

Математические методы в педагогике

Формы учета результатов измерений
Числовые характеристики распределений

Е.А.Коняева

- Для наглядного представления экспериментальных данных используются различные приемы облегчающие прежде всего визуальный анализ полученной в эксперименте информации. К таким приемам относят таблицы, ряды распределений, графики, гистограммы. Их применяют с той целью чтобы полученные экспериментальные данные представить наглядным образом и можно было бы в явной форме увидеть характерные особенности и результаты эксперимента

- Первичный экспериментальный материал полученный педагогом нуждается в соответствующей обработке. Обработка начинается с упорядочения и систематизации собранных данных.
- Процесс систематизации результатов эксперимента объединение их в относительно однородные группы по некоторому признаку называется **группировкой**.
 - **Формы группировки**

- Таблицы
- Статистические ряды
- Ряд распределения (полигон распределения)
- Гистограмма

Таблицы

Наиболее распространенной формой группировки экспериментальных данных являются **статистические таблицы.**

классы	леворукие	Праворукие	сумма
3	23	2	25
4	20	4	24
5	22	11	33
6	22	8	30
суммы	87	23	110

- **Статистические ряды** – это числовые значения признака, расположенного в определенном порядке.
- **Вариационным рядом распределения** называют двойной ряд чисел, показывающий, каким образом числовые значения признака связаны с их повторяемостью в данной выборке.

- Например, проведено тестирование интеллекта по тесту Векслера у 25 школьников, и сырые баллы по второму субтесту оказались следующими 6, 9, 5, 7, 10, 8, 9, 10, 8, 11, 9, 12, 9, 8, 10, 11, 9, 10, 8, 10, 7, 9, 10, 9, 11
- ряд можно представить в более удобной, компактной форме (в виде вариационного ряда распределения)

6, 9, 5, 7, 10, 8, 9, 10, 8, 11, 9, 12, 9, 8, 10, 11, 9, 10, 8, 10, 7, 9, 10, 9, 11

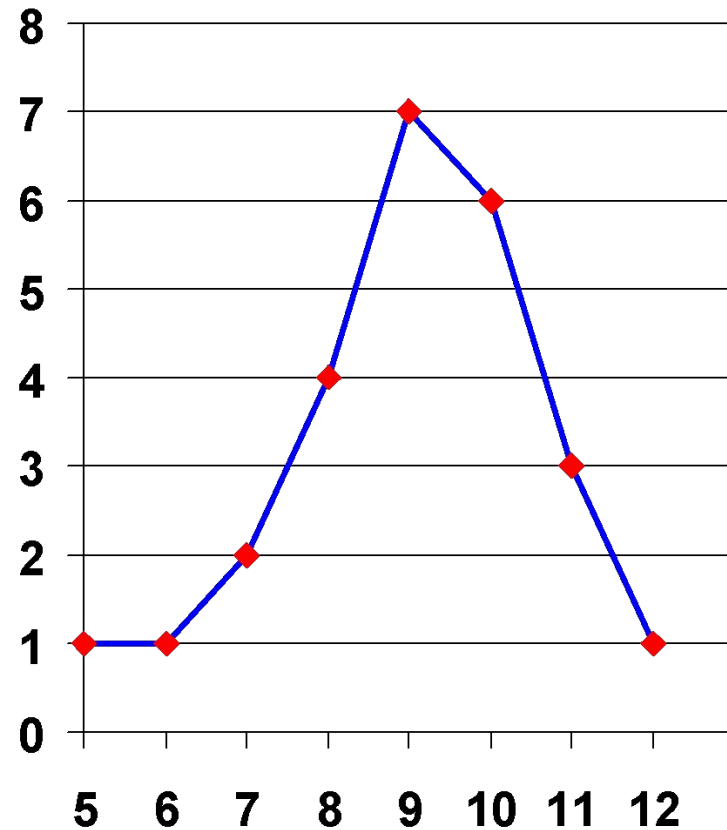
Варианты	x_i	5	6	7	8	9	10	11	12
Частоты вариант	F_i	1	1	2	4	7	6	3	1

При чем, сумма частот вариант равна объему выборки

$$n = \sum_i f_i = 1 + 1 + 2 + 4 + 7 + 6 + 3 + 1 = 25$$

- **Полигон частот распределения**

x_i	5	6	7	8	9	10	11	12
F_i	1	1	2	4	7	6	3	1



Интервальный ряд распределения

Поскольку при анализе ряда распределения сравнивают частоты в разных интервалах, необходимо, чтобы длина интервалов была постоянной. Оптимальное число групп выбирается так, чтобы достаточной мере отразилось разнообразие значений признака в совокупности и в то же время закономерность распределении, его форма не искажалась случайными колебаниями частот. Если групп будет слишком мало, не проявится закономерность вариации; если групп будет чрезмерно много, случайные скачки частот исказят форму распределения.

Чаще всего число групп в ряду распределения определяют по **формуле Стерждесса** (1) или (2):

$$(1) \quad k = 1 + 3,322 \lg N \quad (2) \quad k = 1 + 1,44 \ln N$$

где k – число групп (округляемое до ближайшего целого числа);
 N – численность совокупности.

Зная число групп, рассчитывают длину (размах) интервала по формуле (3):

$$h = \frac{X_{\max} - X_{\min}}{k}$$

где X_{\max} и X_{\min} — максимальное и минимальное значения в совокупности.

В нашем примере по формуле Стерждесса (1) определим число групп:

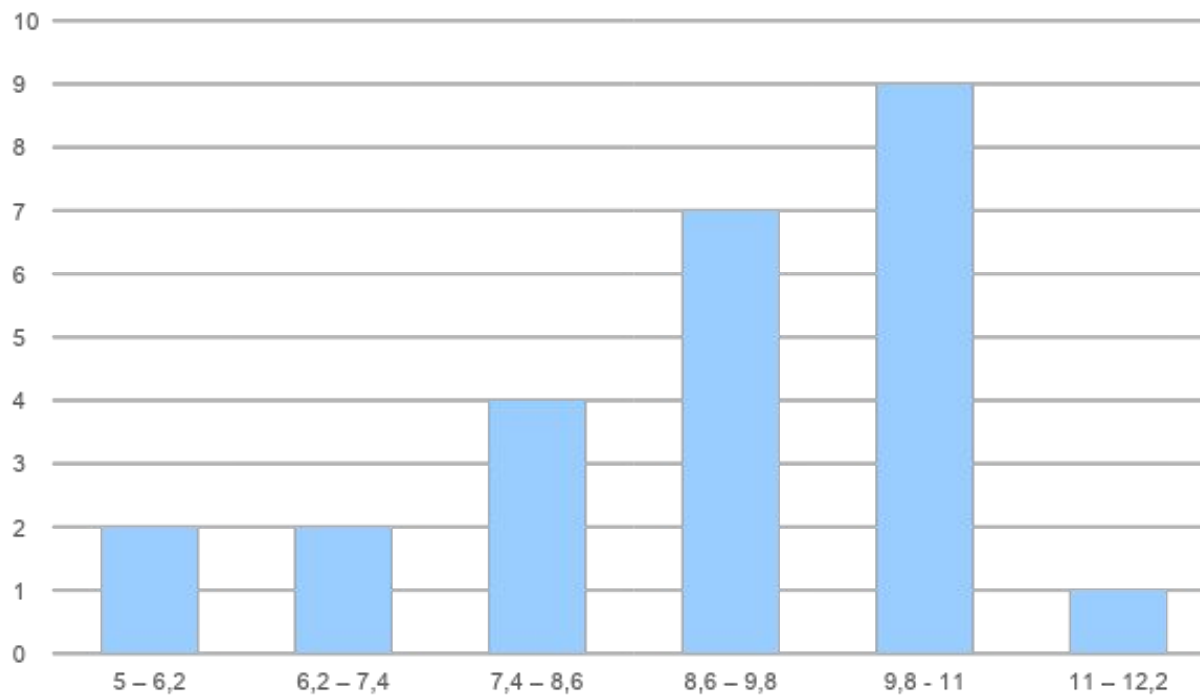
$$k = 1 + 3,322 \lg 25 = 1 + 3.322 * 1.3979 = 5.64 \approx 6$$

Рассчитаем длину (размах) интервала по формуле (3):

$$h = \frac{(12 - 5)}{6} = \frac{7}{6} \approx 1,2$$

Теперь построим интервальный ряд с 6 группами с интервалом 1,2

интервалы	5 – 6,2	6,2 – 7,4	7,4 – 8,6	8,6 – 9,8	9,8 - 11	11 – 12,2
частоты	2	2	4	7	9	1



Числовые характеристики распределений

Мода

Медиана

Среднее арифметическое

Дисперсия

Среднее квадратическое отклонение (стандартное отклонение)

Мода

Мода — это такое числовое значение которое встречается в выборке наиболее часто. Мода обозначается **Mod X**.

Так например в ряду значений (2, 6, 6, 8, 9, 9, 9, 10) модой является 9 потому что 9 встречается чаще любого другого числа.

- Моду находят согласно следующим правилам
1. В том случае когда все значения в выборке встречаются одинаково часто принято считать что этот выборочный ряд не имеет моды.

Например 5, 5, 6, 6, 7, 7 — в этой выборке моды нет

2. Когда два (и более) соседних (смежных) значения имеют одинаковую частоту и их частота больше частот любых других значений мода вычисляется как среднее арифметическое этих двух значений.

- Например в выборке 1, 2, 2, 2, 5, 5, 5, 6 частоты рядом расположенных значений 2 и 5 совпадают и равняются 3. Эта частота больше чем частота других значений 1 и 6 (у которых она равна 1)

Следовательно модой этого ряда будет величина

$$ModX = \frac{2+5}{2} = 3,5$$

3. Если два (и более) несмежных (не соседних) значения в выборке имеют равные частоты которые больше частот любого другого значения то выделяют две или более моды.

Например в ряду 10, 11, 11, 11, 12, 13, 14, 14, 14, 17 модами являются значения 11 и 14. В таком случае говорят что выборка является бимодальной.

$$Mod_1 X = 11$$

$$Mod_2 X = 14$$

- Могут существовать и так называемые мультимодальные распределения имеющие более двух вершин (мод).

Медиана

Медиана — это значение, которое делит **упорядоченное** множество данных пополам. Обозначается **Med X**.

- Найдем медиану выборки 9, 3, 5, 8, 4, 11, 13

Решение: Сначала упорядочим выборку по величинам входящих в нее значений. Получим 3, 4, 5, 8, 9, 11, 13. Т.о. медиана – 8.

$$\text{Med } X = 8$$

- Найдем медиану выборки 20, 9, 13, 1, 4, 11

Упорядочим выборку 1, 4, 9, 11, 13, 20. Поскольку здесь имеется четное число элементов, то существует две «середины» — 9 и 11. В этом случае медиана определяется как среднее арифметическое этих значений.

$$MedX = \frac{9+11}{2} = \frac{20}{2} = 10$$

Среднее арифметическое (оценка математического ожидания)

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

x_i - каждое наблюдаемое значение признака,
 i – индекс, указывающий на порядковый номер данного значения признака
 n – количество наблюдений

Дисперсия

Рассмотрим еще одну очень важную числовую характеристику выборки, называемую **дисперсией**. Дисперсия представляет собой наиболее часто используемую меру рассеяния случайной величины (переменной). **Дисперсия** это среднее арифметическое квадратов отклонений значений переменной от ее среднего значения.

$$\bar{D} = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{X})^2$$

Среднее квадратическое отклонение

$$\delta = \sqrt{D}$$

Пример

Определить моду, медиану, среднее арифметическое, дисперсию, среднее квадратическое отклонение выборки 2, 4, 6, 8, 10.

Mod X – нет, т.к. все значения встречаются одинаково часто

Med X = 6

Среднее арифметическое $\bar{X} = \frac{2 + 4 + 6 + 8 + 10}{5} = \frac{30}{5} = 6$

Объем выборки $n = 5$

Дисперсию удобно находить, построив таблицу

x_i	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
2	-4	16
4	-2	4
6	0	0
8	2	4
10	4	16
	Σ	40

$$D = \frac{40}{5} = 8$$

$$\delta = \sqrt{D} = \sqrt{8} \approx 2,82$$

Задачи для самостоятельного решения

- Получены данные о числе цветных телевизоров, продаваемых ежегодно в магазинах электроники в течение 26 дней: 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.
1. Построить вариационный ряд
 2. Построить полигон частот распределения
 3. Построить интервальный статистический ряд
 4. Построить гистограмму
 5. Найти моду, медиану, среднее арифметическое, дисперсию, среднее квадратическое отклонение.

- Двадцати пяти подросткам, отобранным случайным образом, показали блок телевизионной коммерческой рекламы о новых сортах жевательной резинки и попросили оценить рекламу в баллах от 0 до 100. Получены следующие результаты: 89, 75, 59, 96, 88, 71, 43, 62, 88, 96, 75, 71, 62, 60, 96, 89, 59, 53, 43, 87, 91, 91, 60, 81, 50
1. Построить вариационный ряд
 2. Построить полигон частот распределения
 3. Построить интервальный статистический ряд
 4. Построить гистограмму
 5. Найти моду, медиану, среднее арифметическое, дисперсию, среднее квадратическое отклонение