## **Введение**

В практических наблюдениях мы обычно имеем совокупность наблюдений х1, х2, ... , хn, на основе которых требуется сделать те или иные выводы. Часто этих наблюдений много, поэтому возникает задача их компактного описания. В идеале таким описанием могло бы быть утверждение, что х1, х2, ... , хn являются выборкой, т.е. независимыми реализациями случайной величины ξ с известным законом распределения F(x). Это позволило бы теоретически произвести расчеты всех необходимых исследователю характеристик наблюдаемого явления.

Однако далеко не всегда мы можем утверждать, что х1, х2, ... , хn являются независимыми и одинаково распределенными случайными величинами. Во-первых, это необходимо проверить, а во-вторых, часто заведомо известно, что это не так. Поэтому для компактного описания совокупности наблюдений используют другие методы – методы описательной статистики.

## **1.** **Методы описательной статистики**

Методами описательной статистики называются методы описания выборок х1, х2, ... , хn с помощью различных показателей и графиков. Достоинство методов описательной статистики в том, что ее простые и довольно информативные статистические показатели избавляют от необходимости просмотра большого количества значений выборки.

## 1 Показатели описательной статистики

Показатели, описывающие выборку можно разбить на несколько групп:

1. Показатели положения описывают положение данных (или середины совокупности) на числовой оси:

* Минимальный и максимальный элементы выборки
* Выборочные верхний и нижний квартили
* Среднее
* Выборочная медиана
* Выборочная мода

1. Показатели разброса описывают степень разброса данных относительно своего центра (насколько кучно основная масса данных группируется около середины совокупности)

* Дисперсия выборки
* Выборочное среднее квадратическое отклонение (стандартное отклонение)
* Размах
* Коэффициент эксцесса

1. Показатели асимметрии описывают симметричность распределения данных около своего центра

* Коэффициент асимметрии
* Положение выборочной медианы относительно выборочного среднего и относительно выборочных квартилей
* Гистограмма

1. Показатели, описывающие закон распределения, дают представление о законе распределения данных

* Гистограмма
* Выборочная функция распределения
* Таблица частот

Из перечисленных выше характеристик на практике по традиции чаще всего используют выборочные среднее, медиану и дисперсию (или стандартное отклонение). Однако для получения более точных и достоверных выводов необходимо использовать и другие показатели.

Особое внимание следует обратить на наличие в выборке выбросов – грубых, сильно отличающихся от основной массы, наблюдений. Большинство традиционных статистических методов весьма чувствительны к отклонениям от условий применимости метода. Поэтому выбросы могут не только исказить значение выборочных показателей, но и привести к ошибочным выводам. Подозрение о присутствии таких наблюдений должно возникнуть, если выборочная медиана сильно отличается от выборочного среднего, хотя в целом совокупность симметрична, или, если положение медианы сильно несимметрично относительно минимального и максимального элементов выборки. Проще всего обнаружить выбросы с помощью перехода от выборки к вариационному ряду или гистограмме с большим числом интервалов группировки.

## **2 Порядок выполнения работы**

### **2.1 Исходные данные**

Исходными данными является набор реализаций случайной величины (например, значения какой-либо величины, полученные при измерении). Размер выборки - n шт. Исходные данные оформить в виде таблицы (таблица 1).

Таблица 1 – Исходные данные

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Номер реализации | Значение | Номер реализации | Значение | Номер реализации | Значение | Номер реализации | Значение |
| 1... |  |  |  |  |  | ...n |  |

### **2.2 Построение вариационного ряда**

Для удобства работы с данными выборку преобразуют в вариационный ряд – ряд, в котором элементы выборки упорядочиваются по возрастанию.

Этапы выполнения:

1. Найти наименьший элемент ряда Xmin
2. Найти наибольший элемент ряда Xmax
3. Записать ряд, начиная с наименьшего элемента Xmin и заканчивая наибольшим Xmax (таблица 2)
4. Для упрощения процедуры обработки и с целью уменьшения ошибок при вычислениях необходимо вычесть из каждого элемента ряда постоянное число (например, округленное Xmin) и использовать в расчетах не сами размеры, а их отклонениями. Получившиеся отклонения записать в таблицу 2.

Таблица 2 – Вариационный ряд с отклонениями относительно x0 = <значение>[[1]](#footnote-1)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Номер элемента | Элемент | Отклонение | Номер элемента | Элемент | Отклонение |
| 1... |  |  | n |  |  |

### **2.3 Группировка данных**

Этапы выполнения:

1. Разбить весь диапазон R = Xmax – Xmin на r интервалов. Число интервалов r устанавливают в зависимости от числа наблюдений n:

|  |  |
| --- | --- |
| n | r |
| 40-100  100-500  5000-10000 | 7-9  8-12  10-16 |

При небольших выборках .



1. Назначить длину интервалов. Длину интервалов Δx чаще всего выбирают одинаковой: Δx = R/r. Ее округляют до значения, удобного для графического отображения.
2. Назначить нижнюю границу xн первого интервала (в отклонениях от x0). Она должна быть меньше xmin и удобной с позиции графического отображения. Результат занести в таблицу 3.
3. Назначить нижние xн и верхние xв границы всех оставшихся интервалов (в отклонениях от x0). Результаты занести в таблицу 3.
4. Определить число размеров, попадающих в интервал mi. Условие попадания размера xj в интервал xiн < xj ≤ xiв. Результаты занести в таблицу 3. Полученные результаты проверить по условию .



### **2.4 Определение частостей**

Отношение частоты mi к общему числу наблюдений n называется частостью:



Частость представляет собой эмпирическую оценку вероятности попадания результатов наблюдений Хj в i интервал.

Определить частости и результаты занести в таблицу 3.

Полученные результаты проверить по условию .



### **2.5 Определение эмпирической плотности вероятностей**

Эмпирическая плотность вероятностей равна:



Определить эмпирическую плотность вероятности, результаты занести в таблицу 3.

Таблица 3 – Расчетные данные

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Номера интервалов | Границы интервалов, <размерность> | | Частота, mi | Частость, | Эмпирическая плотность вероятности pi | Середина интервала xi |
| xн | xв |
| 1... |  |  |  |  |  |  |
|  | | |  |  |  | |

Для дальнейших геометрических построений необходимы значения середины интервалов xi. Определить их, результаты занести в таблицу 3.

### **2.6 Построение полигона**

Этапы выполнения[[2]](#footnote-2):

1. Определить масштабы по осям абсцисс и ординат, исходя их соотношения :R = 5 ≈ 8.



1. На оси абсцисс отложить интервалы значений измеряемой величины.
2. В серединах интервалов отметить ординаты, пропорциональные частостям.
3. Полученные точки соединить прямыми линиями.

Пример полигона приведен на рисунке 1.

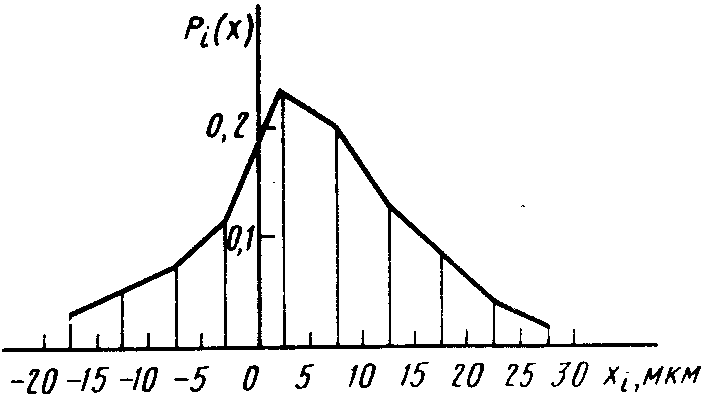


Рисунок 1 – Пример полигона

### **Построение гистограммы распределения**

Этапы выполнения:

1. Повторить пункты 1-2 из 2.5.
2. Над каждым интервалом по оси абсцисс построить прямоугольник, высота которого пропорциональна эмпирической плотности вероятностей.

Пример гистограммы распределения приведен на рисунке 2.

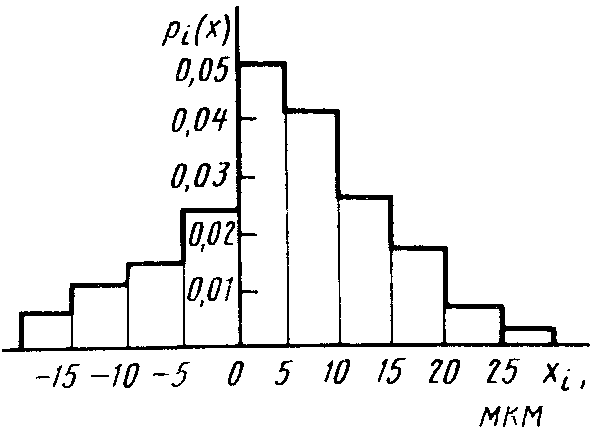


Рисунок 2 – Пример гистограммы распределения

### **2.7 Построение эмпирической функции распределения**

В середине каждого интервала по оси абсцисс ордината возрастает скачком на значение, соответствующее .



Этапы выполнения:

1. Повторить пункты 1-2 из 2.5.
2. В середине интервала 1 отметить скачок, равный . Провести горизонтальную линию от получившейся точки до середины следующего интервала.



1. В середине интервала 2 отметить скачок от горизонтальной линии, полученной в п.2, равный . Провести горизонтальную линию от получившейся точки до середины следующего интервала.



1. Повторить пункт 2 для остальных интервалов.

Значения для каждого интервала называют кумулятивной частостью, а сумму - кумулятивной частотой.



Пример гистограммы эмпирической функции распределения приведен на рисунке 3.

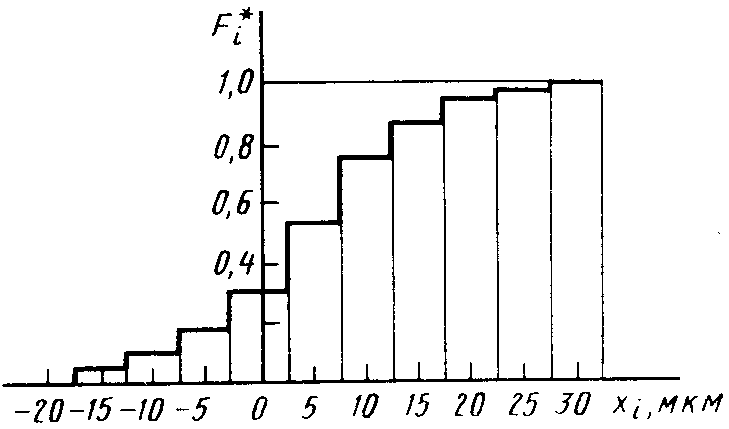


Рисунок 3 – Пример эмпирической функции распределения

### **2.8 Расчет параметров распределения**

С помощью гистограммы распределения можно рассчитать параметры распределения:

1. Для среднего арифметического



1. Для выборочной дисперсии



1. Для оценки центрального момента третьего порядка



1. Для оценки центрального момента четвертого порядка



Однако все расчеты можно значительно упростить, если все отклонения размеров yi выражать относительными величинами в долях ширины интервала Δx (целыми числами), а за начало отсчета отклонений принять условный нуль x0, равный середине интервала, имеющего наибольшую частоту mi:



Относительные начальные моменты в этом случае определяются:



Возвращаясь к размерностям измеряемой величины, получим:



Результаты расчета относительных начальных моментов удобнее всего свести в таблицу 4.

Таблица 4 – Расчетные данные моментов

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Номер интервалов | Середина интервала xi, <размерность> | yi |  |  |  | mi | miyi |  |  |  |
| 1  ... |  |  |  |  |  |  |  |  |  |  |
| Суммы | | | | | | |  |  |  |  |
| Начальные моменты | | | | | | |  |  |  |  |
| Обозначения | | | | | | | а1 | а2 | а3 | а4 |

Этапы выполнения:

1. Определить все произведения в таблице 4.
2. Определить все суммы в таблице 4.
3. Определить относительные начальные моменты в таблице 4.
4. Пересчитать полученные моменты для размерности измеряемой величины.
5. Рассчитать коэффициенты эксцесса и асимметрии.
6. Определить графически верхний и нижний квартили.
7. Определить графически выборочную медиану.
8. Определить выборочную моду (по вариационному ряду).
9. Определить размах.

### **2.9 Оформление результатов**

Полученные результаты свести в таблицу 5 (нечисловые результаты заменить ссылками на рисунки и приложения). На рисунках в приложениях нанести (если возможно) параметры.

Таблица 5 – Показатели описательной статистики для выборки n=< >

|  |  |
| --- | --- |
| Показатель | Значение |
| 1. Показатели положения  * Минимальный элемент выборки MIN * Максимальный элемент выборки MAX * выборочный верхний квартиль x0,75 * выборочный нижний квартиль x0,25 * среднее  * выборочная медиана Md * выборочная мода Mo  1. Показатели разброса  * дисперсия выборки S2 * выборочное среднее квадратическое отклонение S * размах R * коэффициент эксцесса kэкс |  |
| 1. Показатели асимметрии  * Коэффициент асимметрии kас * положение выборочной медианы относительно выборочного среднего и относительно выборочных квартилей * гистограмма | Приложение 2  Приложение 2 |
| 1. Показатели, описывающие закон распределения  * гистограмма * выборочная функция распределения * таблица частот | Приложение 2  Приложение 3  Таблица 3 |

### **Список использованной литературы**

1. Мелник М. Основы прикладной статистики: Пер. с англ. – М.: Энергоатомиздат, 1983. – 416 с.
2. Рудзит Я.А., Плуталов В.Н. Основы метрологии, точность и надежность в приборостроении: Учеб. пособие для студентов приборостроительных специальностей вузов. – М.: Машиностроение, 1991. – 304 с.
3. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере/ Под ред. В.Э.Фигурнова. – М.: ИНФРА-М, 1998. – 528 с.

1. - При оформлении работы выражения в угловых скобках заменить на конкретные значения [↑](#footnote-ref-1)
2. - Все графики выполнить на миллиметровке и оформить в виде приложений к отчету по практической работе [↑](#footnote-ref-2)