

Статистический анализ данных в MS Excel

- 1. Обзор и характеристика основных статистических функций, входящих в MS Excel.**
- 2. Работа с пакетом анализа данных в MS Excel.**

Литература:

- 1. Г.И. Просветов Анализ данных с помощью Excel. Задачи и решения. М: 2009**
- 2. А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов Статистический анализ данных в MS Excel М: 2012**

Понятие анализа данных

Анализ данных – область математики и информатики, занимающаяся построением и исследованием наиболее общих математических методов и вычислительных алгоритмов извлечения знаний из экспериментальных данных.

Анализ данных – это процесс исследования, фильтрации, преобразования и моделирования данных с целью извлечения полезной информации и принятия решения.

Статистические функции MS Excel

Все статистические функции, входящие в MS Excel можно разбить на **восемь** подразделов:

1. Предварительная обработка данных;
2. Определение характеристик положения;
3. Определение корреляции, ковариации;
4. Определение характеристик рассеивания
5. Интервальное оценивание (определение вероятности попадания дискретной случайной величины в интервал);
6. Определения параметров распределения непрерывной случайной величины;
7. Определение параметров распределения дискретной случайной величины;
8. Построение уравнения регрессии и прогнозирования.

Предварительная обработка данных

1. Подсчет количества значений (**СЧЕТ**).
Определение экстремальных значений совокупности данных (**МАКС, МИН**)
Подсчет частот из массива данных, попадающих в заданные интервалы (**ЧАСТОТА**)
Оценка относительного положения точки (**ПРОЦЕНТРАНГ**)
Определение величины, соответствующей ее относительному положению (**ПЕРСЕНТИЛЬ**)
Определение числа перестановок (**ПЕРЕСТ**)
Определение ранга чисел в списке чисел (**РАНГ**)

Предварительная обработка данных

Массив данных

	A
1	5,978508
2	3,902837
3	27,36091
4	7,631094
5	8,351607
6	7,423104
7	20,14716
8	6,044401
9	16,88405
10	19,54965
11	

СЧЕТ

fx =СЧЁТ(A1:A10)		
D	E	F
	10	

МАКС

fx =МАКС(A1:A10)		
D	E	F
	27,36091	

ПРОЦЕНТРАНГ

fx =ПРОЦЕНТРАНГ(A1:A10;A8)			
D	E	F	G
	0,222		

ЧАСТОТА

fx {=ЧАСТОТА(A1:A10;D2:D7)}			
D	E	F	G
5		1	
9		5	
15		0	
20		2	
25		1	
30		1	

ПЕРСЕНТИЛЬ

fx =ПЕРСЕНТИЛЬ(A1:A10;0,5)		
D	E	F
	7,991351	

РАНГ

fx =РАНГ(A4;A1:A10)		
D	E	F
	6	

0-5	5-9	9-15	15-20	20-25	25-30
1	5	0	2	1	1

Определение характеристик положения

1. **Определение среднего (СРЗНАЧ, СРГЕОМ)**
Определение моды в интервале данных или массиве (МОДА)
Определение медианы (МЕДИАНА)
Определение квартилей (КВАРТИЛЬ)

Определение характеристик положения

Массив данных

	A
1	5,978508
2	3,902837
3	27,36091
4	7,631094
5	8,351607
6	7,423104
7	20,14716
8	6,044401
9	16,88405
10	19,54965
11	

СРЗНАЧ

fx	
=СРЗНАЧ(A1:A10)	
E	F
12,32733	

СРГЕОМ

fx	
=СРГЕОМ(A1:A10)	
E	F
10,19584	

МЕДИАНА

fx	
=МЕДИАНА(A1:A10)	
E	F
7,991351	

МОДА

E2	fx	
	=МОДА(A1:B10)	
A	B	
5,978508	2	
3,902837	5	3
27,36091	6	
7,631094	3	
8,351607	3	
7,423104	3	
20,14716	3	
6,044401	3	
16,88405	3	
19,54965	3	

КВАРТИЛЬ

fx		
=КВАРТИЛЬ(A1:A10;3)		
E	F	G
18,88325		

Определение характеристик рассеивания

1. **Определение среднего линейного отклонения (СРОТКЛ)**
Определение суммы квадратов отклонения (ДИСП)
Вычисление стандартного (среднего квадратического) отклонения (СТАНДОТКЛОН)
Определения асимметрии распределения (СКОС)
Определения эксцесса (ЭКСЦЕСС)

Определение характеристик рассеивания

Массив данных

	A
1	5,978508
2	3,902837
3	27,36091
4	7,631094
5	8,351607
6	7,423104
7	20,14716
8	6,044401
9	16,88405
10	19,54965
11	

СРОТКЛ

fx	=СРОТКЛ(A1:A10)	
	E	F
	6,926489	

КВАДРОТКЛ

fx	=КВАДРОТКЛ(A1:A10)		
	E	F	G
	572,7512		

ДИСП

fx	=ДИСП(A1:A10)	
	E	F
	63,63902	

СТАНДОТКЛОН

fx	=СТАНДОТКЛОН(A1:A10)		
	E	F	G
	7,977407		

СКОС

fx	=СКОС(A1:A10)	
	E	F
	0,810077	

ЭКСЦЕСС

fx	=ЭКСЦЕСС(A1:A10)		
	E	F	G
	-0,70134		

Зависимость случайных величин

1. **Определение ковариации (КОВАР)**
Определение коэффициента корреляции (КОРРЕЛ)

Зависимость случайных величин

Массив данных

	A	B
1	5,978508	3,060508
2	3,902837	16,12113
3	27,36091	21,77261
4	7,631094	10,68774
5	8,351607	24,13386
6	7,423104	1,916842
7	20,14716	21,84935
8	6,044401	-2,84641
9	16,88405	21,68298
10	19,54965	21,24717

КОВАР

fx	=КОВАР(A1:A10;B1:B10)		
	E	F	G
	45,75339		

КОРРЕЛ

fx	=КОРРЕЛ(A1:A10;B1:B10)		
	E	F	G
	0,636838		

Интервальное оценивание

1. **Определение доверительного интервала для среднего (ДОВЕРИТ)**
Определение вероятности попадания дискретной случайной величины в интервал (ВЕРОЯТНОСТЬ)

Интервальное оценивание

Массив данных

	A
1	5,978508
2	3,902837
3	27,36091
4	7,631094
5	8,351607
6	7,423104
7	20,14716
8	6,044401
9	16,88405
10	19,54965
11	

ДОВЕРИТ

f _x			
	E	F	G
	6,985885		

ВЕРОЯТНОСТЬ

E2		f _x	=ВЕРОЯТНОСТЬ(A1:A10;B1:B10;6;14)				
A	B	C	D	E	F	G	H
5,978508	0,1						
3,902837	0,1			0,35			
27,36091	0,1						
7,631094	0,1						
8,351607	0,05						
7,423104	0,1						
20,14716	0,15						
6,044401	0,1						
16,88405	0,1						
19,54965	0,1						

Определение параметров распределения непрерывных случайных величин

1. Определение значения функции распределения и функции плотности нормального распределения (**НОРМРАСПР**)
2. Определение аргумента по значению функции распределения (**НОРМОБР**)
3. Определение вероятности статистики z при проверке гипотезы о равенстве статистической оценки математического ожидания заданному значению (**ZТЕСТ**)
4. Определение значений функций распределения отличных от нормальных (**ЛОГНОРМРАСП,**
СТЬЮДРАСП...)
5. Проверка гипотезы о равенстве дисперсий (**ФТЕСТ**)

Определение параметров распределения непрерывных случайных величин

НОРМРАСП

fx =НОРМРАСП(1;10;20;ИСТИНА)		
C	D	E
		0,326355

НОРМОБР

НОРМОБР			
Вероятность	0,326355		= 0,326355
Среднее	10		= 10
Стандартное_откл	20		= 20
			= 0,99998778

Массив данных

	A	B
1	5,978508	6,997678
2	3,902837	-2,77683
3	27,36091	12,44257
4	7,631094	22,76474
5	8,351607	21,9835
6	7,423104	27,33133
7	20,14716	-11,8359
8	6,044401	7,658188
9	16,88405	20,95023
10	19,54965	-0,86701

ZТЕСТ

ZТЕСТ			
Массив	A1:A10		= {5,97850842797197;3,90283688221;
X	7		= 7
Сигма	6,5		= 6,5
			= 0,004774178

ФТЕСТ

ФТЕСТ			
Массив1	A1:A10		= {5,97850842797197;3,90283688221...
Массив2	B1:B10		= {6,99767840866116;-2,7768316815...
			= 0,166776448

Построение уравнения регрессии и прогнозирование

1. Определение параметров линейной регрессии (**ЛИНЕЙН**)
2. Определение значений результативного признака по линейному уравнению регрессии (**ТЕНДЕНЦИЯ**)
3. Определение значения уравнения регрессии вида $y=b_0+b_1x$ в заданной точке (**ПРЕДСКАЗ**)

Построение уравнения регрессии и прогнозирование

Массив данных

	A	B
1	5,978508	6,997678
2	3,902837	-2,77683
3	27,36091	12,44257
4	7,631094	22,76474
5	8,351607	21,9835
6	7,423104	27,33133
7	20,14716	-11,8359
8	6,044401	7,658188
9	16,88405	20,95023
10	19,54965	-0,86701

ЛИНЕЙН

ЛИНЕЙН

Известные_значения_y B1:B10 = {6,99767840866116;-2,77683168...}

Известные_значения_x A1:A10 = {5,97850842797197;3,902836882218}

Конст ИСТИНА = ИСТИНА

Статистика ЛОЖЬ = ЛОЖЬ

= {-0,373274072200147;15,066325560}

ТЕНДЕНЦИЯ

ТЕНДЕНЦИЯ

Известные_значения_y B1:B10 = {6,99767840866116;-2,7768316815...}

Известные_значения_x A1:A10 = {5,97850842797197;3,9028368822182}

Новые_значения_x 37 = {37}

Конст ИСТИНА = ИСТИНА

= {1,25518488950893}

ПРЕДСКАЗ

ПРЕДСКАЗ

x 37 = 37

Известные_значения_y B1:B10 = {6,99767840866116;-2,776831681...}

Известные_значения_x A1:A10 = {5,97850842797197;3,902836882218}

= 1,25518489

Работа с пакетом анализа данных в MS Excel.

The screenshot displays the 'Параметры Excel' (Excel Options) dialog box in Russian. The 'Настройка' (Customization) tab is active, and the 'Настройки' (Customizations) list on the left has 'Пакет анализа' (Data Analysis Tools) selected. The main pane shows a list of available add-ins, with 'Пакет анализа' checked. A tooltip for 'Пакет анализа' is visible, stating: 'Содержит инструменты для анализа научных и финансовых данных' (Contains tools for analyzing scientific and financial data). The background shows the 'Анализ данных' (Data Analysis) task pane and a file explorer window with a list of Excel files.

Параметры Excel

Управление надстройками Microsoft Office.

Настройки

Имя	Расположение	Тип
Активные надстройки приложений		
Пакет анализа	C:\...ANALYS32.XLL	H
Неактивные надстройки приложений		
VBA для помощника по Интернету	C:\...ry\HTML.XLAM	H
Дата (списки смарт-тегов)	C:\...Tag\MOFL.DLL	C
Имя (получатели сообщений Outlook)	C:\...g\FNAME.DLL	C
Инструменты для евро	eurotool.xlam	H
Колонтитулы	C:\...OFFRHD.DLL	I
Мастер подстановок	lookup.xlam	H
Мастер суммирования	sumif.xlam	H
Настраиваемые XML-данные	C:\...OFFRHD.DLL	I
Невидимое содержимое	C:\...OFFRHD.DLL	I
Пакет анализа - VBA	C:\...PVBAEN.XLAM	H
Поиск решения	solver.xlam	H
Скрытые листы	C:\...OFFRHD.DLL	I
Скрытые строки и столбцы	C:\...OFFRHD.DLL	I

Надстройки, связанные с документами
Отсутствуют надстройки, связанные с документами

Отключенные надстройки приложений

Надстройка: Пакет анализа
Издатель: Microsoft Corporation
Расположение: C:\Program Files\Microsoft Office\Office12\Library\Analysis\AM...
Описание: Содержит инструменты для анализа научных и финансовых данных

Управление: Надстройки Excel **Перейти...**

Надстройки

Доступные надстройки:

- VBA для помощника по Интернету
- Инструменты для евро
- Мастер подстановок
- Мастер суммирования
- Пакет анализа
- Пакет анализа - VBA
- Поиск решения

Пакет анализа

Содержит инструменты для анализа научных и финансовых данных

Отправить
Опубликовать
Закреть

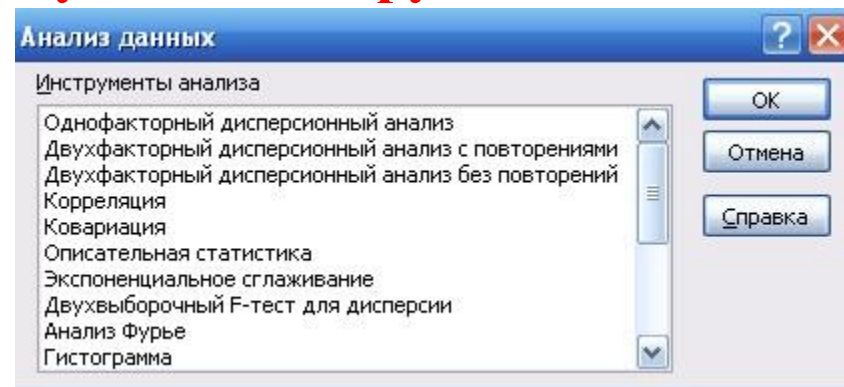
Данные для сортировки.xlsx
Задание к билету 30.xlsx
Копирование формул.xlsx
Данные для сортировки.xlsx
Решение задачи.xlsx

Параметры Excel | Выход из Excel

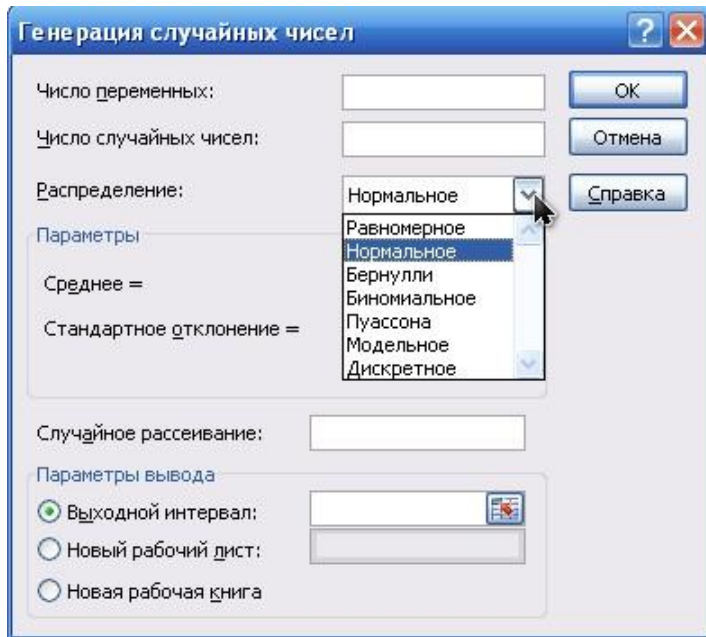
Работа с пакетом анализа данных в MS Excel.

В пакет анализа данных входят следующие инструменты:

1. Генерация случайных чисел
2. Выборка
3. Гистограмма
4. Описательная статистика
5. Скользящее среднее
6. Экспоненциальное сглаживание
7. Ковариационный анализ
8. Корреляционный анализ
9. Двухвыборочный F-тест для дисперсий
10. Двухвыборочный z-тест для средних
11. Парный двухвыборочный t-тест для средних
12. Двухвыборочный t-тест с одинаковыми дисперсиями
13. Двухвыборочный t-тест с разными дисперсиями
14. Дисперсионный анализ
15. Регрессия
16. Ранг и перцентиль
17. Анализ Фурье



Генерация случайных чисел



Окно инструмента Генерация случайных чисел содержит следующие основные параметры:

-Число переменных При помощи этого параметра можно получать многомерную выборку (количество столбцов)

-Число случайных чисел Определяется число точек данных (число реализаций), которое вы хотите генерировать для каждой переменной

-Случайное рассеивание Вводится произвольное значение, для которого необходимо генерировать случайные числа. Применяется для повторной генерации (повторное получение той же совокупности)

Выборка

В пакете Анализ данных инструмент Выборка используется для создания выборки из генеральной совокупности, рассматривая входной диапазон как генеральную совокупность

Выборка

Входные данные

Входной интервал:

Метки

Метод выборки

Периодический

Период:

Случайный

Число выборок:

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

OK

Отмена

Справка

Гистограмма

Гистограмма [?] [X]

Входные данные

Входной интервал:

Интервал карманов:

Метки

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

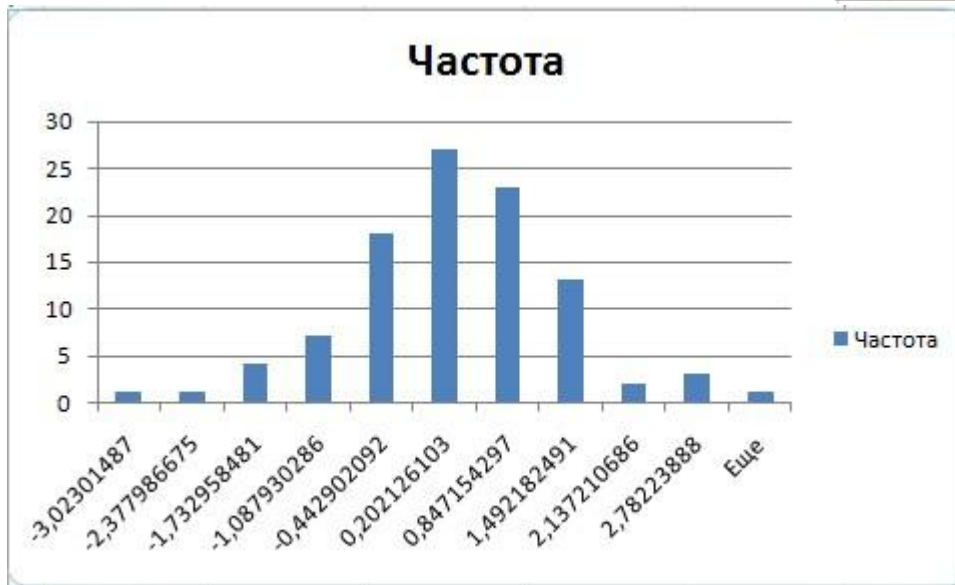
Парето (отсортированная гистограмма)

Интегральный процент

Вывод графика

OK
Отмена
Справка

I	J	Карман	Частота	интегральны	Карман	Частота	интегральны
-3,02301	1	-3,02301	1	1,00%	0,202126	27	27,00%
-2,37799	1	-2,37799	1	2,00%	0,847154	23	50,00%
-1,73296	4	-1,73296	4	6,00%	-0,4429	18	68,00%
-1,08793	7	-1,08793	7	13,00%	1,492182	13	81,00%
-0,4429	18	-0,4429	18	31,00%	-1,08793	7	88,00%
0,202126	27	0,202126	27	58,00%	-1,73296	4	92,00%
0,847154	23	0,847154	23	81,00%	2,782239	3	95,00%
1,492182	13	1,492182	13	94,00%	2,137211	2	97,00%
2,137211	2	2,137211	2	96,00%	-3,02301	1	98,00%
2,782239	3	2,782239	3	99,00%	-2,37799	1	99,00%
Еще	1	Еще	1	100,00%	Еще	1	100,00%



Гистограмма применяется для графического изображения интервального вариационного ряда

Описательная статистика

Описательная статистика

Входные данные

Входной интервал: \$A\$7:\$A\$100

Группирование: по столбцам по строкам

Метки в первой строке

Параметры вывода

Выходной интервал: \$E\$15

Новый рабочий лист:

Новая рабочая книга

Итоговая статистика

Уровень надежности: 95 %

К-ый наименьший: 1

К-ый наибольший: 1

OK
Отмена
Справка

Столбец1	
Среднее	-0,00026
Стандарт	0,111611
Медиана	0,022991
Мода	#Н/Д
Стандарт	1,082113
Дисперси	1,170968
Экссесс	0,844429
Асимметр	0,194017
Интервал	6,026676
Минимум	-2,59941
Максимум	3,427267
Сумма	-0,02426
Счет	94

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$s_x = \frac{\sigma}{\sqrt{n}}$$

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

$$R = x_{\max} - x_{\min}$$

Описательная статистика использует совокупность методов, позволяющих делать научно обоснованные выводы о числовых параметрах распределения генеральной совокупности по случайной выборке из нее