

## ЛАБОРАТОРНАЯ РАБОТА 6. РАСЧЕТ СТАТИСТИЧЕСКИХ ХАРАКТЕРИСТИК ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

**Цель:** провести первичную статистическую обработку данных, ознакомиться с работой инструментов «Описательная статистика» и «Гистограмма» пакета Анализа данных MS Excel.

### 6.1. Определение основных статистических характеристик с помощью инструмента «Описательная статистика»

Исследуется результаты ежедневных измерений некоторой случайной величины. Получены данные за первые 150 суток года. Занесите полученные данные в таблицу

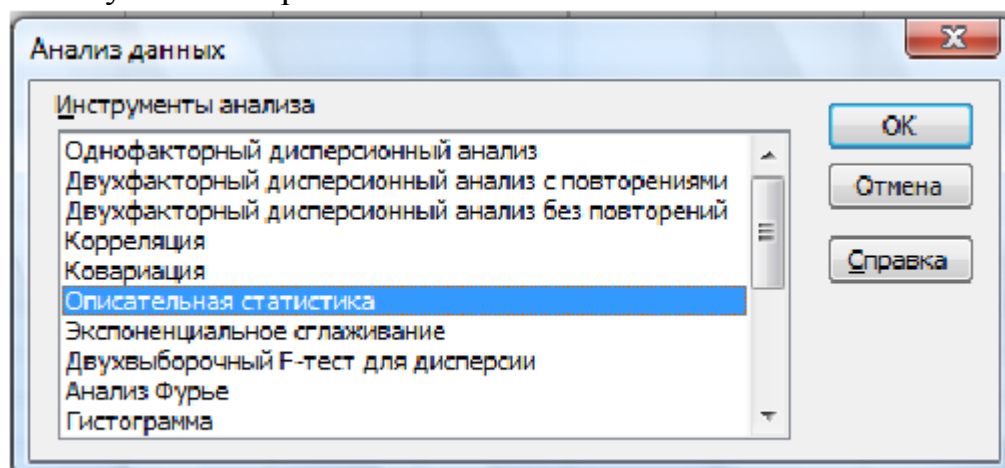
3	5	4	4	5	8	2	3	1	6	6	1	2	5	5
4	4	4	3	4	5	5	2	2	3	4	3	2	4	4
8	10	1	4	3	3	2	5	7	5	3	6	7	5	6
1	4	6	4	5	4	5	7	6	5	3	5	5	8	7
7	5	5	4	5	3	3	6	3	5	2	2	2	6	2
5	6	8	4	4	8	3	6	4	4	5	5	7	5	5
3	5	4	5	5	4	7	6	9	3	3	5	6	6	3
4	5	2	6	7	5	5	4	2	5	4	2	6	2	7
5	5	8	5	3	5	2	5	3	7	4	6	3	6	0
4	4	4	5	2	7	7	3	1	1	3	6	5	7	6

	A	B
1	день	результаты измерений
2	1	3
3	2	4
4	3	8
5	4	1
6	5	7
7	6	5
8	7	4
9	8	5
10	9	4
11	10	4

Рассчитаем статистические характеристики экспериментальных данных.

Для этого воспользуемся соответствующим инструментом из «Пакета анализа».

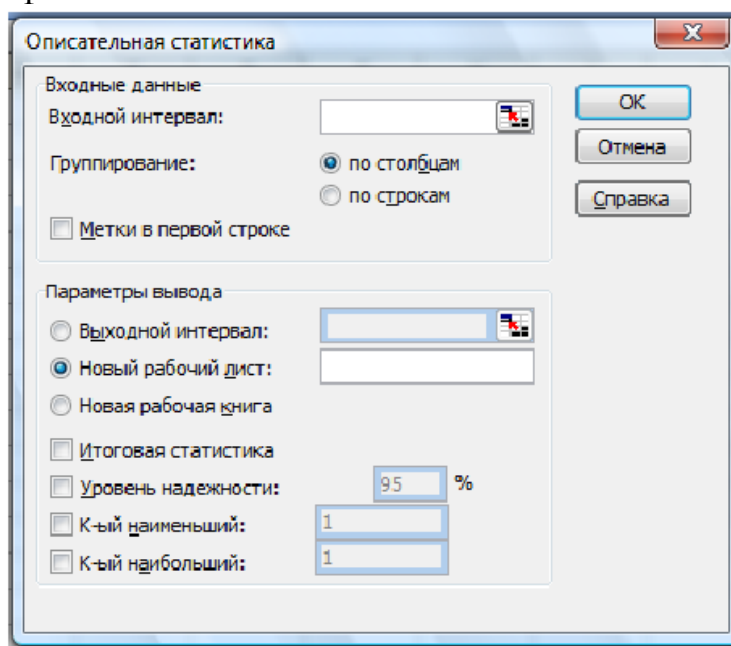
Откройте окно «Анализ данных» (позиция меню «Сервис» => «Анализ данных...»), выберите строку «Описательная статистика» и нажмите кнопку «ОК». Откроется окно «Описательная статистика». Назначение полей этого окна можно узнать в справочной системе Microsoft Excel.



Окно «АНАЛИЗ ДАННЫХ»

В окне «Описательная статистика» следует задать следующие параметры:

- «Входной интервал» – диапазон ячеек с результатами измерений (столбец В);
- «Параметры вывода» – выбрать позицию «Выходной интервал» и указать ячейку (какую пожелаете), с которой будет вставлен блок рассчитанных статистических характеристик;
- установить флажок «Итоговая статистика»;
- установить флажок «Уровень надежности» со значением 95 %;
- установить флажок «К-ый наименьший» со значением 2;
- установить флажок «К-ый наибольший» со значением 2.



Окно ОПИСАТЕЛЬНАЯ СТАТИСТИКА

*Входной интервал(диапазон).* Ссылка на диапазон, содержащий анализируемые данные. Ссылка должна состоять не менее чем из двух смежных диапазонов данных, данные в которых расположены по строкам или столбцам.

*Группирование.* Установите переключатель в положение «По столбцам» или «По строкам» в зависимости от расположения данных во входном диапазоне.

*Метки в первой строке/Метки в первом столбце.* Если первая строка исходного диапазона содержит названия столбцов, установите переключатель в положение Метки в первой строке. Если названия строк находятся в первом столбце входного диапазона, установите переключатель в положение Метки в первом столбце. Если входной диапазон не содержит меток, то необходимые заголовки в выходном диапазоне будут созданы автоматически.

*Уровень надежности.* Установите флажок, если в выходную таблицу необходимо включить строку для уровня надежности. В поле введите требуемое значение. Например, значение 95% вычисляет уровень надежности среднего с уровнем значимости 0,05.

*К-ый наибольший.* Установите флажок, если в выходную таблицу необходимо включить строку для k-го наибольшего значения для каждого диапазона данных. В соответствующем окне введите число k. Если k равно 1, эта строка будет содержать максимум из набора данных.

*К-ый наименьший.* Установите флажок, если в выходную таблицу необходимо включить строку для k-го наименьшего значения для каждого диапазона данных. В соответствующем окне введите число k. Если k равно 1, эта строка будет содержать минимум из набора данных.

**Выходной диапазон.** Введите ссылку на левую верхнюю ячейку выходного диапазона. Этот инструмент анализа выводит два столбца сведений для каждого набора данных. Левый столбец содержит метки статистических данных; правый столбец содержит статистические данные. Состоящий из двух столбцов диапазон статистических данных будет выведен для каждого столбца или для каждой строки входного диапазона в зависимости от положения переключателя Группирование.

Если хотим вывести результаты расчета на новый лист, то установите переключатель, чтобы открыть новый лист в книге и вставить результаты анализа, начиная с ячейки A1. Если в этом есть необходимость, введите имя нового листа в поле, расположенном напротив соответствующего положения переключателя.

Если хотим вывести результаты расчета в новой книге, то установите переключатель, чтобы открыть новую книгу и вставить результаты анализа в ячейку A1 на первом листе в этой книге.

**Итоговая статистика.** Установите флажок, если в выходном диапазоне необходимо получить по одному полю для каждого из следующих видов статистических данных, представленных в таблице

Наименование показателя при расчетах с использованием «Описательной статистики»	Примечания
Среднее	Среднее значение
Стандартная ошибка	Абсолютная ошибка выборки
Медиана	Число, которое является серединой множества чисел, то есть половина чисел имеют значения большие, чем медиана, а половина чисел имеют значения меньшие, чем медиана
Мода	Наиболее часто встречающееся или повторяющееся значение в массиве или интервале данных
Стандартное отклонение	Среднее квадратическое отклонение
Дисперсия выборки	Дисперсия
Эксцесс	Эксцесс
Асимметричность	Асимметрия
Интервал	Размах варьирования
Минимум	Минимум
Максимум	Максимум
Сумма	Сумма всех значений в выборке
Счет	Количество значений в выборке
Наибольший	k-тое наибольшее значение для каждого диапазона данных
Наименьший	k-тое наименьшее значение для каждого диапазона данных
Уровень надежности	Уровень надежности

Нажмите кнопку «ОК». На текущем листе, начиная с ячейки из поля «Выходной интервал» появятся рассчитанные статистические параметры.

### **Задания.**

1. Разобраться со смыслом показателей, полученных с помощью инструмента «Описательная статистика».
2. К таблице показателей, полученных с помощью инструмента «Описательная статистика», добавить справа еще один столбец, в котором рассчитать те же самые показатели с использованием встроенных статистических функций Microsoft Excel. Проверить их совпадение.

### Соответствие статистических показателей встроенным функциям

Статистический показатель	Встроенная функция Microsoft Excel
Среднее	СРЗНАЧ(диапазон)
Стандартная ошибка (среднего)	Расчет по формуле из приложения 1
Медиана	МЕДИАНА(диапазон)
Мода	МОДА(диапазон)
Стандартное отклонение	СТАНДОТКЛОН(диапазон)
Дисперсия выборки	ДИСП(диапазон)
Экцесс	ЭКСЦЕСС(диапазон)
Асимметричность	СКОС(диапазон)
Интервал	Расчет по формуле из приложения 1
Минимум	МИН(диапазон)
Максимум	МАКС(диапазон)
Сумма	СУММ(диапазон)
Счет	СЧЁТ(диапазон)
Наибольшее (k)	НАИБОЛЬШИЙ(диапазон; k)
Наименьшее (k)	НАИМЕНЬШИЙ(диапазон; k)
Уровень надежности	ДОВЕРИТ(альфа; станд_откл; размер) или по формуле из приложения 1

- По экспериментальным данным построить диаграмму изменения экспериментальных данных.

### 6.2. Построение гистограммы распределения частот

«Пакет анализа» позволяет быстро вычислить выборочные и интегральные частоты попадания данных в указанные интервалы (в терминологии Microsoft Excel – «карманы») значений, а также строить гистограммы распределения значений в выборке экспериментальных данных для каждого «кармана». Освоим эту возможность.

Построим гистограмму частот появления значений. Для этого предварительно необходимо сформировать список интервалов значений (границ интервалов, интервалы карманов). Выберем размер интервала, равный 1, то есть, определим частоту вхождения 0, 1, 2, .... Из полученных ранее данных известно, что минимальное значение, составляет 0, максимальное – 10. В свободных ячейках листа сформируйте список со значениями границ интервалов (от 0 до 10 с шагом 1).

Следует отметить, что для определения количества «карманов» гистограммы  $k$  рекомендуется использовать формулу Стерджесса:

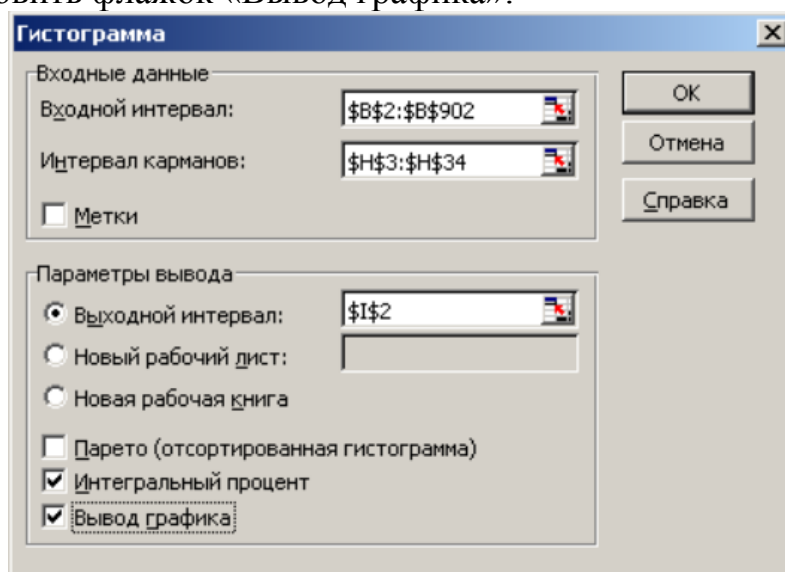
$$k = \sqrt[1 + 3.32 \cdot \lg_{10}(n)]{\quad},$$

где  $n$  – число экспериментальных данных;  $\sqrt{\dots}$  – операция округления вверх.

Откройте окно «Анализ данных» (позиция меню «Сервис» => «Анализ данных...»), выберите строку «Гистограмма» и нажмите кнопку «ОК». Откро-

ется окно «Гистограмма». Назначение полей этого окна можно узнать из помощи по Microsoft Excel. В окне «Гистограмма» следует задать следующие параметры:

- «Входной интервал» – диапазон ячеек с данными (столбец В);
- «Интервал карманов» – сформированный ранее диапазон ячеек со значениями границ интервалов;
- «Параметры вывода» – выбрать позицию «Выходной интервал» и указать ячейку (какую пожелаете), с которой будет вставлен блок рассчитанных частот;
- установить флажок «Интегральный процент»;
- установить флажок «Вывод графика».



Нажмите кнопку «ОК». На текущем листе, начиная с ячейки из поля «Выходной интервал» появятся рассчитанные выборочные и интегральные частоты, а также будет построена гистограмма частот.

### Задание

Даны значения процента содержания Na в интактных зубах у пациентов: 0,48; 0,56; 0,54; 0,57; 0,47; 0,5; 0,59; 0,6; 0,67; 0,68; 0,7; 0,69; 0,74; 0,75; 0,53; 0,58; 0,86; 0,51; 0,88; 0,6; 0,87; 0,65; 0,69; 0,71; 0,68; 0,5; 0,61; 0,76; 0,77; 0,61; 0,85; 0,59; 0,88; 0,64; 0,51; 0,86; 0,91; 0,78; 0,52; 0,49; 0,81; 0,55; 0,62; 0,63; 0,73; 0,72; 0,72; 0,66; 0,8; 0,79; 0,82; 0,84; 0,75; 0,83; 0,84; 0,83; 0,72; 0,73; 0,73; 0,62; 0,67; 0,81; 0,63; 0,84; 0,64; 0,66; 0,67; 0,67; 0,66; 0,68; 0,71; 0,76; 0,63; 0,66; 0,64; 0,66; 0,65; 0,68; 0,76; 0,78; 0,77; 0,68; 0,72; 0,73; 0,74; 0,79; 0,78; 0,77; 0,76; 0,7; 0,69; 0,72; 0,73; 0,69; 0,71; 0,68; 0,7; 0,71; 0,75; 0,69.

Определите основные статистические показатели с помощью инструментов «Описательная статистика» и «Гистограмма».