

Отклонения. Частота в массиве

Урок по вероятности и статистике

Сколько деталей изготовил каждый рабочий за 5 дней?

| День недели | Дневная выработка | |
|-------------|-------------------|------------------|
| | 1-й рабочий (X) | 2-й рабочий (Y) |
| | (кол-во деталей) | (кол-во деталей) |
| Понедельник | 52 | 61 |
| Вторник | 54 | 40 |
| Среда | 50 | 55 |
| Четверг | 48 | 50 |
| Пятница | 46 | 44 |

Решение

| День недели | Дневная выработка | |
|---------------|-------------------|------------------|
| | 1-й рабочий (X) | 2-й рабочий (Y) |
| | (кол-во деталей) | (кол-во деталей) |
| Понедельник | 52 | 61 |
| Вторник | 54 | 40 |
| Среда | 50 | 55 |
| Четверг | 48 | 50 |
| Пятница | 46 | 44 |
| Итого: | 250 | 250 |

Вывод: количество деталей одинаково.

Средняя производительность труда

- Рассчитаем, сколько деталей в день производил в среднем каждый рабочий (среднюю производительность труда).

Для этого найдём **среднее арифметическое** числовых наборов X и Y .

Вывод

| День недели | Дневная выработка | |
|-------------------------------|-------------------|------------------|
| | 1-й рабочий (X) | 2-й рабочий (Y) |
| | (кол-во деталей) | (кол-во деталей) |
| Понедельник | 52 | 61 |
| Вторник | 54 | 40 |
| Среда | 50 | 55 |
| Четверг | 48 | 50 |
| Пятница | 46 | 44 |
| Итого: | 250 | 250 |
| Среднее арифметическое | 50 | 50 |

Производительность труда за день у обоих рабочих тоже одинаковая.

Найдём медианы числовых наборов X и Y.

| День недели | Дневная выработка | |
|-------------------------------|-------------------|------------------|
| | 1-й рабочий (X) | 2-й рабочий (Y) |
| | (кол-во деталей) | (кол-во деталей) |
| Понедельник | 52 | 61 |
| Вторник | 54 | 40 |
| Среда | 50 | 55 |
| Четверг | 48 | 50 |
| Пятница | 46 | 44 |
| Итого: | 250 | 250 |
| Среднее арифметическое | 50 | 50 |

Найдём медианы числовых наборов X и Y.

| День недели | Дневная выработка | |
|-------------------------------|-------------------|------------------|
| | 1-й рабочий (X) | 2-й рабочий (Y) |
| | (кол-во деталей) | (кол-во деталей) |
| Понедельник | 52 | 61 |
| Вторник | 54 | 40 |
| Среда | 50 | 55 |
| Четверг | 48 | 50 |
| Пятница | 46 | 44 |
| Итого: | 250 | 250 |
| Среднее арифметическое | 50 | 50 |
| Медиана | 50 | 50 |

Медианы тоже получились одинаковые.

На данном примере мы увидели, что **с помощью средних характеристик сравнение выполнить не всегда возможно.**

Как поступить?

Отклонение

- В данном случае критерием сравнения может выступать **стабильность** работы токарей – у какого токаря количество произведённых им деталей в день менее отличается друг от друга, тот работает стабильнее.
- Если количество производимых в день деталей сильно разнится, то в какие-то дни токарь работает не в полную силу, производит меньше деталей, а в какие-то дни навёрстывает упущенное, а это всегда сказывается на качестве продукции.
- Стабильность можно оценивать с помощью **отклонений элементов числового набора от среднего значения** (отклонение – это разность между числом из данного набора и средним арифметическим этого набора)

Найдём суммы отклонений

| День недели | Дневная выработка | |
|-------------------------------|-------------------|------------------|
| | 1-й рабочий (X) | 2-й рабочий (Y) |
| | (кол-во деталей) | (кол-во деталей) |
| Понедельник | 52 | 61 |
| Вторник | 54 | 40 |
| Среда | 50 | 55 |
| Четверг | 48 | 50 |
| Пятница | 46 | 44 |
| Итого: | 250 | 250 |
| Среднее арифметическое | 50 | 50 |
| Медиана | 50 | 50 |

Сумма отклонений

- Логично предположить, что чем меньше будет разброс (отклонения от среднего значения) – тем **стабильнее** работает токарь.
- Но **когда набор чисел велик, рассматривать отклонения практически неудобно**, нужно описать разнообразие чисел в наборе одним числом.
- Попробуем найти сумму отклонений.
- В сумме получилось 0 (т.к. при вычислении “среднего разброса” часть отклонений входит в сумму со знаком “+”, часть со знаком “-” и в сумме всегда получается 0). Следовательно сумма отклонений не может нести информацию о разбросе.
- Какой же выход?

- Можно суммировать **квадраты отклонений** (они всегда неотрицательны).
- **Чем меньше сумма квадратов отклонений, тем меньше разброс чисел относительно среднего значения, тем более стабилен набор.**

Сумма квадратов отклонений

- Итак, рассчитаем сумму квадратов отклонений для нашего примера.

| День недели | Дневная выработка | |
|-------------------------------|-------------------|------------------|
| | 1-й рабочий (X) | 2-й рабочий (Y) |
| | (кол-во деталей) | (кол-во деталей) |
| Понедельник | 52 | 61 |
| Вторник | 54 | 40 |
| Среда | 50 | 55 |
| Четверг | 48 | 50 |
| Пятница | 46 | 44 |
| Итого: | 250 | 250 |
| Среднее арифметическое | 50 | 50 |
| Медиана | 50 | 50 |

Сумма квадратов отклонений

- Итак, рассчитаем сумму квадратов отклонений для нашего примера.

| День недели | Дневная выработка | |
|-----------------------------------|-------------------|------------------|
| | 1-й рабочий (X) | 2-й рабочий (Y) |
| | (кол-во деталей) | (кол-во деталей) |
| Понедельник | 52 | 61 |
| Вторник | 54 | 40 |
| Среда | 50 | 55 |
| Четверг | 48 | 50 |
| Пятница | 46 | 44 |
| Итого: | 250 | 250 |
| Среднее арифметическое | 50 | 50 |
| Медиана | 50 | 50 |
| Сумма квадратов отклонений | 40 | 282 |

Кто работает более стабильно?

- **Вывод:** первый токарь работает более стабильно, у него меньше сумма квадратов отклонений. Вероятно, работодатель предпочтёт взять на работу его.
- В данном примере рабочие работали **одинаковое** количество дней.

Дисперсия

Дисперсия — в статистике это "среднее квадратов отклонений от среднего". Чтобы её вычислить нужно:

1. Вычесть среднее значение из каждого числа
2. Возвести каждый результат в квадрат (так получатся квадраты разностей)
3. Найти среднее значение квадратов разностей

Сбор и группировка статистических данных

Частота

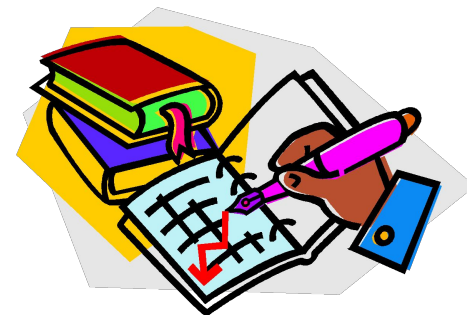




Сбор и группировка статистических данных



- Для изучения различных общественных и социально-экономических явлений, а также некоторых процессов, происходящих в природе, проводят **специальные статистические исследования**.
- Всякое статистическое исследование начинается с целенаправленного **сбора информации** об изучаемом явлении или процессе.
- Для обобщения и систематизации данных, полученных в результате наблюдения, их по какому-либо признаку **разбивают на группы и результаты группировки сводят в таблицы**.





Этапы обработки статистической информации:

Статистическое наблюдение.

Сбор данных:

По результатам теста, выполненного 40 учениками 9 класса, отмечено число верно решенных заданий:

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

Группировка собранных данных:

Для того, чтобы было удобно анализировать собранные данные, **упорядочим** и **сгруппируем** числовой 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 частот:

| | | | | | | | | | | |
|---------------------------------|---|---|---|---|---|---|---|---|---|---|
| Число верно решенных заданий | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Частота | 1 | 1 | 1 | 2 | 5 | 6 | 8 | 7 | 5 | 4 |

- Заметим, что сумма частот равна общему числу проверяемых работ, т.е. 40.
- Вообще, **если собранные данные представлены в виде таблицы частот, то сумма частот равна общему числу данных в ряду.**



Относительная частота

Иногда составляют таблицу, в которой для каждого данного указывается **относительная частота**.

Относительная частота - отношение частоты к общему числу данных в ряду, выраженное в процентах.

| | | | | | | | | | | | |
|------------------------------|-----|-----|-----|---|------|----|----|------|------|----|-------|
| Число верно решенных заданий | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | Сумма |
| Частота | 1 | 1 | 1 | 2 | 5 | 6 | 8 | 7 | 5 | 4 | 40 |
| Относительная частота, % | 2,5 | 2,5 | 2,5 | 5 | 12,5 | 15 | 20 | 17,5 | 12,5 | 10 | ? |



Какой всегда должна быть сумма относительных частот?

Анализ собранной информации

При анализе полученных данных используются различные статистические показатели – среднее арифметическое, размах, мода, медиана и т. д.

Проанализируем результаты проведенной проверки работ:

Используя составленную таблицу частот, найдите:

- Среднее арифметическое;
- Размах;
- Моду;
- Медиану.

Сделайте выводы



| | | | | | | | | | | |
|------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| Число верно решенных заданий | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Частота | 1 | 1 | 1 | 2 | 5 | 6 | 8 | 7 | 5 | 4 |





Проверьте себя:

- **Среднее арифметическое:**

$$(0*1+1*1+2*1+3*2+4*5+5*6+6*8+7*7+8*5+9*4)/40 = 232/40 = \mathbf{5,8}$$

Значит, в среднем учащиеся выполнили по 5,8 заданий, т.е. примерно 2/3 общего объема работы.

- **Размах:**

Наибольшее число верно выполненных заданий – 9, наименьшее – 0, значит размах ряда равен $9 - 0 = \mathbf{9}$, т.е. различие в числе верно выполненных заданий достаточно велико.

- **Мода:**

Из таблицы частот очевидно, что чаще всего встречаются работы, в которых верно выполнены 6 заданий, т.е. мода равна **6**.

- **Медиана:**

Т.к. в ряду 40 чисел, то медиана равна среднему арифметическому 20-го и 21-го чисел соответствующего упорядоченного ряда. Из таблицы частот следует, что на 20-ом и 21-ом местах в таком ряду будет число 6. Значит и медиана ряда равна **6**.

Домашнее задание

| День недели | Дневная выработка | |
|-------------|-------------------|------------------|
| | 1-й рабочий (X) | 2-й рабочий (Y) |
| | (кол-во деталей) | (кол-во деталей) |
| Понедельник | 53 | 60 |
| Вторник | 55 | 41 |
| Среда | 50 | 55 |
| Четверг | 47 | 48 |
| Пятница | 45 | 46 |

Рассчитаем, сколько деталей изготовил каждый из рабочих за 5 дней.

- Найти среднюю производительность труда
- Найти медианы числовых рядов X и Y
- найти квадраты отклонений

* Найти дисперсию