

Лабораторная работа.

Использование электронных таблиц MS Excel для вычисления статистических оценок параметров статистических совокупностей

Цель работы – научить находить с помощью MS Excel выборочные характеристики статистических данных.

Математическая статистика подразделяется на две основные области: описательную и аналитическую статистику. **Описательная статистика** охватывает методы описания статистических данных, представления их в форме таблиц, распределений. Аналитическая статистика или теория статистических выводов ориентирована на обработку данных, полученных в ходе эксперимента, с целью формулировки выводов, имеющих прикладное значение для самых различных областей человеческой деятельности.

1. Использование MS Excel для вычисления числовых характеристик статистических совокупностей

2.1. Использование специальных функций MS Excel

В мастере функций MS Excel имеется ряд специальных функций, предназначенных для вычисления выборочных характеристик.

Функция **СУММ** (категория **Математические**) позволяет вычислить сумму значений.

Функция **СЧЕТ** (категория **Статистические**) вычисляет объём выборки n (количество элементов в выборке).

Функция **СРЗНАЧ** (категория **Статистические**) вычисляет среднее арифметическое из нескольких массивов (аргументов) чисел. Аргументы *число1*, *число2*, ... – это от 1 до 30 массивов, для которых вычисляется среднее.

Функция **ДИСП.В** (категория **Статистические**) позволяет оценить несмещенную дисперсию по выборочным данным.

Функция **СТАНДОТКЛОН.В** (категория **Статистические**) вычисляет несмещенное среднее квадратичное (стандартное) отклонение (S).

Функция **МОДА.ОДН** (категория **Статистические**) вычисляет наиболее часто встречающееся значение в выборке.

Функция **МЕДИАНА** (категория **Статистические**) позволяет получать медиану заданной выборки. Медиана – это элемент выборки, число элементов выборки со значениями больше которого и меньше которого равно.

Функция **КВАРТИЛЬ.ВКЛ** (категория **Статистические**) вычисляет квартили распределения. Функция имеет формат **КВАРТИЛЬ.ВКЛ(массив; значение)**:

Массив – интервал ячеек, содержащих значения СВ.

Значение определяет, какая квартиль должна быть найдена (0 – минимальное значение, 1 – нижняя квартиль, 2 – медиана, 3 – верхняя квартиль, 4 – максимальное значение распределения).

Функция **ЭКССЕСС** (категория **Статистические**) вычисляет оценку эксцесса по выборочным данным.

Функция **СКОС** (категория **Статистические**) позволяет оценить асимметрию выборочного распределения.

Ошибка выборочной средней вычисляется по формуле $m = \frac{S}{\sqrt{n}}$ с помощью стандартных функций и может быть представлена в следующем виде: = СТАНДОТКЛОН.В (A1:AN) / КОРЕНЬ (СЧЕТ(A1:AN)), где функция **КОРЕНЬ** находится в категории **Математические**.

Доверительный интервал для среднего вычисляется с помощью функции **ДОВЕРИТ** (категория **Статистические**). Функция **ДОВЕРИТ** встречается в двух вариациях:

1) **ДОВЕРИТ.НОРМ** применяется в случае, когда среднее квадратическое отклонение σ известно, например, если измерения производятся одним и тем же прибором при одних и тех же условиях;

2) **ДОВЕРИТ.СТЮДЕНТ**, когда среднее квадратическое отклонение σ неизвестно. В обоих случаях функции имеют одинаковый формат:

Альфа – уровень значимости (например, 0,05).

Станд_откл вычисляется с помощью функции **СТАНДОТКЛОН.В**.

Размер – объём выборки, вычисляется с помощью функции **СЧЕТ**.

Пример. Провести статистический анализ методом описательной статистики доходов населения в регионе 1 и регионе 2.

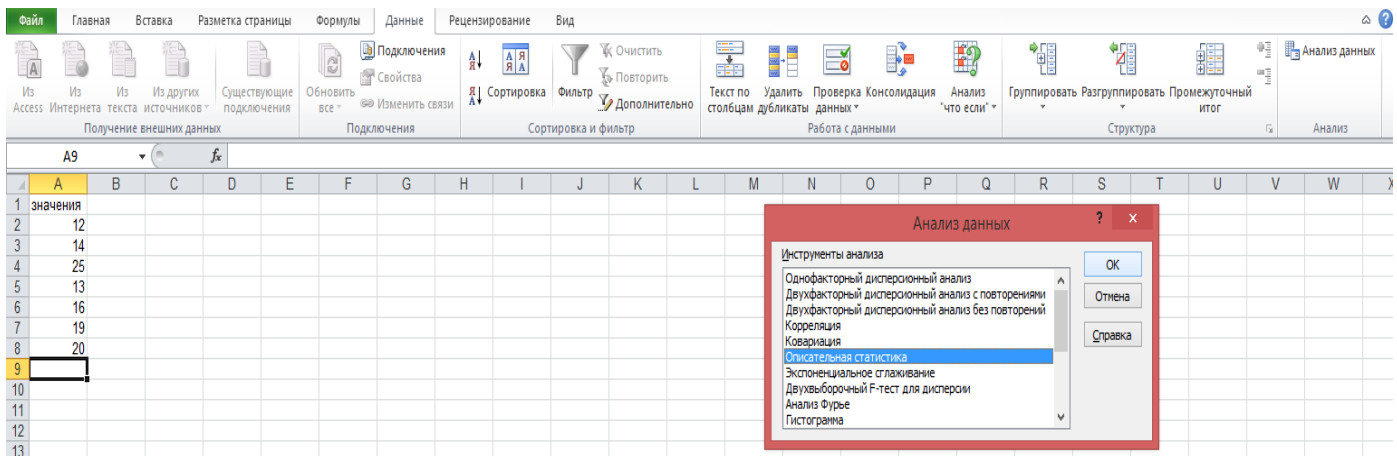
1 регион	2 регион
1	49
1	49
1	49
1	49
1	49
1	51
1	51
1	51

1 регион	2 регион	
1	51	
491	51	
500	500	сумма
10	10	счет
50	50	среднее
24010	1,11	дисперсия
154,95	1,05	стандартное отклонение
1	49	мода
1	50	медиана
1	49	квартили
1	51	квартили
10	-2,57	эксцесс
3,16	0	скос (асимметрия)
49	0,33	ошибка выборочной средней
50 ± 110,84	50 ± 0,75	доверительный интервал

2. Использование инструмента Пакет анализа MS Excel

Пакет анализа Excel может быть использован для решения задач статистической обработки выборочных данных. Для использования статистического пакета анализа данных необходимо:

- указать курсором мыши на вкладку ленты **Данные** и щелкнуть левой кнопкой мыши;
- в крайнем правом углу вкладки **Данные** выбрать команду **Анализ данных** (если команда Анализ данных отсутствует, то необходимо установить в Excel пакет анализа данных);



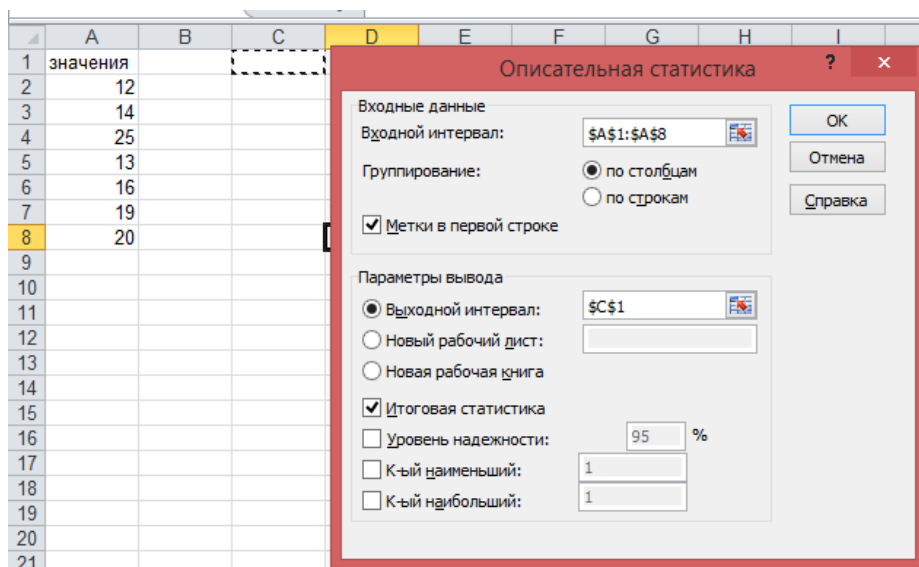
- выбрать строку **Описательная статистика** и нажать кнопку **ОК**:

Входной интервал – диапазон данных (ввести ссылки на ячейки, содержащие анализируемые данные).

Выходной интервал вводят ссылку на ячейку, в которую будут выведены результаты анализа.

Группирование – переключатель в положение по столбцам или по строкам.

Итоговая статистика – установить флажок в этом поле и нажать **ОК**.



	A	B	C	D
1	значения		значения	
2	12			
3	14			
4	25		Среднее	17
5	13		Стандартная ошибка	1,745743
6	16		Медиана	16
7	19		Мода	#Н/Д
8	20		Стандартное отклонение	4,618802
9			Дисперсия выборки	21,33333
10			Эксцесс	-0,11309
11			Асимметричность	0,781453
12			Интервал	13
13			Минимум	12
14			Максимум	25
15			Сумма	119
16			Счет	7

	A	B	C	D	E	F	G	H	I	J
1	1 регион	2 регион								
2	1	49								
3	1	51								
4	1	49								
5	1	51								
6	1	49								
7	1	51								
8	1	49								
9	1	51								
10	1	49								
11	491	51								

Описательная статистика

Входные данные: \$A\$1:\$B\$11

Выходной интервал: \$D\$2

Группирование: по столбцам по строкам

Метки в первой строке

Параметры вывода: Выходной интервал: \$D\$2

Новый рабочий лист:

Новая рабочая книга:

Итоговая статистика

Уровень надежности: 95 %

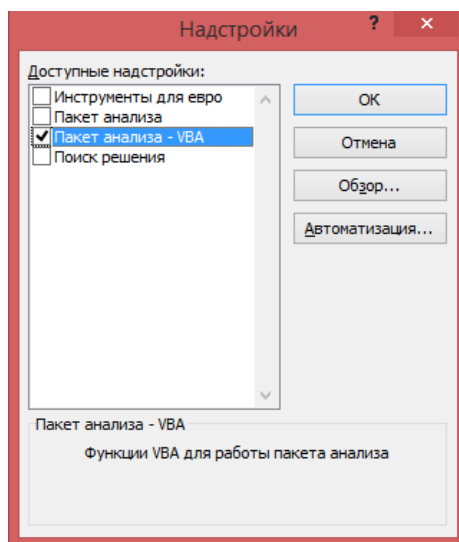
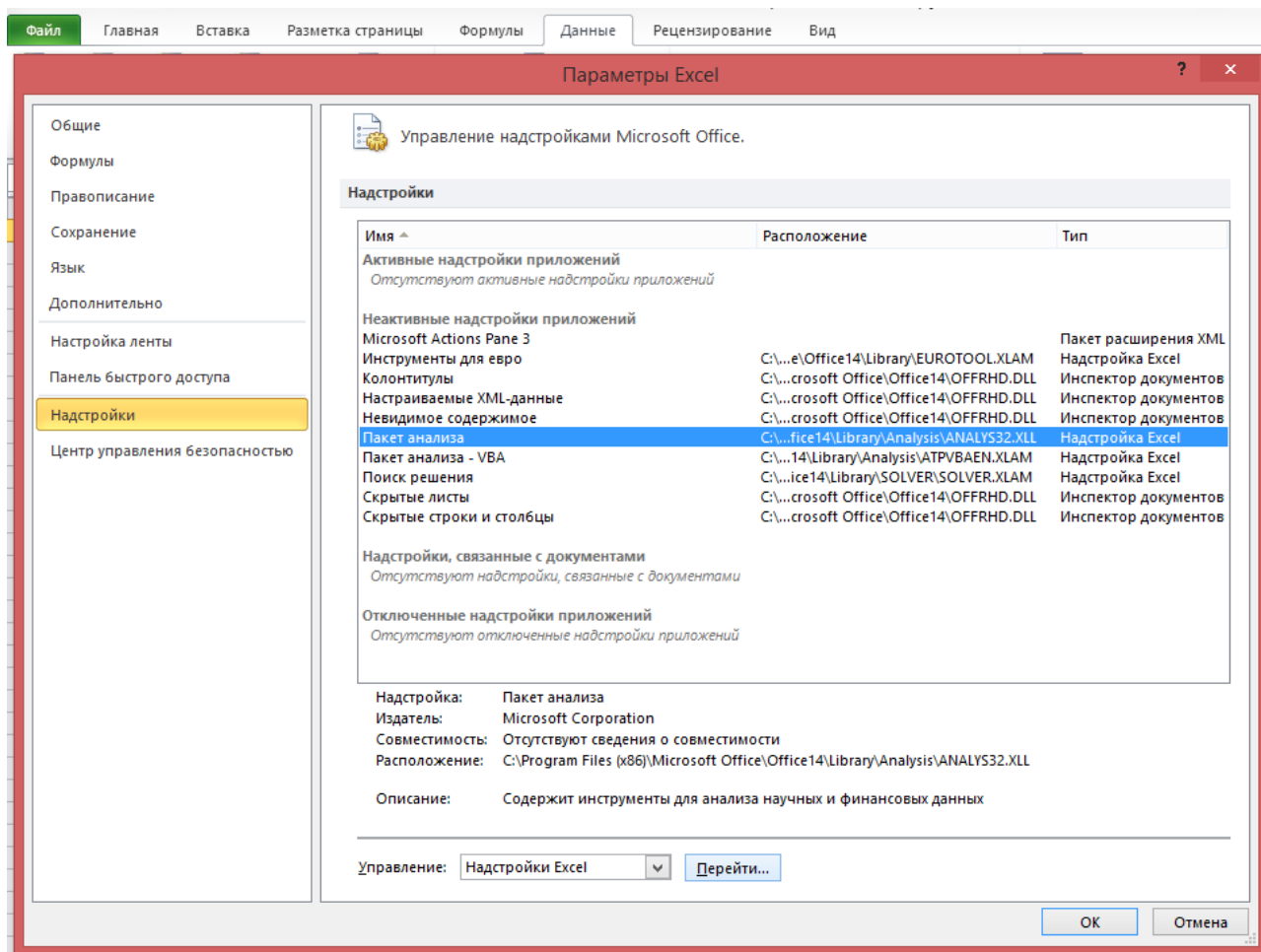
К-ый наименьший: 1

К-ый наибольший: 1

	A	B	C	D	E	F	G
1	1 регион	2 регион					
2	1	49		1 регион		2 регион	
3	1	51					
4	1	49		Среднее	50	Среднее	50
5	1	51		Стандартная ошибка	49	Стандартная ошибка	0,3333
6	1	49		Медиана	1	Медиана	50
7	1	51		Мода	1	Мода	49
8	1	49		Стандартное отклонение	154,95	Стандартное отклонение	1,0541
9	1	51		Дисперсия выборки	24010	Дисперсия выборки	1,1111
10	1	49		Эксцесс	10	Эксцесс	-2,571
11	491	51		Асимметрично	3,1623	Асимметрично	0
12				Интервал	490	Интервал	2
13				Минимум	1	Минимум	49
14				Максимум	491	Максимум	51
15				Сумма	500	Сумма	500
16				Счет	10	Счет	10
17				Уровень надежности (95,0%)	110,85	Уровень надежности (95,0%)	0,7541

Замечание. Если пакет **Анализ данных** не установлен в MS Excel, то для его установки необходимо выполнить следующие шаги:

- выбрать в пункте главного меню **Файл** команду **Параметры**, затем **Надстройки**;
- нажать в появившемся диалоговом окне **Перейти**;
- установить в появившемся списке флажок **Пакет анализа** и нажать **ОК**.



Задачи для самостоятельного решения

Задача 1. Даны измерения роста групп студентов: 175, 192, 183, 192, 188, 182, 176, 178, 187, 180, 188, 175, 178, 178, 190, 187, 188, 190, 185, 188, 185, 178, 180, 185, 182, 180, 187, 178, 178, 185, 188, 190, 192, 190, 178, 176, 178, 188, 195, 190, 185, 182, 185, 180, 178, 180, 190, 183, 185, 185, 188 и студенток 164, 160, 157,

166, 162, 160, 161, 159, 160, 163, 170, 171. Определить основные выборочные характеристики.

Задача 2. В начальной школе в одном из классов были получены следующие данные о массе девочек: 21,8; 24; 21,8; 20; 21,8; 20; 19,3; 20,8; 19,3; 23,8; 20; 24; 20; 19,3; 20; 21,8; 20,8; 20,8; 23,8; 23,8; 20; 20,8; 24; 24,5; 23,8; 23,8; 20,8; 20; 24; 24; 24,5; 19,3; 20,8; 20; 20,8; 24,5; 20; 20. Найти среднее значение, размах, моду и медиану.