложение).

Функцию распределения нормального закона с параметрами m_x и σ можно выразить через $\Phi_{N(0,1)}(x)$ следующим образом

$$F(x) = \Phi_{N(0,1)}\left(\frac{x - m_x}{\sigma}\right).$$

Вероятность попадания случайной величины на отрезок $[x_1, x_2]$ определяется в виде:

$$P(x_1 < x \le x_2) = F(x_2) - F(x_1) = \Phi_{N(0,1)}\left(\frac{x_2 - m_x}{\sigma}\right) - \Phi_{N(0,1)}\left(\frac{x_1 - m_x}{\sigma}\right).$$

Функция $\Phi_{N(0,1)}(x)$ имеет свойство

$$\Phi_{N(0,1)}(-x) = 1 - \Phi_{N(0,1)}(x).$$

Часто на практике необходимо определить значение x_p из уравнения $\Phi(x_p) = p$. Т.е. необходимо определить значение аргумента по заданному значению функции $\Phi(x_p)$, которое задается в таблице стандартного нормального распределения (смотри таблицу II приложения). Как видно из таблицы, значения функции $\Phi(x_p)$ задаются начиная со значения равного 0.5. В случае, когда необходимо определить x_p из уравнения $\Phi(x_p) = p$, при $p < 0.5$, то можно воспользоваться указанным выше свойством. Значение x_p называется **квантилью уровня p.**

Пример 1. Пусть необходимо определить x_p, из условия $\Phi_{N(0,1)}(x_p) = 0.05$.

Решение. По определению функции распределения нужно найти x_p из условия $P(x < x_p) = 0.05$. Так как значение $0.05 < 0.5$, переходим к вероятности противоположного события по формуле $P(x > x_p) = 1 - P(x < x_p) = 1 - 0.05 = 0.95$.

$1 - P(x < x_p) = \Phi_{N(0,1)}(-x_p) = 0.95$, из таблицы $x_p = -1.64$. Квантили нормального закона связаны $-x_p = -x_{1-p}$.

Часто в таблицах приложений приводится не значение функции стандартного нормального закона распределения **N(0,1),** а интеграл вероятностей или функция Лапласа, определяемая в виде

$$\Phi(x)=\frac{2}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt \text{ или } \Phi_1(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}dt,$$

где $\Phi_1(x)=1/2\Phi(x)$.

Для функции Лапласа справедливы равенства

$\Phi(-x)=-\Phi(x)$ и $\Phi_1(-x)=-\Phi_1(x)$.

Функция распределения нормального стандартного закона распределения связана с функцией Лапласа следующими соотношениями:

$\Phi_{N(0,1)}(x)=1/2(\Phi(x)+1)$ или

$\Phi_{N(0,1)}(x)=1/2(2\Phi_1(x)+1)$.

Пример 2. Определить вероятность того, что случайная величина **X** примет значение в интервале **(158, 162),** если известно, что она распределена по нормальному закону с параметрами $m_x=168,\ \sigma=5.92$.

Решение. Так как $P(\alpha<x<\beta)=\frac{1}{2}\left[\Phi\left(\frac{\beta-m_x}{\sigma}\right)-\Phi\left(\frac{\alpha-m_x}{\sigma}\right)\right]$. Вычисляем $t_1=(158-168)/5.92=-1.69$; $t_2=(162-168)/5.92=-1.014$. По таблице функции Лапласа $\Phi(x)$ имеем

$\Phi(t_2)=\Phi(-1.014)=-\Phi(1.014)=-0.6894$;

$\Phi(t_1)=-\Phi(1.690)=-0.9090.$

Следовательно, $P(158\le x\le 162)=(1/2)[-0.6894+0.9090]=0.1106$. По таблице 2 приложения функции распределения стандартного нормального закона распределения имеем

$\Phi_{N(0,1)}(t_2)=\Phi_{N(0,1)}(-1.014)=1-\Phi_{N(0,1)}(1.014)=1-0.84375=0.15625$,

$\Phi_{N(0,1)}(t_1)=\Phi_{N(0,1)}(-1.69)=1-\Phi_{N(0,1)}(1.69)=1-0.95449=0.0455$,

$\Phi_{N(0,1)}(t_2)-\Phi_{N(0,1)}(t_1)=0.15625-0.0455=0.11075$.

Вероятность того, что абсолютная величина отклонения СВ от своего математического ожидания меньше любого $\varepsilon>0$ определяется формулой $P(|x-m_x|<\varepsilon)=2\Phi_{N(0,1)}(\varepsilon/\sigma)-1$.

4.2.3. Индивидуальное задание №2

Построить интервальный ряд и выполнить пункты задания по математической статистике:

1. Представить исходную выборку в виде статистического ряда и изобразить его графически. Привести график эмпирической функции распределения.

2. Определить моду и медиану.

3. Определить точечные оценки для среднего арифметического, дисперсии, среднеквадратического отклонения.

4. Определить квартили Q_1, Q_2, Q_3.

5. Установить, является ли распределение симметричным, используя коэффициент асимметрии и графический способ Box and Whisker Plot (смотри [8], стр 90 – 91.)

6. Определить интервальные оценки для математического ожидания с уровнями значимости $\alpha = 0,05$ и $\alpha = 0,01$.

Варианты заданий

№ 1	4,7	6,7	5,6	6,0	4,5	5,9	4,0	5,3
	5,1	6,1	5,3	3,5	4,2	5,4	4,0	5,3
	3,9	5,5	4,6	7,1	3,1	4,0	4,8	5,5
	3,9	3,4	3,6	4,9	5,0	7,9	5,2	4,7
№ 2	15,6	10,9	11,5	14,5	10,8	13,9	16,8	11,1
	11,9	14,8	8,6	16,2	13,6	10,1	14,8	15,0
	12,0	10,3	8,8	7,3	12,9	14,4	11,6	14,0
	11,1	11,1	11,9	10,1	16,1	14,3	11,3	14,0
№ 3	8,1	6,6	3,8	6,6	7,9	6,5	6,3	8,8
	7,0	9,2	8,6	6,0	7,8	8,9	7,3	4,2
	9,3	7,5	4,4	7,7	3,8	6,7	7,4	6,7
	5,0	4,6	7,4	0,4	7,9	4,0	5,7	1,4
№ 4	7,9	8,0	12,2	15,0	9,2	12,2	9,3	8,6
	13,7	12,0	10,2	13,3	9,0	8,6	10,9	15,3
	11,8	11,4	14,4	9,3	7,7	13,0	10,0	11,5
	12,0	13,6	15,6	11,3	13,7	9,6	8,4	8,7

4.2.4. Методические указания к выполнению индивидуального задания № 2

Основы выборочного метода. Представление результатов наблюдений

Математическая статистика (МС) для своих выводов использует наблюдения за случайным явлением с целью получения заключения об источнике наблюдений. При этом выбирается интересующее исследователя свойство (например, производительность труда, здоровье людей и т.д.) и определяются переменные **x, y, …,** характеризующие это свойство (например, число деталей за смену, показатели сердечно-сосудистой системы и т.д.). Затем случайным образом выбираются объекты из источника наблюдений, на которых измеряются выбранные показатели. Результаты наблюдений обозначаются малыми латинскими буквами **x, y, z, …** в отличие от переменных, характеризующих свойство, которые обозначаются большими латинскими буквами **X, Y, Z, …** .

Таким образом, полный набор всех значений **N**, которое принимает случайная величина **X**, называется **генеральной совокупностью**, где число **N** может быть конечным и бесконечным. Часть генеральной совокупности из **n** элементов, отобранных случайным образом, называется **выборкой**. Число **n** называется **объемом выборки**. Различают выборки большого и малого объема. Выборки с **n ≤ 20** называются **малыми**, выборки **n > 20** называют **большими.** Это является условным и зависит от решаемой задачи. Выборка называется **репрезентативной**, если она дает достаточное представление об особенностях генеральной совокупности.

Статистические методы представляют собой методы для получения выводов о генеральной совокупности через выборку.

Методы, касающиеся сбора данных называются **описательной статистикой**.

Методы, касающиеся заключений о генеральной совокупности называются **статистическими выводами**.

Поскольку конечная цель статистических исследований - получение заключений о генеральной совокупности по выборке с определенной вероятностью, то методы описательной статистики могут рассматриваться как методы первичной обработки результатов наблюдений и представления их в виде удобном для дальнейшего анализа.

Чтобы правильно обработать исходный ряд наблюдений, необходимо выяснить какого типа количественную случайную переменную предстоит исследовать.

В зависимости от типа множества значений, которые может принимать случайная переменная, выделяют два типа: **дискретный** и **непрерывный**.

Случайная переменная **дискретного** типа может принимать лишь изолированные значения. Случайная переменная **непрерывного** типа – любое значение из некоторого интервала.

К **непрерывным** относятся переменные, значения которых получены в результате измерений (иногда они могут быть округлены и представлены в виде целых чисел). Такие переменные как вес, рост, температура, расстояние, время и т.д. являются непрерывными.

Дискретными являются переменные, значения которых определяются в виде некоторых подсчетов.

Первоначально выборки представляют в виде таблицы, состоящей из двух строк, в первой даны номера измерений, во второй – их результаты.

Таблица 1

i – номер измерений	**1**	**2**	**…**	**N**
результат измерений	$\mathbf{x_1}$	$\mathbf{x_2}$	**…**	$\mathbf{x_n}$

Таблица такого вида называется **простым статистическим рядом**. Далее этот ряд преобразуют в **вариационный ряд**, где все наблюдения представляются в порядке возрастания, т.е. в виде:

$$\mathbf{x_{1\,min}, x_2, \ldots, x_{n\,max}},$$

где $\mathbf{x_1 \le x_2 \le \ldots \le x_{nmax}}$.

На следующем этапе данный вариационный ряд представляется в виде статистического ряда. **Статистический ряд** для дискретной переменной это сгруппированный или частотный ряд вида:

Таблица 2

Значение переменной x	**Частота**	**Частость (относит. Частота)**	**Накопленная частота**	**Накопленная частость**
	$\mathbf{m_i}$	$\mathbf{m_i / n}$	$\mathbf{\sum\limits_{x_i<x} m_i}$	$\mathbf{\sum\limits_{x_i<x} m_i / n}$
	$\mathbf{\sum m_i = n}$	$\mathbf{\sum \frac{m_i}{n} = 1}$		

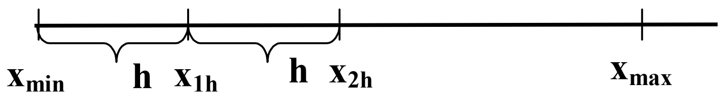
Для **непрерывных** вариационных рядов результаты наблюдений представляются в виде **интервального** ряда.

Таблица 3

Класс границ	Частота m_i	Средняя точка класса x_i	Частости (относит. частота)	Накопленные частоты	Накопленные частости
x_i, x_{i+k}	m_i	$\bar{x}_{i_k}$	m_i / n	$\sum\limits_{x_i<x} m_i$	$\sum\limits_{x_i<x} m_i / n$
	$\sum\limits_i m_i = n$		$\sum\limits_i m_i / n = 1$		

Табл. 1 и 2 дают полный вид статистических рядов (они могут представляться и в усеченном виде (в зависимости от решаемой задачи).

Класс границ для интервального ряда можно изобразить на числовой оси.



Из рисунка видно, что граничные точки классов отстоят друг от друга на одну и ту же величину, равную шагу **h**, который равен $h = (x_{max} - x_{min}) / k$, где **k** - число классов. Чтобы определить величину этого шага, необходимо установить на какое количество классов **k** разбивается данный ряд наблюдений. Это рекомендуется сделать в соответствие со следующими формулами: $k = 1 + 3 \cdot 322 \lg n$ или $k \le 5 \lg n$, где **n** – число наблюдений (объем выборки). В литературе предлагается также выбирать число классов в зависимости от объема выборки. Для малых выборок $k = 5 \div 7$, для больших $k = 10 \div 20$. Выбор числа классов является важным моментом. При слишком малом **k** гистограмма не будет отражать особенностей распределения, при слишком большом **k** гистограмма будет излишне изрезанной. Значения **h** и **k** обычно округляются до ближайшего целого. За начало первого интервала берется точка x_{min} или $x_{min} - h/2$.

Средняя точка класса определяется в виде $\bar{x}_{i_k} = x_{i_h} + h/2$.

Частота m_i представляет собой количество наблюдений, соответствующих данному наблюдению для дискретной переменной для сгруппированного ряда или число наблюдений, попавших в данный интервал для интервального ряда. Значение соответствующей частоты, деленной на объем выборки характеризует **частость** попадания x_i в частичные интервалы.

Закон больших чисел в форме Бернулли утверждает, что если эксперимент повторяется **n** раз при одинаковых условиях, то частость

$\mathbf{m_i / n}$ сходится по вероятности к $\mathbf{p_i}$, то есть $\frac{\mathbf{m_i}}{\mathbf{n}} \xrightarrow{p} \mathbf{p_i}$. Следовательно, $\mathbf{m_i / n}$ является приближенными значениями вероятности $\mathbf{p_i}$. В отличие от теоретического закона распределения СВ **X**, для выборки определяется эмпирический закон распределения.

Сгруппированный и интервальный ряды представляют эмпирическое распределение частот. Для наглядного представления эмпирических распределений **сгруппированного** ряда строится график, где по оси X откладываются значения переменной, а по оси У значения частот (частостей). Полученный точки соединяют ломаной линией. Этот график называется **полигоном**.

Интервальный ряд графически представляется в виде **гистограммы**. Гистограмма представляет собой ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы $[\mathbf{x_i}, \mathbf{x_{i+k}}]$, а их высоты равны $\mathbf{m_i / n}$ или $\mathbf{m_i}$, **h** – длина интервала $[\mathbf{x_i}, \mathbf{x_{i+k}}]$ или шаг. В первом случае площадь гистограммы равна **1**, во втором случае -объему выборки **n**. Результаты в столбцах «накопленные частоты» и «накопленные частости» используются для определения эмпирической функции распределения, с помощью которой также задается частотное распределение.

Эмпирическая функция распределения определяется по значениям **накопленных частостей** из следующих соотношений:

$$\mathbf{F_n^*(x) = 1/n \sum_{x_i < x} m_i},$$

где суммируются частости тех элементов выборки, для которых выполняется неравенство $\mathbf{x_i < x}$ ($\mathbf{x}$ – некоторое значение). Из приведенных формул следует, что

$$\mathbf{F_n^*(x) = 0}, \quad \text{если } \mathbf{x \le x_1},$$

$$\mathbf{F_n^*(x) = 1}, \quad \text{если } \mathbf{x > x_n}.$$

На промежутке $(\mathbf{x_1}, \mathbf{x_n})$ $\mathbf{F_n^*(x)}$ представляет собой неубывающую кусочно-постоянную функцию. Согласно теореме Гливенко эмпирическая функция распределения $\mathbf{F_n^*(x)}$ является хорошей оценкой генеральной функции распределения $\mathbf{F(x)}$ при $\mathbf{n \to \infty}$.

Пример. Получить статистические ряды для следующих наблюдений и представить графически их частотные распределения.

а) производится наблюдение над величиной урожая пшеницы на 10 опытных участках, результаты представлены в табл. 4.

Таблица 4

№ уч	1	2	3	4	5
X	**15.2**	**19.1**	**17.1**	**20.8**	**18.4**
№ уч	**6**	**7**	**8**	**9**	**10**
X	**16.4**	**22.0**	**20.4**	**17.6**	**18.2**

в) имеем выборку 21 пар обуви
39, 41, 40, 41, 44, 40, 42, 41, 43, 39, 42, 41, 42, 38, 42, 41, 43, 41, 39, 40, 40.

Данные выборки представляют собой **непрерывный ряд** наблюдений - случай а) и **дискретный ряд** наблюдений – случай в). Поэтому для случая а) нужно построить статистический ряд в виде интервального ряда, в случае в) в виде сгруппированного ряда.

Решение

Случай **а**

Представляем данный простой статистический ряд в виде вариационного ряда. Имеем:

15.2, 16.4, 17.1, 17.6, 18.2, 18.4, 19.1, 20.4, 20.8, 22.0.

Объем выборки $\mathbf{n = 10}$.

Для представления вариационного ряда в виде интервального ряда необходимо определить ширину интервала **h.**

$$\mathbf{h = \frac{x_{max} - x_{min}}{k}}, \ \mathbf{k} \text{ – число классов.}$$

Число классов **k** определяем по формуле

$$\mathbf{k = 1 + 3.322 \lg n = 1 + 3.322 \lg 10 = 4.322}.$$

В результате округления получаем $\mathbf{k = 4}$.

$$\mathbf{h = \frac{22 - 15.2}{4} = 1.75 \approx 2}.$$

Интервальный ряд представлен табл. 5.

Таблица 5

Границы классов	Частоты m_i	Средние точки классов	Частости m_i / n	Накоплен. частоты	Накоплен. частости
15-17	**2**	**16**	**0.2**	**2**	**0.2**
17-19	**4**	**18**	**0.4**	**6**	**0.6**
19-21	**3**	**20**	**0.3**	**9**	**0.9**
21-23	**1**	**22**	**0.1**	**10**	**1**
	$\sum m_i = 10$		$\sum m_i / n = 1$		

Накопленные частоты определяются по формуле $\sum_{x_i < x} m_i$, то есть суммируются, те значения частот, которые соответствуют соотношению $x_i < x$. При заполнении данного столбца на первое место записывается частота, соответствующая первому интервалу. В данном случае это $m_1 = 2$. Далее суммируются последовательно частоты первого и второго интервалов и т.д. В результате имеем значение **6**, далее **9** и **10**.

Столбец "накопленные частости" определяется делением соответствующих накопленных частот на **10**, то есть на значение, равное общему объему выборки. Графическое представление полученного интервального ряда (гистограмма) представлена на рис. 2.

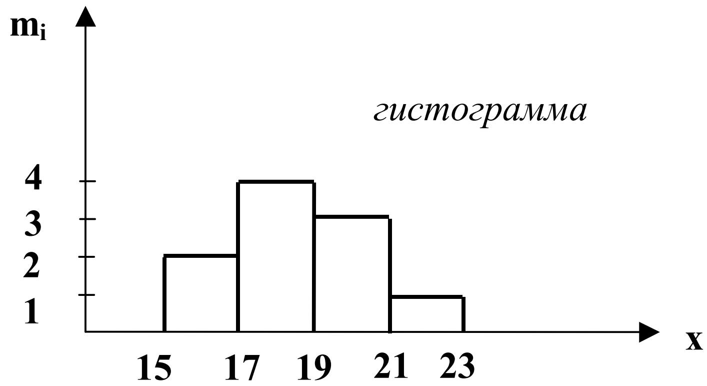


Рис. 2. Графическое представление интервального ряда

На рис. 3 представлено графическое изображение эмпирической функции распределения (кумулята). При построении кумуляты по оси X откладываются начальная точка и верхние точки классов, а по оси У значение накопленных частостей. Полученные точки соединяются кривой.

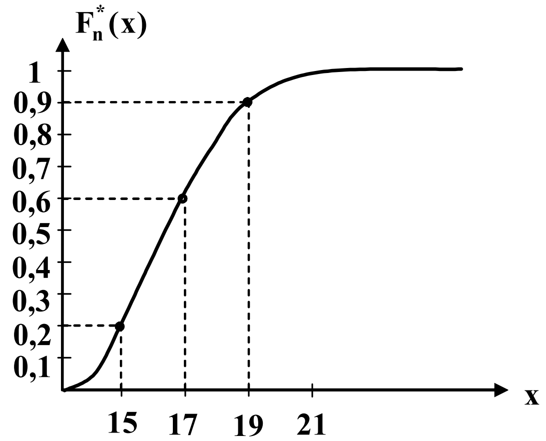


Рис. 3. Графическое представление эмпирической функции распределения $F_n^*(x)$ для случайных переменных непрерывного типа

Гистограмма и кумулята представляют собой графики эмпирических частотных распределений.

Случай **б.** Вариационный ряд имеет вид:

38, 39, 39, 39, 40, 40, 40, 40, 40, 41, 41, 41, 41, 41, 41, 42, 42, 42, 42, 43, 43, 44.

Сгруппированный ряд представлен табл. 6.

Таблица 6

Размеры обуви	**Частоты**	**Частости**	**Накоплен. частоты**	**Накоплен. частости**
38	**1**	**1/21**	**1**	**1/21**
39	**3**	**3/21**	**4**	**4/21**
40	**4**	**4/21**	**8**	**8/21**
41	**6**	**6/21**	**14**	**14/21**
42	**4**	**4/21**	**18**	**18/21**
43	**2**	**2/21**	**20**	**20/21**
44	**1**	**1/21**	**21**	**1**
	$\sum_i m_i = 21$	$\sum_i \frac{m_i}{n} = 1$		

Для графического изображения частотных распределений необходимо построить полигон и кумуляту.

Для построения полигона по оси X откладываются возможные значения переменной, по оси У частоты (частости).

В данном случае полигон имеет вид

Рис. 4. Графическое представление сгруппированного ряда

Эмпирическая функция распределения изображается **кумулятой,** которая представляет собой неубывающую ступенчатую функцию, скачки которой происходят в точках, соответствующих возможным значениям переменной и равны значению частости в этой точке. По оси X откладываются значения переменной, по оси У соответствующие накопленные частости.

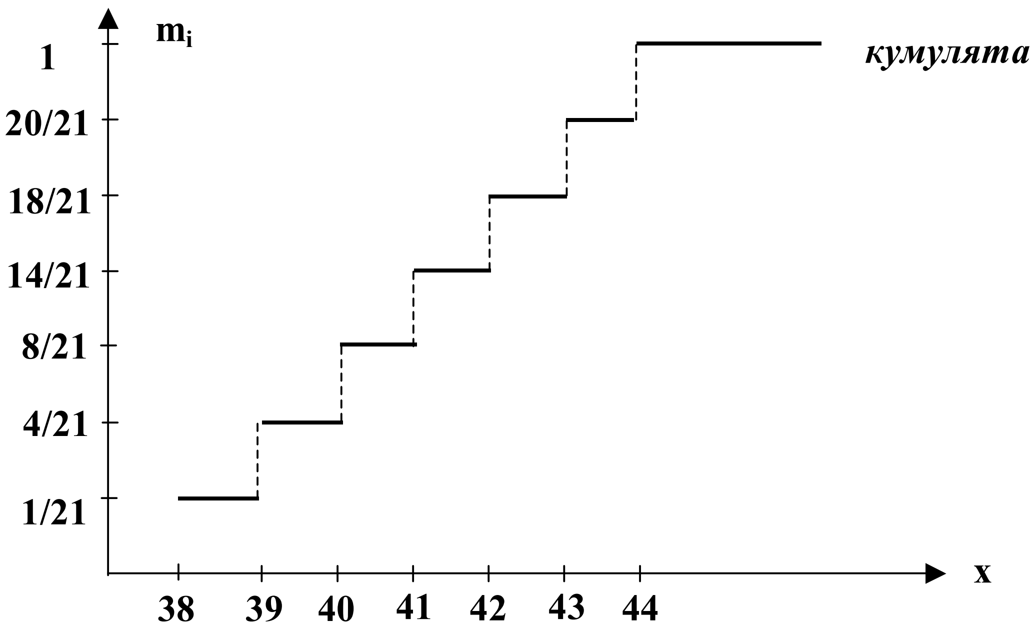


Рис. 5. Графическое представление эмпирической функции распределения $\mathbf{F_n^*(x)}$ для случайной величины дискретного типа

На практике эмпирическая функция распределения используется, когда при принятии решения необходимо знать текущий итог.

Числовые характеристики выборочных распределений

Эмпирические распределения, представляемые для дискретных вариационных рядов в виде полигона и кумуляты, а для непрерывных – в виде гистограммы и кумуляты, позволяют предположить вид генерального распределения.. Кроме этого на практике часто используются числовые характеристики для описания ряда наблюдений. Данные характеристики позволяют проанализировать ряд наблюдений другим способом. Они включают меры расположения, меры рассеивания и меры, характеризующие форму распределения.

Вычисление этих характеристик зависит от того как представлены результаты наблюдений: в виде статистических рядов или в виде простого статистического ряда (исходной выборки). Как было уже сказано, статистический ряд является такой формой представления, когда данные группируются и определяются частоты. В дальнейшем будем называть статистический ряд сгруппированным рядом. Простой статистический ряд не является сгруппированным.

Меры расположения

Меры расположения дают информацию о среднем значении, наиболее часто встречающемся значении и о значении, являющимся серединой данных измерений и т.д. Эти меры не дают информации о том, как разбросаны наблюдения относительно среднего числа. К ним относятся **мода, медиана, среднее арифметическое, процентили и квартили.**

Меры расположения в несгруппированном ряду наблюдений

Мода – наиболее часто встречающееся значение в ряду наблюдений. Для определения моды ряд наблюдений необходимо представить в виде вариационного ряда и определить наиболее часто встречающееся наблюдение. Ряд наблюдений может иметь два и более таких значений. В первом случае ряд называется **бимодальным**, во втором - **мультимодальным**.

Медиана – величина, соответствующая середине в вариационном ряду (то есть число наблюдений меньше медианы равно числу наблюдений больше медианы). Чтобы определить медиану необходимо выполнить следующее:

- представить исходный ряд наблюдений в виде вариационного;
- если объем выборки число нечетное, то медиана это серединный элемент вариационного ряда, номер которого определяется в виде $(\mathbf{n}+\mathbf{1})/\mathbf{2}$;
- если объем выборки число четное, то медиана равна среднему значению двух серединных элементов вариационного ряда.

Среднее арифметическое значение, обозначаемое $\overline{\mathbf{X}}$ (вместо $\mathbf{X}$ может быть любая буква латинского алфавита) определяется в виде:

$$\overline{\mathbf{X}} = \frac{\mathbf{1}}{\mathbf{n}}(\mathbf{x}_1 + \mathbf{x}_2 + \ldots + \mathbf{x}_n) = \frac{\mathbf{1}}{\mathbf{n}}\sum_{i=1}^{n}\mathbf{x}_i .$$

Мода и медиана называются **структурными средними**, так как они связаны с рядом частот, то есть со структурой распределения. Следует отметить, что на значение среднего арифметического оказывают влияние все значения выборки. Медиана не подвержена такому влиянию.

Пример 1. Для данного вариационного ряда, представляющего собой объем акций 10 наиболее действующих предприятий (в тысячах) определить моду, медиану и среднее арифметическое

651 681 789 827 903 903 904 922 979 1514.

Решение.

Мода: наиболее часто встречающееся значение **903** миллиона.

Медиана: так как объем выборки n = 10, то значение $\mathbf{\frac{903+903}{2}=903}$ является медианой.

Среднее арифметическое:

$$\bar{X}=\frac{1}{10}(651+681+789+827+903+903+904+922+979+1514)=907,3.$$

Процентили являются мерами расположения, которые делят наблюдения на 100 частей. Можно определить **99** процентилей. **Р-ый** процентиль – это такая величина, что **Р %** данных являются меньше этой величины и **(100-Р) %** являются больше.

Процентили широко используются в различного рода отчетах. Для того чтобы определить **Р-ый** процентиль, необходимо выполнить следующее:

- представить результаты наблюдений в виде вариационного ряда;
- вычислить номер Р-го процентиля в вариационном ряду

$$\mathbf{i=\frac{P}{100}(n)},$$

где **Р** – значения процентиля,

n – объем выборки,

i – номер процентиля в ряду наблюдений.

- определить значение Р-го процентиля:

а) если **i** – целое, то **Р-ый** процентиль является средней величиной **i – го** и **(i+1) – го** наблюдений в вариационном ряду.

в) если **i** не является целым числом, то номер **Р-го** процентиля определяется как целая часть от значения **(i+1)**.

Пример 2. Определить **30-ый** процентиль для следующего ряда наблюдений:

14 12 19 23 5 13 28 17.

Решение. Вариационный ряд:

5 12 13 14 17 19 23 28,

$$\mathbf{i=\frac{30}{100}\cdot 8=2.4}.$$

Так как **i** не является целым числом, то номер **30-го** процентиля в данном вариационном ряду определяется как целая часть от значения **2.4+1=3.4**, то есть **3.** Следовательно, **30-ым** процентилем является значение $\mathbf{x_3 = 13}$.

Квартили это меры расположения, которые делят ряд наблюдений на **4** части. Для этого необходимо **3** квартиля, которые обозначаются $\mathbf{Q_1}$**,** $\mathbf{Q_2}$**,** $\mathbf{Q_3}$.

$\mathbf{Q_1}$ является **25-ым** процентилем, то есть $\mathbf{Q_1 = P_{25}}$.

$\mathbf{Q_2}$ является **50-ым** процентилем, то есть $\mathbf{Q_2 = P_{50}}$.

$\mathbf{Q_3}$ - это **75-ый** процентиль, то есть $\mathbf{Q_3 = P_{75}}$.

Пример 3. Определить $\mathbf{Q_1, Q_2, Q_3}$ для следующей выборки:

109 121 122 129 106 116 125 114.

Решение. Вариационный ряд:

106 109 114 116 121 122 125 129.

Так как $\mathbf{Q_1 = P_{25}}$, то определяем 25-ый процентиль. Для $\mathbf{n = 8}$

$$\mathbf{i = \frac{25}{100} \cdot 8 = 2}.$$

Так как $\mathbf{i}$ – целое число, то $\mathbf{P_{25}}$ определяется как среднее второго и третьего значений в вариационном ряду.

$$\mathbf{P_{25} = \frac{109 + 114}{2} = 111.5}.$$

$\mathbf{Q_1 = 111.5}$, $\mathbf{Q_2 = P_{50}}$ и является медианой. Так как $\mathbf{n}$ – четное, то

$$\mathbf{Q_2 = \frac{116 + 121}{2} = 118.5},\ \mathbf{Q_3 = P_{75}},$$

$$\mathbf{i = \frac{75}{100} \cdot 8 = 6},\ \mathbf{P_{75} = \frac{122 + 125}{2} = 123.5},\ \mathbf{Q_3 = 123.5}.$$

Меры рассеивания

Меры расположения дают информацию об особых точках в ряду наблюдений.

Меры рассеивания характеризуют разброс или дисперсию в выборке.

Меры расположения вместе с мерами рассеивания дают более полное числовое описание наблюдений.

Одними из основных показателей, характеризующих рассеивание в выборке, являются **выборочная дисперсия** и **среднеквадратическое отклонение.** Для несгруппированных данных выборочная дисперсия определяется в виде:

$$\mathbf{S^2 = \frac{1}{n - 1}\sum_{i=1}^{n}(x_i - \overline{X})^2},$$

где $\mathbf{n}$ – объем выборки,

$\mathbf{\overline{X}}$ - среднее арифметическое,

$\mathbf{x_i}$ – значение **i-го** наблюдения.

Среднеквадратическое отклонение

$$\mathbf{S} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2} = \sqrt{S^2}\,.$$

Следует отметить, что среднее арифметическое $\overline{\mathbf{X}}$, дисперсия $\mathbf{S^2}$ и среднее квадратическое отклонение **S**, определяемые по выборке, являются оценками соответственно для матожидания **M(x),** дисперсии **D(x)** и среднеквадратического отклонения **σ**, являющихся параметрами генеральной совокупности. При этом оценки должны удовлетворять определенным свойствам (эффективности, состоятельности и несмещенности см. [1, 2]), поэтому в формуле для выборочной дисперсии (для выполнения условия несмещенности) вводится сомножитель

$$\frac{1}{n-1}.$$

Вычислительные формулы

В практических задачах кроме представленных выше могут использоваться следующие вычислительные формулы для выборочной дисперсии и стандартного отклонения.

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right), \qquad S = \sqrt{S^2}\,.$$

Если среднее арифметическое задано, то

$$S^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n(\overline{X})^2\right).$$

Смысл среднеквадратического отклонения (стандартного отклонения) может быть установлен по его применению. Два наиболее используемых способа этого применения: это **эмпирическое правило** и **теорема Чебышева**.

Эмпирическое правило справедливо для наблюдений с нормальным законом распределения и утверждает, что

68% данных лежит внутри интервала $\overline{X} \pm 1S$.

95% данных лежит внутри интервала $\overline{X} \pm 2S$.

99.7% данных лежит внутри интервала $\overline{X} \pm 3S$.

Теорема Чебышева применяется для всех распределений. Она утверждает, что внутри **k** стандартных отклонений от среднего находится

$(1-1/k^2)$ значений выборки.

Пример. Определить внутри какого интервала лежит по крайней мере 85 % всех величин, если известно $\overline{X}=16.85;\quad S=4.7$.

Решение

$$0.85=1-1/k^2;\qquad 0.15=1/k^2;\quad k^2=6.667;\quad k=2.58.$$

Полученный интервал определяется как $\overline{X}\pm 2.58\cdot 4.7$, то есть границами являются **4.72** и **28.98**.

Формулы для среднего арифметического дисперсии и среднеквадратического отклонения

Для **сгруппированных данных** указанные числовые характеристики определяются в виде:

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{k}x_i m_i,\qquad S^2=\frac{1}{n-1}\sum_{i=1}^{k}(x_i-\overline{X})^2 m_i,\quad S=\sqrt{S^2},$$

где x_i – средние точки классов; k – число классов.

Вычислительные формулы для сгруппированных данных

$$S^2=\frac{1}{n-1}\left(\sum_{i=1}^{k}m_i x_i^2-\frac{\left(\sum m_i x_i\right)^2}{n}\right),\qquad S=\sqrt{S^2},$$

где n – объем выборки,
k – число классов,
m_i – частоты,
x_i – средние точки в классе.

Структурные средние для сгруппированных данных

Для определения моды M_0 и медианы M_e рассмотрим фрагмент гистограммы, соответствующий интервалу с наибольшей частотой и двумя соседними интервалами.

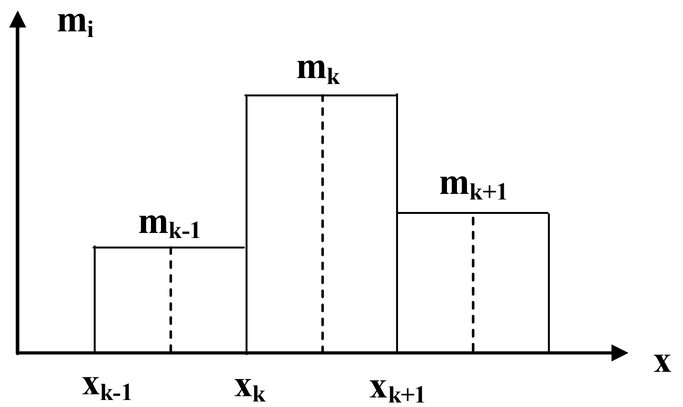


Рис. 6. Фрагмент гистограммы для определения моды M_0

$$M_0 = x_k + \frac{m_k - m_{k-1}}{2m_k - (m_{k-1} + m_{k+1})} \cdot \Delta x,$$

где $\Delta x = x_{k+1} - x_k = h$.

Медиана определяется по формуле

$$M_e = x_{l+1} + \frac{n/2 - F_l(x)}{m_{l+1}} \Delta x,$$

где $F_l(x) = \sum\limits_{x_i < x} m_i$ выбирается из условия $F_l(x) \le n/2$ и $F_{l+1}(x) > n/2$,

где n – объем выборки, $\Delta x = h$,

m_{l+1} - частота интервала, x_{l+1} - нижняя граница интервала, для которого выполняется $F_l(x) \le n/2$ и $F_{l+1}(x) > n/2$ (см. интервальную таблицу).

Пример. Вычислить среднее арифметическое, медиану, моду и среднеквадратическое (стандартное) отклонение для следующего интервального ряда, представленного табл. 7.

Таблица 7

Класс границ	Частоты	Средние точки класса	Накопленные частоты
10-15	6	12.5	6
15-20	22	17.5	28
20-25	35	22.5	63
25-30	29	27.5	92
30-35	16	32.5	108
35-40	8	37.5	116
40-45	4	42.5	120
45-50	2	47.5	122

$$n = \sum m_i = 122.$$

Решение.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i = \frac{1}{122}\sum_{i=1}^{8} x_i m_i = \frac{3130}{122} = 25.66.$$

Медиана

$$M_e = x_{l+1} + \frac{n/2 - F_l^*(x)}{m_{l+1}} \Delta x = 20 + \frac{122/2 - 28}{35} \cdot 5 =$$

$$= 20 + 0.943 \cdot 5 = 20 + 4.714 = 24.714.$$

Мода

$$M_0 = 20 + \frac{(35 - 22) \cdot 5}{70 - (22 + 29)} = 20 + \frac{13 \cdot 5}{70 - 51} = 23.42.$$

Дисперсия и стандартное отклонение могут быть определены как будет показано далее. Вначале для определения дисперсии используем исходную формулу $\mathbf{S^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{X})^2 m_i}$. Вычисления представлены в табл. 8.

Таблица 8

Класс границ	Частоты	Средняя точка класса x_i	$x_i - \overline{X}$	$(x_i - \overline{X})^2$	$m_i(x_i - \overline{X})^2$
10-15	6	12.5	-13.16	173.19	1039.14
15-20	22	17.5	-8.16	66.59	1464.98
20-25	35	22.5	-3.16	9.99	349.65
25-30	29	27.5	1.84	3.39	98.31
30-35	16	32.5	6.84	46.79	748.64
35-40	8	37.5	11.84	140.19	1121.52
40-45	4	42.5	16.84	283.59	1134.36
45-50	2	47.5	21.84	476.99	953.98

$$\sum m_i = 122$$

$$\sum m_i(x_i - \overline{X})^2 = 6910.58$$

$$S^2 = \frac{\sum (x_i - \overline{X})^2 m_i}{n-1} = \frac{6910.58}{121} = 57.11,$$

$$S = \sqrt{57.11} = 7.56.$$

Определим дисперсию, используя вычислительную формулу. Вычисления представлены в табл. 9.

Таблица 9

Класс границ	Частоты	Средняя точка класса x_i	$m_i x_i$	mx_i^2
10-15	6	12.5	75.0	937.50
15-20	22	17.5	385.0	6737.50
20-25	35	22.5	787.0	17718.75
25-30	29	27.5	797.0	21931.25
30-35	16	32.5	520.0	16900.00
35-40	8	37.5	300.0	11250.00
40-45	4	42.5	170.0	7225.00
45-50	2	47.5	95.0	4512.50

$$\sum m_i = 122$$

$$\sum m_i x_i = 3130.0$$

$$\sum m_i x_i^2 = 87212.50.$$

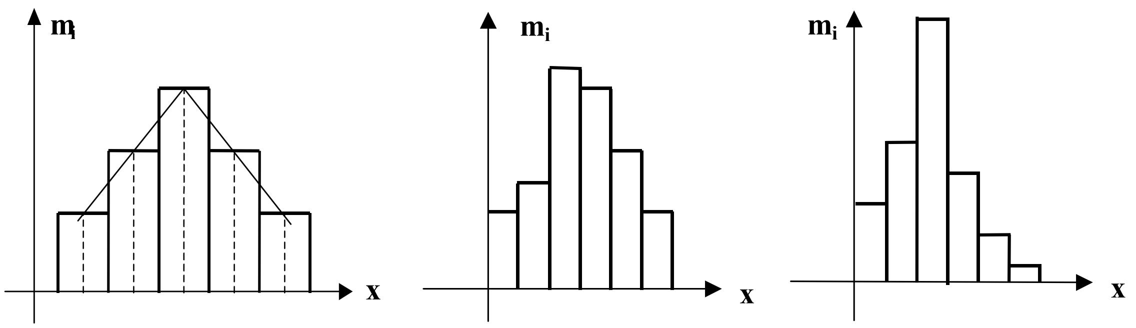
$$S^2 = \frac{\sum m_i x_i^2 - \frac{\left(\sum m_i x_i\right)^2}{n}}{n-1} = \frac{87212.50 - (3130.0)^2 / 122}{121} = \frac{6910.04}{121} = 57.11$$

$S = \sqrt{57.11} = 7.56$.

Меры формы

К мерам формы относятся числовые характеристики, позволяющие сделать вывод о форме частотных распределений выборочных данных.

Рассмотрим числовые характеристики, которые позволяют оценить симметричность распределения. Если относительно наибольшей частоты остальные частоты для соответствующих интервалов приблизительно равны, то такое распределение частот будет изображено гистограммой или полигоном, имеющими **симметричную** форму (случай а) на рис. 6. Случаи в) и с) приводятся соответственно левосторонняя и правосторонняя эмпирические асимметричные распределения.



а) Симметричное распределение частот

в) Левая асимметрия

с) Правая асимметрия

Рис. 7

Коэффициент асимметрии

В данном разделе приводится одна из формул для определения коэффициента асимметрии, который используется для определения симметричности распределения.

$$A_x = \frac{3(\overline{X} - M_e)}{S},$$

где A_x – коэффициент асимметрии,
S – среднеквадратическое отклонение,
M_e – медиана,
$\overline{X}$ - среднее арифметическое.

Если значение этого коэффициента имеет положительное значение, то распределение имеет правостороннюю (положительную) асимметрию. Если значение $\mathbf{A_x}$ отрицательно, то распределение имеет левостороннюю (отрицательную асимметрию. Если $\mathbf{A_x} \approx 0$, то распределение симметрично. Для симметричных распределений выполняется условие: $\overline{\mathbf{X}} \approx \mathbf{M_0} \approx \mathbf{M_e}$.

Методы статистических выводов

Рассмотрим один из методов статистических выводов, часто используемый на практике, метод оценки параметров генеральной совокупности. Другие методы статистических выводов такие как методы проверки гипотез и оценки взаимосвязи показателей можно найти в [1,2].

Методы оценки параметров генеральной совокупности делятся на точечные и интервальные.

Точечными оценками параметров генеральной совокупности являются выборочные оценки этих параметров, которые рассматривались нами ранее. Точечные оценки параметров распределения, полученных при выборке, не дают информацию о степени близости их значений теоретическому значению оцениваемого параметра. В этом случае нельзя ответить на вопрос, какую ошибку мы совершаем, принимая вместо теоретического значения параметра его оценку. В связи с этим во многих случаях следует определять не точечную оценку теоретического параметра, а **интервальную оценку**, которая основана на определении некоторого интервала, внутри которого с определенной вероятностью находится истинное значение параметра. то есть следует определить интервал, который с вероятностью **Р** охватывает оцениваемое значение параметра.

Одним из наиболее важных для практики генеральных параметров является математическое ожидание или генеральное среднее. Например, управляющий крупной компанией хочет установить, какое количество дней в среднем сотрудники его фирмы пропустили по болезни. Если в фирме тысячи работников, то напрямую получить это число сложно. Используя статистические методы можно определить по выборке среднее арифметическое пропущенных дней и взять эту среднюю арифметическую в качестве оценки генерального среднего и сделать вывод о всей генеральной совокупности по полученному среднему значению. Среднее арифметическое будет являться точечной оценкой генерального среднего, но как точно оценивает оно это генеральное среднее, мы не знаем.

Интервальные оценки дают ответ на этот вопрос.

Если среднее значение, найденное по результатам выборки, является точечной оценкой математического ожидания **M(x),** то вероятность того, что действительное значение оцениваемой величины **M(x)** лежит в пределах $(\overline{X} - \Delta < M(x) < \overline{X} + \Delta)$ определяется в виде:

$$P(\overline{X} - \Delta < M(x) < \overline{X} + \Delta) = 1 - \alpha,$$

где $P = 1 - \alpha$ - доверительная вероятность,

α - уровень значимости,

Δ - точность оценки.

Выбор доверительной вероятности определяется конкретными условиями задачи. Обычно используются значения **P** = **0.9, 0.95, 0.99.**

Для данных значений доверительной вероятности уровень значимости соответственно равен **0.1, 0.05, 0.01.** Понимать доверительную вероятность следует следующим образом. Если из генеральной совокупности делать выборки, а по ним определять оценки истинного параметра, то в α **%** случаев эти оценки не будут принадлежать указанному интервалу, который называется доверительным.

На практике важную роль играет длина доверительного интервала. Следует отметить, что чем меньше длина доверительного интервала, тем точнее оценка. Если длина доверительного интервала велика, то оценка мало пригодна для практики. Длина доверительного интервала связана с объемом выборки и доверительной вероятностью, что видно из формул, приводимых ниже.

Интервальное оценивание математического ожидания случайной величины

1. Исходим из следующих предположений:

- генеральная выборка распределена по нормальному закону;
- выборка бесповторная;
- значение генеральной дисперсии **D(x)** известно.

В этом случае

$$P\left(\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < M(x) < \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha,$$

где $\overline{X}$ – выборочное среднее,

$Z_{\alpha/2}$ – квантиль нормального распределения или такое значение (см. таблицу функции распределения **Ф(x)** стандартизированного нормального распределения **N(0,1)**), при котором

$\Phi(Z_{\alpha/2}) = 1 - \dfrac{\alpha}{2}$, $\varepsilon = \dfrac{\sigma}{\sqrt{n}}$ – средняя квадратическая ошибка,

$\Delta = Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ – точность оценивания.

Для наиболее употребительных значений доверительной вероятности **1-α = 0.9, 0.95, 0.99, 0.9973** квантили стандартизованного нормального распределения $\mathbf{Z_\alpha}$ соответственно равны:

$\mathbf{Z_{\alpha/2}}$ **= 1.64, 1.96, 2.58, 3, 3.37.**

2. Исходим из следующих предположений:

- генеральная совокупность распределена по нормальному закону;
- выборка бесповоротная;
- значение D(x) неизвестно, а вместо него используется ее оценка, определяемая по выборке в виде:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2}\ .$$

В этом случае

$$P\left(\overline{X} - t_{\alpha,k}\frac{S}{\sqrt{n}} < M(x) < \overline{X} + t_{\alpha,k}\frac{S}{\sqrt{n}}\right) = 1 - 2\alpha,$$

где $t_{\alpha,k}$ – значение квантиля статистики t, распределенной по закону Стьюдента;

k – число степеней свободы или число элементов выборки минус число связей, которые наложены на элементы выборки **k = n – 1**;

α – уровень значимости.

Пример 1. Необходимо установить в течение скольких лет компании данной страны ведут торговлю с Индией. Случайно выбранные **44** фирмы показали, что ведут торговлю в среднем **10.455** лет. Предположим, что известно генеральное среднеквадратическое отклонение и оно равно 7.7 лет. Построить **90 %** доверительный интервал для генерального среднего.

Решение. Имеем **n = 44**; $\overline{X} = 10.55$; **σ = 7.7**. Так как известно генеральное среднеквадратическое отклонение **σ**, то доверительный интервал определяется по формуле

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le M(x) \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

$P = 1 - \alpha = 0.9;\quad \alpha/2 = 0.05$. По таблице функции распределения стандартизованного нормального распределения определяем $Z_{\alpha/2} = -1.645$. Доверительный интервал вычисляется как

$$10.455 - 1.645\cdot\frac{7.7}{\sqrt{44}} \le M(x) \le 10.455 + 1.645\frac{7.7}{\sqrt{44}},$$

$$10.455 - 1.91 \le M(x) \le 10.455 + 1.91,\ 8.545 \le M(x) \le 12.365,$$
$$P(8.545 \le M(x) \le 12.365) = 0.9.$$

Точечная оценка генерального среднего составляет **10.455** лет. Интервальная оценка указывает, что генеральное среднее находится между **8.545** и **12.365**. Точность оценивания **1.91.**

Пример 2. На основании обработки данных выборки из 145 значений получено $\overline{X} = 31.40$; $S = 3.26$. Определить доверительный интервал для генеральной средней с $P = 0.99$.

Решение. Доверительный интервал вычисляется по формуле

$$\overline{X} - t_{\alpha,k}\frac{S}{\sqrt{n}} \le M(x) \le \overline{X} + t_{\alpha,k}\frac{S}{\sqrt{n}}.$$

Определяем $t_{\alpha,k} = t_{0.01,144}$ по таблице распределения Стьюдента (см. таблицу в приложении) $t_{0.01,144} = 2.58$. Границы доверительного интервала

$$31.40 - 2.58\frac{3.26}{\sqrt{145}} \le M(x) \le 31.40 + 2.58\frac{3.26}{\sqrt{145}},$$
$$31.40 - 0.70 \le M(x) \le 31.40 + 0.70,\ \ 30.70 \le M(x) \le 32.10.$$

Генеральное среднее будет находиться в интервале **(30.70, 32.10).**

5. ИТОГОВЫЙ КОНТРОЛЬ

5.1. Требования для сдачи зачёта

К зачёту допускаются только те студенты, у которых сданы и зачтены все индивидуальные задания.

Студенты, обучающиеся по КЗФ, сдают зачёт во время зимней экзаменационной сессии по билетам (в устной или письменной форме). Каждый билет содержит несколько типовых задач (по усмотрению преподавателя от 5 до 10). Зачёт считается сданным, если выполнено не менее 60% заданий билета.

Студенты, обучающиеся с использованием ДОТ, сдают зачёт в тестовой форме (on-line режим). Зачётный тест содержит 20 тестовых заданий различного уровня сложности. Структура теста представлена в таблице:

Тип задания	**Количество в тесте**	**Уровень сложности**
Задание на выбор единственного ответа	8	1
Задание на выбор мно-	4	2

жественных ответов		
Задание на установление последовательности	4	2
Задание на установление соответствия	2	3
Задание для краткого ответа	2	3

Зачёт выставляется по сумме набранных баллов за задания теста и по результатам сдачи индивидуальных домашних заданий.

5.2. Образцы билетов к экзамену для КЗФ

Билет № 0

1. Найдите функцию распределения случайной величины ξ, заданной функцией плотности

$$f(x)=\begin{cases}\dfrac{2}{\pi}\cos^2 x, & |x|\le\dfrac{\pi}{2};\\[2mm] 0, & |x|>\dfrac{\pi}{2}.\end{cases}$$

2. Для системы случайных величин (ξ,η) $D[\xi]=5$, $D[\eta]=6$, $\operatorname{cov}(\xi,\eta)=-3$. Найдите $D[2\xi-\eta+1]$.

3. В железнодорожном составе 25 вагонов, груженных углем двух сортов: 15 вагонов содержит 60% угля первого сорта и 40% угля второго сорта; 10 вагонов – 85% первого сорта и 15% второго сорта. Случайно взятый для анализа кусок угля оказался второго сорта. Найдите вероятность того, что он взят из вагона первой группы

4. Контрольная работа состоит из трёх вопросов. На каждый вопрос приведено 4 ответа, один из которых правильный. Составьте закон распределения случайной величины ξ – числа правильных ответов при простом угадывании. Найдите математическое ожидание и дисперсию этой случайной величины.

5. Даны две независимые случайные величины: ξ распределена по нормальному закону $f(x)=\dfrac{1}{5\sqrt{2\pi}}e^{-\frac{(x-1)^2}{50}}$, η распределена равномерно на отрезке $[0, 1]$. Найдите математическое ожидание и дисперсию ξη и ξ + η.

6. В результате $n=14$ опытов была получена несмещённая оценка для дисперсии нормальной случайной величины $\tilde{s}^2=45$. Найдите доверительный интервал для дисперсии при доверительной вероятности $\gamma=0,98$.

5.4. Образцы билетов к экзамену для ДОТ

Билет № <u>0</u>

Задания на выбор единственного ответа

Задание 1. Достоверным событием называется событие, …
1. которое происходит всегда;
2. которое уже произошло;
3. которое в результате данного эксперимента обязательно произойдёт;
4. вероятность которого равна 1.

Ответ:

Задание 2. Пусть A – некоторое событие. Тогда вероятность противоположного ему события $\overline{A}$ равна…

1. $-P(A)$
2. $P(A)$
3. $\dfrac{1}{P(A)}$
4. $1-P(A)$

Ответ:

Задание 3. Математическое ожидание случайной величины, распределённой по показательному закону с плотностью $f(x)=2e^{-2x}$, $x\in[0,\infty)$, равно…

1. 0,5
2. 0,25
3. 2
4. 4
5. –2

Ответ:

Задание 4. При транспортировке ящика с 10 шарами (3 белых и 7 чёрных) один шар потерялся. После этого из ящика вынимается один шар. Тогда вероятность появления белого шара равна…

1.0,3
2. 0,7
3. 0,27
4. 0,21
5. 0,09

Ответ:

Задание 5. Дискретная случайная величина задана рядом распределения

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

Математическое ожидание этой случайной величины равно…
1. 2,5
2. 2,6
3. 3
4. 2,3

Ответ:

Задание 6. Непрерывная случайная величина подчиняется нормальному закону с параметрами $a=2$, $\sigma=3$. Тогда с вероятностью 0,9973 все возможные значения этой случайной величины принадлежат интервалу
1. $(-3;3)$
2. $(-7;11)$
3. $(-\infty;\infty)$
4. $(-9;9)$

Ответ:

Задание 7. Теорема сложения вероятностей имеет вид $P(A+B)=P(A)+P(B)$, если…
1. A и B несовместны
2. A и B совместны
3. A и B независимы
4. A и B зависимы

Ответ:

Задание 8. Случайная величина распределена равномерно на отрезке $[2,5]$. Тогда функция плотности (дифференциальная функция распределения) имеет вид…

1. $y=\begin{cases} 0, & x\le 0, \\ \dfrac{1}{3}, & 2<x\le 5, \\ 0, & x>5 \end{cases}$

2. $y=\begin{cases} 0, & x\le 0, \\ 3, & 2<x\le 5, \\ 0, & x>5 \end{cases}$

3. $y = \begin{cases} 0, & x \le 0, \\ \dfrac{5-x}{3}, & 2 < x \le 5, \\ 1, & x > 5 \end{cases}$

4. $y = \begin{cases} 0, & x \le 0, \\ \dfrac{1}{3}, & 2 < x \le 5, \\ 1, & x > 5 \end{cases}$

Ответ:

Задания на выбор множественных ответов

Задание 9. Выберите верные утверждения, если $F(x)$ – функция распределения случайной величины ξ

1. $F(x)$ – неубывающая функция;
2. $F(x)$ – возрастающая функция;
3. областью определения $F(x)$ является $[0;1]$;
4. областью определения $F(x)$ является $(-\infty;+\infty)$;

Ответ:

Задание 10. Дискретная случайная величина ξ задана функцией распределения $F(x) = \begin{cases} 0 & x \le 1, \\ 0{,}35 & 1 < x \le 2, \\ 0{,}85 & 2 < x \le 3, \\ 1 & x > 3 \end{cases}$

Тогда числовыми характеристиками данной случайной величины являются…

1. 1,8
2. 6,94
3. 1
4. 3,7

Ответ:

Задание 11. На трёх карточках написаны цифры 1, 2, 3. Две из них, одна за другой, вынимаются. Элементарным исходом этого опыта **не** является

1. 12
2. 21
3. 22
4. 11

Ответ:

Задание 12. Математическое ожидание равно 2 у случайной величины

1. $F(x) = 1 - e^{-2x}$
2. $P(k) = \dfrac{2^k}{k!} e^{-2}$
3. $f(x) = \dfrac{1}{2\sqrt{2\sigma}} e^{-\frac{(x-1)^2}{8}}$
4. $P_6(m) = C_6^m \left(\dfrac{1}{3}\right)^m \left(\dfrac{2}{3}\right)^{6-m}$

Ответ:

Задания на установление последовательности

Задание 13. Из урны, содержащей 4 белых и 6 чёрных шаров, случайным образом и без возвращения извлекается 3 шара. Случайная величина ξ – число белых шаров в выборке. Расположите значения 0, 1, 2, 3 случайной величины ξ в порядке возрастания их вероятностей

1. 0
2. 1
3. 2
4. 3

Ответ:

Задание 14. Задан закон распределения дискретной случайной величины

x_i	–1	1	2	3
p_i	0,2	0,4	0,3	0,1

Расположите числовые характеристики в следующей последовательности:
Мода; математическое ожидание; дисперсия; начальный момент второго порядка

1. 1,49
2. 1

3. 1,1
4. 2,7
5. 3
6. 2,3

Ответ:

Задание 15. Расположите случайные величины в следующем порядке равномерное распределение; показательное распределение; биномиальное распределение; распределение Пуассона

1. $f(x) = 2,5e^{-2,5x}, \ x \geq 0$
2. $P(m) = \dfrac{3^m}{m!}e^{-3}$
3. $f(x) = \begin{cases} 0,5 & x \in [2;4] \\ 0 & x \notin [2;4] \end{cases}$
4. $P_{10}(m) = C_{10}^m \left(0,63\right)^m \left(0,37\right)^{10-m}$

Ответ:

Задание 16. Вероятность того, что студент сдаст экзамен по истории, равна 0,9; по экономике – 0,9; по математике – 0,8. расположите в порядке возрастания вероятностей следующие события:

1. студент сдаст только один экзамен;
2. студент сдаст хотя бы один экзамен.
3. студент сдаст три экзамена;
4. студент сдаст по крайней мере два экзамена;

Ответ:

Задания на установление соответствия

Задание 17. Вероятности попадания в цель для 1-го, 2-го, 3-го стрелков соответственно равны $p_1 = 0,8$; $p_2 = 0,7$; $p_3 = 0,9$. Каждый из стрелков делает по одному выстрелу. Установите соответствие:

СОБЫТИЕ	ВЕРОЯТНОСТЬ
1. никто из стрелков не попал в цель	1. 0,504
2. все стрелки попали в цель	2. 0,398
3.два стрелка попали в цель	3. 0,092
4. один стрелок попал в цель	4. 0,006

Ответ:

Задание 18. Из урны, содержащей 4 белых и 6 чёрных шаров, случайным образом и без возвращения извлекается 3 шара. Случайная величина ξ – число чёрных шаров в выборке. Установите соответствие между значениями случайной величины ξ и их вероятностями.

Значения:	**Вероятности:**
1. 0	1. 3/10
2. 1	2. 1/6
3. 2	3. 1/30
4. 3	4. 1/2
	5. 1/10

Ответ:

Задания для краткого ответа

Задание 19. Радист вызывает корреспондента. Вероятность того, что вызов будет принят, равна 0.6. Найдите вероятность того, что корреспондент ответит лишь на четвёртый вызов.

Ответ:

Задание 20. Непрерывная случайная величина ξ задана интегральной функцией распределения

$$F(x)=\begin{cases} 0, & x\le 2 \\ (x-2)^2, & 2<x\le 3 \\ 1, & x>3 \end{cases}$$

Найдите $P(2{,}5<\xi<3{,}5)$.

Ответ:

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Литература обязательная

1. Константинова Л.И. Теория вероятностей и математическая статистика. Учебное пособие для студентов ИДО. – Томск, 2005. – 140 с.
2. Вентцель Е.С. Теория вероятностей. – М.: «Высшая школа», 1998. – 575 с.
3. Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика. – М.: Инфра-М., 1997. – 300 с.

6.2. Литература дополнительная

1. Бекишев Г.А., Митрофанов Е.Н., Семенов А.Т., Соболев В.Ф. Теория вероятностей и математическая статистика. Учебное пособие для студентов заочной формы обучения. – Новосибирск, 1997. – 173 с.
2. Вентцель Е.С., Овчаров Л.А. Теория вероятностей. Сб. задач. – М.: Высшая школа, 1965. – 363 с.
3. Сборник задач по математике для ВТУЗов. Теория вероятностей и математическая статистика. – М.: Наука, 1990. – 431 с.

6.3. Учебно-методические пособия

4. Константинова Л.И. Теория вероятностей и математическая статистика. Часть I. Рабочая тетрадь. – Томск, 2000. – 48 с.
5. Константинова Л.И. Теория вероятностей и математическая статистика. Часть II. Рабочая тетрадь. – Томск, 2000. – 44 с.

6.4. Internet-ресурсы

1. Сайт ТПУ.– Режим доступа: http://www.tpu.ru, вход свободный.

ПРИЛОЖЕНИЕ

Таблица I

Функция Лапласа $\Phi(u) = \frac{1}{\sqrt{2\pi}}\int_0^u e^{-t^2/2}dt$

u	0	l	2	3	4	5	6	7	8	9
0.0	0.0000	0040	0080	0120	0160	0199	0239	0279	0314	0359
0.1	0398	0438	0178	0517	0557	0596	0616	0575	0714	0753
0.2	0793	0832	0871	0910	0448	0937	1026	1034	1103	1141
0.3	1179	1217	1255	1293	1331	1368	1405	1443	1480	1517
0.4	1554	1591	1628.	1654	1700	1736	1772	1808	1841	1879
0.5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0.6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0.7	2580	2611	2642	2673	2703	2734	2764	2794	2823	2852
0.8	2881	2910	2939	2967	29Э5	3023	3051	3078	3103	3133
0.9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1.0	3413	3437	3461	3485	3508	3531	3554	3577	3599	3621
1.1	3643	ЗС65	3G86	3708	3729	3749	3770	3790	3810	3830
1.2	3849	3869	3883	3907	3925	3944	3932	3980	3997	4015
1.3	4032	4049	4066	4082	4039	4115	4131	4147	4162	4177
1.4	4192	4207	4222	4236	4251	4265	4279	4292	4305	4319
1.5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1.6	4452	4463	4474	4184	4495	4505	4515	4525	4535	4545
1.7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1.8	4641	4019	4656	4654	4671	4673	4636	4693	4699	4706
1.9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2.0	47725	47778	47831	47882	47932	47981	48030	48077	48124	481699
2.1	48214	48257	48300	48341	48582	48422	48461	48500	48537	48574
2.2	48610	48645	48679	48713	48746	48778	48809	48840	48870	48899
2.3	48928	48956	48983	49010	49036	490S1	49086	49111	49135	49158
2.4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2.5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2.6	49534	49547	49560	49573	49585	49397	49609	49621	49532	49643
2.7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49733
2.8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2.9	49813	49519	49825	49831	49836	49841	49846	49851	49856	49861
3	0.4986	49903	49931	49952	49966	49977	49984	49939	49993	49995

4.0	0.499968
4.5	0.499997
5.0	0.49999997

Таблица II

Функция распределения нормированного нормального распределения

$$\Phi(x) = (1/\sqrt{2\pi})\int_{-\infty}^{x} e^{-t^2/2}dt$$

x	0	1	2	3	4	5	6	7	8	9
0.0	0,50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0.1	53983	54380	54776	55172	55567	55962	56356	56749	57142	57535
0.2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0.3	61791	62172	62552	62930	63307	63683	64058	64431	64803	65173
0.4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68793
0.5	69146	69497	69847	70194	70540	70884	71226	71566	71904	72240
0.6	72575	72907	73237	73565	73891	74215	74537	74857	75175	75490
0.7	75804	76115	76424	76730	77035	77337	77637	77935	78230	78524
0.8	78814	79103	79389	79673	79955	80234	80511	80785	81057	81327
0.9	81594	81859	82121	82381	82639	82894	83147	83398	83646	83891
1.0	84134	84375	84614	84850	85083	85314	85543	85769	85993	86214
1.1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1.2	88493	88686	88877	89065	89251	89435	89617	89796	89973	90147
1.3	90320	90490	90658	90824	90988	91149	91308	91466	91621	91774
1.4	91924	92073	92220	92364	92507	92647	92786	92922	93056	93189
1.5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1.6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449
1.7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327
1.8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062
1.9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670
2.0	97725	97778	97831	97882	97932	97982	98030	98077	98124	98169
2.1	98214	98257	98300	98341	98382	98422	98461	98500	98537	98574
2.2	98610	98645	98679	98713	98745	98778	98609	98840	98870	98899
2.3	98928	98956	98983	99010	99036	99061	99066	99111	99134	99158
2.4	99180	99202	99224	99245	99266	99286	99305	99324	09343	99361
2.5	99379	99396	99413	99430	99446	99461	99477	99492	99500	99520
2.6	99534	99547	99560	99573	99585	99598	90609	99621	99632	99648
2.7	99653	99664	99674	99683	99693	99702	99711	99720	99728	99736
2.8	99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
2.9	99813	99819	99825	99831	99836	99841	99846	99851	99856	99861
3.0	99865	99869	99874	99878	99882	99885	99889	99893	99896	99900
3.1	99903	99906	99910	99913	99916	99918	99921	99924	99926	99929
3.2	99931	99934	99936	99938	99940	99942	99944	99946	99948	99959
3.3	99952	99953	99955	99957	99958	99960	99961	99962	99964	99966
3.4	99966	99968	99969	99970	99971	99972	99973	99974	99975	99976
3.5	99977	99978	99978	99979	99980	99981	99981	99982	99983	99983
3.6	99984	99985	99985	99986	99986	99987	99987	99988	99988	99989
3.7	99989	99990	99990	99990	99991	99991	99992	99992	99992	99993
3.8	99993	99993	99993	99994	99994	99994	99994	99995	99995	99995
3.9	99995	99995	99996	99996	99996	99996	99996	99996	99997	99997
4.0	99997	99998	99999	99999	99999					

Таблица III

Квантили $t_p(n)$ порядка p распределения Стьюдента с n степенями свободы

p	0.750	0.900	0.950	0.975	0.990	0.995	0.999
1	1.0000	30777	6.3138	12.7062	31.8205	63.8567	318.3068
2	0.8185	1.8856	2.9200	4.3027	6.9646	9.9246	22.3271
3	7649	6377	2.3534	3.1824	4.5407	5.8409	10.2145
4	7407	5332	2.1318	2.7764	3.7469	4.6041	7.1732
5	7267	4759	2.0150	2.5706	3.3649	4.0321	5.8934
7	7111	4149	8946	3646	2.9980	3.4995	4.7853
8	7064	3968	8595	3060	2.8965	3.3554	4.5008
9	7027	3830	8331	2622	2.8214	3.2498	4.2968
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693	4.1437
11	6974	3634	7959	2010	7181	3.1058	4.0247
12	6955	3562	7823	1788	6810	3.0545	3.9296
13	6938	3502	7709	1604	6503	3.0123	3.8520
14	6924	3450	7613	1448	6245	2.9768	7874
15	6912	3406	7530	1314	6025	2.9467	7328
16	6901	3368	7459	1199	5835	9208	6862
17	6892	3334	7396	1098	5669	8982	6458
18	6884	3304	7341	1009	5524	8784	6105
19	6876	3277	7291	0930	5395	8609	5794
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453	3.5518
21	6864	3232	7207	0796	5177	8314	5272
22	6858	3212	7171	0739	5083	8188	5050
23	6853	3195	7139	0687	4999	8073	4850
24	6848	3178	7109	0639	4922	7669	4668
25	6844	3163	7081	0595	4851	7874	4502
26	6840	3150	7056	0555	4786	7787	4350
27	6837	3137	7033	0518	4727	7707	4210
28	6834	3125	7011	0484	4671	7633	4082
29	6830	3114	6991	0452	4620	7564	3962
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500	3.3852
40	6805	3031	6839	0211	4233	7045	3069
50	6794	2987	6759	0086	4033	6778	2614
60	6786	2958	6706	0003	3901	6603	2317
70	6780	2938	6669	1.9944	3808	6479	2108
80	6776	2922	6641	9904	3739	6387	1953
90	6772	2910	6820	9867	3685	6316	1833
100	6770	2901	6602	9840	3642	6259	1737
200	6757	2858	6525	9719	3451	6006	1315
500	6750	2832	6479	9647	3338	5857	1066
∞	6745	2816	6449	9600	3263	5758	0902

Таблица IV

Плотность стандартного нормального распределения $\varphi(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}$

x	Сотые доли									
	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0.0	0.3989	0.3989	0.3989	0.3988	0.3986	0.3984	0.3982	0.3980	0.3977	0.3973
0.1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0.2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0.3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3687
0.4	3683	3668	3653	3637	3621	3605	3588	3572	3555	3538
0.5	3521	3503	3485	3467	3448	3429	3410	339	3372	3352
0.6	3332	3312	3292	3271	3251	3230	3209	3«7	3166	3144
0.7	3123	ЗС1	3079	3056	3034	3011	2989	2966	2943	2920
0.8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0.9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1.0	0.2420	0.2396	0.2371	0.2347	0.2323	0.2299	0.2275	0.2251	0.2227	0.2203
1.1	2179	2165	2131	2107	2083	2059	2036	2012	1989	1965
1.2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1.3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1.4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1.5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1.6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1.7	0940	0926	0909	0893	0878	0863	0848	0833	0818	0804
1.8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1.9	0656	0644	0632	0620	0608	0596	0564	0573	0562	0551
2.0	0.0540	0.0529	0.0519	0.0508	0.0498	0.0488	0.0478	0.0468	0.0469	0.0449
2.1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2.2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2.3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2.4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2.5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2.6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2.7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2.8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2.9	0060	0058	0056	0055	0083	0051	0050	0048	0047	0046
3.0	0.0044	0.0043	0.0042	0.0040	0.0039	0.0038	0.0037	0.0036	0.0035	0.0034
3.1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3.2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3.3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3.4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3.5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3.6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3.7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003

3.8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3.9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Таблица V

Значения t при разных уровнях значимости и данном числе степеней свободы ν

ν	Уровни значимости дли одностороннего критерия (Q)							
	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
	Уровни значимости для одностороннего критерия (2Q)							
	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
1	3,08	6,31	12,71	31,82	63,66	127,32	318,30	636,61
2	1,89	2,92	4,30	6,96	9,92	14,09	22,33	31.60
3	1,64	2,35	3,18	4,54	5,84	7,45	10,21	12,92
4	1,53	2,13	2,78	3,75	4,60	5,60	7,17	8,61
5	1,48	2,02	2,57	3,36	4,03	4,77	5,89	6,87
6	1,44	1,94	2,45	3,14	3,71	4,32	5,21	5,96
7	1,41	1,89	2,36	3,00	3,50	4,03	4,79	5,41
8	1,40	1,86	2,31	2,90	3,36	3,83	4,50	5,04
9	1,38	1,83	2,26	2,82	3,25	3,69	4,30	4,78
10	1,37	1,81	2,23	2,76	3,17	3,58	4,14	4,59
11	1,36	1,80	2,20	2,72	3,11	3,50	4,02	4,44
12	1,36	1,78	2,18	2,68	3,05	3,43	3,93	4,32
13	1,35	1,77	2,16	2,68	3,01	3,37	3,85	4,22
14	1,34	1,76	2,14	2,62	2,98	3,33	3,79	4,14
15	1,34	1,75	2,13	2,60	2,95	3,29	3,73	4,07
16	1,34	1,75	2,12	2,58	2,92	3,25	3,69	4,02
17	1,33	1,74	2,11	2,57	2,90	3,22	3,65	3,97
18	1,33	1,73	2,10	2,55	2,88	3,20	3,61	3,92
19	1,33	1,73	2,09	2,54	2,86	3,17	3,58	3,88
20	1,33	1,72	2,09	2,53	2.85	3,15	3,55	3,85
21	1,32	1,72	2,08	2,52	2,83	3,14	3,53	3,82

22	1,32	1,72	2,07	2,51	2,82	3,12	3.51	3,79
23	1,32	1,71	2,07	2,50	2,81	3,10	3,48	3,77
24	1,32	1,71	2,06	2,49	2,80	3,09	3,47	3,75
25	1,32	1,71	2,06	2,49	2,79	3,08	3,45	3,73
26	1,32	1,71	2,06	2,48	2,78	3,07	3,44	3,71
27	1,31	1,70	2,05	2,47	2,77	3,06	3,42	3,69
28	1,31	1,70	2,05	2,47	2,76	3,05	3,41	3,67
29	1,31	1,70	2,05	2,46	2,76	3,04	3,40	3,66
30	1,31	1,70	2,04	2,46	2,75	3,03	3,39	3,65
40	1,30	1,68	2,02	2,42	2,70	2,97	3,31	3,55
60	1,30	1,67	2,00	2,39	2,66	2,91	3,23	3,46
120	1,29	1,66	1,98	2.36	2,62	2,86	3,16	3,37
∞	1,28	1,64	1,96	2,33	2,58	2,81	3,09	3,29

Учебное издание

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания и индивидуальные задания

Составители

КОНСТАНТИНОВА Людмила Ивановна
МОЛДОВАНОВА Евгения Александровна

Рецензент
кандидат физ.-мат. наук,
доцент кафедры ПМ ИК

Г.Е. Шевелев

Редактор *С.В. Ульянова*

Компьютерная верстка *Е.А. Молдованова*

Отпечатано в Издательстве ТПУ в полном соответствии
с качеством предоставленного оригинал-макета

Подписано к печати . Формат 60×84/16. Бумага «Снегурочка».
Печать Xerox. Усл.печ.л. 1,16. Уч.-изд.л. 1,05.
Заказ . Тираж экз.

Национальный исследовательский Томский политехнический университет
Система менеджмента качества
Издательства Томского политехнического университета сертифицирована
NATIONAL QUALITY ASSURANCE по стандарту BS EN ISO 9001:2008

ИЗДАТЕЛЬСТВО ТПУ. 634050, г. Томск, пр. Ленина, 30.
Тел./факс: 8(3822)56-35-35, www.tpu.ru