Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования «Тульский государственный университет»

Институт прикладной математики и компьютерных наук

Кафедра «Информационная безопасность»

**КУРСОВАЯ РАБОТА**

по дисциплине

«Теория вероятностей и математическая статистика»

на тему

«Случайные величины. Распределение Стьюдента (t-распределение)»

|  |  |  |  |
| --- | --- | --- | --- |
| Выполнила: | ст. гр. 230711 | \_\_\_\_\_\_\_\_\_\_\_  (подпись) | Павлова В.С. |
| Проверила: | к. т. н, доц. каф. ПМиИ | \_\_\_\_\_\_\_\_\_\_\_  (подпись) | Родионова Г.А. |

Тула, 2023

**ЗАДАНИЕ**

на курсовую работу по дисциплине

«Теория вероятностей и математическая статистика»

Павловой Виктории Сергеевны

230711

студента гр. \_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Тема курсовой работы

«Случайные величины. Распределение Стьюдента (t-распределение)»

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Исходные данные  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Задание получил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись)

(ФИО)

Задание выдал \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

21.02.2023 г.

21.02.2023 г.

21.02.2023 г.

21.02.2023 г.

21.02.2023 г.

21.02.2023 г.

21.02.2023 г.

21.02.2023 г.

(ФИО)

(подпись)

Дата выдачи задания \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

21.02-28.02 – Получение и ознакомление с заданием

01.03-22.03 – Изучение литературы и других исходных материалов

23.03-03.05 – Изучение теории, раскрывающей тему курсовой работы  
04.05-17.05 – Реализация практической части курсовой работы  
18.05-24.05 – Анализ результатов   
25.05-07.06 – Оформление пояснительной записки и сдача на проверку  
28.06.2023 – Защита курсовой работы

21.02-28.02 – Получение и ознакомление с заданием

01.03-22.03 – Изучение литературы и других исходных материалов

23.03-03.05 – Изучение теории, раскрывающей тему курсовой работы

21.02-28.02 – Получение и ознакомление с заданием

01.03-22.03 – Изучение литературы и других исходных материалов

23.03-03.05 – Изучение теории, раскрывающей тему курсовой работы

21.02-28.02 – Получение и ознакомление с заданием

01.03-22.03 – Изучение литературы и других исходных материалов

23.03-03.05 – Изучение теории, раскрывающей тему курсовой работы

21.02-28.02 – Получение и ознакомление с заданием

01.03-22.03 – Изучение литературы и других исходных материалов

23.03-03.05 – Изучение теории, раскрывающей тему курсовой работы

21.02-28.02 – Получение и ознакомление с заданием

01.03-22.03 – Изучение литературы и других исходных материалов

23.03-03.05 – Изучение теории, раскрывающей тему курсовой работы

21.02-28.02 – Получение и ознакомление с заданием

01.03-22.03 – Изучение литературы и других исходных материалов

23.03-03.05 – Изучение теории, раскрывающей тему курсовой работы

21.02-28.02 – Получение и ознакомление с заданием

01.03-22.03 – Изучение литературы и других исходных материалов

23.03-03.05 – Изучение теории, раскрывающей тему курсовой работы

График выполнения КР \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Рекомендации и особые отметки \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_» \_\_\_\_\_\_\_\_20\_\_г

# **Содержание**

[Введение 4](#_Toc136096769)

[1 Случайная величина и её характеристики 6](#_Toc136096770)

[1.1 Одномерная случайная величина 6](#_Toc136096771)

[1.2 Основные виды распределения случайных величин 10](#_Toc136096772)

[1.3 Распределение Стьюдента 13](#_Toc136096773)

[2 Применение t-распределения при решении практических задач 20](#_Toc136096774)

[2.1 Оценка среднего значения генеральной совокупности 21](#_Toc136096775)

[2.2 Использование t-распределения для вычисления вероятности 22](#_Toc136096776)

[2.3 Задача об оценке статистической значимости (анализ регрессии) 23](#_Toc136096777)

[2.4 Проверка гипотезы о среднем значении показателя в генеральной совокупности 25](#_Toc136096778)

[2.5 Задача о проверке гипотезы о значимости коэффициента регрессии 25](#_Toc136096779)

[Заключение 28](#_Toc136096780)

[Список использованных источников 29](#_Toc136096781)

# **Введение**

Математическая дисциплина, которая изучает случайные явления и вероятности их возникновения, называется теорией вероятностей. Она имеет широкий спектр применений, начиная от криптографии и финансовой аналитики и заканчивая разработками в сфере искусственного интеллекта. История теории вероятностей уходит в далекое прошлое, и её основоположником считается математик Блез Паскаль. В 1654 году он начал свои исследования вероятностей, решая задачу о разделе выигрыша в азартной игре.

За несколько сотен лет теория вероятностей развилась в самостоятельную научную дисциплину, строго формализованную и включающую в себя различные математические методы, теории и анализ случайных процессов. Идеи о понятии распределения вероятностей возникли ещё в 17-ом веке у Пьера де Ферма и Блеза Паскаля, но основополагающую работу в этой области сделал Абрахам де Муавр в 18-ом веке, опубликовав 1733 году свою работу под названием «Доклад о шансах», в которой впервые применил анализ и статистику для описания случайных событий и введения понятия распределения вероятностей. Он показал, как можно использовать биномиальное распределение для приближения вероятности великого числа независимых случайных испытаний.

Окончательно теория вероятностей как математическая наука оформилась в 30-х годах 20-го века, когда Андреем Николаевичем Колмогоровым была предложено аксиоматическое определение вероятности. Колмогоров также формализовал понятие случайной величины и установил строгие математические основы для изучения вероятностных явлений. С помощью теории вероятностей стало возможным описывать случайные величины и их распределения.

Одно из наиболее изучаемых распределений случайных величин в теории вероятностей – нормальное распределение, также известное как распределение Гаусса, названное в честь математика Карла Фридриха Гаусса, который изучил его основные свойства в 18-ом веке, например то, что нормальное распределение подходит для моделирования большого количества случайных процессов. В частном случае, когда известна только выборка, но неизвестна генеральная совокупность, такую модификацию нормального распределения называют распределением Стьюдента или t-распределением.

Темой данной курсовой работы является распределение Стьюдента, тесно связанное с нормальным распределением. Целью работы является изучение свойств этого распределения и решение практических задач, с ним связанных. Актуальность исследования распределение Стьюдента обоснована тем, что оно более устойчиво к выбросам и отклонениям от среднего значения в сравнении с нормальным, а значит может более эффективно использоваться в качестве инструмента для моделирования и анализа случайных процессов в теории вероятностей и математической статистике.

# **1 Случайная величина и её характеристики**

## 1.1 Одномерная случайная величина

Понятие случайной величины играет важную роль в теории вероятностей и является ключевым инструментом, позволяющим формально описывать случайные явления и их вероятностные свойства. ***Одномерная случайная величина*** — это скалярная функция, определенная на пространстве элементарных событий, сопоставляющая каждому элементарному событию число или, иначе говоря, это величина, значение которой зависит от того, каким элементарным событием закончился случайный опыт [1]. Случайные величины обозначаются заглавными буквами латинского алфавита, например . Среди общепринятых обозначений можно выделить также , которое часто используется для обозначения случайной величины, имеющей нормальное (гауссово) распределение, и , которое используется для обозначения распределения Стьюдента (t-распределение).

Под ***элементарным исходом*** или элементарным событием понимается любой простейший, то есть неразделимый в контексте рассматриваемого опыта, исход. Множество всех возможных значений или результатов случайного эксперимента называют ***пространством элементарных исходов*** и обозначают символом *.* Иначе говоря, множество исходов опыта образует пространство элементарных исходов, если верно следующее [2]:

* в результате опыта один из исходов обязательно происходит;
* появление одного из исходов исключает появление всех прочих;
* в рамках данного опыта элементарных исход неразделим.

Более строго данную связь можно выразить следующим образом: , а множество . Чтобы задать некоторую случайную величину , необходимо поставить в соответствие каждому элементарному исходу числовое значение , которое случайная величина примет, если в результате опыта произойдет именно этот исход. Например, если каждому элементарному исходу , , ставится в соответствие число , то получим случайную величину, определяемую следующим образом (таблица 1):

Таблица 1 – Соответствие между элементарными исходами и значениями

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | … |  |
|  | 1 | 2 | … |  |

Таким образом, скалярную функцию , заданную на пространстве элементарных исходов , и называют случайной величиной, если для любого множество элементарных исходов, удовлетворяющих условию , является событием.

Случайные величины могут быть разделены на ***дискретные***, то есть такие, которые принимают набор отдельных изолированных значений (конечное (счётное) или бесконечное число), или ***непрерывные***, набор значений которых целиком заполняет некоторый интервал [3].

Основной характеристикой случайных величин является ***закон распределения случайной величины***, также называемый распределением её вероятностей. Общим законом распределения и для дискретных, и для непрерывных случайных величин является ***функция распределения*** – функция для случайной величины , значение которой в точке равно вероятности события , то есть события, состоящего из тех и только тех элементарных исходов , для которых :

Типичный вид функции распределения приведён на рисунке 2.

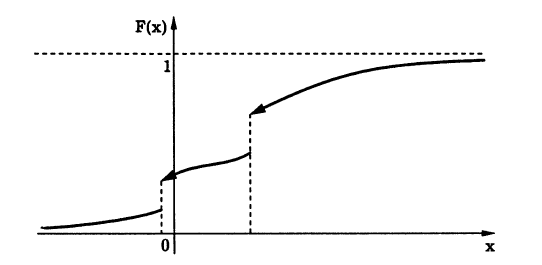
****

Рисунок 2 – Вид графика функции распределения

Из того факта, что функция распределения – это вероятность, следует, что она обладает следующими свойствами:

1. Функция распределения – это неубывающая функция: при ;
2. Множество значений принадлежит отрезку [0, 1], откуда следует, что , ;
3. Вероятность попадания в интервал ( можно определить следующим образом: ;

Законом распределения дискретной случайной величины называют таблицу, состоящую из двух строк, в одной из которых перечислены все возможные значения случайной величины, а во второй – вероятности, что случайная величина примет эти значения, то есть . При этом , то есть сумма по строке вероятностей должна равняться единице [2]. Общий вид закона распределения дискретной случайной величины приведён в таблице 2.

Таблица 2 – Закон распределения дискретной случайной величины в общем виде

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | … |  | … |  |
|  |  |  | … |  | … |  |

График функции распределения дискретной случайной величины представлен на рисунке 3. График представляет собой ступенчатую линию, где каждая ступень соответствует возможному значению случайной величины, а её высота (высота скачка в данной точке) определяет вероятность возникновения этого значения.

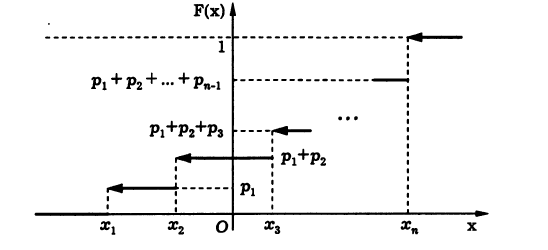
**

Рисунок 3 – График функции распределения дискретной случайной величины

Для непрерывной случайной величины функцию распределения можно выразить с помощью ***функции плотности распределения*** следующим образом:

Из этого следует, что функцию можно называть дифференциальным законом распределения непрерывной случайной величины, а – интегральным, причём для этих законов верно следующее:

Функция плотности обладает следующими свойствами:

1. , поскольку не убывает;
2. ;
3. Вероятность попадания в интервал ( можно определить следующим образом ;
4. .

График функции плотности распределения приведён на рисунке 4. Согласно свойству 3, площадь криволинейной трапеции, обозначенной на рисунке штриховкой, равна вероятности попадания непрерывной случайной величины в полуинтервал [. Аналогичным образом в силу свойства 2 площадь под всей кривой распределения равна единице.

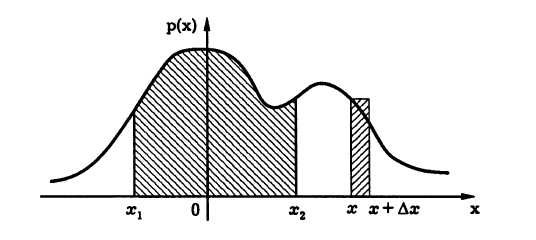


Рисунок 4 – График функции плотности распределения непрерывной случайной величины

## 1.2 Основные виды распределения случайных величин

Среди наиболее часто встречающихся на практике распределений можно выделить следующие [2]:

1. ***Биномиальное распределение*** дискретной случайной величины, задаваемое следующим образом:, где . Его закон распределения приведён в таблице 3.

Таблица 3 – Ряд биномиального распределения

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | … |  | … |  |
|  |  |  | … |  | … |  |

1. ***Распределение Пуассона*** дискретной случайной величины, определяемое следующим образом: , где а – параметр распределения Пуассона. Его закон распределения приведён в таблице 4.

Таблица 4 – Ряд распределения Пуассона

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | 2 |  | n |  |
|  |  |  |  |  |  |  |

1. ***Равномерное распределение*** непрерывной случайной величины, плотность распределения которой определяется выражением:

Исходя из того факта, что , закон распределения равномерного распределения записывается следующим образом:

Графики обеих функций представлены на рисунке 5.

|  |  |
| --- | --- |
| А) | Б) |

Рисунок 5 – График функции: а) плотности равномерного распределения; б) закона равномерного распределения

1. ***Экспоненциальное распределение*** непрерывной случайной величины, функция плотности распределения которого имеет параметр и определяется как:

Случайная величина, распределённая по экспоненциальному (показательному) закону, имеет функцию распределения вида:

Графики данных функций представлены на рисунке 6.

|  |  |
| --- | --- |
| А) | Б) |

Рисунок 6 – График функции: а) плотности показательного распределения; б) закона показательного распределения

1. ***Нормальное распределение*** непрерывной случайной величины, функция плотности распределения которого имеет вид:

Как показано на рисунке 7, данное распределение, в частности, вид его графика, зависит от двух параметров – и . Его закон распределения записывается следующим образом:

|  |  |
| --- | --- |
| А) | Б) |

Рисунок 7 – График функции: а) плотности распределения Гаусса; б) закона распределения Гаусса

## 1.3 Распределение Стьюдента

Распределение Стьюдента (или t-распределение) используется для моделирования случайных величин, когда имеется ограниченное количество наблюдений и некоторая неопределенность в данных. Распределение Стьюдента получило свое название в честь псевдонима, используемого Уильямом Сили Госсетом, который в начале 20-го века внёс значительный вклад в решение проблемы, связанной с оценкой среднего значения на основе ограниченного количества наблюдений [4]. Вопрос, которым занимался Стьюдент, был связан с построением так называемых ***доверительных интервалов*** – диапазонов значений, в которых с определённой вероятностью содержится истинное значение параметра генеральной совокупности (например, среднего значения). Ранее для этого использовалось распределение Гаусса, однако при малом объеме выборки ( наблюдений) и неизвестном стандартном отклонении генеральной совокупности, использование нормального распределения становилось неточным – с этим и столкнулся Госсет, а после разработал распределение, основанное на t-статистике, которое позволяло делать более точные оценки и строить доверительные интервалы при использовании малых выборок и неизвестного стандартного отклонения генеральной совокупности. Он опубликовал свои исследования под псевдонимом «Стьюдент», откуда и произошло название распределения.

***Распределением Стьюдента (t-распределением)*** называют распределение, функция плотности вероятности которого имеет вид:

где – случайная величина, – число степеней свободы и – гамма-функция.

Под гамма-функцией понимается гамма-функцией Эйлера – специальная функция в математике, введённая Леонардом Эйлером для обобщения понятия факториала на комплексные и действительные числа. Она имеет вид: [5]

Гамма-функция определена для всех комплексных чисел с положительной вещественной частью, то есть для . Одним из интересных свойств гамма-функции является возможность применения к ней так называемой формулы понижения [5]:

Это позволяет вычислять значения гамма-функции для разных аргументов, используя уже известные значения. Эйлерова гамма-функция также обладает симметрией, поэтому к ней применима формула отражения:

Формула отражения позволяет вычислять значения гамма-функции для отрицательных аргументов или аргументов, удаленных от положительных целых чисел. В контексте теории вероятностей гамма-функция играет важную роль в определении и вычислении функций распределения и плотности распределения некоторых случайных величин, в частности, распределения Стьюдента. Гамма-функция входит в функцию плотности распределения Стьюдента через отношение . Это отношение возникает из применения гамма-функции для нормировки t-статистики, которая определяется как отношение выборочного среднего и стандартной ошибки. Используя гамма-функцию, мы нормируем t-статистику, чтобы получить функцию плотности распределения Стьюдента.

Как видно из формулы функции распределения, t-распределение имеет один параметр, называемый ***степенью свободы*** распределения Стьюдента. Степень свободы влияет на ширину и форму распределения и является мерой количества независимых наблюдений, используемых для оценки распределения. Чем больше степень свободы, тем ближе t-распределение приближается к нормальному. При этом, при меньшей степени свободы, распределение имеет более плоскую форму.

График функции t-распределения, то есть ***кривая распределения Стьюдента*** имеет колоколообразную форму, напоминающую нормальное распределение, но с более тяжёлыми «хвостами». При распределение стремится к нормальному закону распределения. На рисунке 8 показан график функции t-распределения при значениях .

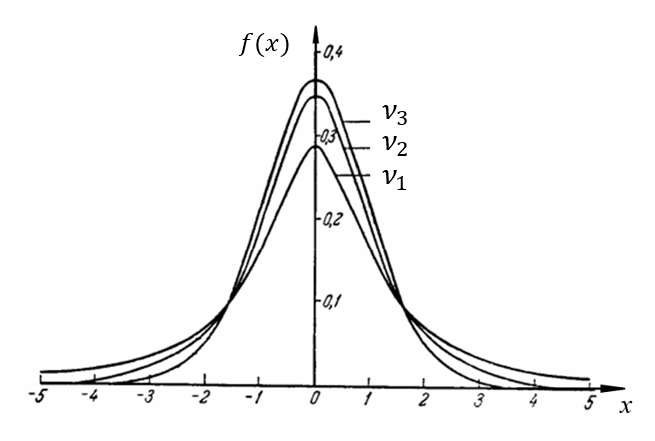


Рисунок 8 – График функции плотности t-распределения с числом степеней свободы

***Функция распределения*** t-распределения в общем случае имеет вид:

Она не имеет аналитического решения в данном случае, поэтому для вычисления её значений часто используются таблицы или специальные программы. Её график схож с графиком функции распределения нормального распределения, как показано на рисунке 9:

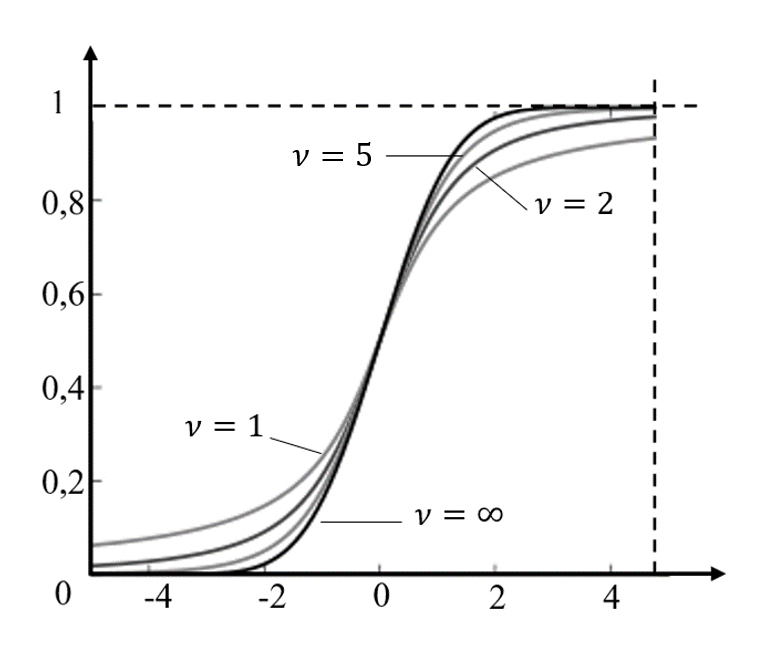


Рисунок 9 – График функции распределения t-распределения

Числовые характеристики распределения Стьюдента зависят от его степени свободы:

1. ***Математическое ожидание*** t-распределения определено и равно нулю при любом :
2. ***Дисперсия*** t-распределения определяется следующей формулой:
3. ***Среднеквадратичное отклонение*** t-распределения равно квадратному корню из дисперсии, то есть:
4. ***Центральный момент*** t-распределения n-го порядка определяется следующим образом:
5. ***Начальный момент*** для t-распределения с ν степенями свободы не имеет аналитической формулы для . Это связано с тем, что функция плотности вероятности t-распределения не интегрируется аналитически для степеней выше , а данная характеристика вычисляется численными методами или аппроксимациями.
6. ***Мода*** равна значению при .
7. ***Медиана*** t-распределения также равна нулю.

Для анализа точности оценок статистических параметров на основе выборочных данных можно использовать формулу для ***вероятности попадания случайной величины в доверительный интервал*** , которая выглядит следующим образом:

где – вероятность отклонения результатов эксперимента от нулевой гипотезы при условии, что она верна. Для нахождения значения ***уровня значимости*** необходимо знать уровень доверия для построения доверительного интервала. Уровень доверия обычно задается в процентах и представляет собой вероятность того, что истинное значение параметра лежит внутри построенного доверительного интервала. Например, для уровня доверия 95% соответствующее значение будет равно 0.05, так как это означает, что необходимо рассмотреть только 5% случаев, когда истинное значение параметра лежит за пределами доверительного интервала, а в остальных 95% случаев это значение лежит внутри этого интервала. Формула для нахождения уровня значимости может быть выражена следующим образом:

Если среднее значение выборки попадает в заданный доверительный интервал (то есть интервал, вероятность попадания среднего значения выборки в который определена заранее), то мы не можем отвергнуть нулевую гипотезу о том, что различия между двумя выборками статистически не значимы на уровне значимости 5%. Для того чтобы отклонить нулевую гипотезу и доказать статистическую значимость различий между двумя выборками, среднее значение выборки должно выходить за пределы заданного доверительного интервала [6].

Одним из важных понятий, связанных с t-распределением, является ***t-статистика*** – мера различия между двумя выборками данных. Она часто используется в статистических тестах гипотез, чтобы определить, является ли различие между выборками значимым или случайным. Если t-статистика больше критического значения, то можно сделать вывод о статистически значимом различии между выборками. Она может быть определена как отношение оценки параметра модели к её стандартной ошибке. В этом случае формула будет иметь вид:

где – оценка параметра модели (числовое значение, которое определяет взаимосвязь между независимыми и зависимыми переменными в рамках выбранной модели), – истинное значение параметра модели, а – стандартная ошибка оценки параметра модели.

Существует некоторое значение, которое разделяет вероятностное распределение Стьюдента на две области и называется оно ***квантилем***. Выражаясь более формально, t-квантиль уровня – это такое число , для которого вероятность , где T представляет распределение Стьюдента с степенями свободы [7]. Квантиль порядка также записывают как .

В контексте задач систематизации и обработки результатов наблюдений массовых случайных явлений для выявления существующих закономерностей необходимо рассмотреть такое понятие, как ***совокупность всех подлежащих изучению объектов*** или возможных результатов наблюдений, проводимых в неизменных условиях над одним объектом, называется генеральной совокупностью. Выборочной совокупностью или ***выборкой*** называется совокупность объектов, отобранных случайным образом из генеральной совокупности. Число объектов в совокупности называется ее объемом. Считается, что объем генеральной выборки бесконечен. Конкретные значения выборки, полученные в результате наблюдений, называют реализацией выборки и обозначают .

# **2 Применение t-распределения при решении практических задач**

Среди задач, в которых для решения используется t-распределение, можно выделить следующие наиболее практически значимые:

* ***Оценка среднего значения*** генеральной совокупности при неизвестной дисперсии: если известен только размер выборки и выборочное среднее значение, то можно использовать распределение Стьюдента для оценки значимости различий между выборочным и генеральным средним.
* ***Расчет доверительного интервала для среднего значения*** генеральной совокупности: распределение Стьюдента также позволяет вычислить доверительный интервал для среднего значения генеральной совокупности с известной точностью.
* ***Оценка статистической значимости*** различий между двумя выборками: если известны средние значения и стандартные отклонения двух выборок, то можно использовать распределение Стьюдента для проверки гипотезы о равенстве этих средних значений.
* ***Оценка коэффициента корреляции***: если известны выборочные коэффициенты корреляции и размер выборки, то можно использовать распределение Стьюдента для проверки гипотезы о значимости корреляции.
* ***Оценка параметров линейной регрессии***: распределение Стьюдента используется для проверки гипотезы о значимости коэффициентов линейной регрессии при небольшом размере выборки.

Рассмотрим далее несколько конкретных примеров задач, решение которых связанно с t-распределением и t-статистикой Стьюдента.

## 2.1 Оценка среднего значения генеральной совокупности

**Задача**. Оценить средний рост студентов в университете среди случайным образом выбранных 10 студентов. Данные выборки = [170, 165, 172, 168, 175, 172, 169, 171, 173, 170] [8]

**Решение**. Среднее значение выборки вычислим по формуле:

Таким образом, средний рост студентов в выборке равен см. Следом вычислим стандартное отклонение выборки:

Следующим шагом будет вычисление значения стандартной ошибки среднего по формуле:

Получим, что значение стандартной ошибки среднего составляет примерно см. При надежности 95% уровень значимости и число степеней свободы . Из таблиц получаем значение t-критерия, равное . Теперь нам необходимо определить доверительный интервал , который показывает, с какой вероятностью истинное среднее значение генеральной совокупности находится в заданном интервале. Для вычисления доверительного интервала нам необходимо использовать формулу:

Мы получили, что доверительный интервал составляет , что равно Таким образом, с 95% вероятностью мы можем утверждать, что истинное среднее значение роста студентов в университете находится в этом интервале.

**Ответ**: оценка среднего роста студентов в университете среди случайно выбранных 10 студентов со 95% вероятностью лежит в интервале

## 2.2 Использование t-распределения для вычисления вероятности

**Задача**. Найти вероятность того, что случайная величина с десятью степенями свободы лежит в пределах [9].

**Решение**. Известно, что . Для решения задачи необходимо сначала перейти от случайной величины к соответствующей ей случайной величине с распределением Стьюдента. По таблицам t-распределения найдём вероятности соответствующих интервалов и .

Для распределения Стьюдента со степенями свободы существует также таблица значений, которая позволяет найти квантили порядка для определенных значений и уровня значимости . Исходя из них в данном случае уровень значимости . Теперь необходимо найти значения t-статистики для интервалов, соответствующих квантилям и . Находим, что первый из них соответствует t-статистике, равной , а соответствует t-статистике, равной .

Наконец, для вычисления искомой вероятности необходимо привести интервал к виду , где и соответствуют найденным значениям t-статистики. Для этого используется формула линейной интерполяции и искомую вероятность можно вычислить путем взвешенного среднего вероятностей попадания в каждый из интервалов при условии, что случайная величина находится в интервале

**Ответ**:

## 2.3 Задача об оценке статистической значимости (анализ регрессии)

**Задача**. Были случайным образом выбраны 8 учеников, для которых получены следующие выборки:

время на домашнюю работу (в часах) X = [2, 4, 3, 1, 5, 6, 7, 8];

результаты экзамена (в баллах) Y = [60, 70, 65, 55, 75, 80, 85, 90].

Необходимо оценить зависимость между количеством времени, проведенным на дому за учебой, и результатами экзамена по математике [10].

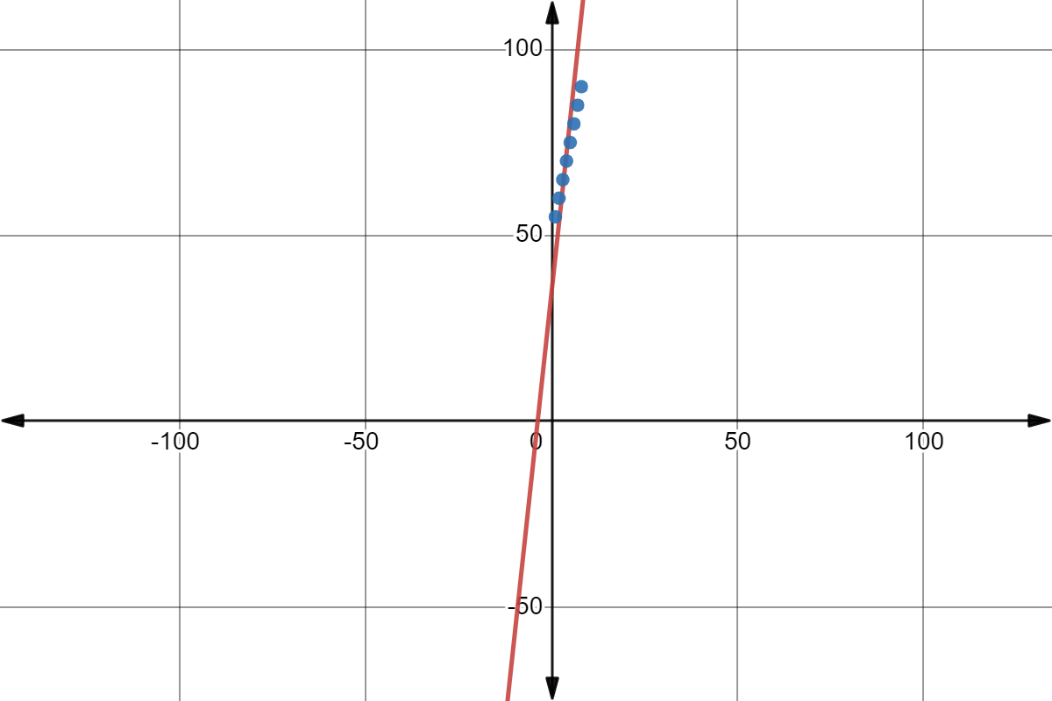
**Решение**. Для оценки параметров линейной регрессии используем распределение Стьюдента. Тогда искомая модель будет иметь вид: , где и – неизвестные коэффициенты, которые необходимо найти.

Сперва найдем по известным формулам для каждой переменной такие характеристики, как выборочное среднее и стандартное отклонение:

Затем рассчитаем выборочный коэффициент корреляции r между x и y:

Теперь мы можем получить оценки коэффициентов и :

Таким образом, наша модель линейной регрессии будет иметь вид . Графически это выглядит следующим образом:

**

Проверим значимость коэффициента наклона с помощью t-теста. Для этого нужно рассчитать значение t-статистики по формуле:

Значение t-статистики соответствует уровню значимости при числе степеней свободы . Это означает, что полученный коэффициент наклона является статистически значимым.

**Ответ**: является статистически значимым.

## 2.4 Проверка гипотезы о среднем значении показателя в генеральной совокупности

**Задача**. Дана выборка из 10 наблюдений. Необходимо проверить гипотезу о том, что среднее значение показателя в генеральной совокупности равно 50. Известно, что стандартное отклонение в выборке составляет 5, а выборочное среднее равно 48 [8].

**Решение**. Тогда для проверки гипотезы используется t-критерий Стьюдента. Формула для расчета значения t-статистики имеет вид:

Далее нам необходимо найти критическое значение t-статистики для заданного уровня значимости и числа степеней свободы. В данном случае число степеней свободы равно . Положим, что уровень значимости α = 0.05. Тогда критическое значение t-статистики можно найти с помощью таблицы распределения Стьюдента .

Следовательно, мы не можем отвергнуть гипотезу о равенстве среднего значения 50 на уровне значимости α = 0.05.

**Ответ:** гипотеза верна.

## 2.5 Задача о проверке гипотезы о значимости коэффициента регрессии

**Задача**. Проверить гипотезу о значимости коэффициента регрессии для независимой переменной x при уровне значимости для следующих данных [8]: ,

**Решение.**  Мы хотим проверить гипотезу о значимости коэффициента регрессии для независимой переменной, то есть проверить, является ли коэффициент значимым или нет. Пусть у нас есть гипотеза (нулевая гипотеза), где – коэффициент регрессии для независимой переменной x. Альтернативная ей гипотеза . Таким образом задача сводится к проверке, является ли коэффициент регрессии значимым. Для проверки можно воспользоваться t-статистикой Стьюдента:

Где – оценка коэффициента регрессии для x, – среднеквадратическое отклонение резидуальных ошибок, n – число наблюдений.

Найдем оценку коэффициента регрессии и среднеквадратическое отклонение следующим образом:

И, наконец, определим t-статистику Стьюдента:

Согласно табличным данным, на уровне значимости и степеней свободы критическое значение t-статистики равно . Так как вычисленное значение , то есть оно больше критического, то мы можем отвергнуть нулевую гипотезу о том, что коэффициент регрессии равен нулю на уровне значимости . То есть, можно сделать вывод о том, что найденный коэффициент регрессии статистически значимо отличается от нуля и имеет влияние на зависимую переменную y.

**Ответ:** существует статистически значимая связь между зависимой переменной y и независимой переменной x.

# **Заключение**

В ходе выполнения данной курсовой работы был рассмотрен такой раздел теории вероятностей, как случайные величины и их распределения, в частности, t-распределение Стьюдента – изучены его основные свойства и общие характеристики. Также были рассмотрены основные виды практических статистических и вероятностных задач, связанных с t-распределением, а также приведено решение нескольких примеров таких задач.

Можно заключить, что благодаря работе Уильяма Госсета и его вкладу в разработку распределения Стьюдента, статистики и исследователи получили важный инструмент, позволяющий делать более точные оценки и строить доверительные интервалы при работе с ограниченными выборками данных. Это имеет важное значение в различных областях, где доступ к большим объемам данных ограничен или невозможен.

# **Список использованных источников**

1. Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х томах. Т. 1: Пер. с англ. – М.: Мир, 1984. – 528 с.
2. Печинкин А.В., Тескин О.И., Г.М. Цветкова и др.; Под ред. Зарубина В.С., Крищенко А.П. Теория вероятностей: Учеб. для вузов. - 3-е изд., испр. изд. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2004. - 456 с.
3. Теория вероятностей и математическая статистика: учеб. пособие / Е. А. Трофимова, Н. В. Кисляк, Д. В. Гилёв; [под общ. ред. Е. А. Трофимовой ; М-во образования и науки Рос. Федерации, Урал. федер. ун-т. – Екатеринбург: Изд-во Урал. ун-та, 2018. – 160 с.
4. Гнеденко Б.В. Курс теории вероятностей: Учебник. – Изд. 8-е., испр. и доп. изд. – М.: Едиториал УРСС, 2005. – 448 с.
5. Зорич В. А. Математический анализ. Часть II. — Изд. 9-е, испр. — М.: МЦНМО, 2019. — xii+676 с. Библ.: 57 назв. Илл.: 41.
6. Понятие t-распределения в тестах на статистическую значимость // RadioProg URL: https://radioprog.ru/post/930 (дата обращения: 21.05.2023).
7. Фишер Р. Applications of «Student’s» distribution (англ.) // metron. – 1925. – Vol. 5. – P. 90 – 104.
8. Леман, Э.Л. Проверка статистических гипотез / перевод с англ. Ю.В. Прохорова. – 2-е изд., испр. – М.: Наука, 1979. – 408 с.
9. Поллард. Д. Справочник по вычислительным методам статистики / Дж. Поллард; Перевод с англ. В. С. Занадворова. – М.: Финансы и статистика, 1982. – 344 с.
10. Кобзарь А.И. Прикладная математическая статистика [Текст]: для инженеров и науч. работников / А. И. Кобзарь. – Изд. 2-е, испр. – Москва : Физматлит, 2012. – 813 с.