

Тема лекции №3

**Статистические
методы анализа
данных параметров
транспортного
процесса.**

Цель лекции – изучить статистические методы анализа данных параметров транспортного процесса.

План лекции.

1. Статистические методы анализа данных.
2. Методы анализа данных в MS Excel.
3. Прикладной пакет Statistica.
4. Решение задач в пакете Statistica.

1. Статистические методы анализа данных.

Статистика изучает большие массивы информации и устанавливает закономерности, которым подчиняются случайные массовые явления.

Под *математической статистикой* понимается раздел математики, посвященный математическим методам сбора, систематизации, обработки и интерпретации статистических данных.

Прикладная статистика – ориентированные на прикладную деятельность статистические методы анализа реальных данных, а также методологии организации статистических исследований и их компьютерной обработки. Теоретическая база – теория вероятностей и математическая статистика.

Анализ данных – позволяет подобрать информацию, которая поможет ответить на все вопросы исследований и проверить гипотезы.



В теории статистику принято условно различать на:

- описательную
- аналитическую.

Описательная статистика связана с планированием исследования, сбором информации и представлением полученных результатов в виде статистических показателей.

Удобная форма представления статистической информации - таблицы, графики.

Задача аналитической статистики - выявить причинные связи, оценить влияние исследуемых факторов и сделать надлежащие выводы, на основании которых могут быть приняты ответственные решения.

Типовые задачи анализа данных.

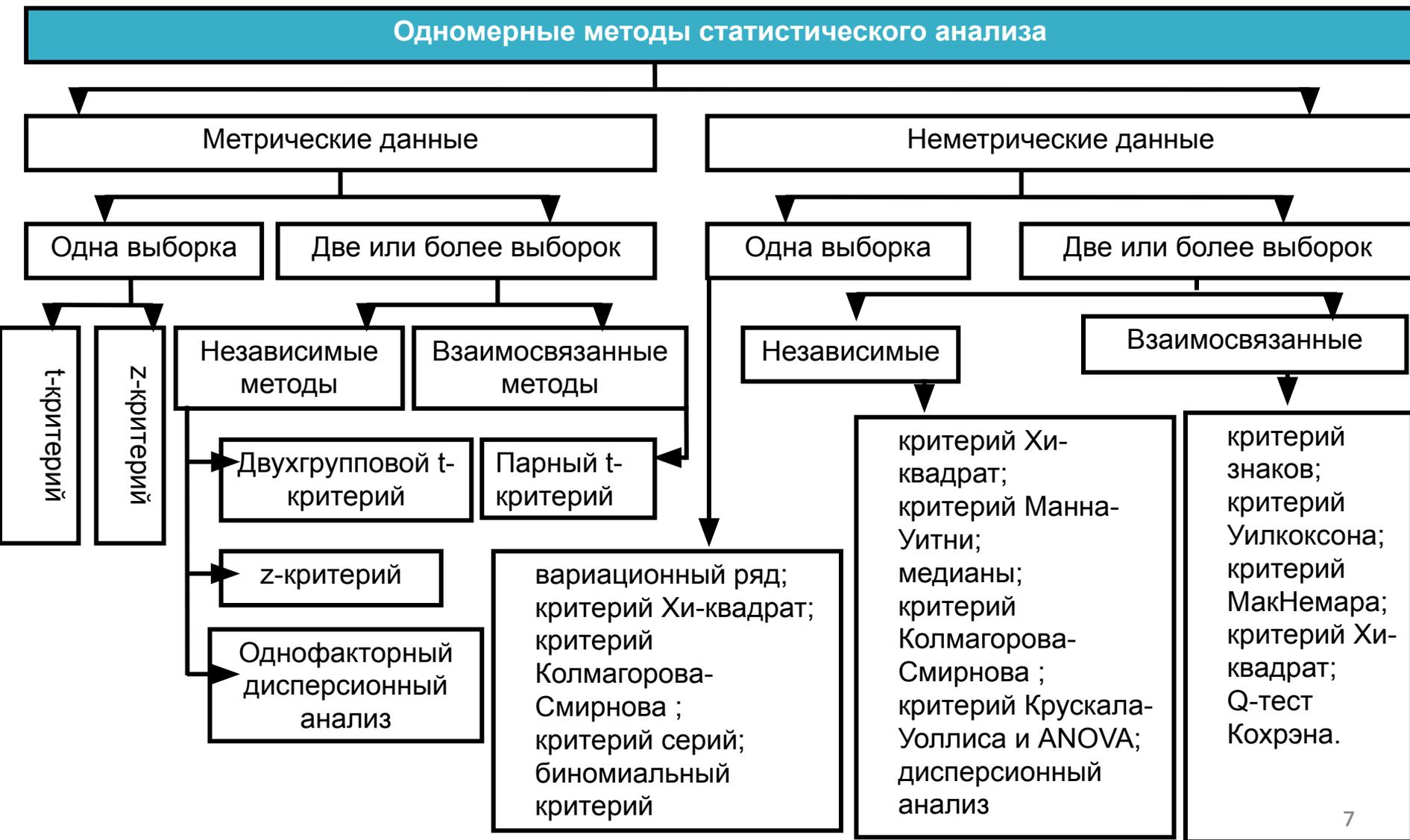
Одномерный анализ:

- Сравнение математических ожиданий;
- Сравнение дисперсий;
- Оценивание параметров распределений;
- Установление закона распределения;
- Отбраковка данных.

Многомерный анализ:

- Исследование зависимостей между признаками;
- Классификация объектов;
- Снижение размерности пространства признаков.

Классификация методов анализа данных



Классификация методов анализа данных



Основные задачи статистического анализа:

- статистическая проверка гипотез;
- определение числа наблюдений и получение выборки;
- определение характеристик генеральной совокупности на основе характеристик выборочной совокупности;
- построение уравнений корреляционной связи (уравнений регрессии);
- создание модели наблюдений (закон распределения);
- оценка параметров модели;
- изучение согласия между моделью и наблюдениями;
- реальное решение задач посредством оценки параметров и критериев значимости.

Способы

Группировка – разбиение совокупности

Табулирование предполагает простой подсчет количества случаев.

Ранжирование позволяет разделить

Математически **распределение частот** является функцией, которая в первую очередь определяет для каждого показателя идеальное значение, так как эта величина обычно уже измерена.

- Группировка
- Табулирование
- Ранжирование
- Распределение частот
- Интервальное распределения частот
- Статистические ряды
- Графическое представление данных

Меры центральной тенденции

- Мода

Мода — это наиболее часто встречающийся вариант ряда.

- Медиана

Медиана — это значение признака, которое лежит в основе ранжированного ряда и делит этот ряд на две равные по численности части.

- Среднее арифметическое значение

- Среднее геометрическое

- Среднее гармоническое

Среднее геометрическое получается от перемножения данных величин и извлечения из

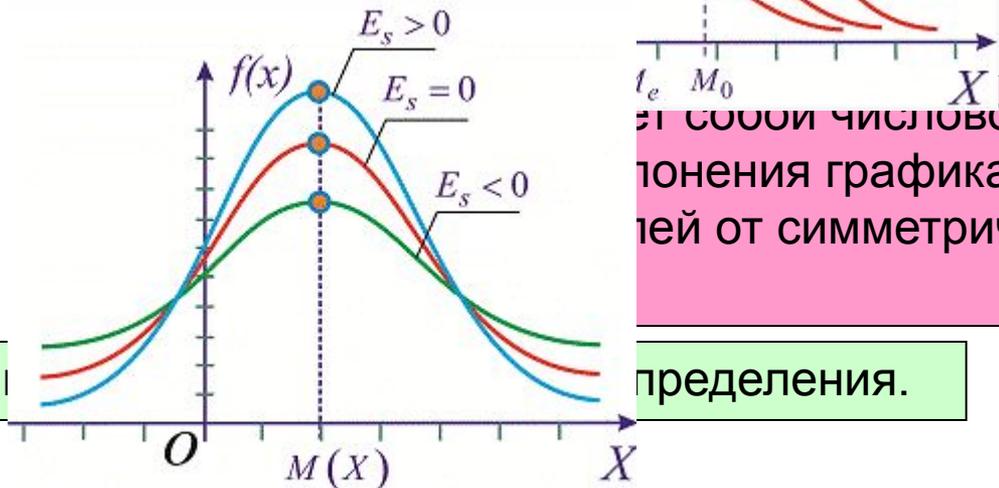
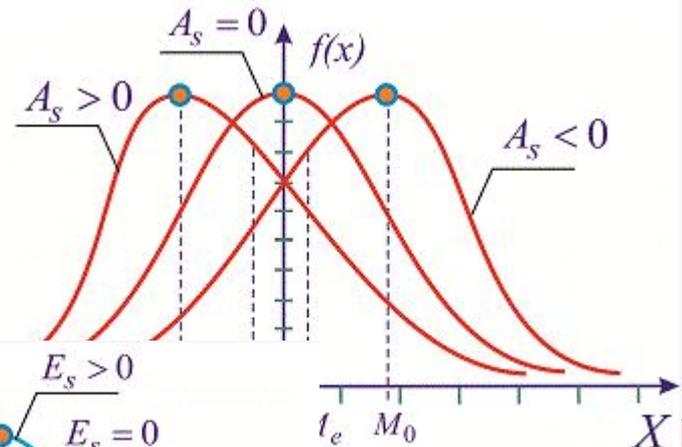
Средним гармоническим нескольких положительных чисел называется число, обратное среднему арифметическому.

Медиана		●	●
Среднее		●	●

Меры изменчивости (вариативности)

- Размах
- Квартильный размах
- Дисперсия
- Стандартное отклонение
- Коэффициент вариации
- Асимметрия
- Эксцесс

Квартильный размах – разница между верхней и нижней квантилями.



...ет собой числовое
тонения графика
тей от симметричного

Эксцесс — показатель

пределения.

Совокупность – группа объектов, предметов или явлений, объединенных каким-либо общим признаком или свойством качественной или количественной характеристики (генеральная или выборочная совокупность).

Выборка или выборочная совокупность — часть генеральной совокупности элементов, которая охватывается экспериментом (наблюдением, опросом).

Характеристики выборки:

- Качественная характеристика выборки — что именно мы выбираем и какие способы построения выборки мы для этого используем.
- Количественная характеристика выборки — сколько случаев выбираем, другими словами объём выборки.

Необходимость выборки:

- Объект исследования очень обширный.
- Существует необходимость в сборе первичной информации.

Заметим, что из генеральной совокупности можно отобрать огромное число выборок. Например, при генеральной совокупности N , равной 100 элементам, можно извлечь выборки объемом $n = 10$ в количестве 10^9 вариантов (1)

Характеристики совокупностей

Наименование характеристики	Для генеральной совокупности	Для выборки
Количество элементов	N	n
Частота	M	m
Частость (доля)	$p = \frac{M}{N}$	$w = \frac{m}{n}$
Среднее	$\mu = \frac{\sum_{i=1}^N x_i}{N}$	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
Дисперсия	$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$	$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$
Стандартное отклонение	$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$	$S_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$

При проведении выборочного наблюдения необходимо соблюдать следующие требования:

- единицы совокупности должны быть: легко различимы; не перекрывать друг друга; образовывать всю совокупность;
- выбор единиц совокупности должен соответствовать целям наблюдения;
- они должны быть удобны для работы;
- должна существовать возможность их перечисления (составление перечня);
- выборочная совокупность должна быть репрезентативной (представительской), т.е. давать представление обо всей совокупности для этого используется метод случайного отбора.

Процесс построения выборки - из большей по размеру генеральной совокупности извлекается выборка для проведения измерений и подробного анализа.

При этом полагается, что выборка является *репрезентативной* (представительной).

Суть репрезентативности выборки – выборка (часть целого) должна достоверно отражать генеральную совокупность (само целое).

Этому соответствует одинаковость частот проявления признака (свойства) как для выборки, так и для всей совокупности, т.е. кривые распределения должны быть идентичными (положение центра, характер формы кривой). Различие только по размаху вариации (дисперсии) – генеральная совокупность должна иметь меньший разброс относительно среднего

Для того, чтобы выборка была **репрезентативной** (хорошо представлять элементы ГС), она должна быть отобрана случайно.

Случайность отбора элементов в выборку достигается соблюдением принципа равной возможности каждого элемента ГС быть отобранным в выборку.

Нарушение принципов случайного выбора приводит к серьезным ошибкам.

Любое число, полученное на основе выборки, носит название «**выборочная статистика**» (или просто «статистика»).

Пусть получена выборка объема n . Над этим массивом исходных данных выполняется операция ранжирования, т.е. экспериментальные данные выстраиваются в порядке возрастания:

$$x_1 < x_2 < x_3 < \dots < x_k; \quad k \leq n;$$

причем значение x_i встречается n_i раз:

$$n_1 + n_2 + \dots + n_k = n;$$

n_i — частота варианта

(количество появлений значений x_i);

$w_i = \frac{n_i}{n}$ — относительная частота варианта или частость ;

обязательно выполняется
$$\sum_{i=1}^k w_i = 1;$$

размах выборки $R = x_{\max} - x_{\min} = x_k - x_1$.

Вариационным рядом называется ранжированный в порядке возрастания ряд значений (вариантов) с соответствующими им

Значения x_i	x_1	x_2	...	x_k
Частота n_i	n_1	n_2	...	n_k
Частоты $w_i = n_i/n$	w_1	w_2	...	w_k

Данный вариационный ряд носит название дискретного вариационного ряда (его члены принимают отдельные изолированные значения).

Построение дискретного вариационного ряда нецелесообразно, когда число значений в выборке велико или признак имеет непрерывную природу, т.е. может принимать любые значения в пределах некоторого интервала. В этом случае строят интервальный вариационный ряд.

Вид интервального ряда:

Интервалы вариантов	$x_1 - x_2$	$x_2 - x_3$...	$x_{k-1} - x_k$
Частота n_i	n_1	n_2	...	n_{k-1}
Частоты $w_i = n_i/n$	w_1	w_2	...	w_{k-1}

Статистический метод определения объема выборки

Для бесповторного отбора

$$n_x = \frac{N \cdot t^2 \cdot \sigma^2}{N \cdot \Delta x^2 + t^2 \cdot \sigma^2};$$

Для повторного отбора

$$n_x = \frac{t^2 \cdot \sigma^2}{\Delta x^2},$$

где σ^2 – дисперсия генеральной совокупности;

N – размер генеральной совокупности;

Δx – доверительный интервал (предельная ошибка);

t – критерий Стьюдента или табулированная константа, табличные значения этой величины следующие: $t=1,96$, при $\alpha=0,05$; $t=2,58$, при $\alpha=0,01$.

Особенность представленных формул :

- в первом случае можно вести расчет, отталкиваясь от известного нам объема самой генеральной совокупности N .
- вторая формула позволяет получить результат, формально игнорируя её количественный размер.

При планировании выборочного исследования предполагается заранее, что известны следующие данные:

- величина допустимой ошибки выборки Δx (доверительного интервала);
- вероятность выводов по результатам наблюдения (величина t -критерия при заданной доверительной вероятности P или уровне значимости α).

Величина σ^2 , характеризующая дисперсию признака в генеральной совокупности, чаще всего бывает неизвестна. Поэтому используют следующие приближенные способы оценки генеральной дисперсии.

1. Можно провести пробное исследование (обычно небольшого объема), на базе которого определяется величина дисперсии этой выборки, используемой в качестве оценки σ^2 :

$$\sigma^2 = \frac{\sum (x_i - \bar{x}_{\text{проб}})^2}{n_{\text{проб}} - 1},$$

где $\bar{x}_{\text{проб}}$ - среднее арифметическое по результатам пробного исследования; $n_{\text{проб}}$ - число единиц, попавших в пробное исследование.

По данным нескольких таких маломасштабных экспериментов выбирается наибольшее значение дисперсии, которое и будет использовано при проведении полного исследования.

2. Можно использовать данные прошлых выборочных наблюдений, проводившихся в аналогичных целях, т.е. дисперсия, полученная по их результатам, применяется в качестве оценки генеральной дисперсии.
3. Если распределение признака в генеральной совокупности может быть отнесена к нормальному закону распределения, то размах вариации примерно равен 6σ (крайние значения отстоят в ту и другую сторону от средней на расстоянии 3σ для $P=99,7\%$), т.е. $R=6\sigma$, откуда $\sigma=1/6R$, где $R=x_{\max} - x_{\min}$.

2. Методы анализа данных в MS Excel.

Программа MS Excel обладает:

- специальным набором функций, которые позволяют вычислять функции распределения случайных величин;
- средствами графического представления данных (постройка диаграмм);
- собственным языком программирования (VBA), с помощью которого можно задавать сложные расчетные алгоритмы;
- набором элементов управления, которые можно внедрять в рабочие листы электронных таблиц;
- удобным способом сохранения данных в виде электронных таблиц;
- использование формул в ячейках для вычисляемых полей.

Файл MS Excel представляет собой **книгу**, которая состоит из набора листов.

Каждый **лист** представляет собой таблицу ячеек.

Каждая **ячейка** может хранить информацию и адресуется именем столбца и номером строки.

Ячейки могут быть вычисляемы, т.е. содержать формулу вычисления по другим ячейкам или их диапазону.

Каждый лист имеет программный модуль, который содержит функции-обработчики событий с данным листом.

Функции расчета

1. Функция МИН.

МИН(число1;числ

Функция МИН на

2. Функция НАИМ.

НАИМЕНЬШИЙ(1

Функция НАИМЕ

наименьшее зн

3. Функция МАКС.

МАКС(число1;чис

Функция МАКС на

4. Функция НАИБО.

НАИБОЛЬШИЙ(M

Функция НАИБОЛ

наибольшее зн

=НАИБОЛЬШИЙ(С3:С32;1)			мые при	
№	Объем заказа, т	Интервал, ч	F	G
1	3,40	8,51		
2	19,10	4,63	Мин.значение	1,111
3	5,23	5,94	Наименьшее	1,106763
4	7,30	0,13	Макс.значение	19,255
5	8,46	9,80	Наибольшее	19,250163
6	2,16	6,05		
7	16,59	4,03		
8	9,14	3,24		
9	10,65	7,87		
10	3,59	9,64		
11	10,34	0,46		
12	12,63	4,03		
13	7,14	8,24		
14	14,78	4,77		
15	15,36	8,22		
16	16,53	7,44		
17	19,25	4,72		
18	5,57	1,23		
19	11,49	6,64		
20	8,55	8,99		
21	14,27	4,47		
22	1,28	3,75		
23	16,32	0,10		
24	6,31	4,72		
25	14,22	9,01		
26	3,92	3,56		
27	13,47	6,46		
28	3,72	0,21		
29	1,11	5,13		
30	2,94	1,23		

го)

го)

5. Функция КВАРТИЛЬ.

КВАРТИЛЬ(массив;k).

Функция КВАРТИЛЬ рассчитывает значение k-го квартиля вариационного ряда.

Функция КВАРТИЛЬ рассчитывает значение k-го квартиля вариационного ряда.

- минимальное значение
- **первую квартиль**, если k=1
- значение медианы, если k=2
- **третью квартиль**, если k=3
- максимальное значение, если k=4

— Функция КВАРТИЛЬ не работает с массивом данных, если он не ранжирован

=КВАРТИЛЬ(С3:С32;0)						
В	С	Д	Е	Ф	Г	
№	Объем заказа, т	Интервал, ч				
1	11,54	3,20				
2	16,35	1,68		Мин.значение	0,20	
3	7,46	1,22		Наименьшее	0,195285	
4	16,40	3,86		Макс.значение	18,02	
5	16,99	8,57		Наибольшее	18,02052	
6	15,03	3,80				
7	15,80	9,94		КВАРТИЛЬ(0)	0,195285	
8	11,56	6,20		КВАРТИЛЬ(4)	18,02052	
9	6,95	8,17				
10	6,69	9,87				
11	3,97	3,04				
12	5,76	9,81				
13	18,02	2,36				
14	16,93	4,37				
15	3,69	1,52				
16	16,56	8,85				
17	8,02	1,31				
18	15,87	3,32				
19	4,67	8,82				
20	12,73	1,51				
21	4,97	8,51				
22	11,80	3,27				
23	2,38	7,10				
24	15,09	8,51				
25	11,15	9,09				
26	10,81	9,38				
27	0,20	2,70				
28	16,29	5,78				
29	13,04	6,67				
30	15,76	4,30				

6. Функция СРЗНАЧ.

СРЗНАЧ(число1;число2;...).

Функция СРЗНАЧ рассчитывает значение арифметической множества данные

7. Функция СРГАРМ.

СРГАРМ(число1;число2;...).

Функция СРГАРМ рассчитывает значение гармонической множества данные

8. Функция СРГЕОМ.

СРГЕОМ(число1;число2;...).

Функция СРГЕОМ рассчитывает среднее положительных чисел.

9. Функция МОДА.

МОДА(число1;число2;...).

Функция МОДА отображает наиболее частотный интервал данных.

10. Функция МЕДИАНА.

МЕДИАНА(число1;число2;...).

Функция МЕДИАНА рассчитывает значение медианы вариационного ряда.

№	Объем заказа, т	Интервал, ч		
1	17,65	5,45		
2	15,53	0,50		
3	12,27	9,85	Среднее значение	10,03
4	10,54	5,16	Среднее гармонич.	4,360098
5	12,76	4,19	Среднее геометрич.	7,466286
6	12,18	2,61	Мода	2,7
7	6,84	5,81	Медиана	11,42
8	11,22	2,71		
9	14,18	2,68		
10	6,63	2,46		
11	11,63	4,51		
12	1,62	0,27		
13	0,88	6,64		
14	4,82	2,41		
15	12,88	4,25		
16	12,96	5,09		
17	2,25	1,17		
18	19,76	2,22		
19	13,83	5,22		
20	15,95	1,25		
21	2,17	6,07		
22	14,99	1,59		
23	7,79	1,45		
24	9,38	0,96		
25	19,80	1,66		
26	0,72	6,89		
27	7,01	8,28		
28	17,26	3,78		
29	2,70	4,01		
30	2,70	3,69		

Функции MS Excel, используемые при расчете показателей разброса

1. Функция ДИСП.

ДИСП(число1;число2;...).

Функция ДИСП оценивает генеральную выборку.

Функция ДИСП рассчитывает дисперсию при условии, что данные образуют **выборочную** совокупность.

В случае, если совокупность является **генеральной**, воспользоваться функцией ДИСПР.

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

Часто генеральную дисперсию обозначают D_x .

fx =ДИСП(С3:С17)

В	С	Д	Е
№	Объем, т		
1	5,6		
2	2,2	Дисперсия	23,0841
3	1,5		
4	5		
5	12		
6	11		

fx =ДИСПР(С3:С17)

В	С	Д	Е
№	Объем, т		
1	5,6		
2	2,2		
3	1,5	Генеральная дисп.	21,54516
4	5		
5	12		
6	11		
7	18		
8	2		
9	1		
10	3,4		
11	2,2		
12	5,6		
13	10		
14	8		
15	7,2		

3. Функция СТАНДОТКЛОН.

СТАНДОТКЛОН(число1;число2;...).

Функция СТАНДОТКЛОН оценивает генеральное отклонение (стандарт) по выборке.

Функция СТАНДОТКЛОН рассчитывает исходные данные образуют выборочную совокупность является генеральной, то функцией СТАНДОТКЛОНП.

Формула: =СТАНДОТКЛОН(С3:С17)

№	Объем, т		
1	5,6		
2	2,2	Дисперсия	23,0841
3	1,5	Генеральная дисп.	21,54516
4	5	Стандартн. Отклонение	4,804591
5	12		
6	11		
7	18		
8	2		
9	1		
10	3,4		

4. Функция СТАНДОТКЛОНП.

СТАНДОТКЛОНП(число1;число2;...).

Функция СТАНДОТКЛОНП вычисляет стандартное отклонение генеральной совокупности.

Формула: =СРОТКЛ(С3:С17)

№	Объем, т		
1	5,6		
2	2,2		
3	1,5		
4	5		
5	12		
6	11	Среднее линейное отклонение	3,776
7	18		
8	2		
9	1		
10	3,4		
11	2,2		
12	5,6		
13	10		
14	8		
15	7,2		

5. Функция СРОТКЛ.

СРОТКЛ(число1;число2;...).

Функция СРОТКЛ вычисляет среднее невзвешенное отклонение множества данных.

Функция Excel, используемая при расчете показателя асимметрии

Функция СКОС.

СКОС(число1;число2;...).

Функция СКОС оценивает коэффициент асимметрии в выборке.

$$A_x = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s} \right)^3$$

Если данные образуют не выборочную совокупность, то асимметрию необходимо рассчитывать по стандартной формуле:

$$A_x = \frac{\mu_3}{\sigma^3}$$

fx = СКОС(С3:С17)			
В	С	Д	Е
№	Объем, т		
1	5,6		
2	2,2		
3	1,5		
4	5		
5	12		
6	11		
7	18	Асимметричность	1,062572
8	2		
9	1		
10	3,4		
11	2,2		
12	5,6		
13	10		
14	8		
15	7,2		

Функция Excel, используемая при расчете показателя распределения

Функция ЭКСЦЕСС.

ЭКСЦЕСС(число1;число2;...).

Функция ЭКСЦЕСС оценивает эксцесс по

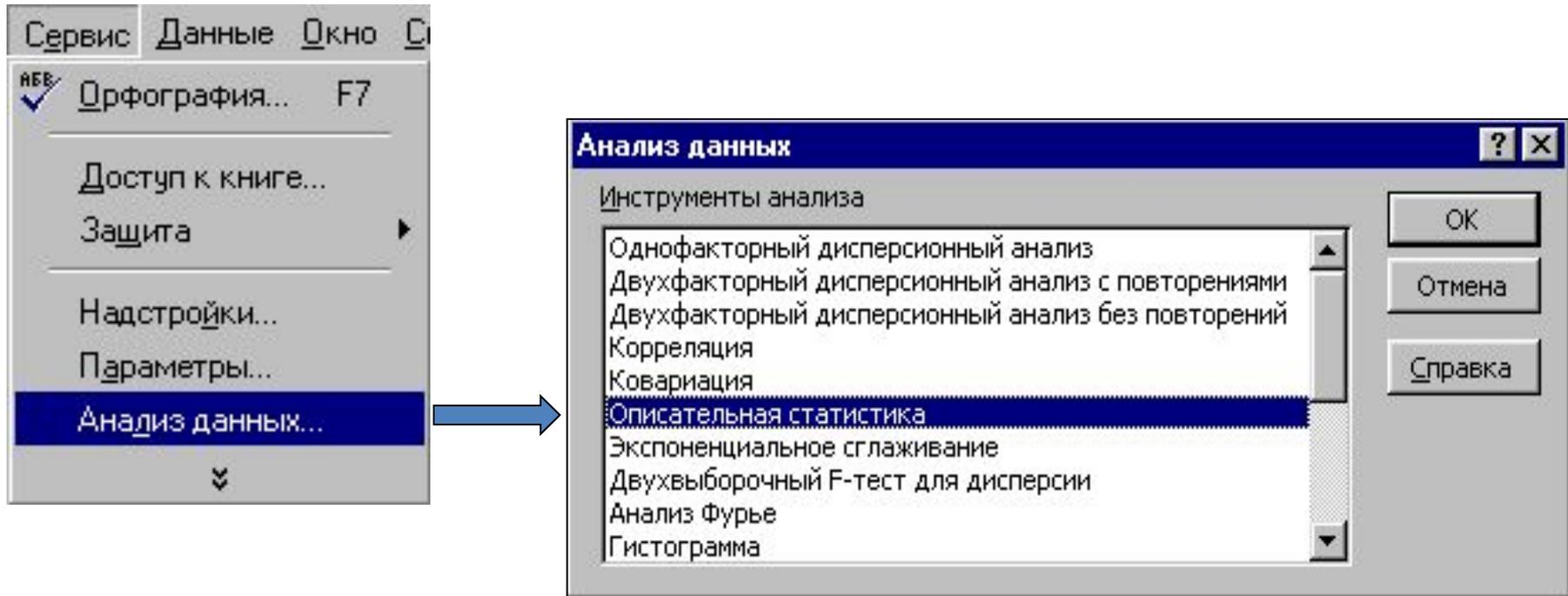
$$E_x = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \right] \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^3$$

Если данные образуют не выборочную, совокупность, то эксцесс необходимо рассчитать по стандартной формуле:

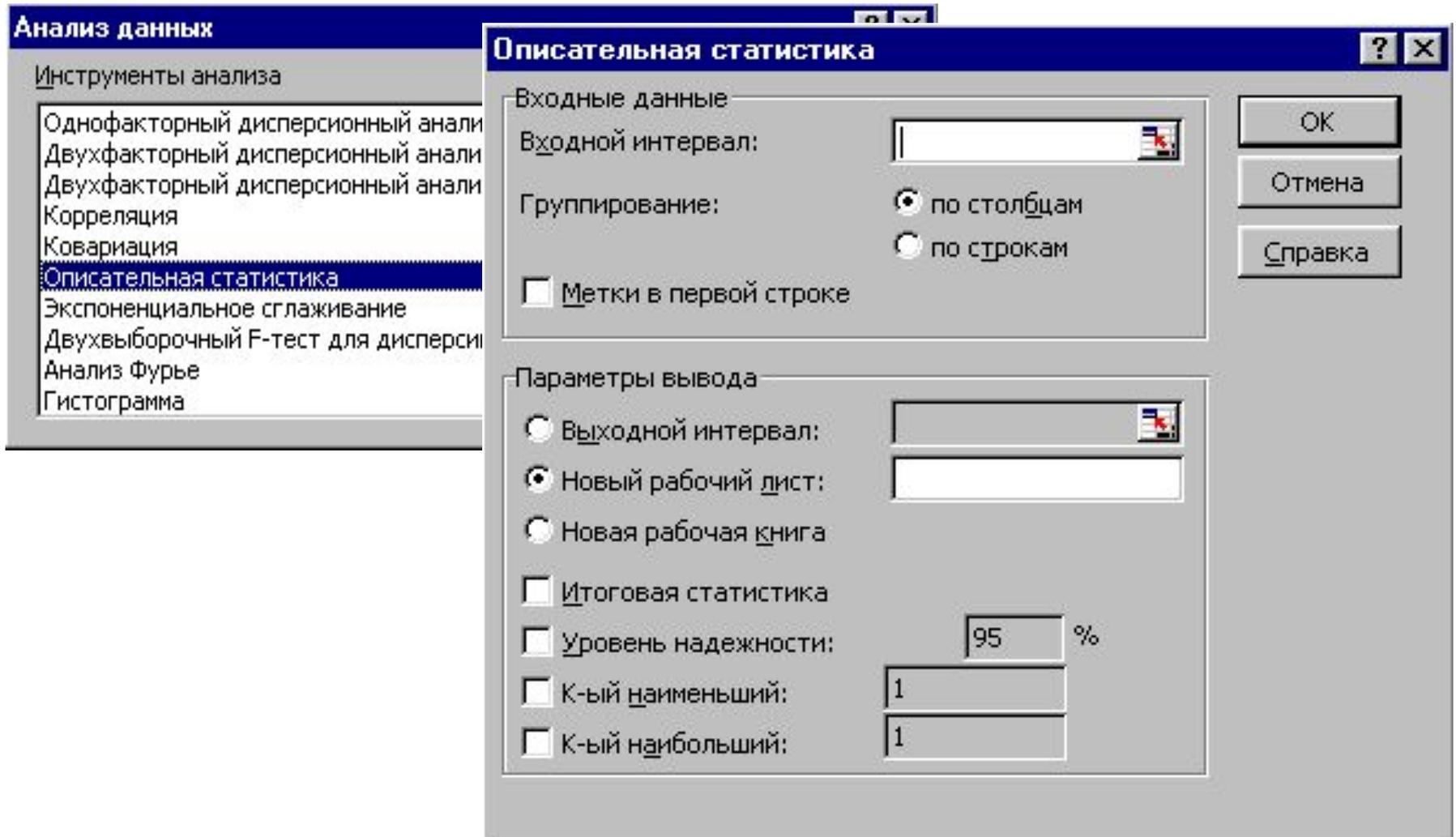
$$E_x = \frac{\mu_4}{\sigma^4} - 3$$

fx = ЭКСЦЕСС(С3:С17)			
В	С	Д	Е
№	Объем, т		
1	5,6		
2	2,2		
3	1,5		
4	5		
5	12		
6	11		
7	18		
8	2	ЭКСЦЕСС	0,912655
9	1		
10	3,4		
11	2,2		
12	5,6		
13	10		
14	8		
15	7,2		

Выход в режим «Описательная статистика»



Справочная информация по технологии работы в режиме «Описательная статистика»



Ввод данных

№	Объем, т
1	5,6
2	2,2
3	1,5
4	5
5	12
6	11
7	18
8	2
9	1
10	3,4
11	2,2
12	5,6
13	10
14	8
15	7,2

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

Результаты

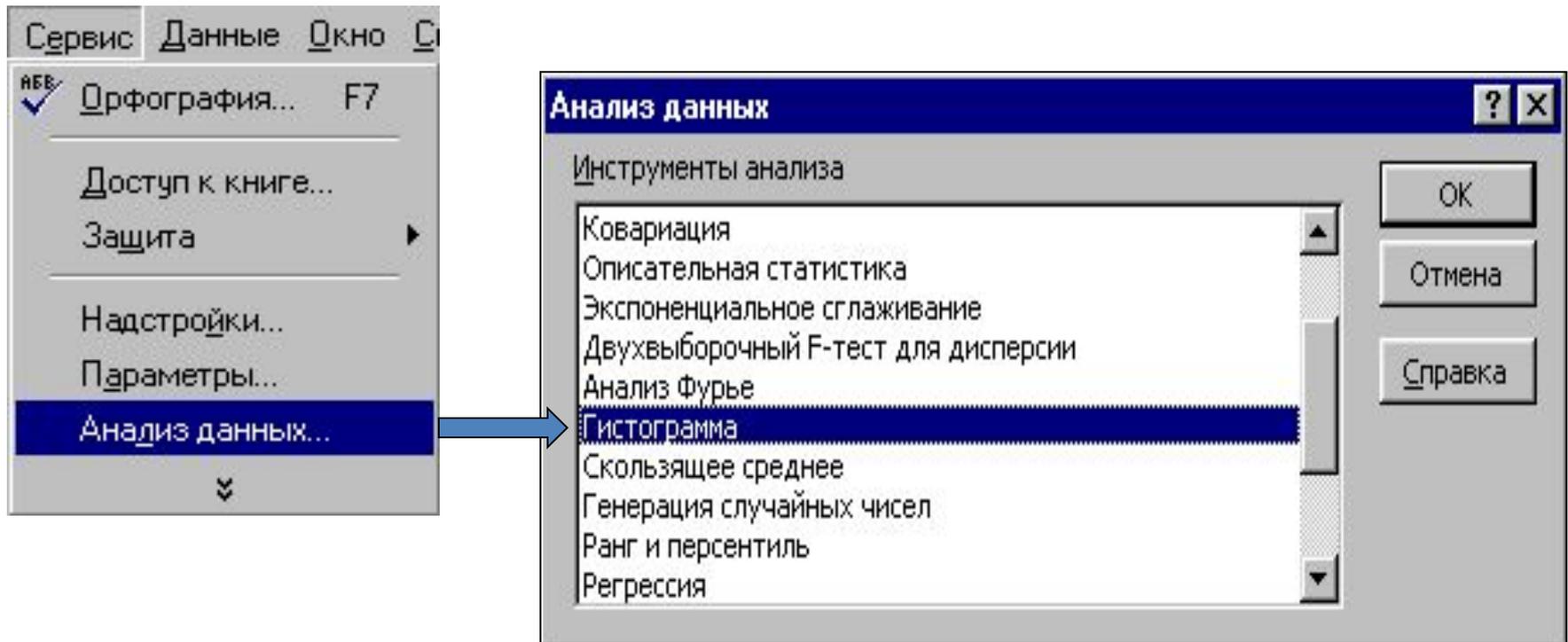
Средняя ошибка выборки
(показатель **Стандартная ошибка**)

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

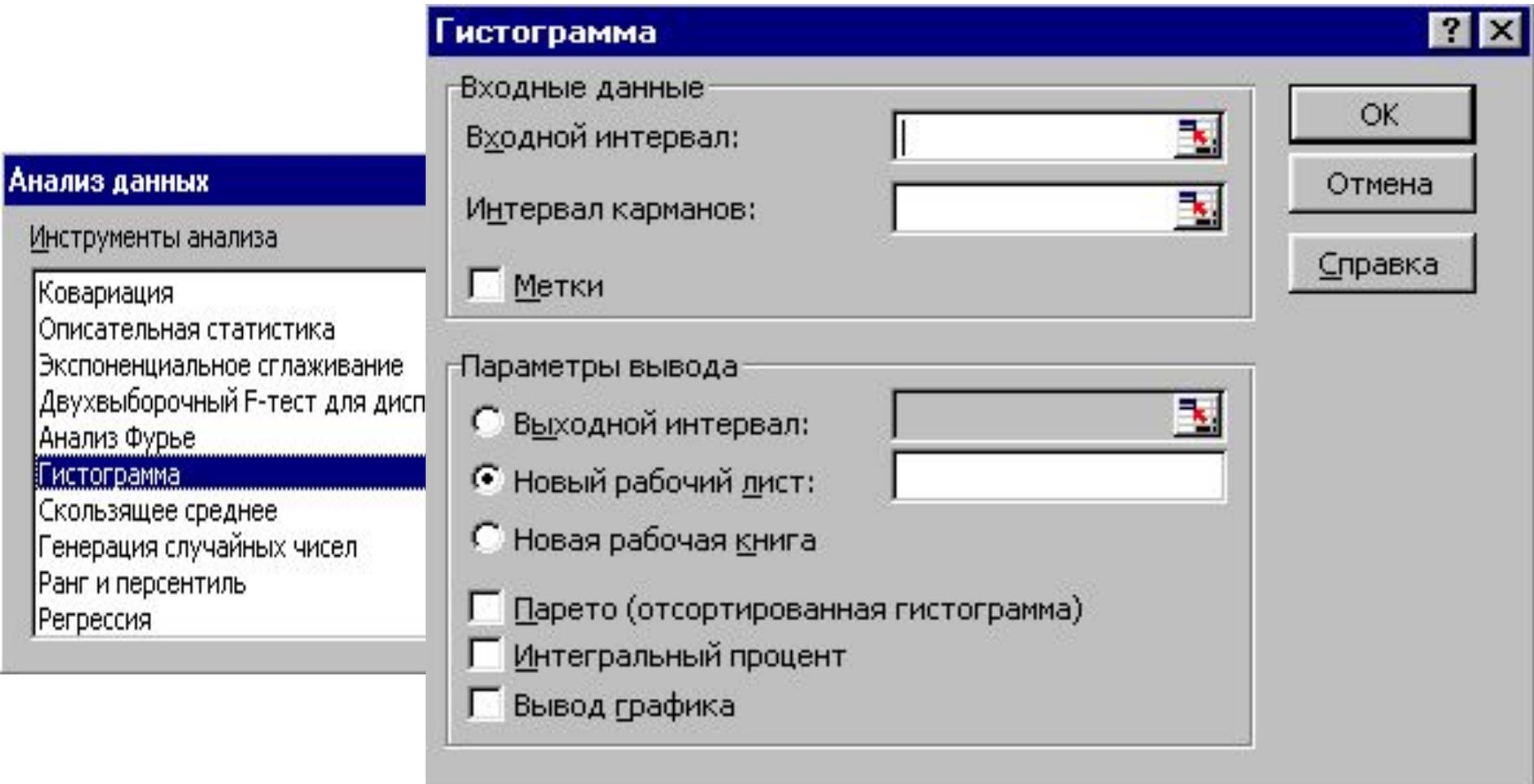
=E8/КОРЕНЬ(E16)

	C	D	E	
	Объем, т	Объем, т		
	5,6			
	2,2	Среднее	6,313333	
	4,5	Стандартная ошибка	1,24054	
	5	Медиана	5,6	
	12	Мода	5,6	
8	6	11	Стандартное отклонение	4,804591
9	7	18	Дисперсия выборки	23,0841
10	8	2	Эксцесс	0,912655
11	9	1	Асимметричность	1,062572
12	10	3,4	Интервал	17
13	11	2,2	Минимум	1
14	12	5,6	Максимум	18
15	13	10	Сумма	94,7
16	14	8	Счет	15
17	15	7,2		

Справочная информация по технологии работы в режиме «Гистограмма»



Режим *Гистограмма* служит для вычисления частот попадания данных в указанные границы интервалов, а также для построения гистограммы **интервального** вариационного ряда распределения.



Ввод данных

№	Объем, т
1	5,6
2	2,2
3	1,5
4	5
5	12
6	11
7	18
8	2
9	1
10	3,4
11	2,2
12	5,6
13	10
14	8
15	7,2

Гистограмма

Входные данные

Входной интервал:

Интервал карманов:

Метки

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

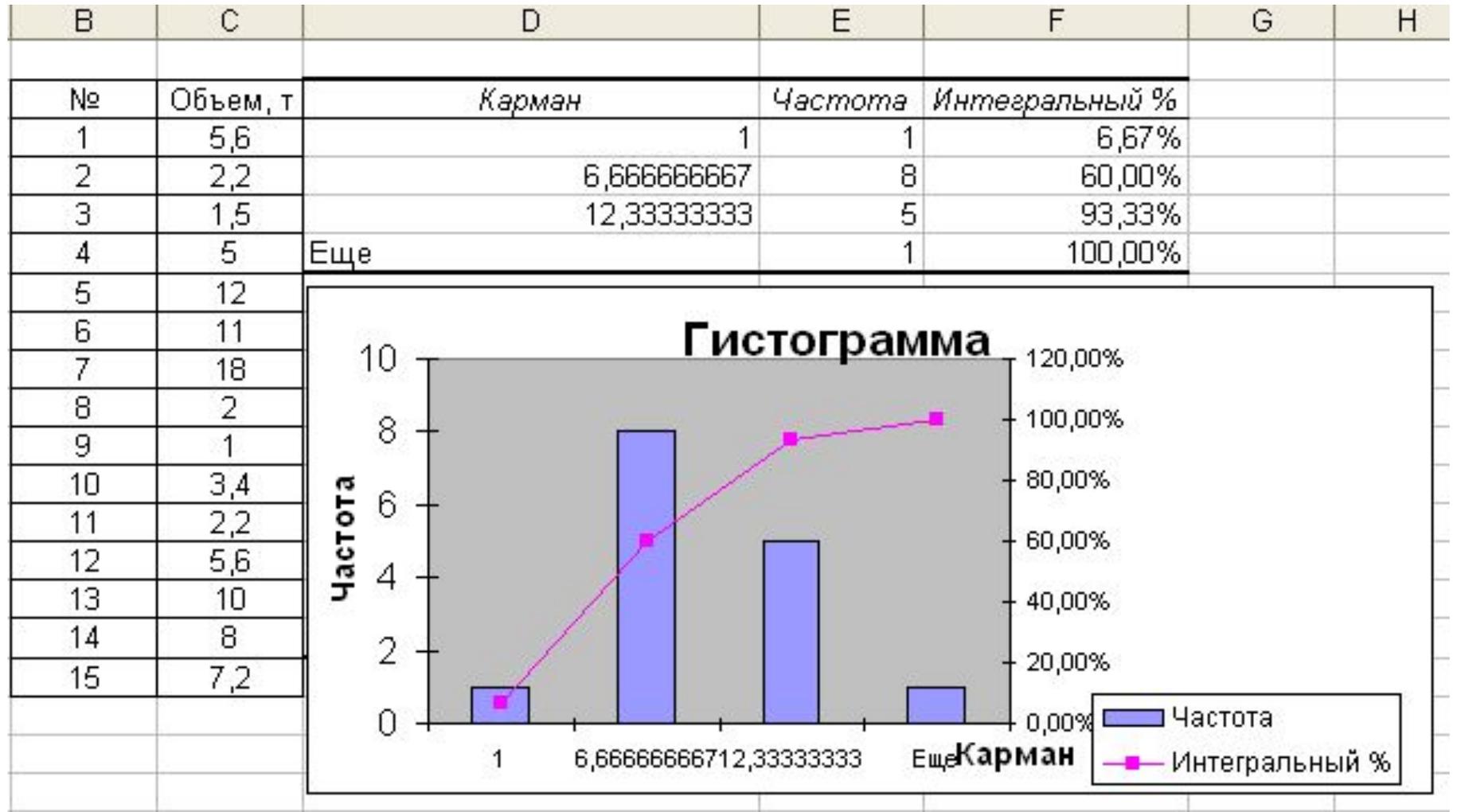
Парето (отсортированная гистограмма)

Интегральный процент

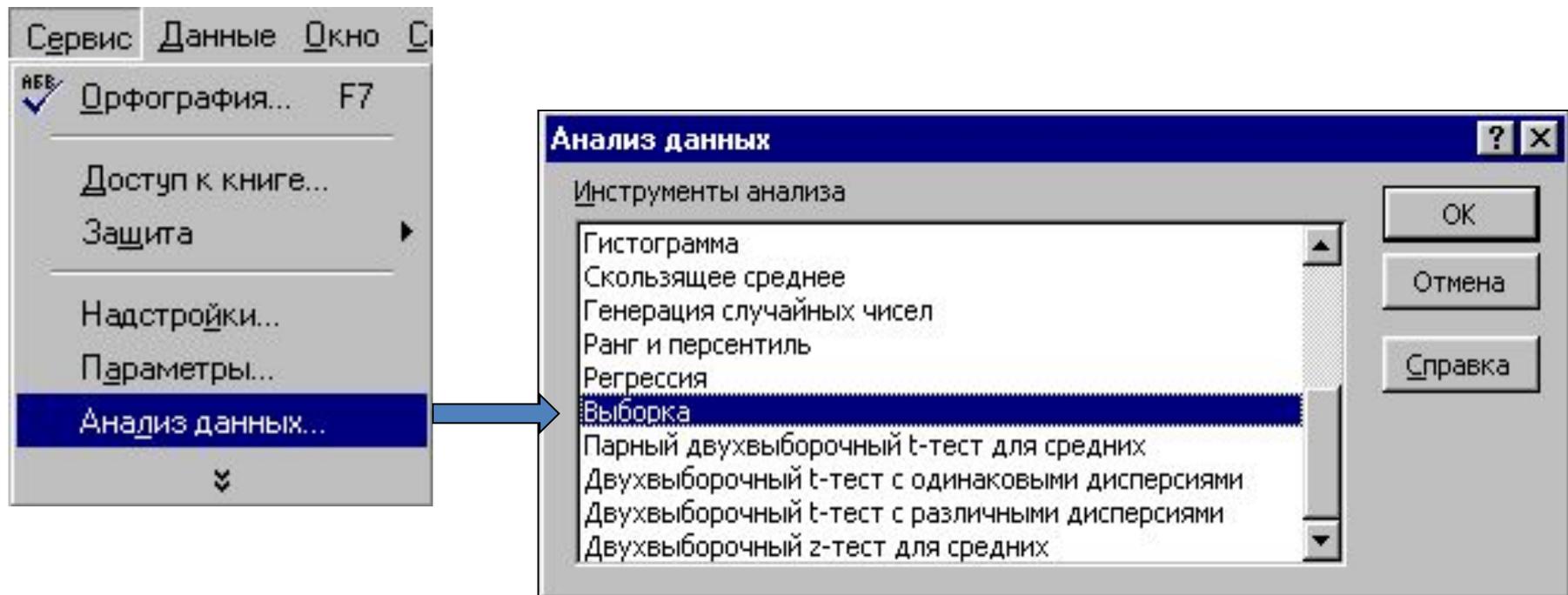
Вывод графика

OK
Отмена
Справка

Результат



Справочная информация по технологии работы в режиме «Выборка»



Режим **Выборка** служит для формирования выборки из генеральной совокупности на основе схемы случайного отбора, а также из периодических данных.

Анализ данных

Инструменты анализа

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

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

№	Объем заказа, т	
1	13,32	
2	8,94	
3	19,48	
4	12,58	
5	8,65	
6	14,27	
7	10,21	
8	13,46	
9	18,72	
10	18,36	
11	19,98	
12	19,54	
13	0,40	
14	19,11	
15	16,86	
16	0,14	
17	11,85	0,11
18	1,25	1,71
19	11,19	0,85
20	1,29	1,12
21	9,80	9,11
22	3,35	4,81
23	4,67	0,39
24	10,05	8,16
25	10,38	2,01
26	1,42	7,38
27	6,23	9,99
28	9,15	0,16
29	2,70	8,21

Выборка

Входные данные

Входной интервал: 

Метки

Метод выборки

Периодический

Период:

Случайный

Число выборок:

Параметры вывода

Выходной интервал: 

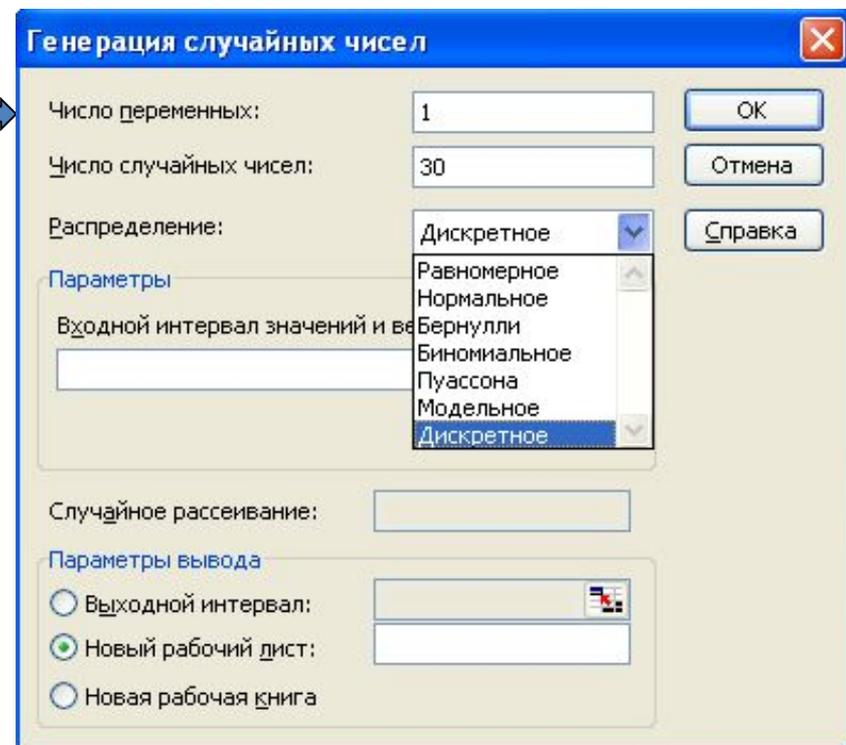
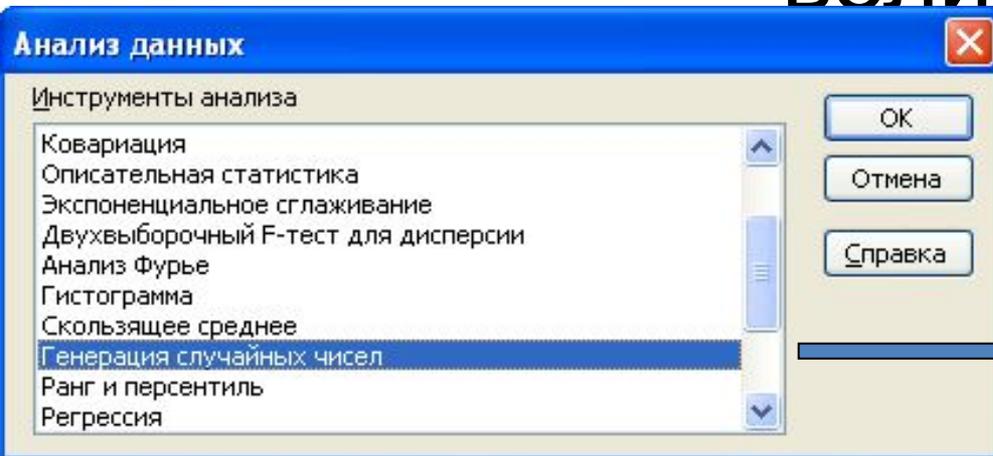
Новый рабочий лист:

Новая рабочая книга

Результаты «Выборки»

№	Объем	Интервал,	
	заказа, т	ч	
1	3,41	5,51	
2	1,61	3,53	
3	1,94	3,13	
4	1,02	7,93	
5	8,89	4,58	
6	0,54	2,91	
7	12,06	4,17	
8	11,53	6,24	
9	19,36	5,48	2,174184883
10	2,17	1,48	13,44647655
11	12,18	5,60	1,173000408
12	18,61	10,00	14,91323109
13	1,60	1,10	16,81658056
14	6,34	3,98	12,17958133
15	7,66	0,03	12,06488985
16	1,35	1,96	7,655889812
17	1,17	1,61	3,41437238
18	13,45	1,84	6,794989303
19	4,06	7,84	
20	3,87	8,57	
21	14,88	5,22	
22	14,68	1,85	
23	14,91	2,49	
24	8,51	9,54	
25	16,82	7,59	
26	5,93	7,75	
27	6,79	9,05	
28	0,75	5,39	
29	2,70	2,49	

Функции генерации случайных величин



Функция генерации равномерного распределения на отрезке

Возвращает равномерно распределенное случайное число, большее либо равное 0 и меньше 1

Синтаксис

- Чтобы получить случайное число a и b , можно использовать формулу $\text{СЛЧИС}()*(b-a)+a$
- Если требуется использовать функцию генерации случайных чисел при каждом вычислении, можно ввести в строку формулы и нажать клавишу F9, чтобы получить случайное число.

=СЛЧИС()*10+1			СЛЧИС()	
B	C	D	C	D
	Случайные числа		Случайные числа	
	4,617672		0,0919	
	8,185748		51487	
	4,023923		74374	
	1,407393		32682	
	5,292164		03134	
	1,520684		65739	
	8,087347		38771	
	9,825033		47437	
	3,845716		33269	
	6,461479		36456	
	8,508965		73114	

Генерация случайных чисел по равномерному закону распределения

Приведенная реализация случайной величины с интервалом $[0, 1]$ к реализации величины с параметром расположения a и формы b осуществляется на основании соотношения:

$$R(a, b) = a + b \cdot R_{01}$$

$$a = \bar{x} - \sqrt{3 \cdot s^2},$$

$$b = \sqrt{12 \cdot s^2},$$

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2,$$

где $R(a, b)$ – равномерно
величина с параметр
параметром формы b
 R_{01} – случайная велич
в интервале от 0 до 1.

где \bar{x} – среднее арифметическое значение;
 s^2 – выборочная дисперсия без поправки;
 n – объём выборки.

Генерация случайных чисел по нормальному закону распределения

Нормально распределенная случайная величина N_{01} с нулевым математическим ожиданием и средним квадратическим отклонением 1 генерируется на основании связи с равномерным распределением

R_{01} :

$$N_{01} = -6 + \sum_{i=1}^{12} R_{01i}.$$

Случайная величина $N(\mu, \sigma)$, распределена по нормальному закону с параметром расположения μ и параметром масштаба σ , приводится с N_{01} на основании соотношения:

$$N(\mu, \sigma) = \mu + \sigma \cdot N_{01_{49}}$$

Генерация случайных чисел по экспоненциальному закону распределения

Значения экспонентно распределенной случайной

вел D2 $\text{fx} = -30,25 * \text{LN}(C2)$

	A	B	C	D	E	F	G	H	I
осл	q	L	I						
ра			0,485008	21,88858					
со			0,275711	38,97418					
			0,019037	119,8315			0,45	297,8	0,5
			0,793963	6,979244			20	649,9	60
			0,936177	1,995003		Средние	q	L	I
			0,306189	35,80242			10,225	473,85	30,25
			0,699669	10,80372	станд_откл		3,26	58,68	9,92
где E			0,079071	76,7565					
ра			0,93797	1,937113					
ма			0,519115	19,83279					
			0,531461	19,12179					
			0,499172	21,01783					

$$b = \bar{x}.$$

3. Прикладной пакет Statistica.

ПП STATISTICA – это универсальная интегрированная система, предназначенная для статистического анализа и обработки данных.

Содержит многофункциональную систему для работы с данными, широкий набор статистических модулей, в которых собраны группы логически связанных между собой статистических процедур, специальный инструментарий для подготовки отчетов, мощную графическую систему для визуализации данных, систему обмена данными с другими Windows-приложениями.

С помощью реализованных в системе STATISTICA языков программирования (SQL, STATISTICA BASIC), снабженных специальными средствами поддержки, легко создаются законченные пользовательские решения и встраиваются в различные другие приложения или вычислительные среды.

История создания пакета Statistica

- Система STATISTICA производится фирмой StatSoft Inc. (США), основанной в 1984 г. в городе Тулса (США). Первые программные продукты фирмы (PsyhoStat-2,3) были предназначены для обработки социологических данных.
- В 1985 г. StatSoft выпускает первую систему статистического анализа для компьютеров Apple Macintosh (StatFast) и статистический пакет для IBM PC (STATS+).
- В 1986 г. начинается работа по созданию интегрированных статистических пакетов комплексной обработки данных.
- В 1991 г. выходит первая версия системы STATISTICA/DOS. Эта программа представляла собой новое направление развития статистического программного обеспечения, так как в ней реализован графически ориентированный подход к анализу данных, могла анализировать фактически неограниченный объем данных.
- В 1992 г. вышла версия STATISTICA для Macintosh.
- В 1994 г. выходит версия STATISTICA 4.5 для Windows, которая сразу же занимает лидирующее положение среди статистических пакетов.

STATISTICA



Решение задач с помощью ПП Statistica (Base)

- **Описательные и внутригрупповые статистики, разведочный анализ данных**
- **Корреляции**
- **Быстрые основные статистики и блоковые статистики**
- **Интерактивный вероятностный калькулятор**
- **T-критерии (и другие критерии групповых различий)**
- **Таблицы частот, сопряженности, флагов и заголовков, анализ многомерных откликов**
- **Множественная регрессия**
- **Непараметрические статистики**
- **Дисперсионный анализ (ANOVA/MANOVA)**
- **Подгонка распределений**

Описательные статистики и графики

Программа вычисляет практически все используемые описательные статистики общего характера: медиану, моду, квартили, заданные пользователем процентиля, среднее значение и стандартное отклонение, квартильный размах, доверительные интервалы для среднего, асимметрию и эксцесс (и их стандартные ошибки), гармоническое и геометрическое среднее.

Доступны разнообразные графики и диаграммы, в т.ч. различные виды диаграмм размаха и гистограмм, гистограммы двумерных распределений (трехмерные и категоризованные), двух- и трехмерные диаграммы рассеяния с помеченными подмножествами данных, нормальные и полунормальные вероятностные графики и графики с исключенным трендом, графики квантиль-квантиль, вероятность-вероятность и т.д.

Имеется набор критериев для подгонки нормального распределения к данным (критерии Колмогорова-Смирнова, Лилиефорса и Шапиро-Уилкса).

4. Решение задач в пакете Statistica.