*Морозова Анжелика Владимировна,*
*Поздняков Сергей Николаевич*

# ПРАКТИЧЕСКАЯ СТАТИСТИКА С ПРОГРАММОЙ «АВТОГРАФ»

Среди ученых и методистов есть разные мнения о том, почему курс статистики не приживается в отечественной школе.

Не вдаваясь в исторический экскурс, отметим те изменения, которые могут положительно повлиять на судьбу этого курса.

Первое – появление доступных всем желающим больших необработанных массивов данных. Такие массивы данных могут быть полезны при изучении конкретного предмета, например, географии, астрономии, экономики. Кроме того, проникновение автоматизированных информационных систем в управление школой приводит к появлению массивов данных, характеризующих процесс обучения. Умение превращать необозримые массивы данных в небольшое число понятных параметров и составляет основу статистики.

Наконец, самым важным для организации работы над этой темой является то, что появилось программное обеспечение учебного назначения, с помощью которого статистическая обработка данных не требует от человека затрат времени на рутинные вычисления.

Об одном таком средстве – программе «Автограф», точнее одном её режиме – мы расскажем в этой статье. Материал представим в форме урока «Описательная статистика: мода, медиана, среднее, квартили, децили, процентили, ящичковая диаграмма, гистограмма, многоугольник частот».

## СОДЕРЖАНИЕ УРОКА

### *Беседа*

Рассмотрим пример с результатами контрольной работы классов А, Б и В. Перед тем как построить диаграммы они могли выглядеть как страницы классного журнала (рис. 1).

Если мы запишем оценки в том порядке, как они идут в журнале, они дадут нам мало возможностей для анализа. Такие наборы называется исходными:

3255445343232523433344 исходный набор оценок для класса А;

24423555523522543253552 исходный набор оценок для класса Б;

34555543234454443542543 исходный набор оценок для класса В.

Для статистического анализа качества успеваемости номера (и фамилии) учеников нам не важны, а оценки удобно расположить по возрастанию. Такая последовательность называется упорядоченным или ранжированным рядом:

22233333333344444445555 ранжированный ряд оценок для класса А;

22222223333344455555555 ранжированный ряд оценок для класса Б;

22333334444444445555555 ранжированный ряд оценок для класса В.

Разобьем каждую из этих последовательностей на две части равной длины (левую и правую). Если это не удается сделать (как в нашем случае, когда число оценок нечётное), одно число оставим посередине. Это число называется медианой. Ниже медиана выделена жирным цветом:

22233333333 **3** 44444445555 медиана оценок контрольной для класса А;

22222223333 **3** 44455555555 медиана оценок контрольной для класса Б;

22333334444 **4** 44445555555 медиана оценок контрольной для класса В.

Предположим, что в каждом классе добавился один отличник, как изменятся медианы? Новые ряды оценок:

22233333333 **34** 44444455555 оценки для класса А+;

22222223333 **34** 44555555555 оценки для класса Б+;

22333334444 **44** 44455555555 оценки для класса В+.

В случае чётного числа оценок в качестве медианы берётся среднее между двумя средними значениями. В нашем случае это:

**3,5** – медиана оценок контрольной для класса А с добавленным отличником,

**3,5** – медиана оценок контрольной для класса Б с добавленным отличником,

**4** – медиана оценок контрольной для класса В с добавленным отличником.

Медиана характеризует оценку среднего человека (это в точности совпадает со смыслом названия, если число человек нечётное), в отличие от среднего значения оценки. Наиболее наглядно это видно на следующем примере.

*Пример*

Если из 10 человек один получает по сто тысяч в месяц, а остальные получают одинаковую и не столь большую зарплату, например, десять тысяч, то медиана зарплат будет равна десяти тысячам, а средняя зарплата – девятнадцати тысячам.

**Определение.** *Медианой* числового набора называется число, для которого половина элементов набора будет не больше его, а остальные не меньше.

| № | Фамилия | |
|---|---|---|
| 1 | Xxxxxxxxxxxx | 3 |
| 2 | Xxxxxxxxxxxx | 2 |
| 3 | Xxxxxxxxxxxx | 5 |
| 4 | Xxxxxxxxxxxx | 5 |
| 5 | Xxxxxxxxxxxx | 4 |
| 6 | Xxxxxxxxxxxx | 4 |
| 7 | Xxxxxxxxxxxx | 5 |
| 8 | Xxxxxxxxxxxx | 3 |
| 9 | Xxxxxxxxxxxx | 4 |
| 10 | Xxxxxxxxxxxx | 3 |
| 11 | Xxxxxxxxxxxx | 2 |
| 12 | Xxxxxxxxxxxx | 3 |
| 13 | Xxxxxxxxxxxx | 5 |
| 14 | Xxxxxxxxxxxx | 2 |
| 15 | Xxxxxxxxxxxx | 3 |
| 16 | Xxxxxxxxxxxx | 4 |
| 17 | Xxxxxxxxxxxx | 3 |
| 18 | Xxxxxxxxxxxx | 4 |
| 19 | Xxxxxxxxxxxx | 3 |
| 20 | Xxxxxxxxxxxx | 3 |
| 21 | Xxxxxxxxxxxx | 3 |
| 22 | Xxxxxxxxxxxx | 4 |
| 23 | Xxxxxxxxxxxx | 4 |

| № | Фамилия | |
|---|---|---|
| 1 | Xxxxxxxxxxxx | 2 |
| 2 | Xxxxxxxxxxxx | 4 |
| 3 | Xxxxxxxxxxxx | 4 |
| 4 | Xxxxxxxxxxxx | 2 |
| 5 | Xxxxxxxxxxxx | 3 |
| 6 | Xxxxxxxxxxxx | 3 |
| 7 | Xxxxxxxxxxxx | 5 |
| 8 | Xxxxxxxxxxxx | 5 |
| 9 | Xxxxxxxxxxxx | 5 |
| 10 | Xxxxxxxxxxxx | 2 |
| 11 | Xxxxxxxxxxxx | 3 |
| 12 | Xxxxxxxxxxxx | 5 |
| 13 | Xxxxxxxxxxxx | 2 |
| 14 | Xxxxxxxxxxxx | 2 |
| 15 | Xxxxxxxxxxxx | 5 |
| 16 | Xxxxxxxxxxxx | 4 |
| 17 | Xxxxxxxxxxxx | 3 |
| 18 | Xxxxxxxxxxxx | 2 |
| 19 | Xxxxxxxxxxxx | 5 |
| 20 | Xxxxxxxxxxxx | 3 |
| 21 | Xxxxxxxxxxxx | 5 |
| 22 | Xxxxxxxxxxxx | 5 |
| 23 | Xxxxxxxxxxxx | 2 |

| № | Фамилия | |
|---|---|---|
| 1 | Xxxxxxxxxxxx | 3 |
| 2 | Xxxxxxxxxxxx | 4 |
| 3 | Xxxxxxxxxxxx | 5 |
| 4 | Xxxxxxxxxxxx | 5 |
| 5 | Xxxxxxxxxxxx | 5 |
| 6 | Xxxxxxxxxxxx | 5 |
| 7 | Xxxxxxxxxxxx | 4 |
| 8 | Xxxxxxxxxxxx | 3 |
| 9 | Xxxxxxxxxxxx | 2 |
| 10 | Xxxxxxxxxxxx | 3 |
| 11 | Xxxxxxxxxxxx | 4 |
| 12 | Xxxxxxxxxxxx | 4 |
| 13 | Xxxxxxxxxxxx | 5 |
| 14 | Xxxxxxxxxxxx | 4 |
| 15 | Xxxxxxxxxxxx | 4 |
| 16 | Xxxxxxxxxxxx | 4 |
| 17 | Xxxxxxxxxxxx | 3 |
| 18 | Xxxxxxxxxxxx | 5 |
| 19 | Xxxxxxxxxxxx | 4 |
| 20 | Xxxxxxxxxxxx | 2 |
| 21 | Xxxxxxxxxxxx | 5 |
| 22 | Xxxxxxxxxxxx | 4 |
| 23 | Xxxxxxxxxxxx | 3 |

**Рис. 1**

В случае если таким свойством обладает множество чисел, берут среднее из них.

*Пример*. Для набора д22233333333 **34** 44444455555 указанным свойством обладают все числа из промежутка (3; 4). Средним из них будет 3,5.

Разобьём теперь упорядоченные по возрастанию оценки на четыре равные по количеству группы. Разделяющие их оценки называются квартилями:

22233 **3** 33333 **3** 44444 **4** 45555 квартили оценок контрольной для класса А;

22222 **2** 23333 **3** 44455 **5** 55555 квартили оценок контрольной для класса Б;

22333 **3** 34444 **4** 44445 **5** 55555 квартили оценок контрольной для класса В.

Первый (левый) квартиль называется нижним квартилем, второй квартиль совпадает с медианой, третий квартиль называется верхним.

**Определение.** Аналогично определению медианы *нижним квартилем* называется число, для которого четверть элементов будет не больше его, а остальные три четверти – не меньше.

В случае если таким свойством обладает множество чисел, берут среднее из них.

***Упражнение***. Дайте самостоятельно определение верхнему квартилю.

**Задание.** Определите по аналогии понятия децилей (деление упорядоченного набора на 10 равных частей) и процентилей (деление упорядоченного набора на 100 равных частей).

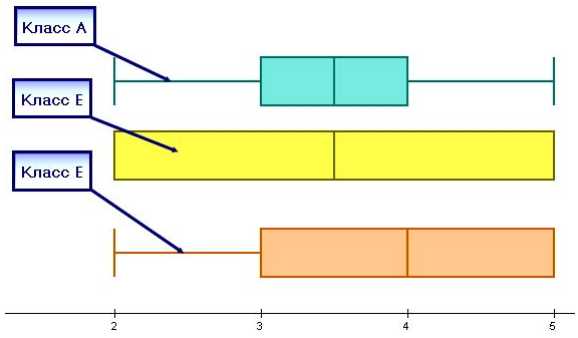



**Рис. 2**

*Замечание.* В общем случае квартили, децили, процентили и прочие деления упорядоченного набора на равные части называются квантилями.

*Пример.* Для класса с добавленным отличником квартили определяются так:

22233 **3** 33333 **34** 44444 **4** 55555 оценки для класса А+;

22222 **2** 23333 **34** 44555 **5** 55555 оценки для класса Б+;

22333 **3** 34444 **44** 44455 **5** 55555 оценки для класса В+.

| | Нижний квартиль | Медиана | Верхний квартиль |
|---|---|---|---|
| Класс А+ | **3** | **3,5** | **4** |
| Класс Б+ | **2** | **3,5** | **5** |
| Класс В+ | **3** | **4** | **5** |

Эти данные часто изображаются в виде так называемой *ящичковой диаграммы* (рис. 2), в которой:

1) левая и правая границы показывают наименьшую и наибольшую оценки, длина диаграммы называется диапазоном оценки (в нашем примере диапазон равен 3: от 2 до 5);

2) границами «ящичка» являются нижний и верхний квартили;

3) медиана обозначена поперечной чертой на «ящичке».

Таким образом, внутрь ящичка попадает ровно половина всех оценок. Положение ящичка показывает характер расположения оценок «средних» учеников.

Ящичковая диаграмма позволяет быстро увидеть особенности каждого класса: в классе А ученики примерно одного уровня, в классе Б, наоборот, ученики очень различаются по уровню, в классе В ученики имеют больший средний уровень по сравнению с первыми двумя классами, а по разбросу они занимают промежуточное значение между классами А и Б.

Кроме ранжирования оценок их можно представить в виде так называемого *вариационного ряда* – таблицы, в которой каждой оценке сопоставляется частота её появления в наборе:

| | Класс А | Класс Б | Класс В |
|---|---|---|---|
| Оценка «2» | 3 | 7 | 2 |
| Оценка «3» | 9 | 5 | 5 |
| Оценка «4» | 7 | 3 | 9 |
| Оценка «5» | 4 | 8 | 7 |

По таблице можно определить ещё одну характеристику набора данных – *моду*, то есть самую «популярную» оценку. Для класса А это тройка, для класса Б – пятёрка, а для класса В – четвёрка.

**Определение.** *Модой* набора чисел называется наиболее часто встречающееся число набора. Если таких чисел два, набор называется бимодальным, три – тримодальным и т. д.

*Замечание.* Познакомьтесь со статьей по биологии, описывающей различные стратегии поведения ящериц, и связанное с ними тримодальное распределение их численности (http://www.nature.ok.ru/models/rock_paper_scissors.htm).

Вариационный ряд обычно представляется графически в виде *точечной диаграммы*, когда каждому элементу исходного ряда сопоставляется точка. Вместо такого представления часто используется близкое к нему изображение в виде *гистограммы*, в которой результаты изображаются столбиками шириной 1 и высотой, равной частоте данного числа в наборе. Иногда вместо гистограммы строится *многоугольник частот*, когда верхние точки соседних столбиков соединяются отрезками прямых. Так, например, результаты контрольной для класса А изображены на рис. 3 всеми тремя способами.

Для сравнительного анализа иногда одну из гистограмм «переворачивают», то есть зеркально отражают относительно оси абсцисс. Например, на рис. 4 приведены результаты сравнения гистограмм результатов контрольной для классов Б и В.

Во всех рассмотренных случаях мы имели дело с *дискретными наборами* данных, однако можно представить себе наблюдения величин, которые меняются *непрерывно*, например, замеры температуры, давления, веса или роста учеников. В этом случае диапазон изменения величины разбивается на равные промежутки (их количество обычно мало отличается от 10) и измерения, попавшие в один промежуток, рассматриваются как равные.

Например мы может предположить, что оценки учеников, рассматриваемые выше, на самом деле были более точными, например 2,3; 3,1; 4,8; 5,7, ..., но мы учитывали только целую часть оценки, отбрасывая дробную часть. В этом случае

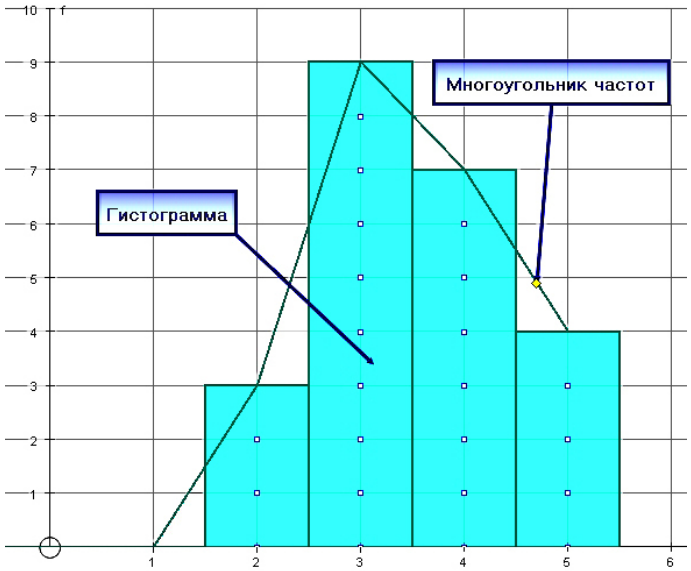



**Рис. 3**

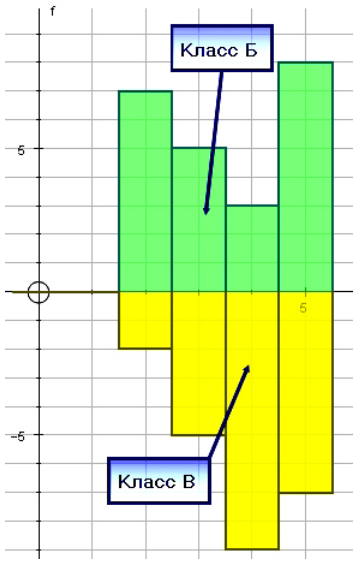



**Рис. 4**

гистограмма будет изображена по-другому – сдвинется на 0,5 единиц вправо, а площадь любой части фигуры между вертикальными прямыми $x=a$ и $x=b$ будет показывать количество оценок, попавших в промежуток [a; b) (рис. 5).

*Замечание.* В нашем примере, если $a$ и $b$ не являются целыми, эти значения будут приближенными, в предположении, что все оценки внутри каждого промежутка между последовательными целыми числами распределены равномерно.

Изменится и расположение медианы и квартилей (а значит, и ящичковая диаграмма будет иной). Теперь медиана $Me$ будет располагаться так, чтобы прямая $x = Me$ делила площадь на две равные части, соответственно прямые, проведенные через квартили разделят площадь на четыре равные части (рис. 6).

**Задачи**

1. Найдите точное положение медианы и квартилей для распределения оценок класса А в предположении, что они были получены округлением действительных чисел.

2. Постройте ящичковые диаграммы для распределений оценок в классах Б и В в предположении, что они получены из «непрерывно-распределенных» оценок отбрасыванием цифр после запятой.

## ИСПОЛЬЗОВАНИЕ ПРОГРАММЫ «АВТОГРАФ» ДЛЯ ПОДДЕРЖКИ КУРСА ПРАКТИЧЕСКОЙ СТАТИСТИКИ

Программа АвтоГраф является русификацией одноименной программы, созданной в Великобритании. Программу можно рассматривать как справочник, составленный из динамических моделей и инструментов для поддержки курса математики как в средней школе, так и в колледже. В программе есть три разных режима, которые представляют разные разделы математики: режим 1D – для изучения статистики и теории вероятностей, режим 2D – для изучения основ математического анализа и планиметрии, режим 3D – для изучения стереометрии и аналитической геометрии. В этой статье мы коснемся лишь первого из этих режимов.

### *Режим 1D*

Работа в этом режиме начинается с выбора листа, на котором будут изображаться в графической форме изучаемые объекты (рис. 7). При изучении статистики основными объектами являются наборы чисел. Поэтому следующим шагом является ввод данных (рис. 8).

Последовательность оценок учащихся, описанная в начале урока, называется исходными данными. Далее программа сама определит частоты оценок: если в меню

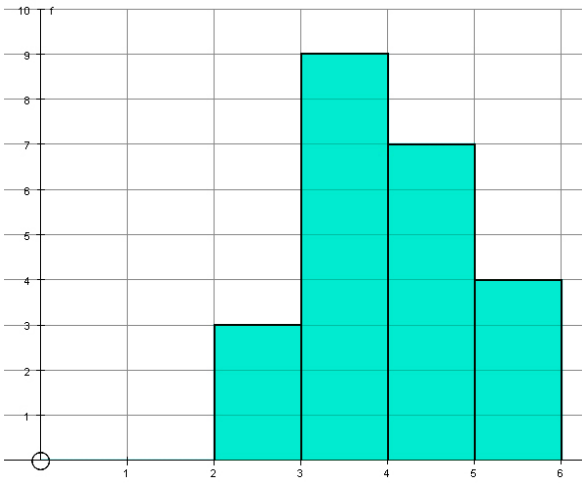


**Рис. 5**

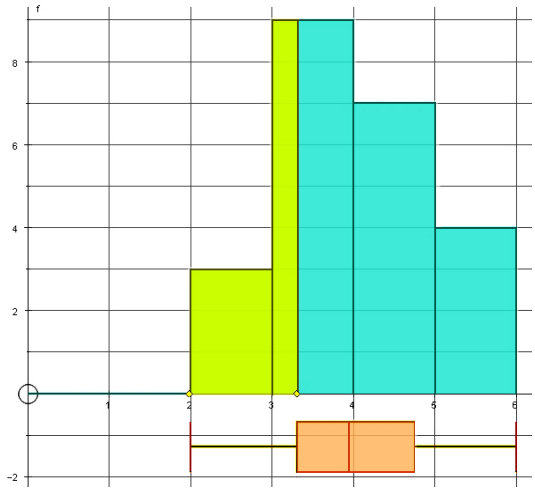


**Рис. 6**

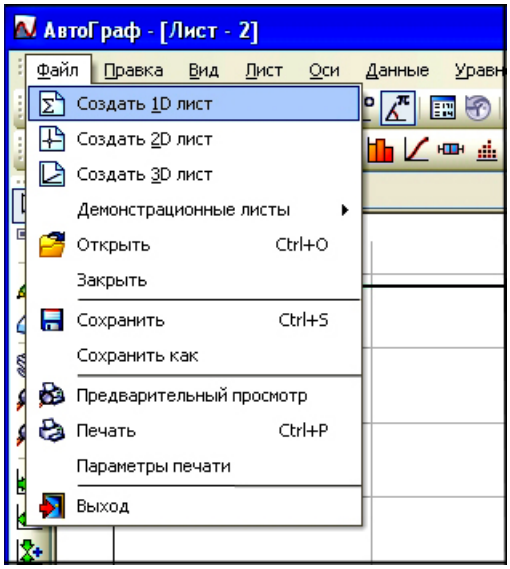



**Рис. 7**

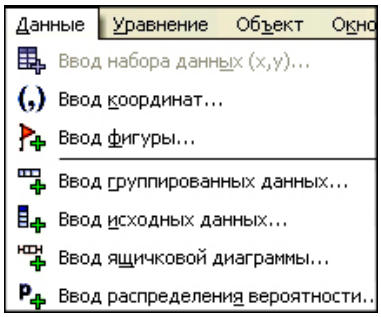


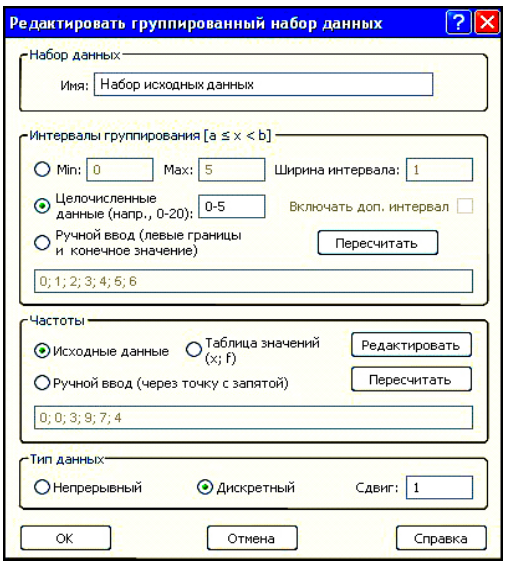

**Рис. 8**

выбрать опцию «ввод группированных данных», а в открывшемся окне – вариант «исходные данные», то появится окно для ввода (в левой части рис. 9 *б*). Данные можно ввести вручную или с помощью копирования и вставки через буфер (Ctrl C – копирование, Ctrl V – вставка), если они были представлены столбцом в текстовом редакторе.

Исходные данные при вводе будут группированы так, чтобы в дальнейшем можно было работать не с самими данными, а с частотами их появления в наборе. Заметим, что если частоты данных были уже подсчитаны, то данные можно вводить парами, указывая для каждой оценки частоту её появления в наборе (на рис. 9 *а* видно, что этому соответствует опция, которая называется таблицей значений ($x$; $f$)).

Все введенные наборы получают автоматически имя «Набор исходных данных №», которое можно заменить на произвольное имя. Для того чтобы выбрать набор для дальнейшей работы (для построения диаграмм, вычисления средних и пр.), нужно выбрать его имя в нижней части экрана (рис. 10), при этом слева от имени появляется крестик в кружочке (закрашенный для набора сгруппированных данных,

*а)*




*б)*

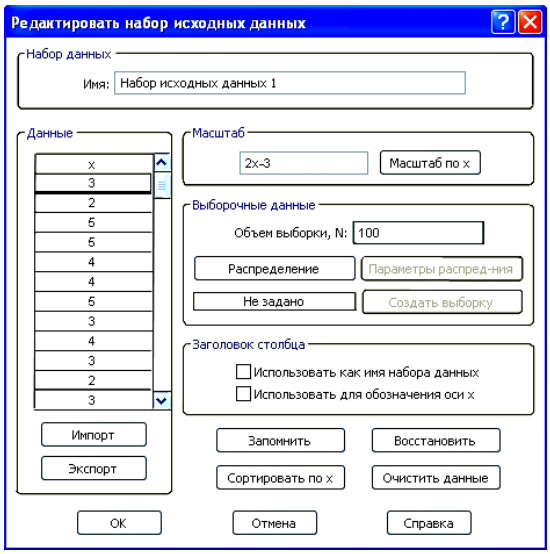



**Рис. 9**

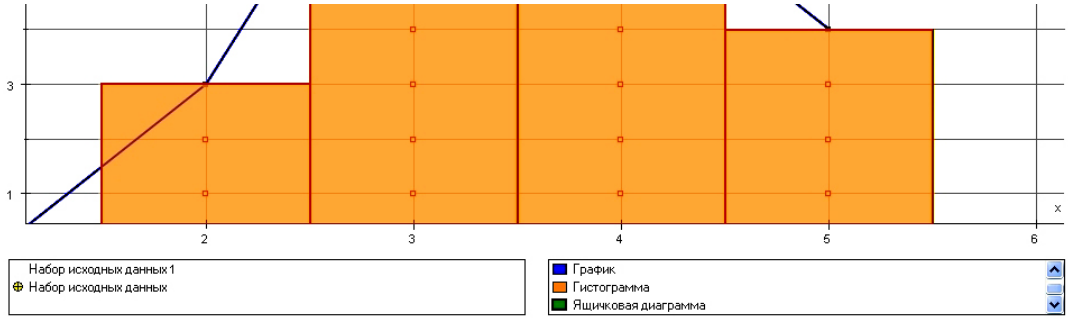



**Рис. 10**

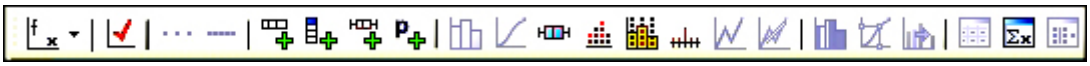


**Рис. 11**

и незакрашенный – если данные не были сгруппированы).

После выбора набора с ним можно производить операции, которые собраны в меню «Объект», а также вынесены на панель в виде пиктограмм (рис. 11, 12).

Наиболее важными для нашего урока являются следующие операции: построение гистограммы, построение многоугольника частот (графика), построение ящичковой диаграммы, вывод статистических результатов (рис. 12).

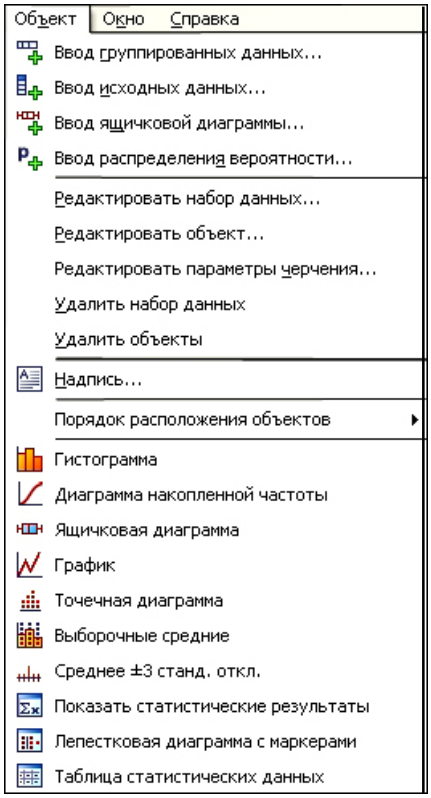



**Рис. 12**

### *Экспериментально-исследовательская работа*

В заключение урока полезно провести экспериментально-статистическую работу, данные для которой можно сгенерировать с помощью программы АвтоГраф.

Например, возьмем в качестве исходного набора результаты 100-кратного подбрасывания монеты (1 – орёл, 0 – решка) или по-другому – одновременного подбрасывания 100 монет.

Собрать такие данные с помощью натурного эксперимента весьма затруднительно. Используем возможности моделирования случайных величин, имеющиеся в программе АвтоГраф. Для этого при вводе исходных данных нужно выбрать опцию «Распределение» (тематика данного урока не предполагает изучение понятия распределения, поэтому дальнейшие действия надо рассматривать просто как запуск генератора случайных чисел) и из предложенных вариантов выбрать дискретное распределение. Далее в параметрах распределения нужно указать наибольшее и наименьшее целые числа, которые будут генерироваться (для моделирования монетки нужно взять два последовательных целых числа, например 0 и 1, а для

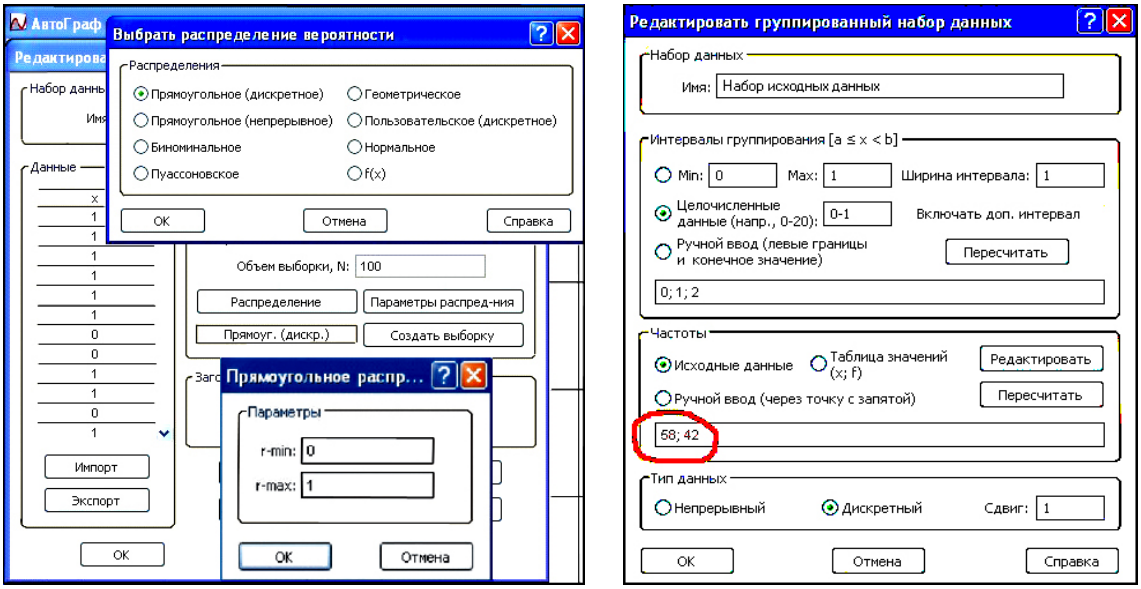



**Рис. 13**

моделирования, например, игральной кости можно было бы взять 1 и 6) и указать количество бросаний монеты (объём выборки). Мы сделаем 100 бросаний.

После группировки данных мы увидим, сколько у нас выпало орлов и сколько решек (на рис. 13 обведены выпавшие числа орлов и решек: 58 и 42). Теперь будем повторять эксперимент, создавая новые выборки и записывая результаты. Интуитивно ясно, что число орлов будет примерно равно числу решек. Однако также очевидно, что выпадение одинакового числа орлов и решек маловероятно. Представляет интерес исследовать статистическую закономерность выпадения числа орлов при бросании, например, 100 монет. Однако собрать такие данные трудно даже при автоматическом «выбрасывании» монет.

Здесь нас опять выручит генератор данных. Описанный выше процесс многократного бросания набора монет описывается биномиальным распределением.

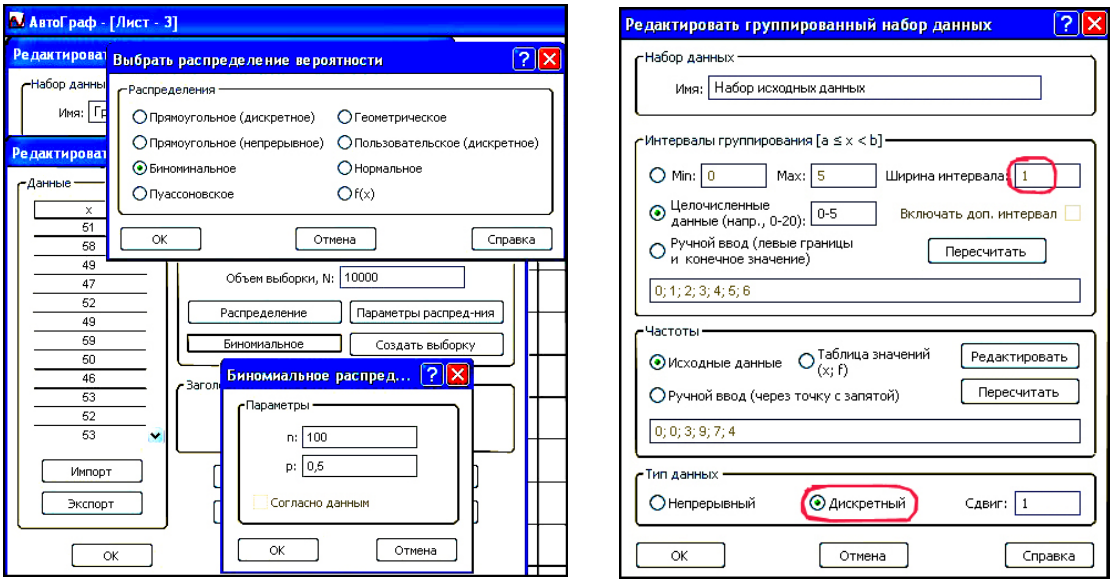



**Рис. 14**

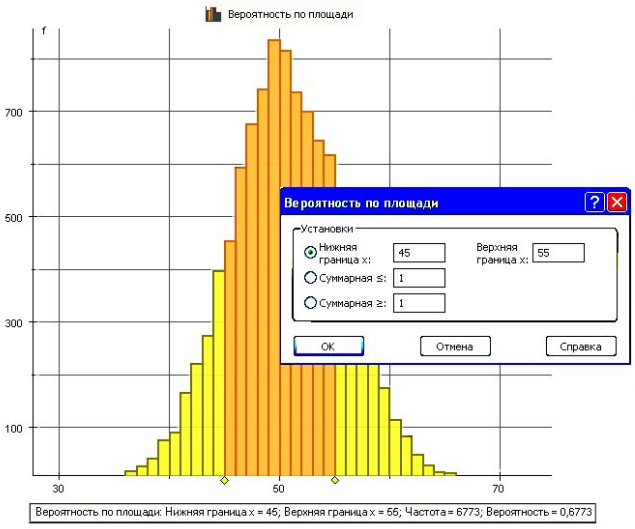



**Рис. 15**

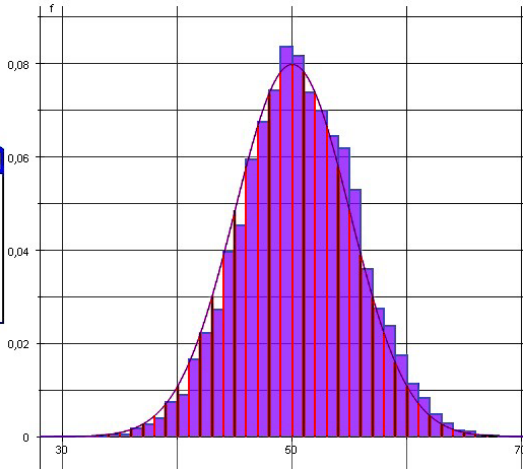



**Рис. 16**

Выбрав это распределение, зададим в качестве параметров $n = 100$ (бросаем 100 монет), $p = 0{,}5$ (берём «правильные» монеты, то есть вероятность выпадения орла равна 1/2, или иначе, орёл и решка равновероятны), объём выборки возьмём 10000. Это означает, что эксперимент по выбрасыванию 100 монет будет повторен 10000 раз (!). В качестве типа данных возьмём дискретный, а для того чтобы точно учесть выпадающие числа, ширину интервала, в котором не различаем числа (то есть берём приближённые значения), считаем 1 (рис. 14).

На изображенной на рис. 15 гистограмме показан результат 10000 бросаний 100 «правильных» монет (выделена область гистограммы, для которой число выпавших орлов не меньше 45, но меньше 55). Таких результатов оказалось 68%. Для построения такой области нужно воспользоваться опцией «вероятность по площади», которая доступна при выделенной гистограмме.

Однако статистическая обработка на этом не заканчивается, и можно разобрать следующий шаг – приблизить построенную гистограмму, изображаемую графиком ступенчатой функции графиком непрерывной функции. Оказывается это можно сделать с помощью функции $e^{-x^2}$. Представляет интерес подбор параметров для функции $Ae^{-\left(\frac{x-m}{\sigma}\right)^2}$ так, чтобы её график как можно лучше описывал гистограмму, а точнее, чтобы площадь подграфика функции как можно лучше приближала площадь гистограммы (рис. 16).

Такое распределение называется нормальным, а его обсуждение заслуживает отдельного урока.



***Морозова Анжелика Владимировна,***
***ассистент кафедры ВМ-2***
***СПбГЭТУ «ЛЭТИ»,***

***Поздняков Сергей Николаевич,***
***доктор педагогических наук,***
***профессор кафедры ВМ-2***
***СПбГЭТУ «ЛЭТИ».***