

## Практическое занятие 12

**Тема:** Генеральная совокупность и выборка. Эмпирическая и теоретическая функции распределения. Полигон и гистограмма.

**Цель:** изучить основные понятия математической статистики; познакомить с методами решения задач.

### Общие теоретические положения и решение типовых задач

Установление закономерностей, которым подчинены массовые случайные явления, основано на изучении методами теории вероятностей статистических данных - результатов наблюдений. Современная математическая статистика разрабатывает способы определения числа необходимых испытаний до начала исследования (планирование эксперимента), в ходе исследования (последовательный анализ) и решает многие другие задачи. Современную математическую статистику определяют как науку о принятии решений в условиях неопределенности.

Пусть требуется изучить совокупность однородных объектов относительно некоторого качественного или количественного признака, характеризующего эти объекты. Например, если имеется партия деталей, то качественным признаком может служить стандартность детали, а количественным - контролируемый размер детали.

Различают генеральную и выборочную совокупности:

*Генеральной совокупностью* называют совокупность всех мысленно возможных объектов данного вида, над которыми проводятся наблюдения с целью получения конкретных значений случайной величины, или совокупность результатов всех мыслимых наблюдений, проводимых в неизменных условиях над одной из случайных величин, связанных с данным видом объектов.

*Выборочной совокупностью* называют часть отобранных объектов из генеральной совокупности.

*Объемом совокупности* (выборочной или генеральной) называют число объектов этой совокупности. Например, если из 1000 деталей отобрано для обследования 100 деталей, то объем генеральной совокупности  $N = 1000$ , а объем выборки  $n = 100$ . Число объектов генеральной совокупности  $N$  значительно превосходит объем выборки  $n$ .

При составлении выборки можно поступать двумя способами: после того как объект отобран и над ним произведено наблюдение, он может быть возвращен (*повторная выборка*) либо не возвращен в генеральную совокупность (*бесповторная выборка*). На практике обычно пользуются бесповторным случайным отбором.

Для того чтобы по данным выборки можно было достаточно уверенно судить об интересующем признаке генеральной совокупности, необходимо, чтобы объекты выборки правильно его представляли. Другими словами, выборка должна правильно представлять пропорции генеральной совокупности. Это требование коротко формулируют так: выборка должна быть *репрезентативной* (представительной), т.е. каждый объект выборки

отобран случайно из генеральной совокупности, и все объекты имеют одинаковую вероятность попасть в выборку.

Обычно полученные наблюдаемые данные представляют собой множество расположенных в беспорядке чисел. Просматривая это множество чисел, трудно выявить какую-либо закономерность их варьирования (изменения). Для изучения закономерностей варьирования значений случайной величины опытные данные подвергают обработке.

**Пример 1.** На телефонной станции проводились наблюдения над числом  $X$  неправильных соединений в минуту. Наблюдения в течение часа дали следующие результаты:

3; 1; 3; 1; 4; 2; 2; 4; 0; 3; 0; 2; 2; 0; 2; 1; 4; 3; 3; 1; 4; 2; 2; 1; 1; 2; 1; 0; 3; 4; 1; 3; 2; 7; 2; 0; 0; 1; 3; 3; 1; 2; 4; 2; 0; 2; 3; 1; 2; 5; 1; 1; 0; 1; 1; 2; 2; 1; 1; 5.

Число  $X$  является дискретной случайной величиной, а полученные о ней сведения представляют собой статистические (наблюдаемые) данные.

Расположим приведенные данные в порядке неубывания и сгруппировав их, получим ранжированный ряд данных наблюдения.

0; 0; 0; 0; 0; 0; 0; 0; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 2; 3; 3; 3; 3; 3; 3; 3; 3; 3; 3; 4; 4; 4; 4; 4; 4; 5; 5; 7.

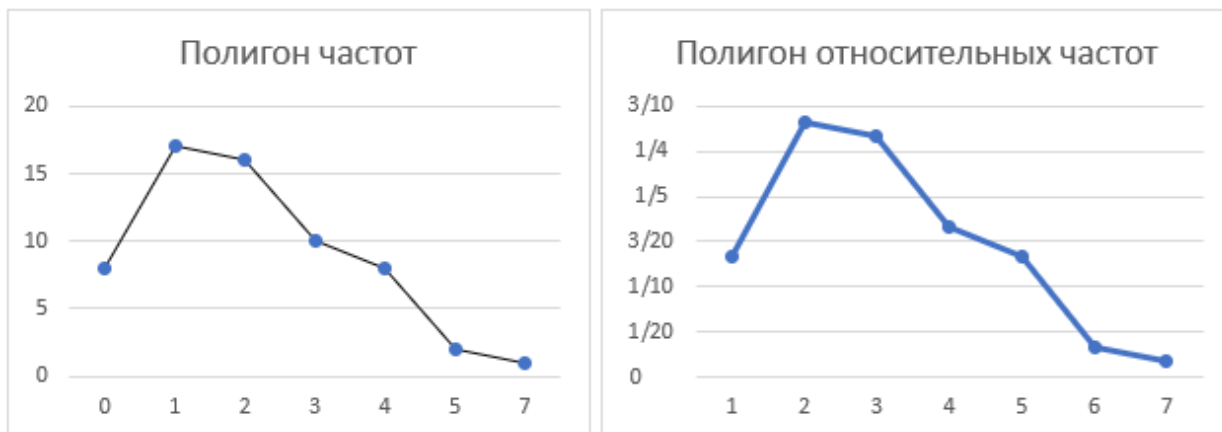
Из ряда чисел видно, что все 60 значений случайной величины разбиты на семь групп, в пределах каждой из которых все значения случайной величины одинаковы. Таким образом, имеется семь различных значений случайной величины: 0; 1; 2; 3; 4; 5; 7. Каждое такое значение называют *вариантой*, обозначают  $x_i$ . Для каждой группы сгруппированного ряда данных можно подсчитать их численность, т.е. определить сколько раз встречается соответствующий вариант в ряде наблюдений. Такие числа называют *частотой варианта* обозначают  $n_i$ . В ряде случаев представляет практический интерес относительная частота того или иного варианта, называемая *частостью*, обозначают  $w_i$ ,  $w_i = n_i / n$ .

Подсчитав частоты и частости для каждого варианта, наблюдаемые данные представляют в виде таблицы, которую называют дискретным вариационным рядом. В первой строке расположены варианты, во второй- соответствующие частоты или частости.

Для рассмотренного примера ряд имеет вид:

$x_i$	0	1	2	3	4	5	7
$n_i$	8	17	16	10	6	2	1
$w_i$	8/60	17/60	16/60	10/60	6/60	2/60	1/60

По данным дискретного вариационного ряда строят полигон частот (относительных частот): ломаную, отрезки которой соединяют точки  $(x_i; n_i)$  ( $(x_i; w_i)$ ).



Если изучаемая случайная величина является непрерывной, то ранжирование и группировка наблюдаемых значений зачастую не позволяют выделить характерные черты варьирования ее значений. Это объясняется тем, что отдельные значения случайной величины могут, как угодно, мало отличаться друг от друга, и поэтому в совокупности наблюдаемых данных одинаковые значения величины могут встречаться редко, а частоты вариантов мало отличаются друг от друга. Нецелесообразно также построение дискретного ряда для дискретной случайной величины, число возможных значений которой велико. В подобных случаях следует построить интервальный (вариационный) ряд распределения. Для построения такого ряда весь интервал варьирования наблюдаемых значений случайной величины разбивают на ряд частичных интервалов и подсчитывают частоту попадания значений величины в каждый частичный интервал.

**Пример 2.** При измерении диаметра валиков после шлифовки получены следующие результаты:

6,75; 6,77; 6,77; 6,73; 6,76; 6,74; 6,70; 6,75; 6,71; 6,72; 6,77; 6,79; 6,71; 6,78;  
 6,73; 6,70; 6,73; 6,77; 6,75; 6,74; 6,71; 6,70; 6,78; 6,76; 6,81; 6,69; 6,80; 6,80;  
 6,77; 6,68; 6,74; 6,70; 6,70; 6,74; 6,77; 6,83; 6,76; 6,76; 6,82; 6,77; 6,71; 6,74;  
 6,77; 6,75; 6,74; 6,75; 6,77; 6,72; 6,74; 6,80; 6,75; 6,80; 6,72; 6,78; 6,70; 6,75;  
 6,78; 6,78; 6,76; 6,77; 6,74; 6,74; 6,77; 6,73; 6,74; 6,77; 6,74; 6,75; 6,74; 6,76;  
 6,76; 6,74; 6,74; 6,74; 6,74; 6,76; 6,74; 6,72; 6,80; 6,76; 6,78; 6,73; 6,70; 6,76;  
 6,76; 6,77; 6,75; 6,78; 6,72; 6,76; 6,78; 6,68; 6,75; 6,73; 6,82; 6,73; 6,80; 6,81;  
 6,71; 6,82; 6,77; 6,80; 6,80; 6,70; 6,70; 6,82; 6,72; 6,69; 6,73; 6,76; 6,74; 6,77;  
 6,72; 6,76; 6,78; 6,78; 6,73; 6,76; 6,80; 6,76; 6,72; 6,76; 6,76; 6,70; 6,73; 6,75;  
 6,77; 6,77; 6,70; 6,81; 6,74; 6,73; 6,77; 6,74; 6,78; 6,69; 6,74; 6,71; 6,76; 6,76;  
 6,77; 6,70; 6,81; 6,74; 6,74; 6,77; 6,75; 6,80; 6,74; 6,76; 6,77; 6,77; 6,81; 6,75;  
 6,78; 6,73; 6,76; 6,76; 6,76; 6,77; 6,76; 6,80; 6,77; 6,74; 6,77; 6,72; 6,75; 6,76;  
 6,77; 6,81; 6,76; 6,76; 6,76; 6,80; 6,74; 6,80; 6,74; 6,73; 6,75; 6,77; 6,74; 6,76;  
 6,77; 6,77; 6,75; 6,76; 6,74; 6,82; 6,76; 6,73; 6,74; 6,75; 6,76; 6,72; 6,78; 6,72;  
 6,76; 6,77; 6,75; 6,78.

Просматривая результаты наблюдений  $n=200$ , находим, что наибольшим значением случайной величины  $x_{\text{наиб}}=6,83$ , а наименьшим  $x_{\text{наим}}=6,68$ . Найдем размах варьирования  $R$ :  $R = x_{\text{наиб}} - x_{\text{наим}} = 6,83 - 6,68 = 0,15$ .

Выберем число интервалов по формуле:  $k \approx 1 + 3,322 \cdot \lg n$ .  
 $k \approx 1 + 3,2 \cdot \lg 200 = 8$ , тогда длина частичного интервала  $h = R/k = 0,15/8 \approx 0,02$

За начало первого интервала рекомендуется брать величину  $x_{\text{нач}} = x_{\text{наим}} - 0,5h$ . В данном случае  $x_{\text{нач}} = 6,67$ .

Конец последнего интервала должен удовлетворять условию  $x_{\text{кон}} - h \leq x_{\text{наиб}} \leq x_{\text{кон}}$ .

Промежуточные интервалы получают, прибавляя к концу предыдущего интервала длину частичного интервала  $h$  ( $h=0,02$ ).

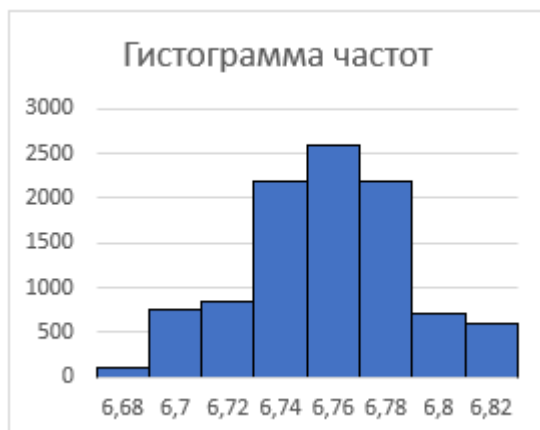
Просматривая результаты наблюдений, определяем, сколько значений признака попало в каждый конкретный интервал. При этом в интервал включают значения случайной величины, большие или равные нижней границе и меньшие верхней границы.

В таблице частота  $n_i$ , показывает, в скольких наблюдениях случайная величина приняла значения, принадлежащие тому или иному интервалу, причем нижний конец интервала входит в него, а верхний - нет. Такие частоты обычно называют интервальными, а их отношение к общему числу наблюдений - интервальными частотами. При вычислении интервальных частот округление результатов следует проводить таким образом, чтобы общая сумма частот была равна 1.

Для данного примера интервальный вариационный ряд имеет вид:

№	$x_i - x_{i+1}$	$n_i$	$w_i$	$n_i/h$	$w_i/h$
1	6,67-6,69	2	0,01	100	0,5
2	6,69-6,71	15	0,075	750	3,75
3	6,71-6,73	17	0,085	850	4,25
4	6,73-6,75	44	0,22	2200	11
5	6,75-6,77	52	0,26	2600	13
6	6,77-6,79	44	0,22	2200	11
7	6,79-6,81	14	0,07	700	3,5
8	6,81-6,83	12	0,06	600	3,0
	$\Sigma$	200	1		

По данным интервального ряда строят гистограмму частот или гистограмму относительных частот: ступенчатая фигура, состоящая из прямоугольников, основания которых - частичные интервалы, высоты равны отношению частоты к длине частичного интервала (плотность частоты) (частоты к длине частичного интервала (плотность частоты)).



Для гистограммы частот: площадь каждого прямоугольника равна частоте интервала, сумма площадей всех прямоугольников равна объему выборки.

Для гистограммы частостей: площадь каждого прямоугольника равна частоте интервала, сумма площадей всех прямоугольников равна 1.

Пусть известно статистическое распределение частот количественного признака  $X$ .

Эмпирической функцией распределения (функцией распределения выборки) называют функцию определяющую для каждого значения  $x$  относительную частоту события  $X < x$ , т.е.  $F^*(x) = n_x/n$ .

В отличие от эмпирической функции распределения выборки функцию распределения  $F(x)$  генеральной совокупности называют теоретической функцией распределения. Различие между эмпирической и теоретической функциями состоит в том, что теоретическая функция  $F(x)$  определяет вероятность события  $X < x$ , а эмпирическая функция определяет относительную частоту этого же события.

Эмпирическая функция обладает всеми свойствами  $F(x)$ :

- 1) ее значения принадлежат отрезку  $[0, 1]$ ;
- 2) неубывающая;
- 3) если  $x_i$  - наименьшая варианта, то  $F^*(x) = 0$  при  $x \leq x_i$   
если  $x_k$  - наибольшая варианта, то  $F^*(x) = 1$  при  $x > x_k$

Эмпирическая функция распределения выборки служит для оценки теоретической функции распределения генеральной совокупности.

**Пример 3.** Построить эмпирическую функцию по данному распределению выборки:

$x_i$	2	6	10
$m_i$	12	18	30

Объем выборки  $n = 12 + 18 + 30 = 60$ .  $x_{\text{наим}} = 2$ .

Значение  $X < 2$  не наблюдалось, значит  $F^*(x) = 0$  при  $x \leq 2$

Значение  $X < 6$  наблюдалось 12 раз, значит  $F^*(x) = 12/60 = 1/5$  при  $x \leq 6$

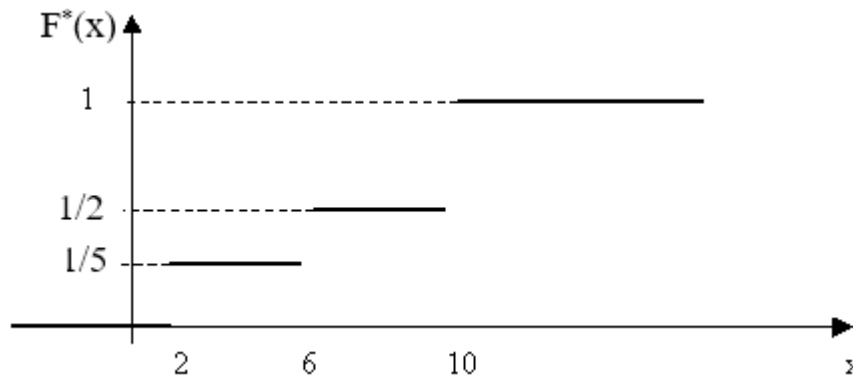
Значение  $X < 10$  наблюдалось  $12 + 18 = 30$  раз, значит  $F^*(x) = 30/60 = 1/2$  при  $x \leq 10$

Так как  $x_{\text{наиб}} = 10$ , то при  $X > 10$  наблюдались все значения, и  $F^*(x) = 1$  при  $x > 10$

Искомая эмпирическая функция имеет вид:

$$F^*(x) = \begin{cases} 0 & \text{при } x \leq 2 \\ 1/5 & \text{при } 2 < x \leq 6 \\ 1/2 & \text{при } 6 < x \leq 10 \\ 1 & \text{при } x > 10 \end{cases}$$

График строится так же, как и график интегральной функции распределения.

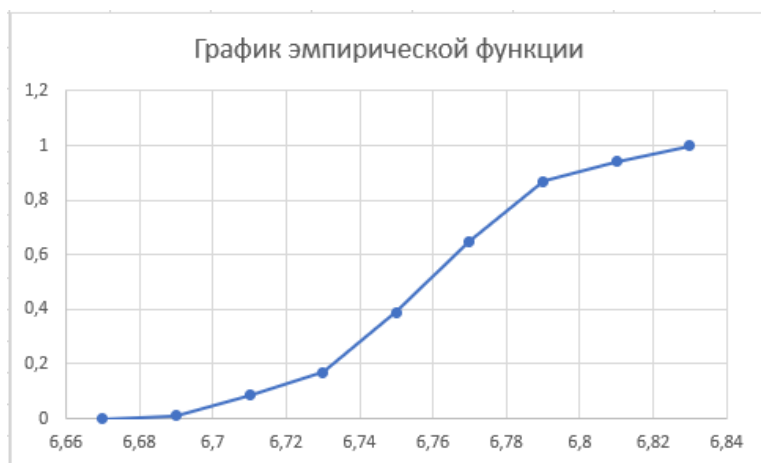


Если результаты наблюдений представлены в виде интервального вариационного ряда, то в качестве  $x$  принимают концы частичных интервалов и, пользуясь данным выше определением вычисляют значения эмпирической функции. Причем, при  $X < x_{\text{нач}}$   $F^*(x) = 0$ , а при  $X > x_{\text{кон}}$   $F^*(x) = 1$ .

Для рассмотренного примера получим таблицу:

$x$	6,67	6,69	6,71	6,73	6,75	6,77	6,79	6,81	6,83
$F^*(x)$	0	0,01	0,085	0,17	0,39	0,65	0,87	0,94	1

Так как таблица определяет функцию не полностью, то при изображении графика доопределяем функцию, соединяя точки графика, соответствующие концам интервалов, отрезками. График эмпирической функции для интервального вариационного ряда есть непрерывная линия.



Выборочным аналогом дифференциальной функции  $f(x)$  является функция  $f^*(x) = \frac{F^*(x+\Delta x) - F^*(x)}{\Delta x}$ , где  $F^*(x + \Delta x) - F^*(x)$  — есть частость попадания наблюдаемых значений СВ  $X$  в интервал  $[x; x + \Delta x)$ , следовательно,  $f^*(x)$  характеризует плотность частоты на этом интервале.

Выборочная дифференциальная функция  $f^*(x) = p_i/h$ , при  $x \leq x_{\text{нач}}$  и  $x \geq x_{\text{кон}}$   $f^*(x) = 0$ .

При построении графика выборочной функции плотности в качестве  $x$  принимают середину каждого частичного интервала. Удобно совмещать на одной координатной плоскости гистограмму частот с графиком выборочной плотности.

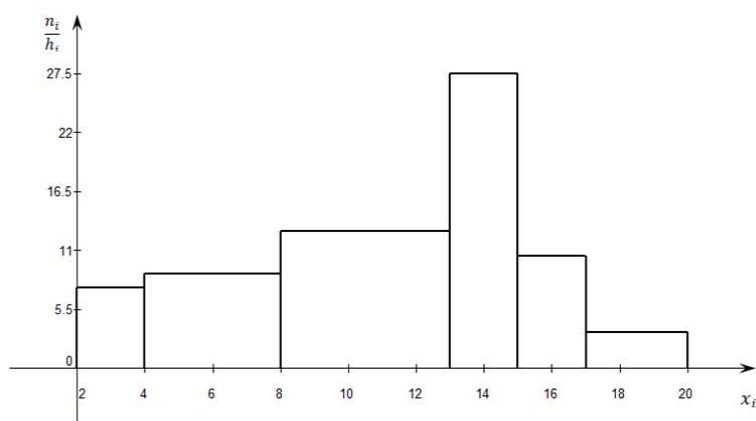
**Пример 4.** Построить гистограмму частот (случай неравных интервалов).

$x_i$	2-4	4-8	8-13	13-15	15-17	17-20
$n_i$	15	35	64	55	21	10

Решение

Вычислим плотности частоты:

Интервалы, $X_i - X_{i+1}$	$n_i$	Длина интервала, $h_i$	Плотность частоты, $n_i/h_i$
2 – 4	15	2	7.500
4 – 8	35	4	8.750
8 – 13	64	5	12.800
13 – 15	55	2	27.500
15 – 17	21	2	10.500
17 – 20	10	3	3.333
Итого	200	--	--



### Задачи для аудиторной работы

1. Построить полигон частот и полигон относительных частот

$x_i$	2	7	8	15	16	17
$n_i$	15	35	64	55	21	10

2. Построить гистограмму частот и относительных частот

$x_i$	2-5	5-8	8-11	11-14	14-17	17-20
$n_i$	15	35	64	55	21	10

### Задачи для самостоятельного решения

- По данной таблице составить интервальный вариационный ряд.
- По сгруппированным данным построить: полигон относительных частот; гистограмму относительных частот; график эмпирической функции распределения; на чертеже гистограммы её теоретический аналог  $f(x)$ ; на чертеже эмпирической функции  $F^*(x)$  её теоретический аналог  $F(x)$ .

1.

42	40	35	29	58	46	39	46	45	38	53	31	39	53	50	21	57	25	30	47
23	56	42	39	46	45	45	47	25	48	56	22	40	41	41	33	45	34	43	37
29	41	36	31	35	63	56	42	40	48	39	33	39	20	37	34	33	38	40	61
45	45	43	39	26	14	36	55	48	34	43	44	37	37	31	58	39	47	56	41
57	31	27	46	42	43	47	44	46	37	34	45	30	46	37	33	28	56	42	29

2.

73	85	108	55	61	95	93	67	86	96	101	87	90	75	92	93	95	94	87
77	59	98	68	94	78	86	84	82	87	70	67	74	70	80	75	71	89	85
70	75	69	89	61	101	91	70	73	86	75	79	79	68	74	87	79	72	67
92	69	86	96	70	69	69	85	81	84	92	95	61	65	83	69	92	74	68
69	87	79	74	75	73	81	68	81	75	81	78	95	76	72	82	74	69	75

3.

60	71	62	57	81	55	59	47	75	56	61	60	63	65	59	61	65	58	76	49
65	64	59	76	58	52	70	77	67	50	65	53	56	64	55	77	51	61	73	64
45	53	45	58	57	60	48	71	33	65	50	80	58	67	71	51	51	49	66	63
67	60	67	61	58	36	75	47	68	63	77	75	62	75	70	75	66	53	63	60
68	67	55	75	71	59	77	58	65	57	55	28	74	71	47	73	40	45	37	66

4.

21	13	28	19	20	14	24	23	18	15	32	14	15	20	16	17	14	22	21	24
7	22	17	17	26	22	21	21	14	23	24	18	25	18	20	21	20	22	7	31
18	14	22	17	5	20	20	11	17	19	19	3	15	16	19	7	25	13	20	15
16	12	19	16	16	22	21	7	14	21	20	26	17	14	14	14	10	26	12	9
12	11	15	19	13	15	2	6	21	9	23	16	16	21	11	14	19	19	28	12

5.

43	34	32	33	29	39	40	37	38	38	29	28	35	29	37	34	30	30	37	40
26	33	38	26	30	32	28	44	30	35	20	34	28	34	24	26	34	37	36	39
29	39	31	33	32	28	37	32	37	32	35	39	30	31	33	38	39	20	37	28
34	37	33	25	43	32	36	35	33	28	35	31	25	38	41	37	24	33	42	30
28	34	36	26	33	32	35	43	33	32	33	31	29	33	27	44	33	24	29	41

6.

21	13	28	19	20	14	24	23	18	15	32	14	15	20	16	17	14	22	21	24
7	22	17	17	26	22	21	21	14	23	24	18	25	18	20	21	20	22	7	31
18	14	22	17	5	20	20	11	17	19	19	3	15	16	19	7	25	13	20	15
16	12	19	16	16	22	21	7	14	21	20	26	17	14	14	14	10	26	12	9
12	11	15	19	13	15	2	6	21	9	23	16	16	21	11	14	19	19	28	12

7.

28	31	34	36	21	26	37	35	33	28	37	47	32	37	22	22	26	35	26	47
33	33	34	40	11	36	36	32	26	38	32	44	47	26	49	32	36	35	49	30
51	28	42	30	40	46	28	44	14	27	30	18	28	41	34	35	48	36	43	14
40	22	41	45	27	37	26	42	45	21	22	26	33	36	29	20	34	39	29	37
46	35	43	49	24	44	24	52	23	31	54	60	36	31	38	37	47	22	40	36

8.

32	46	29	7	32	40	36	37	41	44	36	37	49	37	41	23	29	23	25	39
29	45	31	14	29	31	36	53	38	37	35	28	29	44	43	44	20	31	32	32
31	45	42	38	35	29	39	33	40	26	42	43	36	39	36	32	50	37	23	49
26	43	34	30	46	65	41	16	24	27	58	34	47	36	39	24	46	30	8	22
19	26	37	24	38	33	38	22	25	27	44	30	45	28	38	26	24	33	49	40

9.

22	21	26	20	22	18	28	23	26	29	18	21	24	25	25	21	23	24	20	25
19	25	28	21	23	27	24	19	19	21	25	21	16	19	17	27	23	15	25	27
19	21	24	28	23	24	21	18	28	23	23	26	19	24	18	27	20	29	23	32
23	30	22	21	29	22	25	20	23	23	18	23	21	30	26	32	21	19	18	23
20	28	16	23	18	20	27	25	30	24	25	23	25	29	22	29	26	22	25	22

10.

47	45	56	44	47	50	62	63	3S	56	38	46	52	54	55	52	51	56	45	42
39	56	61	45	50	41	53	46	60	42	55	44	34	39	35	32	49	58	60	55
40	45	53	62	50	3S	44	51	53	62	51	5S	41	52	38	65	42	72	60	49
49	67	4S	46	63	42	55	51	49	49	39	49	46	67	58	41	45	50	71	39
44	62	34	50	38	56	61	53	43	66	54	51	54	64	48	48	57	46	65	55

### ***Контрольные вопросы***

1. Сформулировать определение генеральной совокупности, выборки.
2. Что называется частотой и относительной частотой (частостью) варианта?
3. Алгоритм составления дискретного вариационного ряда.
4. Полигон частот или частостей.
5. Алгоритм составления интервального вариационного ряда.
6. Гистограмма частот и гистограмма частостей, их геометрический смысл.
7. Эмпирическая функция распределения и ее графики для дискретного и интервального вариационных рядов.
8. Выборочная дифференциальная функция распределения (выборочная плотность) и ее график.