

ВЫБОРОЧНЫЙ МЕТОД

ВОПРОСЫ:

1. Задачи математической статистики.
2. Генеральная и выборочная совокупности.
3. Статистический вариационный ряд распределения.
4. Статистический интервальный ряд распределения.

ВОПРОС 4.

Статистический интервальный ряд распределения

Представление результатов наблюдений в виде вариационного ряда удобно лишь в случае ограниченного количества (не более 10-20) различающихся между собой вариант в выборочной совокупности.

Если количество вариант существенно больше, то результаты представляют в виде *статистического интервального ряда распределения*.

В этом случае разбивают вариационный ряд на несколько интервалов и указывают число значений в каждом из них.

АЛГОРИТМ ПОСТРОЕНИЯ ИНТЕРВАЛЬНОГО РЯДА

Пусть имеется вариационный ряд:

$$x_{min} = x_1, x_2, \dots, x_n = x_{max}$$

Разобьем варианты на отдельные интервалы, т. е. проведем их *группировку*.

Для этого требуется:

1. Определить число интервалов m

Согласно формуле Стерджеса:

$$m = 1 + 3,322 \lg n,$$

где n — общее число наблюдений.

2. Определить величину (ширину) интервала

$$\Delta = \frac{x_{\max} - x_{\min}}{m}$$

где $x_{\max} - x_{\min}$ — разность между наибольшим и наименьшим значением признака (размах).

3. Определить границы интервалов

За начало первого интервала рекомендуется брать величину

$$x_{нач} = x_{\min} - \frac{\Delta}{2}$$

Замечание.

За начало первого интервала можно также брать величину $x_{\min}$

Границы интервалов получаются добавлением шага (величины интервала) к предыдущей границе.

4. Определить частоты (относительные частоты)

Частоты (n_i) – числа, показывающие сколько раз встречаются варианты из данного интервала.

Относительные частоты (W_i) – отношение частот к общему числу наблюдений , т.е.

$$W_i = n_i / n .$$

5. Сгруппированный ряд представить в виде таблицы

В первой строке указывается последовательность интервалов, во второй – соответствующие им частоты.

Пример 1. Изучался рост (в см) студентов 1 курса. Из очень большого количества студентов извлечена случайная выборка объема 50:

173, 169, 189, 181, 178, 158, 185, 166, 177, 161, 179, 194, 182, 174, 170, 179, 175, 171, 183, 186, 197, 163, 167, 174, 177, 183, 185, 194, 169, 191, 166, 177, 175, 183, 190, 187, 181, 177, 174, 162, 179, 174, 187, 183, 178, 174, 166, 177, 169, 185.

Найти статистический интервальный ряд распределения роста студентов.

Решение

1. Упорядочим данные выборки по возрастанию (ранжируем выборку):

158, 161, 162, 163, 166, 166, 166, 167, 169, 169, 169, 170, 171, 173, 174, 174, 174, 174, 174, 175, 175, 177, 177, 177, 177, 177, 178, 178, 179, 179, 179, 181, 181, 182, 183, 183, 183, 183, 185, 185, 185, 186, 187, 187, 189, 190, 191, 194, 194, **197**

2. Определим число интервалов:

$$m = 1 + 3,322 \lg 50 \approx 6,644 \approx 7.$$

3. Вычислим длину интервала:

$$\Delta = \frac{x_{\max} - x_{\min}}{m} = \frac{197 - 158}{7} \approx 5,57 \approx 6$$

Округляем с точностью чисел, заданных в условии задачи, т.е. до целых.

4. Определим границы интервалов:

$$x_{нач} = 158 - \frac{6}{2} = 155.$$

Рассматриваемый интервал разделим на 7 интервалов с шириной $\Delta = 6$.
Полученные результаты представим в таблице.

X	[155;161)	[161;167)	[167;173)	[173;179)	[179;185)	[185;191)	[191;197]
n_i							

5. Определим частоту n_i для каждого интервала.

158, 161, 162, 163, 166, 166, 166, 167, 169, 169, 169, 170, 171, 173, 174, 174, 174, 174, 174, 175, 175, 177, 177, 177, 177, 177, 177, 178, 178, 179, 179, 179, 181, 181, 182, 183, 183, 183, 183, 185, 185, 185, 186, 187, 187, 189, 190, 191, 194, 194, 197

Интервальный ряд распределения частот

X	[155;161)	[161;167)	[167;173)	[173;179)	[179;185)	[185;191)	[191;197]
n_i	1	6	6	15	10	8	4

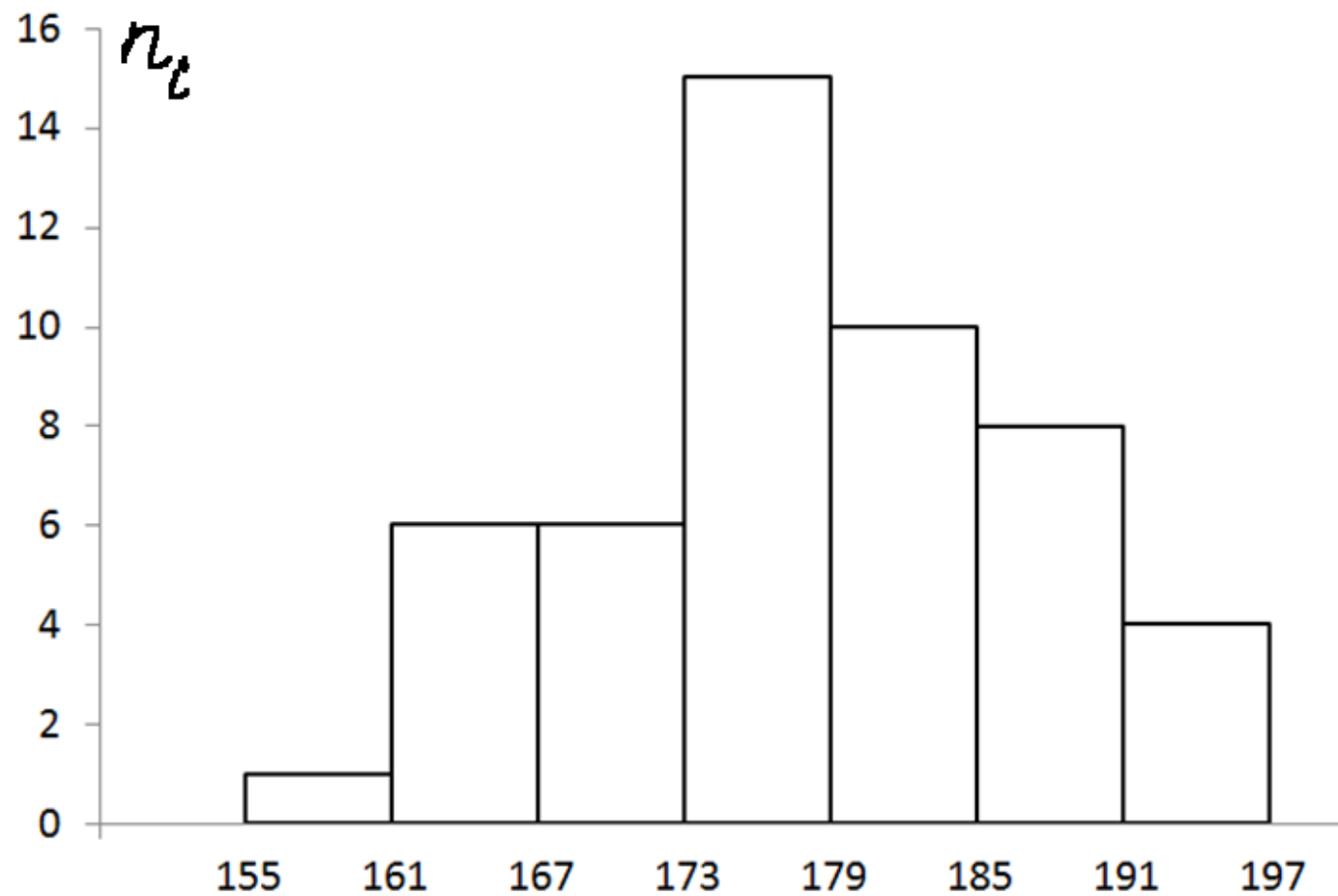
Интервальный ряд распределения относительных частот

X	[155;161)	[161;167)	[167;173)	[173;179)	[179;185)	[185;191)	[191;197]
W_i	0,02	0,12	0,12	0,3	0,2	0,16	0,08

ГИСТОГРАММА

Гистограмма служит для изображения интервальных вариационных рядов и представляет собой ступенчатую фигуру из прямоугольников с основаниями, равными интервалам значений признака и высотами, равными частотам n_i (относительным частотам W_i) интервалов.

Например, гистограмма роста студентов



Числовые характеристики интервального вариационного ряда

СРЕДНЯЯ АРИФМЕТИЧЕСКАЯ

Средняя арифметическая интервального вариационного ряда определяется по формуле:

$$\bar{x} = \frac{\sum_{i=1}^m x_i^* n_i}{n}$$

где

x_i^* — середины интервалов интервального вариационного ряда;

n_i — частоты интервалов;

m — число интервалов.

Так как $W_i = \frac{n_i}{n}$ — относительные частоты интервалов, то очевидно, что

$$\bar{x} = \sum_{i=1}^m x_i^* W_i$$

МЕДИАНА

Для интервального вариационного ряда :

$$\widetilde{Me} = a_i + h \frac{\frac{n}{2} - T}{n_i}$$

где a_i – начало медианного интервала, т.е. интервала, в котором содержится срединный элемент;

h – длина медианного интервала;

n – объем выборки;

T – сумма частот интервалов, предшествующих медианному;

n_i – частота медианного интервала.

М О Д А

Для интервального вариационного ряда:

$$\widetilde{M_o} = a_i + h \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}$$

где a_i — начало модального интервала, т.е. интервала с наибольшей частотой;

h — длина модального интервала;

n_i — частота модального интервала;

n_{i-1} , n_{i+1} — частоты предыдущего и последующего интервалов.

ДИСПЕРСИЯ

Дисперсия s^2 интервального вариационного ряда:

$$s^2 = \frac{\sum_{i=1}^m (x_i^* - \bar{x})^2 n_i}{n}$$

или

$$s^2 = \frac{1}{n} \sum_{i=1}^m (x_i^*)^2 n_i - (\bar{x})^2$$

Замечание. Суммирование выполняется по интервалам.

СРЕДНЕЕ КВАДРАТИЧЕСКОЕ ОТКЛОНЕНИЕ

Среднее квадратическое отклонение s — арифметическое значение корня квадратного из дисперсии:

$$s = \sqrt{s^2}$$

Пример 2. Для интервального ряда, полученного в примере 1

X	[155;161)	[161;167)	[167;173)	[173;179)	[179;185)	[185;191)	[191;197]
n_i	1	6	6	15	10	8	4

вычислить числовые характеристики:

а) среднюю арифметическую;

б) дисперсию;

в) среднее квадратическое отклонение;

г) моду;

д) медиану.

Решение

Вычислим середины интервалов интервального вариационного ряда x_i^* как *среднее арифметическое концов интервала*:

$$x_1^* = \frac{155 + 161}{2} = 158$$

$$x_5^* = \frac{179 + 185}{2} = 182$$

$$x_2^* = \frac{161 + 167}{2} = 164$$

$$x_6^* = \frac{185 + 191}{2} = 188$$

$$x_3^* = \frac{167 + 173}{2} = 170$$

$$x_7^* = \frac{191 + 197}{2} = 194$$

$$x_4^* = \frac{173 + 179}{2} = 176$$

Заполним таблицу:

Номер интервала	1	2	3	4	5	6	7
Границы интервала	155-161	161-167	167-173	173-179	179-185	185-191	191-197
Частота, n_i	1	6	6	15	10	8	4
Накопленная частота	1	7	13	28	38	46	50
Середина интервала, x_i^*	158	164	170	176	182	188	194

Вычислим числовые характеристики:

а) Средняя арифметическая.

$$\begin{aligned}\bar{x} &= \frac{1}{50} (158 * 1 + 164 * 6 + 170 * 6 + 176 * 15 \\ &\quad + 182 * 10 + 188 * 8 + 194 * 4) = \\ &= 178,04 \approx 178\end{aligned}$$

Таким образом, средний рост студентов 1 курса составляет 178 см (результат оставляем с той же точностью, что и сами данные).

б) Дисперсия

Дисперсию вычислим по формуле:

$$s^2 = \frac{1}{n} \sum_{i=1}^m (x_i^*)^2 n_i - (\bar{x})^2$$

$$\begin{aligned} \overline{x^2} &= \frac{1}{50} (158^2 * 1 + 164^2 * 6 + 170^2 * 6 + 176^2 * \\ &\quad * 15 + 182^2 * 10 + 188^2 * 8 + 194^2 * 4) \\ &= 31778,32 \end{aligned}$$

$$s^2 = 31778,32 - 178^2 = 94,32$$

в) Среднее квадратическое отклонение

$$s = \sqrt{94,32} \approx 9,7$$

Значит, рост студентов в выборке отклоняется от среднего значения 178 см в среднем на 9,7 см.

з) *Мода*

Модальный интервал, т.е. интервал с наибольшей частотой: [173; 179).

Начало модального интервала $a_4 = 173$.

Длина модального интервала $h = 179 - 173 = 6$.

Частота модального интервала $n_4 = 15$.

Частота предыдущего интервала $n_3 = 6$.

Частота последующего интервала $n_5 = 10$.

Получим:

$$\widetilde{M}_o = 173 + 6 * \frac{15 - 6}{(15 - 6) + (15 - 10)} == 176,86 \approx 177$$

д) Медиана

Медианному интервалу соответствует *первая* накопленная частота, большая $n/2$.

Так как $n = 50$, то $n/2 = 50/2 = 25$.

Первая накопленная частота, большая 25, соответствует четвертому интервалу, т.е. $[173; 179)$. Этот интервал – медианный, в нем содержится срединный элемент.

Начало медианного интервала $a_4 = 173$; длина медианного интервала $h = 6$;

объем выборки $n = 50$;

сумма частот интервалов, предшествующих медианному, $T = n_1 + n_2 + n_3 = 1 + 6 + 6 = 13$;

частота медианного интервала $n_4 = 15$.

Получим:

$$\widetilde{Me} = 173 + 6 * \frac{\frac{50}{2} - 13}{15} = 177,8 \approx 178.$$