

# Проведение выборочного эксперимента с использованием OpenOffice.org Calc

## Оглавление

|                                                                               |    |
|-------------------------------------------------------------------------------|----|
| Проведение выборочного эксперимента с использованием OpenOffice.org Calc..... | 1  |
| Способы случайного отбора в выборку.....                                      | 1  |
| Табличное представление данных.....                                           | 2  |
| Шаг 1. Импорт данных.....                                                     | 2  |
| Шаг 2. Разбиение экспериментальных данных на интервалы группировки.....       | 10 |
| Шаг 3. Определение границ интервалов группировки.....                         | 11 |
| Шаг 4. Вычисление срединных значений интервалов.....                          | 15 |
| Шаг 5. Распределение вариантов выборки по интервалам группировки.....         | 16 |
| Графическое представление данных.....                                         | 19 |
| Гистограмма.....                                                              | 19 |
| Полигон частот.....                                                           | 21 |
| Полигон накопленных частот.....                                               | 22 |
| Определение числовых характеристик выборки.....                               | 24 |
| Среднее арифметическое (математическое ожидание).....                         | 24 |
| Мода.....                                                                     | 25 |
| Медиана.....                                                                  | 25 |
| Дисперсия.....                                                                | 26 |
| Стандартное отклонение.....                                                   | 27 |

Выборочное наблюдение применяется в тех случаях, когда использование сплошного наблюдения невозможно либо из-за большого массива данных (к примеру, при изучении пассажиропотоков, рыночных цен, семейных бюджетов), либо из-за экономической нецелесообразности (оценка качества товаров, связанная с их уничтожением, например, дегустация, испытание кирпичей на прочность и т.п.).

Статистические единицы, отобранные для наблюдения, составляют выборочную совокупность или выборку, а весь их массив - генеральную совокупность (ГС).

Качество результатов выборочного наблюдения зависит от репрезентативности выборки, то есть от того, насколько она представительна в ГС. Для обеспечения репрезентативности выборки необходимо соблюдать принцип случайности отбора единиц, который предполагает, что на включение единицы ГС в выборку не может повлиять какой-либо иной фактор, кроме случая.

## Способы случайного отбора в выборку

1. Собственно случайный отбор или «метод лото», когда статистическим величинам присваиваются порядковые номера, заносимые на определенные предметы (например, бочки), которые затем перемешиваются в некоторой емкости (например, в мешке) и выбираются наугад. На практике этот способ осуществляют с помощью генератора случайных чисел или математических таблиц случайных чисел.

2. Механический отбор, согласно которому отбирается каждая (N/n)-я величина генеральной совокупности. Например, если она содержит 100 000 величин, а требуется выбрать 1 000, то в выборку попадет каждая  $100\,000 / 1000 = 100$ -я величина. Причем, если они не ранжированы, то первая выбирается наугад из первой сотни, а номера других будут на сотню больше. Например, если первой оказалась единица № 19, то следующей должна быть № 119, затем № 219, затем № 319 и т.д. Если единицы генеральной совокупности ранжированы, то

первой выбирается № 50, затем № 150, затем № 250 и так далее.

3. Отбор величин из неоднородного массива данных ведется стратифицированным (расслоенным) способом, когда генеральная совокупность предварительно разбивается на однородные группы, к которым применяется случайный или механический отбор.

4. Особый способ составления выборки представляет собой серийный отбор, при котором случайно или механически выбирают не отдельные величины, а их серии (последовательности с какого-то номера по какой-то подряд), внутри которых ведут сплошное наблюдение.

Качество выборочных наблюдений зависит и от типа выборки: **повторная** или **бесповторная**:

1. При повторном отборе попавшие в выборку статистические величины или их серии после использования возвращаются в генеральную совокупность, имея шанс попасть в новую выборку. При этом у всех величин генеральной совокупности одинаковая вероятность включения в выборку.

2. Бесповторный отбор означает, что попавшие в выборку статистические величины или их серии после использования не возвращаются в генеральную совокупность, а потому для остальных величин последней повышается вероятность попадания в следующую выборку.

Бесповторный отбор дает более точные результаты, поэтому применяется чаще. Но есть ситуации, когда его применять нельзя или невозможно (изучение пассажиропотоков, потребительского спроса и т.п.), и тогда ведется повторный отбор.

## Табличное представление данных

Необходимо провести исследование аудитории некоторого издания (журнала, газеты). Естественно, все регионы по экономическим соображениям подключать к исследованию не будем. Поэтому будет исследоваться не вся генеральная совокупность, а выборочная.

Итак, предположим, что Вы провели анкетирование покупателей. В итоге, получаем следующие экспериментальные данные (возраст анкетированного):

16; 17; 50; 30; 31; 40; 43; 44; 30; 17; 18; 55; 65; 24; 27; 43; 44; 44; 17; 55; 34; 45; 42; 49; 48; 62; 61; 23; 25; 29; 32; 36; 37; 32; 39; 50; 58; 52; 41; 31; 32; 47; 43; 29; 51; 66; 70; 21; 22; 53; 24; 27; 19; 19; 18; 43; 44; 32; 21; 24; 25; 17; 20

### Шаг 1. Импорт данных

Чтобы не заносить эти данные вручную в нашу таблицу, скопируйте их и вставьте в Блокнот (Пуск — Программы — Стандартные — Блокнот, Рисунок 3).

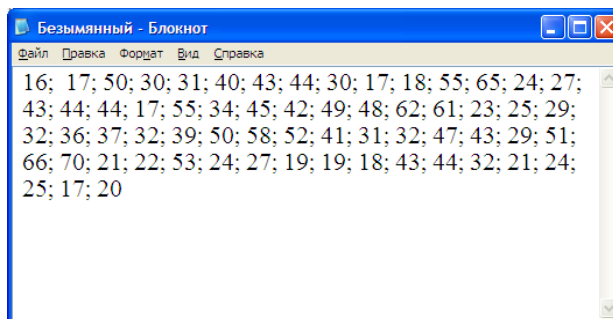


Рисунок 1.

Сохраните файл с именем **data.csv** (Рисунок 2). Расширение **csv** означает «данные, разделенные запятыми» (*comma separated values*).

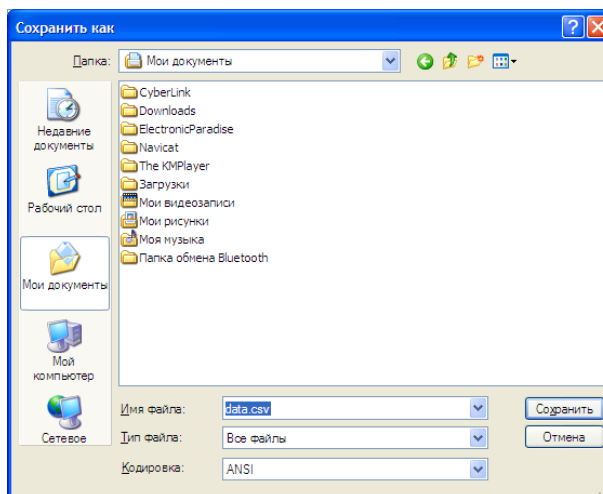


Рисунок 2. Диалоговое окно для сохранения файла

Теперь запускаем программу OOo Calc (Пуск — Программы — OpenOffice.org — OOo Calc), после чего выполним команду **Файл — Открыть** (Рисунок 3).

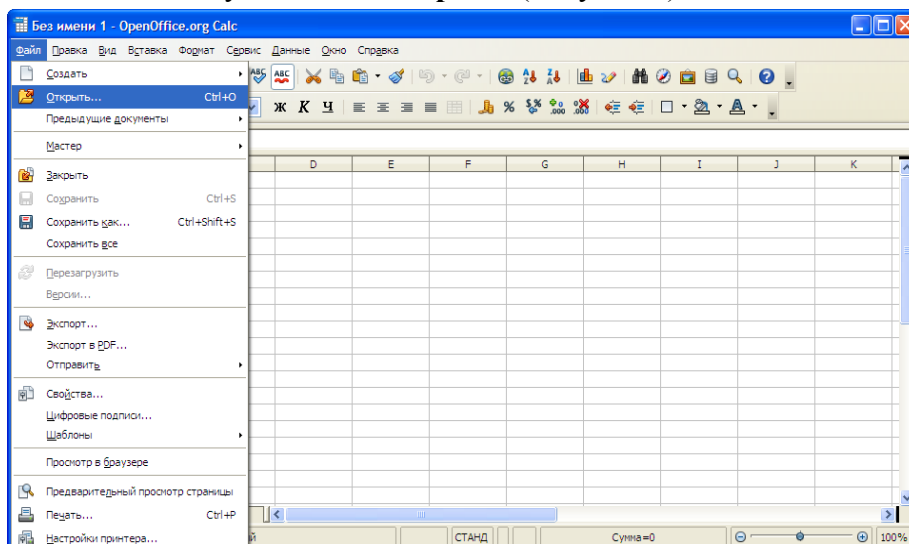


Рисунок 3.

В появившемся диалоговом окне необходимо выбрать файл **data.csv** и нажать на кнопку «**Открыть**» (Рисунок 4), после чего появится диалоговое окно (Рисунок 5).

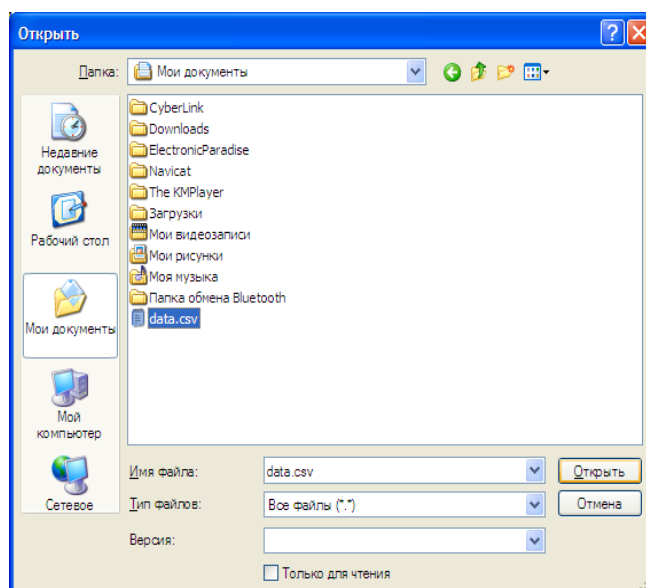


Рисунок 4. Диалоговое окно для открытия файла

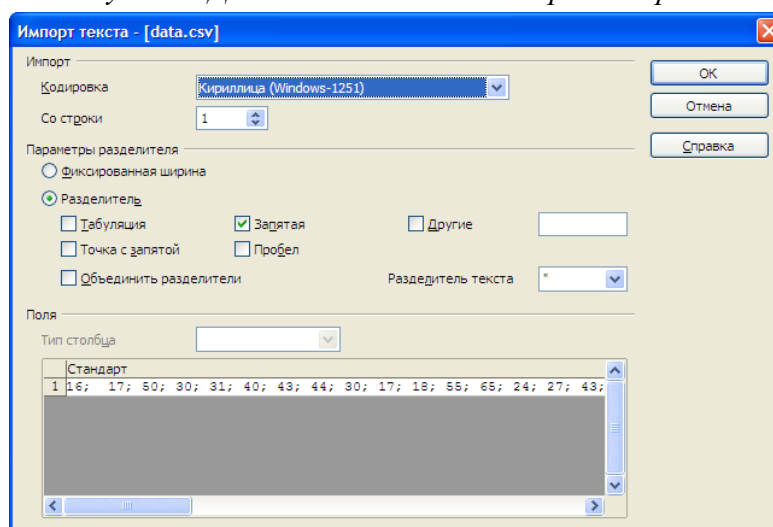


Рисунок 5. Диалоговое окно "Импорт текста"

В данном окне необходимо указать разделителя текста в файле для автоматической генерации столбцов электронной таблицы (в нашем случае это будет **Точка с запятой**, Рисунок 6). Возможен предварительный просмотр разметки импортированных данных в нижней части диалогового окна (Рисунок 6). После установки параметров нажмите кнопку **ОК**.

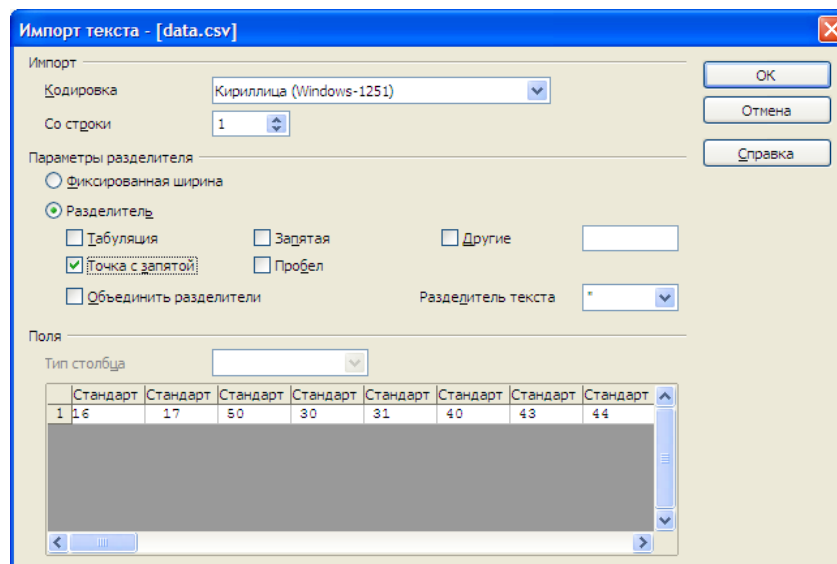


Рисунок 6. Диалоговое окно "Импорт текста"

Окно OOo Calc примет вид, как показано на Рисунок 7.

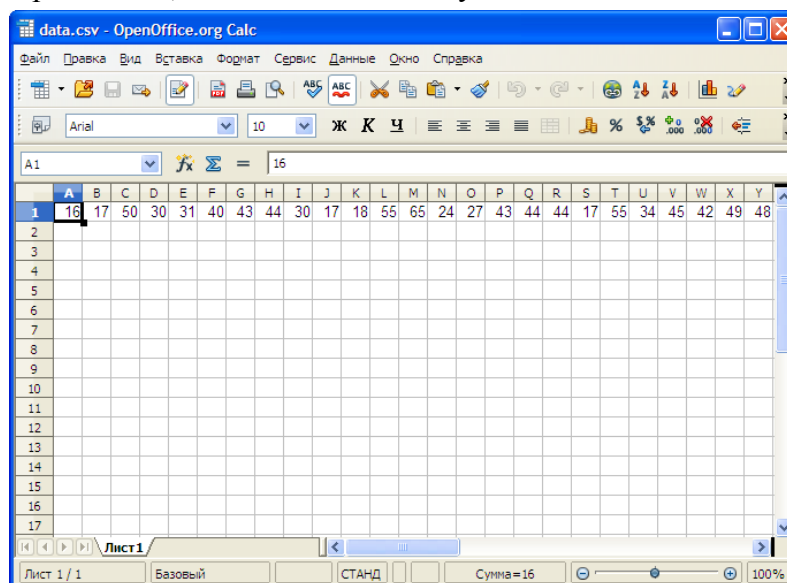


Рисунок 7.

Теперь нам необходимо сделать так, чтобы строка с данными стала столбцом («вращение» таблицы):

1. Для это выделите строку с данными, выберите **Правка — Вырезать** (Рисунок 8).

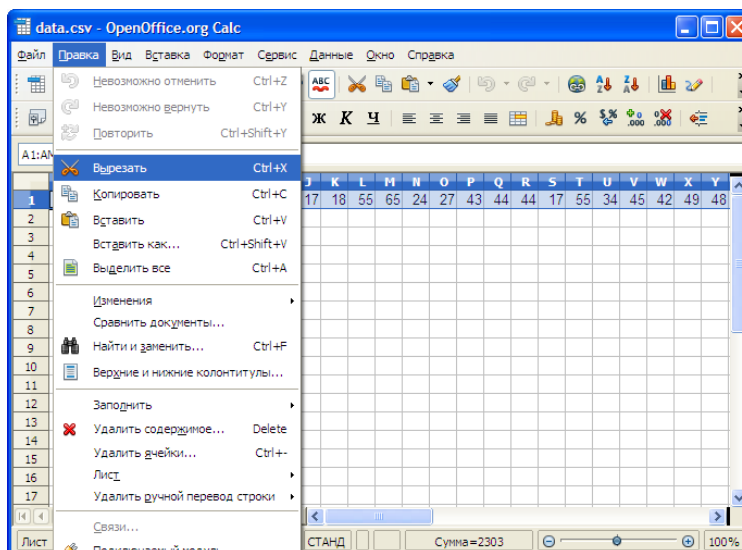


Рисунок 8.

2. Щелкните ячейку, которая в итоге должна стать верхней слева (ячейка A1).
3. Выберите команду **Правка — Вставить как...**
4. В диалоговом окне установите флажки **Вставить все** и **Транспонировать** (Рисунок 9).

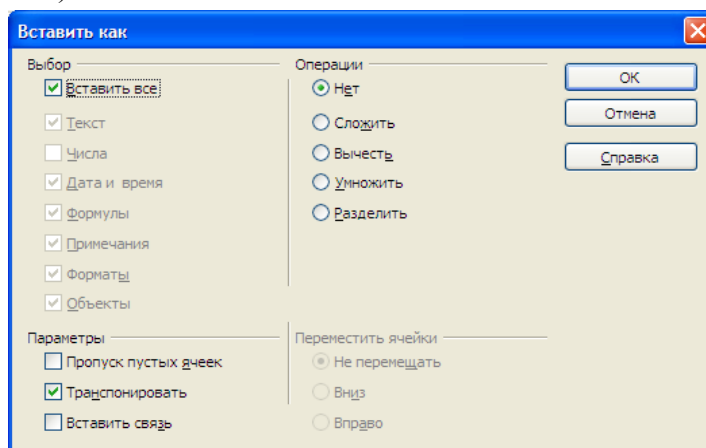


Рисунок 9.

5. Если теперь нажать кнопку **ОК**, все столбцы и строки будут транспонированы (Рисунок 10).

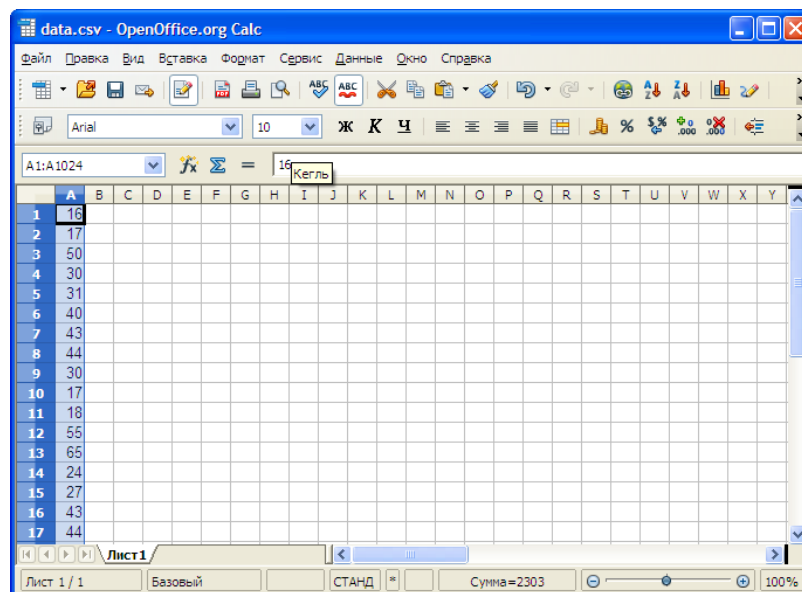


Рисунок 10. Вид электронной таблицы после транспонирования строки

Вставьте новую строку в электронную таблицу (Рисунок 11). В ячейку A1 введите текст «Экспериментальные данные».

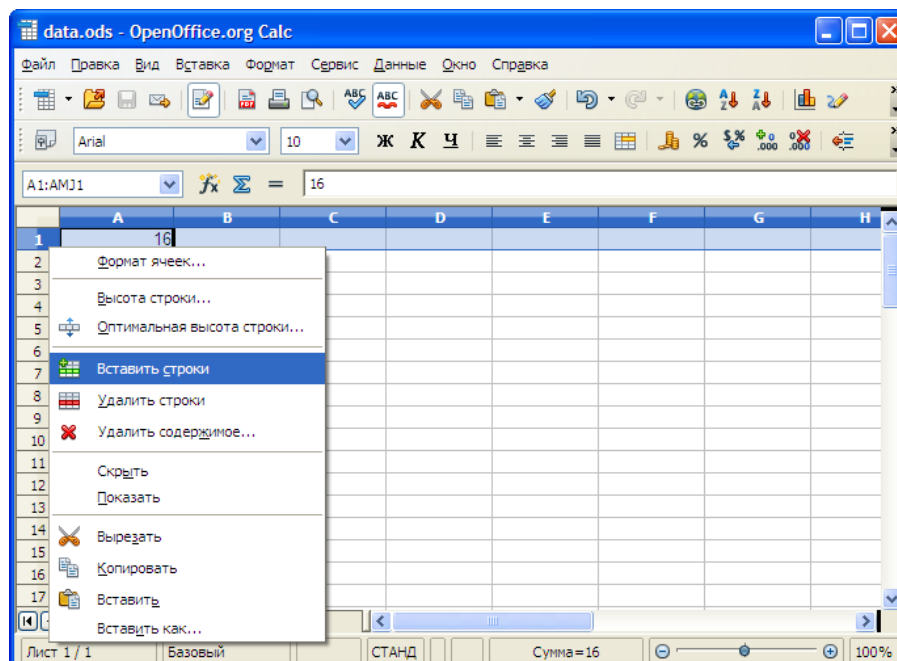


Рисунок 11. Вставка строки

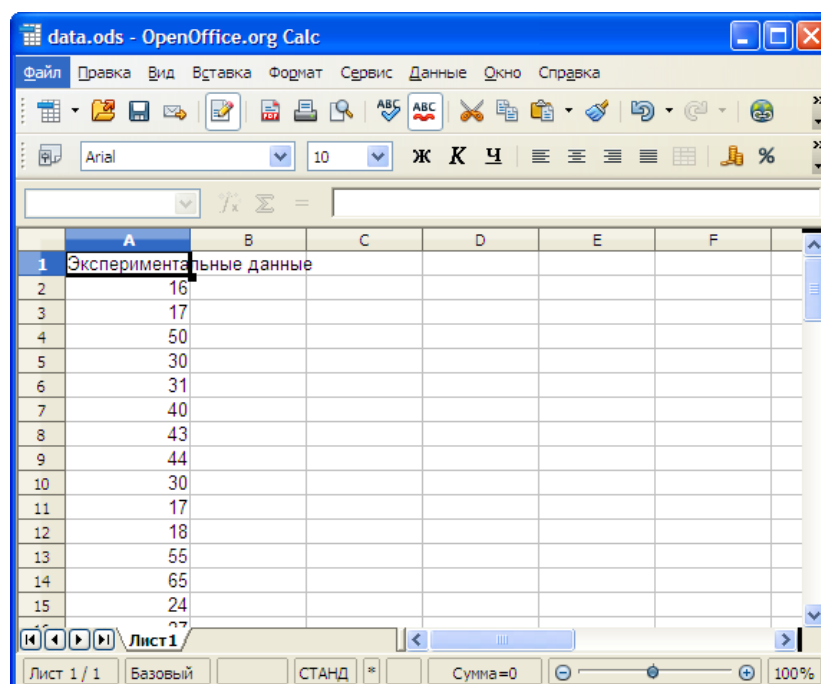


Рисунок 12. Ввод данных

Измените формат ячеек первой строки так, чтобы текст ячейки переносился по словам (выделите всю строку, вызовите **Формат — Ячейки**, Рисунок 13).

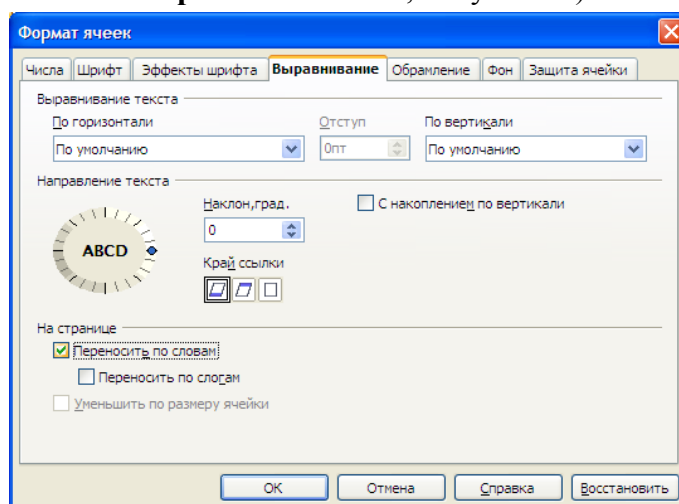


Рисунок 13. Изменение формата ячейки

Теперь необходимо отсортировать данные в порядке возрастания или убывания их значений. Для этого выделите диапазон ячеек, начиная с ячейки **A2**, и заканчивая ячейкой **A64**. Вызовите диалоговое окно для сортировки данных (**Данные — Сортировка**, Рисунок 14), установите параметры сортировки (**По возрастанию**), нажмите **ОК**.



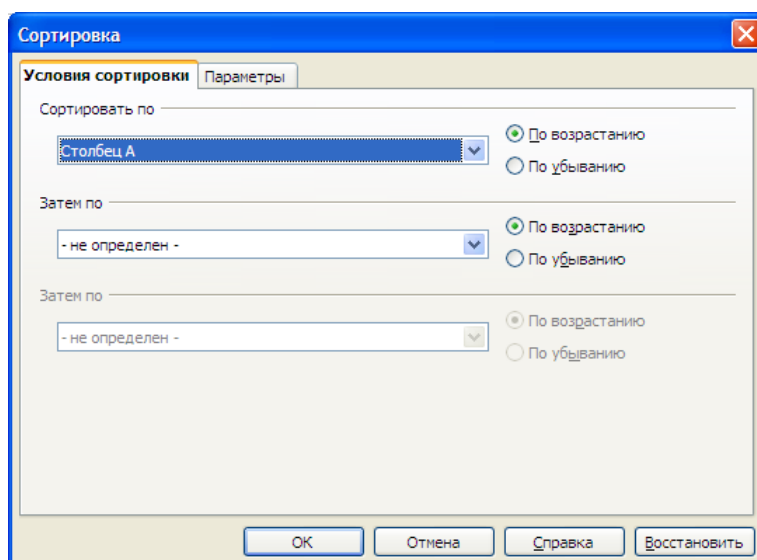


Рисунок 14. Сортировка данных

Сохраните электронную таблицу в своем каталоге (Файл — Сохранить как...). В диалоговом окне выберите Тип файла **Электронная таблица ODF (.ods)** (Рисунок 15).

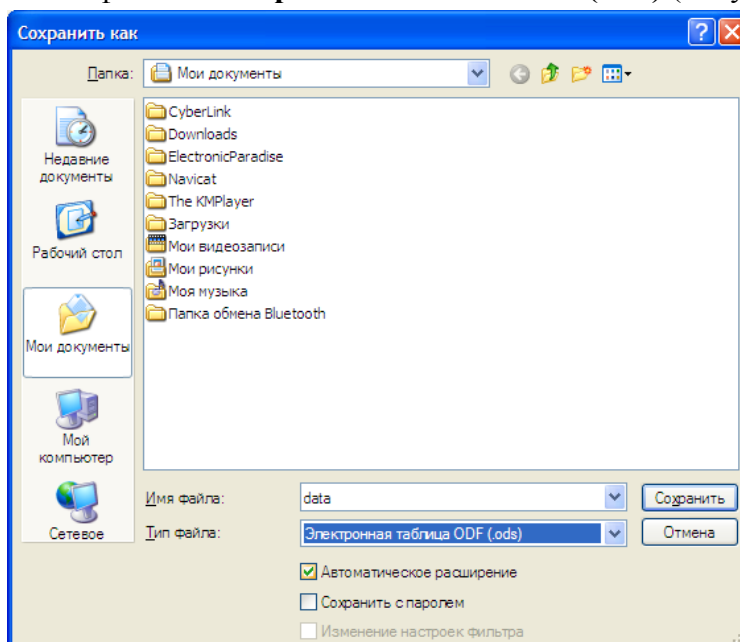


Рисунок 15. Диалоговое окно для сохранения файла

## Шаг 2. Разбиение экспериментальных данных на интервалы группировки

Первая задача, которую необходимо решить при группировке, состоит в том, чтобы разбить весь диапазон варьирования признака в выборке (между минимальной и максимальной вариантами выборки) на интервалы группировки. Эта задача требует определения числа интервалов группировки и ширины каждого из них.

**2.1.** Объем нашей выборки ( $n$ ) равен **63**, следовательно выбираем количество интервалов в диапазоне от **7** до **10** (Таблица 1).

Таблица 1. Выбор числа интервалов группировки

| Объем выборки, $n$ | Число интервалов, $k$ |
|--------------------|-----------------------|
| 25 – 40            | 5 – 6                 |

|           |         |
|-----------|---------|
| 40 – 60   | 6 – 8   |
| 60 – 100  | 7 – 10  |
| 100 – 200 | 8 – 12  |
| > 200     | 10 – 15 |

В ячейку **B1** введите текст «Количество интервалов», в ячейку **B2** введите число **7**.

Теперь найдем ширину каждого из интервалов (она будет одинаковой для всех интервалов). Ширина интервалов вычисляется по формуле:

$$h = \frac{x_{\max} - x_{\min}}{k - 1}$$

Для этого в ячейку **C1** введите текст «Ширина интервалов», а в ячейку **C2** введите формулу:

$$=(\text{MAX}(\text{A2:A64})-\text{MIN}(\text{A2:A64})) / (\text{B2}-1)$$

Данная формула вычисляем максимальное значение из диапазона с данными, минимальное значение, вычисляет разницу между ними, и делит на количество интервалов минус один.

Если вы ввели формулу правильно, то электронная таблица примет вид, как показано на Рисунок 16.

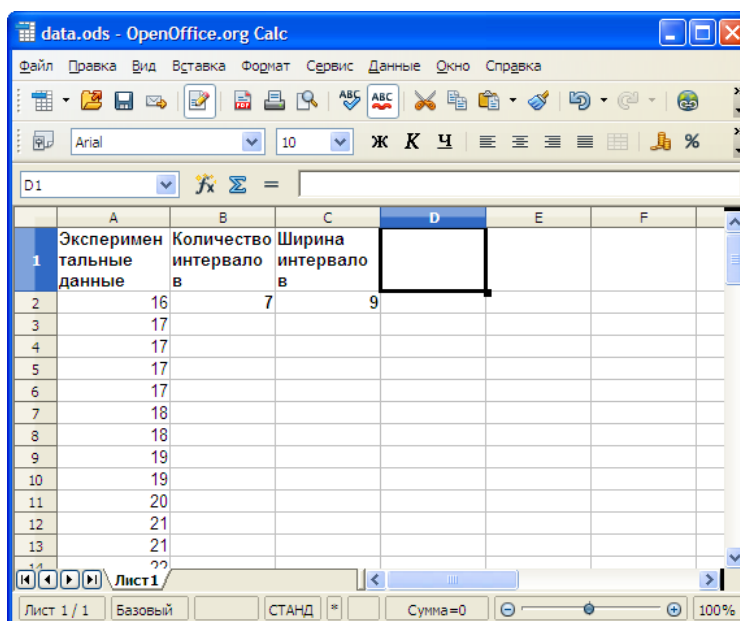


Рисунок 16.

В ячейку **D1** введите текст «Объем выборки», а в ячейку **D2** введите число **63** (Рисунок 17). Это понадобится для дальнейших вычислений.

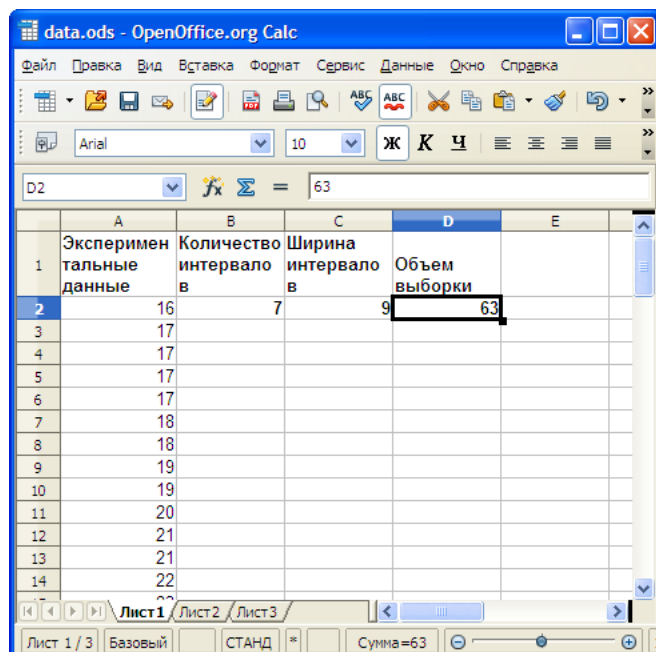


Рисунок 17.

### Шаг 3. Определение границ интервалов группировки.

Таблицу с интервалами будем составлять на втором листе таблицы. Если в Вашей таблице только один лист, то щелкните правой кнопкой мыши на ярлычке первого листа и выберите пункт **Добавить листы** (Рисунок 18). В диалоговом окне устанавливаем значения, как показано на Рисунок 19.

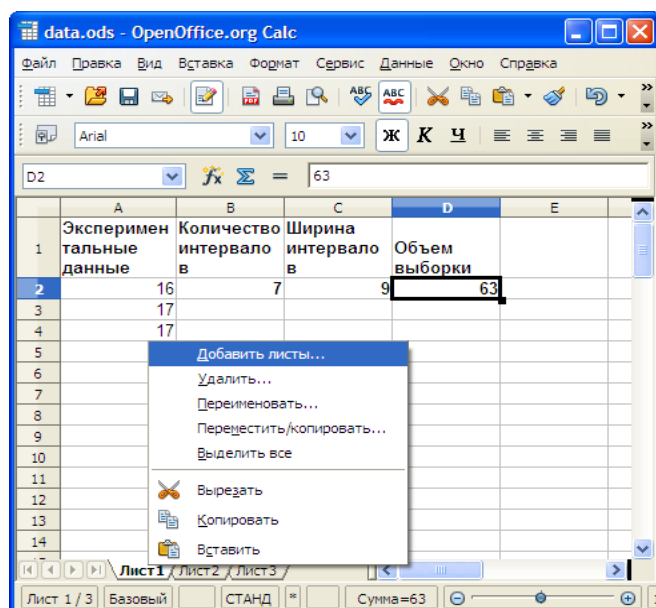


Рисунок 18. Добавление листов

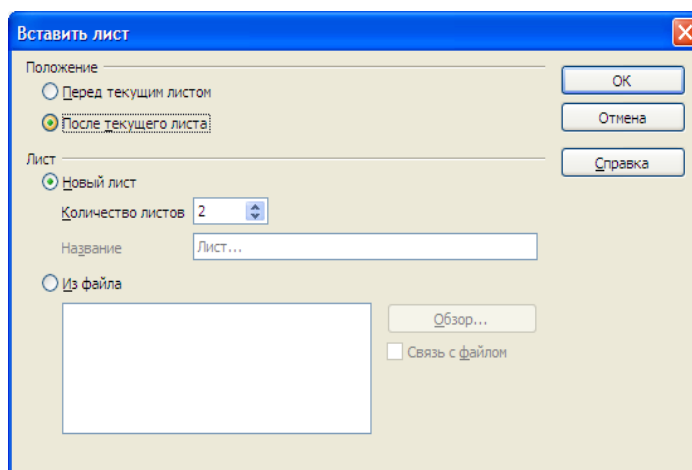


Рисунок 19. Диалоговое окно для вставки новых листов

Переходим на **Лист 2**. В ячейку **A1** вводим текст «Номер интервала,  $i$ », в ячейку **B1** – «Границы интервалов», в ячейку **B2** – «Нижняя,  $x_{ni}$ », в ячейку **C2** – «Верхняя,  $x_{vi}$ » (Рисунок 20).

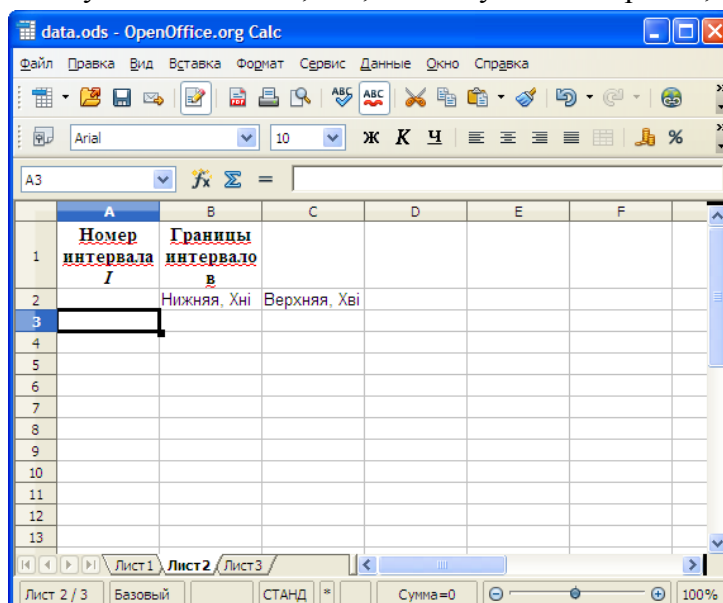


Рисунок 20.

Теперь ячейку **B1** необходимо объединить с ячейкой **C1**, а ячейку **A1** – с ячейкой **A2** (Рисунок 21). В столбце **A** введите номера интервалов, используя автозаполнение (Рисунок 22).

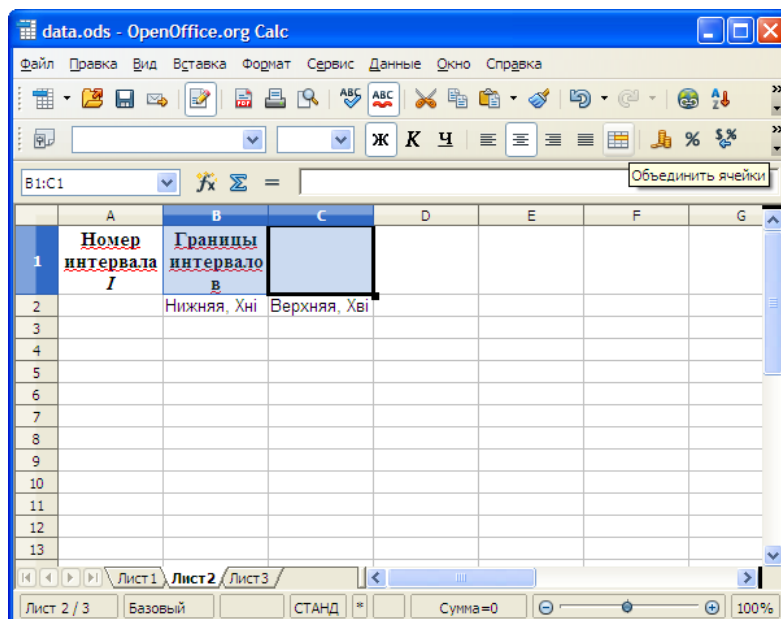


Рисунок 21. Объединение ячеек

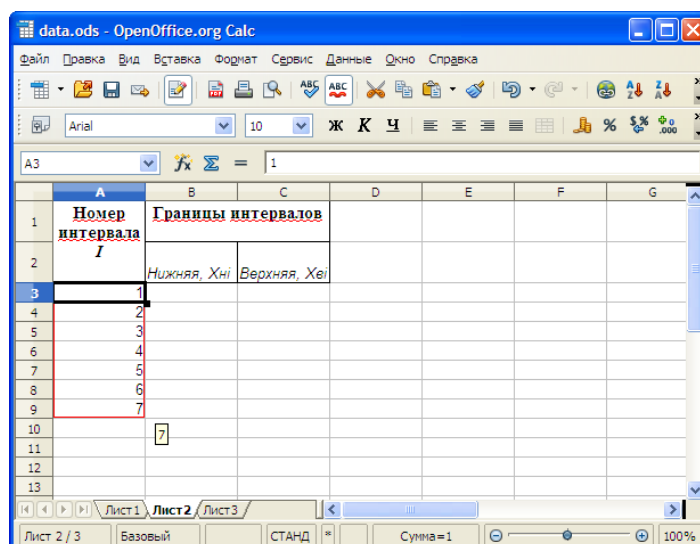


Рисунок 22.

Нижняя граница первого интервала выбирается так, чтобы минимальная варианта выборки  $x_{min}$  попадала примерно в середину этого интервала. Для вычисления нижней границы первого интервала воспользуемся формулой:

$$x_{H1} = x_{min} - \frac{h}{2}$$

Прибавив к этой величине ширину интервала, найдем нижнюю границу второго интервала. Это будет одновременно и верхняя граница  $x_{B1}$  предыдущего (первого) интервала.

В ячейку **B3** введем формулу для нахождения ее нижней границы:

$$= \text{Лист1.A2} - \text{Лист1.C2} / 2$$

где **Лист1.A2** — минимальное значение вариационного ряда, **Лист1.C2** — шаг интервалов выборки.

Значения нижних границ для следующих интервалов будут больше предыдущей нижней границы на величину шага интервала. Таким образом, для ячейки B4 формула будет следующей:

$$=B3+Лист1.\$C\$2$$

где **B3** — нижняя граница предыдущего интервала, **Лист1.\\$C\\$2** — шаг интервала. В формуле мы используем абсолютную адресацию ячейки **C2** для того, чтобы скопировать данную формулу в следующие ячейки, и чтобы ячейка **C2** в формуле была зафиксирована, и ее адрес не менялся при копировании. Скопируйте данную формулу на следующие ячейки (Рисунок 23).

|    | A                                            | B                             | C                              | D | E |
|----|----------------------------------------------|-------------------------------|--------------------------------|---|---|
| 1  | <u>Номер</u><br><u>интервала</u><br><i>I</i> | <u>Границы интервалов</u>     |                                |   |   |
| 2  |                                              | <i>Нижняя, X<sub>ни</sub></i> | <i>Верхняя, X<sub>ви</sub></i> |   |   |
| 3  | 1                                            | 11,5                          |                                |   |   |
| 4  | 2                                            | 20,5                          |                                |   |   |
| 5  | 3                                            | 29,5                          |                                |   |   |
| 6  | 4                                            | 38,5                          |                                |   |   |
| 7  | 5                                            | 47,5                          |                                |   |   |
| 8  | 6                                            | 56,5                          |                                |   |   |
| 9  | 7                                            | 65,5                          |                                |   |   |
| 10 |                                              |                               |                                |   |   |
| 11 |                                              |                               |                                |   |   |
| 12 |                                              |                               |                                |   |   |

Рисунок 23.

Для вычисления верхних границ интервалов нам необходимо к соответствующей нижней границе интервала прибавить величину шага интервала. Для ячейки **C3** формула будет следующей:

$$=B3+Лист1.\$C\$2$$

Введите данную формулу в электронную таблицу и скопируйте ее на остальные ячейки (Рисунок 24).

| Номер интервала I | Нижняя, X <sub>ни</sub> | Верхняя, X <sub>ви</sub> |
|-------------------|-------------------------|--------------------------|
| 1                 | 11,5                    | 20,5                     |
| 2                 | 20,5                    | 29,5                     |
| 3                 | 29,5                    | 38,5                     |
| 4                 | 38,5                    | 47,5                     |
| 5                 | 47,5                    | 56,5                     |
| 6                 | 56,5                    | 65,5                     |
| 7                 | 65,5                    | 74,5                     |

Рисунок 24.

#### Шаг 4. Вычисление срединных значений интервалов

Срединные значения интервалов группировки  $x_i$  отстоят от нижних границ на величину, равную половине ширины интервалов:

$$x_i = x_{H_i} - \frac{h}{2},$$

где  $x_{H_i}$  – нижняя граница  $i$ -го интервала.

Ячейку **D1** объедините с ячейкой **D2** и введите текст «Срединные значения,  $x_i$ ». В ячейку **D3** введите формулу:

$$=B3+Лист1.$C$2/2$$

Скопируйте содержимое ячейки **D3** в ячейки **D4:D9** (Рисунок 25).

| Номер интервала I | Нижняя, X <sub>ни</sub> | Верхняя, X <sub>ви</sub> | Срединные значения, X <sub>i</sub> |
|-------------------|-------------------------|--------------------------|------------------------------------|
| 1                 | 11,5                    | 20,5                     | 16                                 |
| 2                 | 20,5                    | 29,5                     | 25                                 |
| 3                 | 29,5                    | 38,5                     | 34                                 |
| 4                 | 38,5                    | 47,5                     | 43                                 |
| 5                 | 47,5                    | 56,5                     | 52                                 |
| 6                 | 56,5                    | 65,5                     | 61                                 |
| 7                 | 65,5                    | 74,5                     | 70                                 |

Рисунок 25.

## Шаг 5. Распределение вариантов выборки по интервалам группировки.

На основании первичных данных необходимо распределить варианты выборки по интервалам группировки, то есть, подсчитать частоты интервалов  $n_i$ , сумма которых всегда равна объему выборки  $n$ .

Итак, объединяем ячейку **E1** с ячейкой **E2** вводим текст «Частота интервала,  $n_i$ ». А в ячейку **E3** вводим формулу:

$$=COUNTIF(\text{Лист1.A2:A64}; "<"\&C3)-COUNTIF(\text{Лист1.A2:A64}; "<"\&B3)$$

Таким образом, мы подсчитываем количество значений в выборке, которые меньше верхней границы текущего интервала и вычитаем из этого значения количество значений выборки, которые меньше нижней границы.

Скопируйте формулу в остальные ячейки и проверьте правильность вычислений путем ссуммировани всех частот (Рисунок 26).

| Номер интервала I | Границы интервалов |                   | Срединные значения, $X_i$ | Частота интервала, $n_i$ |
|-------------------|--------------------|-------------------|---------------------------|--------------------------|
|                   | Нижняя, $X_{ни}$   | Верхняя, $X_{ви}$ |                           |                          |
| 1                 | 11,5               | 20,5              | 16                        | 10                       |
| 2                 | 20,5               | 29,5              | 25                        | 13                       |
| 3                 | 29,5               | 38,5              | 34                        | 11                       |
| 4                 | 38,5               | 47,5              | 43                        | 14                       |
| 5                 | 47,5               | 56,5              | 52                        | 9                        |
| 6                 | 56,5               | 65,5              | 61                        | 4                        |
| 7                 | 65,5               | 74,5              | 70                        | 2                        |
| 10                |                    |                   |                           | 63                       |

Рисунок 26.

Чтобы вычислить накопленную частоту интервала необходимо последовательно росуммировать частоты в направлении от первого интервала к последнему.

Объедините ячейки **F1** и **F2** и введите текст «Накопленные частоты».

Для первого интервала накопленная частот а будет равна частоте данного интервала (ячейка **F3**). В ячейку же **F4** введите формулу:

$$=F3+E4$$

Скопируйте данную формулу на оставшиеся ячейки (Рисунок 27).



|    | B                                  | C                                   | D                         | E                        | F                                        |
|----|------------------------------------|-------------------------------------|---------------------------|--------------------------|------------------------------------------|
| 1  | <u>Границы интервалов</u>          |                                     | Срединные значения, $X_i$ | Частота интервала, $n_i$ | <u>Накопл. частоты</u><br>$N_{\Sigma i}$ |
| 2  | <u>Нижняя, <math>X_{ni}</math></u> | <u>Верхняя, <math>X_{vi}</math></u> |                           |                          |                                          |
| 3  | 11,5                               | 20,5                                | 16                        | 10                       | 10                                       |
| 4  | 20,5                               | 29,5                                | 25                        | 13                       | 23                                       |
| 5  | 29,5                               | 38,5                                | 34                        | 11                       | 34                                       |
| 6  | 38,5                               | 47,5                                | 43                        | 14                       | 48                                       |
| 7  | 47,5                               | 56,5                                | 52                        | 9                        | 57                                       |
| 8  | 56,5                               | 65,5                                | 61                        | 4                        | 61                                       |
| 9  | 65,5                               | 74,5                                | 70                        | 2                        | 63                                       |
| 10 |                                    |                                     |                           | 63                       |                                          |
| 11 |                                    |                                     |                           |                          |                                          |

Рисунок 27. Определение накопленных частот

Объедините ячейки **G1** и **G2** и введите текст «Относительные частоты,  $w_i$ ».

Вычисляем относительную частоту интервала (отношение частоты к объему выборки), для чего в ячейку **G3** введите формулу:

$$=E3/Лист1.\$D\$2$$

Скопируйте данную формулу на оставшиеся ячейки (Рисунок 28). Сумма всех относительных частот всегда равна единице. Проверьте правильность вычислений путем подсчета данной суммы (Рисунок 28).

|    | B                                  | C                                   | D                         | E                        | F                                        | G                                     |
|----|------------------------------------|-------------------------------------|---------------------------|--------------------------|------------------------------------------|---------------------------------------|
| 1  | <u>Границы интервалов</u>          |                                     | Срединные значения, $X_i$ | Частота интервала, $n_i$ | <u>Накопл. частоты</u><br>$N_{\Sigma i}$ | <u>Относительные частоты</u><br>$w_i$ |
| 2  | <u>Нижняя, <math>X_{ni}</math></u> | <u>Верхняя, <math>X_{vi}</math></u> |                           |                          |                                          |                                       |
| 3  | 11,5                               | 20,5                                | 16                        | 10                       | 10                                       | 0.16                                  |
| 4  | 20,5                               | 29,5                                | 25                        | 13                       | 23                                       | 0.21                                  |
| 5  | 29,5                               | 38,5                                | 34                        | 11                       | 34                                       | 0.17                                  |
| 6  | 38,5                               | 47,5                                | 43                        | 14                       | 48                                       | 0.22                                  |
| 7  | 47,5                               | 56,5                                | 52                        | 9                        | 57                                       | 0.14                                  |
| 8  | 56,5                               | 65,5                                | 61                        | 4                        | 61                                       | 0.06                                  |
| 9  | 65,5                               | 74,5                                | 70                        | 2                        | 63                                       | 0.03                                  |
| 10 |                                    |                                     |                           | 63                       |                                          | 1                                     |
| 11 |                                    |                                     |                           |                          |                                          |                                       |

Рисунок 28. Определение относительных частот

Относительные частоты показывают доли (удельные веса) членов совокупности с одинаковым значением признака (для дискретных рядов) или попадающие в один интервал (для непрерывных).

Накопленные относительные частоты вычисляются как накопленная частота интервала, поделенная на объем выборки. Объедините ячейки **H1** и **H2**, введите текст «Накопленные относительные частоты», а в ячейку **H3** введите следующую формулу и скопируйте на

оставшиеся ячейки (Рисунок 29):

$$=F3/Лист1.SD\$2$$

|    | Границы интервалов |                   | Срединные значения, $X_i$ | Частота интервала, $n_i$ | Накопл. частоты $N_{\Sigma i}$ | Относительные частоты $n_i$ | Накопл. относит. частоты $F_i$ |
|----|--------------------|-------------------|---------------------------|--------------------------|--------------------------------|-----------------------------|--------------------------------|
|    | Нижняя, $X_{ni}$   | Верхняя, $X_{vi}$ |                           |                          |                                |                             |                                |
| 3  | 11,5               | 20,5              | 16                        | 10                       | 10                             | 0,16                        | 0,16                           |
| 4  | 20,5               | 29,5              | 25                        | 13                       | 23                             | 0,21                        | 0,37                           |
| 5  | 29,5               | 38,5              | 34                        | 11                       | 34                             | 0,17                        | 0,54                           |
| 6  | 38,5               | 47,5              | 43                        | 14                       | 48                             | 0,22                        | 0,76                           |
| 7  | 47,5               | 56,5              | 52                        | 9                        | 57                             | 0,14                        | 0,9                            |
| 8  | 56,5               | 65,5              | 61                        | 4                        | 61                             | 0,06                        | 0,97                           |
| 9  | 65,5               | 74,5              | 70                        | 2                        | 63                             | 0,03                        | 1                              |
| 10 |                    |                   |                           | 63                       |                                | 1                           |                                |

Рисунок 29. Вычисление накопленных относительных частот

В итоге, для последнего интервала значение накопленной относительной частоты будет равно единице.

## Графическое представление данных

Для повышения наглядности эмпирических распределений, используется их графическое представление. Наиболее распространенными способами графического представления являются гистограмма, полигон частот и полигон накопленных частот

### Гистограмма

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которой служат частичные интервалы длиной  $h$ , а высоты равны отношению  $n_i / h$  (частота интервала к ширине интервала).

Для построения гистограммы частот на оси абсцисс откладывают частичные интервалы, а над ними проводят отрезки, параллельные оси абсцисс на расстоянии (высоте)  $n_i / h$ . Площадь  $i$ -го прямоугольника равна  $h * n_i / h = n_i$  – сумме частот вариант  $i$ -го интервала, поэтому площадь гистограммы частот равна сумме всех частот, т.е. объему выборки.

В случае гистограммы относительных частот по оси ординат откладываются относительные частоты, на оси абсцисс – частичные интервалы, над ними проводят отрезки, параллельные оси абсцисс на высоте  $w_i / h$ . Площадь гистограммы относительных частот равна сумме всех относительных частот, то есть единице.

Итак, для построения гистограммы частот необходимо вычислить отношение  $n_i / h$ . Для этого объедините ячейки **I1** и **I2** и введите текст « $n_i / h$ ». В ячейку **I3** введите формулу:

$$=E3/Лист1.SCS2$$

Скопируйте формулу на оставшиеся ячейки таблицы (Рисунок 30).

| Номер интервала I | Границы интервалов |                   | Срединные значения, $X_i$ | Частота интервала, $n_i$ | Накопл. частоты $N_i$ | Относительные частоты $H_i$ | Накопл. относит. частоты $F_i$ | $n_i / h$ |
|-------------------|--------------------|-------------------|---------------------------|--------------------------|-----------------------|-----------------------------|--------------------------------|-----------|
|                   | Нижняя, $X_{ни}$   | Верхняя, $X_{ви}$ |                           |                          |                       |                             |                                |           |
| 1                 | 1                  | 11,5              | 20,5                      | 16                       | 10                    | 0,16                        | 0,16                           | 1,11      |
| 2                 | 2                  | 20,5              | 29,5                      | 25                       | 13                    | 0,21                        | 0,37                           | 1,44      |
| 3                 | 3                  | 29,5              | 38,5                      | 34                       | 11                    | 0,17                        | 0,54                           | 1,22      |
| 4                 | 4                  | 38,5              | 47,5                      | 43                       | 14                    | 0,22                        | 0,76                           | 1,56      |
| 5                 | 5                  | 47,5              | 56,5                      | 52                       | 9                     | 0,14                        | 0,9                            | 1         |
| 6                 | 6                  | 56,5              | 65,5                      | 61                       | 4                     | 0,06                        | 0,97                           | 0,44      |
| 7                 | 7                  | 65,5              | 74,5                      | 70                       | 2                     | 0,03                        | 1                              | 0,22      |

Рисунок 30.

Теперь выделите диапазон ячеек **I3:I9**. Вызовите пункт меню **Вставка — Диаграмма** (Рисунок 31).

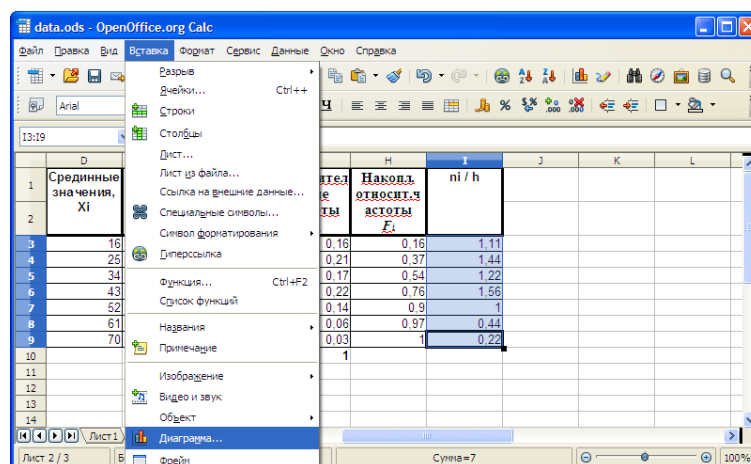


Рисунок 31. Вставка диаграммы

В появившемся диалоговом окне на **Шаге 1** выберите тип диаграммы (Гистограмма — Обычная, Рисунок 32). Перейдите к следующему шагу (нажмите **Далее**).

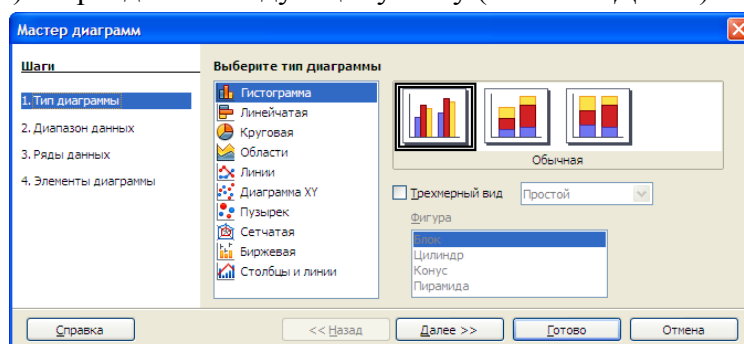


Рисунок 32. Выбор типа диаграммы

На втором шаге оставляем всё, как есть, а на третьем шаге указываем адрес ячейки, в котором содержится название для данных гистограммы (Рисунок 33).

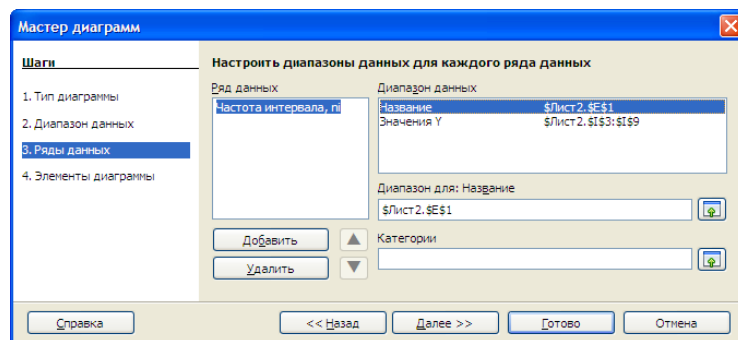


Рисунок 33. Третий шаг построения гистограммы  
На четвертом шаге указываем остальные элементы диаграммы (Рисунок 34).

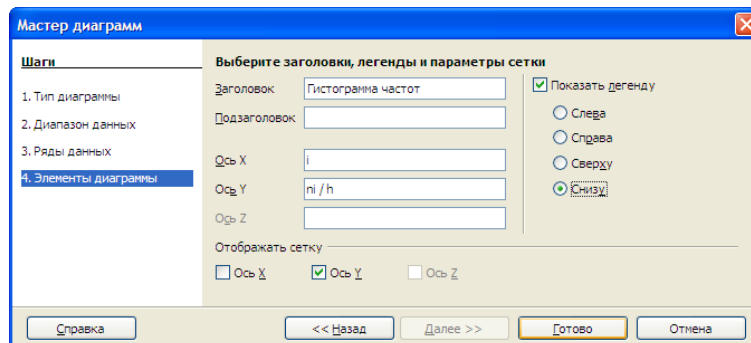


Рисунок 34. Четвертый шаг построения диаграммы  
Если все сделано правильно, то гистограмма примет вид, как показано на Рисунок 35.

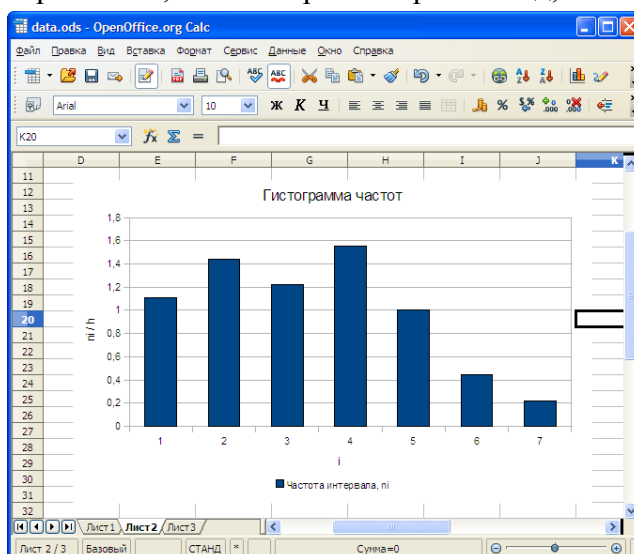


Рисунок 35. Итоговый вид гистограммы

## Полигон частот

Полигон частот представляет собой ломаную, соединяющую точки, соответствующие средним значениям интервалов группировки и частотам этих интервалов. Для построения полигона частот выделите диапазон **E3:E9** и вызовите пункт меню **Вставка — Диаграмма**. В диалоговом окне на первом шаге построения укажите тип **Линии — Линии и точки** (Рисунок 36), перейдите к третьему шагу, оставив на втором всё без изменений.

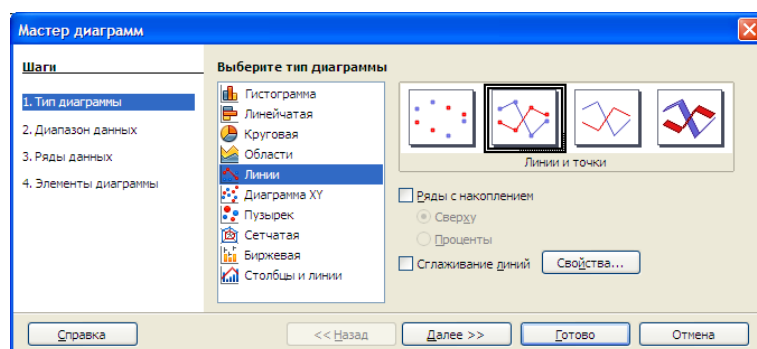


Рисунок 36. Первый шаг построения полигона частот

На третьем шаге необходимо указать значения для оси X. Для этого в поле **Категории** укажите диапазон, содержащий срединные значения (Рисунок 37). В поле **Название** укажите ячейку E1.

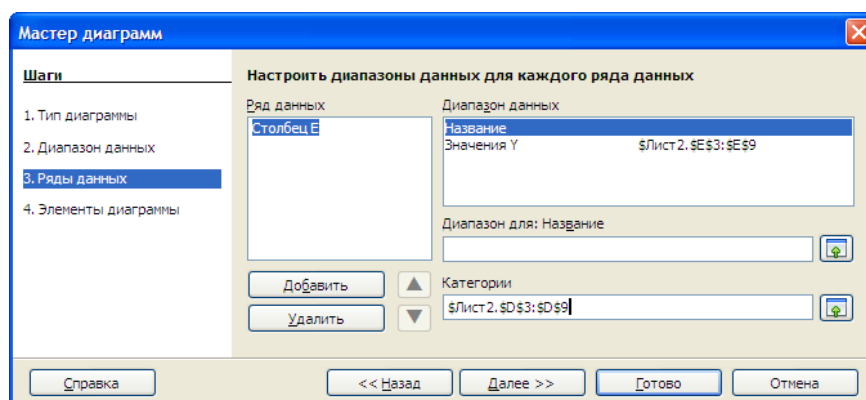


Рисунок 37. Третий шаг построения полигона частот

На последнем шаге введите название диаграммы, подписи для горизонтальной и вертикальной осей (Рисунок 38). Нажмите **Готово**.

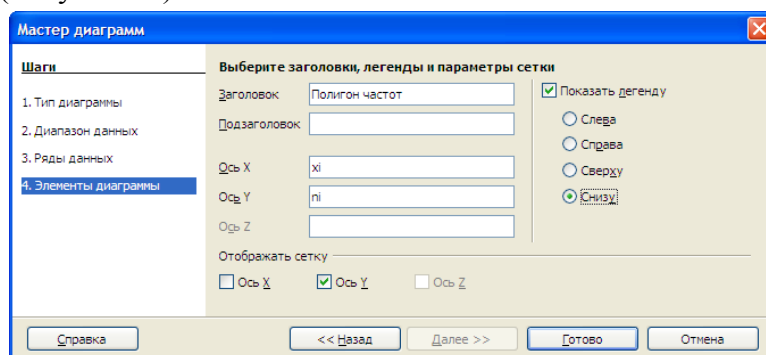


Рисунок 38. Четвертый шаг построения полигона частот

Если все сделано правильно, то полигон частот будет выглядеть, как показано на Рисунок 39.

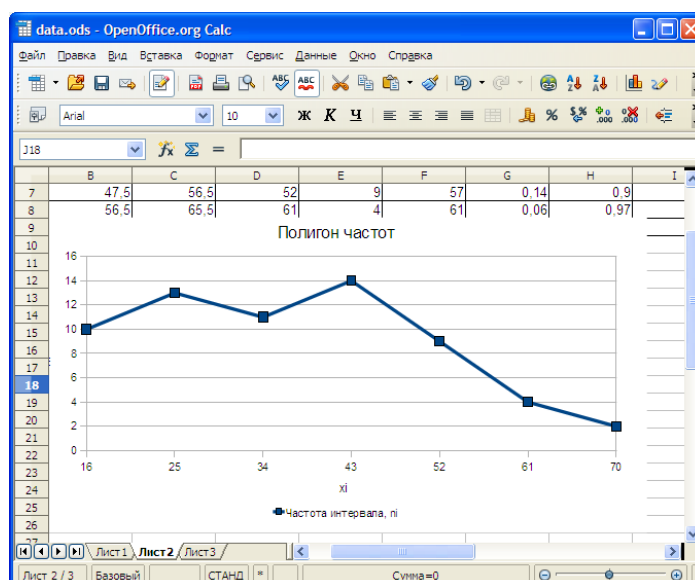


Рисунок 39. Итоговый вид полигона частот

## Полигон накопленных частот

Полигон накопленных частот получается при соединении отрезками прямых точек, координаты которых соответствуют верхним границам интервалов группировки и накопленным частотам. Если по оси ординат откладывать накопленные относительные частоты, то полученный график называется полигоном накопленных относительных частот.

Для построения полигона накопленных частот выделите диапазоны C3:C9 и F3:F9, удерживая клавишу Ctrl (Рисунок 40). Вызовите диалоговое окно для вставки диаграммы, укажите тип **Линии и точки**.

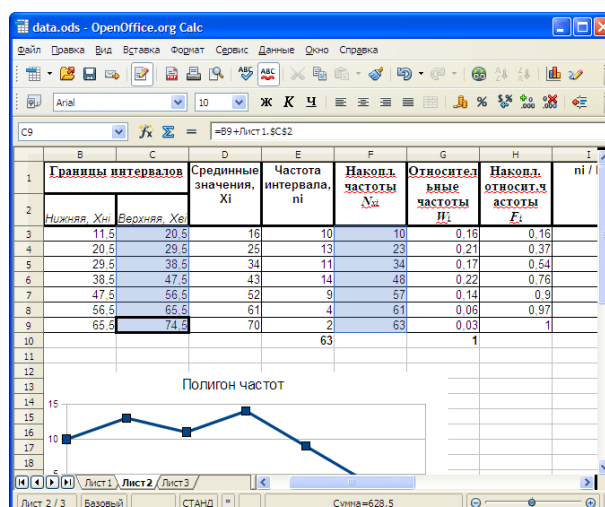


Рисунок 40.

На втором шаге установите флажок в поле **Первый столбец как подпись** (Рисунок 41).

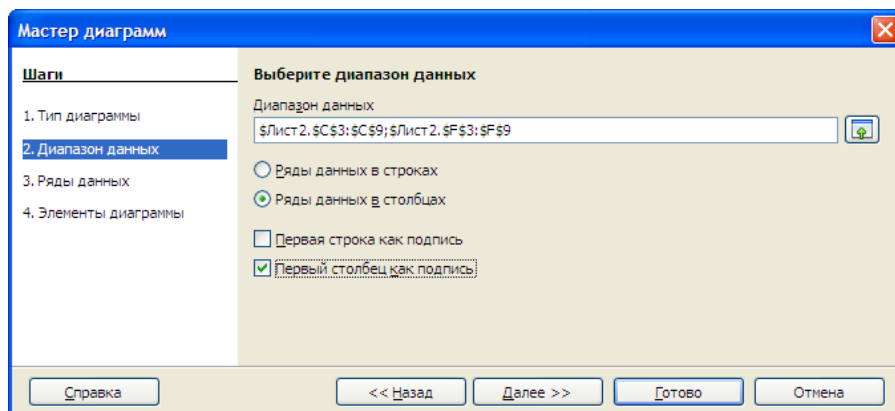


Рисунок 41.

На третьем шаге в поле Название укажите ячейку F1 (Рисунок 42).

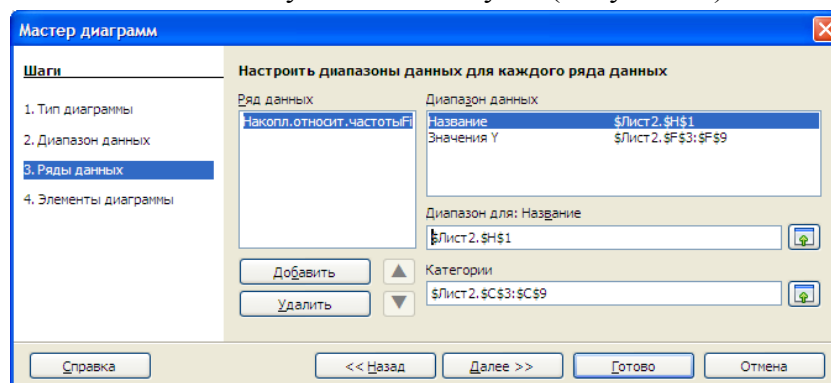


Рисунок 42.

На последнем шаге введите остальные элементы диаграммы (Рисунок 43).

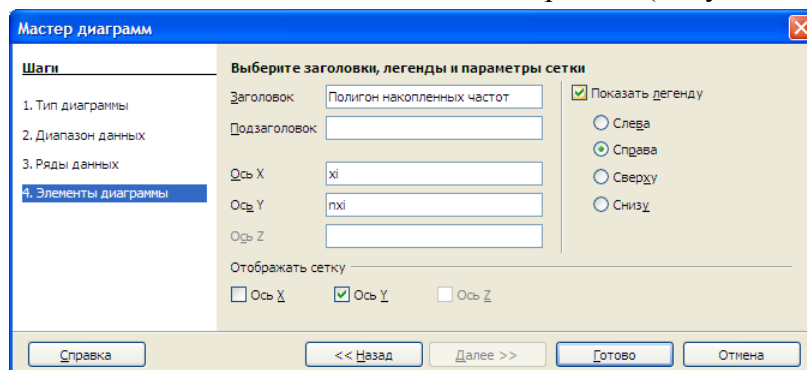


Рисунок 43.

Если все сделано правильно, то полигон накопленных частот примет вид (Рисунок 44):

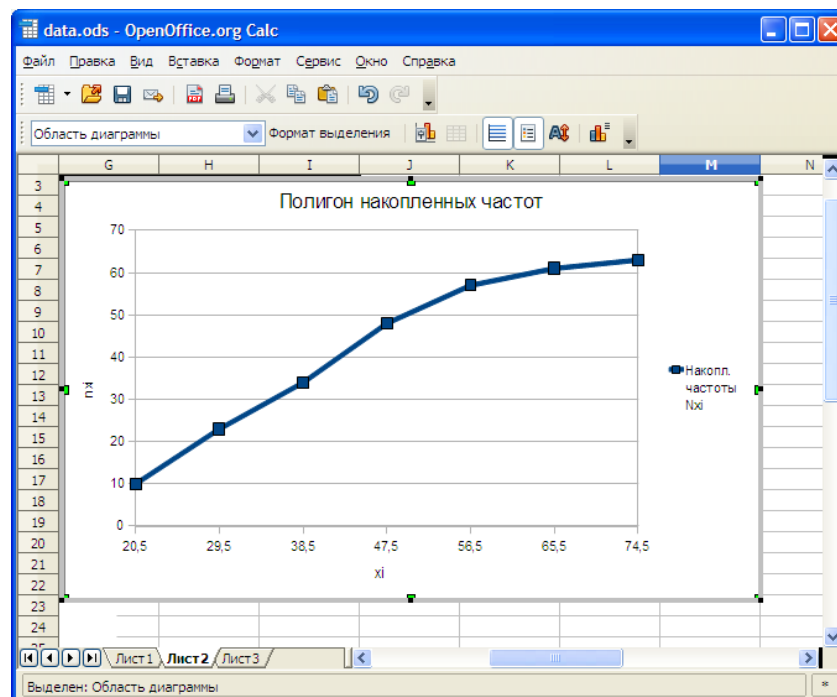


Рисунок 44. Полигон накопленных частот

## Определение числовых характеристик выборки

Характеристики положения определяют положение центра эмпирического распределения. Чаще всего употребляются такие характеристики положения, как среднее арифметическое, медиана и мода.

### Среднее арифметическое (математическое ожидание)

Среднее арифметическое – такое значение признака, сумма отклонений от которого выборочных значений признака равна нулю (с учетом знака отклонения).

Для несгруппированных данных среднее арифметическое определяется по следующей формуле:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n},$$

где  $n$  — объем выборки;  $x_i$  — варианты выборки.

Для сгруппированных же данных среднее арифметическое можно найти по формуле:

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2 + \dots + n_k x_k}{n},$$

где  $n$  — объем выборки;  $k$  — число интервалов группировки;  $n_i$  — частота  $i$ -го интервала;  $x_i$  — срединное значение  $i$ -ого интервала.

Итак, для нахождения среднего арифметического в ячейке A11 введите текст «Среднее арифметическое», а в соседнюю ячейку введите формулу:

$$=SUMPRODUCT(D3:D9;E3:E9)/Лист1.D2$$

Где функция **SUMPRODUCT** вычислит сумму произведений указанных диапазонов.

### Мода

Мода представляет собой значение признака, встречающееся в выборке наиболее часто. Ряд называется унимодальным, если в нем только одно модальное значение и полимодальным в противном случае. Для полимодального ряда значение моды не определяют.



Для дискретного ряда мода находится по определению. Например, если имеем следующие данные измерений кол-ва :

10, 15, 20, 22, 22, 23, 25, 25, 26, 26, 26, 27, 28, 28, 30, 30, 31 ,

то чаще всего встречается значение 26, значит данный ряд унимодальный и мода равна 26. В данном случае мы бы использовали функцию **MODE** из категории **Статистические**.

В тех случаях, когда данные сгруппированы, в первую очередь определяют модальный интервал (интервал группировки с наибольшей частотой). В нашем случае это четвертый интервал (значение частоты — 14).

Для определения моды в интервальном ряду используется следующая формула:

$$Mo = x_{Mo_n} + h \frac{n_{Mo} - n_{Mo-1}}{(n_{Mo} - n_{Mo-1}) + (n_{Mo} - n_{Mo+1})} ,$$

где  $x_{Mo_n}$  — нижняя граница модального интервала;  $h$  — ширина интервала группировки;  $n_{Mo}$  — частота модального интервала;  $n_{Mo-1}$  — частота интервала, предшествующего модальному;  $n_{Mo+1}$  — частота интервала, следующего за модальным.

В ячейку **A13** введите текст «Модальный интервал». В ячейку **B13** – «4». В ячейку **A14** введите текст «Мода», а в ячейку **A15** – формулу:

$$=B6+Лист1.C2*(E6-E5)/(E6-E5+E6-E7)$$

## Медиана

Медианой называется такое значение признака  $X$ , когда ровно половина значений экспериментальных данных меньше ее, а вторая половина — больше.

Для ранжированного ряда с нечетным числом величин (например, 1, 2, 4, 4, 6, 7, 8, 8, 10) медианой будет величина, которая расположена в центре ряда, т.е. пятая величина.

Для ранжированного ряда с четным числом величин (например, 1, 5, 7, 10, 11, 14) медианой будет средняя арифметическая величина, которая рассчитывается из двух смежных величин. В данном случае медиана равна  $(7+10) / 2 = 8,5$ .

Таким образом, для нахождения медианы сначала необходимо определить ее порядковый номер (ее положение в ранжированном ряду) по формуле

$$N_{Me} = \frac{n+1}{2} , \text{ где } n - \text{объем совокупности.}$$

Для сгруппированных по интервалам данных значение медианы определяют по накопленным частотам дискретного вариационного ряда, для чего требуется указать интервал нахождения медианы в интервальном ряду распределения. Медианным называют первый интервал, где сумма накопленных частот превышает половину наблюдений от общего числа всех наблюдений.

Внутри медианного интервала медиана определяется по следующей формуле:

$$Me = x_{Me_n} + h_{me} \frac{0,5n - n_{x_{Me-1}}}{n_{Me}} ,$$

где  $x_{Me_n}$  — нижняя граница медианного интервала;  $0,5n = n$  — половина объема выборки;  $h_{me}$  — ширина медианного интервала; — накопленная частота интервала, предшествующего медианному,  $n_{x_{Me-1}}$   $n_{Me}$  — частота медианного интервала.

В ячейку **A17** введите формулу:

$$=E10/2$$

Это и есть половина объем выборки. Как видно из таблицы, именно в третьем интервале впервые накопленная частота больше данного значения. В ячейку **B17** введите номер данного интервала.

В ячейку **B18** введите формулу для численного определения медианы:

$$=B5+Лист1.C2*(B16-F4)/E5$$

Если все формулы введены правильно, то значения характеристик вариационного ряда будут следующие (Рисунок 45):

|    | A                  | B     | C | D | E |
|----|--------------------|-------|---|---|---|
| 10 |                    |       |   |   |   |
| 11 | Среднее ариф.      | 36,71 |   |   |   |
| 12 |                    |       |   |   |   |
| 13 | Модальный интервал | 4     |   |   |   |
| 14 | Мода               | 41,88 |   |   |   |
| 15 |                    |       |   |   |   |
| 16 | N                  | 31,5  |   |   |   |
| 17 | Медианный интервал | 3     |   |   |   |
| 18 | Медиана            | 36,45 |   |   |   |
| 19 |                    |       |   |   |   |
| 20 |                    |       |   |   |   |
| 21 |                    |       |   |   |   |
| 22 |                    |       |   |   |   |
| 23 |                    |       |   |   |   |

Рисунок 45. Числовые характеристики

## Дисперсия

Дисперсия и стандартное отклонение являются важнейшими характеристиками рассеяния.

Дисперсией называется средний квадрат отклонения значений признака от среднего арифметического. Дисперсия, вычисляемая по выборочным данным, называется выборочной дисперсией и обозначается  $\sigma^2$ .

Для несгруппированных данных значение дисперсии вычисляется по формуле:

$$\sigma_B^2 = \frac{\sum_{i=1}^n (x_i - \bar{x}_B)^2}{n}.$$

где в числителе — сумма квадратов отклонений значений признака  $x_i$  от среднего арифметического  $\bar{x}$ . Для получения среднего квадрата отклонений эта сумма поделена на объем выборки  $n$ .

Для несгруппированных данных существуют стандартные функции для вычисления дисперсии (**ДИСП** – в MS Excel и **VAR** – в OOo Calc).

Для сгруппированных данных дисперсия определяется по формуле:

$$\sigma_B^2 = \frac{\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2}{n}.$$

Здесь  $x_i$  — срединные значения интервалов группировки;  $\bar{x}$  — среднее арифметическое.

В ячейку **B20** введите формулу:

**=SUMPRODUCT(D3:D9;D3:D9;E3:E9)/E10-B11\*B11**

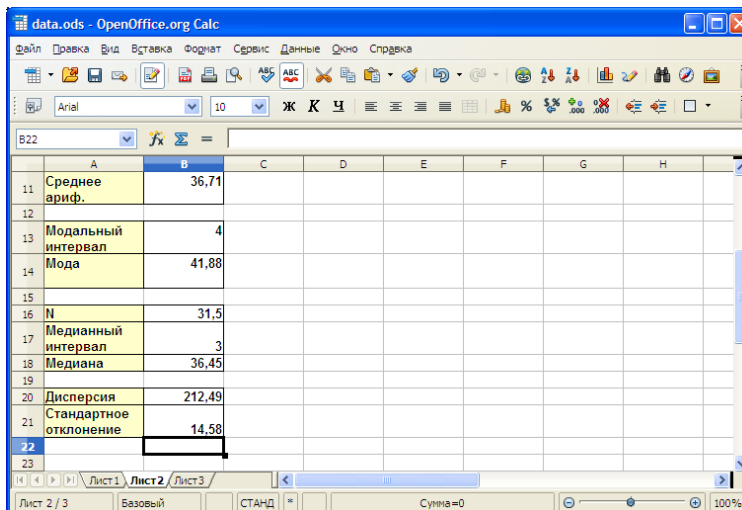
## Стандартное отклонение

Стандартным отклонением (или средним квадратическим отклонением) называется корень квадратный из дисперсии.

Для определения стандартного отклонения в ячейку B21 введите формулу:

$$=SQRT(B20)$$

Если все формулы введены правильно, то результаты вычислений будут следующими (Рисунок 46):



The screenshot shows the OpenOffice.org Calc application window with a spreadsheet titled 'data.ods'. The spreadsheet contains statistical data in columns A and B. The results are as follows:

|    | A                      | B      |
|----|------------------------|--------|
| 11 | Среднее ариф.          | 36,71  |
| 12 |                        |        |
| 13 | Модальный интервал     | 4      |
| 14 | Мода                   | 41,88  |
| 15 |                        |        |
| 16 | N                      | 31,5   |
| 17 | Медианный интервал     | 3      |
| 18 | Медиана                | 36,45  |
| 19 |                        |        |
| 20 | Дисперсия              | 212,49 |
| 21 | Стандартное отклонение | 14,58  |
| 22 |                        |        |
| 23 |                        |        |

Рисунок 46. Результаты вычислений