

Основы статистического описания

Дано: 200 чисел в таблице 1.

Таблица 1 - изначальные данные

1,3477	0,2103	0,1986	0,6075	0,1855	1,2369	0,6342	0,6085
0,1149	0,0589	2,4146	0,3616	0,0507	0,4288	0,5271	0,0256
0,0122	0,0372	3,1630	0,0052	0,1318	0,2368	0,3951	0,1228
1,2179	1,1764	0,7849	0,0127	1,1532	0,6347	0,3706	1,4618
0,6058	2,2735	0,9660	2,5070	1,2873	0,2850	0,4288	1,5013
0,6288	0,3467	0,0852	0,3289	0,8315	0,5832	0,9193	0,2175
0,7987	0,3619	0,9346	1,1240	2,0330	0,4499	0,8190	0,3876
1,5254	0,0460	0,2162	0,0940	0,3219	0,0553	0,5683	0,7826
0,1124	0,4599	0,8962	0,4216	0,5704	0,5462	0,0617	0,3739
1,1240	0,2457	0,8243	0,7718	2,2571	3,1750	0,1205	0,1577
0,2793	0,2657	0,3568	0,3157	0,5193	0,0412	0,0783	0,5974
0,3485	0,3693	0,1615	4,1198	0,6488	0,6333	0,3160	2,4637
0,3243	0,6258	1,0332	2,1218	0,5510	1,0726	0,2653	0,8948
0,5117	0,5252	0,4743	0,4009	0,4780	0,2046	1,9297	0,2017
0,0243	0,2475	0,4533	1,5822	1,0943	0,1892	0,0776	0,6677
0,0492	0,7961	0,7037	1,6622	0,6615	0,3959	0,1086	0,4650
2,1563	0,9641	0,0725	0,0578	0,1003	0,7716	0,5468	0,7078
0,9448	0,1065	0,8215	0,0667	2,3079	2,9904	0,2937	0,1147
0,2728	0,9434	0,4326	0,2749	2,9410	0,2141	0,1783	0,2735
0,1445	0,0090	0,1459	0,1555	0,4759	1,5617	0,8086	0,4629
0,0746	4,9853	1,2168	1,0016	0,8417	1,1090	0,2977	0,9793
0,9855	2,6596	0,6743	0,4680	0,0179	1,9120	0,4759	0,4485
0,8310	1,2612	0,1081	2,7027	1,4983	0,1689	0,4115	0,0489
0,5970	0,2522	1,3052	0,4892	0,1613	0,0472	2,6204	0,4666
0,4302	0,7117	0,1903	1,5936	1,1317	0,1480	0,9583	0,6624

Найти: выборочное математическое ожидание, дисперсию, коэффициент асимметрии и эксцесс, выборочную медиану, построить гистограмму.

Решение:

Сначала произведем упорядочение выборки по возрастанию. Минимальный элемент выборки равен 0,0052, максимальный элемент равен 4,9853. Для частей.

$$L = \frac{5,0000 - 0,0000}{20} = 0,2500.$$

Посчитав количество чисел, входящих в каждый интервал, получаем данные из таблицы 2.

Таблица 2 - данные, разбитые по интервалам

№	Интервалы до объединения	Количество данных в интервале	Интервалы после объединения	Количество данных в интервале
1	(0,000-0,250)	57	(0,000-0,250)	57
2	(0,250-0,500)	46	(0,250-0,500)	46
3	(0,500-0,750)	28	(0,500-0,750)	28
4	(0,750-1,000)	24	(0,750-1,000)	24
5	(1,000-1,250)	13	(1,000-1,250)	13
6	(1,250-1,500)	6	(1,250-1,750)	12
7	(1,500-1,750)	6		
8	(1,750-2,000)	2	(1,750-2,500)	10
9	(2,000-2,250)	3		
10	(2,250-2,500)	5		
11	(2,500-2,750)	4	(2,500-3,250)	8
12	(2,750-3,000)	2		
13	(3,000-3,250)	2		
14	(3,250-3,500)	0	Ошибка	исключен
15	(3,500-3,750)	0		
16	(3,750-4,000)	0		
17	(4,000-4,250)	1		
18	(4,250-4,500)	0		
19	(4,500-4,750)	0		
20	(4,750-5,000)	1		

После сортировки видно, что значения 4,1198 и 4,9853 находятся в хвосте распределения и отделены от основной части выборки пустыми интервалами. Поэтому рассматриваем их как грубые ошибки опыта и в дальнейших расчетах не участвуют. Объем выборки после исключения грубых ошибок: $n = 198$.

Далее производим расчеты: частоты попадания выборочных значений в i -й интервал, середины интервалов, эмпирической функции и промежуточных величин для среднего значения, дисперсии, коэффициента асимметрии и эксцесса.

Таблица 3 - результаты расчетов и промежуточные расчеты

p_i^*	$\tilde{x}_i$	$F^*(\tilde{x}_i)$	$\tilde{x}_i p_i^*$	$\tilde{x}_i^2 p_i^*$	$\tilde{x}_i^3 p_i^*$	$\tilde{x}_i^4 p_i^*$
0,287879	0,1250	0,000000	0,035985	0,004498	0,000562	0,000070
0,232323	0,3750	0,287879	0,087121	0,032670	0,012251	0,004594
0,141414	0,6250	0,520202	0,088384	0,055240	0,034525	0,021578
0,121212	0,8750	0,661616	0,106061	0,092803	0,081203	0,071052
0,065657	1,1250	0,782828	0,073864	0,083097	0,093484	0,105169
0,060606	1,5000	0,848485	0,090909	0,136364	0,204545	0,306818
0,050505	2,1250	0,909091	0,107323	0,228062	0,484631	1,029842
0,040404	2,8750	0,959596	0,116162	0,333965	0,960148	2,760427

Исходя из результатов расчетов, приведенных в таблице 3, получаем:

$$d_1 = \sum_{i=1}^{m_1} \tilde{x}_i p_i^* = 0,705808, d_2 = \sum_{i=1}^{m_1} \tilde{x}_i^2 p_i^* = 0,966698,$$

$$d_3 = \sum_{i=1}^{m_1} \tilde{x}_i^3 p_i^* = 1,871350, d_4 = \sum_{i=1}^{m_1} \tilde{x}_i^4 p_i^* = 4,299551.$$

Выборочные центральные моменты третьего и четвертого порядка:

$$m_3 = d_3 - 3\bar{x}d_2 + 2\bar{x}^3 = 0,527658.$$

$$m_4 = d_4 - 4\bar{x}d_3 + 6\bar{x}^2d_2 - 3\bar{x}^4 = 1,161241.$$

Среднее значение:

$$\bar{x} = d_1 = 0,705808.$$

Выборочная дисперсия:

$$S^2 = d_2 - \bar{x}^2 = 0,468533.$$

Выборочный коэффициент асимметрии:

$$A^* = \frac{m_3}{S^3} = 1,645288.$$

Экцесс:

$$E^* = \frac{m_4}{S^4} - 3 = 2,289827.$$

Так как после исключения грубых ошибок объем выборки четный, выборочная медиана равна:

$$M_e^* = \frac{x_{(99)} + x_{(100)}}{2} = 0,4751.$$

На основании таблицы 2 построена гистограмма распределения чисел по интервалам.

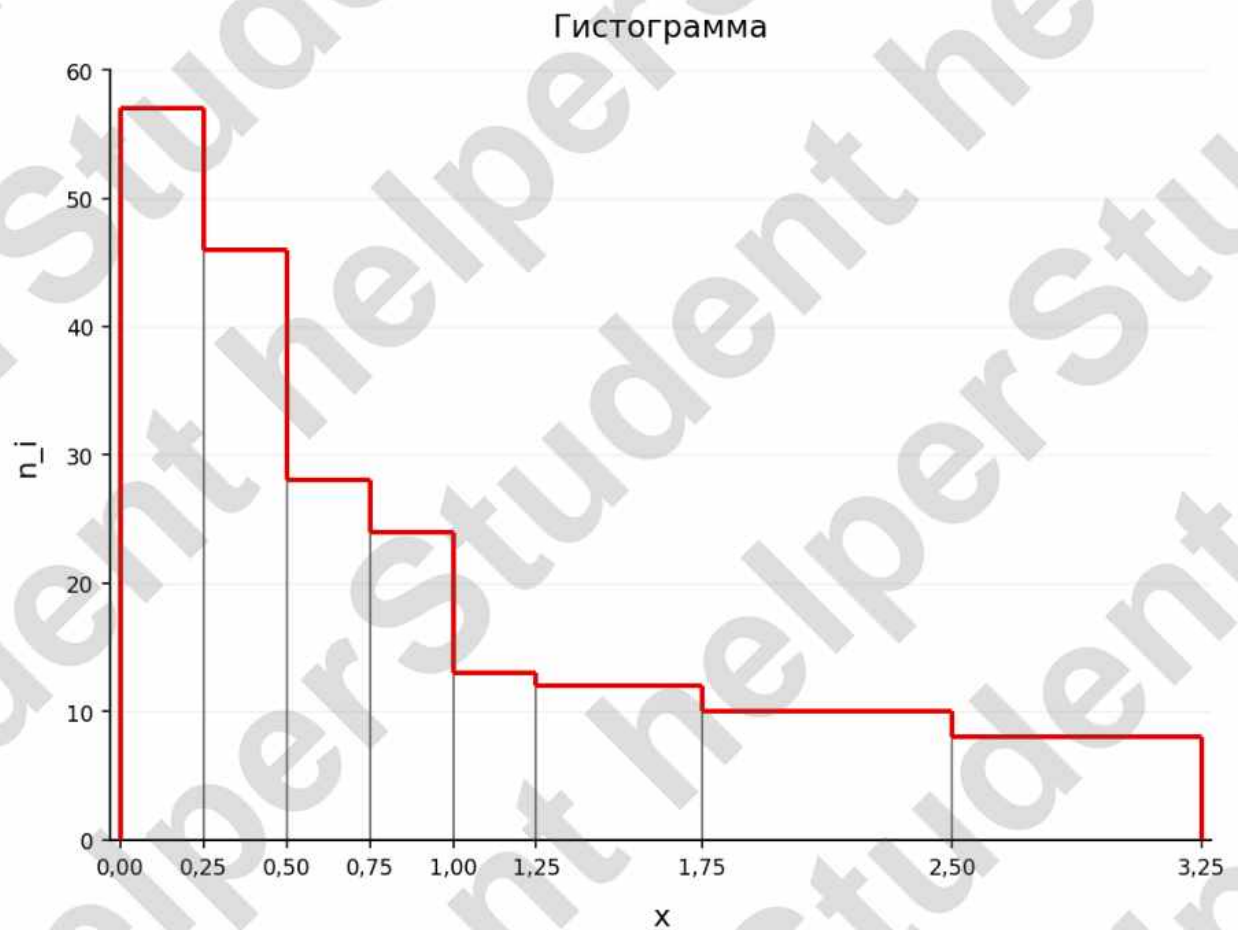


Рисунок 1 - гистограмма распределения чисел по интервалам.