

Статистический анализ данных в MS Excel

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

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

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	16,88105
10	19,54965
11	

СЧЕТ

fx	=СЧЁТ(A1:A10)	
D	E	F
	10	

МАКС

fx	=МАКС(A1:A10)	
F	F	
27,36091		

ПРОЦЕНТРАНГ

fx	=ПРОЦЕНТРАНГ(A1:A10:A8)		
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)		
E	F	G	
7,991351			

РАНГ

fx	=РАНГ(A4;A1:A10)	
E	F	
6		

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

Предварительная обработка данных

Массив данных СЧЕТ МАКС

ЧАСТОТА 0 -5 5 -9 9 -15 15 -20 20 -25

25 -30 1 5 0 2 1 1 ПРОЦЕНТРАНГ

ПЕРСЕНТИЛЬ РАНГ

Определение характеристик положения

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

=СРЗНАЧ(A1:A10)	
E	F

12,32733

СРГЕОМ

=СРГЕОМ(A1:A10)	
E	F

10,19584

МЕДИАНА

=МЕДИАНА(A1:A10)	
E	F

7,991351

МОДА

E2

F2

=МОДА(A1:B10)

A	B	C	D	E	F
5,978508	2				
3,902837	5				
27,36091	6				
7,631094	3				
8,351607	3				
7,423104	3				
20,14716	3				
6,044401	3				
16,88405	3				
19,54965	3				

3

КВАРТИЛЬ

=КВАРТИЛЬ(A1:A10;3)		
F	F	G

18,88325

Определение характеристик рассеивания

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	16,88105
10	19,54965
11	

СРОТКЛ

f_x =СРОТКЛ(A1:A10)

E

F

6,926489

КВАДРОТКЛ

f_x =КВАДРОТКЛ(A1:A10)

E

F

G

572,7512

ДИСП

f_x =ДИСП(A1:A10)

E

F

63,63902

СТАНДОТКЛОН

f_x =СТАНДОТКЛОН(A1:A10)

E

F

G

7,977107

СКОС

f_x =СКОС(A1:A10)

E

F

0,810077

ЭКСЦЕСС

f_x =ЭКСЦЕСС(A1:A10)

E

F

-0,70134

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 (Z ТЕСТ)
4. Определение значений функций распределения отличных от нормальных (ЛОГНОРМРАСП, СТЬЮДРАСП...)
5. Проверка гипотезы о равенстве дисперсий (ФТЕСТ)

Определение параметров распределения непрерывных случайных величин

НОРМРАСП

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

НОРМОБР

Вероятность	0,326355	=	1,076757
Среднее	10	=	10
Стандартное откл	20	=	20
		=	1,98958775

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

Массив	A1:A10	=	-5,978508+2797197;3,9028368221...
X	7	=	7
Сигма	1,5	=	6,0
		=	0,104774175

ФТЕСТ

Массив1	A1:A10	=	-5,978508+2797197;3,9028368221...
Массив2	B1:B10	=	-0,997678+0076116;-2,7768316015
		=	0,165776115

Надстройка «Анализ данных»

В состав Microsoft Excel входит набор средств анализа данных (так называемый **пакет анализа**), предназначенный для решения сложных статистических и инженерных задач. Для анализа данных с помощью этих инструментов следует указать входные данные и выбрать параметры.

Анализ будет выполнен с помощью подходящей статистической или инженерной макрофункции, а результат будет помещен в выходной диапазон.

Команда «Данные» - «Анализ данных». Если этой команды нет в меню, необходимо загрузить надстройку **Пакет анализа** (аналогично надстройке «Поиск решения»).



Спасибо за внимание:

Бабенко О Н
