Владимирский государственный университет
Кафедра Информационных систем и информационного менеджмента

МЕТОДИЧЕСКИЕ УКАЗАНИЯ К ЛАБОРАТОРНЫМ РАБОТАМ ПО ДИСЦИПЛИНЕ «Основы теории стохастических систем»

Владимир 2012

Оглавление

Часть 1. Настройка Пакета Анализа Microsoft Excel 3

1. Описательная статистика 7

2. Гистограмма 10

3. Генерация случайных чисел 14

4. Выборка 21

5. Ранг и персентиль 22

Задания к первой части лабораторных работ: 24

Часть 2. Статистическая проверка гипотез 25

6. Двухвыборочный z-тест для средних 26

7. Двухвыборочныи t-тест с одинаковыми дисперсиями 29

8. Двухвыборочный t-тест с различными дисперсиями 32

9. Парный двухвыборочный t-тест для средних 35

10. Двухвыборочный F-тест для дисперсий 37

Задания ко второй части лабораторных работ: 39

Часть 3. Определение статистических оценок многомерных случайных процессов 40

11. Однофакторный дисперсионный анализ 40

12. Двухфакторный дисперсионный анализ с повторениями 41

13. Двухфакторный дисперсионный анализ без повторений 44

14. Корреляция 46

15. Ковариация 47

16. Регрессия 49

17. Скользящее среднее 54

18. Экспоненциальное сглаживание 56

19. Анализ Фурье 57

Задания к третьей части лабораторных работ: 60

Часть 1. Настройка Пакета Анализа Microsoft Excel

В состав Microsoft Excel входит надстройка ***Пакет анализа***, которая содержит 19 статистических процедур и около 50 функций. Функции в основном относятся к категориям инженерных и финансовых и поэтому здесь не рассматриваются. Статистические процедуры, содержащиеся в надстройке ***Пакет анализа***, предоставляют широкий спектр средств для статистического анализа начиная от простой описательной статистики или сглаживания данных и заканчивая анализом Фурье и проведением различных тестов. Полный список этих средств и их краткое описание представлены в табл.1 (названия средств приводятся в соответствии со списком из диалогового окна ***Анализ данных)***.

Таблица 1. Статистические средства надстройки Пакет анализа

Средство	***Описание***
Однофакторный дисперсионный анализ	Используется для проверки гипотезы о равенстве математических ожиданий двух или более выборок
Двухфакторный дисперсионный анализ без повторений	Двухфакторный дисперсионный анализ на основе одной выборки
Двухфакторный дисперсионный анализ с повторениями	Двухфакторный дисперсионный анализ на основе нескольких выборок
Корреляция	Вычисляет корреляционную матрицу
Ковариация	Вычисляет матрицу ковариаций
Описательная статистика	Создает отчет, содержащий статистические характеристики представленной выборки
Экспоненциальное сглаживание	Реализует метод экспоненциального сглаживания данных
Двухвыборочный F-тест для дисперсий	Применяется для сравнения дисперсий двух генеральных совокупностей
Анализ Фурье	Реализует метод быстрого преобразования Фурье (БПФ) для анализа данных
Гистограмма	Используется для анализа распределения выборочных данных и построения гистограмм
Скользящее среднее	Используется для сглаживания данных
Генерация случайных чисел	Генерирует случайные числа, имеющие заданное распределение
Ранг и персентиль	Используется для вычисления рангов и квантилей

Регрессия	Используется для построения линейной регрессии
Выборка	Создает случайную выборку, рассматривая входной диапазон значений как генеральную совокупность
Парный двухвыборочный t-тест для средних	Используется для проверки гипотезы о равенстве математических ожиданий для двумерной выборки данных
Двухвыборочный t-тест с одинаковыми дисперсиями	Служит для проверки гипотезы о равенстве математических ожиданий для двух выборок. Предполагается равенство дисперсий генеральных совокупностей
Двухвыборочный t-тест с разными дисперсиями	Используется для проверки гипотезы о равенстве математических ожиданий для двух выборок. Не требует предположения о равенстве дисперсий генеральных совокупностей
Двухвыборочный z-тест для средних	Используется для проверки гипотезы о различии между математическими ожиданиям двух генеральных совокупностей

Отметим, что эти средства имеют определенные ограничения и иногда удобнее воспользоваться статистическими функциями или другими средствами Excel. Преимуществом функций перед данными средствами является то, что функции автоматически пересчитываются при любых изменениях, сделанных в выборке, тогда как эти средства необходимо выполнять заново, если выборка изменилась. В "оправдание" этих средств скажем, что они сохраняют установки, сделанные пользователем при последнем применении средства, но только в течение одного сеанса работы с Excel.

Средства, которые включены в надстройку ***Пакет анализа***, доступны через команду ***Сервис=>Анапиз данных***. (Если команды ***Анализ данных*** нет в меню ***Сервис,*** подключите эту надстройку. Для этого выполните команду ***Сервис=> Надстройки*** и в открывшемся диалоговом окне ***Надстройки*** в списке ***Доступные надстройки*** установите флажок напротив опции ***Пакет анализа***.) Команда ***Сервис=>Анапиз*** данных открывает одноименное диалоговое окно, в списке ***Инструменты анализа*** которого следует выбрать необходимое средство (рис. 1). После выбора какого-либо средства (и последующего щелчка на кнопке ОК) открывается диалоговое окно этого средства.

В большинстве таких диалоговых окон (на рис. 2 для примера показано диалоговое окно средства ***Описательная статистика***) выделены области ***Входные данные*** и ***Параметры вывода***. В области ***Входные данные*** указывается диапазон ячеек, в котором содержатся данные

(поле ***Входной интервал***), указывается, сгруппированы ли данные, и если сгруппированы, то по столбцам или по строкам (переключатели по столбцам и по строкам). Если задается входной диапазон данных вместе с заголовками, то устанавливается флажок опции Метки в первой строке (столбце). (Если заголовки не задаются, то данным автоматически присваиваются заголовки Столбец1, Столбец2 и т.д. или Строка1, Строка2 и т.д. в зависимости от того, расположены данные в столбцах или в строках.) В некоторых диалоговых окнах в области ***Входные данные*** необходимо указать несколько входных диапазонов (например, в окне Регрессия) либо дополнительные параметры для проведения выбранной статистической процедуры, например доверительный уровень для проведения тестов.

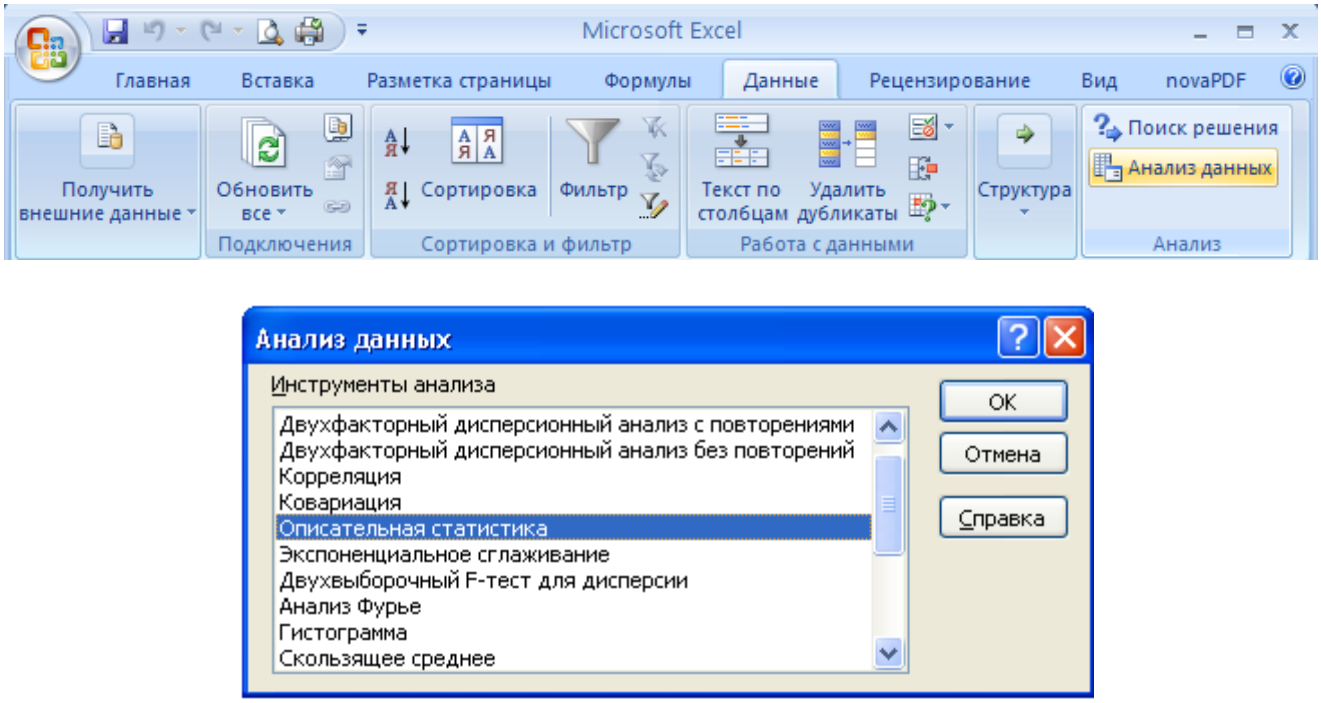


Рис.*1. Диалоговое окно* ***Анализ данных*** *со списком инструментов статистического анализа*

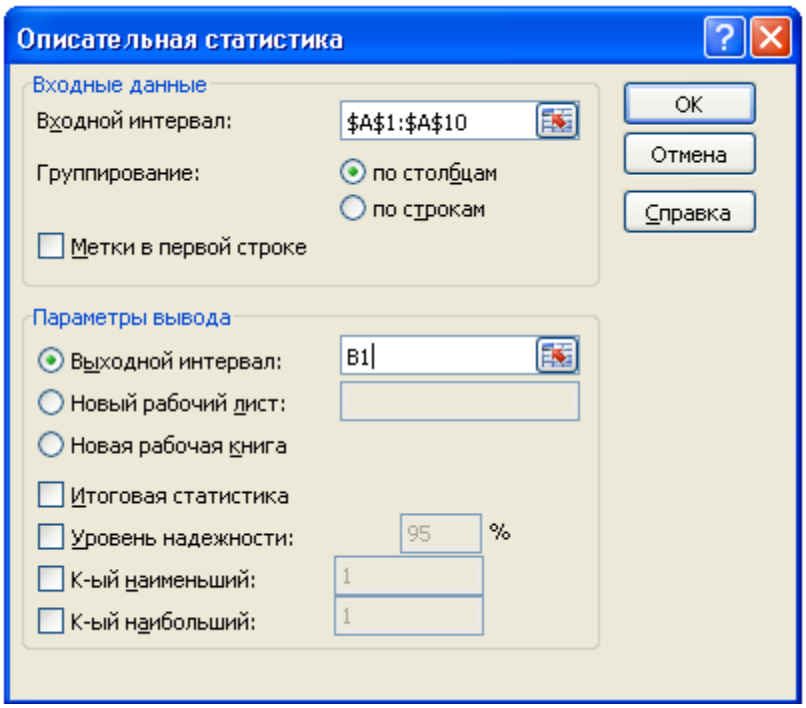


Рис.2. Диалоговое окно средства ***Описательная статистика***

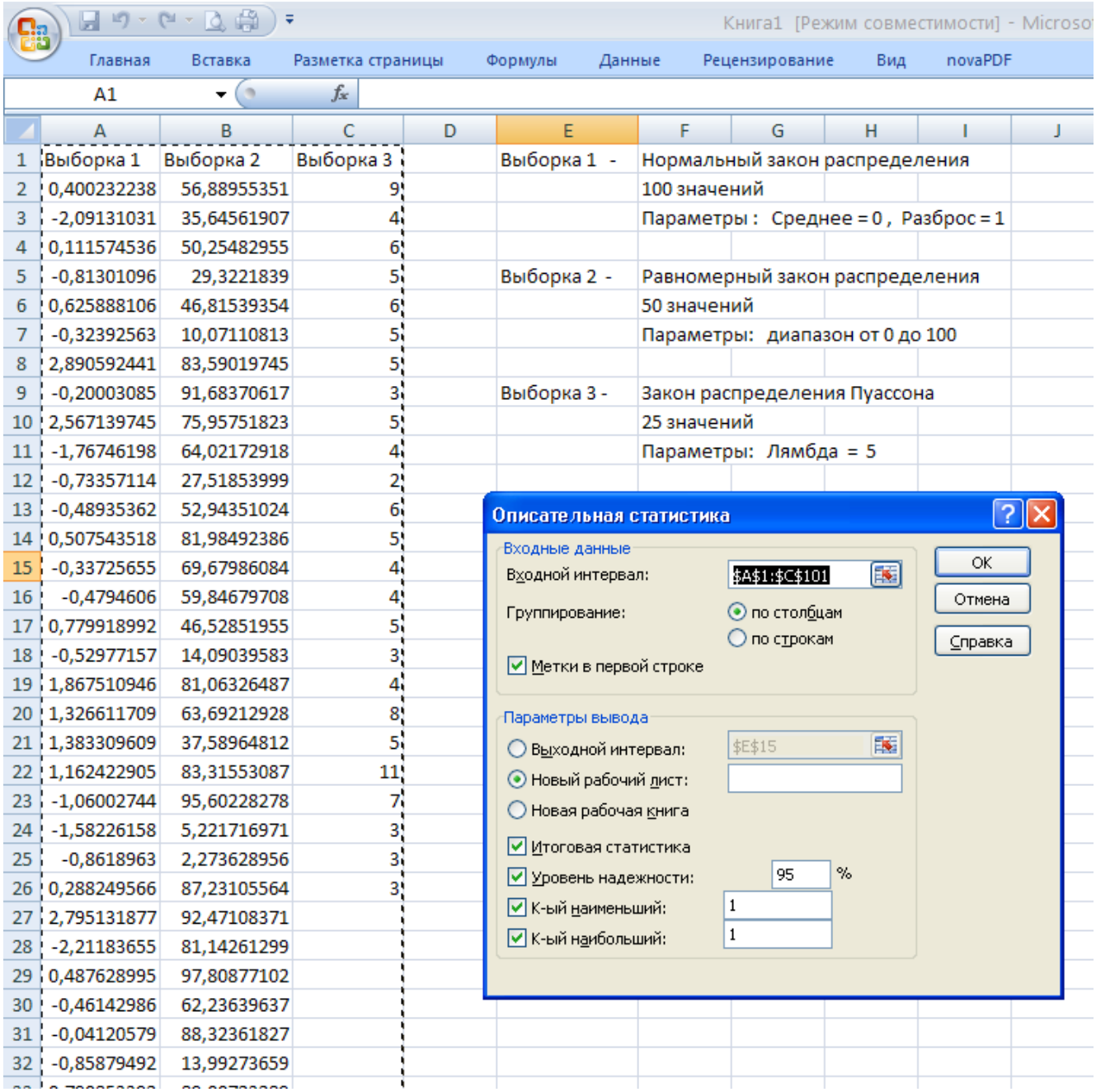
В области ***Параметры вывода***, как правило, надо указать, куда будут выводиться результаты расчетов. Предусмотрено три возможности: на текущий рабочий лист (переключатель ***Выходной интервал***), при этом необходимо указать выходной интервал (достаточно указать адрес одной ячейки, которая определяет верхний левый угол выходного диапазона); на новый рабочий лист текущей рабочей книги начиная с ячейки A1 (переключатель Новый рабочий лист), при этом можно сразу задать имя этому листу; в новую рабочую книгу (переключатель Новая рабочая книга), в этом случае автоматически открывается новая рабочая книга. Также в этой области часто имеются опции, которые указывают, что именно необходимо вывести из возможного набора выходных результатов (например, графики либо дополнительные статистические характеристики).

В некоторых диалоговых окнах имеются другие области, в которых содержатся опции, необходимые для работы данного средства. Эти опции будут приведены при описании конкретных средств. Опции областей ***Входные данные*** и ***Параметры вывода*** будем упоминать только тогда, когда они будут отличаться от описанных выше.

Перейдем к описанию конкретных средств статистического анализа, при этом будем называть их так, как они названы в списке диалогового окна ***Анализ данных***.

1. Описательная статистика

Это средство (вместе со средством Гистограмма, которое будет описано в следующем разделе) является, по-видимому, наиболее часто используемым из всего пакета анализа, поскольку быстро и просто вычисляет основные статистические характеристики одномерных выборок. На рис.3 показан рабочий лист, содержащий три ряда данных (три независимые выборки, имеющие разные распределения) и диалоговое окно ***Описательная статистика***.



	A	B	C
1	Выборка 1	Выборка 2	Выборка 3
2	0,400232238	56,88955351	9
3	-2,09131031	35,64561907	4
4	0,111574536	50,25482955	6
5	-0,81301096	29,3221839	5
6	0,625888106	46,81539354	6
7	-0,32392563	10,07110813	5
8	2,890592441	83,59019745	5
9	-0,20003085	91,68370617	3
10	2,567139745	75,95751823	5
11	-1,76746198	64,02172918	4
12	-0,73357114	27,51853999	2
13	-0,48935362	52,94351024	6
14	0,507543518	81,98492386	5
15	-0,33725655	69,67986084	4
16	-0,4794606	59,84679708	4
17	0,779918992	46,52851955	5
18	-0,52977157	14,09039583	3
19	1,867510946	81,06326487	4
20	1,326611709	63,69212928	8
21	1,383309609	37,58964812	5
22	1,162422905	83,31553087	11
23	-1,06002744	95,60228278	7
24	-1,58226158	5,221716971	3
25	-0,8618963	2,273628956	3
26	0,288249566	87,23105564	3
27	2,795131877	92,47108371	
28	-2,21183655	81,14261299	
29	0,487628995	97,80877102	
30	-0,46142986	62,23639637	
31	-0,04120579	88,32361827	
32	-0,85879492	13,99273659	

Рис. 3. Три выборки и диалоговое окно ***Описательная статистика***

Отметим, что в данном случае имеются выборки разных размеров. Средство ***Описательная статистика*** правильно определяет размеры выборок, игнорируя пустые ячейки. На рис.4 показан рабочий лист с результатами расчетов.

В табл.2 перечислены вычисляемые средством Описательная статистика статистические характеристики выборок, а также функции, которые возвращают те же самые характеристики.

	A	B	C	D	E	F
1	*Выборка 1*		*Выборка 2*		*Выборка 3*	
2						
3	Среднее	0,071886	Среднее	55,67675	Среднее	5
4	Стандартная ошибка	0,111877	Стандартная ошибка	3,977668	Стандартная ошибка	0,41231056
5	Медиана	-0,00011	Медиана	58,36818	Медиана	5
6	Мода	#Н/Д	Мода	#Н/Д	Мода	5
7	Стандартное отклонение	1,118771	Стандартное отклонение	28,12636	Стандартное отклонение	2,06155281
8	Дисперсия выборки	1,251649	Дисперсия выборки	791,0922	Дисперсия выборки	4,25
9	Эксцесс	-0,08594	Эксцесс	-0,96332	Эксцесс	1,97223628
10	Асимметричность	0,245944	Асимметричность	-0,35418	Асимметричность	1,27160622
11	Интервал	5,1186	Интервал	96,78945	Интервал	9
12	Минимум	-2,22801	Минимум	1,019318	Минимум	2
13	Максимум	2,890592	Максимум	97,80877	Максимум	11
14	Сумма	7,188645	Сумма	2783,837	Сумма	125
15	Счет	100	Счет	50	Счет	25
16	Наибольший(1)	2,890592	Наибольший(1)	97,80877	Наибольший(1)	11
17	Наименьший(1)	-2,22801	Наименьший(1)	1,019318	Наименьший(1)	2
18	Уровень надежности(95,0%)	0,221988	Уровень надежности(95,0%)	7,993423	Уровень надежности(95,0%)	0,85096717

Рис. 4. Результаты работы средства ***Описательная статистика***

Таблица 2. Значения, вычисляемые средством Описательная статистика

Значение	***Описание***
Среднее	Выборочное среднее $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ Функция СРЗНАЧ
Стандартная ошибка	Оценка среднеквадратического отклонения выборочного среднего; вычисляется по формуле $\sqrt{\frac{1}{n(n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Медиана	Значение медианы, т.е. квантиля порядка 0,5. Функция МЕДИАНА
Мода	Значение моды. Вычисляется так же, как и функцией МОДА. Если нет одинаковых выборочных значений, то возвращается значение ошибки #Н/Д. Функция МОДА
Стандартное отклонение	Оценка среднеквадратического отклонения генеральной совокупности $s_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$

	Функция СТАНДОТКЛОН
Дисперсия выборки	Оценка дисперсии генеральной совокупности $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ Функция ДИСП
Эксцесс	Выборочный коэффициент эксцесса. Функция ЭКСЦЕСС
Асимметричность	Выборочный коэффициент асимметрии. Функция СКОС
Интервал	Размах выборки. Вычисляется как разность между максимальным и минимальным выборочными значениями.
Минимум	Минимальное выборочное значение. Функция МИН
Максимум	Максимальное выборочное значение. Функция МАКС
Сумма	Сумма выборочных значений. Функция СУММ
Счет	Объем выборки. Функция СЧЁТ
Наибольший (К)	К-е наибольшее значение. Если К = 1, то выводится максимальное (К) выборочное значение. Функция НАИБОЛЬШИЙ
Наименьший (К)	К-е наименьшее значение. Если К = 1, то выводится минимальное (К) выборочное значение. Функция НАИМЕНЬШИЙ
Уровень надежности (X%)	Граница доверительного интервала для неизвестного математического ожидания с доверительным уровнем X%; доверительный интервал строится как выборочное среднее плюс-минус данное значение. Граница вычисляется с помощью распределения Стьюдента, т.е. здесь неявно используется предположение о нормальности распределения генеральной совокупности. Поэтому к данному показателю следует

	относиться осторожно, особенно при малых выборках.

1.1. Опции диалогового окна Описательная статистика

Установка флажка опции ***Итоговая статистика*** указывает, что в итоговом отчете этого средства будут вычислены все статистические характеристики выборки, за исключением границы доверительного интервала для среднего и К-х наибольших и наименьших значений, для которых имеются отдельные опции ***Уровень надежности***, ***К-ый наименьший*** и ***К-ый наибольший***. Если флажок опции ***Итоговая статистика*** не установлен, то выводится только то, что задается с помощью опций ***Уровень надежности***, ***К-ый наименьший*** и ***К-ый наибольший***.

Опция ***Уровень надежности*** указывает, надо ли вычислять границу доверительного интервала для среднего. В поле ввода рядом с этой опцией задается доверительный уровень в процентах.

В полях ввода рядом с опциями ***К-ый наибольший*** и ***К-ый наименьший*** указываются порядки выводимых наибольшего и наименьшего значений. Если эти порядки равны 1, то выводятся соответственно максимальное и минимальное выборочные значения.

2. Гистограмма

Это средство полезно для первичного анализа распределения выборки и построения гистограмм (столбцовых диаграмм эмпирических плотностей вероятностей). В качестве исходных данных нужно указать входной диапазон, содержащий выборочные значения, и интервал карманов. Интервал карманов определяет границы для столбцов гистограммы. Средство ***Гистограмма*** подсчитывает число выборочных значений, попавших в каждый карман (эти числа в выходных данных называются ***Частота***), и по этим числам строит гистограмму. Далее последовательно суммируются частоты (подсчитываются так называемые накапливающие суммы), эти суммы делятся на объем выборки и умножаются на 100. Получается то, что здесь называется ***Интегральный процент***. На самом деле, если убрать проценты (т.е. накапливающие суммы нормировать не на 100%, а на 1), это просто эмпирическая функция распределения. Средство ***Гистограмма*** предоставляет возможность вывести значения интегрального процента в виде графика. В качестве дополнительной возможности предусмотрена сортировка частот по убыванию и построение гистограммы по этим отсортированным частотам.

2.1. Опции диалогового окна Гистограмма

Диалоговое окно ***Гистограмма*** показано на рис.5. В области ***Входные данные*** задаются адрес диапазона ячеек с выборочными значениями (поле ввода ***Входной интервал***) и адрес диапазона, содержащего границы карманов (поле ввода ***Интервал карманов***). Границы карманов должны быть представлены в порядке возрастания. При подсчете количества попаданий выборочных значений в карманы в число попавших в данный карман включаются значения, равные нижней границе кармана и

меньшие верхней границы кармана. Если не указывать интервал границ карманов, будут автоматически созданы равновеликие интервалы, количество которых определяется по формуле Стерджесса:

$$k = [1 + 3{,}322 * \lg(n)] \text{ или } \quad k = [1 + \log_2(n)], \quad \text{где}$$

n – число единиц совокупности, [#] — целая часть числа.

При дробном значении k берется ближайшее к нему целое. Используйте функцию

ОКРУГЛВВЕРХ (число; число_разрядов) .

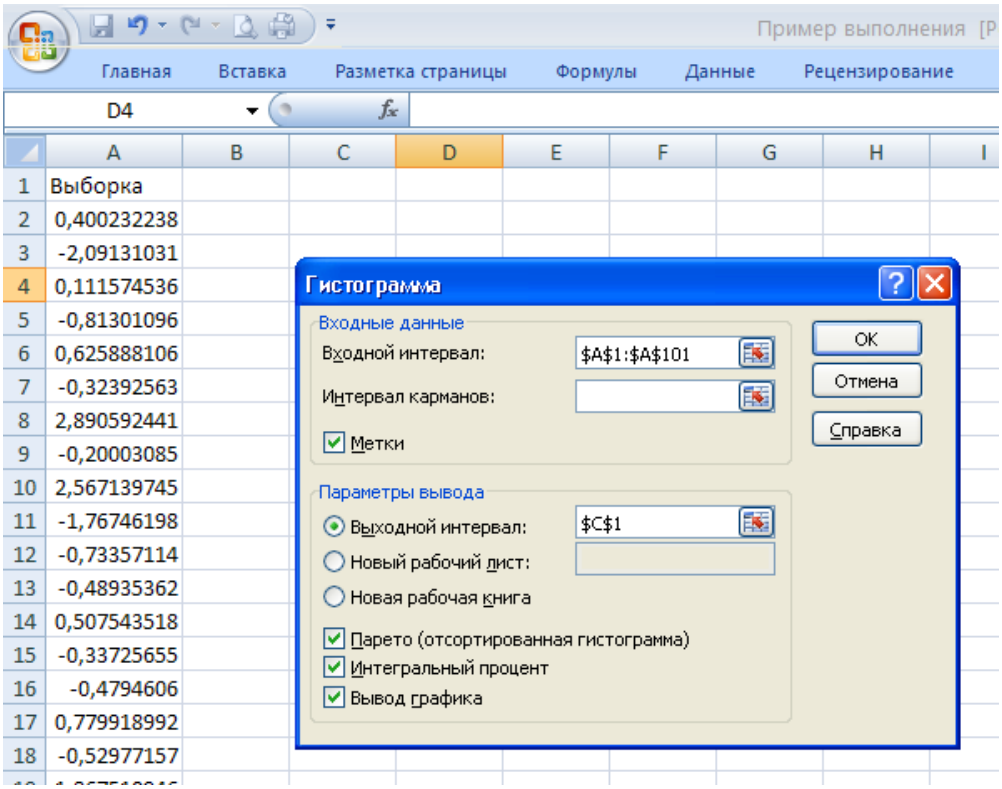
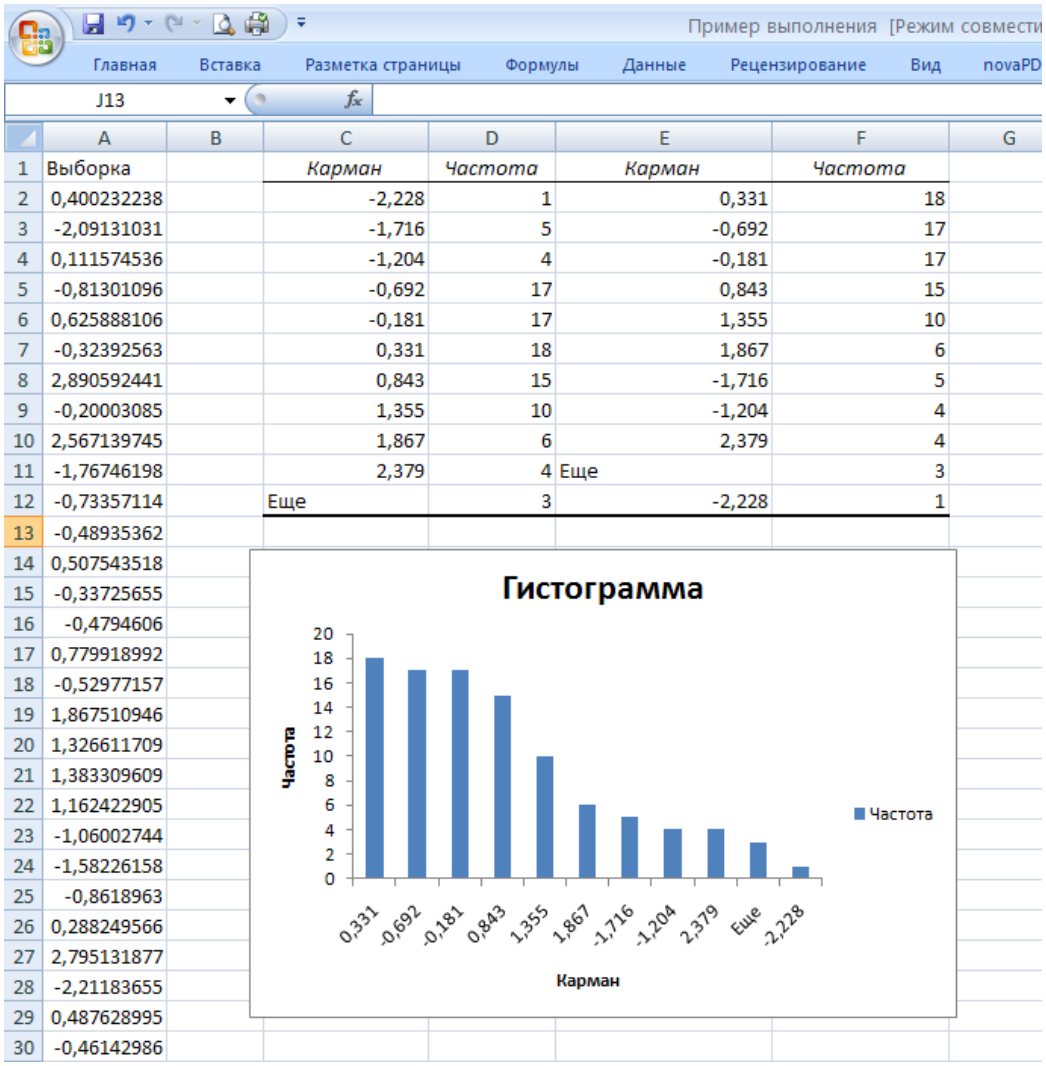


Рис.5. Исходные данные и диалоговое окно Гистограмма

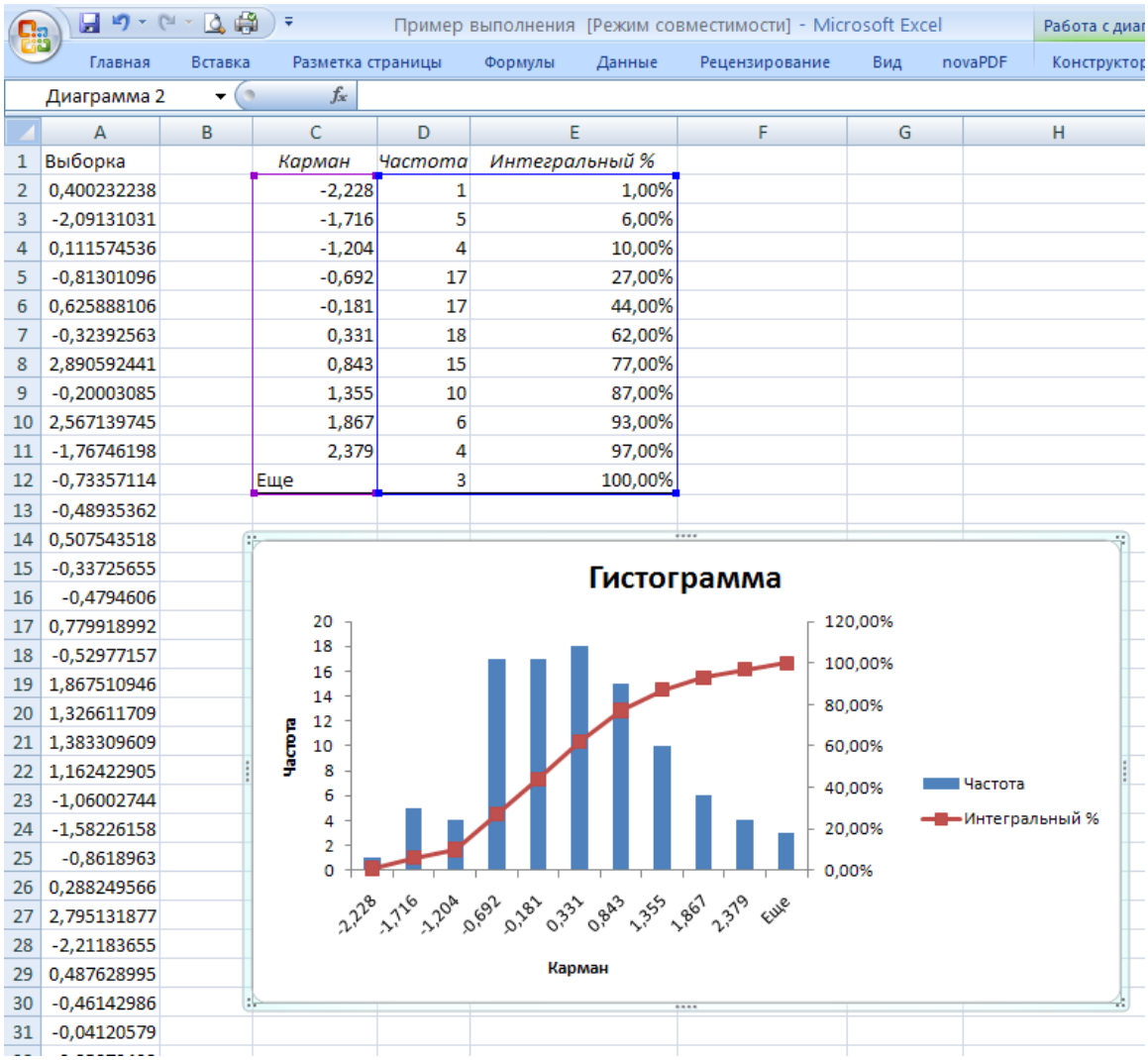
Рассмотрим опции ***Парето*** (отсортированная гистограмма), ***Интегральный процент*** и ***Вывод графика*** из области ***Параметры вывода***. Если установлен только флажок опции ***Парето*** (отсортированная гистограмма), то выводятся таблица частот и таблица отсортированных в порядке убывания частот. Если также установлен флажок опции ***Вывод графика***, выводится гистограмма отсортированных частот, как показано на рис. 6.



	A	B	C	D	E	F	G
1	Выборка		*Карман*	*Частота*	*Карман*	*Частота*	
2	0,400232238		-2,228	1	0,331	18	
3	-2,09131031		-1,716	5	-0,692	17	
4	0,111574536		-1,204	4	-0,181	17	
5	-0,81301096		-0,692	17	0,843	15	
6	0,625888106		-0,181	17	1,355	10	
7	-0,32392563		0,331	18	1,867	6	
8	2,890592441		0,843	15	-1,716	5	
9	-0,20003085		1,355	10	-1,204	4	
10	2,567139745		1,867	6	2,379	4	
11	-1,76746198		2,379	4	Еще	3	
12	-0,73357114		Еще	3	-2,228	1	
13	-0,48935362						
14	0,507543518						
15	-0,33725655						
16	-0,4794606						
17	0,779918992						
18	-0,52977157						
19	1,867510946						
20	1,326611709						
21	1,383309609						
22	1,162422905						
23	-1,06002744						
24	-1,58226158						
25	-0,8618963						
26	0,288249566						
27	2,795131877						
28	-2,21183655						
29	0,487628995						
30	-0,46142986						

Рис..6. Отсортированная гистограмма

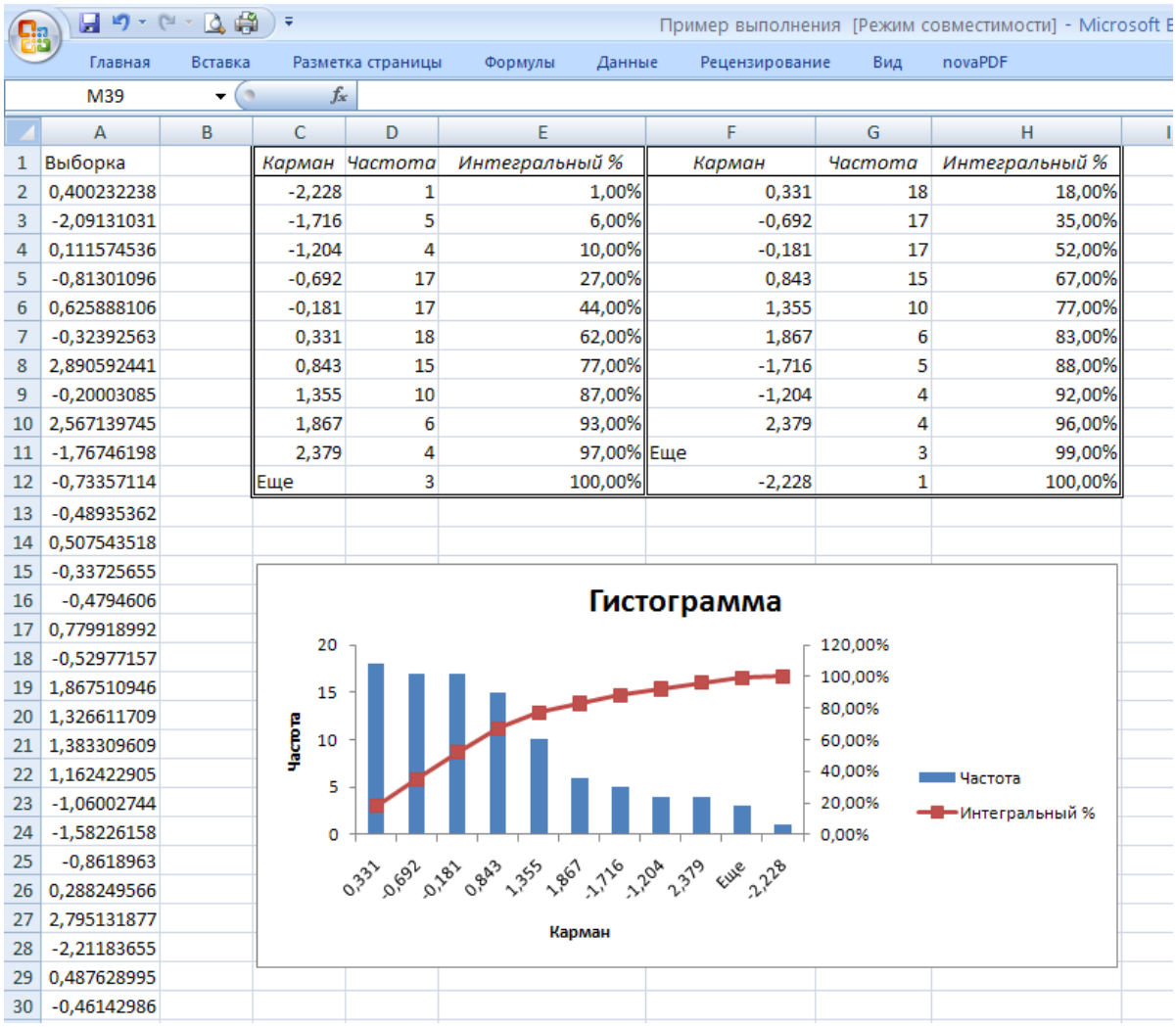
Если установлен только флажок опции ***Интегральный процент***, то выводится таблица, содержащая частоты и значения интегрального процента. Если еще установлен флажок опции ***Вывод графика***, эти данные также отображаются графически, как показано на рис.7.



	A	B	C	D	E
1	Выборка		*Карман*	*Частота*	*Интегральный %*
2	0,400232238		-2,228	1	1,00%
3	-2,09131031		-1,716	5	6,00%
4	0,111574536		-1,204	4	10,00%
5	-0,81301096		-0,692	17	27,00%
6	0,625888106		-0,181	17	44,00%
7	-0,32392563		0,331	18	62,00%
8	2,890592441		0,843	15	77,00%
9	-0,20003085		1,355	10	87,00%
10	2,567139745		1,867	6	93,00%
11	-1,76746198		2,379	4	97,00%
12	-0,73357114		Еще	3	100,00%
13	-0,48935362				
14	0,507543518				
15	-0,33725655				
16	-0,4794606				
17	0,779918992				
18	-0,52977157				
19	1,867510946				
20	1,326611709				
21	1,383309609				
22	1,162422905				
23	-1,06002744				
24	-1,58226158				
25	-0,8618963				
26	0,288249566				
27	2,795131877				
28	-2,21183655				
29	0,487628995				
30	-0,46142986				
31	-0,04120579				

Рис.7. Гистограмма частот и график интегрального процента

Если установлены флажки опций ***Парето*** (отсортированная гистограмма) и ***Интегральный процент***, то выводятся две таблицы: одна содержит неотсортированные частоты и интегральные проценты, вторая — отсортированные частоты и соответствующие интегральные проценты (рис.8). Если также установлен флажок опции ***Вывод графика***, выводятся гистограмма и график интегрального процента, построенные по отсортированным частотам.



	A	B	C	D	E	F	G	H
1	Выборка		*Карман*	*Частота*	*Интегральный %*	*Карман*	*Частота*	*Интегральный %*
2	0,400232238		-2,228	1	1,00%	0,331	18	18,00%
3	-2,09131031		-1,716	5	6,00%	-0,692	17	35,00%
4	0,111574536		-1,204	4	10,00%	-0,181	17	52,00%
5	-0,81301096		-0,692	17	27,00%	0,843	15	67,00%
6	0,625888106		-0,181	17	44,00%	1,355	10	77,00%
7	-0,32392563		0,331	18	62,00%	1,867	6	83,00%
8	2,890592441		0,843	15	77,00%	-1,716	5	88,00%
9	-0,20003085		1,355	10	87,00%	-1,204	4	92,00%
10	2,567139745		1,867	6	93,00%	2,379	4	96,00%
11	-1,76746198		2,379	4	97,00%	Еще	3	99,00%
12	-0,73357114		Еще	3	100,00%	-2,228	1	100,00%
13	-0,48935362							
14	0,507543518							
15	-0,33725655							
16	-0,4794606							
17	0,779918992							
18	-0,52977157							
19	1,867510946							
20	1,326611709							
21	1,383309609							
22	1,162422905							
23	-1,06002744							
24	-1,58226158							
25	-0,8618963							
26	0,288249566							
27	2,795131877							
28	-2,21183655							
29	0,487628995							
30	-0,46142986							

Рис..8. Выходные данные (две таблицы)

Наконец, если установлен флажок только опции Вывод графика, выводятся таблица частот (не отсортированная) и гистограмма.

3. Генерация случайных чисел

Это средство предназначено для генерирования значений случайных чисел, имеющих заданное распределение, т.е. для получения случайных выборок. Средство имеет возможность генерировать случайные числа, имеющие следующие распределения.

- **Равномерное.** Генерируется последовательность равномерно распределенных случайных чисел в заданном интервале, для чего необходимо указать верхнюю и нижнюю границы интервала.
- **Нормальное.** Генерируется последовательность случайных чисел, подчиняющихся нормальному распределению. Задается математическое ожидание и среднеквадратическое отклонение.
- **Бернулли.** Генерируется последовательность случайных чисел, принимающих только значение 0 или 1, в зависимости от заданной вероятности успеха (исхода "1").
- **Биномиальное.** Генерируется последовательность случайных чисел, равная количеству исходов "1" в n независимых испытаниях. В результате каждого из них с вероятностью p может

произойти исход "1" и с вероятностью (1 - *p)* — исход "О". Здесь необходимо задать число испытаний *n* и вероятность *p.*

• **Пуассона.** Генерируется последовательность случайных чисел, подчиняющихся распределению Пуассона с заданным параметром λ .

• **Модельное.** При выборе этого распределения на самом деле генерируются не случайные числа, а повторяющаяся последовательность членов арифметической прогрессии, причем члены прогрессии также могут повторяться заданное число раз. Для этого распределения задаются интервал изменения членов арифметической прогрессии, шаг прогрессии, число повторений членов прогрессии и число повторений этой последовательности чисел.

• **Дискретное.** Генерируется последовательность случайных чисел, подчиняющихся заданному дискретному распределению. Для задания этого распределения необходимо указать диапазон ячеек, состоящий из двух столбцов: в первом столбце содержатся значения, а во втором — вероятности каждого значения. Сумма вероятностей во втором столбце должна быть равна 1.

3.1. Опции диалогового окна Генерация случайных чисел

Диалоговое окно ***Генерация случайных чисел*** при задании различных распределений имеет ряд одинаковых элементов, но наличие некоторых других опций зависит от выбранного типа распределения. Выбор распределения осуществляется в раскрывающемся списке ***Распределение.***

Рассмотрим сначала общие элементы всех диалоговых окон ***Генерация случайных чисел.*** В поле ввода ***Число переменных*** указывается количество генерируемых выборок. Каждая выборка располагается в отдельном столбце. Максимальное количество выборок — 256 (по количеству столбцов в рабочем листе Excel). Если это число не введено, то будет сгенерирована одна случайная выборка, или, если в поле ***Выходной интервал*** указан диапазон ячеек, в котором будут располагаться сгенерированные значения, будут заполнены все столбцы этого диапазона.

В поле ввода ***Число случайных чисел*** задается количество выборочных значений (т.е. объем генерируемых выборок), одно и то же для всех выборок. Если это число не введено, то будет сгенерировано одно значение, или, если в поле ***Выходной интервал*** указан диапазон ячеек, в котором будут располагаться сгенерированные значения, будут заполнены все строки этого диапазона.

В большинстве диалоговых окон ***Генерация случайных чисел*** (кроме окон для модельного и дискретного распределений) имеется поле ввода ***Случайное рассеивание***. Число, введенное в это поле, задает начальное значение, которое будет использовано в алгоритме генерации случайных чисел. Обычно это поле оставляют пустым. Однако, чтобы генерировать одинаковые последовательности случайных чисел, необходимо ввести число из диапазона от 1 до 32 767 (допускаются только целые числа). Тогда в будущем можно получить тот же набор выборочных значений, если в это поле снова ввести то же самое начальное значение.

Все диалоговые окна ***Генерация случайных чисел*** имеют область ***Параметры***; опции этой области зависят от типа выбранного распределения. Назначение большинства этих опций очевидно, но некоторые требуют пояснений.

Равномерное распределение. Диалоговое окно ***Генерация случайных чисел*** для этого распределения показано на рис.9. Здесь в области ***Параметры*** надо задать только верхнюю и нижнюю границы, в пределах которых сосредоточено распределение.

Нормальное распределение. Диалоговое окно ***Генерация случайных чисел*** для этого распределения показано на рис.10. В области ***Параметры*** задаются значения среднего (математического ожидания) и стандартное (среднеквадратическое отклонение). Для стандартного нормального распределения среднее равно 0, а стандартное отклонение — 1.

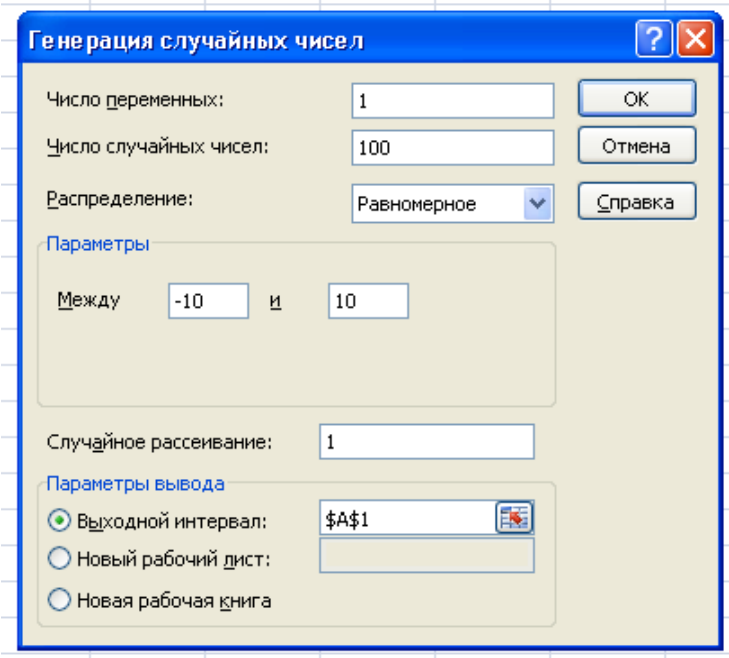


Рис.9. Диалоговое окно для генерирования равномерно распределенных случайных чисел

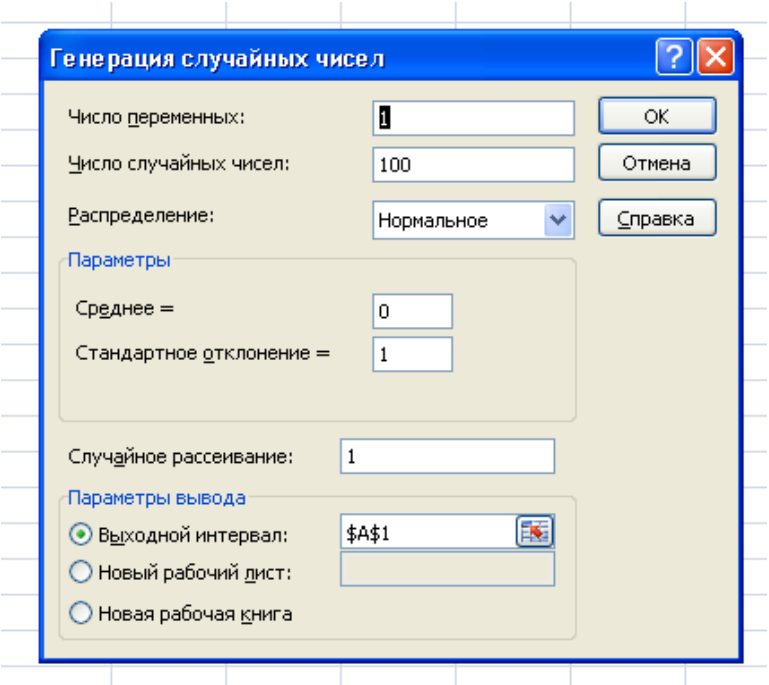


Рис.10. Диалоговое окно для генерирования нормально распределенных случайных чисел

Распределение Бернулли. Диалоговое окно ***Генерация случайных чисел*** для данного случая показано на рис.11. Здесь в области ***Параметры*** задается только один параметр — вероятность *p*.

Биномиальное распределение. Диалоговое окно ***Генерация случайных чисел*** для этого распределения показано на рис.12. Для этого распределения задаются значения вероятности *p* и количество испытаний *n*.

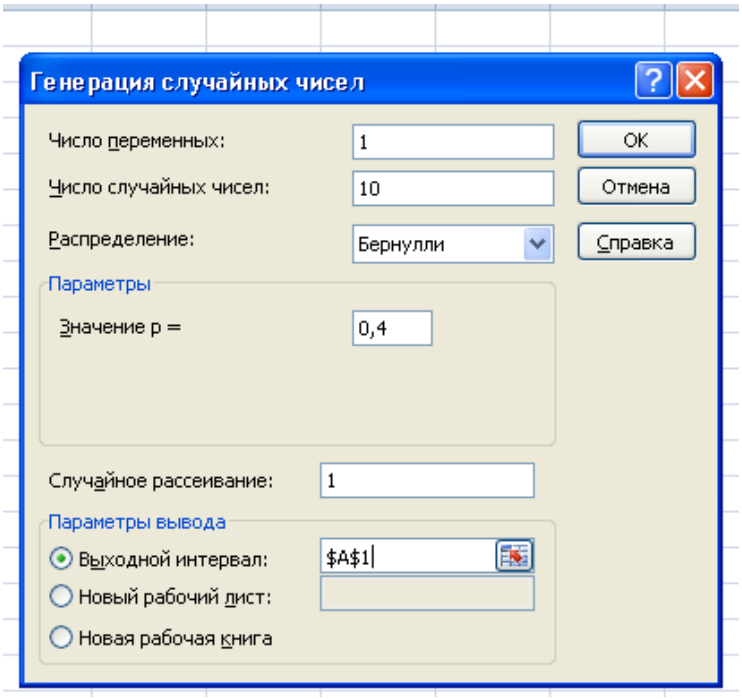


Рис.11. Диалоговое окно для генерирования случайных чисел, имеющих распределение Бернулли

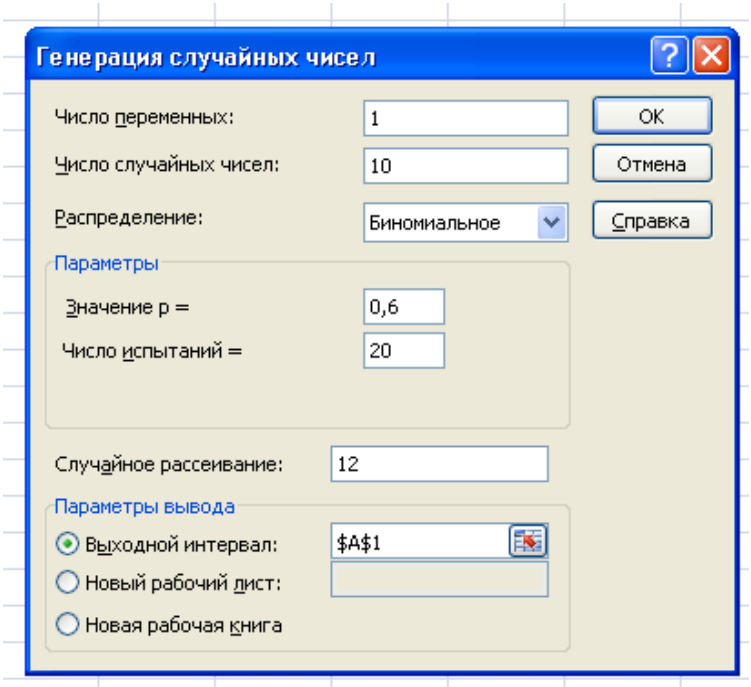


Рис.12. Диалоговое окно для генерирования случайных чисел, имеющих биномиальное распределение

Распределение Пуассона. Диалоговое окно ***Генерация случайных чисел*** для данного случая показано на рис.13. Здесь в области ***Параметры*** задается только один параметр Лямбда.

Модельное распределение. Диалоговое окно ***Генерация случайных чисел*** для этого случая показано на рис.14. Здесь задаются нижняя и верхняя границы чисел, шаг прогрессии, число повторений значений в последовательности и число повторений последовательности. На рис.15 показаны сгенерированные числа с модельным распределением, параметры которого заданы на рис.14.

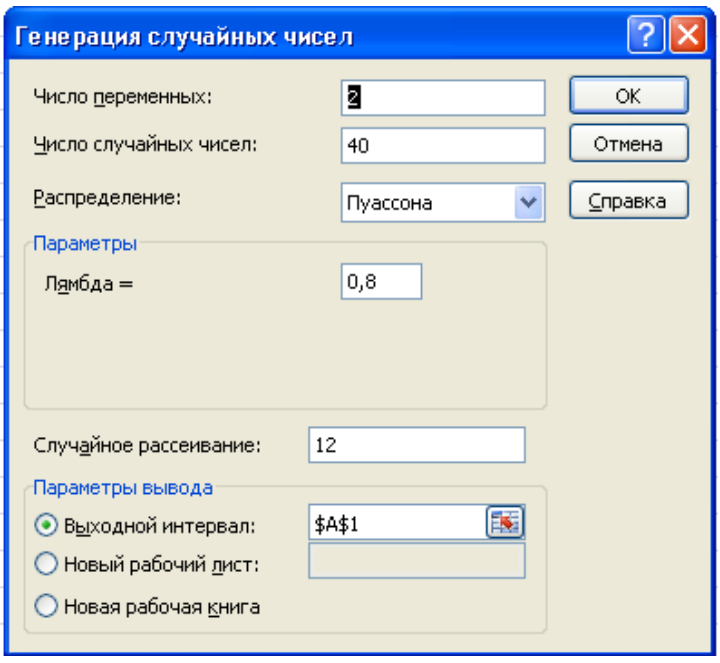


Рис. 13. *Диалоговое окно для генерирования случайных чисел, имеющих распределение Пуассона*

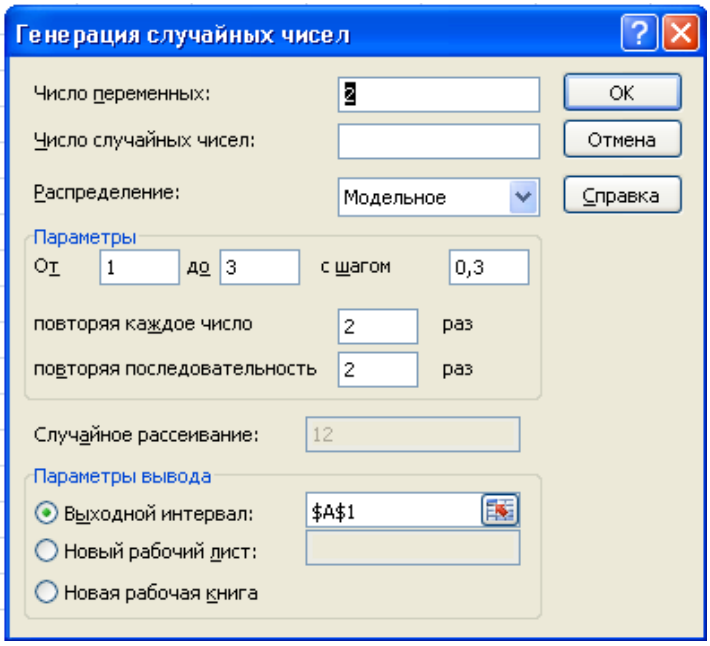
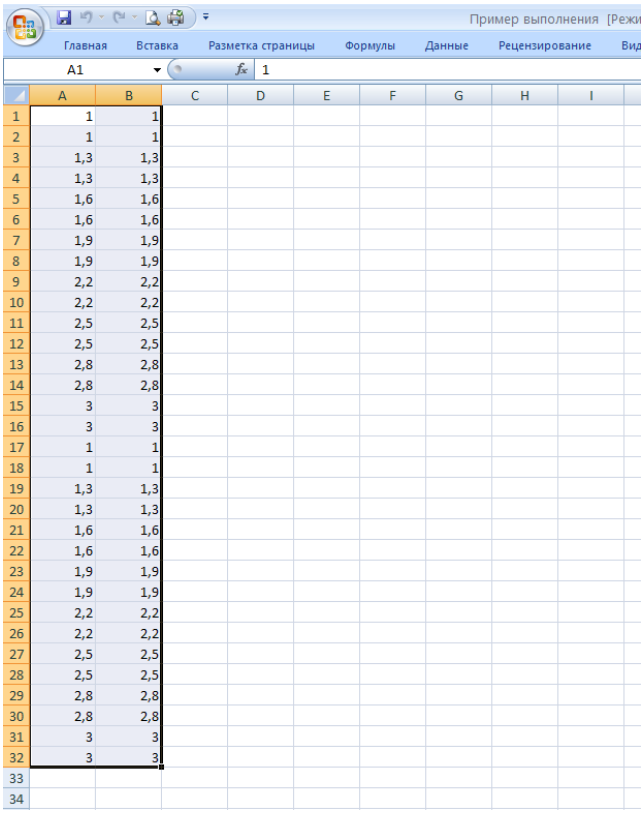


Рис.14. Диалоговое окно для генерирования заданных чисел (модельное распределение)



	A	B
1	1	1
2	1	1
3	1,3	1,3
4	1,3	1,3
5	1,6	1,6
6	1,6	1,6
7	1,9	1,9
8	1,9	1,9
9	2,2	2,2
10	2,2	2,2
11	2,5	2,5
12	2,5	2,5
13	2,8	2,8
14	2,8	2,8
15	3	3
16	3	3
17	1	1
18	1	1
19	1,3	1,3
20	1,3	1,3
21	1,6	1,6
22	1,6	1,6
23	1,9	1,9
24	1,9	1,9
25	2,2	2,2
26	2,2	2,2
27	2,5	2,5
28	2,5	2,5
29	2,8	2,8
30	2,8	2,8
31	3	3
32	3	3
33		
34		

Рис.15. Сгенерированные числа

Дискретное распределение. Диалоговое окно ***Генерация случайных чисел*** для этого типа распределения вместе с необходимыми входными данными показано на рис.16. Для задания дискретного распределения в поле ***Входной интервал*** значений и вероятностей необходимо указать адрес диапазона ячеек, содержащий значения случайной величины и соответствующие им вероятности. Диапазон должен состоять из двух столбцов: левого, содержащего значения, и правого, содержащего

вероятности, как показано на рис.16. Сумма вероятностей должна быть равна 1. На рис.17 представлены сгенерированные числа с распределением, параметры которого заданы на рис.16.

В заключение отметим, что в Excel имеются и другие средства генерирования случайных выборок, например функции СЛЧИС и СЛУЧМЕЖДУ.

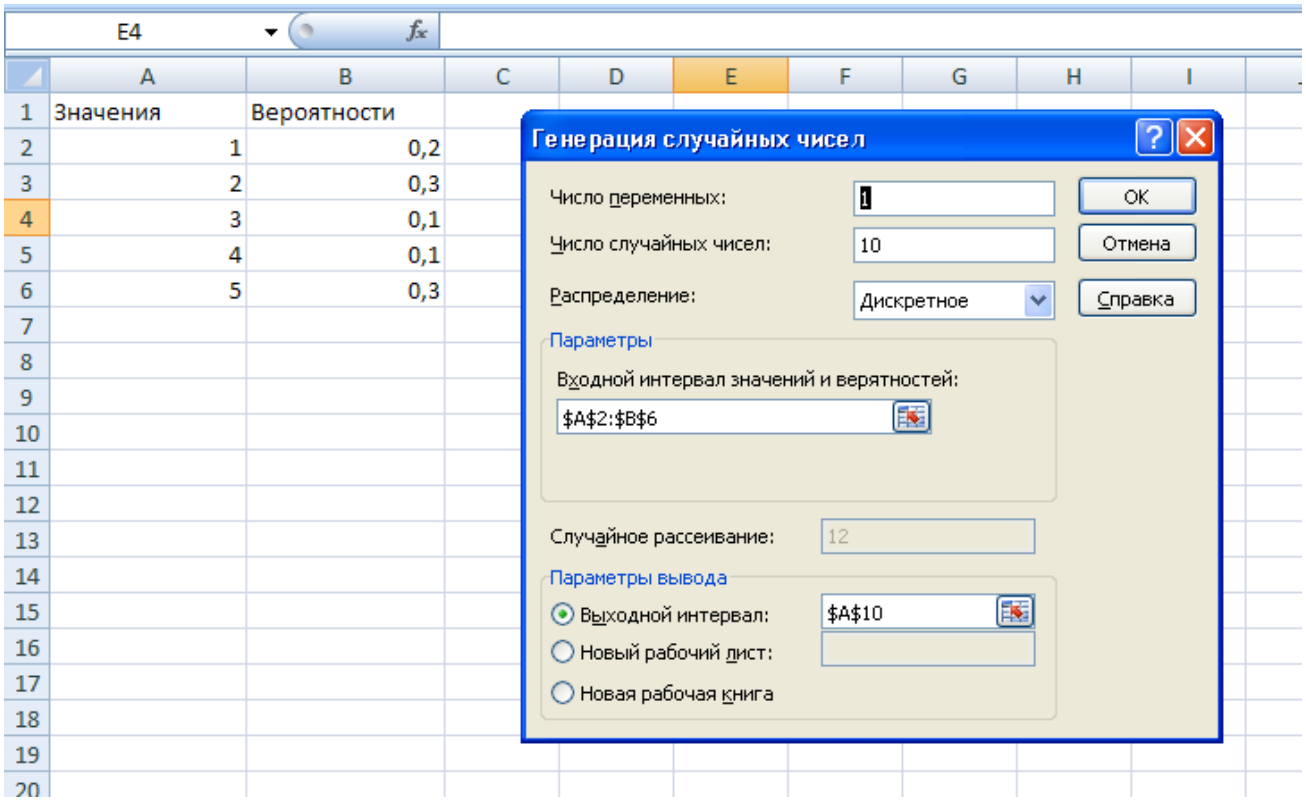
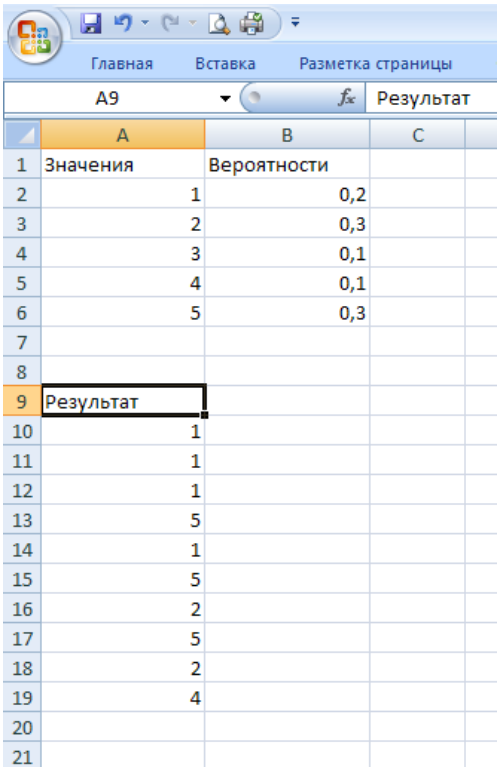


Рис. 16. Диалоговое окно для генерирования случайных чисел, имеющих заданное дискретное распределение



	A	B	C
1	Значения	Вероятности	
2	1	0,2	
3	2	0,3	
4	3	0,1	
5	4	0,1	
6	5	0,3	
7			
8			
9	Результат		
10	1		
11	1		
12	1		
13	5		
14	1		
15	5		
16	2		
17	5		
18	2		
19	4		
20			
21			

Рис. 17. Сгенерированные числа

4. Выборка

Это средство из исходного числового множества выбирает указанное количество чисел, причем либо случайным образом, либо с заданным периодом (например, каждое второе или каждое десятое число). Такую операцию выбора числовых значений из заданного множества можно трактовать как создание выборки заданного объема, если исходное множество рассматривать как генеральную совокупность. Подобная операция часто составляет один из этапов предварительной обработки данных. Например, если исходная выборка слишком велика для обработки или построения диаграмм либо если исходные данные содержат периодическую составляющую, то можно создать выборку, содержащую значения только из отдельных частей периода.

4.1. Опции диалогового окна Выборка

Диалоговое окно ***Выборка*** показано на рис.18. Адрес диапазона ячеек, содержащий исходный набор числовых значений, задается в поле ***Входной интервал***. Если этот диапазон состоит из нескольких столбцов, то значения сначала будут извлекаться из первого столбца, затем из второго столбца и т.д. Средство ***Выборка*** откажется работать (выведет соответствующее окно предупреждения), если среди исходных данных имеются нечисловые значения.

В области ***Метод выборки*** необходимо указать, каким способом будут выбираться значения из исходного множества. Если установлен переключатель ***Периодический***, то из исходного множества будет выбрано каждое *n-е* значение; число *n* задается в поле ввода ***Период***. Количество выбранных

значений будет равно количеству значений в исходном диапазоне, деленному на значение в поле ***Период***. Если установлен переключатель ***Случайный***, значения из исходного множества выбираются случайным образом; количество выбираемых значений задается в поле ***Число выборок***.

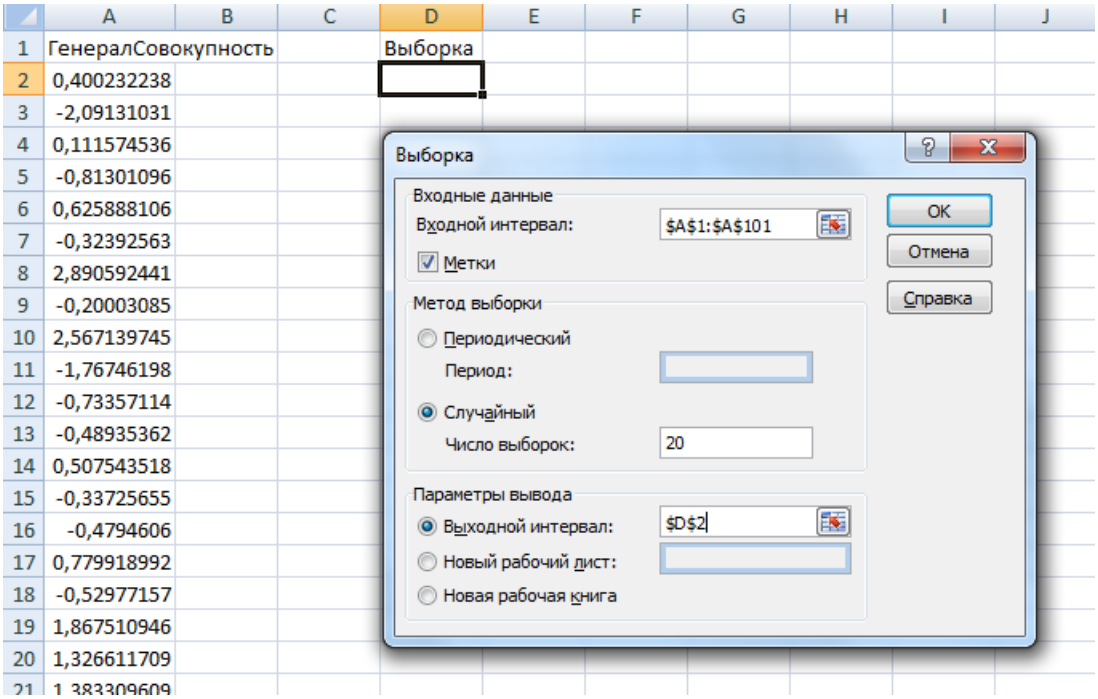


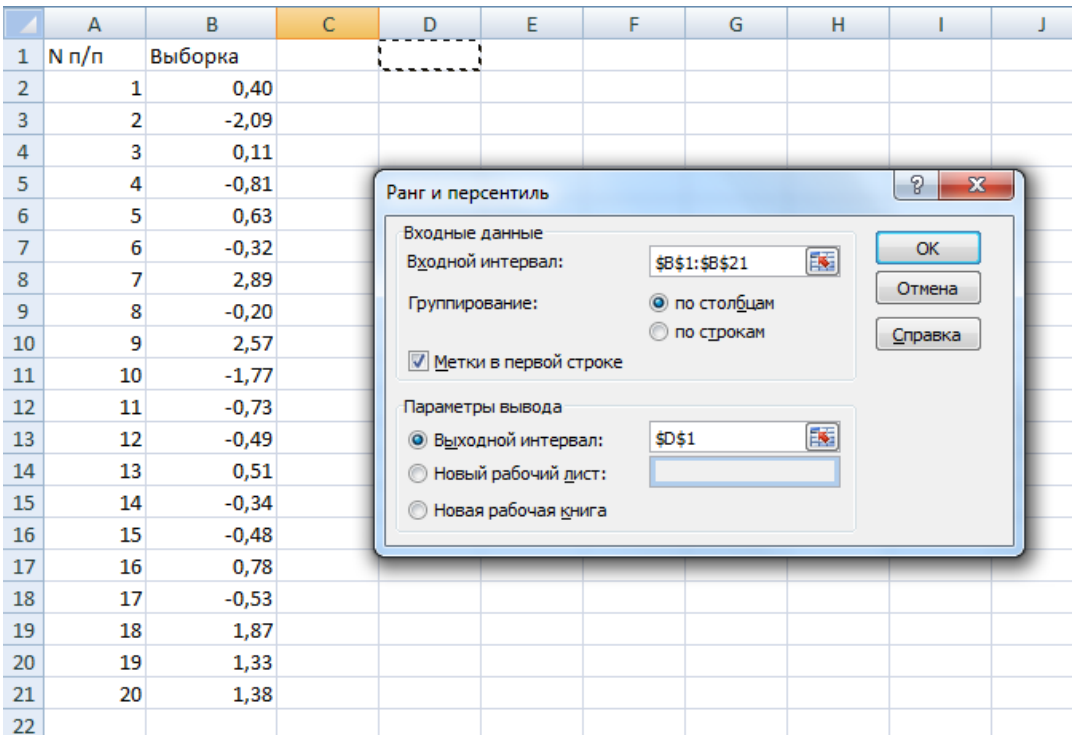
Рис.18 Диалоговое окно Выборка

5. Ранг и персентиль

Это средство позволяет создать таблицу, содержащую порядковый и процентный ранги для каждого значения в заданном наборе данных, при этом значения упорядочиваются в *порядке убывания.* На рис.19 показаны диалоговое окно ***Ранг и персентиль*** и исходные данные, на рис.20 — результат применения этого средства. Итоговая таблица содержит порядковый номер выборочного значения, столбец выборочных значений, отсортированных в порядке убывания, столбец рангов и столбец процентных рангов этих значений, причем наибольшему значению присваивается ранг 1 и процентный ранг 100%, а наименьшему — наибольший ранг и процентный ранг, равный 0%.

Если имеется группа совпадающих значений, то им присваиваются одинаковые ранги, равные рангу первого числа из группы совпадающих значений. Значению, следующему за этой группой, присваивается ранг, больший ранга совпадающих значений на число этих одинаковых значений. Процентный ранг T_i для выборочного значения x_t рассчитывается по формуле

$$T_i = \frac{n - R_i}{n - 1} \cdot 100\%$$, где R_i —ранг значения x_i , рассчитанный при условии упорядочивания данных по убыванию, n — объем выборки.



	A	B	C	D	E	F	G	H	I	J
1	N п/п	Выборка								
2	1	0,40								
3	2	-2,09								
4	3	0,11								
5	4	-0,81								
6	5	0,63								
7	6	-0,32								
8	7	2,89								
9	8	-0,20								
10	9	2,57								
11	10	-1,77								
12	11	-0,73								
13	12	-0,49								
14	13	0,51								
15	14	-0,34								
16	15	-0,48								
17	16	0,78								
18	17	-0,53								
19	18	1,87								
20	19	1,33								
21	20	1,38								
22										

Рис.19 Исходные данные и диалоговое окно ***Ранг и Персентиль***

	A	B	C	D	E	F	G
1	N п/п	Выборка		*Точка*	*Выборка*	*Ранг*	*Процент*
2	1	0,40		7	2,89	1	100,00%
3	2	-2,09		9	2,57	2	94,70%
4	3	0,11		18	1,87	3	89,40%
5	4	-0,81		20	1,38	4	84,20%
6	5	0,63		19	1,33	5	78,90%
7	6	-0,32		16	0,78	6	73,60%
8	7	2,89		5	0,63	7	68,40%
9	8	-0,20		13	0,51	8	63,10%
10	9	2,57		1	0,40	9	57,80%
11	10	-1,77		3	0,11	10	52,60%
12	11	-0,73		8	-0,20	11	47,30%
13	12	-0,49		6	-0,32	12	42,10%
14	13	0,51		14	-0,34	13	36,80%
15	14	-0,34		15	-0,48	14	31,50%
16	15	-0,48		12	-0,49	15	26,30%
17	16	0,78		17	-0,53	16	21,00%
18	17	-0,53		11	-0,73	17	15,70%
19	18	1,87		4	-0,81	18	10,50%
20	19	1,33		10	-1,77	19	5,20%
21	20	1,38		2	-2,09	20	0,00%
22							

Рис.20 Результаты вычислений

Задания к первой части лабораторных работ:

1. Установить и настроить ***Пакет анализа*** в *Microsoft Excel.*

2. Сгенерировать значения случайных чисел, имеющих заданные распределения:

- Равномерное
- Нормальное (в качестве среднего взять порядковый номер студента по списку)
- Бернулли
- Биномиальное
- Пуассона
- Модельное
- Дискретное

3. Изучить параметры, задаваемые для каждого вида распределения. Сохранить каждый вид распределения на отдельном листе в документе *Microsoft Excel.*

4. Получить ***Описательную статистику*** значений случайных чисел для каждого вида распределения:

- с установкой флажка опции ***Итоговая статистика,***
- без установки флажка опции ***Итоговая статистика.***

Результаты сохранить на соответствующем листе *Microsoft Excel.* Изучить набор вычисляемых средством ***Описательная статистика*** статистических характеристик выборок, а также набор функций, которые возвращают те же самые характеристики.

5. Построить 3 гистограммы для каждого вида распределения со следующими параметрами:

- Уставлен флажок ***Парето*** (отсортированная гистограмма).
- Уставлен флажок ***Интегральный процент***.
- Уставлен флажок ***Вывод графика***.

Результаты сохранить на соответствующем листе *Microsoft Excel.*

6. Подсчитать порядковый и процентный ранги для каждого значения для нормального распределения, распределения Пуассона и распределения Бернулли. Результаты сохранить на соответствующем листе *Microsoft Excel.*

Контрольные вопросы

1. Перечислите виды распределений случайных чисел, которые можно сгенерировать в *Microsoft Excel.*

2. Чем различается построение Гистограммы с установкой флажка опции ***Итоговая статистика*** от построения без установки флажка?

3. Какие статистические функции представлены в ***Описательной статистике***?

4. Как будет выглядеть Гистограмма в случае если:

Уставлен флажок ***Парето*** (отсортированная гистограмма)?

Уставлен флажок ***Интегральный процент***?

Уставлен флажок ***Вывод графика***?

Часть 2. Статистическая проверка гипотез

Общие сведения

Проверяемую гипотезу обычно называют нулевой (или основной) и обозначают Н0. Наряду с нулевой гипотезой Н0 рассматривают альтернативную, или конкурирующую, гипотезу Н1, являющуюся логическим отрицанием Н0. Например, σэ – σт = 0. Принятие Н1 свидетельствует о наличии различий. Например, σэ – σт ≠ 0.

Правило, по которому гипотеза Н0 отвергается или принимается, называется **статистическим критерием** или **статистическим тестом.**

При проверке гипотезы, т.к. все данные у нас имеют вероятностную природу, неизбежны ошибки. При проверке гипотез возможны ошибки двух родов:

1 рода. в результате статистической проверке нулевая гипотеза (Н0) отклоняется, но на самом деле она верна;

2 рода. когда нулевая гипотеза (Н0) не отклоняется, но на самом деле верна альтернативная гипотеза (Н1).

Ошибки тесно связаны с понятием уровня статистической значимости. Уровень значимости – это вероятность ошибки первого рода при принятии решения. Для обозначения уровня значимости используют α = 0,05; α = 0,01; α = 0,001. Вероятность ошибки (α) показывает процент ошибки, допустимый при статистическом исследовании.

В науке приняты следующие уровни значимости:

α <0,01 – высокозначимый уровень

α <0,05 – значимый уровень

0,05< α <0,1 – не значимо, но есть тенденция (квазизначимый уровень)

Например, в группе из 100 учеников проводилось исследование по уровню утомляемости в начале учебного года и в конце учебного года. Была принята Н0, заключающееся в том, что существенных различий между уровнем утомляемости в начале года и в конце года нет (с уровнем значимости 0,05). Это означает, что различия могут быть только в пяти процентах от общего количества испытуемых, т.е. различия могут наблюдаться у пяти детей из этих ста.

Этапы принятия статистического решения

1. Формулировка Н0 и Н1.

2. Определение объема выборки.

3. Выбор соответствующего уровня значимости (≈ 0,05, 0,01, 0,001).

4. Выбор статистического метода (критерия), который зависит от типа решаемой психологической задачи.

5. Вычисление соответствующего эмпирического значения по экспериментальным данным, согласно выбранному статистическому методу (критерию).

6. Нахождение по таблице для выбранного статистического метода критических значений, соответствующих уровню значимости α = 0,05 и α =0,01 или уровень допустимого значения β = 0,95, β = 0,99.

7. Построение оси значимости и нанесение на нее табличных критических значений и эмпирических значений измеряемой величины.

8. Формулировка принятия решения (принятие гипотезы).

Уровень значимости прямо зависит от того, каким числом степеней свободы обладает данный коэффициент или параметр. Число независимых величин, участвующих в образовании того или другого параметра, называется числом степеней свободы этого параметра. Оно равно общему числу величин, по которым вычисляется параметр, минус число условий, связывающих эти величины. Число степеней свободы и способы его определения всегда даются в окончательных формулах, которыми пользуется исследователь при статистической обработке своих материалов.

6. Двухвыборочный z-тест для средних

Это средство применяется для проверки гипотезы о равенстве (неравенстве) математических ожиданий двух независимых генеральных совокупностей (большие независимые выборки), имеющих нормальное распределение, при известных дисперсиях этих распределений. Пусть имеются две независимые выборки x_1 , x_2, ..., x_n и y_1 , y_2 , ... y_m объемом соответственно *n и m,* извлеченные из совокупностей, имеющих нормальные распределения с известными дисперсиями σ_1^2 и σ_2^2 и неизвестными математическими ожиданиями соответственно μ_1 и μ_2. Проверяется нулевая гипотеза H_o: $\mu_1 - \mu_2 = \delta$ (δ задано). Z-тест позволяет проверить гипотезу H_o против разных конкурирующих гипотез: H_1: $\mu_1 \neq \mu_2 + \delta$ или H_1: $\mu_1 > \mu_2 + \delta$, либо H_1: $\mu_1 < \mu_2 + \delta$. Критериальная статистика вычисляется по формуле

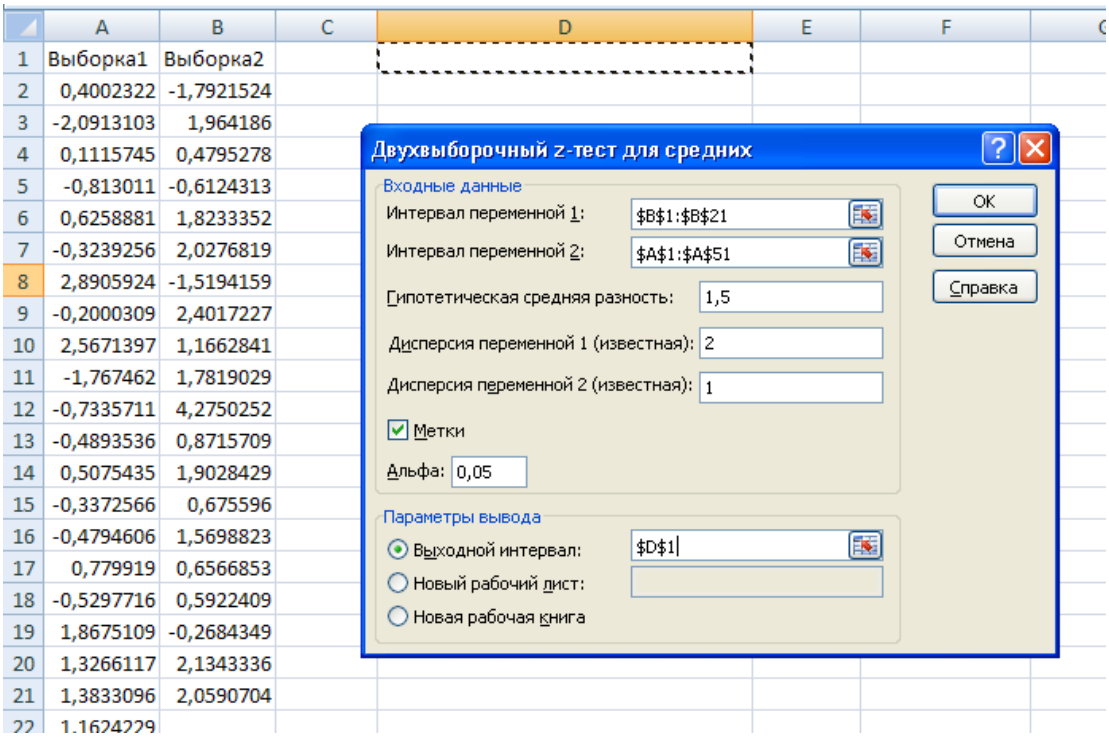
$$z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}},$$

где $\bar{x}$ и $\bar{y}$ — выборочные средние соответственно первой и второй выборок.

Для выборок из нормально распределенных генеральных совокупностей критериальная статистика z имеет стандартное нормальное распределение. Поэтому при заданном уровне значимости α критическая область строится на основе стандартного нормального распределения — вычисляется квантиль ***t*** порядка (1 — α) для проверки гипотезы о равенстве, либо квантиль ***t*** порядка (1 - α/2) для проверки гипотез неравенства. Нулевая гипотеза о равенстве принимается, если $|z| < t$ (в противном случае отвергается); гипотеза H_o при конкурирующей гипотезе H_1: $\mu_1 > \mu_2 + \delta$ принимается, если $z < t$; и при конкурирующей гипотезе H_1: $\mu_1 < \mu_2 + \delta$ нулевая гипотеза принимается при выполнении неравенства $z > -t$.

Рассмотрим пример. Имеется две выборки[1] объемом соответственно 50 и 20 значений, показанные на рис. 21. Обе имеют нормальное распределение, первая — стандартное (т.е. $\mu_1 = 0$ и $\sigma_1^2 = 1$), а для второй $\mu_2 = 1$ и $\sigma_2^2 = 2$. Проверим с помощью средства ***Двухвыборочныи z-тест для средних*** нулевую гипотезу, что $\mu_2 - \mu_1 = 1{,}5$ для разных случаев конкурирующих гипотез. Заполненное диалоговое окно для этого примера также показано на рис.21а.

[1] Выборки получены с помощью средства ***Генерация случайных чисел***

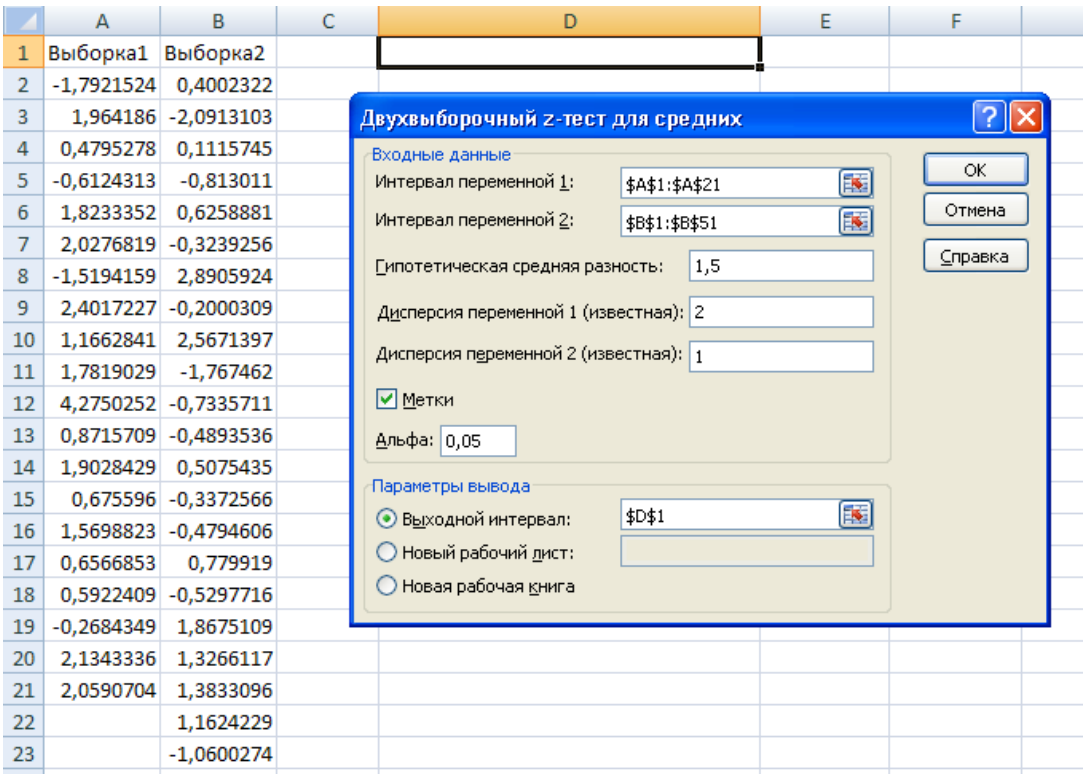


	A	B
1	Выборка1	Выборка2
2	0,4002322	-1,7921524
3	-2,0913103	1,964186
4	0,1115745	0,4795278
5	-0,813011	-0,6124313
6	0,6258881	1,8233352
7	-0,3239256	2,0276819
8	2,8905924	-1,5194159
9	-0,2000309	2,4017227
10	2,5671397	1,1662841
11	-1,767462	1,7819029
12	-0,7335711	4,2750252
13	-0,4893536	0,8715709
14	0,5075435	1,9028429
15	-0,3372566	0,675596
16	-0,4794606	1,5698823
17	0,779919	0,6566853
18	-0,5297716	0,5922409
19	1,8675109	-0,2684349
20	1,3266117	2,1343336
21	1,3833096	2,0590704
22	1,1624229	

Рис.21а Исходные данные и диалоговое окно ***Двухвыборочный z-тест для средних***

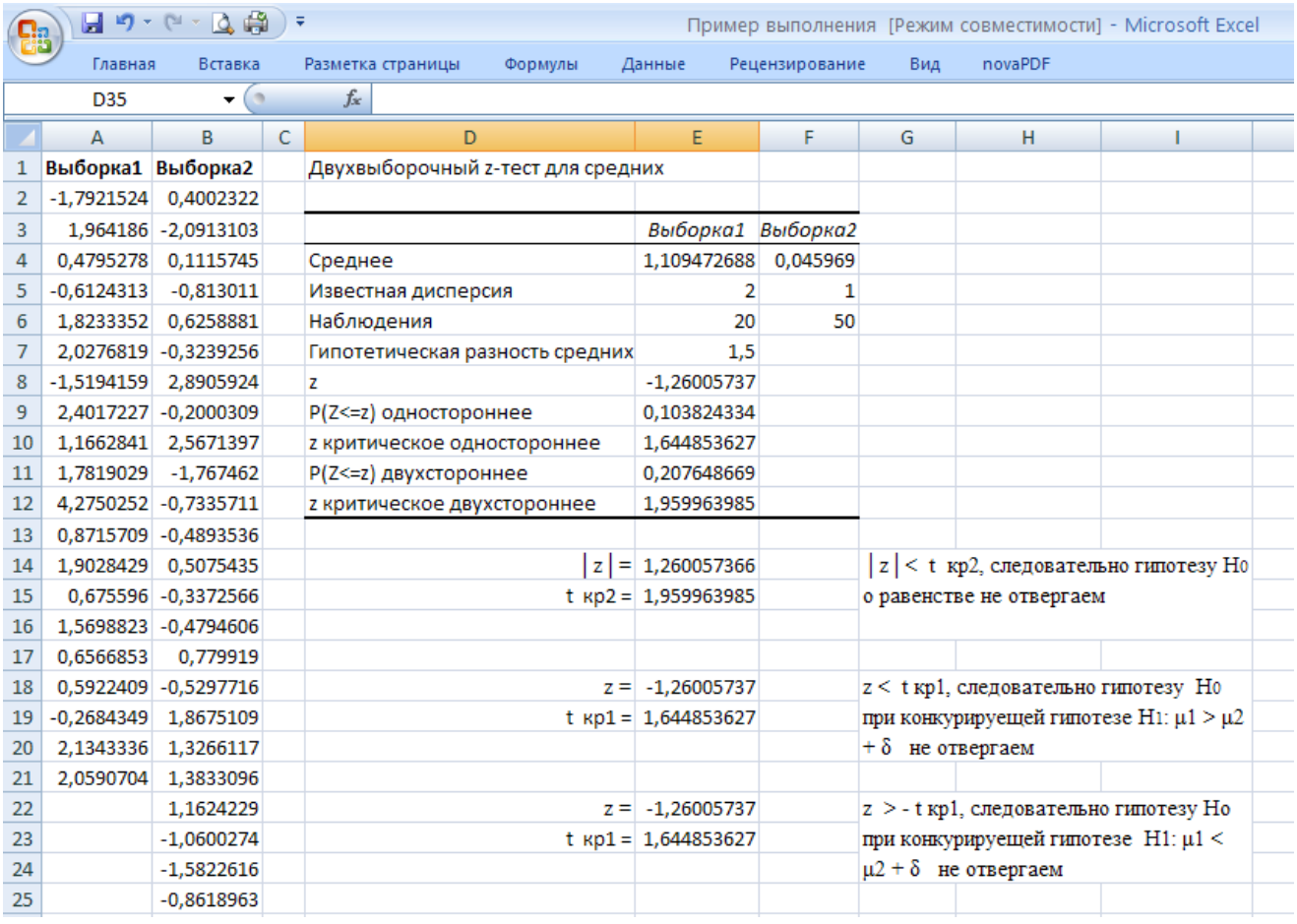
Отметим, что средство требует, чтобы δ, значение которого задается в поле ***Гипотетическая средняя разность***, было неотрицательно. Поэтому первым (в поле ввода ***Интервал переменной 1***) задается адрес диапазона ячеек, содержащий выборку с большим математическим ожиданием, а затем в поле ***Интервал переменной 2*** указывается адрес второй выборки. В полях ввода ***Дисперсия переменной 1*** и ***Дисперсия переменной 2*** вводятся значения дисперсий соответственно первой и второй выборок. В поле ***Альфа*** вводится значение уровня значимости α.

Для удобства анализа результатов сформируем исходные данные таким образом, чтобы выборка с большим математическим ожиданием расположилась в первом столбце. В нашем примере необходимо поменяем местами исходные выборки. Итак, имеем две выборки объемом соответственно 20 и 50 значений, показанные на рис. 21б. Обе имеют нормальное распределение, параметры первой — $\mu_2 = 1$ и $\sigma_2^2=2$; вторая имеет стандартное распределение (т.е. $\mu_1 = 0$ и $\sigma_1^2= 1$). Проверим с помощью средства ***Двухвыборочный z-тест для средних*** нулевую гипотезу, что $\mu_1 - \mu_2 = 1{,}5$ для разных случаев конкурирующих гипотез. Заполненное диалоговое окно для этого примера также показано на рис.21б. Результат вычислений средства ***Двухвыборочный z-тест для средних*** показан на рис.22.



	A	B
1	Выборка1	Выборка2
2	-1,7921524	0,4002322
3	1,964186	-2,0913103
4	0,4795278	0,1115745
5	-0,6124313	-0,813011
6	1,8233352	0,6258881
7	2,0276819	-0,3239256
8	-1,5194159	2,8905924
9	2,4017227	-0,2000309
10	1,1662841	2,5671397
11	1,7819029	-1,767462
12	4,2750252	-0,7335711
13	0,8715709	-0,4893536
14	1,9028429	0,5075435
15	0,675596	-0,3372566
16	1,5698823	-0,4794606
17	0,6566853	0,779919
18	0,5922409	-0,5297716
19	-0,2684349	1,8675109
20	2,1343336	1,3266117
21	2,0590704	1,3833096
22		1,1624229
23		-1,0600274

Рис.21б Исходные данные и диалоговое окно ***Двухвыборочный z-тест для средних***



	A	B	C	D	E	F	G	H	I
1	**Выборка1**	**Выборка2**		Двухвыборочный z-тест для средних					
2	-1,7921524	0,4002322							
3	1,964186	-2,0913103			*Выборка1*	*Выборка2*			
4	0,4795278	0,1115745		Среднее	1,109472688	0,045969			
5	-0,6124313	-0,813011		Известная дисперсия	2	1			
6	1,8233352	0,6258881		Наблюдения	20	50			
7	2,0276819	-0,3239256		Гипотетическая разность средних	1,5				
8	-1,5194159	2,8905924		z	-1,26005737				
9	2,4017227	-0,2000309		P(Z<=z) одностороннее	0,103824334				
10	1,1662841	2,5671397		z критическое одностороннее	1,644853627				
11	1,7819029	-1,767462		P(Z<=z) двухстороннее	0,207648669				
12	4,2750252	-0,7335711		z критическое двухстороннее	1,959963985				
13	0,8715709	-0,4893536							
14	1,9028429	0,5075435		|z| =	1,260057366		|z| < t кр2, следовательно гипотезу Н0		
15	0,675596	-0,3372566		t кр2 =	1,959963985		о равенстве не отвергаем		
16	1,5698823	-0,4794606							
17	0,6566853	0,779919							
18	0,5922409	-0,5297716		z =	-1,26005737		z < t кр1, следовательно гипотезу Н0		
19	-0,2684349	1,8675109		t кр1 =	1,644853627		при конкурирующей гипотезе Н1: μ1 > μ2		
20	2,1343336	1,3266117					+ δ не отвергаем		
21	2,0590704	1,3833096							
22		1,1624229		z =	-1,26005737		z > - t кр1, следовательно гипотезу Но		
23		-1,0600274		t кр1 =	1,644853627		при конкурирующей гипотезе Н1: μ1 <		
24		-1,5822616					μ2 + δ не отвергаем		
25		-0,8618963							

Рис.22. Результат вычислений

В итоговой таблице приводятся следующие данные:

- ***Среднее*** — выборочные средние выборок.
- ***Известная дисперсия*** — дисперсии выборок, которые указаны в диалоговом окне.
- ***Наблюдения*** — объемы выборок.
- ***Гипотетическая разность средних*** — значение δ, которое задано в диалоговом окне.
- ***z*** — значение критериальной статистики.
- ***P(Z<=z) одностороннее*** — вероятность $P(X \le z)$, где X — случайная величина, распределенная по стандартному нормальному закону, z — подсчитанное значение критериальной статистики.
- ***z критическое одностороннее*** — значение квантиля порядка $(1 - \alpha/2)$.
- ***P(Z<=z) двухстороннее*** — вероятность $P(|X| \le |z|)$, где X — случайная величина, распределенная по стандартному нормальному закону, z — подсчитанное значение критериальной статистики.
- ***z критическое двухстороннее*** — значение квантиля порядка $(1 - \alpha)$.

Как видно из результатов расчета, в данном примере нет оснований отвергать нулевую гипотезу при любых конкурирующих гипотезах.

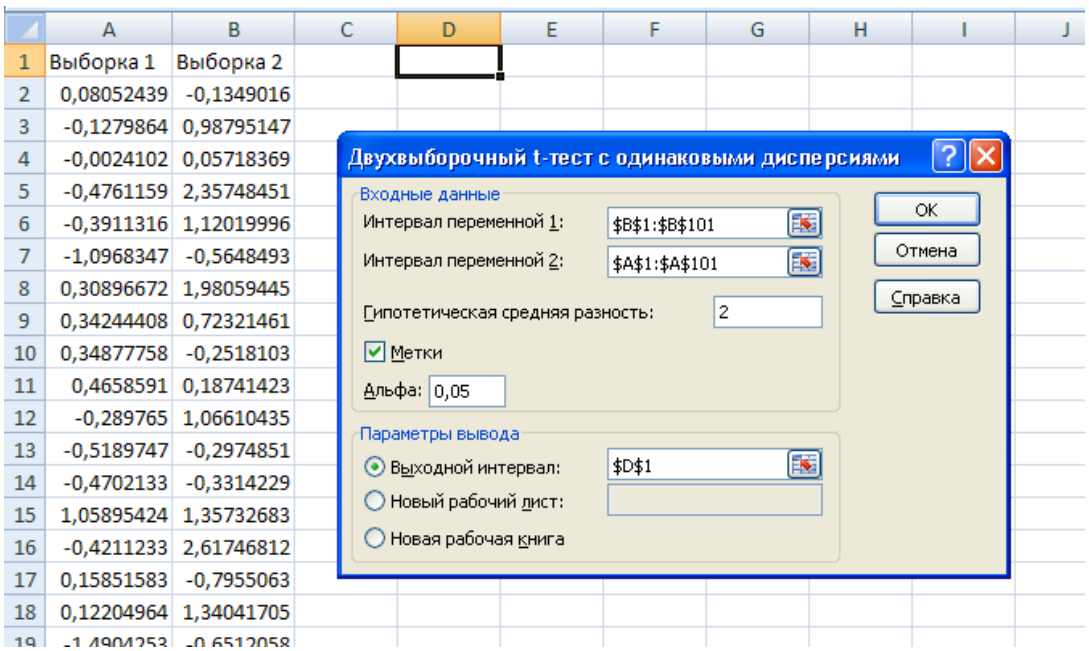
Статистическая функция ***ZTECT*** вычисляет вероятность ***P(Z ≤ z) двухстороннее***.

7. Двухвыборочныи t-тест с одинаковыми дисперсиями

Это средство реализует критерий проверки гипотезы о равенстве (неравенстве) математических ожиданий распределений двух независимых генеральных совокупностей (малые независимые выборки), имеющих нормальные распределения с неизвестными дисперсиями в предположении, что дисперсии равны. Этот критерий, называется *t*-тестом или тестом Стьюдента.

Рассмотрим выходные данные, вычисляемые этим средством, на примере проверки нулевой гипотезы H_o: $\mu_1 - \mu_2 = \delta$ (δ задано) против разных конкурирующих гипотез: H_1: $\mu_1 \neq \mu_2 + \delta$ или H_1: $\mu_1 > \mu_2 + \delta$, либо H_1: $\mu_1 < \mu_2 + \delta$ (μ_1 и μ_2— неизвестные математические ожидания выборок). Исходные данные и заполненное диалоговое окно ***Двухвыборочныи t-тест с одинаковыми дисперсиями*** показаны на рис.23а. Выборки извлечены из нормально распределенных генеральных совокупностей с одной и той же дисперсией, равной 1, и математическими ожиданиями 0 и 1 соответственно[2]. Проверим гипотезу, что $\mu_1 - \mu_2 = 2$ (на самом деле $\mu_1 - \mu_2 = 1$).

[2] Выборки получены с помощью средства ***Генерация случайных чисел***

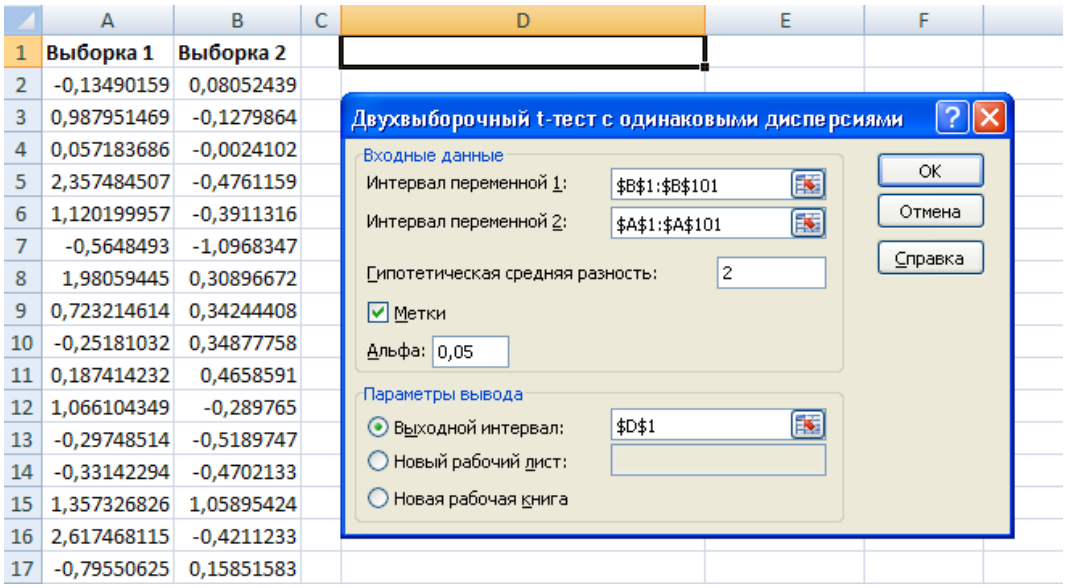


	A	B
1	Выборка 1	Выборка 2
2	0,08052439	-0,1349016
3	-0,1279864	0,98795147
4	-0,0024102	0,05718369
5	-0,4761159	2,35748451
6	-0,3911316	1,12019996
7	-1,0968347	-0,5648493
8	0,30896672	1,98059445
9	0,34244408	0,72321461
10	0,34877758	-0,2518103
11	0,4658591	0,18741423
12	-0,289765	1,06610435
13	-0,5189747	-0,2974851
14	-0,4702133	-0,3314229
15	1,05895424	1,35732683
16	-0,4211233	2,61746812
17	0,15851583	-0,7955063
18	0,12204964	1,34041705

Рис. 23а. Исходные данные и диалоговое окно ***Двухвыборочный t-тест с одинаковыми дисперсиями***

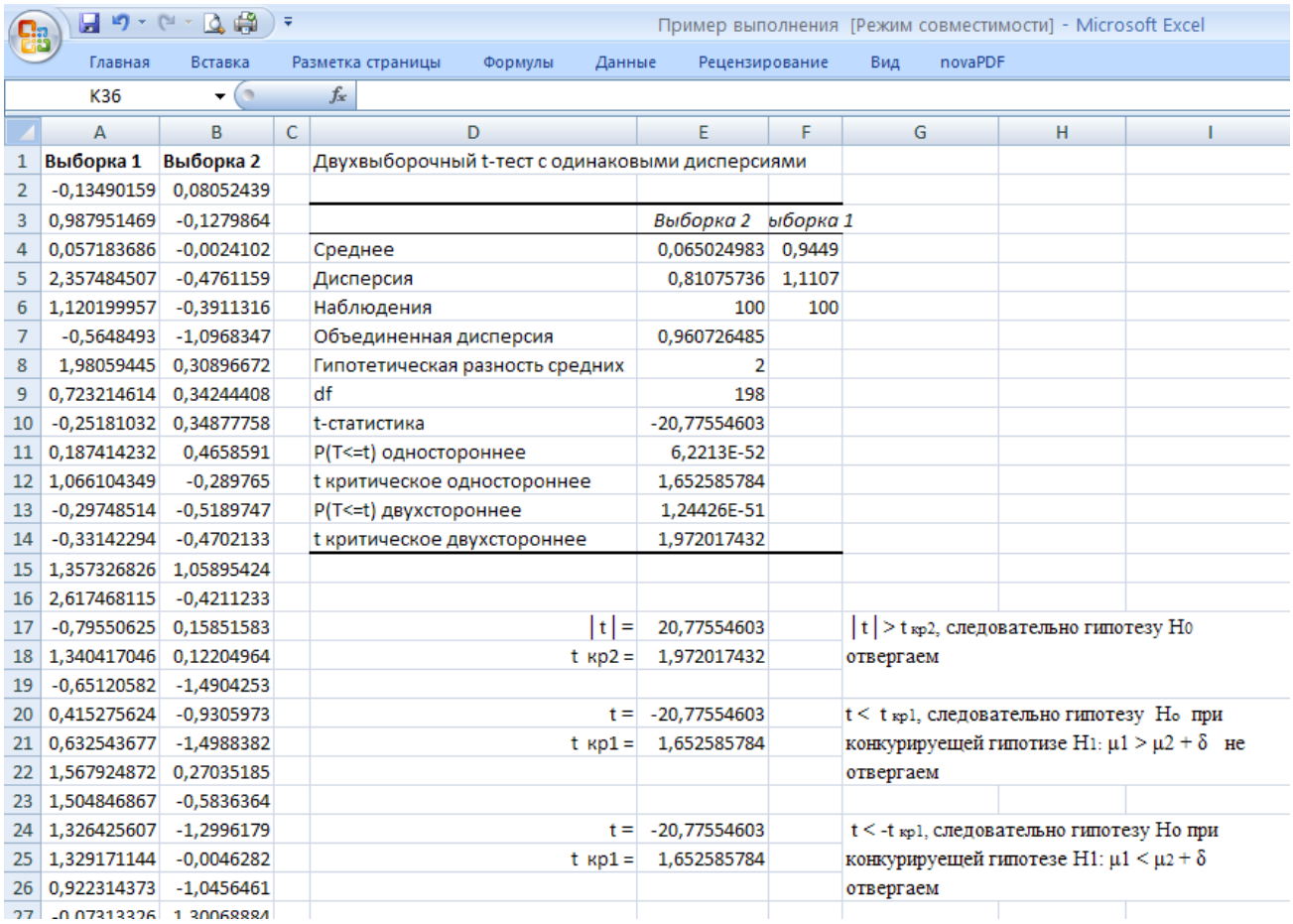
Отметим, что средство требует, чтобы δ, значение которого задается в поле ***Гипотетическая средняя разность***, было неотрицательно. Поэтому первым (в поле ввода ***Интервал переменной 1***) задается адрес диапазона ячеек, содержащий выборку с большим математическим ожиданием, а затем в поле ***Интервал переменной 2*** указывается адрес второй выборки. (Диапазоны должны состоять из одного столбца или одной строки.) В поле ***Альфа*** вводится значение уровня значимости α.

Для удобства анализа результатов сформируем исходные данные таким образом, чтобы выборка с большим математическим ожиданием расположилась в первом столбце. В нашем примере необходимо поменяем местами исходные выборки. Итак, имеем две выборки объемом по 100 значений каждая, показанные на рис. 23б. Выборки имеют одинаковую дисперсию, равную 1, и математическими ожиданиями 1 и 0 соответственно. Результат вычислений средства ***Двухвыборочный t-тест с одинаковыми дисперсиями*** показан на рис.24.



	A	B
1	**Выборка 1**	**Выборка 2**
2	-0,13490159	0,08052439
3	0,987951469	-0,1279864
4	0,057183686	-0,0024102
5	2,357484507	-0,4761159
6	1,120199957	-0,3911316
7	-0,5648493	-1,0968347
8	1,98059445	0,30896672
9	0,723214614	0,34244408
10	-0,25181032	0,34877758
11	0,187414232	0,4658591
12	1,066104349	-0,289765
13	-0,29748514	-0,5189747
14	-0,33142294	-0,4702133
15	1,357326826	1,05895424
16	2,617468115	-0,4211233
17	-0,79550625	0,15851583

Рис. 23б. Исходные данные и диалоговое окно ***Двухвыборочный t-тест с одинаковыми дисперсиями***



	A	B	C	D	E	F	G	H	I
1	**Выборка 1**	**Выборка 2**		Двухвыборочный t-тест с одинаковыми дисперсиями					
2	-0,13490159	0,08052439							
3	0,987951469	-0,1279864			*Выборка 2*	*ыборка 1*			
4	0,057183686	-0,0024102		Среднее	0,065024983	0,9449			
5	2,357484507	-0,4761159		Дисперсия	0,81075736	1,1107			
6	1,120199957	-0,3911316		Наблюдения	100	100			
7	-0,5648493	-1,0968347		Объединенная дисперсия	0,960726485				
8	1,98059445	0,30896672		Гипотетическая разность средних	2				
9	0,723214614	0,34244408		df	198				
10	-0,25181032	0,34877758		t-статистика	-20,77554603				
11	0,187414232	0,4658591		P(T<=t) одностороннее	6,2213E-52				
12	1,066104349	-0,289765		t критическое одностороннее	1,652585784				
13	-0,29748514	-0,5189747		P(T<=t) двухстороннее	1,24426E-51				
14	-0,33142294	-0,4702133		t критическое двухстороннее	1,972017432				
15	1,357326826	1,05895424							
16	2,617468115	-0,4211233							
17	-0,79550625	0,15851583		|t| =	20,77554603		|t| > $t_{кр2}$, следовательно гипотезу H_0		
18	1,340417046	0,12204964		$t_{кр2}$ =	1,972017432		отвергаем		
19	-0,65120582	-1,4904253							
20	0,415275624	-0,9305973		t =	-20,77554603		t < $t_{кр1}$, следовательно гипотезу H_0 при		
21	0,632543677	-1,4988382		$t_{кр1}$ =	1,652585784		конкурирующей гипотизе H_1: $\mu1 > \mu2 + \delta$ не		
22	1,567924872	0,27035185					отвергаем		
23	1,504846867	-0,5836364							
24	1,326425607	-1,2996179		t =	-20,77554603		t < -$t_{кр1}$, следовательно гипотезу Ho при		
25	1,329171144	-0,0046282		$t_{кр1}$ =	1,652585784		конкурирующей гипотезе H1: $\mu1 < \mu2 + \delta$		
26	0,922314373	-1,0456461					отвергаем		
27	-0,07313326	1,30068884							

Рис.24. Результат вычислений

В итоговой таблице приводятся следующие данные.

- ***Среднее*** — выборочные средние для каждой выборки.
- ***Дисперсия*** — несмещенные выборочные оценки дисперсий выборок.
- ***Наблюдения*** — объемы выборок.
- ***Гипотетическая разность средних*** — значение δ, которое задано в диалоговом окне.
- ***Объединенная дисперсия*** — "средняя" оценка дисперсии; рассчитывается по формуле

  $s^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m+2}$, где n и m — объемы выборок, s_i^2 — оценки дисперсий (их значения приводятся в строке ***Дисперсия***).
- ***df*** — число степеней свободы; вычисляется как $(n + m - 2)$.
- ***t-статистика*** — значение критериальной статистики; вычисляется по формуле

  $t = \frac{\sqrt{n+m-2}(\bar{x} - \bar{y} - \delta)}{\sqrt{\frac{n+m}{nm}}\sqrt{(n-1)s_1^2 + (m-1)s_2^2}}$, имеет распределение Стьюдента с ***df*** степенями свободы.

- ***P(T<=t) одностороннее*** — вероятность $P(X \le t)$, где ***X*** — случайная величина, имеющая распределение Стьюдента с ***df*** степенями свободы, ***t*** — подсчитанное значение критериальной статистики.
- ***t критическое одностороннее*** — значение квантиля $t_{кр1}$ порядка $(1 - \alpha)$ распределения Стьюдента с ***df*** степенями свободы.
- ***P(T<=t) двухстороннее*** — вероятность $P(|X|<|t|)$, где ***X*** — случайная величина, имеющая распределение Стьюдента с ***df*** степенями свободы, ***t*** — подсчитанное значение критериальной статистики.
- ***t критическое двухстороннее*** — значение квантиля $t_{кр2}$ порядка (1 - a/2) распределения Стьюдента с ***df*** степенями свободы.

Нулевая гипотеза H_o: $\mu_1 - \mu_2 = \delta$ принимается, если $|t| < t_{кр2}$ (в противном случае отвергается); гипотеза H_o при конкурирующей гипотезе H_1: $\mu_1 > \mu_2 + \delta$ принимается, если $t < t_{кр1}$; при конкурирующей гипотезе H_1: $\mu_1 < \mu_2 + \delta$ нулевая гипотеза принимается при выполнении неравенства $t > -t_{кр1}$.

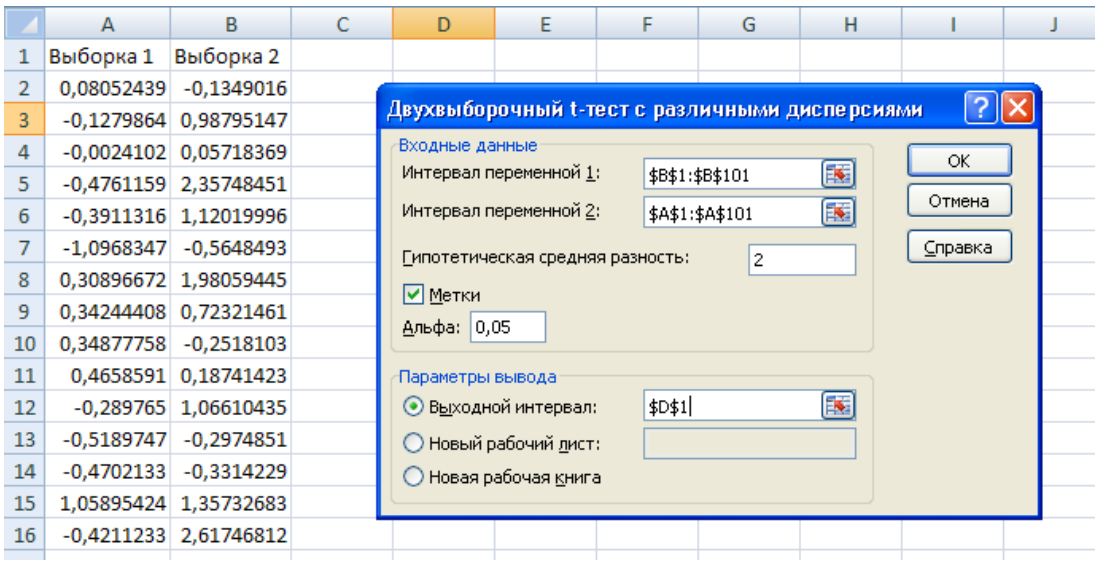
Как видно из результатов расчета, в данном примере нулевую гипотезу H_o: $\mu1 - \mu2 = \delta$ следует отвергнуть при конкурирующей гипотезе о неравенстве $\mu1 - \mu2 \neq \delta$ и конкурирующей гипотезе H_1: $\mu1 < \mu2 + \delta$. При конкурирующей гипотезе H_1: $\mu1 > \mu2 + \delta$ нулевую гипотезу H_o: $\mu1 - \mu2 = \delta$ нет оснований отвергать.

Статистическая функция ***TTECT*** при значении аргумента ***Тип*** = 2 вычисляет вероятности ***P(T<=t) двухстороннее*** и ***P(T<=t) одностороннее***.

8. Двухвыборочный t-тест с различными дисперсиями

Это средство реализует критерий проверки гипотезы о равенстве (неравенстве) математических ожиданий распределений двух независимых генеральных совокупностей, имеющих нормальные распределения с неизвестными и различными дисперсиями. Этот критерий также называется f-тестом или тестом Стьюдента для неравных дисперсий, либо критерием Фишера-Беренса.

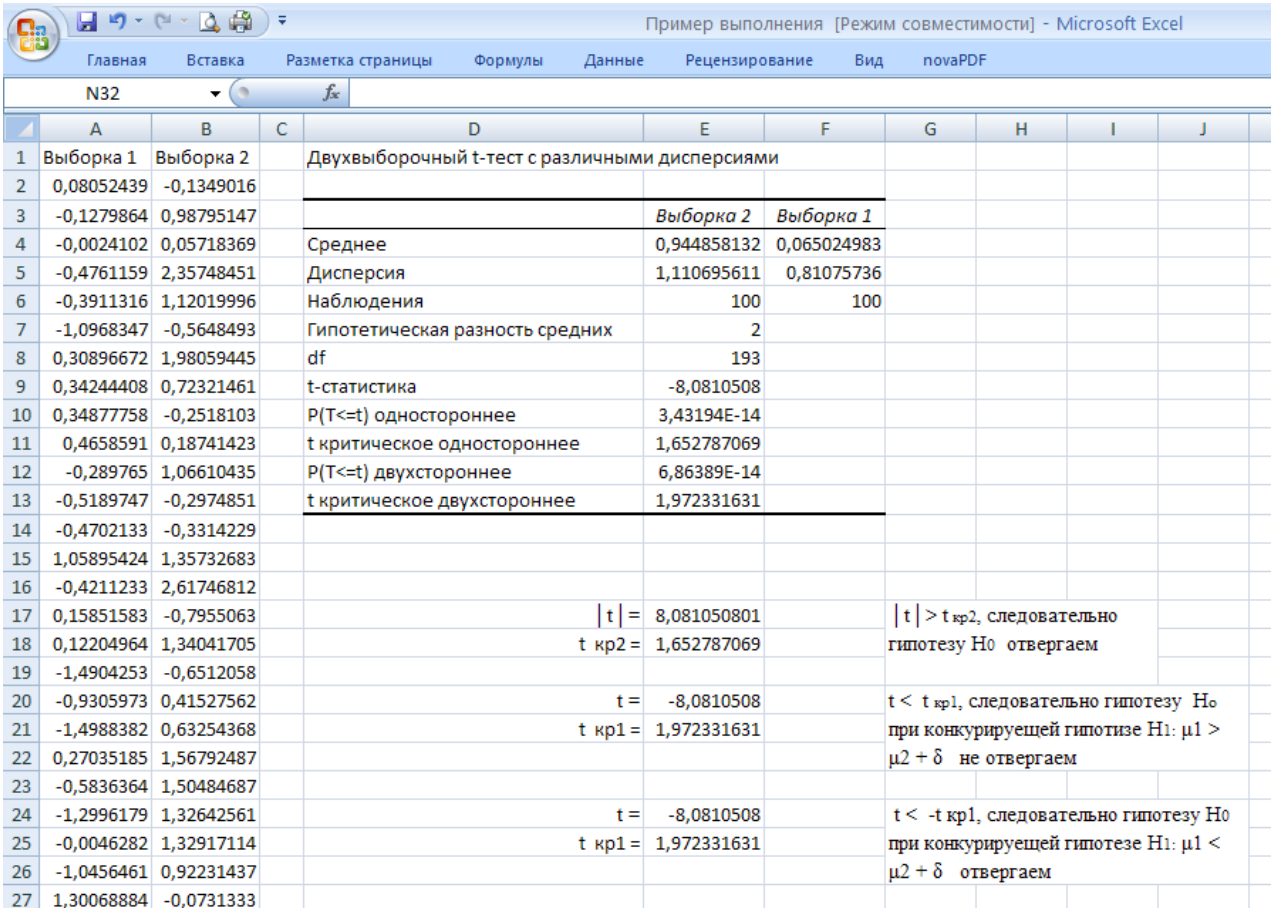
Рассмотрим выходные данные, вычисляемые этим средством, на примере проверки нулевой гипотезы H_o: $\mu_1 - \mu_2 = \delta$ (δ задано) против разных конкурирующих гипотез: H_1: $\mu_1 \neq \mu_2 + \delta$ или H_1: $\mu_1 > \mu_2 + \delta$, либо H_1: $\mu_1 < \mu_2 + \delta$ (μ_1 и μ_2— неизвестные математические ожидания выборок). Повторим тест на примере данных из предыдущего раздела, т.е. выборки извлечены из нормально распределенных генеральных совокупностей с одной и той же дисперсией, равной 1, и математическими ожиданиями соответственно 0 и 1. Проверим гипотезу, что $\mu_1 - \mu_2 = 2$ (на самом деле $\mu_1 - \mu_2 = 1$). Исходные данные и заполненное диалоговое окно ***Двухвыборочный t-тест с различными дисперсиями*** показаны на рис. 25.



	A	B
1	Выборка 1	Выборка 2
2	0,08052439	-0,1349016
3	-0,1279864	0,98795147
4	-0,0024102	0,05718369
5	-0,4761159	2,35748451
6	-0,3911316	1,12019996
7	-1,0968347	-0,5648493
8	0,30896672	1,98059445
9	0,34244408	0,72321461
10	0,34877758	-0,2518103
11	0,4658591	0,18741423
12	-0,289765	1,06610435
13	-0,5189747	-0,2974851
14	-0,4702133	-0,3314229
15	1,05895424	1,35732683
16	-0,4211233	2,61746812

Рис.25 Исходные данные и диалоговое окно ***Двухвыборочный t-тест с различными дисперсиями***

Отметим, что средство требует, чтобы δ, значение которого задается в поле ***Гипотетическая средняя разность***, было неотрицательно. Поэтому первым (в поле ввода ***Интервал переменной 1***) задается адрес диапазона ячеек, содержащий выборку с большим математическим ожиданием, а затем в поле Интервал переменной 2 указывается адрес второй выборки. (Диапазоны должны состоять из одного столбца или одной строки.) В поле ***Альфа*** вводится значение уровня значимости α. Результат вычислений средства ***Двухвыборочный t-тест с различными дисперсиями*** показан на рис.26.



Пример выполнения [Режим совместимости] - Microsoft Excel

Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид novaPDF

N32

	A	B	C	D	E	F	G	H	I	J
1	Выборка 1	Выборка 2		Двухвыборочный t-тест с различными дисперсиями						
2	0,08052439	-0,1349016								
3	-0,1279864	0,98795147			*Выборка 2*	*Выборка 1*				
4	-0,0024102	0,05718369		Среднее	0,944858132	0,065024983				
5	-0,4761159	2,35748451		Дисперсия	1,110695611	0,81075736				
6	-0,3911316	1,12019996		Наблюдения	100	100				
7	-1,0968347	-0,5648493		Гипотетическая разность средних	2					
8	0,30896672	1,98059445		df	193					
9	0,34244408	0,72321461		t-статистика	-8,0810508					
10	0,34877758	-0,2518103		P(T<=t) одностороннее	3,43194E-14					
11	0,4658591	0,18741423		t критическое одностороннее	1,652787069					
12	-0,289765	1,06610435		P(T<=t) двухстороннее	6,86389E-14					
13	-0,5189747	-0,2974851		t критическое двухстороннее	1,972331631					
14	-0,4702133	-0,3314229								
15	1,05895424	1,35732683								
16	-0,4211233	2,61746812								
17	0,15851583	-0,7955063		|t| =	8,081050801		|t| > t кр2, следовательно			
18	0,12204964	1,34041705		t кр2 =	1,652787069		гипотезу H0 отвергаем			
19	-1,4904253	-0,6512058								
20	-0,9305973	0,41527562		t =	-8,0810508		t < t кр1, следовательно гипотезу Ho			
21	-1,4988382	0,63254368		t кр1 =	1,972331631		при конкурирующей гипотизе H1: μ1 >			
22	0,27035185	1,56792487					μ2 + δ не отвергаем			
23	-0,5836364	1,50484687								
24	-1,2996179	1,32642561		t =	-8,0810508		t < -t кр1, следовательно гипотезу H0			
25	-0,0046282	1,32917114		t кр1 =	1,972331631		при конкурирующей гипотезе H1: μ1 <			
26	-1,0456461	0,92231437					μ2 + δ отвергаем			
27	1,30068884	-0,0731333								

Рис.26. Результат вычислений

В итоговой таблице приводятся следующие данные.

- ***Среднее*** — выборочные средние для каждой выборки.
- ***Дисперсия*** — несмещенные выборочные оценки дисперсий выборок.
- ***Наблюдения*** — объемы выборок.
- ***Гипотетическая разность средних*** — значение δ, которое задано в диалоговом окне.
- ***df*** — число степеней свободы; вычисляется по формуле $\frac{\left(\frac{s_1^2}{n}+\frac{s_2^2}{m}\right)^2}{\frac{\left(s_1^2/n\right)^2}{n-1}+\frac{\left(s_2^2/m\right)^2}{m-1}}$, где

  s_1^2 и s_2^2 — несмещенные оценки дисперсий (их значения приводятся в строке ***Дисперсия***),

  n и m — объемы соответственно первой и второй выборок.
- ***t-статистика*** — значение критериальной статистики; вычисляется по формуле $t=\frac{\bar{x}-\bar{y}-\delta}{\sqrt{\frac{s_1^2}{n}+\frac{s_2^2}{m}}}$,

  имеет распределение, близкое к распределению Стьюдента с ***df*** степенями свободы.
- ***P(T<=t) одностороннее*** — вероятность $P(X \le t)$, где X — случайная величина, имеющая распределение Стьюдента с ***df*** степенями свободы, ***t*** — подсчитанное значение критериальной статистики.
- ***t критическое одностороннее*** — значение квантиля $t_{кр1}$ порядка (1 - α) распределения Стьюдента с ***df*** степенями свободы.
- ***P(T<=t) двухстороннее*** — вероятность $P(|X| \le |t|)$, где X — случайная величина, имеющая распределение Стьюдента с ***df*** степенями свободы, ***t*** — подсчитанное значение критериальной статистики.
- ***t критическое двухстороннее*** — значение квантиля $t_{кр2}$ порядка (1 - α/2) распределения Стьюдента с ***df*** степенями свободы.

Нулевая гипотеза H_o: $\mu_1 - \mu_2 = \delta$ принимается, если $|t| < t_{кр2}$, (в противном случае отвергается); гипотеза H_o при конкурирующей гипотезе H_1: $\mu_1 > \mu_2 + \delta$ принимается, если $t < t_{кр1}$, при конкурирующей гипотезе H_1: $\mu_1 < \mu_2 + \delta$ нулевая гипотеза принимается при выполнении неравенства $t > -t_{кр1}$.

Как видно из результатов расчета, в данном примере нулевую гипотезу H_o: μ1 - μ2 = δ следует отвергнуть при конкурирующей гипотезе о неравенстве μ1 - μ2 ≠ δ и конкурирующей гипотезе H_1: μ1 < μ2 + δ. При конкурирующей гипотезе H_1: μ1 > μ2 + δ нулевую гипотезу H_o: μ1 - μ2 = δ нет оснований отвергать.

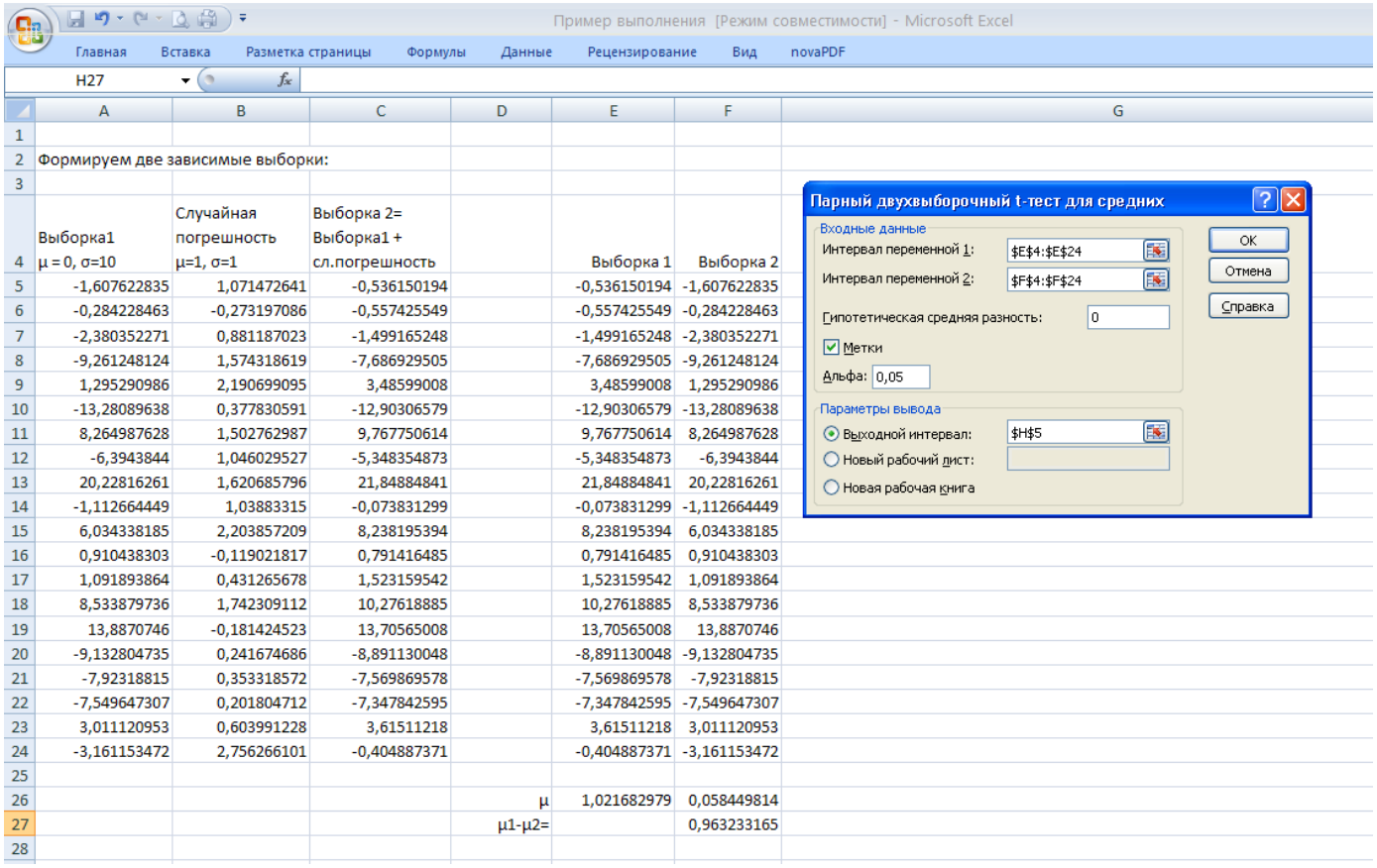
Статистическая функция ***ТТЕСТ*** при значении аргумента ***Тип = 3*** вычисляет вероятности ***P(T<=t) двухстороннее*** и ***P(T<=t) одностороннее.***

9. Парный двухвыборочный t-тест для средних

Это средство реализует критерий проверки гипотезы о равенстве (неравенстве) математических ожиданий распределений двух зависимых выборок, имеющих нормальные распределения. Этот критерий также называется *t*-тестом или тестом Стьюдента для парных наблюдений.

Рассмотрим выходные данные, вычисляемые этим средством, на примере проверки нулевой гипотезы H_o: $\mu_1 - \mu_2 = \delta$ (δ задано) против разных конкурирующих гипотез: H_1: $\mu_1 \neq \mu_2 + \delta$ или H_1: $\mu_1 > \mu_2 + \delta$, либо H_1: $\mu_1 < \mu_2 + \delta$ (μ_1 и μ_2— неизвестные математические ожидания выборок). Рассмотрим пример, когда выборки извлечены из нормально распределенных генеральных совокупностей с математическими ожиданиями соответственно 1 и 0.

Проверим гипотезу, что: $\mu_1 - \mu_2 = 0$ (на самом деле $\mu_1 - \mu_2 = 1$). Исходные данные и заполненное диалоговое окно ***Парный двухвыборочныи t-тест для средних*** показаны на рис. 27.



	A	B	C	D	E	F
1						
2	Формируем две зависимые выборки:					
3						
4	Выборка1 μ = 0, σ=10	Случайная погрешность μ=1, σ=1	Выборка 2= Выборка1 + сл.погрешность		Выборка 1	Выборка 2
5	-1,607622835	1,071472641	-0,536150194		-0,536150194	-1,607622835
6	-0,284228463	-0,273197086	-0,557425549		-0,557425549	-0,284228463
7	-2,380352271	0,881187023	-1,499165248		-1,499165248	-2,380352271
8	-9,261248124	1,574318619	-7,686929505		-7,686929505	-9,261248124
9	1,295290986	2,190699095	3,48599008		3,48599008	1,295290986
10	-13,28089638	0,377830591	-12,90306579		-12,90306579	-13,28089638
11	8,264987628	1,502762987	9,767750614		9,767750614	8,264987628
12	-6,3943844	1,046029527	-5,348354873		-5,348354873	-6,3943844
13	20,22816261	1,620685796	21,84884841		21,84884841	20,22816261
14	-1,112664449	1,03883315	-0,073831299		-0,073831299	-1,112664449
15	6,034338185	2,203857209	8,238195394		8,238195394	6,034338185
16	0,910438303	-0,119021817	0,791416485		0,791416485	0,910438303
17	1,091893864	0,431265678	1,523159542		1,523159542	1,091893864
18	8,533879736	1,742309112	10,27618885		10,27618885	8,533879736
19	13,8870746	-0,181424523	13,70565008		13,70565008	13,8870746
20	-9,132804735	0,241674686	-8,891130048		-8,891130048	-9,132804735
21	-7,92318815	0,353318572	-7,569869578		-7,569869578	-7,92318815
22	-7,549647307	0,201804712	-7,347842595		-7,347842595	-7,549647307
23	3,011120953	0,603991228	3,61511218		3,61511218	3,011120953
24	-3,161153472	2,756266101	-0,404887371		-0,404887371	-3,161153472
25						
26				μ	1,021682979	0,058449814
27				μ1-μ2=		0,963233165
28						

Рис. 27. Исходные данные и диалоговое окно ***Парный двухвыборочныи t-тест для средних***

Отметим, что средство требует, чтобы δ, значение которого задается в поле ***Гипотетическая средняя разность***, было неотрицательно. Поэтому первым (в поле ввода ***Интервал переменной 1***) задается адрес диапазона ячеек, содержащий выборку с большим математическим ожиданием, а затем в поле ***Интервал переменной 2*** указывается адрес второй выборки. (Диапазоны должны состоять из

одного столбца или одной строки.) В поле ***Альфа*** вводится значение уровня значимости α. Результат вычислений средства ***Парный двухвыборочныи t-тест для средних*** показан на рис. 28.

	Выборка 1	Выборка 2
	-0,536150194	-1,607622835
	-0,557425549	-0,284228463
	-1,499165248	-2,380352271
	-7,686929505	-9,261248124
	3,48599008	1,295290986
	-12,90306579	-13,28089638
	9,767750614	8,264987628
	-5,348354873	-6,3943844
	21,84884841	20,22816261
	-0,073831299	-1,112664449
	8,238195394	6,034338185
	0,791416485	0,910438303
	1,523159542	1,091893864
	10,27618885	8,533879736
	13,70565008	13,8870746
	-8,891130048	-9,132804735
	-7,569869578	-7,92318815
	-7,347842595	-7,549647307
	3,61511218	3,011120953
	-0,404887371	-3,161153472
μ	1,021682979	0,058449814
μ1-μ2=		0,963233165

Парный двухвыборочный t-тест для средних

	Выборка 1	*Выборка 2*
Среднее	1,021682979	0,058449814
Дисперсия	73,10150276	69,37172144
Наблюдения	20	20
Корреляция Пирсона	0,995083187	
Гипотетическая разность средних	0	
df	19	
t-статистика	4,977091384	
P(T<=t) одностороннее	4,18371E-05	
t критическое одностороннее	1,729132792	
P(T<=t) двухстороннее	8,36741E-05	
t критическое двухстороннее	2,09302405	

|t| =	4,977091384	|t| > $t_{кр2}$, следовательно гипотезу H_0 отвергаем
$t_{кр2}$ =	2,09302405	
t =	4,977091384	t > $t_{кр1}$, следовательно гипотезу H_0 при конкурирующей гипотизе H_1: μ1 > μ2 + δ отвергаем
$t_{кр1}$ =	1,729132792	
t =	4,977091384	t > -$t_{кр1}$, следовательно гипотезу H_0 при конкурирующей гипотезе H_1: μ1 < μ2 + δ не отвергаем
$t_{кр1}$ =	1,729132792	

Рис.28. Результат вычислений

В итоговой таблице приводятся следующие данные.

- ***Среднее*** — выборочные средние для каждой выборки.
- ***Дисперсия*** — несмещенные выборочные оценки дисперсий выборок.
- ***Наблюдения*** — объемы выборок.
- ***Корреляция Пирсона*** — выборочный коэффициент корреляции; вычисляется по формуле

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

- ***Гипотетическая разность средних*** — значение δ, которое задано в диалоговом окне.
- ***df*** — число степеней свободы, равное $(n - 1)$.
- ***t-статистика*** — значение критериальной статистики; вычисляется по формуле $t = \frac{\bar{d} - \delta}{S_n / \sqrt{n}}$, где

$\bar{d} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)$, $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i - \bar{d})^2$, и имеет распределение Стьюдента с ***df*** степенями свободы.

- ***P(T<=t) одностороннее*** — вероятность $P(X \le t)$, где X — случайная величина, имеющая распределение Стьюдента с *df* степенями свободы, t — подсчитанное значение критериальной статистики.

- ***t критическое одностороннее*** — значение квантиля $t_{кр1}$ порядка $(1 - \alpha)$ распределения Стьюдента с ***df*** степенями свободы.
- ***P(T<=t) двухстороннее*** — вероятность $P(|X| \le |t|)$, где X — случайная величина, имеющая распределение Стьюдента с *df* степенями свободы, t — подсчитанное значение критериальной статистики.
- ***t критическое двухстороннее*** — значение квантиля $t_{кр2}$ порядка $(1 - \alpha/2)$ распределения Стьюдента с *df* степенями свободы.

Нулевая гипотеза H_o: $\mu_1 - \mu_2 = \delta$ принимается, если $|t| < t_{кр2}$, (в противном случае отвергается); гипотеза H_o при конкурирующей гипотезе H_1: $\mu_1 > \mu_2 + \delta$ принимается, если $t < t_{кр1}$, при конкурирующей гипотезе H_1: $\mu_1 < \mu_2 + \delta$ нулевая гипотеза принимается при выполнении неравенства $t > -t_{кр1}$.

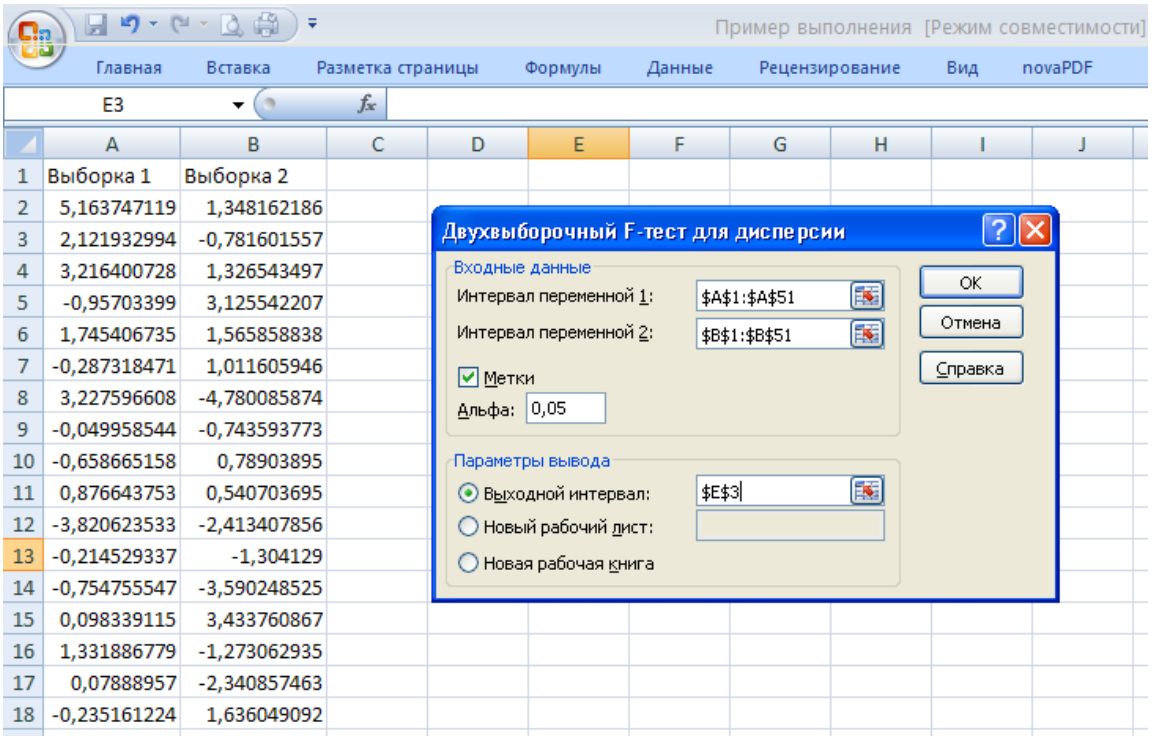
Как видно из результатов расчета, в данном примере нулевую гипотезу H_o: $\mu1 - \mu2 = \delta$ следует отвергнуть при конкурирующей гипотезе о неравенстве $\mu1 - \mu2 \neq \delta$ и конкурирующей гипотезе H_1: $\mu1 > \mu2 + \delta$. При конкурирующей гипотезе H_1: $\mu1 < \mu2 + \delta$ нулевую гипотезу H_o: $\mu1 - \mu2 = \delta$ нет оснований отвергать.

Статистическая функция ***TTECT*** при значении аргумента ***Тип = 1*** вычисляет вероятности ***P(T<=t) двухстороннее*** и ***P(T<=t) одностороннее***.

10. Двухвыборочный F-тест для дисперсий

Это средство реализует критерий Фишера проверки равенства дисперсий двух независимых выборок из нормально распределенных генеральных совокупностей с дисперсиями соответственно σ_1^2 и σ_2^2. Рассмотрим выходные данные, вычисляемые этим средством, на примере проверки нулевой гипотезы H_o: $\sigma_1^2 = \sigma_2^2$ против конкурирующей гипотезы H_1: $\sigma_1^2 \neq \sigma_2^2$.

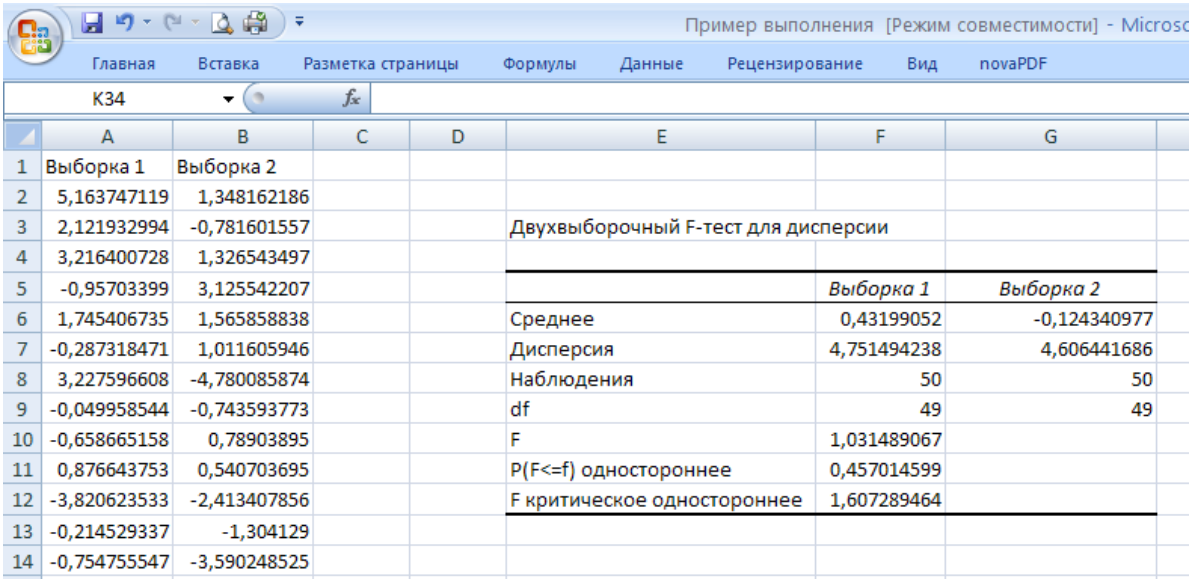
Рассмотрим пример, когда выборки извлечены из нормально распределенных генеральных совокупностей с равными дисперсиями 2. Исходные данные и заполненное диалоговое окно ***Двухвыборочный F-тест для дисперсий*** показаны на рис.29.



	A	B	C	D	E	F	G	H	I	J
1	Выборка 1	Выборка 2								
2	5,163747119	1,348162186								
3	2,121932994	-0,781601557								
4	3,216400728	1,326543497								
5	-0,95703399	3,125542207								
6	1,745406735	1,565858838								
7	-0,287318471	1,011605946								
8	3,227596608	-4,780085874								
9	-0,049958544	-0,743593773								
10	-0,658665158	0,78903895								
11	0,876643753	0,540703695								
12	-3,820623533	-2,413407856								
13	-0,214529337	-1,304129								
14	-0,754755547	-3,590248525								
15	0,098339115	3,433760867								
16	1,331886779	-1,273062935								
17	0,07888957	-2,340857463								
18	-0,235161224	1,636049092								

Рис. 29. Исходные данные и диалоговое окно ***Двухвыборочный F-тест для дисперсий***

Отметим, что первой (в поле ***Входной интервал 1***) должна задаваться выборка, имеющая большую дисперсию. В поле ***Альфа*** вводится значение уровня значимости α. Результат вычислений средства ***Двухвыборочный F-тест для дисперсий*** показан на рис. 30.



	A	B	C	D	E	F	G
1	Выборка 1	Выборка 2					
2	5,163747119	1,348162186					
3	2,121932994	-0,781601557			Двухвыборочный F-тест для дисперсии		
4	3,216400728	1,326543497					
5	-0,95703399	3,125542207				*Выборка 1*	*Выборка 2*
6	1,745406735	1,565858838			Среднее	0,43199052	-0,124340977
7	-0,287318471	1,011605946			Дисперсия	4,751494238	4,606441686
8	3,227596608	-4,780085874			Наблюдения	50	50
9	-0,049958544	-0,743593773			df	49	49
10	-0,658665158	0,78903895			F	1,031489067	
11	0,876643753	0,540703695			P(F<=f) одностороннее	0,457014599	
12	-3,820623533	-2,413407856			F критическое одностороннее	1,607289464	
13	-0,214529337	-1,304129					
14	-0,754755547	-3,590248525					

Рис.30. Результат вычислений

В итоговой таблице приводятся следующие данные.

- ***Среднее*** — выборочные средние для каждой выборки.
- ***Дисперсия*** — несмещенные выборочные оценки дисперсий выборок.
- ***Наблюдения*** — объемы выборок.

- ***df*** — числа степеней свободы, равные *(n - 1)* и (*m* – 1); *n* и *m* — объемы выборок.
- ***F*** — значение критериальной статистики, вычисляемой по формуле

$$F = \frac{S_x^2}{S_y^2}, \text{ где } S_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad S_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 ,$$

  и имеющей F-распределение со степенями свободы $k_1 = n - 1$ и $k_2 = m - 1$.
- ***P(F<=f) одностороннее*** — вероятность P(X ≤ F), где X — случайная величина, имеющая F-распределение с *df* степенями свободы, *F* — подсчитанное значение критериальной статистики.
- ***F критическое одностороннее*** — значение квантиля *t* порядка (1 – α) *F*-распределения с *df* степенями свободы.

Нулевая гипотеза H_o принимается, если $F < F_{кр1}$ (в противном случае отвергается). Как видно из результатов расчета, в данном примере нулевую гипотезу следует принять.

Статистическая функция ***ФТЕСТ*** вычисляет удвоенную вероятность ***P(F<=f) одностороннее***.

Задания ко второй части лабораторных работ:

1. Сгенерировать 2 выборки значений случайных чисел разной длины, имеющих нормальное распределение. С помощью средств **Пакета анализа** получить дисперсии и математические ожидания каждой выборки. Проверить с помощью средства ***Двухвыборочныи z-тест для средних*** нулевую гипотезу, что μ2 — μ1 = (порядковый номер студента по списку) для разных случаев конкурирующих гипотез при α=0,05.

2. Сгенерировать 2 выборки значений случайных чисел одинаковой длины, имеющих нормальное распределение, с одной и той же дисперсией. . С помощью средств **Пакета анализа** получить дисперсии и математические ожидания каждой выборки. Проверить с помощью средства ***Двухвыборочныи t-тест с одинаковыми дисперсиями*** нулевую гипотезу, что μ1 - μ2 = (порядковый номер студента по списку) для разных случаев конкурирующих гипотез при α=0,05.

3. Сгенерировать 2 выборки значений случайных чисел одинаковой длины, имеющих нормальное распределение и разные дисперсии. . С помощью средств **Пакета анализа** получить дисперсии и математические ожидания каждой выборки. Проверить с помощью средства ***Двухвыборочный t-тест с различными дисперсиями*** нулевую гипотезу, что μ1 - μ2 = (порядковый номер студента по списку) для разных случаев конкурирующих гипотез при α=0,05.

4. Сгенерировать 2 выборки значений случайных чисел одинаковой длины, имеющих нормальное распределение и разные дисперсии. С помощью средств **Пакета анализа** получить дисперсии и математические ожидания каждой выборки. Проверить с помощью средства ***Парный двухвыборочныи t-тест для средних*** нулевую гипотезу, что μ1 - μ2 = (порядковый номер студента по списку) для разных случаев конкурирующих гипотез при α=0,05.

5. Сгенерировать 2 выборки значений случайных чисел одинаковой длины, имеющих нормальное распределение и разные дисперсии. С помощью средств **Пакета анализа** получить дисперсии и математические ожидания каждой выборки. Проверить с помощью средства ***Двухвыборочный F-тест для дисперсий*** нулевую гипотезу, что μ1 - μ2 = (порядковый номер студента по списку) для разных случаев конкурирующих гипотез при α=0,05.

6. Сгенерировать 2 выборки значений случайных чисел разной длины, имеющих нормальное распределение и разные дисперсии. С помощью средств **Пакета анализа** получить дисперсии и математические ожидания каждой выборки. Проверить с помощью средства ***Двухвыборочный F-тест для дисперсий*** нулевую гипотезу, что μ1 - μ2 = (порядковый номер студента по списку) для разных случаев конкурирующих гипотез при α=0,05.

Сравнить результаты из п/п 3.5 и 3.6.

Контрольные вопросы

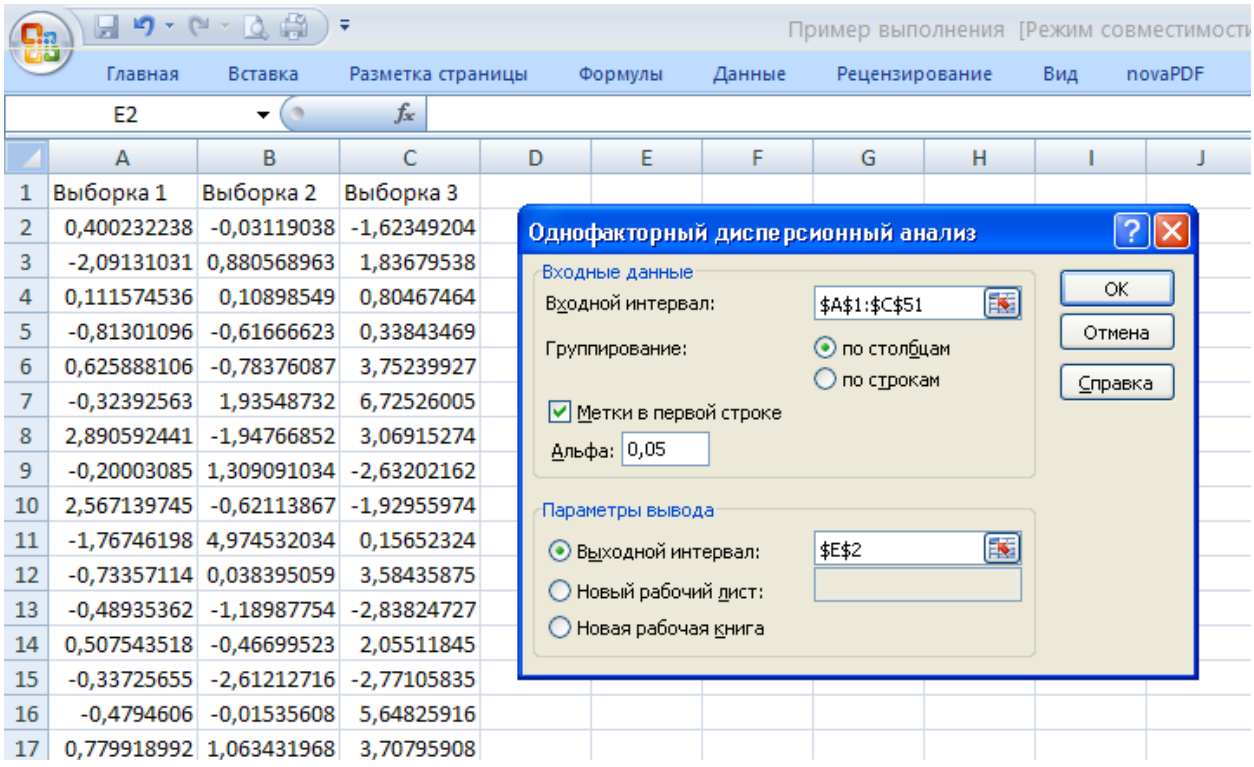
1. Что такое альтернативная или конкурирующая гипотеза?
2. Что называется статистическим критерием?
3. Дайте определение ошибок первого и второго рода.

Часть 3. Определение статистических оценок многомерных случайных процессов

11. Однофакторный дисперсионный анализ

Это средство реализует критерий проверки гипотезы о равенстве математических ожиданий нескольких независимых выборок, построенный на основе дисперсионного анализа. Здесь покажем применение средства ***Однофакторный дисперсионный анализ*** и опишем его выходные данные.

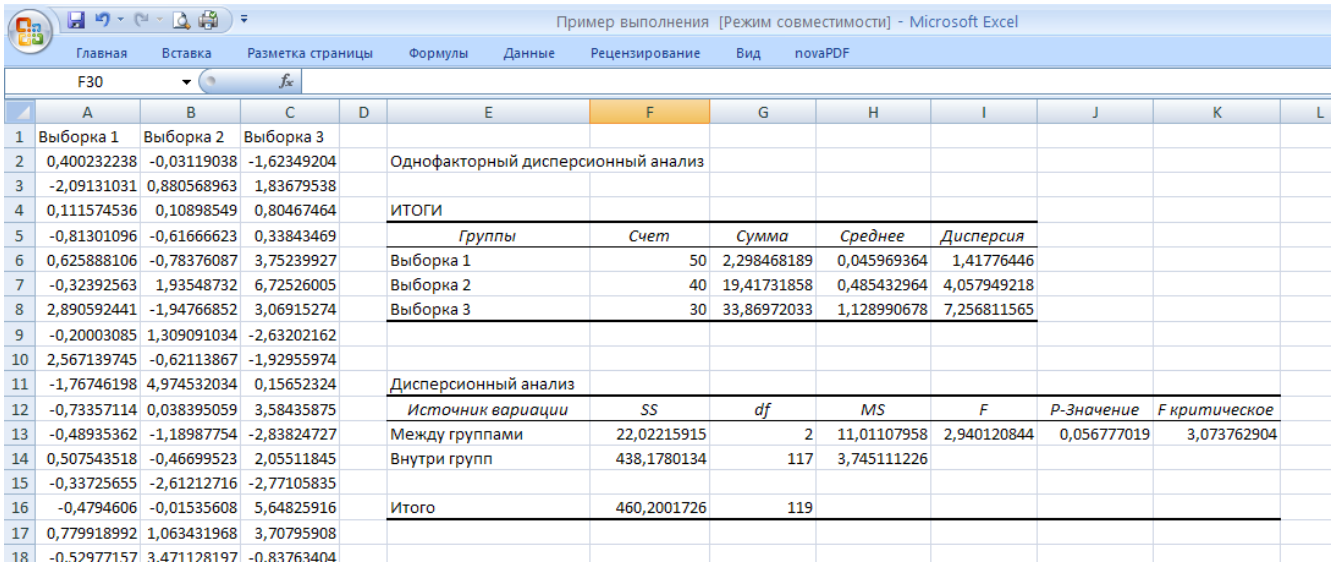
На рис. 31 показаны три выборки, имеющие нормальное распределение с математическими ожиданиями 0, 0,5 и 1 и среднеквадратическими отклонениями 1, 2 и 3 соответственно. Объемы выборок — 50, 40 и 30 значений. (Выборки сгенерированы с помощью средства ***Генерация случайных чисел***.) На рис. 31 также показано заполненное диалоговое окно ***Однофакторный дисперсионный анализ***. Обращаем внимание, что все три выборки задаются в виде одного диапазона ячеек. В случае, когда выборки имеют разные размеры, диапазон задается в соответствии с наибольшей выборкой и неизбежно содержит пустые ячейки. Но средство правильно определяет объемы выборок. Также отметим, что в данном случае результаты анализа будут выводиться на отдельный рабочий лист с именем ***Результаты***, который автоматически вставится в текущую рабочую книгу.



	A	B	C
1	Выборка 1	Выборка 2	Выборка 3
2	0,400232238	-0,03119038	-1,62349204
3	-2,09131031	0,880568963	1,83679538
4	0,111574536	0,10898549	0,80467464
5	-0,81301096	-0,61666623	0,33843469
6	0,625888106	-0,78376087	3,75239927
7	-0,32392563	1,93548732	6,72526005
8	2,890592441	-1,94766852	3,06915274
9	-0,20003085	1,309091034	-2,63202162
10	2,567139745	-0,62113867	-1,92955974
11	-1,76746198	4,974532034	0,15652324
12	-0,73357114	0,038395059	3,58435875
13	-0,48935362	-1,18987754	-2,83824727
14	0,507543518	-0,46699523	2,05511845
15	-0,33725655	-2,61212716	-2,77105835
16	-0,4794606	-0,01535608	5,64825916
17	0,779918992	1,063431968	3,70795908

Рис. 31. Исходные данные и диалоговое окно ***Однофакторныи дисперсионный анализ***

На рис. 32 показаны результаты, выводимые средством ***Однофакторныи дисперсионный анализ***. Они представлены в виде двух таблиц, озаглавленных ***ИТОГИ*** и ***Дисперсионный анализ***. В таблице ***ИТОГИ*** выводятся основные статистические характеристики выборок: в столбце ***Счет*** — объемы выборок, в столбце ***Сумма*** — суммы выборочных значений, в столбцах ***Среднее*** и ***Дисперсия*** — соответственно выборочные средние и дисперсии.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Выборка 1	Выборка 2	Выборка 3									
2	0,400232238	-0,03119038	-1,62349204		Однофакторный дисперсионный анализ							
3	-2,09131031	0,880568963	1,83679538									
4	0,111574536	0,10898549	0,80467464		ИТОГИ							
5	-0,81301096	-0,61666623	0,33843469		*Группы*	*Счет*	*Сумма*	*Среднее*	*Дисперсия*			
6	0,625888106	-0,78376087	3,75239927		Выборка 1	50	2,298468189	0,045969364	1,41776446			
7	-0,32392563	1,93548732	6,72526005		Выборка 2	40	19,41731858	0,485432964	4,057949218			
8	2,890592441	-1,94766852	3,06915274		Выборка 3	30	33,86972033	1,128990678	7,256811565			
9	-0,20003085	1,309091034	-2,63202162									
10	2,567139745	-0,62113867	-1,92955974									
11	-1,76746198	4,974532034	0,15652324		Дисперсионный анализ							
12	-0,73357114	0,038395059	3,58435875		*Источник вариации*	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*	
13	-0,48935362	-1,18987754	-2,83824727		Между группами	22,02215915	2	11,01107958	2,940120844	0,056777019	3,073762904	
14	0,507543518	-0,46699523	2,05511845		Внутри групп	438,1780134	117	3,745111226				
15	-0,33725655	-2,61212716	-2,77105835									
16	-0,4794606	-0,01535608	5,64825916		Итого	460,2001726	119					
17	0,779918992	1,063431968	3,70795908									
18	-0,52977157	3,471128197	-0,83763404									

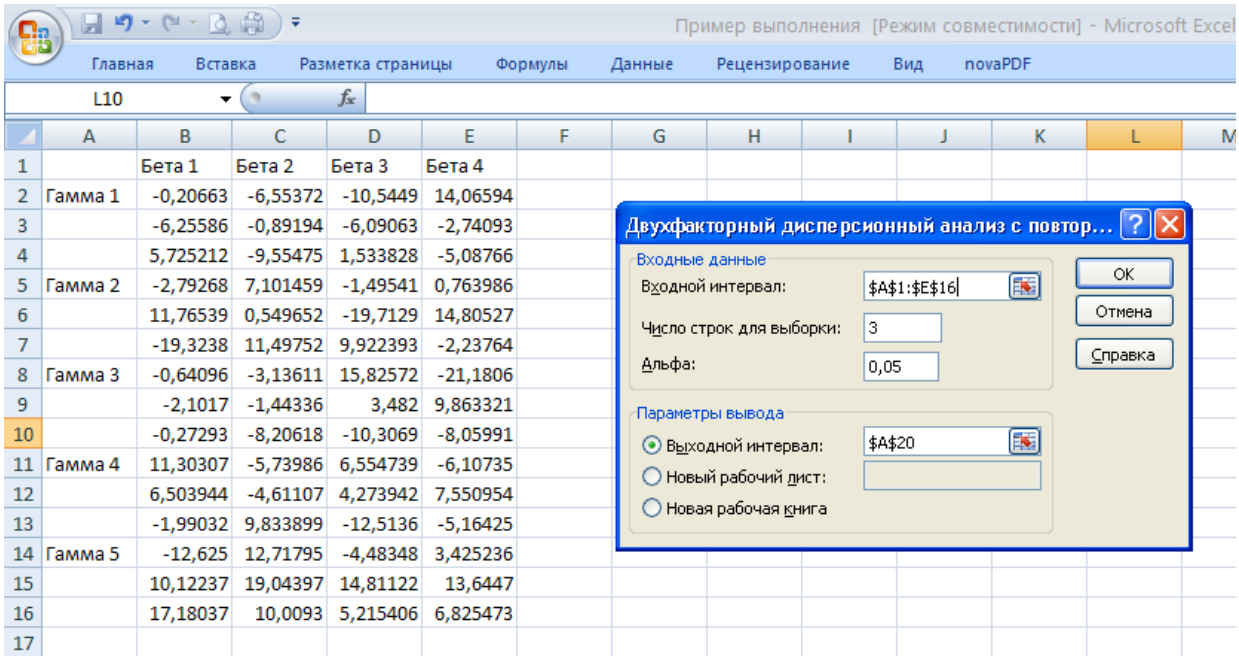
Рис. 32. Результат вычислений

Значения в первых четырех столбцах таблицы ***Дисперсионный анализ*** повторяют значения из дисперсионной таблицы. В столбце ***SS*** приведены суммы квадратов (межгрупповая, внутригрупповая и полная); в столбце ***df*** —значения степеней свободы, а в столбце ***MS*** — дисперсии, межгрупповая и внутригрупповая. В столбце ***F*** записано значение критериальной статистики, в столбце ***P-Значение*** — значение вероятности $P(X \geq x)$, где X — случайная величина, имеющая ***F***-распределение с ***df*** степенями свободы. В столбце ***F критическое*** приводится критическое значение ***t***, рассчитанное в соответствии с заданным уровнем значимости (параметр ***Альфа***).

Нулевая гипотеза о равенстве математических ожиданий всех выборок принимается, если выполняется неравенство $F \leq F_{критическое}$. В нашем примере эту гипотезу отвергать не следует.

12. Двухфакторный дисперсионный анализ с повторениями

Здесь рассмотрим структуру входных данных для работы с этим средством и опишем выходные результаты. Структура входных данных представлена на рис. 33: в строке 1 показаны обозначения уровней фактора β; в столбце A — обозначения уровней фактора γ; в данном случае имеется три выборки, поэтому под общим обозначением уровней фактора γ записаны три строки числовых данных. Таким образом, в диапазоне, например, C8:C10 содержатся выборочные значения, соответствующие второму уровню фактора β и третьему уровню фактора γ.

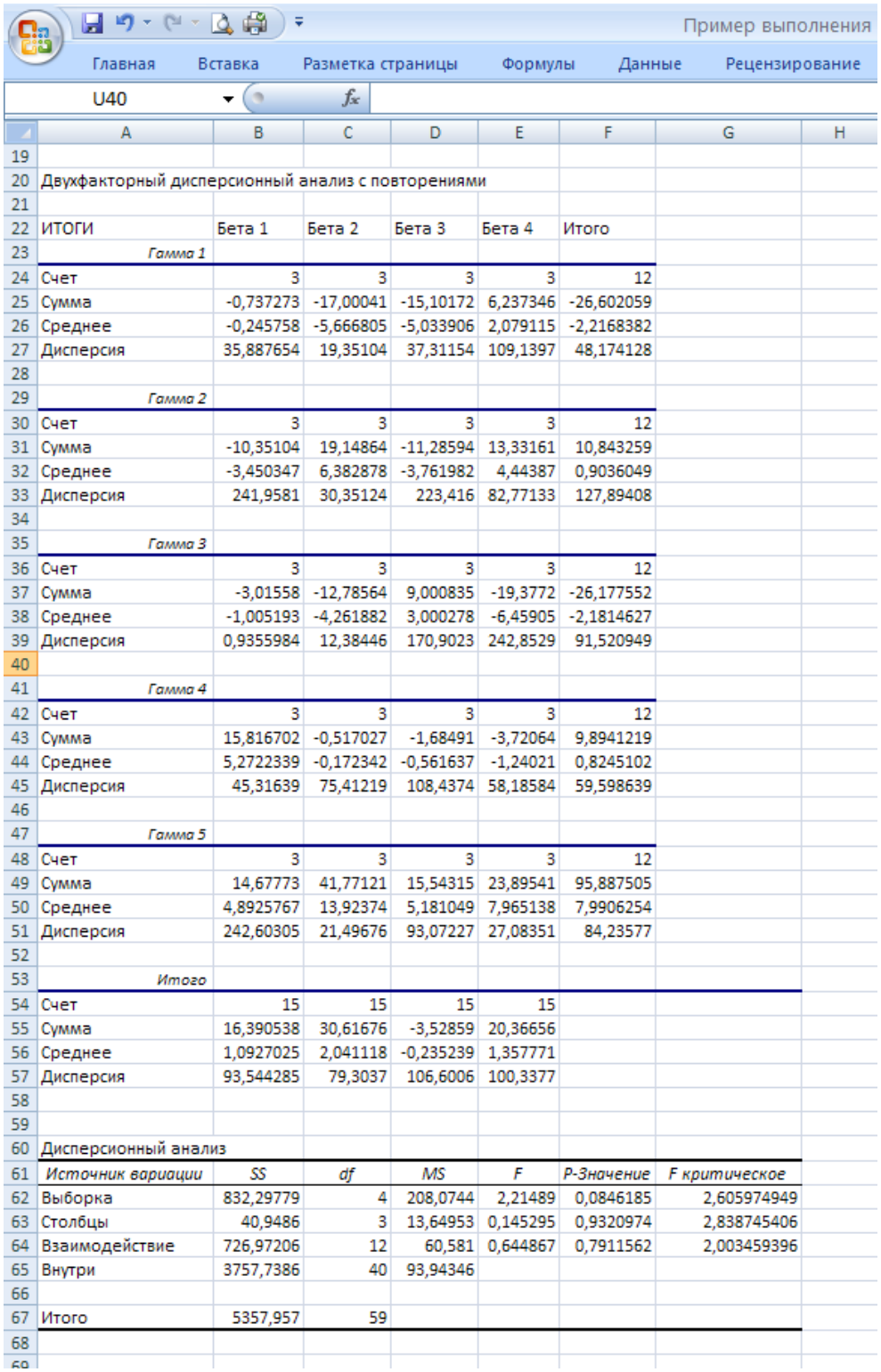


	A	B	C	D	E
1		Бета 1	Бета 2	Бета 3	Бета 4
2	Гамма 1	-0,20663	-6,55372	-10,5449	14,06594
3		-6,25586	-0,89194	-6,09063	-2,74093
4		5,725212	-9,55475	1,533828	-5,08766
5	Гамма 2	-2,79268	7,101459	-1,49541	0,763986
6		11,76539	0,549652	-19,7129	14,80527
7		-19,3238	11,49752	9,922393	-2,23764
8	Гамма 3	-0,64096	-3,13611	15,82572	-21,1806
9		-2,1017	-1,44336	3,482	9,863321
10		-0,27293	-8,20618	-10,3069	-8,05991
11	Гамма 4	11,30307	-5,73986	6,554739	-6,10735
12		6,503944	-4,61107	4,273942	7,550954
13		-1,99032	9,833899	-12,5136	-5,16425
14	Гамма 5	-12,625	12,71795	-4,48348	3,425236
15		10,12237	19,04397	14,81122	13,6447
16		17,18037	10,0093	5,215406	6,825473
17					

Рис. 33. Исходные данные и диалоговое окно ***Двухфакторный дисперсионный анализ с повторениями***

Диалоговое окно рассматриваемого здесь средства показано на рис.33. В поле ***Входной интервал*** указывается диапазон ячеек, содержащий входные данные, включая заголовки. В поле ***Число строк для выборки*** указывается количество рассматриваемых выборок, в данном случае введено число 3. В поле ***Альфа***, как обычно, указывается значение уровня значимости.

На рис.34 показаны выходные результаты работы данного средства, выведенные на отдельный рабочий лист. Выходные результаты сгруппированы в несколько таблиц. В первой таблице, озаглавленной ***ИТОГИ*** и состоящей из нескольких подтаблиц (по количеству уровней фактора γ), приводятся статистические характеристики выборочных значений, соответствующих каждому сочетанию уровней фактора β и фактора γ: количество выборочных значений (строка ***Счет***), сумма выборочных значений (строка ***Сумма***), выборочное среднее (строка ***Среднее***) и выборочная дисперсия (строка ***Дисперсия***). В столбце ***Итого*** подтаблиц выводятся такие же статистические характеристики выборочных значений, соответствующие одному уровню фактора γ: количество выборочных значений, выборочное среднее и выборочная дисперсия (вычисляется по всем значениям данного уровня относительно общего среднего). В конце таблицы ***ИТОГИ*** выводится подтаблица ***Итого***, в которой приведены те же характеристики, но подсчитанные по выборочным значениям для каждого уровня фактора β.



	A	B	C	D	E	F	G	H
19								
20	Двухфакторный дисперсионный анализ с повторениями							
21								
22	ИТОГИ	Бета 1	Бета 2	Бета 3	Бета 4	Итого		
23	*Гамма 1*							
24	Счет	3	3	3	3	12		
25	Сумма	-0,737273	-17,00041	-15,10172	6,237346	-26,602059		
26	Среднее	-0,245758	-5,666805	-5,033906	2,079115	-2,2168382		
27	Дисперсия	35,887654	19,35104	37,31154	109,1397	48,174128		
28								
29	*Гамма 2*							
30	Счет	3	3	3	3	12		
31	Сумма	-10,35104	19,14864	-11,28594	13,33161	10,843259		
32	Среднее	-3,450347	6,382878	-3,761982	4,44387	0,9036049		
33	Дисперсия	241,9581	30,35124	223,416	82,77133	127,89408		
34								
35	*Гамма 3*							
36	Счет	3	3	3	3	12		
37	Сумма	-3,01558	-12,78564	9,000835	-19,3772	-26,177552		
38	Среднее	-1,005193	-4,261882	3,000278	-6,45905	-2,1814627		
39	Дисперсия	0,9355984	12,38446	170,9023	242,8529	91,520949		
40								
41	*Гамма 4*							
42	Счет	3	3	3	3	12		
43	Сумма	15,816702	-0,517027	-1,68491	-3,72064	9,8941219		
44	Среднее	5,2722339	-0,172342	-0,561637	-1,24021	0,8245102		
45	Дисперсия	45,31639	75,41219	108,4374	58,18584	59,598639		
46								
47	*Гамма 5*							
48	Счет	3	3	3	3	12		
49	Сумма	14,67773	41,77121	15,54315	23,89541	95,887505		
50	Среднее	4,8925767	13,92374	5,181049	7,965138	7,9906254		
51	Дисперсия	242,60305	21,49676	93,07227	27,08351	84,23577		
52								
53	*Итого*							
54	Счет	15	15	15	15			
55	Сумма	16,390538	30,61676	-3,52859	20,36656			
56	Среднее	1,0927025	2,041118	-0,235239	1,357771			
57	Дисперсия	93,544285	79,3037	106,6006	100,3377			
58								
59								
60	Дисперсионный анализ							
61	*Источник вариации*	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*	
62	Выборка	832,29779	4	208,0744	2,21489	0,0846185	2,605974949	
63	Столбцы	40,9486	3	13,64953	0,145295	0,9320974	2,838745406	
64	Взаимодействие	726,97206	12	60,581	0,644867	0,7911562	2,003459396	
65	Внутри	3757,7386	40	93,94346				
66								
67	Итого	5357,957	59					
68								

Рис.34. Выходные результаты, работы, средства ***Двухфакторный дисперсионный анализ с повторениями***

В нижней части выходных результатов приведена дисперсионная таблица. Здесь в первом столбце, обозначенном ***SS***, выведены суммы квадратов: соответственно SS_1, SS_2, SS_3, SS_4 и в строке ***Итого — SS***. В столбце ***df*** приведены степени свободы сумм квадратов, а в столбце ***MS*** — значения соответствующих дисперсий. В столбце ***F*** вычислены значения критериальных статистик, т.е. отношения дисперсий s_1^2, s_2^2, s_3^2 к дисперсии s_4^2.

В столбце ***P-Значение*** вычисляются вероятности P(X ≥ *F),* где X — случайная величина, имеющая *F*-распределение со степенями свободы, значения которых приведены в столбце ***df*** : первое значение степени свободы — из соответствующей строки этого столбца, а второе — всегда из четвертой строки, ***F*** — значение из столбца **F.** Например, значение в ячейке F43, можно вычислить по формуле Excel =***FРАСП(E43;C43;C46)***. Эти значения используются для проверки гипотез о значимом влиянии факторов или их взаимного влияния: если вероятность больше заданного уровня значимости, то нулевая гипотеза об отсутствии влияния принимается, в противном случае — отвергается.

В столбце ***F критическое*** вычисляются критические значения, соответствующие заданному в диалоговом окне ***Двухфакторный дисперсионный анализ с повторениями*** уровню значимости α. Эти значения вычисляются как квантили порядка (1 – α) F-распределения со степенями свободы, значения которых определяются так же, как при вычислении вероятностей из столбца ***P-Значение***. Например, значение в ячейке G43 (см. рис. 34) можно вычислить по формуле Excel =***FРАСПОБР(0,05;C43;C46)***. Эти значения используются для проверки гипотез о значимом влиянии факторов или их взаимного влияния: если значение в этом столбце больше значения в столбце ***F*** той же строки, то нулевая гипотеза об отсутствии влияния принимается, в противном случае — отвергается. Здесь принимаются все три нулевые гипотезы об отсутствии влияния факторов β и γ и их взаимного влияния. Однако значение в столбце ***F*** третьей строки (соответствует взаимному влиянию факторов) значительно больше аналогичных значений для отдельных факторов, и на это необходимо обратить внимание.

13. Двухфакторный дисперсионный анализ без повторений

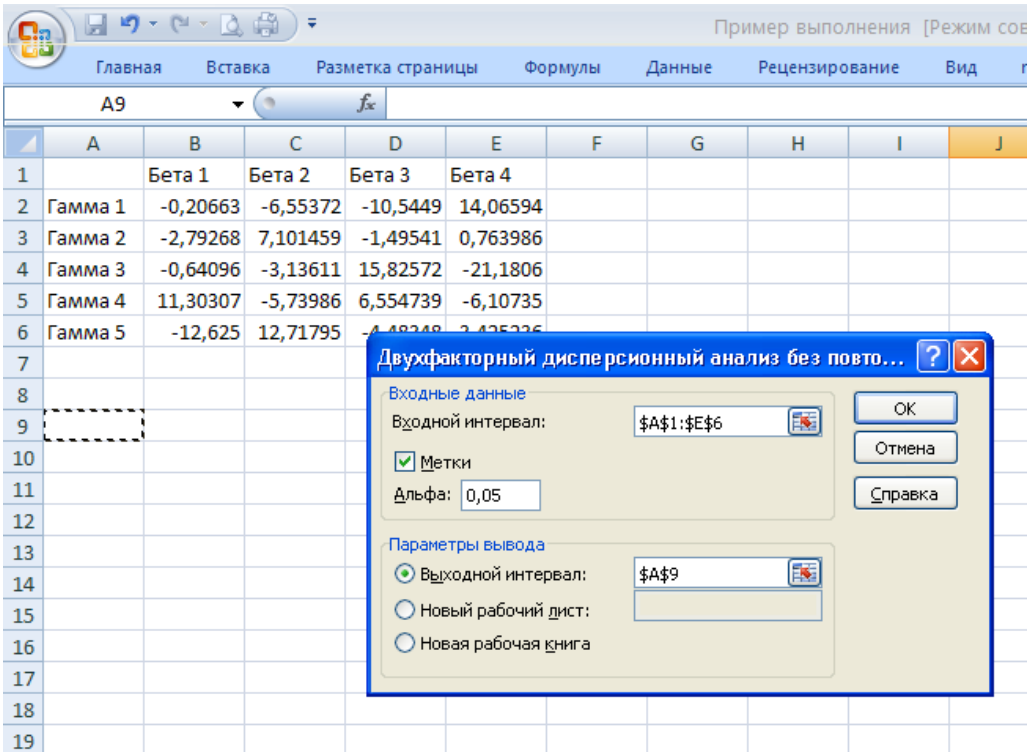
Структура входных данных показана на рис. 35: в строке 1 приводятся обозначения уровней фактора β; в столбце A — обозначения уровней фактора γ; в диапазоне, обозначенном этими заголовками, введены числовые данные.

Диалоговое окно этого средства показано на рис. 35. В поле ***Входной интервал*** указывается диапазон ячеек, содержащий входные данные; если в этот диапазон включены заголовки строк и столбцов, то следует установить флажок опции ***Метки***. В поле ***Альфа*** указывается значение уровня значимости.

На рис.36 представлены выходные результаты работы данного средства, выведенные на отдельный рабочий лист. Выходные результаты сгруппированы в две таблицы. В первой таблице, озаглавленной ***ИТОГИ***, приводятся статистические характеристики выборочных значений, соответствующих каждому уровню фактора β (группировка по столбцам) и каждому уровню фактора γ (группировка по строка): количество выборочных значений (столбец ***Счет***), сумма выборочных значений (столбец ***Сумма***), выборочное среднее (столбец ***Среднее***) и выборочная дисперсия (столбец ***Дисперсия***).

В нижней части выходных результатов приведена дисперсионная таблица. Здесь в первом столбце, обозначенном ***SS***, выведены суммы квадратов: соответственно SS_1, SS_2, SS_3 и в строке ***Итого — SS***. В столбце ***df*** приведены степени свободы сумм квадратов, а в столбце ***MS*** — значения

соответствующих дисперсий. В столбце ***F*** вычислены значения критериальных статистик, т.е. отношения дисперсий s_1^2 и s_2^2 к дисперсии s_3^2 .



	A	B	C	D	E
1		Бета 1	Бета 2	Бета 3	Бета 4
2	Гамма 1	-0,20663	-6,55372	-10,5449	14,06594
3	Гамма 2	-2,79268	7,101459	-1,49541	0,763986
4	Гамма 3	-0,64096	-3,13611	15,82572	-21,1806
5	Гамма 4	11,30307	-5,73986	6,554739	-6,10735
6	Гамма 5	-12,625	12,71795		

Рис.35. Исходные данные и диалоговое окно ***Двухфакторныи дисперсионный анализ без повторений***

Двухфакторный дисперсионный анализ без повторений

ИТОГИ	*Счет*	*Сумма*	*Среднее*	*Дисперсия*
Гамма 1	4	-3,23933	-0,80983	116,4722
Гамма 2	4	3,577346	0,894337	19,28346
Гамма 3	4	-9,13192	-2,28298	229,4894
Гамма 4	4	6,010609	1,502652	77,31263
Гамма 5	4	-0,9653	-0,24132	117,5793
Бета 1	5	-4,9622	-0,99244	72,61953
Бета 2	5	4,389729	0,877946	73,51386
Бета 3	5	5,856642	1,171328	105,0228
Бета 4	5	-9,03276	-1,80655	169,9412

Дисперсионный анализ

Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Строки	35,23287	4	8,808216	0,064093	0,991432679	3,259166727
Столбцы	31,25424	3	10,41808	0,075807	0,971862057	3,490294821
Погрешность	1649,157	12	137,4297			
Итого	1715,644	19				

Рис.36. Выходные результаты, работы, средства ***Двухфакторныи дисперсионный анализ без повторений***

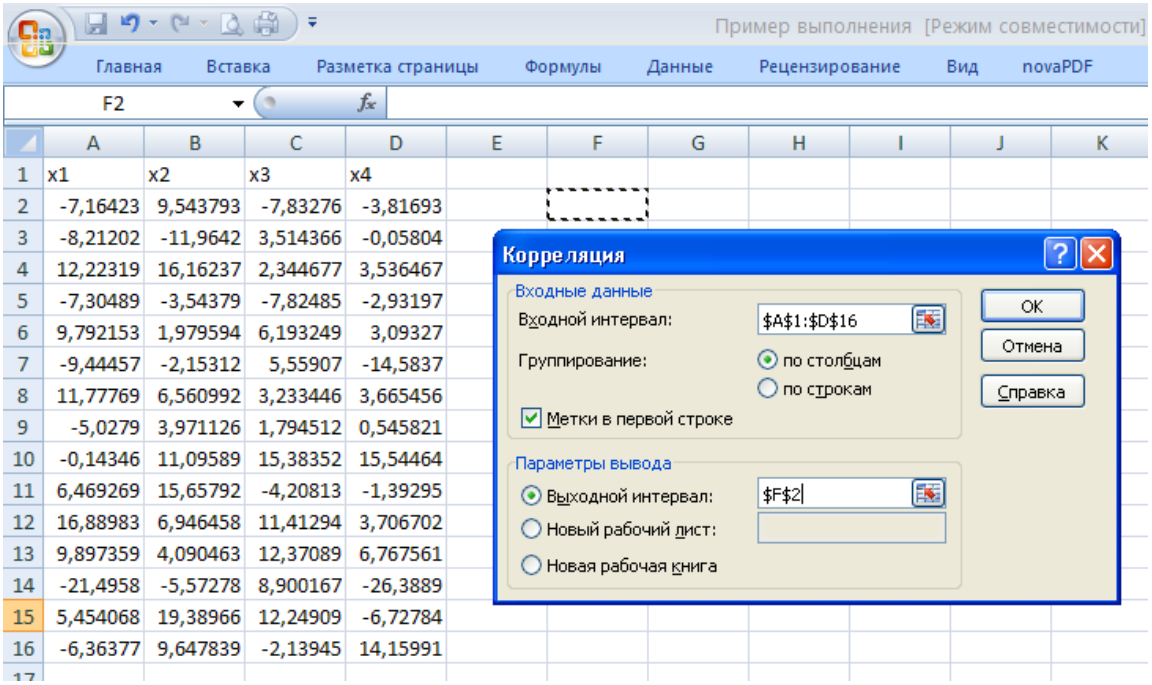
В столбце ***P-Значение*** вычисляются вероятности $P(X \geq F)$, где X — случайная величина, имеющая *F*-распределение со степенями свободы, значения которых приведены в столбце **df**: первое значение степени свободы — из соответствующей строки этого столбца, а второе — всегда из третьей строки, ***F*** — значение из столбца **F**. Например, значение в ячейке E18 можно вычислить по формуле

Excel =***FРАСП(D18;C18;C20)***. Эти значения используются для проверки гипотез о значимом влиянии факторов: если вероятность больше заданного уровня значимости, то нулевая гипотеза об отсутствии влияния принимается, в противном случае — отвергается.

В столбце ***F критическое*** вычисляются критические значения, соответствующие заданному в диалоговом окне ***Двухфакторный дисперсионный анализ без повторений*** уровню значимости α. Эти значения вычисляются как квантили порядка (1 – α) *F*-распределения со степенями свободы, значения которых определяются так же, как при вычислении вероятностей из столбца ***P-Значение***. Например, значение в ячейке G18 (см. рис. 36) можно вычислить по формуле Excel = ***FРАСПОБР(0,05;C18;C20)***. Эти значения используются для проверки гипотез о значимом влиянии факторов или их взаимного влияния: если значение в этом столбце больше значения в столбце ***F*** той же строки, то нулевая гипотеза об отсутствии влияния принимается, в противном случае — отвергается. Здесь принимаются обе нулевые гипотезы об отсутствии влияния факторов β и γ.

14. Корреляция

Это средство вычисляет корреляционную матрицу компонентов многомерной выборки. Диагональные элементы матрицы равны единице, а внедиагональные — коэффициентам корреляции соответствующих компонентов. На рис.37 показаны многомерная выборка, имеющая совместное нормальное распределение, причем первая пара компонентов зависима с коэффициентом корреляции 0,5. С таким же коэффициентом корреляции зависимы третий и четвертый компоненты выборки. Первая и вторая пара компонентов между собой независимы.



	A	B	C	D
1	x1	x2	x3	x4
2	-7,16423	9,543793	-7,83276	-3,81693
3	-8,21202	-11,9642	3,514366	-0,05804
4	12,22319	16,16237	2,344677	3,536467
5	-7,30489	-3,54379	-7,82485	-2,93197
6	9,792153	1,979594	6,193249	3,09327
7	-9,44457	-2,15312	5,55907	-14,5837
8	11,77769	6,560992	3,233446	3,665456
9	-5,0279	3,971126	1,794512	0,545821
10	-0,14346	11,09589	15,38352	15,54464
11	6,469269	15,65792	-4,20813	-1,39295
12	16,88983	6,946458	11,41294	3,706702
13	9,897359	4,090463	12,37089	6,767561
14	-21,4958	-5,57278	8,900167	-26,3889
15	5,454068	19,38966	12,24909	-6,72784
16	-6,36377	9,647839	-2,13945	14,15991

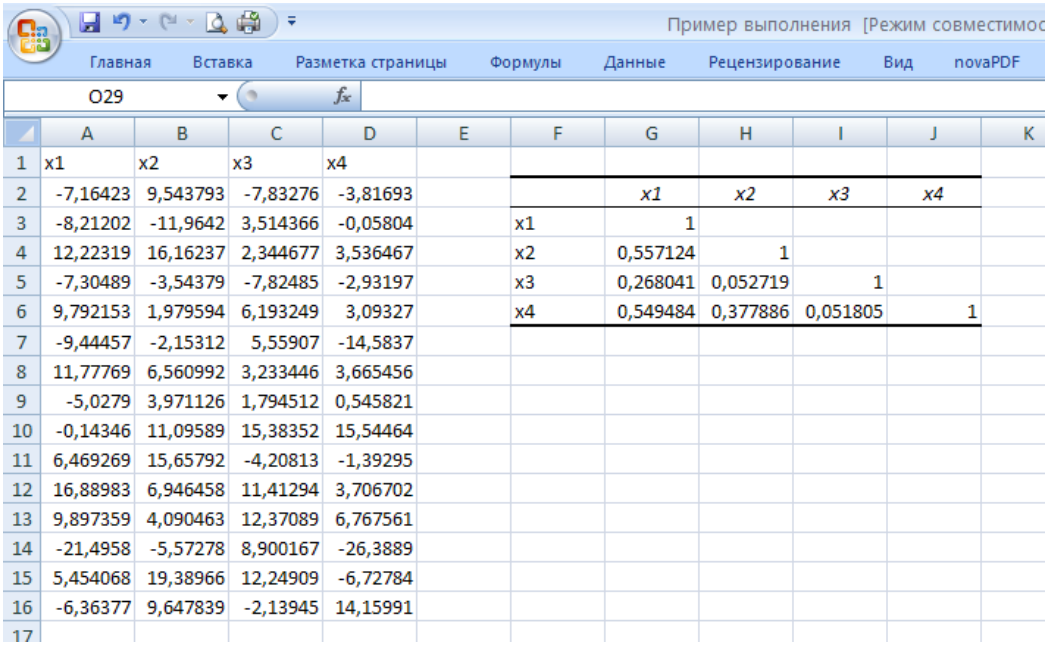
Рис.37. Исходные данные и диалоговое окно ***Корреляция***

Внедиагональные элементы корреляционной матрицы рассчитываются по стандартным формулам: коэффициент корреляции между компонентами *x* и *y* многомерной выборки вычисляется как

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$, где $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, n — объем выборки.

Отметим, что эти же вычисления выполняет функция КОРРЕЛ.

На рис. 38 показан результат применения средства ***Корреляция***. Поскольку корреляционная матрица симметрична, выводится только нижняя ее половина.



	A	B	C	D	E	F	G	H	I	J	K
1	x1	x2	x3	x4							
2	-7,16423	9,543793	-7,83276	-3,81693			*x1*	*x2*	*x3*	*x4*	
3	-8,21202	-11,9642	3,514366	-0,05804		x1	1				
4	12,22319	16,16237	2,344677	3,536467		x2	0,557124	1			
5	-7,30489	-3,54379	-7,82485	-2,93197		x3	0,268041	0,052719	1		
6	9,792153	1,979594	6,193249	3,09327		x4	0,549484	0,377886	0,051805	1	
7	-9,44457	-2,15312	5,55907	-14,5837							
8	11,77769	6,560992	3,233446	3,665456							
9	-5,0279	3,971126	1,794512	0,545821							
10	-0,14346	11,09589	15,38352	15,54464							
11	6,469269	15,65792	-4,20813	-1,39295							
12	16,88983	6,946458	11,41294	3,706702							
13	9,897359	4,090463	12,37089	6,767561							
14	-21,4958	-5,57278	8,900167	-26,3889							
15	5,454068	19,38966	12,24909	-6,72784							
16	-6,36377	9,647839	-2,13945	14,15991							
17											

Рис. 38. *Результат применения средства* ***Корреляция***

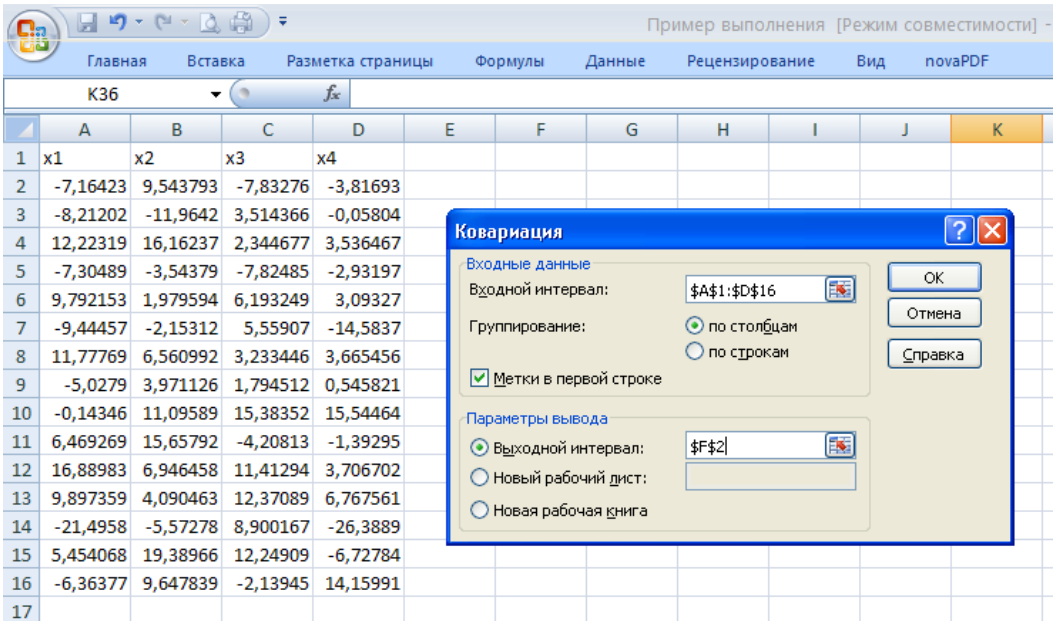
15. Ковариация

Это средство вычисляет ковариационную матрицу компонентов многомерной выборки. Диагональные элементы матрицы равны выборочным дисперсиям, а внедиагональные — ковариациям соответствующих компонентов. На рис.39 показана многомерная выборка, имеющая совместное нормальное распределение, причем первая пара компонентов зависима с коэффициентом корреляции 0,5. С таким же коэффициентом корреляции зависимы третий и четвертый компоненты выборки. Первая и вторая пары компонентов между собой независимы.

Внедиагональные элементы ковариационной матрицы рассчитываются по формулам: ковариация cov(X, Y) между компонентами x и y многомерной выборки вычисляется как

$$\mathrm{cov}(X,Y) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}), \text{ где } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \ n \text{ — объем выборки.}$$

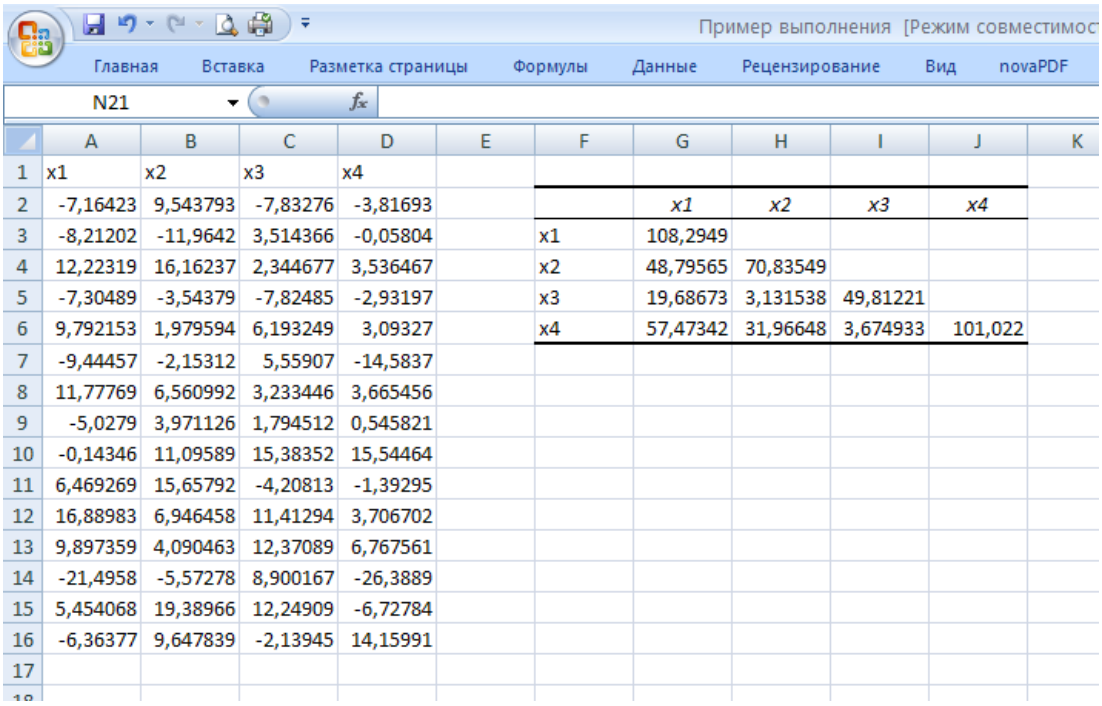
Отметим, что эти же вычисления выполняет функция ***КОВАР***. Диагональные элементы матрицы — выборочные дисперсии — вычисляются по стандартным формулам $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Выборочную дисперсию также вычисляют функции ***ДИСПР*** и ***ДИСПРА***.



	A	B	C	D
1	x1	x2	x3	x4
2	-7,16423	9,543793	-7,83276	-3,81693
3	-8,21202	-11,9642	3,514366	-0,05804
4	12,22319	16,16237	2,344677	3,536467
5	-7,30489	-3,54379	-7,82485	-2,93197
6	9,792153	1,979594	6,193249	3,09327
7	-9,44457	-2,15312	5,55907	-14,5837
8	11,77769	6,560992	3,233446	3,665456
9	-5,0279	3,971126	1,794512	0,545821
10	-0,14346	11,09589	15,38352	15,54464
11	6,469269	15,65792	-4,20813	-1,39295
12	16,88983	6,946458	11,41294	3,706702
13	9,897359	4,090463	12,37089	6,767561
14	-21,4958	-5,57278	8,900167	-26,3889
15	5,454068	19,38966	12,24909	-6,72784
16	-6,36377	9,647839	-2,13945	14,15991
17				

Рис. 39. *Исходные данные и диалоговое окно* ***Ковариация***

На рис.40 показан результат применения средства ***Ковариация***. Поскольку ковариационная матрица симметрична, выводится только нижняя ее половина.



	A	B	C	D	E	F	G	H	I	J	K
1	x1	x2	x3	x4							
2	-7,16423	9,543793	-7,83276	-3,81693			*x1*	*x2*	*x3*	*x4*	
3	-8,21202	-11,9642	3,514366	-0,05804		x1	108,2949				
4	12,22319	16,16237	2,344677	3,536467		x2	48,79565	70,83549			
5	-7,30489	-3,54379	-7,82485	-2,93197		x3	19,68673	3,131538	49,81221		
6	9,792153	1,979594	6,193249	3,09327		x4	57,47342	31,96648	3,674933	101,022	
7	-9,44457	-2,15312	5,55907	-14,5837							
8	11,77769	6,560992	3,233446	3,665456							
9	-5,0279	3,971126	1,794512	0,545821							
10	-0,14346	11,09589	15,38352	15,54464							
11	6,469269	15,65792	-4,20813	-1,39295							
12	16,88983	6,946458	11,41294	3,706702							
13	9,897359	4,090463	12,37089	6,767561							
14	-21,4958	-5,57278	8,900167	-26,3889							
15	5,454068	19,38966	12,24909	-6,72784							
16	-6,36377	9,647839	-2,13945	14,15991							
17											

Рис. 40. Результат применения средства ***Ковариация***

16. Регрессия

Покажем, что для проведения регрессионного анализа может использовать средство Регрессия. В отдельных таблицах оно вычисляет (рис. 42 и 43) следующее:

- методом наименьших квадратов — коэффициенты линейной (относительно этих коэффициентов) функции регрессии; вид функции регрессии определяется структурой исходных данных (подробнее об этом речь идет ниже);
- коэффициент детерминации и связанные с ним величины (таблица ***Регрессионная статистика***);
- дисперсионную таблицу и критериальную статистику для проверки значимости регрессии (таблица ***Дисперсионный анализ***);
- для каждого коэффициента регрессии — среднеквадратическое отклонение и другие его статистические характеристики, позволяющие проверить значимость этого коэффициента и построить для него доверительные интервалы;
- значения функции регрессии и остатки — разности между исходными значениями переменной Y и вычисленными значениями функции регрессии (таблица ***Вывод остатка***);
- вероятности, соответствующие упорядоченным по возрастанию значениям переменной Y (таблица ***Вывод вероятности***).

Кроме того, средство ***Регрессия*** строит три типа графиков, которые будут показаны ниже.

Пусть входной интервал X состоит из ***k*** диапазонов-столбцов, содержащих значения $\{x_{i1}\}$, $\{x_{i2}\}$, ..., $\{x_{ik}\}$ переменных X_1, X_2, ..., X_k. В каждом диапазоне содержится одинаковое количество значений. Входной интервал Y, состоящий из одного диапазона-столбца, должен содержать такое же количество значений. Средство ***Регрессия*** вычисляет коэффициенты функции регрессии вида

$$Y = m_1 X_1 + m_2 X_2 + \ldots + m_k X_k + b.$$

Это уравнение линейной множественной регрессии, если переменные X_i независимы. На основе данного уравнения, используя соответствующие значения переменных X_i, можно получить множество других уравнений регрессии. Например, если в качестве переменных X_i взять значения одной переменной X в степени i (т.е. $X_i = X^i$), получим уравнение полиномиальной регрессии

$$Y = m_1 X_1 + m_2 X^2 + \ldots + m_k X^k + b$$

На рис. 41 показан рабочий лист с исходными данными: входной интервал X состоит из пяти столбцов. В первом столбце представлены значения переменной X_1, во втором — квадраты значений переменной X_1 в третьем — значения второй переменной X_2, в четвертом — квадраты значений переменной X_2, в пятом — произведения значений переменных X_1 и X_2. Таким образом, в данном случае Регрессия будет вычислять значения коэффициентов уравнения регрессии вида

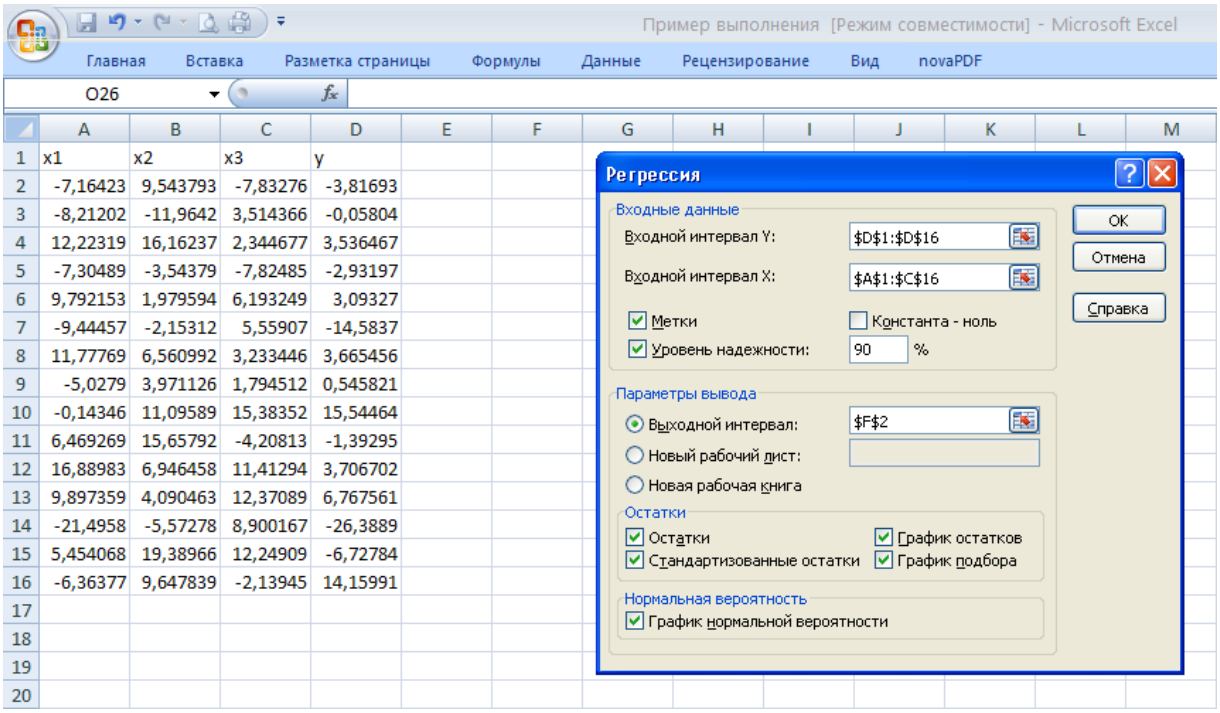
$$Y = m_1 X_1 + m_2 X_1^2 + m_3 X_2 + m_4 X_2^2 + m_5 X_1 X_2 + b$$

Отметим, что значения зависимой переменной Y в столбце F получены по формуле

$$Y = X_1 - m_2 X_1^{\,2} + m_3 X_2 + m_4 X_2^2 + m_5 X_1 X_2 + b$$

Здесь случайная переменная e имеет стандартное нормальное распределение.

Диалоговое окно средства ***Регрессия*** показано на рис. 41. В поле ***Входной интервал*** Y вводится адрес диапазона, содержащего значения зависимой переменной Y. Диапазон должен состоять из одного столбца. В поле ***Входной интервал*** X вводится адрес диапазона, содержащего значения переменной X. Диапазон должен состоять из одного или нескольких столбцов, но не более чем из 16 столбцов. Если указанные в полях ***Входной интервал*** Y и ***Входной интервал*** X диапазоны включают заголовки столбцов, то необходимо установить флажок опции ***Метки*** — эти заголовки будут использованы в выходных таблицах, сгенерированных средством ***Регрессия***.



	A	B	C	D
1	x1	x2	x3	y
2	-7,16423	9,543793	-7,83276	-3,81693
3	-8,21202	-11,9642	3,514366	-0,05804
4	12,22319	16,16237	2,344677	3,536467
5	-7,30489	-3,54379	-7,82485	-2,93197
6	9,792153	1,979594	6,193249	3,09327
7	-9,44457	-2,15312	5,55907	-14,5837
8	11,77769	6,560992	3,233446	3,665456
9	-5,0279	3,971126	1,794512	0,545821
10	-0,14346	11,09589	15,38352	15,54464
11	6,469269	15,65792	-4,20813	-1,39295
12	16,88983	6,946458	11,41294	3,706702
13	9,897359	4,090463	12,37089	6,767561
14	-21,4958	-5,57278	8,900167	-26,3889
15	5,454068	19,38966	12,24909	-6,72784
16	-6,36377	9,647839	-2,13945	14,15991

Рис.41. Исходные данные и диалоговое окно ***Регрессия***

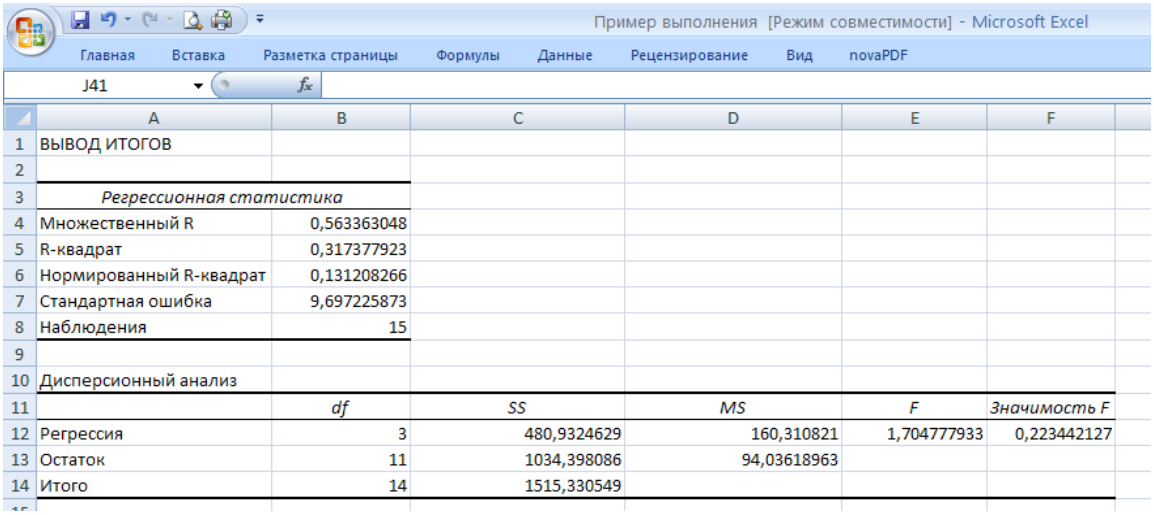
Флажок опции ***Константа - ноль*** следует установить, если в уравнении регрессии константа *b* принудительно полагается равной нулю. Опция ***Уровень надежности*** устанавливается тогда, когда необходимо построить доверительные интервалы для коэффициентов регрессии с доверительным уровнем, отличным от 0,95, который используется по умолчанию. После установки флажка опции ***Уровень надежности*** становится доступным поле ввода, в котором вводится новое значение доверительного уровня.

В области ***Остатки*** имеются четыре опции: ***Остатки, Стандартизованные остатки, График остатков*** и ***График подбора***. Если установлена хотя бы одна из них, то в выходных результатах появится таблица ***Вывод остатка***, в которой будут выведены значения функции регрессии и остатки —

разности между исходными значениями переменной Y и вычисленными значениями функции регрессии. Значения этой таблицы и возможности каждой из опций показаны ниже.

В области ***Нормальная вероятность*** имеется одна опция — ***График нормальной вероятности***; ее установка порождает в выходных результатах таблицу ***Вывод вероятности*** и приводит к построению соответствующего графика.

На рис.42-44 показаны части рабочего листа с выходными результатами средства ***Регрессия***, которые получены на основе исходных данных, приведенных на рис. 41. Рассмотрим подробнее эти результаты.



	A	B	C	D	E	F
1	ВЫВОД ИТОГОВ					
2						
3	*Регрессионная статистика*					
4	Множественный R	0,563363048				
5	R-квадрат	0,317377923				
6	Нормированный R-квадрат	0,131208266				
7	Стандартная ошибка	9,697225873				
8	Наблюдения	15				
9						
10	Дисперсионный анализ					
11		*df*	*SS*	*MS*	*F*	*Значимость F*
12	Регрессия	3	480,9324629	160,310821	1,704777933	0,223442127
13	Остаток	11	1034,398086	94,03618963		
14	Итого	14	1515,330549			

Рис.42. Верхняя часть рабочего листа с выходными результатами

В таблице ***Регрессионная статистика*** приводятся следующие данные.

- ***Множественный R*** — корень из коэффициента детерминации R^2, приведенного в следующей строке. Другое название этого показателя — индекс корреляции, или множественный коэффициент корреляции.

		Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 90,0%*	*Верхние 90,0%*
17	Y-пересечение	-0,625204356	3,563139621	-0,175464456	0,863902944	-8,46762178	7,217213067	-7,024192692	5,773783979
18	x1	0,505943825	0,302531696	1,672366343	0,122622401	-0,159923947	1,171811598	-0,037368253	1,049255904
19	x2	0,108634254	0,360880725	0,30102537	0,769008878	-0,685658866	0,902927373	-0,53946596	0,756734467
20	x3	-0,133012304	0,370948801	-0,358573215	0,726701391	-0,949465109	0,683440501	-0,799193622	0,533169014

	ВЫВОД ОСТАТКА			
26	*Наблюдение*	*Предсказанное y*	*Остатки*	*Стандартные остатки*
27	1	-2,17126413	-1,645669922	-0,191453317
28	2	-6,547203062	6,489165138	0,754934009
29	3	7,002957776	-3,46649074	-0,403283272
30	4	-3,665244893	0,733271924	0,085307108
31	5	3,720348659	-0,62707839	-0,072952806
32	6	-6,376952479	-8,206784488	-0,954757751
33	7	5,61630757	-1,950851448	-0,226957409
34	8	-2,976332946	3,522154029	0,409759006
35	9	-1,538592043	17,08323131	1,987422422
36	10	4,908601651	-6,301550423	-0,73310736
37	11	7,156661227	-3,44995952	-0,401360069
38	12	3,181187371	3,586373281	0,417230121
39	13	-13,2901012	-13,09878604	-1,523881554
40	14	2,611350561	-9,339190618	-1,086499181
41	15	-2,512255059	16,67216591	1,939600052

	ВЫВОД ВЕРОЯТНОСТИ	
26	*Персентиль*	*y*
27	3,333333333	-26,38888725
28	10	-14,58373697
29	16,66666667	-6,727840057
30	23,33333333	-3,816934052
31	30	-2,931972969
32	36,66666667	-1,392948772
33	43,33333333	-0,058037924
34	50	0,545821083
35	56,66666667	3,093270268
36	63,33333333	3,536467036
37	70	3,665456122
38	76,66666667	3,706701707
39	83,33333333	6,767560651
40	90	14,15991085
41	96,66666667	15,54463927

Рис.43. Нижняя часть рабочего листа с выходными результатами

- ***R-квадрат*** — коэффициент детерминации R^2; вычисляется как отношение регрессионной суммы квадратов (ячейка C12) к полной сумме квадратов (ячейка C14).
- ***Нормированный R-квадрат*** вычисляется по формуле $\frac{(n-1)R^2-k}{n-k-1}$, где n —количество значений переменной Y, k— количество столбцов во входном интервале переменной X.
- ***Стандартная ошибка*** — корень из остаточной дисперсии (ячейка D13).
- ***Наблюдения*** — количество значений переменной Y.

Дисперсионная таблица строиться следующим образом. В столбце ***SS*** приводятся суммы квадратов, в столбце ***df*** — число степеней свободы, в столбце ***MS*** — дисперсии. Строка ***Регрессия*** соответствует одноименной строке из таблицы, строка ***Остаток*** — строке ***Остатки*** и строка ***Итого*** — строке ***Полная***. В дисперсионной таблице _приведены формулы, по которым вычисляет соответствующие значения средство Регрессия. В столбце ***F*** вычислено значение критериальной статистики для проверки значимости регрессии. Это значение вычисляется как отношение регрессионной дисперсии к остаточной (ячейки D12 и D13). В столбце ***Значимость F*** вычисляется вероятность полученного значения критериальной статистики. (Эту вероятность с помощью формул Excel можно вычислить как =FРАСПР(E12;B12;B13).) Если эта вероятность меньше, например, 0,05 (заданного уровня значимости), то гипотеза о незначимости регрессии (т.е. гипотеза о том, что все коэффициенты функции регрессии равны нулю) отвергается и считается, что регрессия значима. В данном примере регрессия значима практически с любым уровнем значимости.

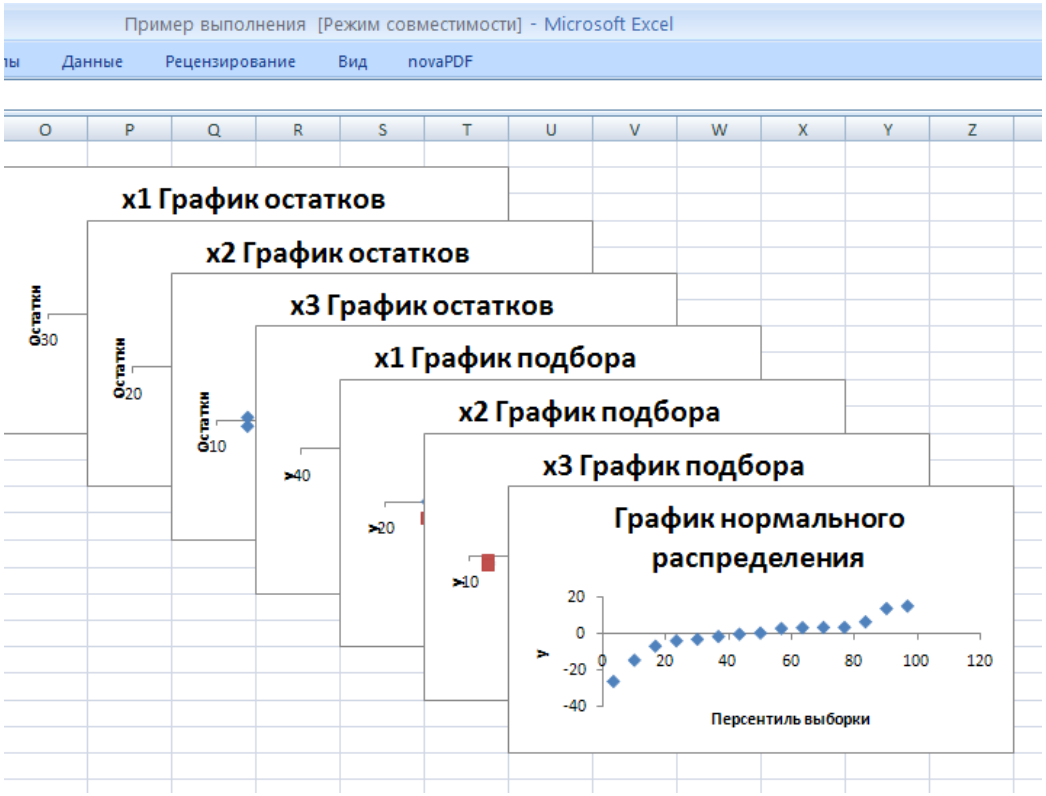


*Рис. 44. Графики, выводимые средством **Регрессия***

В следующей таблице (см. рис.43), в столбце ***Коэффициенты***, записаны вычисленные значения коэффициентов функции регрессии, при этом в строке ***Y-пересечение*** записано значение свободного члена *b*. В столбце ***Стандартная ошибка*** вычислены среднеквадратические отклонения коэффициентов. В столбце ***t-статистика*** записаны отношения значений коэффициентов к их среднеквадратическим отклонениям. Это значения критериальных статистик для проверки гипотез о значимости коэффициентов регрессии. В столбце ***P-Значение*** вычисляются уровни значимости, соответствующие значениям критериальных статистик. (Их можно вычислить с помощью формулы Excel =СТЬЮДЕНТРАСП (ABS(D17);14;2), например, для значения в ячейке E17; второй аргумент в функции СТЬЮДРАСП вычисляется как (*n* — *k* — 1) .) Если вычисленный уровень значимости меньше заданного уровня значимости (например, 0,05), то принимается гипотеза о значимом отличии коэффициента от нуля; в противном случае принимается гипотеза о незначимом отличии коэффициента от нуля. В данном примере только коэффициент *b* незначимо отличается от нуля.

В столбцах ***Нижние 95%*** и ***Верхние 95%*** приводятся границы доверительных интервалов с доверительным уровнем 0,95. Эти границы вычисляются по формулам

Нижние 95% = Коэффициент - Стандартная ошибка × t_a;

Верхние 95% = Коэффициент + Стандартная ошибка × t_a.

Здесь t_a — квантиль порядка α распределения Стьюдента с *(n - k -* 1) степенью свободы. В данном случае α = 0,95. Аналогично вычисляются границы доверительных интервалов в столбцах ***Нижние 90,0%*** и ***Верхние 90,0%***. Отметим, что если в диалоговом окне ***Регрессия*** не устанавливать опцию ***Уровень надежности***, то будут повторены столбцы ***Нижние 95%*** и ***Верхние 95%***.

Рассмотрим таблицу ***Вывод остатка*** из выходных результатов средства ***Регрессия***. Напомним, что эта таблица появляется в выходных результатах только тогда, когда установлена хотя бы одна опция в области ***Остатки*** диалогового окна ***Регрессия***. В столбце ***Наблюдение*** приводятся порядковые номера значений переменной Y. В столбце ***Предсказанное Y*** вычисляются значения функции регрессии $\hat{y}_i = f(x_i)$ для тех значений переменной X, которым соответствует порядковый номер i в столбце ***Наблюдение***. В столбце ***Остатки*** содержатся разности (остатки) $\varepsilon_i = y_i - \hat{y}_i$, а в столбце ***Стандартные остатки*** — нормированные остатки, которые вычисляются как отношения $\varepsilon_i / s_\varepsilon$, где s_ε — среднеквадратическое отклонение остатков. Квадрат величины s_ε вычисляется по формуле $s_\varepsilon^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2$, где $\bar{\varepsilon}$ — среднее остатков. Здесь величину s_ε^2 можно вычислить как отношение двух значений из дисперсионной таблицы: суммы квадратов остатков (ячейка C13) и степени свободы из строки ***Итого*** (ячейка B14).

По значениям таблицы ***Вывод остатка*** средство ***Регрессия*** строит два типа графиков: графики остатков и графики подбора (если установлены соответствующие опции в области ***Остатки*** диалогового окна Регрессия). На рис. 45 показаны образцы этих графиков (графики немного переформатированы по сравнению с оригиналами). Они строятся для каждого компонента переменной X в отдельности. На графиках остатков отображаются остатки, т.е. разности между исходными

значениями Y и вычисленными по функции регрессии для каждого значения компонента переменной X. На графиках подбора отображаются как исходные значения Y, так и вычисленные значения функции регрессии для каждого значения компонента переменной X. (На графиках подбора, представленных на рис.45, эти значения практически совпадают.)

Последней таблицей выходных результатов средства ***Регрессия*** является таблица Вывод вероятности (см. рис. 43). Она появляется, если в диалоговом окне ***Регрессия*** установлена опция ***График нормальной вероятности***. Значения в столбце ***Персентиль*** вычисляются следующим образом. Вычисляется шаг $h = (1/n) \times 100\%$, первое значение равно $h/2$, последнее равно $100 - h/2$. Начиная со второго значения каждое последующее значение равно предыдущему, к которому прибавлен шаг *h*. В столбце Y приведены значения переменной Y, упорядоченные по возрастанию. По данным этой таблицы строится так называемый график нормального распределения (рис. 46). Он позволяет визуально оценить степень линейности зависимости между переменными X и Y.

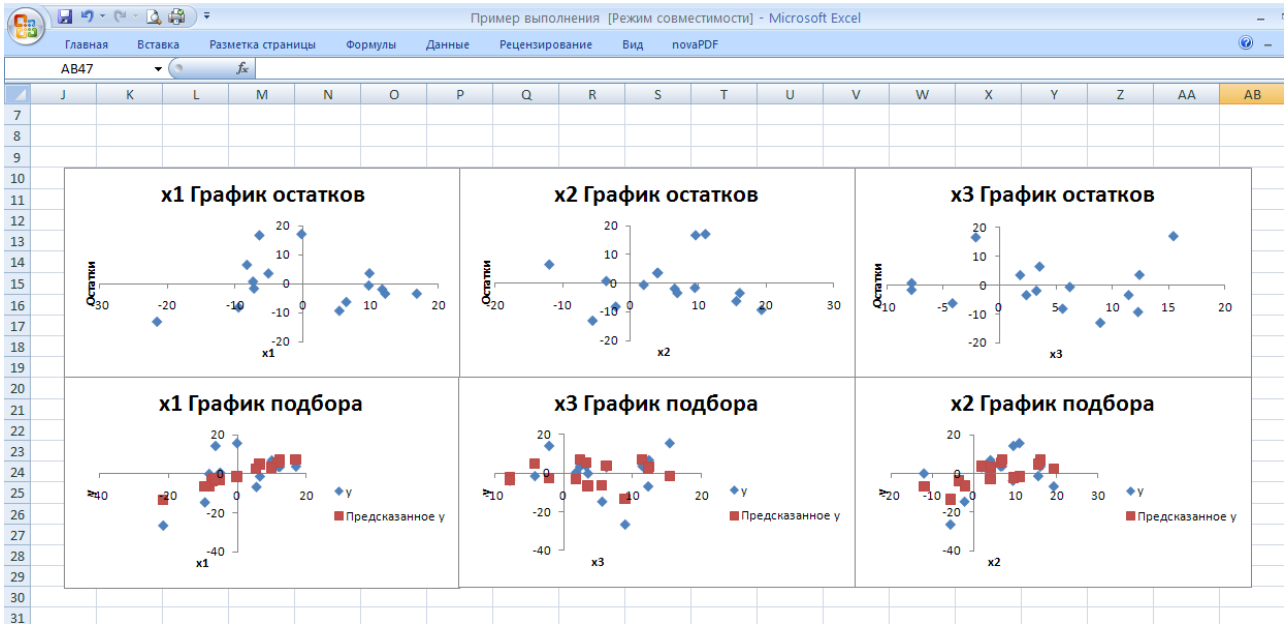


Рис. 45. Примеры графиков остатков и подбора

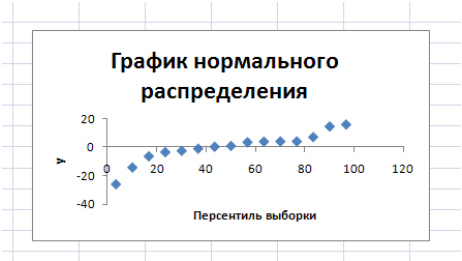


Рис. 46. График нормального распределения

17. Скользящее среднее

Метод скользящего среднего — один из наиболее широко используемых способов сглаживания значений временного ряда. Метод основан на локальном усреднении, когда за новое значение временного ряда берется среднее *k* последовательных значений, ближайших к заменяемому значению.

Пусть имеются дискретные наблюдения y_1 и y_2, ..., y_n и задано число k наблюдений, по которым будет проводиться усреднение. Значение скользящего среднего для значения t вычисляется по формуле $y_t = \frac{1}{k}\sum_{i=0}^{k-1} y_{t-i}$. Отметим, что по этой формуле выполняет вычисления средство ***Скользящее среднее***, но существуют и другие способы вычисления скользящего среднего.

На рис. 47 показаны исходные данные, для которых будут вычисляться скользящие средние, и диалоговое окно ***Скользящее среднее***. В поле ввода ***Входной интервал*** в качестве исходных данных задан диапазон B1:B17. Поскольку этот диапазон содержит заголовок, установлен флажок опции ***Метки в первой строке***. В поле ***Интервал*** вводится число k — количество значений, по которым подсчитывается скользящее среднее. Если этот параметр не задан, то по умолчанию используется значение 3.

Если установлен флажок опций ***Вывод графика***, то будет построен график, отображающий исходные значения y_t и сглаженные скользящим средним значения (рис.48). Если также установлен флажок опции ***Стандартные погрешности***, то к значениям вычисленных средних будет добавлен столбец, в котором будут записаны стандартные погрешности, вычисляемые как сумма квадратов разностей между исходными и расчетными k значения y_t, деленная на число k. Формула Excel, по которой подсчитываются стандартные погрешности, показана на рис.48.

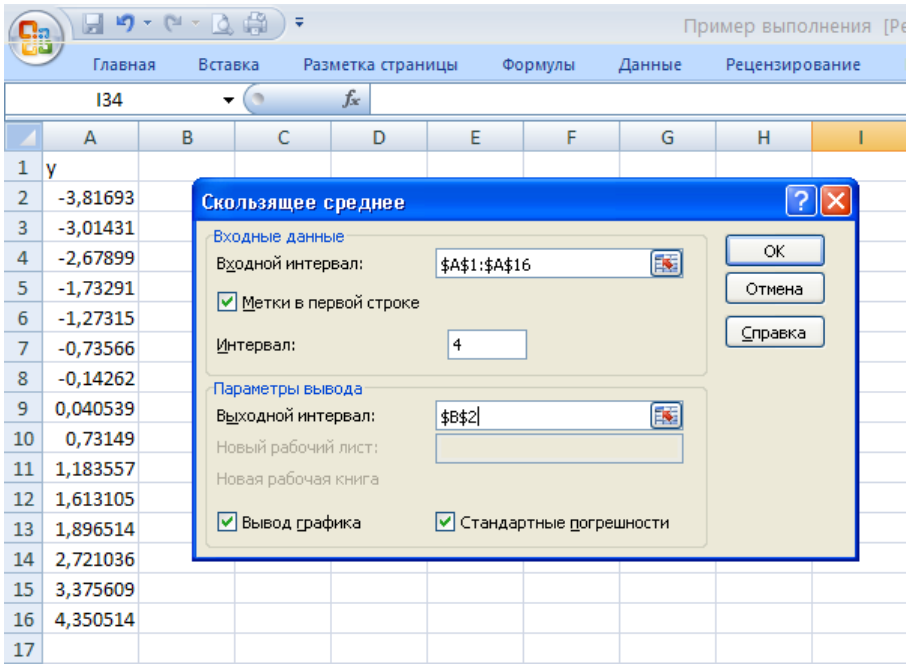
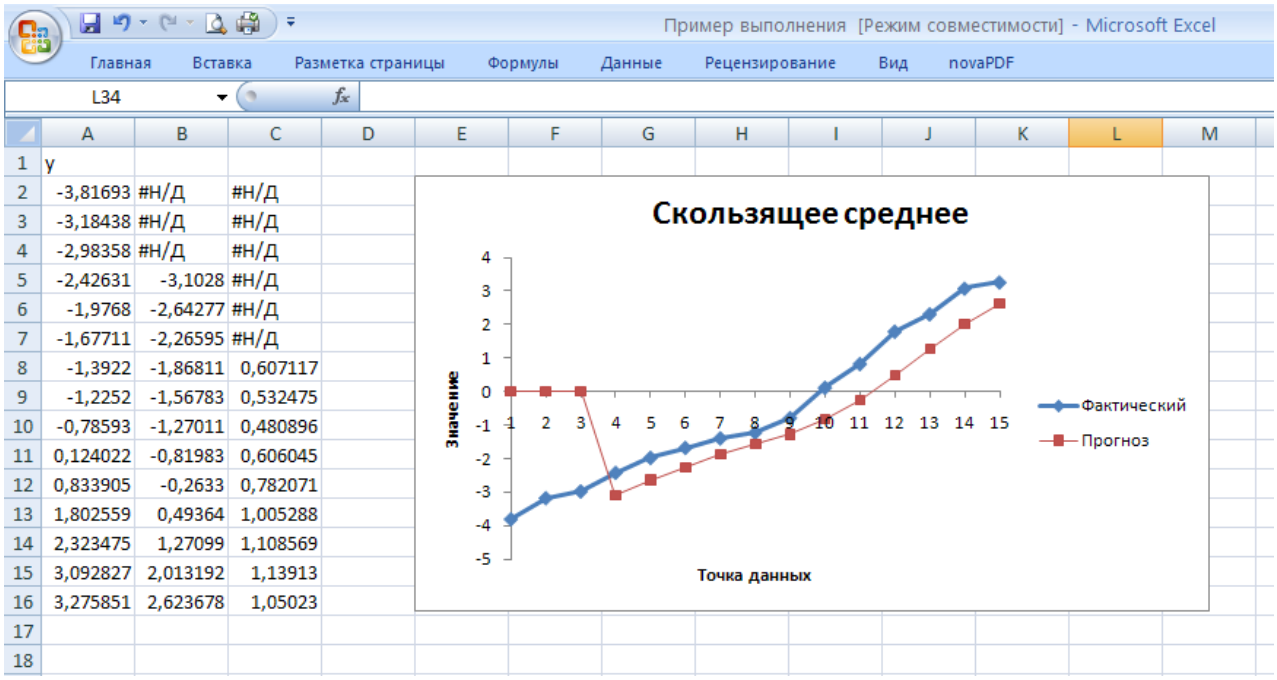


Рис.47. Исходные данные и диалоговое окно ***Скользящее среднее***



	A	B	C
1	y		
2	-3,81693	#Н/Д	#Н/Д
3	-3,18438	#Н/Д	#Н/Д
4	-2,98358	#Н/Д	#Н/Д
5	-2,42631	-3,1028	#Н/Д
6	-1,9768	-2,64277	#Н/Д
7	-1,67711	-2,26595	#Н/Д
8	-1,3922	-1,86811	0,607117
9	-1,2252	-1,56783	0,532475
10	-0,78593	-1,27011	0,480896
11	0,124022	-0,81983	0,606045
12	0,833905	-0,2633	0,782071
13	1,802559	0,49364	1,005288
14	2,323475	1,27099	1,108569
15	3,092827	2,013192	1,13913
16	3,275851	2,623678	1,05023
17			
18			

Рис.48. Результаты вычислений

18. Экспоненциальное сглаживание

Экспоненциальное сглаживание, как и скользящее среднее, используется для выравнивания (сглаживания) значений временных рядов. Если имеются дискретные наблюдения y_1 , y_2, ..., y_n , то сглаженные значения вычисляются по формуле $\hat{y}_{t+1} = \alpha \cdot y_t + (1-\alpha)\hat{y}_t$, где y_t — сглаженное значение для предыдущего *t*, *α*— постоянная сглаживания, также называемая ***фактором затухания*** (это число из интервала (0, 1)).

На рис.49 показаны рабочий лист Excel с исходными данными (данные взяты из примера предыдущего раздела) и диалоговое окно ***Экспоненциальное сглаживание***. В поле ***Входной интервал*** указывается адрес диапазона, содержащего значения y_i. Если этот диапазон включает заголовок, то надо установить флажок опции ***Метки***. В поле ***Фактор затухания*** задается постоянная сглаживания; если она не задана, то по умолчанию используется значение 0,4. Установка флажков опций ***Вывод графика*** и ***Стандартные погрешности*** приводит к построению графика, на котором будут отображаться исходные и сглаженные значения (рис.50), и к выводу дополнительного столбца со значениями погрешностей. Эти погрешности вычисляются как сумма квадратов разностей между тремя последовательными исходными и расчетными значениями, деленная на число 3. Формула Excel, по которой подсчитываются стандартные погрешности, показана на рис. 50.

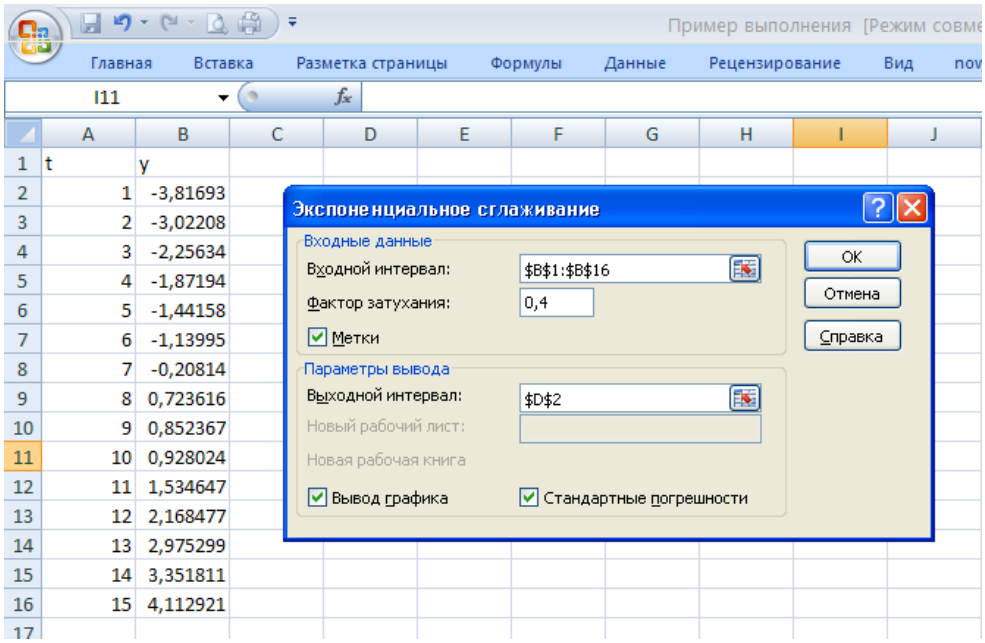


Рис. 49. Исходные данные и диалоговое окно ***Экспоненциальное сглаживание***

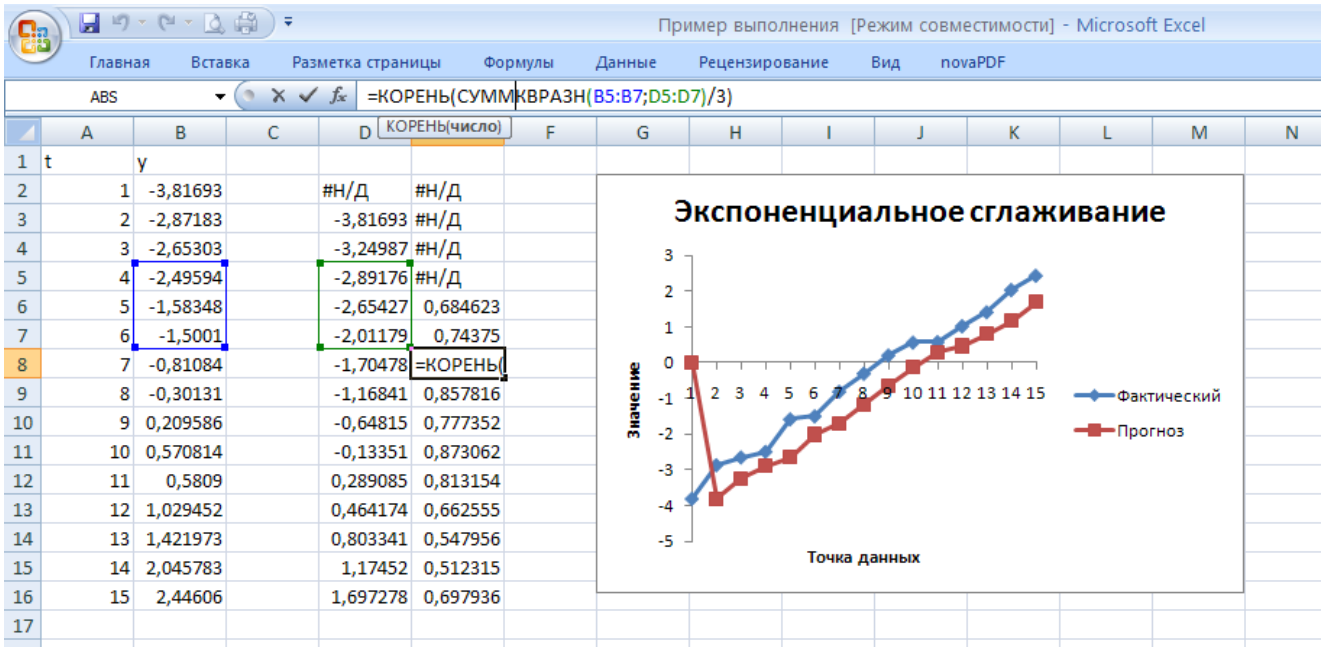


Рис. 50. Результаты вычислений

19. Анализ Фурье

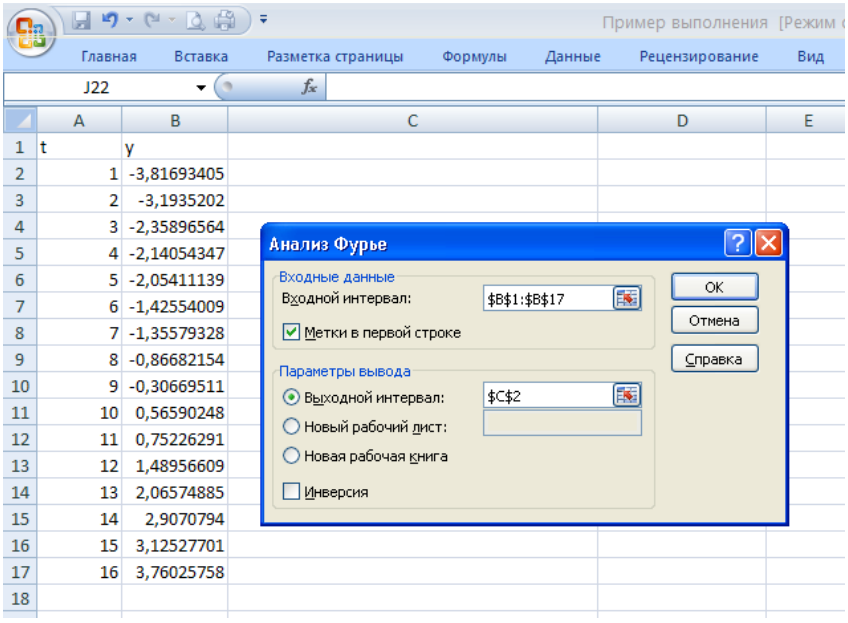
Данное средство выполняет дискретное преобразование Фурье. Это преобразование используется в анализе линейных систем и применяется к временным рядам для выявления периодических (спектральных) составляющих таких рядов.

Если имеются дискретные наблюдения y_1 ,y_2, ..., y_n, то ***прямое дискретное преобразование Фурье*** выполняется в соответствии с формулой $Y_k = \sum_{j=1}^{n} y_j e^{-i\frac{2\pi}{n}jk}$,k = *0,1,..., n-1*. Результаты

преобразования Y_k являются комплексными числами, модуль которых равен амплитуде k-й спектральной составляющей (k-й гармоники), а аргумент комплексного числа Y_k равен фазе этой гармоники. Аналогично определяется *обратное дискретное преобразование Фурье* ($y_j = \sum_{k=0}^{n-1} Y_k e^{i\frac{2\pi}{n}jk}$), которое преобразует спектральное представление временного ряда в действительное.

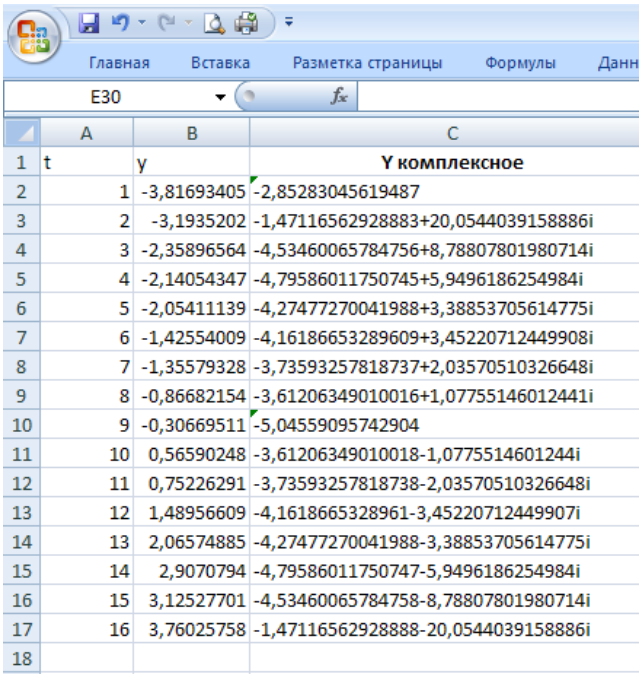
Средство Анализ Фурье выполняет как прямое так и обратное преобразования методом *быстрого преобразования Фурье* (БПФ). Применение метода БПФ диктует условие, чтобы количество исходных значений как для прямого, так и для обратного преобразований, было равно некоторой положительной степени числа 2. Максимальное число значений, которое может обработать средство ***Анализ Фурье***, составляет 4096 (= 2^{12}). Для применения обратного преобразования Фурье исходные значения должны быть в формате комплексных чисел $x + yi$ или $x + yj$ (i и j — обозначение мнимой единицы). Если x является отрицательным числом, перед ним ставится апостроф (').

На рис.51 показаны рабочий лист с исходными данными и диалоговое окно ***Анализ Фурье.*** Результат прямого преобразования Фурье показан на рис. 52. Первое значение (ячейка C2) равно сумме исходных данных.



	A	B	C	D	E
1	t	y			
2	1	-3,81693405			
3	2	-3,1935202			
4	3	-2,35896564			
5	4	-2,14054347			
6	5	-2,05411139			
7	6	-1,42554009			
8	7	-1,35579328			
9	8	-0,86682154			
10	9	-0,30669511			
11	10	0,56590248			
12	11	0,75226291			
13	12	1,48956609			
14	13	2,06574885			
15	14	2,9070794			
16	15	3,12527701			
17	16	3,76025758			
18					

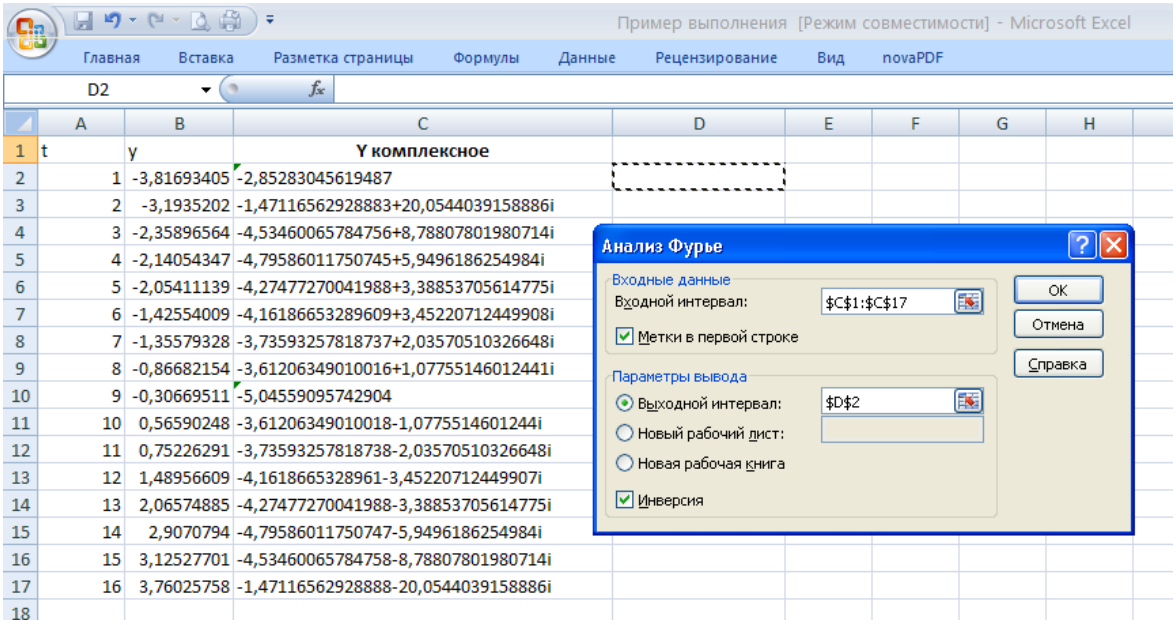
Рис. 51. Исходные данные и диалоговое окно ***Анализ Фурье***



	A	B	C
1	t	y	Y комплексное
2	1	-3,81693405	-2,85283045619487
3	2	-3,1935202	-1,47116562928883+20,0544039158886i
4	3	-2,35896564	-4,53460065784756+8,78807801980714i
5	4	-2,14054347	-4,79586011750745+5,9496186254984i
6	5	-2,05411139	-4,27477270041988+3,38853705614775i
7	6	-1,42554009	-4,16186653289609+3,45220712449908i
8	7	-1,35579328	-3,73593257818737+2,03570510326648i
9	8	-0,86682154	-3,61206349010016+1,07755146012441i
10	9	-0,30669511	-5,04559095742904
11	10	0,56590248	-3,61206349010018-1,0775514601244i
12	11	0,75226291	-3,73593257818738-2,03570510326648i
13	12	1,48956609	-4,1618665328961-3,45220712449907i
14	13	2,06574885	-4,27477270041988-3,38853705614775i
15	14	2,9070794	-4,79586011750747-5,9496186254984i
16	15	3,12527701	-4,53460065784758-8,78807801980714i
17	16	3,76025758	-1,47116562928888-20,0544039158886i
18			

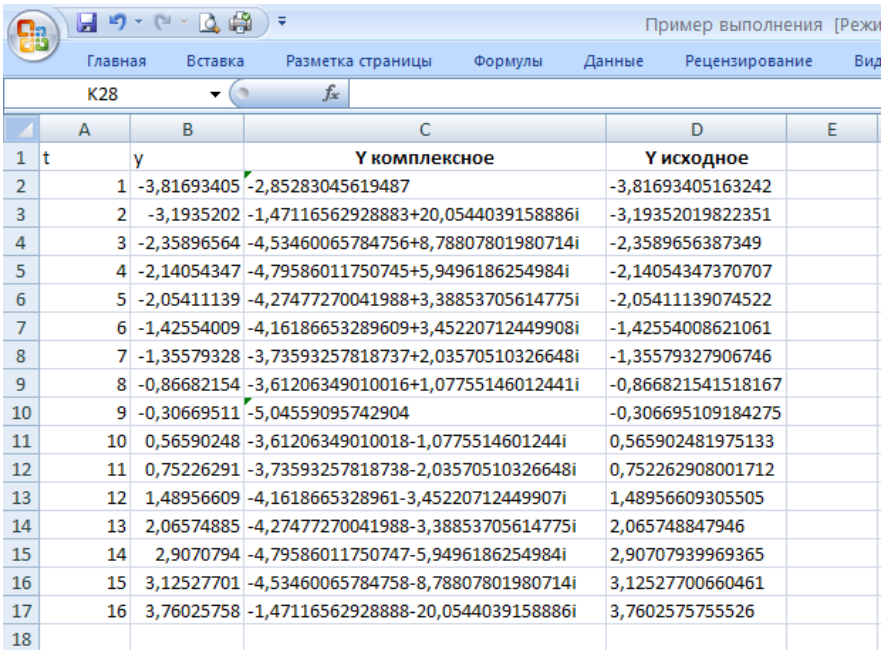
Рис. 52. Результат прямого преобразования Фурье

На рис. 53 показаны рабочий лист с исходными данными (результат прямого преобразования Фурье) для обратного преобразования и диалоговое окно ***Анализ Фурье***, в котором установлен флажок опции ***Инверсия***. Результат обратного преобразования Фурье показан на рис.54; он совпадает с первоначальными данными из столбца B.



	A	B	C
1	t	y	Y комплексное
2	1	-3,81693405	-2,85283045619487
3	2	-3,1935202	-1,47116562928883+20,0544039158886i
4	3	-2,35896564	-4,53460065784756+8,78807801980714i
5	4	-2,14054347	-4,79586011750745+5,9496186254984i
6	5	-2,05411139	-4,27477270041988+3,38853705614775i
7	6	-1,42554009	-4,16186653289609+3,45220712449908i
8	7	-1,35579328	-3,73593257818737+2,03570510326648i
9	8	-0,86682154	-3,61206349010016+1,07755146012441i
10	9	-0,30669511	-5,04559095742904
11	10	0,56590248	-3,61206349010018-1,0775514601244i
12	11	0,75226291	-3,73593257818738-2,03570510326648i
13	12	1,48956609	-4,1618665328961-3,45220712449907i
14	13	2,06574885	-4,27477270041988-3,38853705614775i
15	14	2,9070794	-4,79586011750747-5,9496186254984i
16	15	3,12527701	-4,53460065784758-8,78807801980714i
17	16	3,76025758	-1,47116562928888-20,0544039158886i
18			

*Рис. 53. Исходные данные для обратного преобразования Фурье и диалоговое окно **Анализ Фурье***



	A	B	C	D	E
1	t	y	Y комплексное	Y исходное	
2	1	-3,81693405	-2,85283045619487	-3,81693405163242	
3	2	-3,1935202	-1,47116562928883+20,0544039158886i	-3,19352019822351	
4	3	-2,35896564	-4,53460065784756+8,78807801980714i	-2,3589656387349	
5	4	-2,14054347	-4,79586011750745+5,9496186254984i	-2,14054347370707	
6	5	-2,05411139	-4,27477270041988+3,38853705614775i	-2,05411139074522	
7	6	-1,42554009	-4,16186653289609+3,45220712449908i	-1,42554008621061	
8	7	-1,35579328	-3,73593257818737+2,03570510326648i	-1,35579327906746	
9	8	-0,86682154	-3,61206349010016+1,07755146012441i	-0,866821541518167	
10	9	-0,30669511	-5,04559095742904	-0,306695109184275	
11	10	0,56590248	-3,61206349010018-1,0775514601244i	0,565902481975133	
12	11	0,75226291	-3,73593257818738-2,03570510326648i	0,752262908001712	
13	12	1,48956609	-4,1618665328961-3,45220712449907i	1,48956609305505	
14	13	2,06574885	-4,27477270041988-3,38853705614775i	2,065748847946	
15	14	2,9070794	-4,79586011750747-5,9496186254984i	2,90707939969365	
16	15	3,12527701	-4,53460065784758-8,78807801980714i	3,12527700660461	
17	16	3,76025758	-1,47116562928888-20,0544039158886i	3,7602575755526	
18					

Рис. 54. Результат обратного преобразования Фурье

Задания к третьей части лабораторных работ:

Даны 5 факторов (4 фактора – из файла примера – ЛИСТ «Корреляция», 5-ый фактор, имеющий распределение Пуассона, генерируется самостоятельно).

1. Для всех 5-ти факторов построить ковариационную матрицу, корреляционную матрицу. Проанализировать полученные результаты.

2. Получить 2 уравнения регрессии, описывающих влияние факторов на показатель, при условии, что:

a) константа *b* принудительно полагается равной нулю.

b) константа *b* может принимать значения, отличные от нуля.

Сравнить получившиеся коэффициенты регрессии. Сделать вывод о влиянии константы *b* на значения коэффициентов регрессии.

3. Проверить значимости регрессии по F-статистике для случая b). Критическое значение F – статистики можно получить функцией FРАСПОБР(1-α;m-1;n-m), где m-число факторов.

4. Графически оценить степень линейности зависимости между факторами

Контрольные вопросы

1. Что характеризует ковариация?

2. Чему равна ковариация двух независимых случайных величин?

3. Каково максимальное абсолютное значение коэффициента корреляции?

4. Какие случайные величины называются некоррелированными?