

Муниципальное бюджетное образовательное учреждение
«Средняя общеобразовательная школа №31»

Исследовательский проект
по математике:

Математическая статистика в действии

Автор проекта: ученик 10 «Б»
класса Кузьмин Никита

Руководитель проекта:
Резванцева Наталья Валерьевна,
учитель математики

Абакан, 2023

ВВЕДЕНИЕ

Математическая статистика- раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов. При этом статистическими данными называются сведения о числе объектов в какой-либо более или менее обширной совокупности, обладающих теми или иными признаками. В данной работе будут представлены способы применения статистики в жизни, продемонстрированные на примере Excel и Minitab.

Объектом исследования является математическая статистика.

Предметом исследования статистические данные.

Цель работы: рассмотрение статистических данных в различных программах.

Для достижения поставленной цели поставлены следующие **задачи:**

1. Рассмотреть научно-исследовательские работы по данной теме.
2. Раскрыть основные понятия для глубокого анализа данной темы.
3. Проанализировать программы для работы с математической статистикой.
4. Сделать выводы на основе проведённых сравнений.

При выполнении данной работы использованы следующие **методы:** анализ и синтез, обобщение и систематизация материала.

Структура и объем работы. Работа представлена на 15 страницах состоящая из введения, 5 глав, где изложены основные теоретические исследовательские положения работы, заключения и библиографического списка.

ГЛАВА 1. ИСТОРИЯ ПОЯВЛЕНИЯ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Математическая статистика как наука начинается с работ знаменитого немецкого математика Карла Фридриха Гаусса (1777-1855), который на основе теории вероятностей исследовал и обосновал метод наименьших квадратов, созданный им в 1795 г. и примененный для обработки астрономических данных (с целью уточнения орбиты малой планеты Церера). Его именем часто называют одно из наиболее популярных распределений вероятностей – нормальное, а в теории случайных процессов основной объект изучения – гауссовские процессы

В конце XIX в. – начале XX в. крупный вклад в математическую статистику внесли английские исследователи, прежде всего Карл Пирсон (1857-1936) и Рональд Фишер (1890-1962). В частности, Пирсон разработал критерий «хи-квадрат» проверки статистических гипотез, а Фишер – дисперсионный анализ, теорию планирования эксперимента, метод максимального правдоподобия оценки параметров.

В 30-е годы XX в. поляк Ежи Нейман (1894-1977) и англичанин Эгон Пирсон развили общую теорию проверки статистических гипотез, а советские математики академик А.Н. Колмогоров (1903-1987) и член-корреспондент АН СССР Николай Васильевич Смирнов (1900-1966) заложили основы непараметрической статистики. В сороковые годы XX в. румын Абрахам Вальд (1902-1950) построил теорию последовательного статистического анализа [1].

ГЛАВА 2. ОСНОВНЫЕ ОПРЕДЕЛЕНИЯ В МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Перед началом сравнения программ для использования математической статистики, необходимо раскрыть несколько понятий.

Начнём со **среднего арифметического**. Среднее арифметическое- это среднее значение как минимум двух чисел. Чтобы вычислить среднее арифметическое, все числа надо сложить и разделить на общее количество значений.

Общая формула для вычисления среднего арифметического выглядит следующим образом [2]:

$$\text{Среднее арифметическое} = \frac{\text{сумма всех чисел}}{\text{количество слагаемых}}$$

Рис.1. Формула Средней арифметической

Следующее определение- размах ряда. **Размах ряда чисел**- разность между наибольшим и наименьшим из этих чисел [3]. Например, размах ряда 5,24, 6,97, 8,56, 7,32, 6,23 равен $8,56-5,24=3,32$.

Далее идёт мода ряда. **Мода ряда чисел**- число, которое встречается в данном ряду чаще других [3]. Например, в ряде 1, 2, 3, 4, 4, 7,7,7 самое встречающееся число- 7 (она встречается в ряду три раза).

Следующее- медиана. **Медиана**-середина упорядоченного ряда. Медиана делит этот ряд пополам таким образом, что в одной половине стоят все значения меньшие, а в другой все значения большие медианы [4]. Например, чтобы найти медиану в ряде чисел 1, 2, 1, 1, 3, 8, 10, 1, нужно распределить числа в порядке возрастания, т.е. ряд будет выглядеть следующим образом: 1,1,1,1,2,3,8,10. Если в ряде чётное количество показателей, надо взять два средних числа, затем сложить два средних числа и разделить пополам [5], т.е. 1 и 2 являются средними числами, значит медиана ряда будет равна:

$$\frac{1 + 2}{2} = 1,5$$

Рис.2. Нахождение медианы

Последнее определение- статистика объектов нечисловой прямой. **Статистика объектов нечисловой прямой** -раздел математической статистики, в котором статистическими данными являются объекты нечисловой природы, т.е. элементы множеств, не являющихся линейными пространствами, например, ответы на «закрытые» вопросы в социологических анкетах, в которых респондент должен выбрать одну или несколько из фиксированного числа подсказок [6].

Хочется отметить, количество понятий в математической статистике огромно, понимание, которых нужно глубоко понимать многомерный статический анализ, анализ функций (процессов) и временных рядов.

ГЛАВА 3. ОСНОВНЫЕ РАЗНОВИДНОСТИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Особенностью статистики является изучение массовых, случайных явлений в условной неопределенности. Достоверность выводов зависит от числа объектов исследования. На основе анализов и прогнозов вырабатывается оптимальное решение.

Статистика подразделяется на:

- теоретическую** (вырабатывает методы).
- прикладную** (общая, отраслевая (экономическая статистика, метеорологическая, медицинская)) [7].

Так, например, в медицинской статистике систематизируются данные о рождаемости, смертности, заболеваемости и о количестве медицинских учреждений.

Что же до теоретической статистики, существует два вида:

- описательная** (комплекс методов сбора, группировки данных и представления их в виде таблиц, графиков).
- аналитическая** (делает заключения, выводы с целью практического применения) [7].

Хочется отметить, что при написании программ для математической статистики применяется теоретическая статистика, или статистика объектов нечисловой прямой.

ГЛАВА 4. ПРИМЕНЕНИЕ И СРАВНЕНИЕ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В ПРОГРАММАХ

За время существования математической статистики появилось множество программ для работы с ней. Первая программа для работы со статистикой- Excel, которой является самой распространённой из всех перечисленных в данной работе. Далее идёт Minitab- специальная программа для математической статистики. Для сравнения программ на простые операции в математической статистике, были использованы результаты зачёта по геометрии в десятом классе:

	C1	C2	C3	C4-T	C5	C6-T	C7	C8-T
	Номер теоремы	Сдали	Не сдали	Среднее арифметическое		Размах	Медиана	Мода
1	1	23	1	21,6666		5	23	23
2	2	23	1	2,3		5	1	1
3	3	20	4					
4	4	23	1					
5	5	18	6					
6	6	23	1					
7								
8								
9								
10								
11								

Рис.3. Результаты зачёта по геометрии (в Minitab)

1	Номер теоремы	Сдали	Не сдали		Среднее арифметическое	Размах	Медиана	Мода
2	1	23	1		21,66666667	5	23	23
3	2	23	1		2,333333333	5	1	1
4	3	20	4					
5	4	23	1					
6	5	18	6					
7	6	23	1					

Рис.4. Результаты зачёта по геометрии (в Excel)

Колонки обозначались следующим образом:

- номер теоремы.
- количество сдавших учеников (далее- сдали).
- количество не сдавших учеников (далее- не сдали).
- среднее арифметическое.

-размах.

-медиана.

-мода.

Начнём с Excel. Основные команды для работы с математической статистикой:

-МОДА(число1;число2)

-СРЗНАЧА(число1;число2)

-МЕДИАНА(число1;число2)

-МАКС(число1;число2)-МИН(число1;число2)

Далее Minitab. Хочется отметить, что данная программа написана на английском, следовательно, все команды тоже на английском языке:

- MEAN (СРЗНАЧА)

- MEDIAN (МЕДИАНА)

- MODE (МОДА)

- MAXIMUM –MINIMUM (МАКС-МИН)

Начнём с функции **МОДА**. Данная команда анализирует и вычисляет предоставленный нами ряд значений по определению моды, после чего выводит результат в таблицу. В Excel это выглядит следующим образом:

1	Сдали	Не сдали	Мода
2	23	1	=МОДА(A2:A7)
3	23	1	1
4	20	4	
5	23	1	
6	18	6	
7	23	1	

Рис.5. Демонстрация функции МОДА в Excel

Теперь рассмотрим **MODE**. Стоит отметить, что Minitab не позволяет просмотреть алгоритм работы той, или иной функции. Поэтому, чтобы выбрать функцию, нужно:

1.Нажать по меню Stat.

2.Выбрать Display Descriptive Statistics.

3.Сделать двойной щелчок по данным, которые хотите анализировать.

4.Выберите статистику, которая вам необходима.

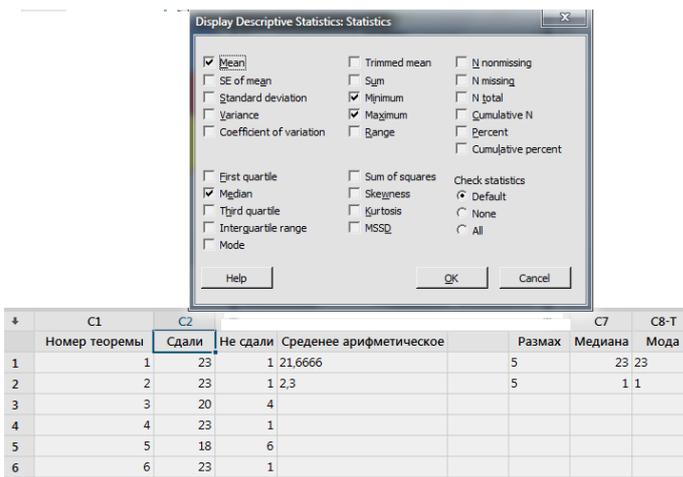


Рис.6. Выбор нужной статистики для наших данных в Minitab
5.Нажмите ОК.

Так работает выбор функции в Minitab.

Далее рассмотрим функцию **МЕДИАНА**. Данная функция также, что и **МОДА**, однако по определению медианы.

1	Сдали	Не сдали	Мода	Медиана
2	23	1		23
3	23	1		1
4	20	4		
5	23	1		
6	18	6		
7	23	1		

Рис.7. Демонстрация функции МЕДИАНА в Excel

Следующая функция **СРЗНАЧА**. Данная функция является суммой всех значений, делённая на их количество, т.е сумма всех значений колонки **СДАЛИ** равна:

$$\frac{23 + 23 + 20 + 23 + 18 + 23}{6} = 21,6$$

Рис.8. Среднее значение колонки Сдали

Сумма всех значений колонки **НЕ СДАЛИ** равна:

$$\frac{1 + 1 + 4 + 1 + 6 + 1}{6} = 2,3$$

Рис.9. Среднее значение колонки Не сдали

1	Сдали	Не сдали		Среднее арифметическое
2	23	1		=СРЗНАЧ(A2:A7)
3	23	1		2,333333333
4	20	4		
5	23	1		
6	18	6		
7	23	1		

Рис.10. Демонстрация функции СРЗНАЧ в Excel

Последняя функция - **МАКС(число1;число2)-МИН(число1;число2)**. Данное уравнение в Excel обозначает размах ряда, т.к в программе отсутствует функция **РАЗМАХ**. Уравнение работает следующим образом:

1.Сначала функция МАКС находит максимальное значение в колонке, т.е наибольшее значение СДАЛИ равно 23.

2.Затем функция МИН находит минимальное значение в колонке, т.е наименьшее значение СДАЛИ равно 18.

3.Производим вычитание значение МАКС от МИН, т.е значение колонки Размах равно 5.

1	Сдали	Не сдали	Размах	Максимальное значение	Минимальное значение
2	23	1	=МАКС(A2:A7)-МИН(A2:A7)	23	18
3	23	1			
4	20	4			
5	23	1			
6	18	6			
7	23	1			

Рис.11. Демонстрация работы уравнения МАКС-МИН

Также работает и в Minitab.

По итогу можно сказать, что,несмотря на различия программ, их результат будет неизменен.

ГЛАВА 5. СРАВНЕНИЕ EXCEL И MINITAB

Исходя из сравнений функций в Excel и Minitab, можно составить список плюсов и минусов программ.

Начнём с Excel. Преимущества:

1. Распространённость. Данная программа установлена на многих устройствах.

2. Понятный интерфейс. Не нужно обладать особыми знаниями, чтобы работать с Excel.

3. Возможность создания собственной функции. То есть, если в Excel нет какой-либо функции, можно заменить уравнение, исходя из определения математической статистики (пример с колонкой РАЗМАХ).

Недостатки:

1. Малое количество встроенных статистических функций, которые предоставляет Excel. Для работы со статистикой вам потребуются специальные платные пакеты (например, пакет STATISTICA).

Что же до Minitab, нужно рассмотреть несколько недостатков:

1. Ограниченный доступ. Программа платная, однако доступна пробная версия на 30 дней. После истечения периода, программа не запустится до приобретения лицензии.

2. Отсутствие русской локализации. К сожалению, в программе доступны несколько языков: английский, французский, немецкий, японский, корейский, китайский и испанский.

Преимущества:

1. Огромный выбор функций. Данная программа содержит все возможные формулы и определения, существующие в математической статистике.

По итогу можно сказать, что Excel- подходящая программа для тех, кому нужна простая статистика, т.е. не требующая подробных исследований.

Однако, если вы изучаете математический анализ, вам потребуется программа, подобная Minitab.

ЗАКЛЮЧЕНИЕ

В ходе данной работы нам удалось выполнить **задачи и цель**, поставленные в введении проекта:

1. Мы рассмотрели научно-исследовательские работы по данной теме.
2. Раскрыли основные понятия для глубокого анализа данного раздела.
3. Рассмотрели статистические данные в различных программах.
4. Проанализировали программы для работы с математической статистикой.
5. Сделали выводы на основе проведённых сравнений.

По итогу можно сказать, что нет разницы между использованием программы Excel и Minitab. Для каждой деятельности потребуется та, или иная программа. Стоит добавить, что на основе сравнений двух программ планируется создание собственной программы для работы с математической статистикой.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Орлов А.И. Математика случая. Вероятность и статистика – основные факты/ Вероятность и статистика нужны всем. Коротко об истории математической статистики.: Учебное пособие– Москва : МЗ-Пресс, 2004. – 170 с.
2. Простая формула, чтобы подсчитать среднее арифметическое//Skysmart, URL: <https://skysmart.ru/articles/mathematic/kak-najti-srednee-arifmeticheskoe>
3. Среднее арифметическое, размах, мода и медиана//Calsc.su, URL: <https://calcs.su/html/math/grade7/srednee-arifmeticheskoe.html>
4. Медиана в статистике//Блог Джафара Алиева, URL: <https://jsoft.ws/?content=median-statistics.html>
5. Почему для анализа лучше использовать медианные значения, а не средние//Лайфхакер, URL: <https://lifehacker.ru/mediana/>
6. Статистика объектов нечисловой прямой //Понятия и категории, URL: <http://ponjatija.ru/node/13785>
7. Математическая статистика. Ее виды, особенности, задачи.//StudFiles, URL: <https://studfile.net/preview/6199100/>