МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ

ВЫСШЕГО ОБРАЗОВАНИЯ

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ им. А.И. ГЕРЦЕНА»



Направление подготовки

09.03.01 – Информатика и вычислительная техника

Профиль «Технологии разработки программного обеспечения»

**Лабораторная работа №3**

**“Числовые характеристики вариационного ряда”**

|  | Работу выполнили:  Балаев Жамал,  Васильева Марина,  Иванов Никита,  Шардт Максим  Рожков Максим  очная форма обучения  курс: 2; группа: ИВТ-1.1 |
| --- | --- |
|  | Научный руководитель:  Профессор Власова Елена Зотиковна |

Санкт-Петербург

2022

**Лабораторная работа № 3**

“Вариационный ряд”

Выполнил Балаев Ж.Б. ИВТ 1.1

Цель работы: вычислить математические характеристики вариационного ряда

Инструменты: ПК, табличный процессор Excel.

Использованные формулы:

1. Средняя арифметическая, которая также называется выборочным средним

*для дискретного вариационного ряда:*

*для интервального ряда:*

1. Дисперсия

1. Среднее квадратическое отклонение

1. Коэффициент вариации

1. Коэффициент асимметрии

1. Эксцесс распределения

1. Медиана – значение признака ряда, относительно которого вариационный ряд делится на две равные по числу вариантов части. Это – 50-й перцентиль.

1. Мода – это значение признака, наиболее часто встречающееся в вариационном ряду

1. Перцентиль – это значение признака в определенной позиции ранжированного ряда, мера относительной позиции варианта в ряду.
2. Формула Стерджесса

,

1. Средняя арифметическая взвешенная

**Лабораторная работа №2**

Задание 1:

Вычислить математические характеристики вариационного ряда, который был построен при выполнении Задания №2 в Лабораторной работе №1.

Характеристики следующие:

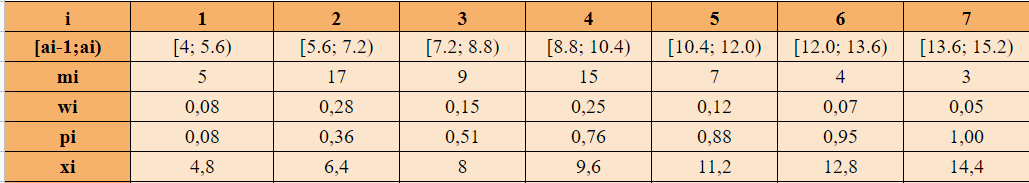
1. Среднее значение признака.
2. Дисперсия и среднее квадратическое отклонение.
3. Коэффициент вариации.
4. Определите, в каком промежутке колеблется плотность работников.
5. Коэффициент асимметрии.
6. Эксцесс

*Решение:*

1. *Среднее значение признака:*

В нашем случае n = 60.

Для удобства расчетов воспользуемся гугл таблицами:



*Табл.1. Исходная таблица*

Найдем для каждого случая:



И находим сумму этих ячеек:



Среднее значения признака **M = 8,736**

1. *Дисперсия и среднее квадратическое отклонение:*

Найдем для каждого случая:



Найдем сумму данных ячеек и умножим на :



Дисперсия **D = 6,518**

Мы взяли корень из найденного выше числа:



Среднее квадратическое отклонение **S = 2,553**

1. *Коэффициент вариации:*

Среднее квадратическое отклонение **S = 2,553**

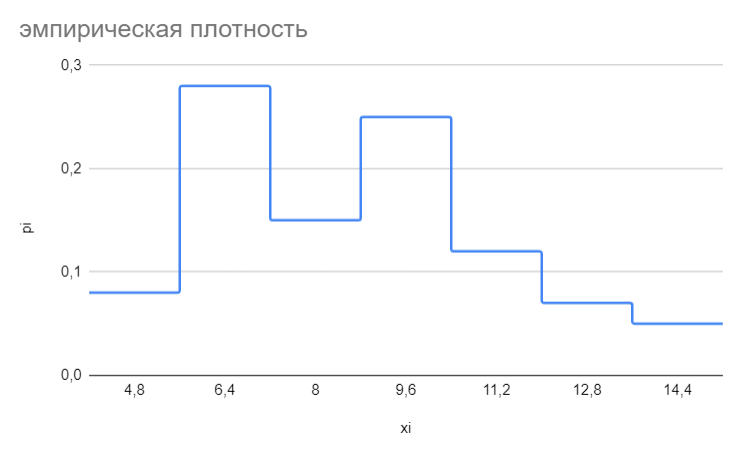
Среднее значения признака **M = 8,736**

Значит, можем найти С:



Коэффициент вариации **С = 29,225%**

1. *Определите, в каком промежутке колеблется плотность работников:*

**

*Рис.1. Эмпирическая плотность*

Исходя из графика, можно сказать, что плотность работников колеблется в интервале **[4;15]**

1. *Коэффициент асимметрии:*

Найдем для каждого случая:



Найдем сумму и разделим на n\*S^3 (где n=60, S = 2,553):



Коэффициент асимметрии **Λ = 0,4034**

1. *Эксцесс:*

Найдем для каждого случая:



Найдем сумму ячеек, разделим на n\*S^4 (где n=60, S = 2,553) и вычтем 3:



Эксцесс **Е=-0,586**

Задание 2:

Дайте ответы на следующие вопросы:

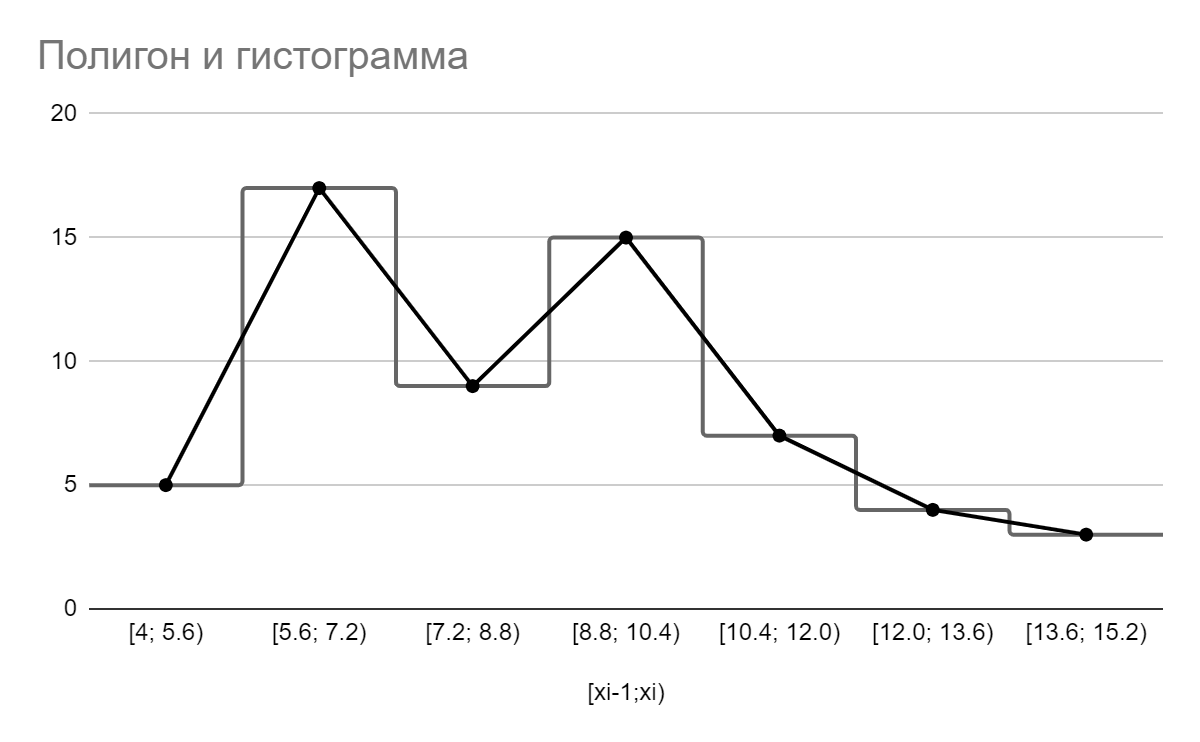
1. Имеются ли различия в обеспеченности хозяйств рабочей силой?

Эти различия несущественные или они весьма большие?

*Ответ:* Различия в обеспеченности хозяйств рабочей силой имеются, и они весьма большие. Рабочая сила распределена по участкам неравномерно. Об этом можно судить из того факта, что на большей части участков (41 участок) трудятся от 5 до 10 человек. В то время как на 14 участках трудятся по 10-15 человек и на 5 участках по 4-5 человек.

1. Является ли рассматриваемое распределение симметричным? Можно ли считать его нормальным?

*Ответ:* Рассматриваемое нами распределение не является симметричным. Оно имеет правостороннюю симметрию. Об этом говорит положительный коэффициент асимметрии(Λ = 0,4034**)**. Также это наглядно видно на гистограмме ряда, которая была построена в прошлой лабораторной работе:



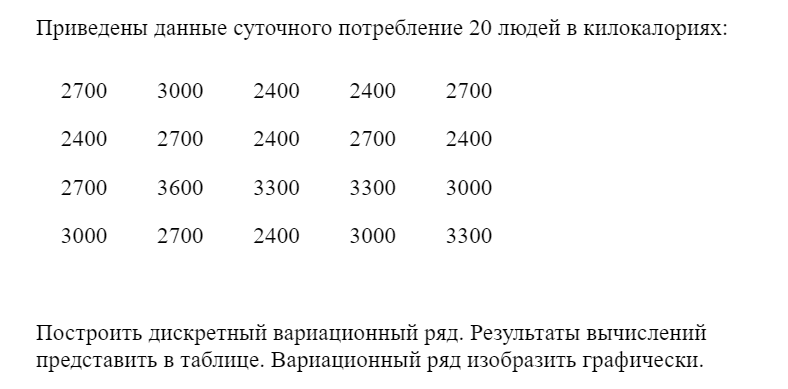
*Рис.2. Полигон и гистограмма*

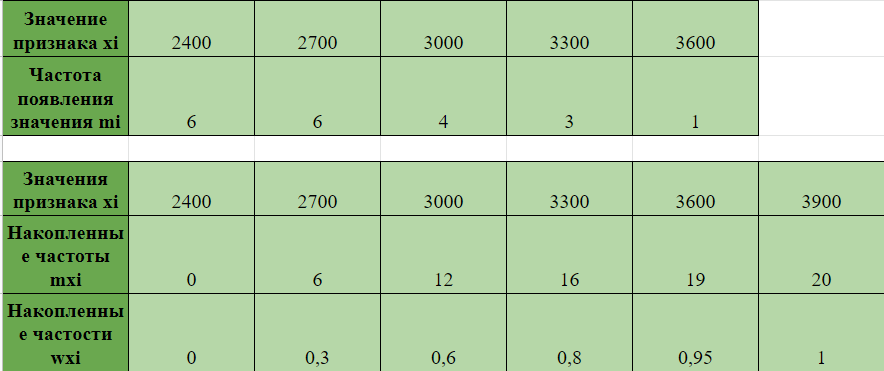
Данное распределение можно считать нормальным, так как известно, что для нормального распределения количество значений, отличающиеся от среднего на число, меньшее чем одно стандартное отклонение, составляет более 68,27 % выборки. В нашем случае, в интервал (6,186-11,286) попадает 71%. Из чего мы делаем вывод, что распределение является нормальным.

Задание 3:

Вычислить математические характеристики вариационного ряда, который был построен при выполнении Задания №3 в Лабораторной работе №1. Проведите анализ этого ряда.

*Решение:*

Напомним условие задачи:



*Табл.2. Исходная таблица*

1. *Среднее значение признака:*

Чтобы найти сумму по всем х, возьмем значения из таблицы и умножим на их количество:

Для нахождения среднего значения сложим полученные значения и разделим на n:



Среднее значения признака **M = 2805**

1. *Дисперсия и среднее квадратическое отклонение:*

Найдем для каждого случая:



Найдем сумму данных ячеек и умножим на :



Дисперсия **D = 128475**

Мы взяли корень из найденного выше числа:



Среднее квадратическое отклонение **S = 358,434**

1. *Коэффициент вариации:*

Среднее квадратическое отклонение **S = 358,434**

Среднее значения признака **M = 2805**

Значит, можем найти С:



Коэффициент вариации **С = 12,778%**

1. *Коэффициент асимметрии:*

Найдем для каждого случая:



Найдем сумму и разделим на n\*S^3 (где n=20, S = 358,434):



Коэффициент асимметрии **Λ = 0,533**

1. *Эксцесс:*

Найдем для каждого случая:



Найдем сумму ячеек, разделим на n\*S^4 (где n=20, S = 358,434) и вычтем 3:



Эксцесс **Е=-0,736**

**Анализ вариационного ряда на основе полученных характеристик:**

Коэффициент асимметрии положителен и больше 0, поэтому можно сделать вывод, что распределение имеет правостороннюю асимметрию. Значение эксцесса отрицательно, что позволяет сделать вывод о сглаженном распределение.

**Лабораторная работа №3**

Задание 1:

В качестве изучаемого признака рассматривается число продаж каждого из 26 случайно выбранных продавцов универмага:

16, 12, 15, 15, 23, 9, 15, 13, 14, 14, 21, 15, 14, 17, 27, 15, 16, 12, 16, 19, 14, 16, 17, 13, 14, 14.

Необходимо сделать следующее:

1. Построить вариационный ряд;
2. Провести анализ построенного вариационного ряда.

Для заданных данных был построен вариационный ряд и вычислены его характеристики:

| **i** | **[ai - 1; ai)** | **mi** | **wi** | **pi** | **xi** |
| --- | --- | --- | --- | --- | --- |
| **1** | [9; 11,8) | 1 | 0,04 | 0,04 | 10,4 |
| **2** | [11,8; 14,6) | 10 | 0,38 | 0,42 | 13,2 |
| **3** | [14,6; 17,4) | 11 | 0,42 | 0,85 | 16 |
| **4** | [17,4; 20,2) | 1 | 0,04 | 0,88 | 18,8 |
| **5** | [20,2; 23) | 1 | 0,04 | 0,92 | 21,6 |
| **6** | [23; 25,8) | 1 | 0,04 | 0,96 | 24,4 |
| **7** | [25,8; 28,6) | 1 | 0,04 | 1,00 | 27,2 |

Для простоты расчетов, сложные вычисления были разбиты на действия:

| **xi\*wi** | **(xi - М)^2** | **(xi - М)^3** | **(xi - М)^4** |
| --- | --- | --- | --- |
| 0,4 | 28,99408284 | -156,1219845 | 840,6568397 |
| 5,076923077 | 6,680236686 | -17,26584251 | 44,62556219 |
| 6,769230769 | 0,04639053254 | 0,00999180701 | 0,00215208151 |
| 0,7230769231 | 9,092544379 | 27,41751843 | 82,67436328 |
| 0,8307692308 | 33,81869822 | 196,6687374 | 1143,70435 |
| 0,9384615385 | 74,22485207 | 639,4756486 | 5509,328665 |
| 1,046153846 | 130,3110059 | 1487,550252 | **16980,95826** |

| **Среднее арифметическое** | | | 15,78461538 |
| --- | --- | --- | --- |
| **Дисперсия** | | | 10,89106964 |
| **Среднее квадратическое отклонение** | | | 3,300162063 |
| **Коэффициент вариации** | | | 20,90745946 |
| **Коэффициент асимметрии** | | | 2,330374638 |
| **Эксцесс распределения** | | | 4,977285161 |

Анализ вариационного ряда на основе полученных характеристик:

Коэффициент асимметрии положителен и больше 0, поэтому можно сделать вывод, что распределение имеет правостороннюю асимметрию. Значение эксцесса положительно, что позволяет сделать вывод об остроконечности распределения

Задание 2:

Менеджер большого универмага зафиксировал суммы денег, которые израсходовали 184 покупателя, посетившие отдел верхней одежды в день сезонной распродажи по сниженным ценам. Зная минимальную и максимальную стоимость покупки, менеджер сгруппировал данные о суммах, израсходованных на покупки, в виде таблицы

*Таблица 1.*

**Суммы денег, израсходованные на покупки товаров в отделе верхней одежды (у.ед.)**

| **Интервалы расходов** | 100-300 | 300-500 | 500-700 | 700-900 | 900-1100 | 1100-1300 |
| --- | --- | --- | --- | --- | --- | --- |
| **Число покупателей (mi)** | 30 | 38 | 50 | 31 | 22 | 13 |
| **Доля покупателей (wi)** | 0.163 | 0.207 | 0.272 | 0.168 | 0.120 | 0.070 |

**Вопросы:**

1. Какова нижняя и верхняя границы интервалов (по таблице 1)?
2. Изучите распределение, представленное в таблице 2.

Задание 3:

*Таблица 2*

| **Интервалы** | **Число регионов** |
| --- | --- |
| До 60 | 10 |
| 60-70 | 29 |
| 70-80 | 2 |
| 80-90 | 13 |
| 90-100 | - |
| Свыше 100 | 6 |

Определите начало первого интервала и правую границу последнего интервала.

1. Для данных Задания №1 определите оптимальную величину интервала и представьте ряд из этого задания в виде интервального ряда.
2. Интервальные ряды бывают с равными и неравными интервалами. Иногда при группировке с равными интервалами сначала определяют число интервалов (групп) z при заданном объеме совокупности, используя формулу:

,

и тогда k в формуле Стерджесса вычисляется по формуле

Для Задания 1 вычислите оптимальную величину интервала по данной формуле и сравните его с интервалом, вычисленным по формуле Стерджесса.

1. Для данных таблицы 1 накопленные частоты и расположите их в таблице в восходящем порядке и в нисходящем порядке. На что они указывают? Поясните.
2. Постройте полигон распределения для Задания 1.
3. Постройте гистограмму и кумулянту для данных таблицы 2.
4. Постройте **огиву** для данных таблицы 2.

Огива строится аналогично кумулянте (но по оси абсцисс откладываются накопленные частоты (частости), а по оси ординат - значения признака).

1. Для данных Задания №1 вычислить:
2. 25-й, 50-й и 90-й перцентили в вариационном ряду.
3. По данным таблицы 2 вычислить медиану (используйте формулу для нахождения медианы внутри медианного интервала).
4. По данным таблицы 2 вычислить моду (используйте формулу для нахождения моды внутри модального интервала).
5. Вычислите среднюю арифметическую для данных Задания №1 по формулам:

* Средней арифметической;
* Средней арифметической взвешанной.

1. Проверить насколько медиана и средняя арифметическая чувствительна к положению крайних значений ряда значений (для данных Задания №1).
2. Может ли быть в одном ряду несколько мод? Обоснуйте.

*Решение:*

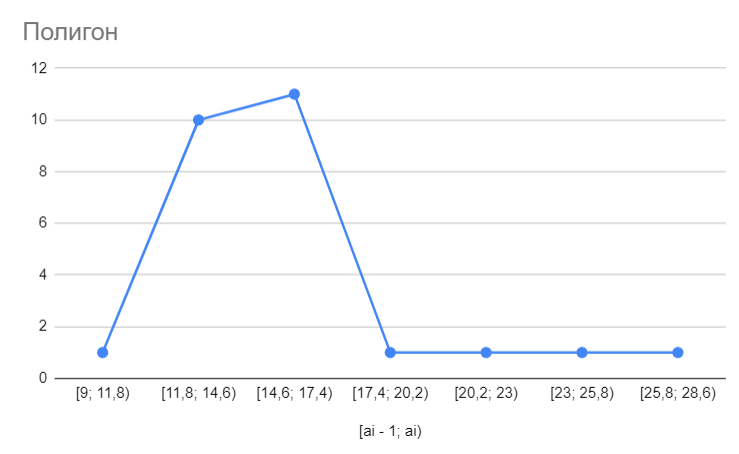
1. Нижняя граница имеет значение - 100, верхняя - 1300
2. Так как вариационный ряд должен иметь одинаковый шаг для каждого интервала, а в нашей таблице на примере второй ячейки мы можем понять, что он равен 70 - 60 = 10. Получается, что нижняя граница интервала первой ячейки равна 60 - 10 = 50, а правая граница последней = 100 + 10 = 110

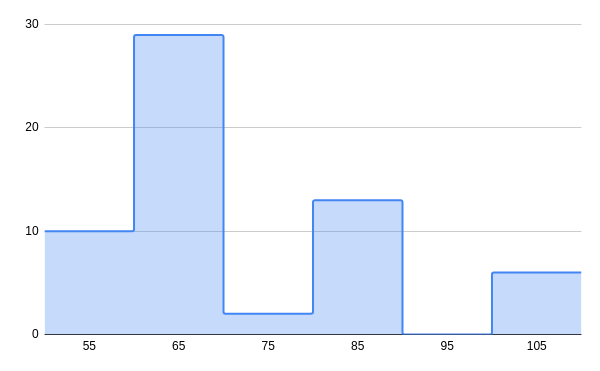
*) = 6.51619307604*

Выполнение ряда см. в самом задании.

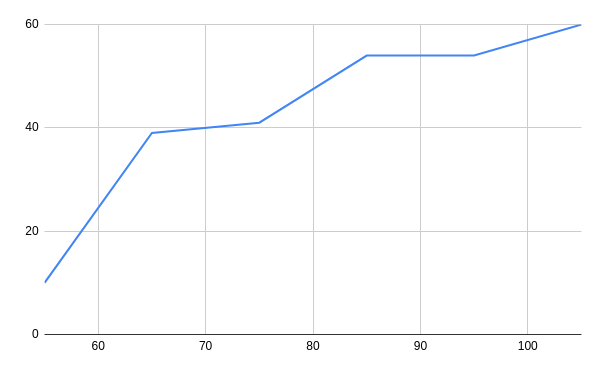


Отклонения от метода вычисления через формулу Стерджесса относительно невелико

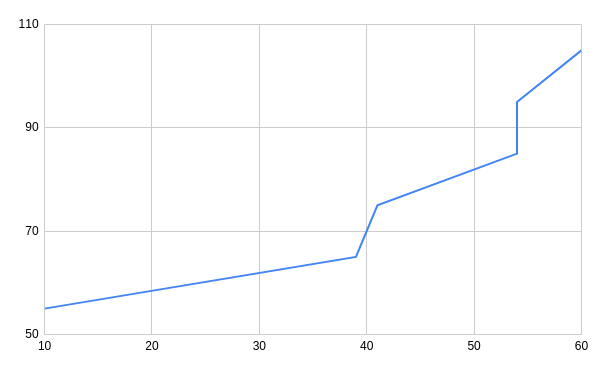
1. Накопленная частота интервалов – это число, полученное последовательным суммированием частот в направлении от первого интервала к последнему, до того интервала включительно, для которого определяется накопленная частота.
3. 
4. Гистограмма:



Кумулянта



1. Огива:



1. P -ый перцентиль вариационного ряда – это значение признака, слева от которого лежит P% вариантов ряда. Позиция P -го перцентиля задается как (n + 1) P /100, где n – число вариантов ряда.

Перцентиль – это значение признака в определенной позиции

ранжированного ряда, мера относительной позиции варианта в ряду

25-й перцентиль

50-й перцентиль

90-й перцентиль

1. Медиана – значение признака ряда, относительно которого вариационный ряд делится на две равные по числу вариантов части. Это – 50-й перцентиль.

1. Мода – это значение признака, наиболее часто встречающееся в вариационном ряду

=64.10434

1. Среднюю арифметическую мы вычислили в задании 1:



Средняя арифметическая взвешенная:



1. Средняя арифметическая довольно чувствительна к влиянию крайних значений переменной и ее значение смещается в их сторону. Медиана же не чувствительна к влиянию крайних значений. Медиана - это середина набора данных, половина из которых лежит ниже (или левее) медианы, а другая половина выше (или правее).
2. Мода — это наиболее часто встречающееся значение в выборке.

В одном ряду может быть две моды и более. Например в выборке 3, 3, 4, 5, 6, 7, 8, 8 модами будут значения 3 и 8. Значение 2 будет называться нижней модой, значение 8 верхней модой. Если два соседних значения встречаются одинаково часто, то мода считается как среднее арифметическое между ними. Например в распределении 3, 4, 4, 5, 5, 6, 7 модой будет значение 4,5 (четыре целых пять десятых) поскольку 4 и 5 находятся рядом и встречаются одинаково часто.

**Вывод:**

В данной работе были выполнены различные задания по анализу данных. В ходе выполнения работы были изучены основные математические характеристики вариационных рядов распределения и способы их вычисления. Изучены следующий характеристики: среднее значение признака, дисперсия, среднее квадратическое отклонение, коэффициент вариации, коэффициент асимметрии, эксцесс. Все изученные характеристики помогают получить более детальное представление о вариационном ряде, глубже его проанализировать.