

ЭЛЕМЕНТЫ СТАТИСТИКИ

Сбор и группировка
статистических данных

Задача

На выборах мэра города будут баллотироваться три кандидата: Алексеев, Иванов, Карпов. Проводя опрос 50 избирателей, выяснили, за кого из кандидатов они собираются голосовать. Получили следующие данные:
И,А,И,И,К,К,И,И,И,А,К,А,А,А,К,К,И,К,А,А,И,К,И,И,К,И,А,И,И,
И,А,И,И,К,И,А,И,К,К,И,К,А,И,И,И,А,А,А,К,И.

Представьте эти данные в виде таблицы частот.
Достаточно ли этих данных, чтобы сделать вывод о предстоящих результатах голосования?

Сбор и группировка статистических данных

Рассмотрим специальные статистические исследования. Всякое статистическое исследование начинается с целенаправленного сбора информации об изучаемом явлении или процессе. Это – *этап статистического наблюдения*. Для обобщения и систематизации данных, полученных в результате статистического наблюдения, их по какому – либо признаку разбивают на группы и результаты группировки сводят в таблицы.

ПРИМЕР

Администрация школы решила проверить математическую подготовку восьмиклассников. С этой целью был составлен тест, содержащий 9 заданий. Работу выполняли 40 учащихся школы. При проверке каждой работы учитель отмечал число верно выполненных заданий. В результате был составлен такой ряд чисел:
6,5,4,0,4,5,7,9,1,6,8,7,9,5,8,7,6,2,5,7,6,3,4,4,5,6,8,6,
7,7,4,3,5,9,6,7,8,6,6,9,8.

Для того, чтобы удобно было анализировать полученные данные, упорядочим этот ряд:

0, 1, 2, 3, 3, 4, 4, 4, 4, 4,
5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6,
7, 7, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 9, 9, 9, 9.

Представим полученные данные в виде таблицы, в которой для каждого числа верно выполненных заданий, записанного в верхней строке, укажем в нижней строке количество появлений этого числа в ряду, т.е. частоту

4

Таблица частот

Число верно выполненных заданий	0	1	2	3	4	5	6	7	8	9
Частота	1	1	1	2	5	6	8	7	5	4

Результаты проведённой проверки работ

Сумма частот равна общему числу проверяемых работ, т.е. 40.

Найдём среднее арифметическое

$$\frac{0 \cdot 1 + 1 \cdot 1 + 2 \cdot 1 + 3 \cdot 2 + 4 \cdot 5 + 5 \cdot 6 + 6 \cdot 8 + 7 \cdot 7 + 8 \cdot 5 + 9 \cdot 4}{40}$$

Среднее арифметическое равно 5,8.

Значит, в среднем учащиеся выполнили по 5,8 заданий, т.е. примерно две трети общего объёма работы.

Наибольшее число верно выполненных учащимися заданий равно 9, а наименьшее равно 0. Значит, размах рассматриваемого ряда данных равен

$$9 - 0 = 9$$

Из таблицы видно, что чаще всего встречаются работы, в которых выполнено 6 заданий, т.е. мода ряда равна 6.

МЕДИАНА РЯДА

Найдём медиану ряда. Т.к. в ряду всего 40 чисел, то медиана равна среднему арифметическому 20 – го и 21 – го членов соответствующего упорядоченного ряда. Для того чтобы определить, в какие группы попадают эти члены, нужно последовательно суммировать частоты и сравнивать с числами 20 и 21.

$$1+1+1+2+5+6=16$$

$$1+1+1+2+5+6 +8=24$$

20-й и 21-ый члены ряда попадают в ту группу, которую составляют учащиеся, верно выполнившие 6 заданий. Значит, медиана ряда равна $(6+6):2=6$.

Таблица относительных частот

10

Число верно выполненных заданий	0	1	2	3	4	5	6	7	8	9
Относительная частота	2,5	2,5	2,5	5	12,5	15	20	17,5	12,5	10

Если в ряду имеется большое число данных и одинаковые значения встречаются редко, то для анализа данных строят интервальный ряд. Для этого разность между наибольшим и наименьшим значениями делят на несколько равных частей и, округляя полученный результат, определяют длину интервала. За начало первого интервала выбирают наименьшее данное или ближайшее к нему целое число, его не превосходящее.

ПРИМЕР

Пусть, на партии из 50 электроламп изучали продолжительность их горения (в часах). По результатам составили таблицу:

Продолжительность горения, ч	Частота
До 200	1
200-400	3
400-600	5
600-800	9
800-1000	16
1000-1200	9
1200-1400	5
1400-1600	2

Найдём среднюю продолжительность горения.
Составим новую таблицу частот, заменив каждый интервал числом, которое является его серединой

Продолжительность горения, ч	Частота
100	1
300	3
500	5
700	9
900	16
1100	9
1300	5
1500	2

Найдём ряд данных

$$(100 \cdot 1 + 300 \cdot 3 + 500 \cdot 5 + 700 \cdot 9 + 900 \cdot 16 + 1100 \cdot 9 + 1300 \cdot 5 + 1500 \cdot 2) : 50 \approx 870$$

Значит, средняя продолжительность горения электроламп приближённо равна 870 часов.

В тех случаях, когда бывает сложно или даже невозможно провести сплошное исследование, его заменяют выборочным. При выборочном исследовании из всей изучаемой совокупности данных, называемой генеральной совокупностью, выбирается определённая её часть, т.е. составляется выборочная совокупность (выборка), которая подвергается исследованию. При этом выборка должна быть представительной, или, как говорят, репрезентативной, т.е. достаточной по объёму и отражающей характерные особенности исследуемой генеральной совокупности.