

СТАТИСТИКА.

Описательная статистика.

Лекция 4. Использование Excel в описательной статистике.

**Авторы: Равичев Л.В., Ломакина И.А.
Кафедра менеджмента и маркетинга
РХТУ им. Д.И.Менделеева.
Москва - 2007**

Показатели описательной статистики

Показатели описательной статистики

Показатели положения

X_{\min}, X_{\max}

Q_1, Q_3

Средние

Мода

Медиана

Показатели разброса

Дисперсия

Стандарт

R

d_k

\bar{d}

Показатели асимметрии

A_x

Положение медианы относительно среднего

Показатели распределения

Полигон

Кумулята

Гистограмма

Экссесс

Функции Excel, используемые при расчете показателей положения

1. Функция МИН.

МИН(число1;число2;...).

Функция МИН нахо

D7		=МИН(B3:B10)			
	A	B	C	D	E
	Объем индивидуального жилищного строительства по районам в Ярославской области				
1					
2	Районы	Площадь, м ²			
3	Большесельский	718			
4	Борисоглебский	1319			
5	Гаврилов-Ямский	919			
6	Даниловский	1321			
7	Любимский	437	Минимальное значение=	437	
8	Некрасовский	2121			
9	Переславский	8872			
10	Тутаевский	730			
11					

Функции Excel, используемые при расчете показателей положения

2. Функция *НАИМЕНЬШИЙ*.

НАИМЕНЬШИЙ(массив;k).

Функция *НАИМЕНЬШИЙ* возвращает наименьшее значение из заданного массива (или диапазона ячеек). Если массив содержит несколько значений, то функция возвращает значение, которое находится в k-ом по порядку наименьшем месте.

D7		=НАИМЕНЬШИЙ(В3:В10;3)			
	A	B	C	D	E
1	Объем индивидуального жилищного строительства по районам в Ярославской области				
2	Районы	Площадь, м ²			
3	Большесельский	718			
4	Борисоглебский	1319			
5	Гаврилов-Ямский	919			
6	Даниловский	1321			
7	Любимский	437	Наименьшее третье	730	
8	Некрасовский	2121			
9	Переславский	8872			
10	Тутаевский	730			
11					

Функции Excel, используемые при расчете показателей положения

3. Функция МАКС.

МАКС(число1;число2;...).

Функция МАКС

D7		= =МАКС(B3:B10)			
	A	B	C	D	E
1	Объем индивидуального жилищного строительства по районам в Ярославской области				
2	Районы	Площадь, м ²			
3	Большесельский	718			
4	Борисоглебский	1319			
5	Гаврилов-Ямский	919			
6	Даниловский	1321			
7	Любимский	437	Максимальное=	8872	
8	Некрасовский	2121			
9	Переславский	8872			
10	Тутаевский	730			
11					

Функции Excel, используемые при расчете показателей положения

4. Функция **НАИБОЛЬШИЙ**.

НАИБОЛЬШИЙ(массив;k).

Функция **НАИБОЛЬШИЙ** возвращает наибольший (к-го) наибольший

D7		=		=НАИБОЛЬШИЙ(В3:В10;4)	
	A	B	C	D	
1	Объем индивидуального жилищного строительства по районам в Ярославской области				
2	Районы	Площадь, м ²			
3	Большесельский	718			
4	Борисоглебский	1319			
5	Гаврилов-Ямский	919			
6	Даниловский	1321			
7	Любимский	437	Наибольший четвертый=	1319	
8	Некрасовский	2121			
9	Переславский	8872			
10	Тутаевский	730			
11					

Функции Excel, используемые при расчете показателей положения

5. Функция КВАРТИЛЬ.

КВАРТИЛЬ(массив;k).

Функция КВАРТИЛЬ расчитывает значение k-го члена ряда.

Функция КВАРТИЛЬ расчитывает значение k-го члена ряда.

- минимальное значение
- **первую квартиль**, если k=1
- значение медианы, если k=2
- **третью квартиль**, если k=3
- максимальное значение

Функция КВАРТИЛЬ
данных она проводит

D17 =				
	A	B	C	D
1	Спрос на спортивные костюмы "Reebok" в фирме "Чемпион" (за март 2003 г.)			
	Размер костюма	Число купленных костюмов	Ранжированный ряд	Накопленные частоты
2				
3	46	57	48	48
4	48	48	57	105
5	50	95	60	165
6	52	60	77	242
7	54	77	95	337
8				
9	Формула	Результат		
10	=КВАРТИЛЬ(B3:B7;0)	48		
11	=КВАРТИЛЬ(B3:B7;1)	57		
12	=КВАРТИЛЬ(B3:B7;2)	60		
13	=КВАРТИЛЬ(B3:B7;3)	77		
14	=КВАРТИЛЬ(B3:B7;4)	95		
15				

Функции Excel, используемые при расчете показателей положения

6. Функция СРЗНАЧ.

СРЗНАЧ(число1;число2;...).

Функция СРЗНАЧ возвращает арифметическое среднее значение аргументов. Аргументы могут быть числами, именами диапазонов, ссылками на диапазоны или массивами.

D7		= СРЗНАЧ(B3:B10)		
	A	B	C	D
1	Объем индивидуального жилищного строительства по районам в Ярославской области			
2	Районы	Площадь, м ²		
3	Большесельский	718		
4	Борисоглебский	1319		
5	Гаврилов-Ямский	919		
6	Даниловский	1321		
7	Любимский	437	Среднее значение=	2055
8	Некрасовский	2121		
9	Переславский	8872		
10	Тутаевский	730		
11				

Функции Excel, используемые при расчете показателей положения

7. Функция СРГАРМ.

СРГАРМ(число1;число2;...).

Функция СРГАРМ рассчитывает значение невзвешенной средней гармонической множества данных. На практике используется редко.

	А	В	С	Д	Е	Ф
1	Цена и сумма выручки от продажи CD "Шедевры русской живописи"					
2	Фирма	Цена x_i , руб.	Сумма выручки W_i , руб.			
3	"Нижига"	290	20300			
4	"Кирилл и Мефодий"	270	27000	Средняя гармоническая невзвешенная	118,42	=СРГАРМ(В3:В5)
5	"Рога и Кольга"	55	16500	Средняя гармоническая взвешенная	135,74	{=СУММ(С3:С5)/СУММ(С3:С5/В3:В5)}
6						

Функции Excel, используемые при расчете показателей положения

8. Функция СРГЕОМ.

СРГЕОМ(число1;число2;...).

Функция СРГЕОМ рассчитывает среднее арифметическое значение массива положительных чисел.

СВ				=	=СРГЕОМ(С4:С7)
	А	В	С		
1	Динамика производства молока в регионе за 1994-1998 гг.				
2	Год	Производство молока, тыс. т.	Коэффициент роста $K_{i/i-1}$		
3	1994	310,12			
4	1995	321,50	1,04		
5	1996	340,70	1,06		
6	1997	315,40	0,93		
7	1998	335,90	1,06		
8	Средний относительный темп роста		1,02		
9					

Функции Excel, используемые при расчете показателей положения

9. Функция МОДА.

МОДА(число1;число2;...).

Функция МОДА отображает наиболее часто встречающееся значение в интервале данных.

	А	В	С	Д	Е
1	Спрос на спортивные костюмы "Реебок" в фирме "Чемпион" (за март 2003 г.)				
2	Размер костюма	Число купленных костюмов			
3	46	57			
4	48	48	Мода=	#Н/Д	=МОДА(А3:А7)
5	50	95	Мода=	#Н/Д	=МОДА(В3:В7)
6	52	60	Мода=	48	=МОДА(А3:А7;В3:В7)
7	54	77	Мода=	50	=ИНДЕКС(А3:А7;ПОИСКПОЗ(МАКС(В3:В7);В3:В7))
8					

Функции Excel, используемые при расчете показателей положения

10. Функция МЕДИАНА.

МЕДИАНА(число1;число2;...).

Функция М
вариацион

D7	=	=МЕДИАНА(B3:B10)		
	A	B	C	D
	Объем индивидуального жилищного строительства по районам в Ярославской области			
1				
2	Районы	Площадь, м ²		
3	Любимский	437		
4	Большесельский	718		
5	Тутаевский	730		
6	Гаврилов-Ямский	919		
7	Борисоглебский	1319	Медиана=	1119,0
8	Даниловский	1321		
9	Некрасовский	2121		
10	Переславский	8872		
11				

Функции Excel, используемые при расчете показателей разброса

1. Функция ДИСП.

ДИСП(число1;число2;...).

Функция ДИСП

D7	=ДИСП(B3:B10)			
	A	B	C	D
1	Объем индивидуального жилищного строительства по районам в Ярославской области			
2	Районы	Площадь, м ²		
3	Любимский	437		
4	Большесельский	718		
5	Тутаевский	730		
6	Борисовский	910		

Функция ДИСП рассчитывает дисперсию при условии, что исходные данные образуют *выборочную* совокупность. В случае, если совокупность является *генеральной*, то необходимо воспользоваться функцией ДИСПР.

Функции Excel, используемые при расчете показателей разброса

2. Функция ДИСПР.

ДИСПР(число1;число2;

Функция ДИСПР вычисляет дисперсию совокупности.

Часто генеральную дисперсию

D7	=ДИСПР(B3:B19)			
	A	B	C	D
1	Объем индивидуального жилищного строительства по районам в Ярославской области			
2	Районы	Площадь, м ²		
3	Любимский	437		
4	Большесельский	718		
5	Тутаевский	730		
6	Гаврилов-Ямский	919		
7	Борисоглебский	1319	Дисперсия=	4513386
8	Даниловский	1321		
9	Некрасовский	2121		
10	Переславский	8872		
11	Брейтовский	632		
12	Мышкинский	218		
13	Некоузский	206		
14	Первомайский	457		
15	Пошехонский	3011		
16	Ростовский	1363		
17	Рыбинский	1389		
18	Угличский	4728		
19	Ярославский	3439		
20				

Функции Excel, используемые при расчете показателей разброса

3. Функция СТАНДОТКЛОН.

СТАНДОТКЛОН(число1;число2;...).

Функция СТАНДОТКЛОН (стандарт) по вы

The screenshot shows an Excel spreadsheet. The formula bar at the top displays the formula `=СТАНДОТКЛОН(В3:В10)` in cell D7. Below the formula bar is a table with the following content:

	А	В	С	Д
1	Объем индивидуального жилищного строительства по районам в Ярославской области			
2	Районы	Площадь, м ²		
3	Любимский	437		
4	Большесельский	718		
5	Тутаевский	730		
6				

Функция СТАНДОТКЛОН рассчитывает стандарт при условии, что исходные данные образуют *выборочную* совокупность. В случае, если совокупность является *генеральной*, то необходимо воспользоваться функцией СТАНДОТКЛОНП.

Функции Excel, используемые при расчете показателей разброса

4. Функция СТАНДОТКЛОНП.

СТАНДОТКЛОНП(число1;число2;...).

Функция СТАНДОТКЛО
генеральной совокупности

D7		= СТАНДОТКЛОНП(B3:B19)		
	A	B	C	D
1	Объем индивидуального жилищного строительства по районам в Ярославской области			
2	Районы	Площадь, м ²		
3	Любимский	437		
4	Большесельский	718		
5	Тутаевский	730		
6	Гаврилов-Ямский	919		
7	Борисоглебский	1319	Стандарт=	2124
8	Даниловский	1321		
9	Некрасовский	2121		
10	Переславский	8872		
11	Брейтовский	632		
12	Мьшкиский	218		
13	Некоузский	206		
14	Первомайский	457		
15	Пошехонский	3011		
16	Ростовский	1363		
17	Рыбинский	1389		
18	Угличский	4728		
19	Ярославский	3439		
20				

Функции Excel, используемые при расчете показателей разброса

5. Функция СРОТКЛ.

СРОТКЛ(число1;число2;...).

Функция
жества д

D7	=	=СРОТКЛ(B3:B10)		
	A	B	C	D
1	Объем индивидуального жилищного строительства по районам в Ярославской области			
2	Районы	Площадь, м ²		
3	Любимский	437		
4	Большесельский	718		
5	Тутаевский	730		
6	Гаврилов-Ямский	919		
7	Борисоглебский	1319	Среднее линейное отклонение=	1721
8	Даниловский	1321		
9	Некрасовский	2121		
10	Переславский	8872		
11				

Функция Excel, используемая при расчете показателя асимметрии

Функция СКОС.

СКОС(число1;число2;...).

Функция СКОС оце

$$A_x =$$

Если данные образуют асимметрию необходимо

D7	=	=СКОС(B3:B10)		
	A	B	C	D
1	Объем индивидуального жилищного строительства по районам в Ярославской области			
2	Районы	Площадь, м ²		
3	Тутаевский	730		
4	Переславский	8872		
5	Некрасовский	2121		
6	Любимский	437		
7	Даниловский	1321	Асимметричность=	2,64
8	Гаврилов-Ямский	919		
9	Борисоглебский	1319		
10	Большесельский	718		
11				

Функция Excel, используемая при расчете показателя распределения

Функция ЭКСЦЕСС.

ЭКСЦЕСС(число1;число2;...).

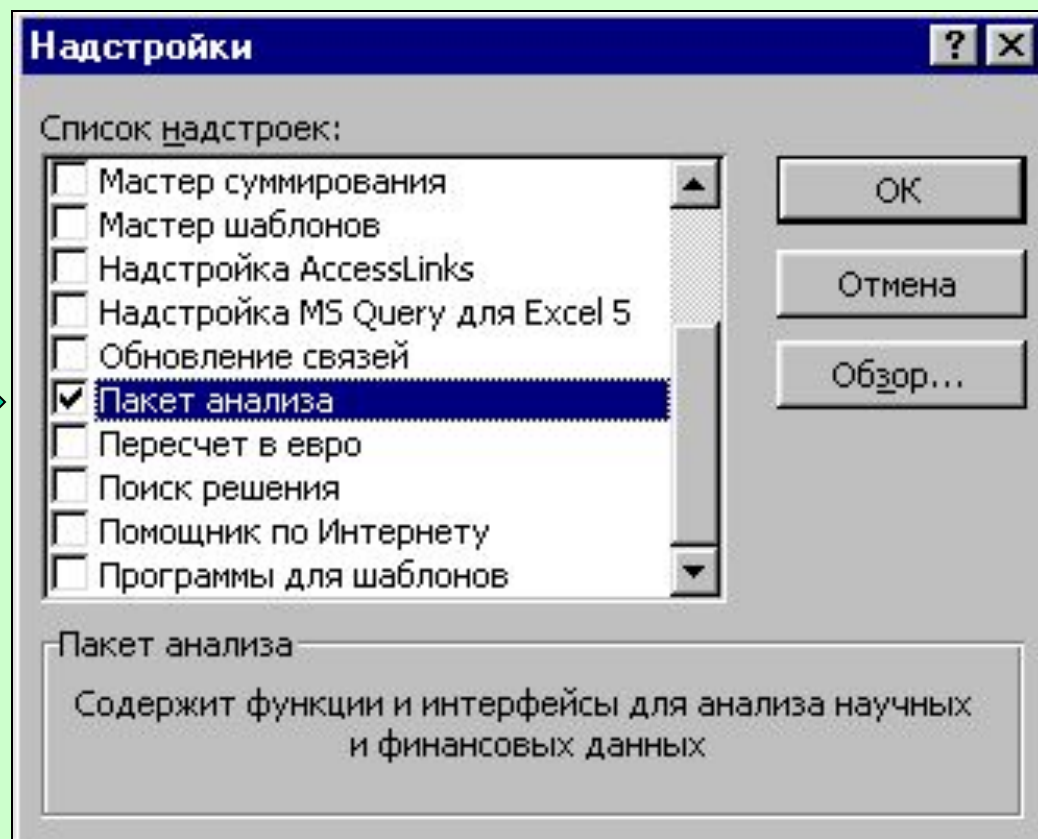
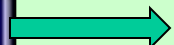
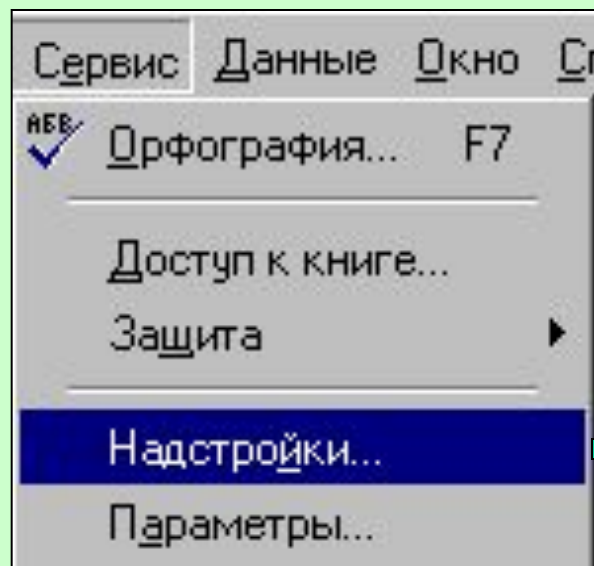
Функция ЭКСЦЕСС оценивает

$$E_x = \left[\frac{n(n+1)}{(n-1)(n-2)} \right]$$

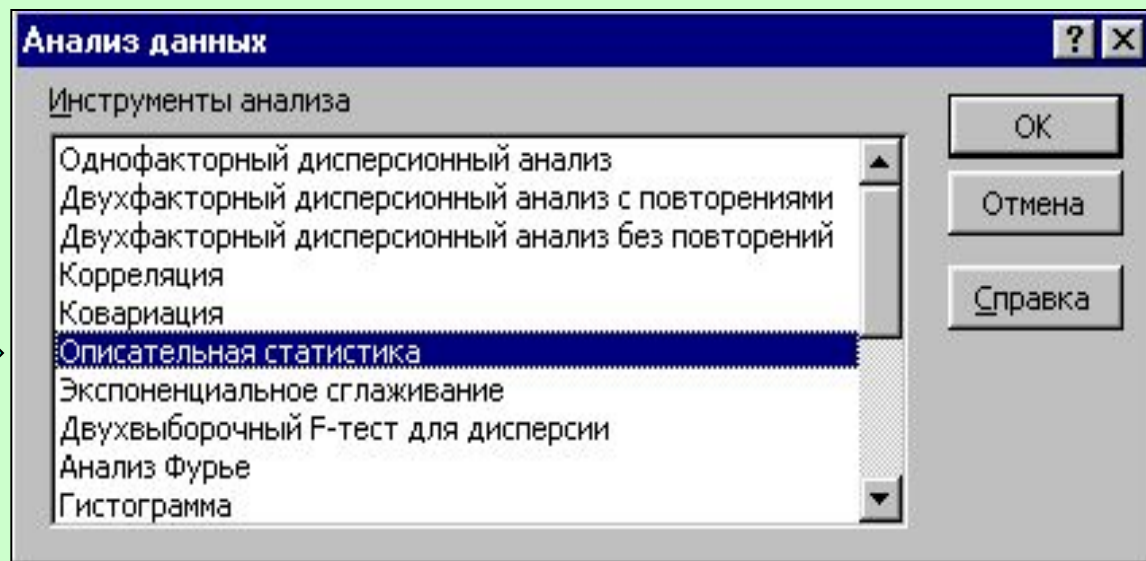
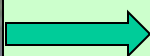
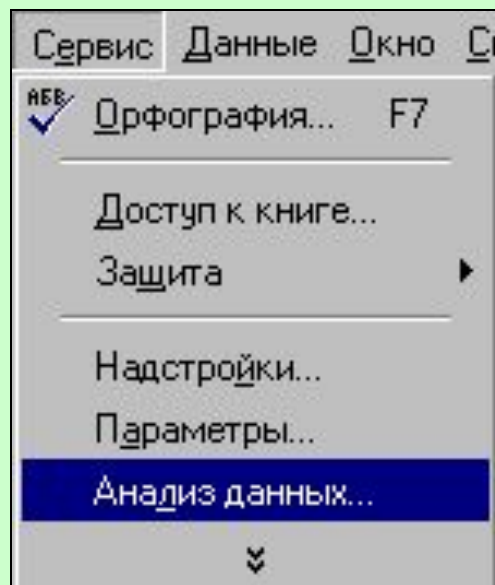
Если данные образуют не эксцесс необходимо расче

D7	= ЭКСЦЕСС(B3:B10)			
	A	B	C	D
1	Объем индивидуального жилищного строительства по районам в Ярославской области			
2	Районы	Площадь, м ²		
3	Тутаевский	730		
4	Переславский	8872		
5	Некрасовский	2121		
6	Любимский	437		
7	Даниловский	1321	Эксцесс=	7,18
8	Гаврилов-Ямский	919		
9	Борисоглебский	1319		
10	Большесельский	718		
11				

Подключение режима «Анализ данных»



Выход в режим «Описательная статистика»



Справочная информация по технологии работы в режиме «Описательная статистика»

The image shows a screenshot of a software interface with two windows. The background window is titled 'Анализ данных' and contains a list of analysis tools under the heading 'Инструменты анализа'. The 'Описательная статистика' tool is selected and highlighted in blue. The foreground window is titled 'Описательная статистика' and contains the following settings:

- Входные данные**
 - Входной интервал: [Empty text box]
 - Группирование: по столбцам, по строкам
 - Метки в первой строке
- Параметры вывода**
 - Выходной интервал: [Empty text box]
 - Новый рабочий лист: [Empty text box]
 - Новая рабочая книга
 - Итоговая статистика
 - Уровень надежности: [95] %
 - К-ый наименьший: [1]
 - К-ый наибольший: [1]

Buttons on the right side of the dialog include 'OK', 'Отмена', and 'Справка'.

Справочная информация по технологии работы в режиме «Описательная статистика»

	А	В	С	Д	Е	Ф	Г	Н	І
1	Объем индивидуального жилищного строительства по районам в Ярославской области								
2	Районы	Площадь, м ²							
3	Тутаевский	730							
4	Переславский	8872							
5	Некрасовский	2121							
6	Любимский	437							
7	Даниловский	1321							
8	Гаврилов-Ямский	919							
9	Борисоглебский	1319							
10	Большесельский	718							
11									
12									
13									
14									
15									
16									
17									

Описательная статистика

Входные данные

Входной интервал:

Группирование: по столбцам
 по строкам

Метки в первой строке

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Итоговая статистика

Уровень надежности: %

К-ый наименьший:

К-ый наибольший:

OK
Отмена
Справка

Справочная информация по технологии работы в режиме «Описательная статистика»

Средняя ошибка выборки
(показатель **Стандартная ошибка**)

$$\mu_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

=E7/КОРЕНЬ(E15)

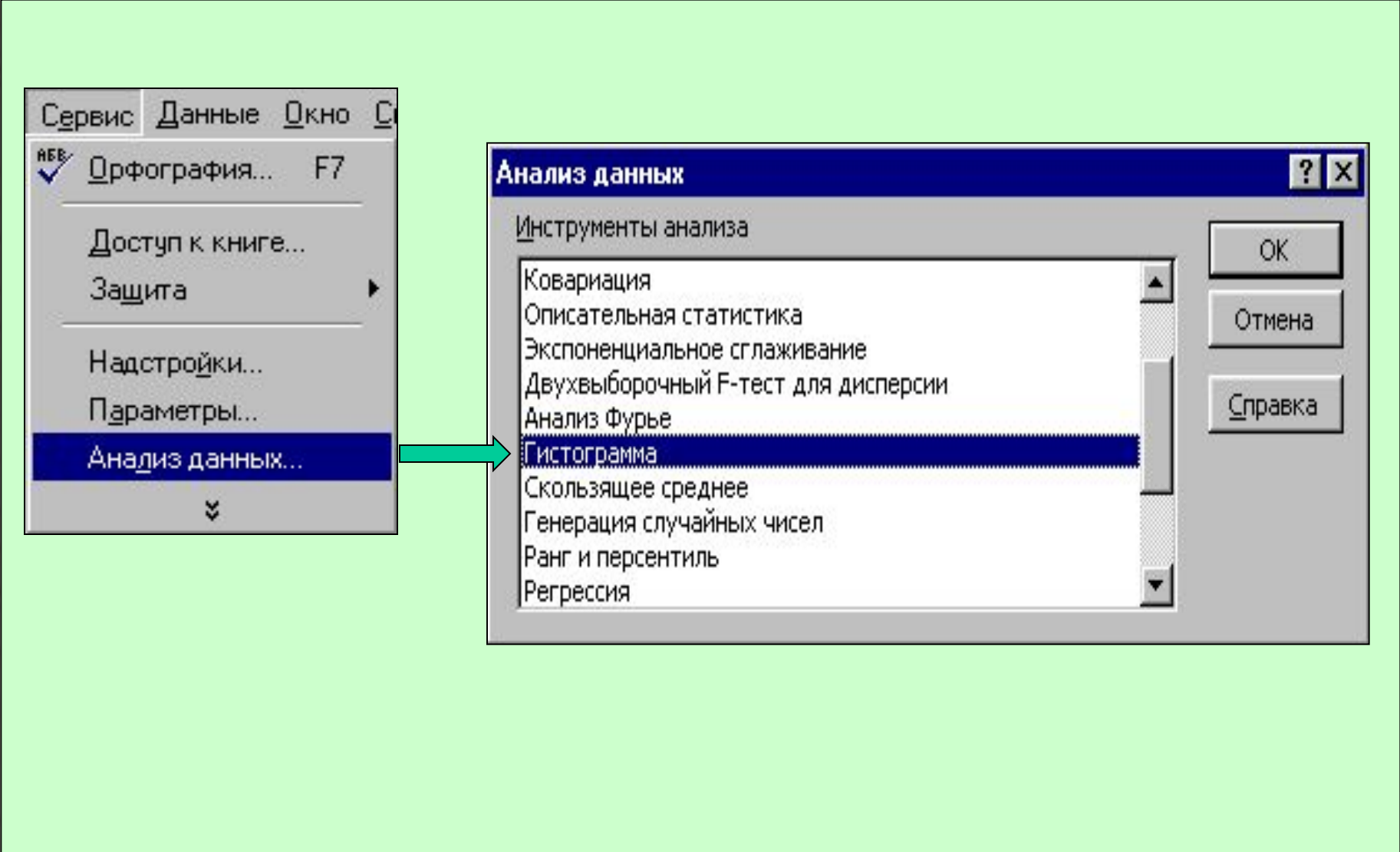
Предельная ошибка выборки
(показатель **Уровень надежности**)

$$\Delta_{\bar{x}} = t\mu_{\bar{x}}$$

=СТЮДРАСПОБР(0,05;E15-1)*E4

			C	D	E
				Площадь, м2	
				Среднее	2054,625
				Стандартная ошибка	991,1025
				Медиана	1119
				Мода	#Н/Д
				Стандартное отклонение	2803,261
				Дисперсия выборки	7858273
				Эксцесс	7,178297
				Асимметричность	2,642829
				Интервал	8435
				Минимум	437
				Максимум	8872
				Сумма	16437
				Счет	8
				Наибольший(1)	8872
				Наименьший(2)	718
				Уровень надежности(95,0%	2343,583

Справочная информация по технологии работы в режиме «Гистограмма»



Справочная информация по технологии работы в режиме «Гистограмма»

Режим *Гистограмма* служит для вычисления частот попадания данных в указанные границы **интервального**

- Анализ данных
- Инструменты анализа
- Ковариация
- Описательная статистика
- Экспоненциальное сглаживание
- Двухвыборочный F-тест
- Анализ Фурье
- Гистограмма**
- Скользящее среднее
- Генерация случайных чисел
- Ранг и перцентиль
- Регрессия

Гистограмма [?] [X]

Входные данные

Входной интервал: [] []

Интервал карманов: [] []

Метки

Параметры вывода

Выходной интервал: [] []

Новый рабочий лист: [] []

Новая рабочая книга

Парето (отсортированная гистограмма)

Интегральный процент

Вывод графика

OK

Отмена

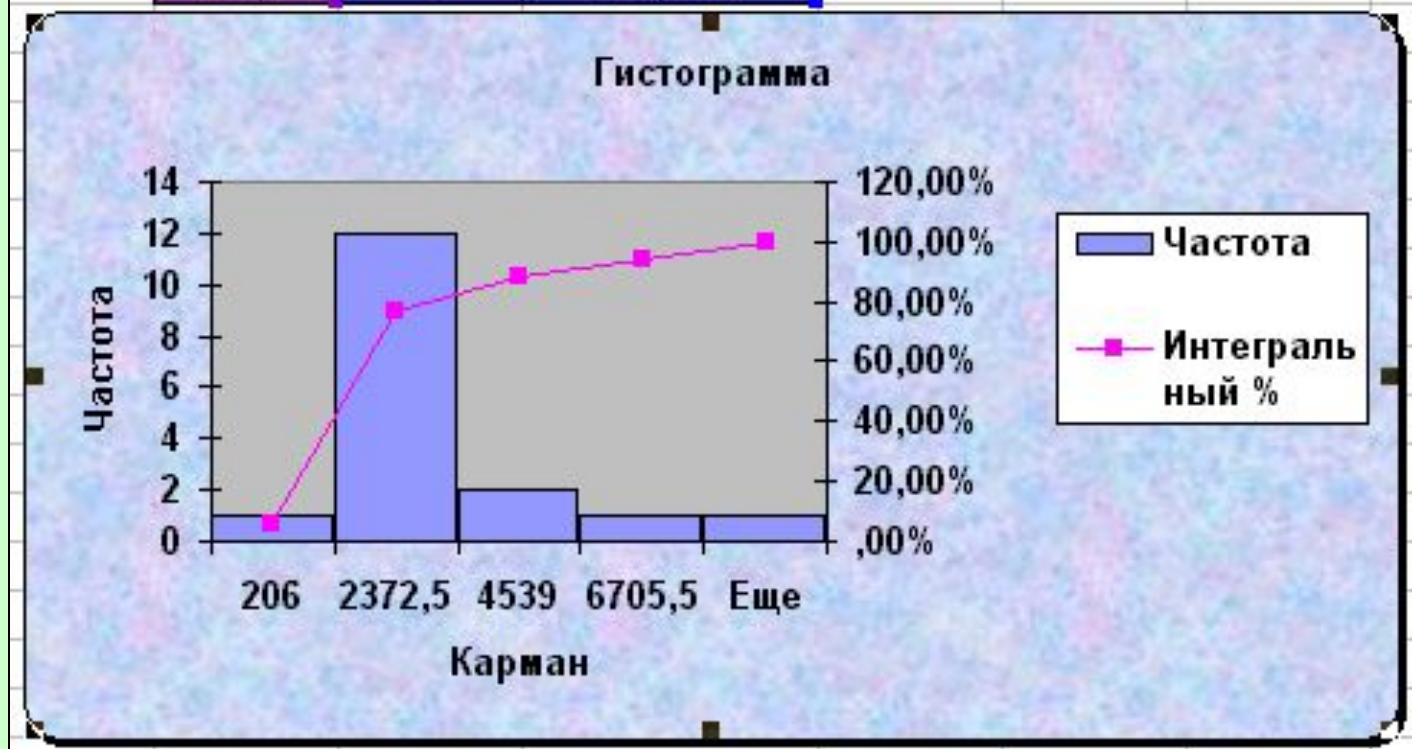
Справка

Справочная информация по технологии работы в режиме «Гистограмма»

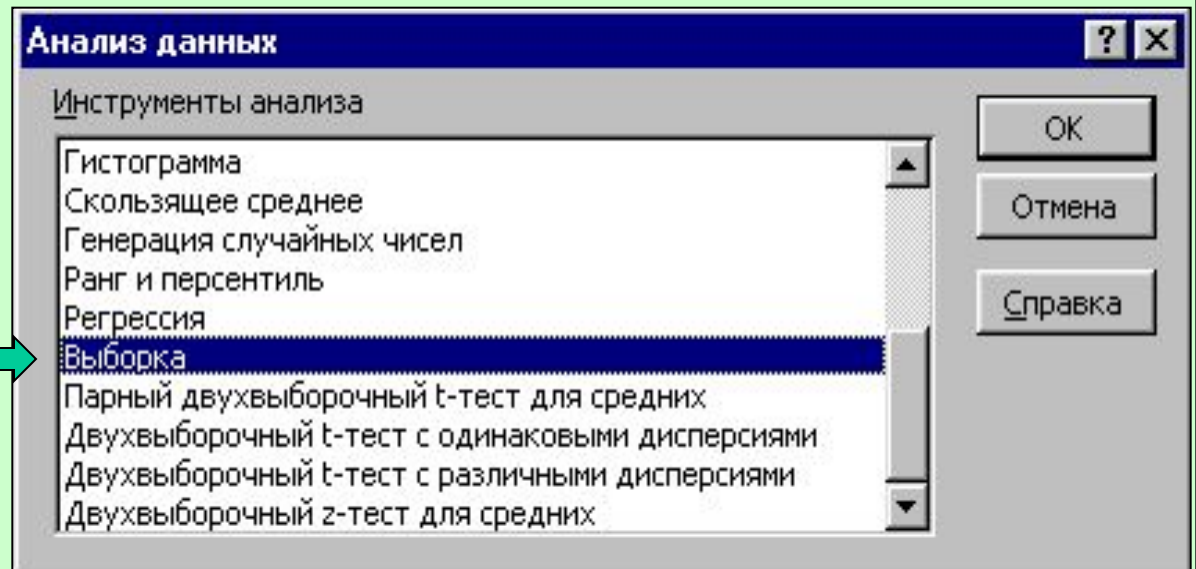
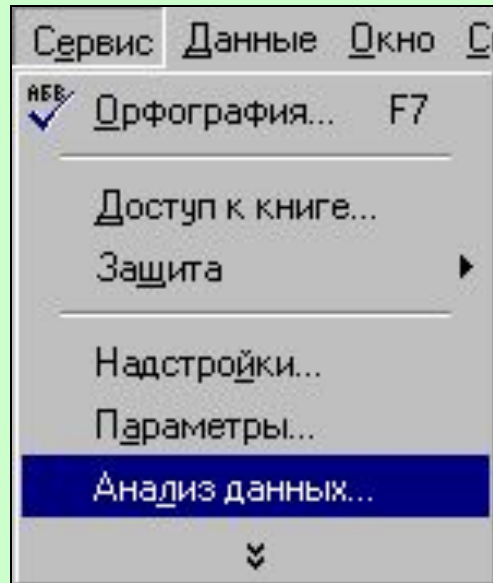
	А	В	С	Д	Е	Ф	Г	Н
1	Объем индивидуального жилищного строительства по районам в Ярославской области		<div data-bbox="749 332 1856 1119"> <h3>Гистограмма</h3> <p>Входные данные</p> <p>Входной интервал: <input type="text" value="\$B\$2:\$B\$19"/></p> <p>Интервал карманов: <input type="text"/></p> <p><input checked="" type="checkbox"/> Метки</p> <hr/> <p>Параметры вывода</p> <p><input checked="" type="radio"/> Выходной интервал: <input type="text" value="\$D\$2"/></p> <p><input type="radio"/> Новый рабочий лист:</p> <p><input type="radio"/> Новая рабочая книга</p> <p><input type="checkbox"/> Парето (отсортированная гистограмма)</p> <p><input checked="" type="checkbox"/> Интегральный процент</p> <p><input checked="" type="checkbox"/> Вывод графика</p> <p>OK Отмена Справка</p> </div>					
2	Районы	Площадь, м ²						
3	Любимский	437						
4	Большесельский	718						
5	Тутаевский	730						
6	Гаврилов-Ямский	919						
7	Борисоглебский	1319						
8	Даниловский	1321						
9	Некрасовский	2121						
10	Переславский	8872						
11	Брейтовский	632						
12	Мышкинский	218						
13	Некоузский	206						
14	Первомайский	457						
15	Пошехонский	3011						
16	Ростовский	1363						
17	Рыбинский	1389						
18	Угличский	4728						
19	Ярославский	3439						

Справочная информация по технологии работы в режиме «Гистограмма»

Карман	Частота	Интегральный %
206	1	5,88%
2372,5	12	76,47%
4539	2	88,24%
6705,5	1	94,12%
Еще	1	100,00%



Справочная информация по технологии работы в режиме «Выборка»



Справочная информация по технологии работы в режиме «Выборка»

Режим *Выборка* служит для формирования выборки из генеральной совокупности на основе исходных данных.

Анализ данных

Инструменты анализа

- Гистограмма
- Скользящее среднее
- Генерация случайных чисел
- Ранг и перцентиль
- Регрессия
- Выборка**
- Парный двухвыборочный t-тест
- Двухвыборочный t-тест с одним
- Двухвыборочный t-тест с разл
- Двухвыборочный z-тест для с

Выборка ? X

Входные данные

Входной интервал:

Метки

Метод выборки

Периодический

Период:

Случайный

Число выборок:

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

OK

Отмена

Справка

Справочная информация по технологии работы в режиме «Выборка»

	A	B	C	D	E	F	G	H
1	Объем индивидуального жилищного строительства по районам в Ярославской области							
2	Районы	Площадь, м²		718				
3	Любимский	437		919				
4	Большесельский	718		1321				
5	Тутаевский	730		8872				
6	Гаврилов-Ямский	919		218				
7	Борисоглебский	1319		457				
8	Даниловский	1321		1363				
9	Некрасовский	2121						
10	Переславский	8872						
11	Брейтовский	632						
12	Мышкинский	218						
13	Некоузский	206						
14	Первомайский	457						
15	Пошехонский	3011						
16	Ростовский	1363						
17								

Справочная информация по технологии работы в режиме «Выборка»

	A	B	C	D	E	F	G	H	
1	Объем индивидуального жилищного строительства по районам в Ярославской области								
2	Районы	Площадь, м²		3011					
3	Любимский	437		218					
4	Большесельский	718		919					
5	Тутаевский	730		1321					
6	Гаврилов-Ямский	919		206					
7	Борисоглебский	1319		8872					
8	Даниловский	1321		919					
9	Некрасовский	2121							
10	Переславский	8872							
11	Брейтовский	632							
12	Мышкинский	218							
13	Некоузский	206							
14	Первомайский	457							
15	Пошехонский	3011							
16	Ростовский	1363							
17									