

Занятие 1

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ. ВЫБОРКА. ВАРИАЦИОННЫЕ РЯДЫ И ИХ ГРАФИЧЕСКОЕ ИЗОБРАЖЕНИЕ

Краткие теоретические сведения

Статистической совокупностью называют множество предметов или явлений, объединенных каким-либо общим признаком или свойством качественного или количественного характера.

Статистическая совокупность называется *генеральной*, если исследованию подвергаются все ее элементы.

Выборочной совокупностью, или **выборкой**, называется совокупность объектов, случайно отобранных из генеральной совокупности.

Объемом совокупности называется число ее объектов. **Объем выборки** – количество объектов, входящих в выборку. Например, если из 10 000 изготовленных деталей для обследования отобрано 100, то объем генеральной совокупности $N = 10\,000$, объем выборки $n = 100$.

Статистические данные представляют собой набор значений некоторой случайной величины X , после обработки которых строятся функции, называемые статистиками, характеризующие случайную величину.

Рассмотрим выборку, извлеченную из генеральной совокупности, в которой значение x_1 наблюдалось n_1 раз, x_2 наблюдалось n_2 раз, ..., x_k наблюдалось n_k раз, причем $\sum_{i=1}^k n_i = n$, где n – объем выборки. Наблюдаемые значения x_i называются **вариантами**.

Упорядочение, расположение вариантов в порядке возрастания (убывания) есть ранжирование вариантов ряда.

Числа n_i называют **частотами**, а их отношение к объему выборки – **относительными частотами** и обозначают $W_i = \frac{n_i}{n}$.

Статистическим распределением выборки называется соответствие между вариантами и их частотами (или относительными частотами).

Статистическое распределение может быть задано, например, с помощью таблицы, в которой указаны варианты и соответствующие им частоты. Статистическое распределение можно задавать также указанием последовательности некоторых интервалов и соответствующих им частот (сумма частот, попавших в этот интервал).

Вариационным рядом называется ранжированный в порядке возрастания (или убывания) ряд вариант с соответствующими им частотами или относительными частотами.

Вариационный ряд называется **дискретным**, если любые его варианты отличаются на постоянную величину (обычно целое число).

Общий вид дискретного вариационного ряда представлен в табл. 1.1, где $i = \overline{1, k}$.

Таблица 1.1

Значение признака (варианты) x_i	x_1	x_2	x_3	...	x_k
Частоты n_i	n_1	n_2	n_3	...	n_k

Вариационный ряд называется **интервальным** (непрерывным), если варианты могут отличаться одна от другой на сколько угодно малую величину.

Общий вид интервального вариационного ряда представлен в табл. 1.2, где $i = \overline{1, l}$.

Таблица 1.2

Значение признака (варианты) x_i	$x_1 - x_2$	$x_2 - x_3$...	$x_{i-1} - x_i$
Частоты n_i	n_1	n_2	...	n_i

Под **частотой интервала** понимают число членов совокупности, варианта которых лежит в данном интервале.

Число интервалов l следует брать не очень большим, чтобы после группировки ряд не был громоздким, и не очень малым, чтобы не потерять особенности распределения признака.

Рекомендуемое число интервалов можно найти по формуле Стерджесса:

$$l = 1 + 3,322 \cdot \lg n,$$

где n – объем выборки,

а величину интервала (ширину интервала) – по формуле

$$h = \frac{x_{\max} - x_{\min}}{l},$$

где $x_{\max} - x_{\min}$ – разность между наибольшим и наименьшим значениями признака, которая называется **размахом выборки**.

Если окажется, что h – дробное число, то за длину интервала следует брать либо ближайшее целое число, либо ближайшую простую дробь.

При изучении вариационных рядов используется также понятие **накопленной частоты** $n_i^{\text{нак}}$, которая показывает, сколько наблюдалось вариант со значением признака, меньшим x .

Отношение накопленной частоты к общему числу наблюдений называют **накопленной частотью**:

$$W_i^{\text{нак}} = \frac{n_i^{\text{нак}}}{n}.$$

Накопленные частоты для дискретного ряда находятся по формуле

$$n_i^{\text{нак}} = n_i + n_{i-1} + \dots + n_1.$$

В случае интервального вариационного ряда накопленные частоты для каждого интервала находятся последовательным суммированием частот всех предшествующих интервалов, включая данный.

Для графического изображения вариационных рядов можно использовать полигон, гистограмму, кумулятивную кривую, огиву.

Полигон используют для изображения дискретного ряда. Для его построения в прямоугольной системе координат наносят точки с координатами (x_i, n_i) или (x_i, W_i) . Ломаная, соединяющая отмеченные точки, называется **полигоном распределения** частот или частостей.

Гистограмма служит только для изображения интервальных вариационных рядов. Если интервалы имеют одинаковую величину, то гистограмма представляет собой ступенчатую фигуру из прямоугольников с основаниями, равными интервалам значений признака, и высотами, равными частотам (частостям) интервалов. Соединив середины верхних оснований прямоугольников отрезками прямой, получим полигон того же распределения.

Кумулятивная кривая – это кривая накопленных частот (частостей).

Для дискретного вариационного ряда **кумулята** – ломаная, соединяющая точки $(x_i, n_i^{\text{нак}})$ или $(x_i, W_i^{\text{нак}})$, $i = \overline{1, k}$.

Для интервального ряда ломаная начинается с точки, абсцисса которой равна началу первого интервала, а ордината – накопленной частоте (частости), равной нулю. Абсциссами других точек этой ломаной являются верхние границы интервалов, а ординатами – накопленные частоты (частости) соответствующих интервалов.

Огива строится аналогично кумуляте. На оси абсцисс наносятся точки, соответствующие накопленным частотам (частостям), а на оси ординат – значения варианты.

Пример 1.1.

В результате тестирования группа студентов из 25 человек набрала баллы: 4, 0, 3, 4, 1, 0, 3, 1, 0, 1, 0, 0, 3, 1, 0, 1, 1, 3, 2, 3, 4, 2, 1, 2, 3. Построить дискретный вариационный ряд. Построить полигон распределения частот и относительных частот, кумуляту и огиву статистического распределения.

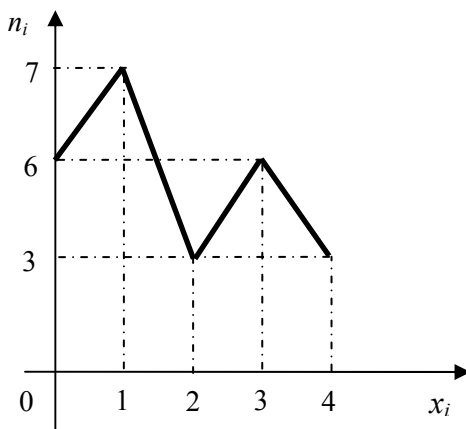
Решение. Проранжируем исходные данные, подсчитаем частоту вариантов: 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4 (табл. 1.3).

Таблица 1.3

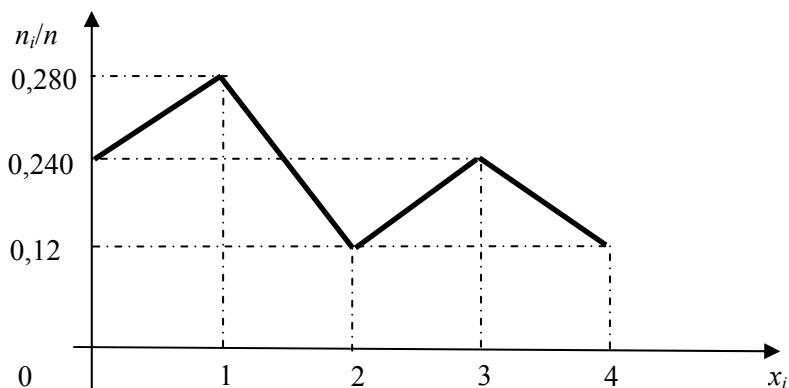
Балл x_i	0	1	2	3	4
Число студентов n_i	6	7	3	6	3

$$\sum_{i=1}^5 n_i = 25.$$

Построим полигон частот.



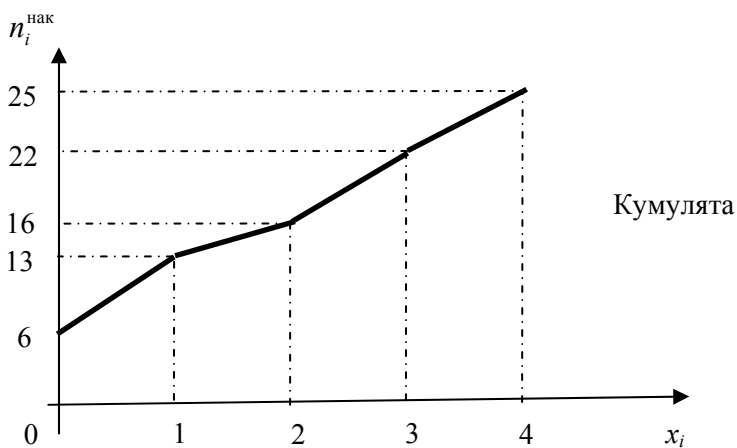
Полигон относительных частот будет иметь следующий вид.

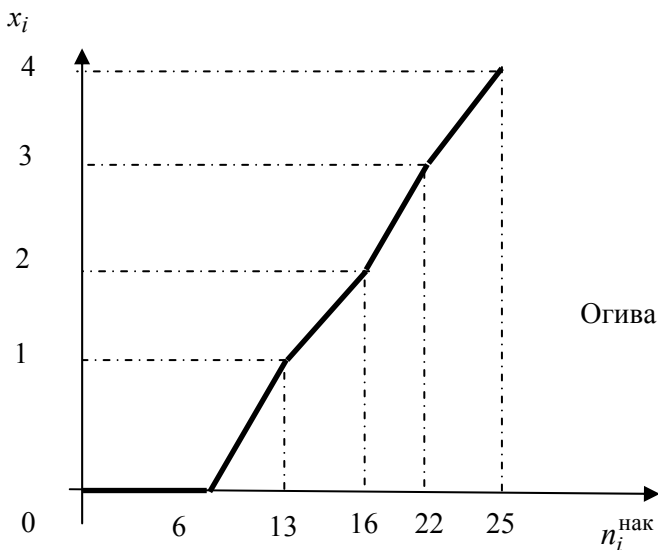


Вычислим накопленные частоты и частоты (табл. 1.4).

Таблица 1.4

x_i	0	1	2	3	4
n_i	6	7	3	6	3
$n_i^{\text{нак}}$	6	13	16	22	25
$W_i^{\text{нак}}$	6/25	13/25	16/25	22/25	1





Пример 1.2.

Результаты измерения производительности труда 100 рабочих имеют следующий вид:

90 75 85 84 83 84 76 76 74 81 78 83 78 82 81 75 92 76 95 92 73
 85 89 78 90 75 83 73 91 74 79 83 81 88 76 83 79 81 80 77 76 79
 81 79 76 83 84 77 86 83 77 82 85 81 84 85 83 82 83 78 84 81 82
 89 81 81 78 74 76 79 84 74 80 71 83 88 75 86 78 86 76 80 87 83
 87 73 84 82 85 85 79 73 75 84 79 81 86 84 82 90.

Построить интервальный вариационный ряд. Построить гистограмму (полигон) частот. Построить кумуляту и огиву.

Решение. В случае, когда число вариантов x_i достаточно велико (>30), составляют интервальный вариационный ряд. Наибольшим значением случайной величины является 95, а наименьшим – 71, т. е. $x_{\max} = 95$, $x_{\min} = 71$.

Для определения величины интервала используем формулу Стерджесса:

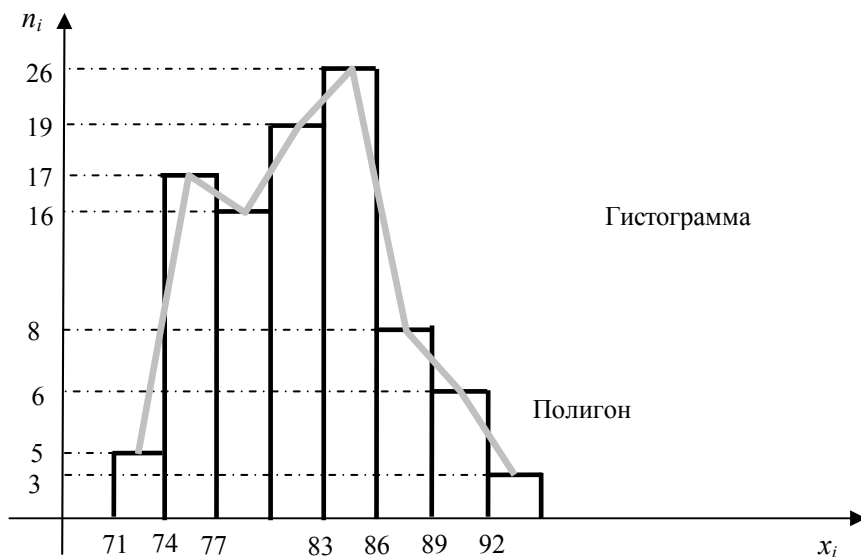
$$h = \frac{95 - 71}{1 + 3,322 \lg 100} = 3,14.$$

Возьмем за ширину интервала $h = 3$.

Интервальный ряд представлен в табл. 1.5.

Таблица 1.5

$[x_{i-1} - x_i)$	[71-74)	[74-77)	[77-80)	[80-83)	[83-86)	[86-89)	[89-92)	[92-95]
n_i	5	17	16	19	26	8	6	3



Гистограмма относительных частот является аналогом дифференциальной функции случайной величины.

Найдем накопленные частоты $n_i^{\text{нак}}$ для каждого из интервалов данного интервального вариационного ряда (табл. 1.6).

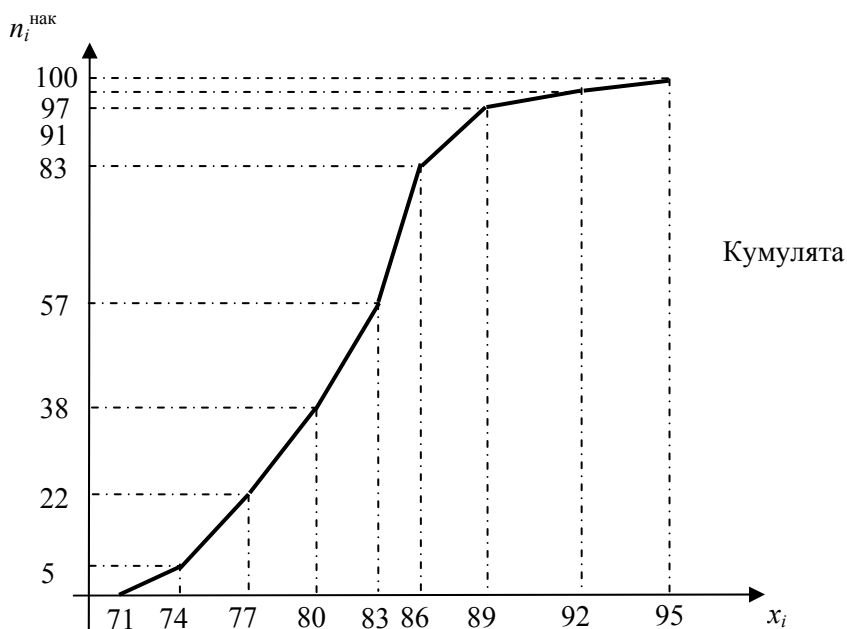
Таблица 1.6

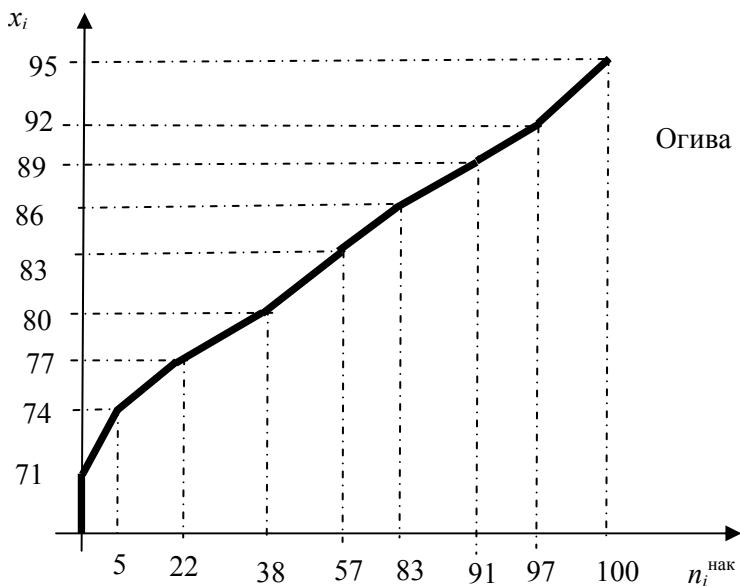
$n_i^{\text{нак}}$	5	22	38	57	83	91	97	100
--------------------	---	----	----	----	----	----	----	-----

Кумулятивный ряд представлен в табл. 1.7.

Таблица 1.7

$[x_{i-1} - x_i)$	[71–74)	[74–77)	[77–80)	[80–83)	[83–86)	[86–89)	[89–92)	[92–95]
$n_i^{\text{нак}}$	5	22	38	57	83	91	97	100





Задачи для аудиторной работы

Задача 1.1.

Наблюдая за толщиной (в мм) 50 слюдяных прокладок, получили следующие результаты:

0,021 0,039 0,030 0,042 0,031 0,028 0,030 0,024 0,033 0,040
 0,031 0,027 0,031 0,031 0,045 0,027 0,034 0,028 0,027 0,046
 0,033 0,030 0,043 0,030 0,028 0,051 0,036 0,031 0,034 0,031
 0,030 0,039 0,031 0,042 0,034 0,036 0,030 0,031 0,033 0,031
 0,030 0,048 0,030 0,033 0,028 0,031 0,036 0,034 0,042 0,037.

Построить интервальный вариационный ряд с равными интервалами (первый интервал 0,020–0,024; второй – 0,024–0,028 и т. д.). Начертить гистограмму частот. Построить кумулятивный ряд.

Задача 1.2.

В течение часа АТС регистрировалось число неправильных соединений в минуту: 1, 2, 1, 1, 0, 2, 1, 1, 1, 3, 1, 1, 1, 4, 2, 2, 2,

2, 0, 1, 4, 3, 3, 1, 1, 0, 0, 1, 2, 2, 4, 5, 3, 3, 3, 2, 2, 2, 1, 1, 2, 1, 0, 2, 2, 2, 1, 1, 3, 3, 4, 0, 2, 2, 2, 1, 3, 1, 3, 2.

Построить:

- а) дискретный вариационный ряд;
- б) полигон распределения частот;
- в) кумуляту и огиву статистического распределения.

Задача 1.3.

По данной выборке

3,86 4,06 3,67 3,97 3,76 3,61 3,96 4,04 3,84 3,94 3,99 3,69 3,76
3,71 3,94 3,82 4,16 3,76 4,00 3,46 3,71 3,81 4,02 4,17 3,72 4,09
3,78 4,02 3,73 3,52 4,03 4,14 3,72 4,33 3,82 4,03 3,62 3,91 3,98
4,08 3,89 3,57 3,88 3,92 3,87 4,01 4,18 4,07 3,93 4,26

построить:

- а) интервальный вариационный ряд;
- б) гистограмму частот;
- в) кумуляту и огиву.

Задачи для внеаудиторной работы

Задача 1.4.

В результате тестирования 60 студентов набрали баллы:

20 19 22 24 21 18 23 17 20 16 15 23 21 24 21 18 23 21 19 20 24
21 20 18 17 22 20 16 22 18 20 17 21 17 19 20 20 21 18 22 23 21
25 22 20 19 21 24 23 21 19 22 21 19 20 23 22 25 21 21.

Построить:

- а) дискретный вариационный ряд;
- б) полигон распределения относительных частот;
- в) огиву и кумуляту.

Задача 1.5.

По данной выборке

111 85 85 91 101 109 86 102 111 98 105 85 112 109 115 99 105
111 94 107 99 107 125 89 104 113 105 88 103 97 115 109 89 108
107 97 106 107 96 108 109 139 116 117 103 127 119 118 125 105