

15 Группировка данных и гистограммы

Как правило, в обширных массивах статистических данных совпадающие значения встречаются редко. Чаще встречаются близкие значения, которые разумно не различать. Чтобы понять, насколько плотно распределены значения на каждом участке числовой прямой, применяют **группировку данных**.

Чтобы сгруппировать данные, нужно разбить числовую прямую на одинаковые промежутки — **интервалы группировки**. Длина интервала называется **шагом группировки**. Затем нужно подсчитать долю значений в каждом интервале. Получаются **частоты значений в интервалах**. Осталось построить диаграмму по полученным данным.

ПРИМЕР 1. В таблице 41 приведены результаты ежедневных измерений атмосферного давления в Москве летом 2019 г.

¹ Криптограф — специалист по шифрованию, расшифровке и защите информации.

Таблица 41. Атмосферное давление летом 2019 г. в Москве

749,4	753,0	746,1	756,8	740,2	738,2	740,3	746,0	740,3	750,3	751,6	756,7
747,3	752,8	745,8	751,7	743,3	742,4	744,6	741,2	736,5	745,8	754,5	754,8
750,8	752,3	748,7	739,3	740,8	743,5	749,1	742,2	740,1	746,5	754,3	752,2
755,2	751,2	749,6	730,5	735,4	745,8	748,3	739,7	744,2	746,2	758,1	752,1
753,7	755,9	749,2	736,8	738,1	745,3	744,7	736,9	742,3	747,1	753,0	
754,1	749,4	749,1	741,3	738,4	743,9	745,1	742,9	735,4	746,8	757,0	
754,3	747,5	752,4	738,9	735,7	742,6	748,3	741,0	741,8	743,4	755,8	
753,8	748,4	753,7	739,8	735,4	740,1	748,9	738,9	746,6	750,1	751,9	

Всего значений 92, повторяющихся среди них мало. Наименьшее значение 730,5, наибольшее — 758,1 мм рт. ст. Шаг группировки выберем 4 мм рт. ст., а первый интервал возьмём 726—730 мм рт. ст. В него не попадает ни одно значение. Последний интервал тоже сделаем пустым.

Если значение попадает на границу двух интервалов, можно отнести его к любому из них — это вопрос договорённости. Мы относим граничное значение к левому интервалу. Например, значение 742,0 мы включили в интервал 738—742. Результаты соберём в таблицу 42.

Получилось 10 интервалов. Самые «населённые» — интервалы от четвёртого до седьмого.

Частоты находятся в правом столбце таблицы. Их выражают в долях единицы. Для частот попадания в интервал верно то же свойство, которое мы формулировали и доказывали для отдельных значений: сумма частот равна единице.

Таблица 42. Группированные данные, шаг 4 мм рт. ст.

№ п/п	Интервал, мм рт. ст.	Количество попаданий в интервал	Частота значения в интервале
1	726—730	0	0
2	730—734	1	0,011
3	734—738	7	0,076
4	738—742	18	0,196
5	742—746	18	0,196
6	746—750	19	0,207
7	750—754	17	0,185
8	754—758	11	0,120
9	758—762	1	0,011
10	762—766	0	0
	Всего	92	1,002

Чтобы с помощью электронной таблицы узнать, сколько всего значений в массиве данных, используйте функцию

СЧЁТ()

С	D	E	F
Данные			
1,4	1,3		12
1,8	2,6		
2,3	3,1		
3,5	2,7		
2,1	1,8		
1,3	2,6		



Задачи

103 В таблице 43 приведены данные о населении всех городов Московской области (данные на 2019 г.). Рассмотрите таблицу.

Таблица 43. Население городов Московской области

№ п/п	Город	Население, тыс. чел.	№ п/п	Город	Население, тыс. чел.
1	Апрелевка	29,1	31	Краснознаменск	42,5
2	Балашиха	490,0	32	Кубинка	20,1
3	Бронницы	22,6	33	Куровское	20,6
4	Верея	5,0	34	Ликино-Дулёво	29,0
5	Видное	71,4	35	Лобня	89,3
6	Волоколамск	19,2	36	Лосино-Петровский	25,3
7	Воскресенск	93,2	37	Луховицы	30,3
8	Высоковск	10,4	38	Лыткарино	58,6
9	Голицыно	17,4	39	Люберцы	207,3
10	Дедовск	30,1	40	Можайск	30,0
11	Дзержинский	56,3	41	Мытища	222,7
12	Дмитров	68,7	42	Наро-Фоминск	63,0
13	Долгопрудный	112,0	43	Ногинск	102,3
14	Домодедово	133,5	44	Одинцово	137,5
15	Дрезна	11,3	45	Озёры	25,0
16	Дубна	75,0	46	Орехово-Зуево	118,0
17	Егорьевск	72,1	47	Павловский Посад	63,9
18	Жуковский	107,9	48	Пересвет	13,6
19	Зарайск	22,9	49	Подольск	304,2
20	Звенигород	21,9	50	Протвино	35,8
21	Ивантеевка	79,3	51	Пушкино	106,8
22	Истра	33,5	52	Пушино	20,8
23	Кашира	48,0	53	Раменское	119,9
24	Клин	79,4	54	Реутов	107,0
25	Коломна	141,1	55	Рошаль	20,2
26	Королев	224,5	56	Руза	13,0
27	Котельники	46,8	57	Сергиев Посад	102,0
28	Красноармейск	26,5	58	Серпухов	124,9
29	Красногорск	171,8	59	Солнечногорск	51,6
30	Краснознаменск	12,9	60	Старая Купавна	22,5

Смотри ниже

№ п/п	Город	Население, тыс. чел.	№ п/п	Город	Население, тыс. чел.
61	Ступино	56,0	68	Истрга	32,4
62	Талдом	12,7	69	Щёлково	124,8
63	Фрязино	60,0	70	Электрогорск	22,8
64	Химки	254,7	71	Электросталь	157,4
65	Хотьково	21,2	72	Электроугли	20,7
66	Черноголовка	21,1	73	Яхрома	14,3
67	Чехов	71,9			

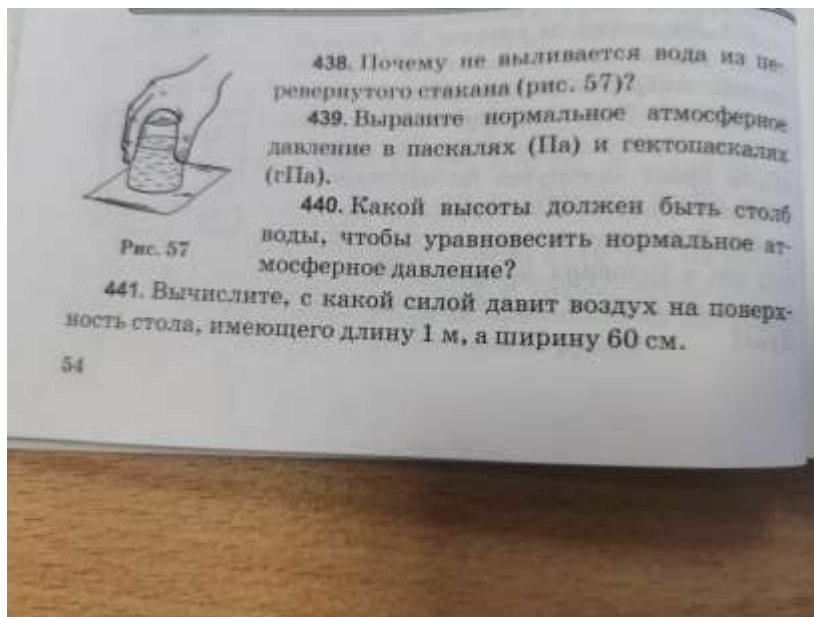
- а) Укажите в таблице три подмосковных города, где численность населения, по вашему мнению, значительно отличается от численности населения большинства городов (значительно больше или значительно меньше).
- б) Удалите из таблицы три найденных в пункте а) города. Найдите наименьшее и наибольшее значения после удаления этих трёх городов.

104 После удаления из таблицы 43 городов Балашихи, Подольска и Верея нужно сгруппировать оставшиеся данные.



- а) Сколько интервалов группировки получится, если сделать начальное значение 10 тыс. чел. при шаге группировки 10 тыс. чел.?
- б) Сколько интервалов группировки получится, если сделать начальное значение 0 тыс. чел., а шаг группировки 25 тыс. чел.?
- в) Какой из этих двух способов группировки вы бы предпочли?
- г) Сколько городов из таблицы 43 попадает в интервал 90—120 тыс. чел.?
- д) Какова частота попадания в интервал 90—120 тыс. чел.?

Физика.



ДЗ по русскому языку 24.01

П. 39 С.106 упражнение 255 (составить устный рассказ по опорным словам). Выполнить задание по данной теме на Учи. ру (23.01 - 24.01)