



Е.И. Стенина

# МЕТОД ГРУППИРОВАНИЯ ДАННЫХ

Екатеринбург  
2015

МИНОБРНАУКИ РОССИИ

ФГБОУ ВПО «УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ ЛЕСОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра инновационных технологий и оборудования деревообработки

Е.И. Стенина

## **МЕТОД ГРУППИРОВАНИЯ ДАННЫХ**

Методические указания  
по выполнению практической работы по дисциплине  
«Методы и средства научных исследований»  
студентами очной формы обучения по направлению  
250400.62 «Технология лесозаготовительных  
и деревообрабатывающих производств»

Екатеринбург  
2015

Печатается по рекомендации методической комиссии ИЛБиДС.  
Протокол № 2 от 9 октября 2014 года.

Рецензент – доцент кафедры ИТОД Тютиков С.С.

Редактор Е.Л. Михайлова  
Оператор компьютерной верстки Е.А. Газеева

---

Подписано в печать 02.03.15		Поз. 20
Плоская печать	Формат 60×84 1/16	Тираж 10 экз.
Заказ №	Печ. л. 0,93	Цена руб. коп.

---

Редакционно-издательский отдел УГЛТУ  
Отдел оперативной полиграфии УГЛТУ

## ВВЕДЕНИЕ

Любое экспериментальное исследование условно можно разделить на три этапа: подготовка эксперимента, планирование и постановка опытов, обработка результатов измерений и их анализ. Множество значений результатов экспериментов (случайных величин), полученных в продублированных опытах, представляет собой статистическую совокупность, которая иногда может содержать десятки и даже сотни наблюдений. Обычная статистическая обработка такого количества результатов становится крайне трудоемким процессом. В этом случае целесообразно применять метод группирования данных.

## 1. ЦЕЛЬ РАБОТЫ

Методические указания предназначены для приобретения студентами практических навыков в статистической обработке результатов эксперимента, содержащих значительное количество значений.

## 2. ПОНЯТИЕ ЭКСПЕРИМЕНТА

Эксперимент – это совокупность опытов, позволяющая установить влияние воздействующих факторов  $x_i$  на выходные параметры объекта исследования  $y_i$  (рис. 1).

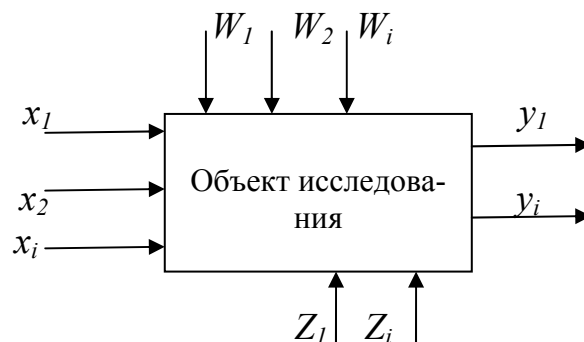


Рис. 1. Схема эксперимента

Фактор – это измеряемая переменная величина, принимающая в некоторый момент времени определенное значение (температура, давление, количество циклов и т.п.).

Постоянными называются факторы, не меняющие своего значения в пределах всего эксперимента ( $W_i, Z_i$ ).

Переменным (варьируемым) называется фактор  $x_i$ , значение которого меняется от опыта к опыту. Каждое значение, принимаемое фактором в опыте, называется уровнем переменного фактора. Диапазон изменения (варьирования) переменных факторов ограничен верхним и нижним уровнями.

Выходным параметром называется результат эксперимента  $y_i$ , который является случайной величиной, так как всегда в большей или меньшей степени содержит ошибки, обусловленные погрешностью приборов, измерений, расчетов и т.п.

Опыт – часть эксперимента, выполненная при определенных значениях одного или нескольких факторов. С целью снижения вероятности ошибки при анализе результатов эксперимента необходимо дублирование каждого опыта.

Математическая статистика – это наука о математических методах обработки, систематизации и использовании результатов наблюдений для научных и практических выводов.

Выборочной статистической совокупностью (или выборкой) называется совокупность, в которой содержится только некоторая часть значений случайной величины  $y_i$ , полученных по результатам экспериментов.

Число значений выходной величины, содержащихся в выборке, называют объемом выборки  $n$ .

### 3. ОБЩИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ РАБОТЫ

При обработке результатов эксперимента (выборки) методом группирования данных необходимо:

- определить наименьшее и наибольшее значения выходной величины (п. 4.1);
- весь диапазон значений случайной величины в выборке необходимо разбить на интервалы (п. 4.2);
- определить границы и середины интервалов (п. 4.3);
- подсчитать количество результатов эксперимента, попадающее в каждый интервал (п. 4.4);
- рассчитать относительную частоту попадания случайной величины в каждый интервал (п. 4.5);
- рассчитать основные статистические показатели выборки (п. 4.6);
- построить гистограмму и кривую распределения случайной величины в выборке (п. 4.7).

Расчеты должны выполняться и оформляться в соответствии с требованиями ЕСКД.

### 4. ПРИМЕР ВЫПОЛНЕНИЯ РАБОТЫ

В результате эксперимента по пропитке кленовых образцов были получены следующие значения общего поглощения  $y_i$ , кг/м<sup>3</sup>:

40,2	50,0	57,7	48,3	49,5	42,9	51,6	53,4	47,7	49,0	46,6	55,8
34,1	20,4	16,7	22,8	23,8	24,0	26,7	21,3	23,6	30,4	27,4	23,0
46,1	22,5	54,2	48,7	56,4	54,0	50,8	60,7	59,4	51,2	55,5	19,2
21,0	35,4	33,2	36,9	40,4	30,5	34,7	35,5	39,1	32,2	33,1	56,0
50,6	54,4	58,9	60,6	51,8	50,2	56,3	58,3	54,2	58,4	55,7	57,1

## 4.1 Определение наименьшего и наибольшего значений выходной величины

Наименьшее и наибольшее значения выходной величины можно определить визуально либо с помощью пакета программ Microsoft Excel 2010.

$$y_{\min} = 16,7 \qquad y_{\max} = 60,7$$

Для того чтобы выполнить эту задачу с помощью пакета программ Microsoft Excel 2010, нужно создать документ в соответствующем формате. В столбец А вносим все значения случайной величины обрабатываемой выборки, выделяем их. Выбираем закладку «Формулы» → «Автосумма» → «Максимум» («Минимум»). В конце столбца появятся соответствующие значения (рис. 2, 3).

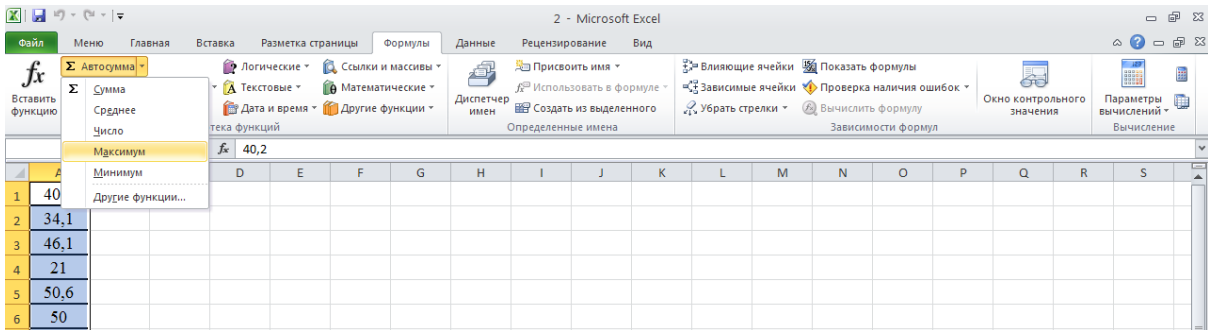


Рис. 2

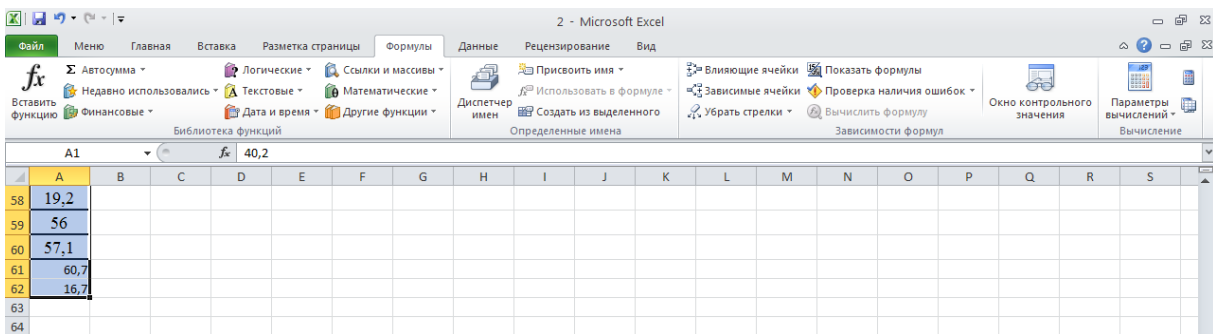


Рис. 3

## 4.2 Расчет интервалов

### 4.2.1 Расчет числа интервалов

Необходимое число интервалов  $k$  определяется по формуле:

$$k = 1 + 3,2 \lg n, \tag{1}$$

где  $n$  – объем выборки.

$$n = 60,$$

$$k = 1 + 3,2 \lg 60 = 6,69.$$

Полученное значение числа интервалов округляется в сторону большего числа. Таким образом,  $k = 7$ .

В случае выполнения расчетов с помощью пакета программ Microsoft Excel 2010 нужно выделить ячейку, выбрать в закладке «Формулы» ячейку «Математические», а в ее выпадающем списке – «log10». В появившемся окне «Аргументы функции» ввести число, равное объему выборки, и нажать на кнопку «ОК» (рис. 4, 5).

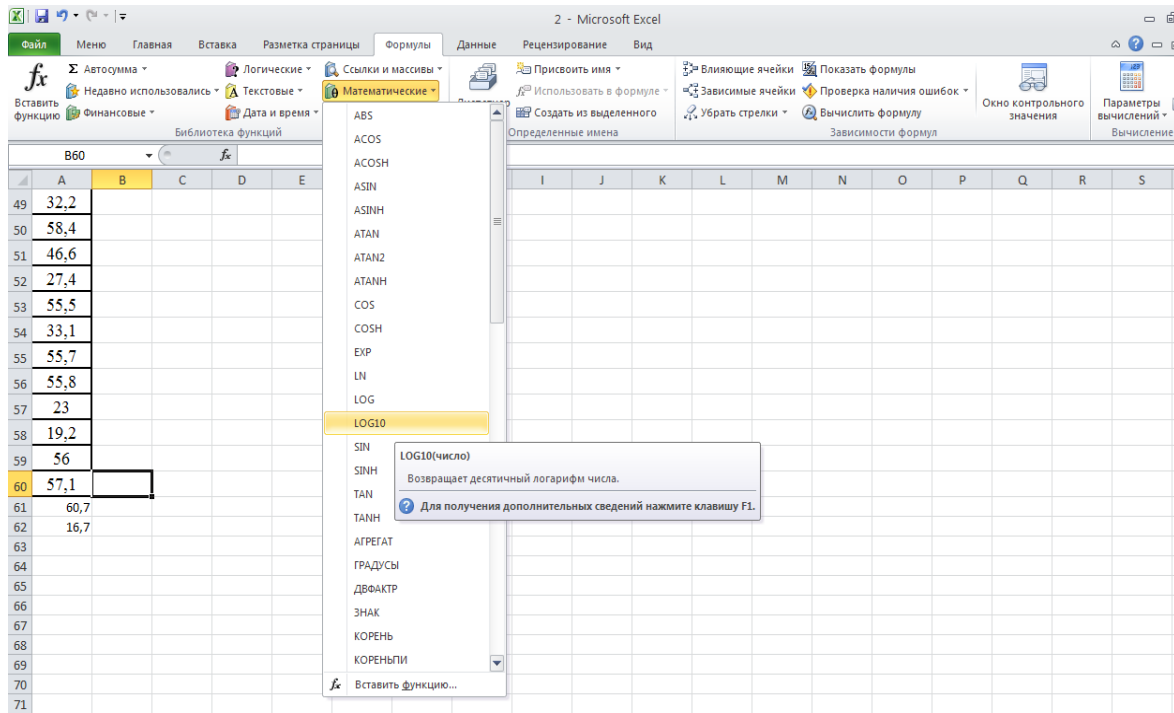


Рис. 4

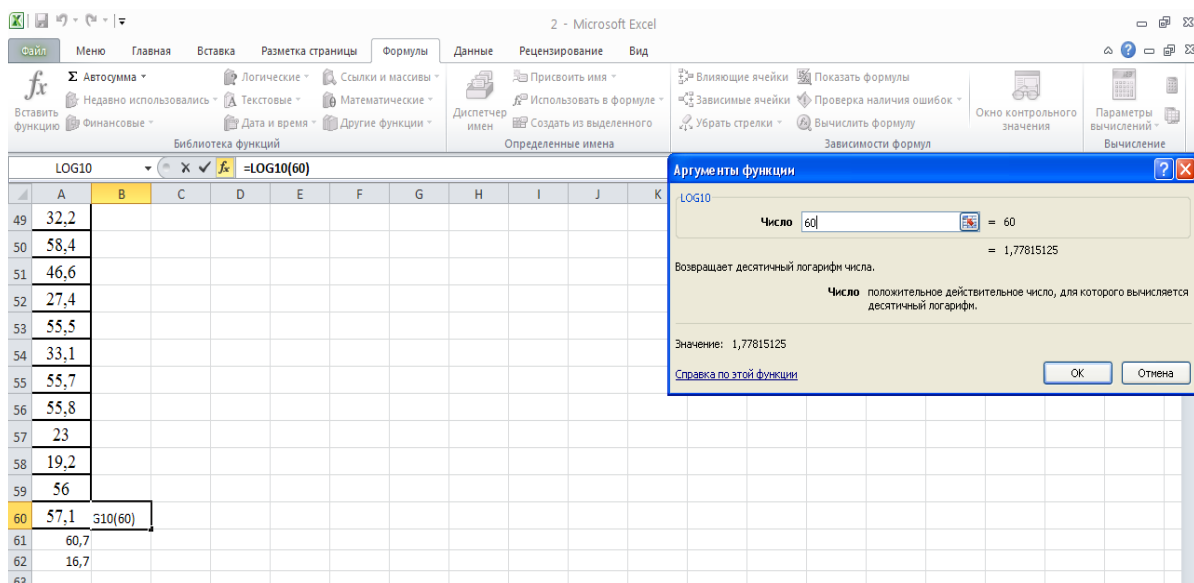


Рис. 5

В выделенной ячейке появится значение десятичного логарифма от числа 60. Выделить соседнюю ячейку, в строке редактора формул ввести формулу (1) и выделить соседнюю ячейку (рис. 6, 7).

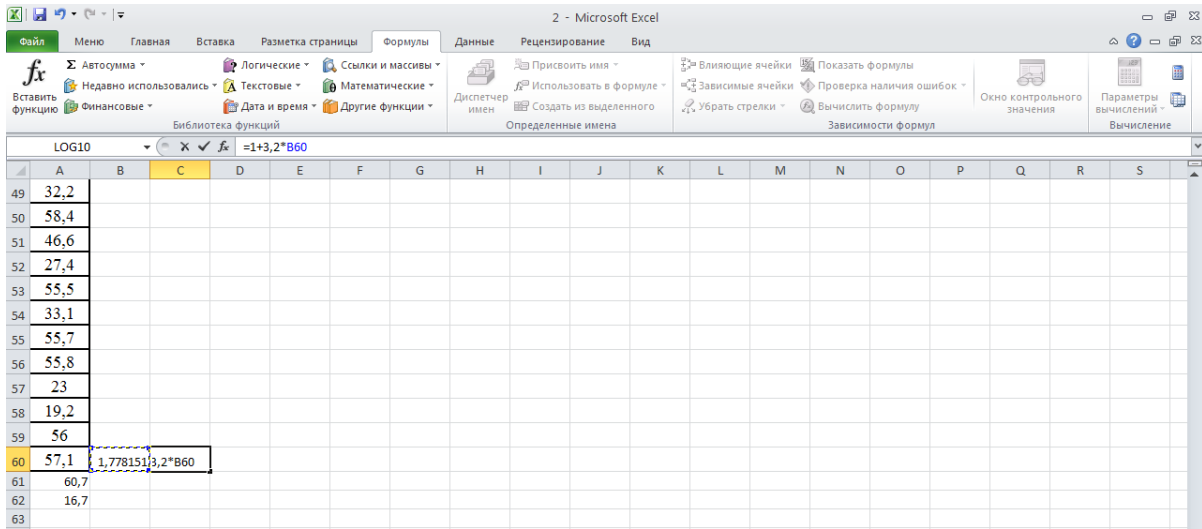


Рис. 6

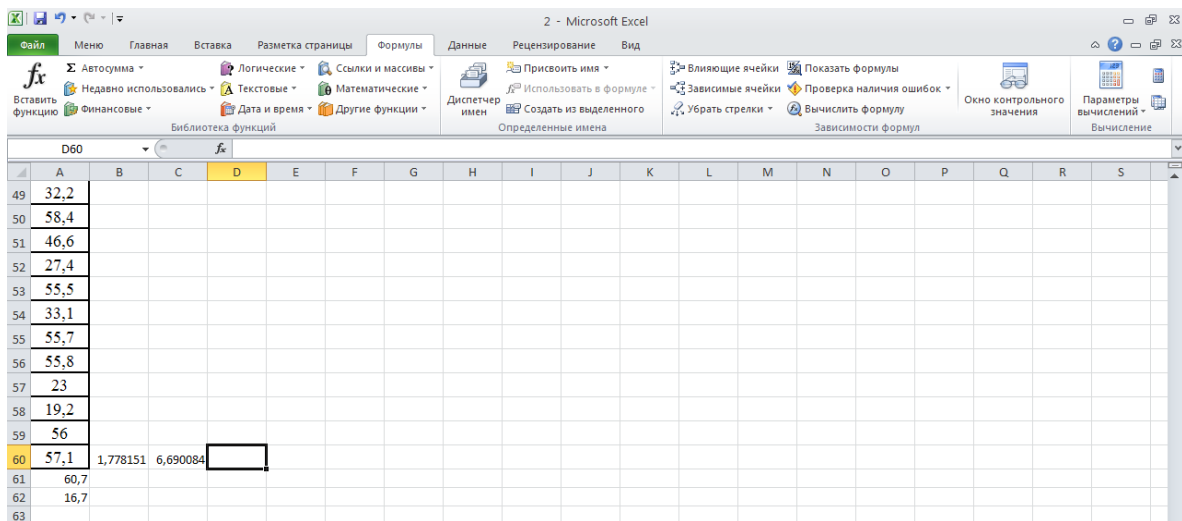


Рис. 7

## 4.2.2 Расчет длины интервала

Чаще всего используют интервалы равной длины. В этом случае длина каждого интервала  $h$  рассчитывается по формуле

$$h = \frac{y_{\min} - y_{\max}}{k}, \quad (2)$$

$$h = \frac{60,7 - 16,7}{7} = 6,29.$$



Данный расчет можно выполнить также с помощью программного пакета Excel. Выделяем ячейку, в строке редактора формул забиваем соответствующую формулу (рис. 8,9).

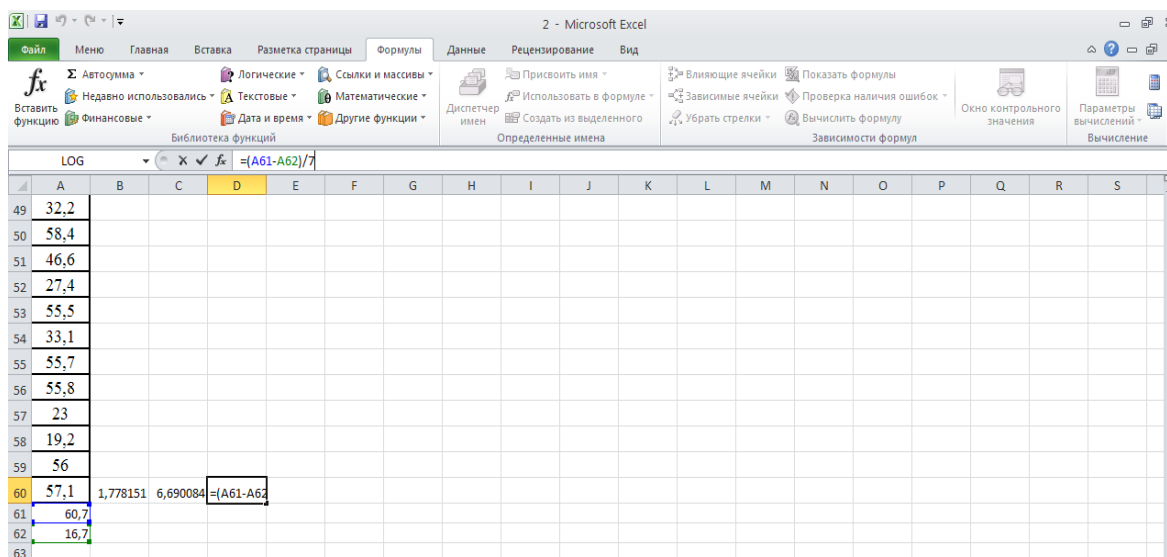


Рис. 8

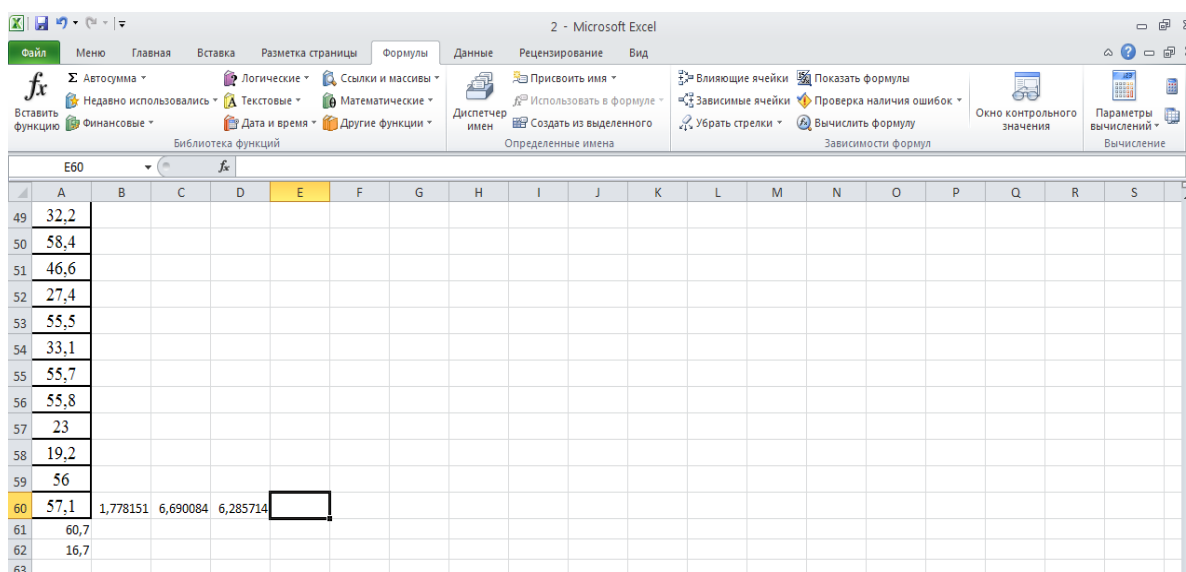


Рис. 9

### 4.3 Определение границ и середин интервалов

Первый интервал будет лежать в пределах  $y_{min} \dots y_1$ , где

$$y_1 = y_{min} + h, \quad (3)$$

$y_1 = 16,7 + 6,29 = 22,99$  и так далее.

Для каждого интервала определяют его середину  $y_i^*$  по формуле

$$y_i^* = \frac{y_{i-1} + y_i}{2}, \quad (4)$$

где  $y_{i-1}$  – нижняя граница интервала;

$y_i$  – верхняя граница интервала.

Так, для первого интервала середина будет равна:

$$y_1^* = \frac{16,7 + 22,99}{2} = 19,85.$$

Для систематизации расчетов их можно свести в таблицу.

Сводная таблица расчетов

Номер интервала	Границы интервала	Середина интервала $y_i^*$	Число значений $y_i$ , попадающих в интервал $m_i$	Относительная частота $p_i$
1	16,7...22,99	19,85	7	0,12
2	22,99...29,28	26,14	6	0,10
3	29,28...35,57	32,43	9	0,15
4	35,57...41,86	38,72	4	0,07
5	41,86...48,15	45,01	4	0,07
6	48,15...54,44	51,30	16	0,27
7	54,44...60,7	57,57	14	0,22

#### 4.4 Расчет числа результатов эксперимента, попадающих в каждый интервал

Расчет числа результатов эксперимента, попадающих в каждый интервал, можно проводить визуально либо используя программный пакет Excel. В столбце А выделить все значения выборки. В закладке «Меню» нажать на кнопку «Фильтр», в ячейке А1 нажать на появившуюся кнопку (рис. 10), затем в выпадающем окне выделить «Числовые фильтры» → «между» (рис. 11). В выпадающем окне «Пользовательский автофильтр» задать границы первого интервала (рис. 12) и нажать кнопку «ОК». На листе останутся только значения числового ряда, попадающие в указанный интервал (рис. 13). Нажать на панели инструментов на кнопку «Отменить», и все значения выборки восстановятся. Подобные манипуляции повторяются для каждого интервала.

Предварительно необходимо принять решение о том, к какому интервалу всегда будут относить значение случайной величины, попадающее на границу интервала.

Сумма всех результатов эксперимента, попадающих в каждый интервал, должна быть равна объему выборки.

$$n = \sum_{i=1}^k m_i, \quad (5)$$

где  $m_i$  – число результатов эксперимента, попадающих в  $i$ -й интервал.  
 $7+6+9+4+4+16+14 = 60$  (см. таблицу).

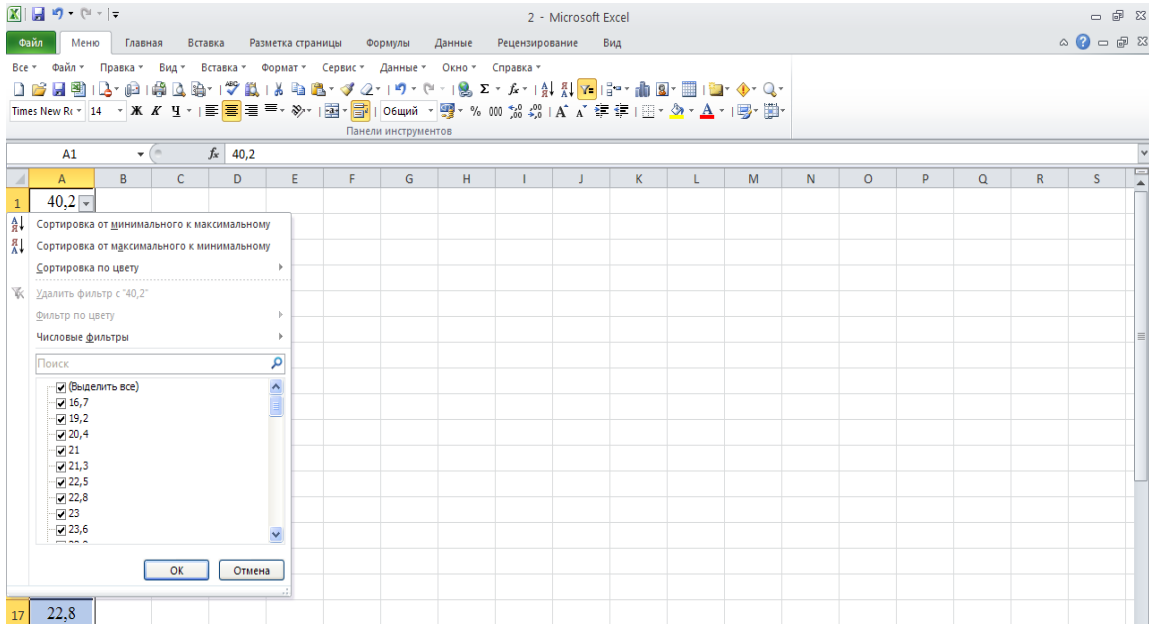


Рис. 10

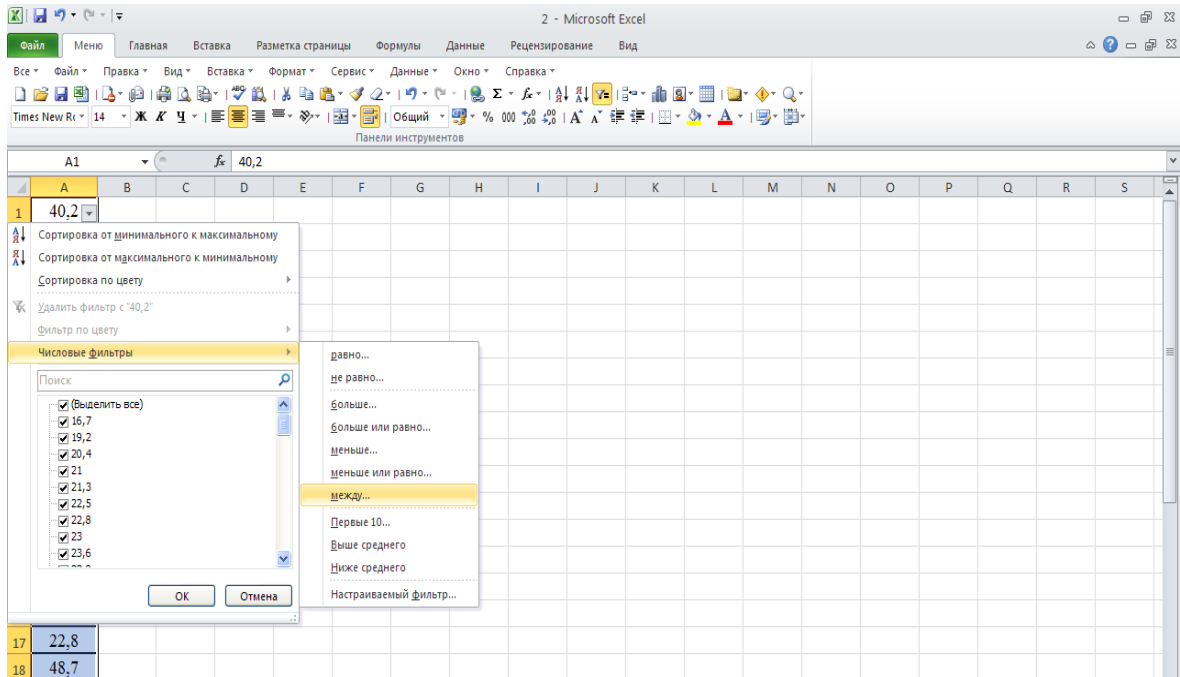


Рис. 11

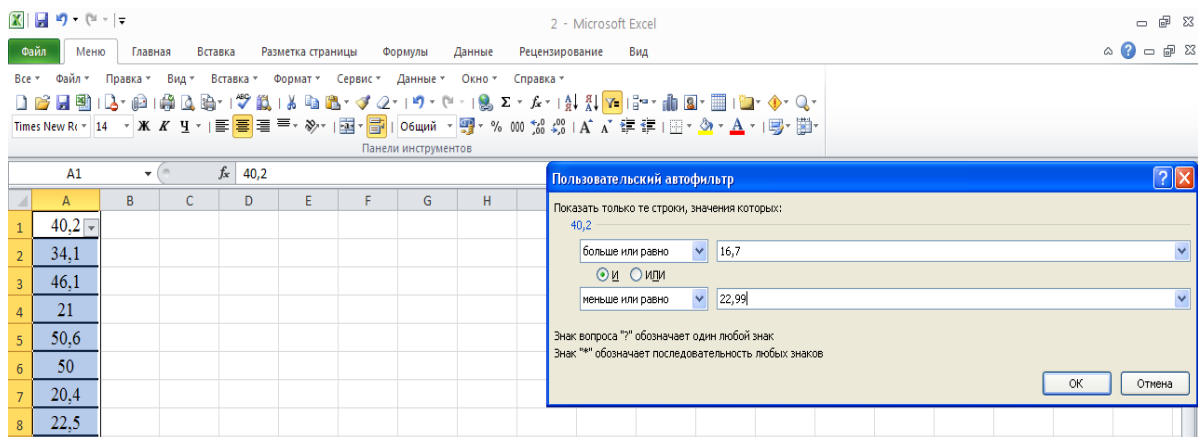


Рис. 12

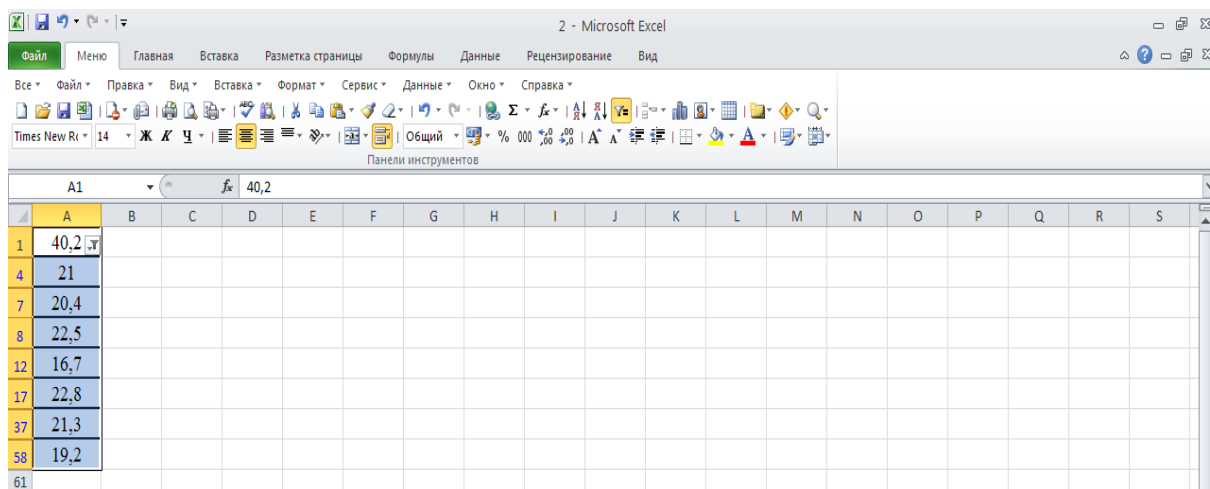


Рис. 13

## 4.5 Расчет относительной частоты попадания случайной величины в каждый интервал

Относительная частота (вероятность) попадания случайной величины в каждый интервал рассчитывается по формуле

$$p_i = \frac{m_i}{n}, \quad (6)$$

$$\sum_{i=1}^k p_i^* = 1. \quad (7)$$

Так, для первого интервала относительная частота равна:

$$p_i = \frac{7}{60} = 0,12.$$

#### 4.6 Расчет основных статистических показателей выборки

К основным статистическим показателям относятся объем выборки  $n$ , выборочное среднее  $\bar{y}$  и выборочная дисперсия  $S^2$ , которые определяются по формулам (8), (9):

$$\bar{y} = \frac{\sum_{i=1}^k y_i \cdot m_i}{n}, \quad (8)$$

$$S^2 = \frac{\sum_{i=1}^k m_i (y_i^* - \bar{y})^2}{n - 1}. \quad (9)$$

Так, для приведенной выше выборки, используя данные таблицы, можно рассчитать:

$$\bar{y} = \frac{19,85 \cdot 7 + 26,14 \cdot 6 + 32,43 \cdot 9 + 38,72 \cdot 4 + 45,01 \cdot 4 + 51,30 \cdot 16 + 57,59 \cdot 14}{60} = 42,49.$$

$$S^2 = \frac{7 \cdot (19,85 - 42,49)^2 + 6 \cdot (26,14 - 42,49)^2 + 9 \cdot (32,43 - 42,49)^2 + 4 \cdot (38,72 - 42,49)^2 + 4 \cdot (45,01 - 42,49)^2 + 16 \cdot (51,30 - 42,49)^2 + 14 \cdot (57,59 - 42,49)^2}{60 - 1}.$$

$$S^2 = 2030,42.$$

#### 4.7 Графическое представление статистического ряда

График, построенный по данным статистического ряда, называется **гистограммой**. Для его построения по оси абсцисс откладывают значения границ интервалов и на каждом из них, как на основании, строят прямоугольник, площадь которого должна быть равна относительной частоте, соответствующей данному интервалу. Следовательно, высота каждого прямоугольника равна  $p_i^*/h$ .

Гистограмму можно построить в пакете программ Microsoft Word, например в версии 2010. Для этого в закладке «Вставка» выбрать иллюстрацию «Диаграмма». В появившейся верхней линейке нажать кнопку «Тип диаграммы» и выбрать «Гистограмма» (рис. 14). В таблице

базы данных внести изменения в соответствии с таблицей. Оставляя курсив наведенным на график, нажать на правую кнопку мышки. В появившемся окне выбрать команду «Параметры диаграммы» (рис. 15).

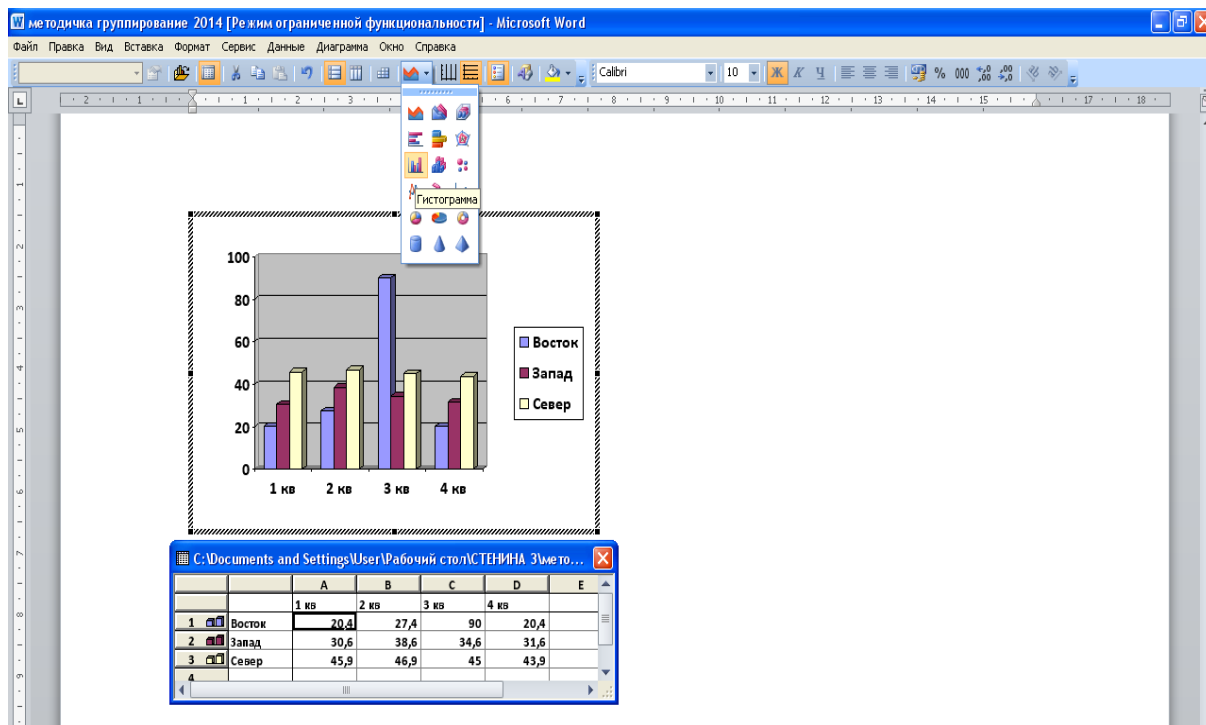


Рис. 14

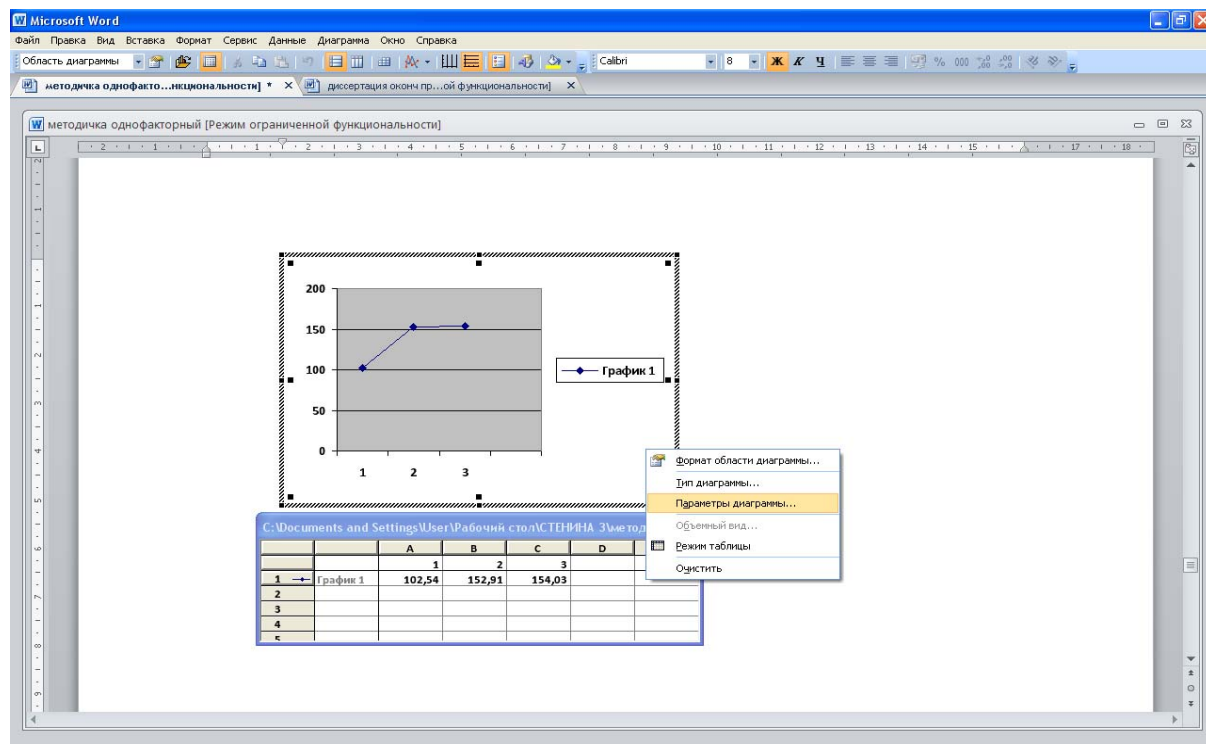


Рис. 15

Используя кнопки окна «Параметры диаграммы», отредактировать график (рис. 16). Оформить подрисуночную подпись (рис. 17).

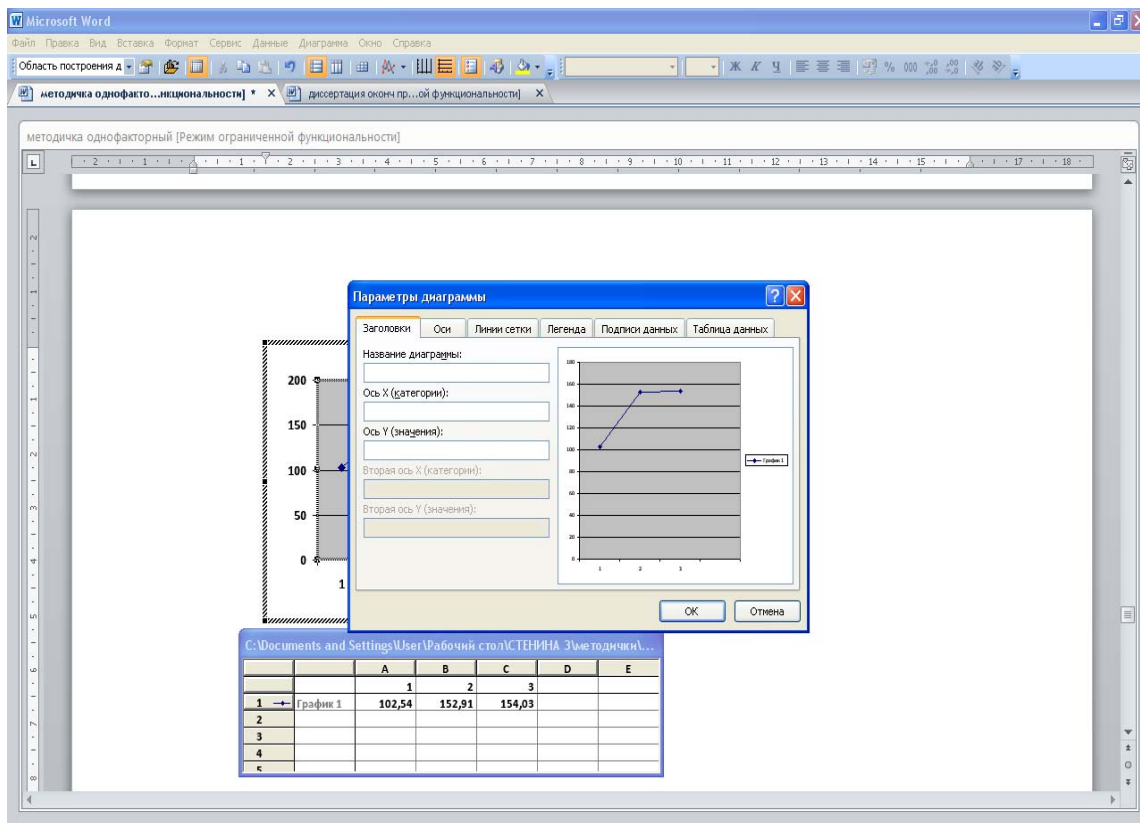


Рис. 16

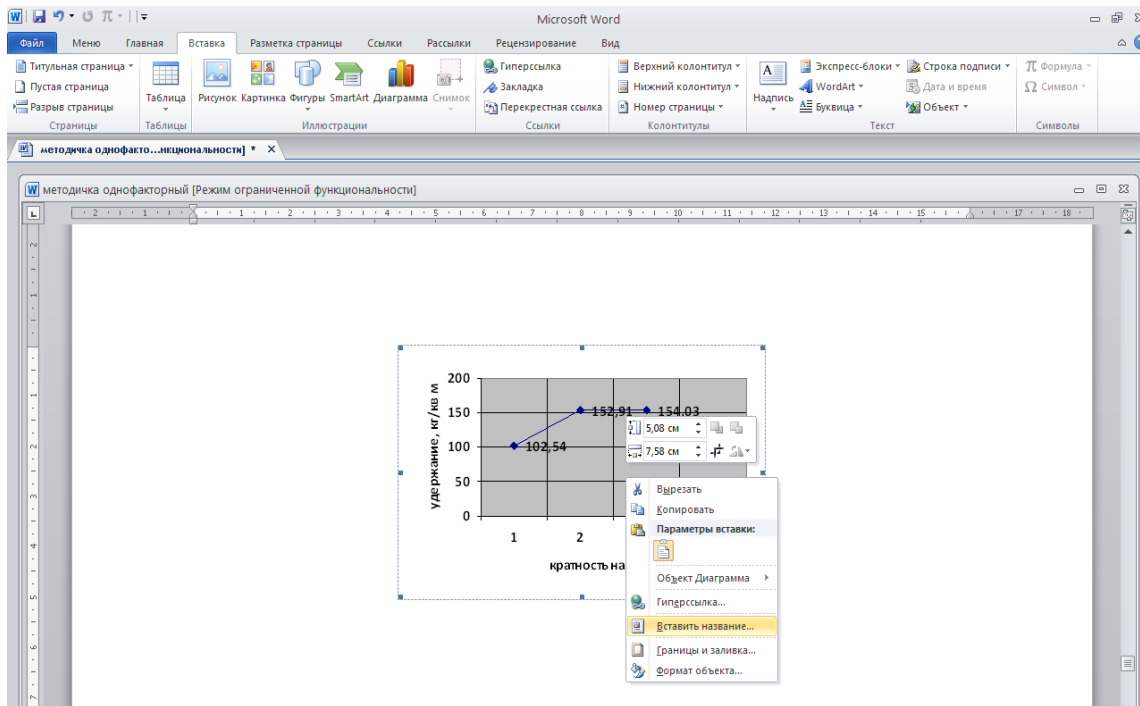


Рис. 17

Кривая распределения представлена на рис. 18.

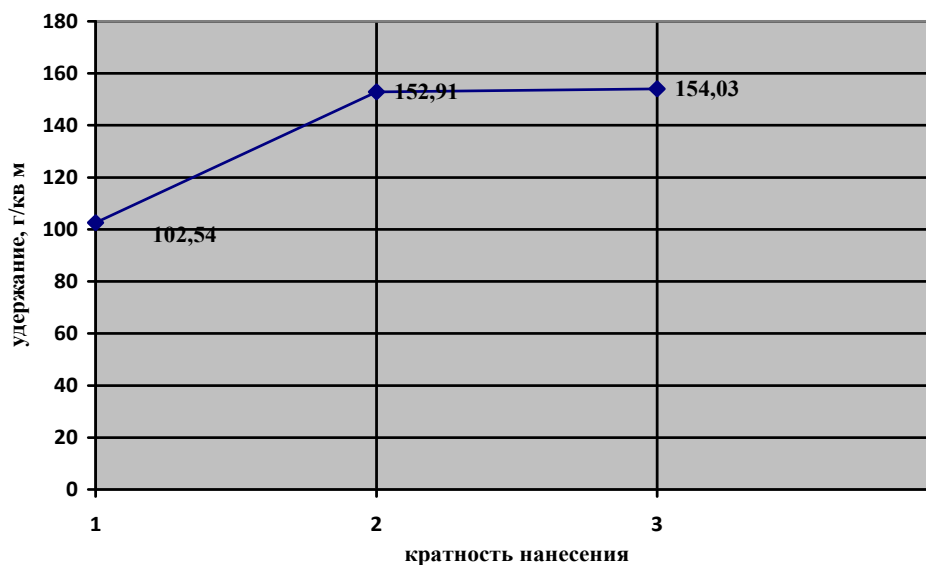


Рис. 18

### РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Пижурин, А.А. Основы научных исследований в деревообработке [Текст]: учебник для студ. вузов, обучающихся по дневн. и заоч. форме спец. 250403 (260200) «Технология деревообработки» и 150405 (170400) «Машины и оборудование лесного комплекса» / А.А. Пижурин, А.А. Пижурин; Моск. гос. ун-т леса. - М.: МГУЛ, 2005. – 305 с.

2. Пижурин, А.А. Научные исследования в деревообработке. Основы научных исследований [Текст]: текст лекций для студ. спец. 2602.00 и 1704.00 / А.А. Пижурин; [ред. Е. Г. Петрова]; Моск. гос. ун-т леса. - М.: МГУЛ, 1999. – 103 с.