

Генеральная совокупность.

Выборка. Вероятностная модель эксперимента

Вероятностная модель ставит в соответствие результатам наблюдений

$$x_1, x_2, \dots, x_n \quad (1)$$

последовательность случайных величин

$$X_1, X_2, \dots, X_n . \quad (2)$$

Предполагается, что случайные величины X_1, X_2, \dots, X_n независимы и имеют одно и то же распределение с функцией распределения $F(x)$.

Полагают, что наблюдения (1) являются значениями величин (2) при осуществлении вероятностного эксперимента. Несмотря на различие объектов (1) и (2), в математической статистике принято называть и то и другое *выборкой из генеральной совокупности*.

Количество наблюдений n называется объемом выборки.

Статистическим закономерностям числовых данных реального эксперимента (1) отвечают вероятностные утверждения для случайных величин (2).

Рассмотрим эксперимент, состоящий в n -кратном бросании игральной кости и построим вероятностную модель эксперимента.

Результаты наблюдений имеют вид

$$x_1, x_2, \dots, x_n ,$$

где $x_i = \overline{1,6}, i = \overline{1,n}$.

Вероятностную модель представим следующим образом:

$$X_1, X_2, \dots, X_n ,$$

где $X_i, i = \overline{1,n}$ - независимые случайные величины, имеющие дискретное распределение и принимающие значения 1,2,3,4,5,6 с равными вероятностями $\frac{1}{6}$.

Рассмотрим урну с N шарами, занумерованными числами от 0 до $N-1$. Опыт заключается в n - кратном извлечении шара «наугад» с

последующим его возвращением в урну. Построим вероятностную модель опыта.

Все возможные исходы опыта могут быть представлены множеством из N^n последовательностей

$$x_1, x_2, \dots, x_n,$$

где $x_i = \overline{0, N-1}, i = \overline{1, n}$.

Вероятностная модель опыта будет иметь вид

$$X_1, X_2, \dots, X_n,$$

где $X_i, i = \overline{1, n}$ - независимые случайные величины, имеющие дискретное распределение и принимающие значения $0, 1, \dots, N-1$ с равными вероятностями $\frac{1}{N}$.

Если число наблюдений велико, то необходима группировка данных, которая выполняется следующим образом. Выделяются все несовпадающие между собой значения выборки, для каждого из таких значений определяется общее число повторений. В результате получаются два ряда чисел. Первый ряд содержит все несовпадающие значения из выборок, расположенные порядке возрастания. Числа второго ряда показывают количество повторений каждого из этих значений в выборке.

Количество повторений соответствующего элемента x_i в выборке называют **частотой** этого элемента и обозначают n_i . **Относительной частотой** элемента x_i называют число $\frac{n_i}{n}$, где n - объем выборки.

$$\sum_i n_i = n, \sum_i \frac{n_i}{n} = 1.$$

Пример. Подсчет числа деревьев определенного вида на 60 делянках дал следующие результаты

15 13 15 14 16 14 12 14 14 15 13 11 13 15 14
 13 15 14 12 14 14 15 15 12 12 13 15 16 14 13
 14 13 15 14 13 14 15 14 15 14 14 13 14 15 13
 14 13 13 11 12 14 13 12 11 15 15 13 13 13 14

Провести группировку данных.

Решение. Группировку осуществим следующим образом. В первую строку таблицы впишем все возможные значения из выборки в порядке возрастания. Просматривая подряд заданную выборку и, считая повторения каждого элемента, заполняем вторую строку. Отсюда сгруппированный ряд имеет следующий вид:

x_i	11	12	13	14	15	16
n_i	3	6	16	19	14	2

Такая форма записи сгруппированного ряда называется **статистическим рядом**.

Если выборка имеет большое число различных элементов, то группировка заключается в том, что диапазон выборки делится на определенное число частей (интервалов), а затем подсчитывается число значений выборки, попавшее в каждый из интервалов. При такой группировке возникает вопрос о длине интервалов Δx и расположении границ. Решение этого вопроса рассмотрим на примере.

Пример. Получены следующие данные измерения толщины кремниевой подложки в микронах

5,39	5,43	5,49	5,42	5,45	5,44	5,64
5,42	5,52	5,35	5,45	5,37	5,54	5,18
5,38	5,45	5,48	5,32	5,48	5,66	5,46
5,47	5,26	5,26	5,44	5,46	5,43	5,61
5,51	5,33	5,55	5,58	5,51	5,42	5,47
5,30	5,43	5,46	5,50	5,29	5,44	5,36
5,40	5,50	5,41	5,56	5,42	5,34	5,57
5,40	5,44	5,55	5,44	5,69	5,33	5,39
5,28	5,47	5,37	5,50	5,60	5,41	5,58
5,43	5,52	5,45	5,37	5,45	5,54	5,44
5,46	5,48	5,54	5,57	5,38	5,40	
5,53	5,34	5,32	5,50	5,46	5,23	
5,55	5,36	5,52	5,44	5,52	5,45	
5,47	5,59	5,39	5,28	5,43	5,47	
5,24	5,45	5,62	5,31	5,41	5,40	

Произвести группировку данных.

Решение. Самое простое предложение - разделить разность между наибольшим ($x_{\max} = x_{(n)} = 5,69$) и наименьшим ($x_{\min} = x_{(1)} = 5,18$) значениями на принятое число частей (например, $k=11$), после чего границы интервалов находятся сразу. В нашем примере $5,69 - 5,18 = 0,51$ и $\Delta x = \frac{0,51}{11} = 0,04636\dots$ Так как здесь $x_{\max} - x_{\min}$ не делится без остатка на k (в пределах принятой точности), то производится округление длины интервала в сторону увеличения. В противном случае общая длина интервалов уменьшилась бы так, что крайние значения выборки не попали бы в него.

Если какое-либо значение попадает на границу интервалов, то его относят к правому интервалу. Так, если границы интервалов: $\dots; 5,30; 5,35; 5,40; 5,45; \dots$, то значение 5,35 относят к интервалу 5,35 - 5,40. Иными

словами этот интервал содержит значения x удовлетворяющие условию $5,35 \leq x < 5,40$.

Принимая $\Delta x = 0,05$ и $k = 11$, получим, что весь диапазон выборки имеет длину $0,05 \cdot 11 = 0,55$.

Следовательно, в качестве границ диапазона значений можно принять либо $5,175 - 5,725$ (расширив диапазон в сторону больших значений, случай а)), либо $5,145 - 5,695$ (расширив его в сторону меньших значений, случай б)), либо какие-нибудь промежуточные значения, например $5,165 - 5,715$. Чтобы пояснить, как отразится на виде распределения тот или иной выбор границ диапазона выборки, приведем ниже результаты группировки для двух рассмотренных случаев.

Границы интервалов	Метки	Частота
Случай а)		
5,175 - 5,225	I	1
5,225 - 5,275	IIII	4
5,275 - 5,325	IIII II	7
5,325 - 5,375	IIII III I	11
5,375 - 5,425	IIII III III I	16
5,425 - 5,475	IIII III III III III III III	30
5,475 - 5,525	IIII III III	14
5,525 - 5,575	IIII III	8
5,575 - 5,625	IIII I	6
5,625 - 5,675	II	2
5,675 - 5,725	I	1

Случай б)		
5,145 - 5,195	I	1
5,195 - 5,245	II	2
5,245 - 5,295	III	5
5,295 - 5,345	III III	8
5,345 - 5,395	III III II	12
5,395 - 5,445	III III III III III	23
5,445 - 5,495	III III III III II	22
5,495 - 5,545	III III IIII	14
5,545 - 5,595	III II	7
5,595 - 5,645	IIII	4
5,645 - 5,695	II	2

Несмотря на видимую несхожесть, полученные ряды отражают одно и то же фактическое распределение.

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$:

$$F^*(x) = \frac{n_x}{n},$$

где n_x – число вариант, меньших x ; n – объем выборки.

Пример. Построить эмпирическую функцию распределения вероятностей по выборке, представленной в таблице 2.

Таблица 2. Сгруппированная выборка.

x_i	2	5	8	12
$\frac{n_i}{n}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{4}{9}$

Решение. Эмпирическая функция распределения делает скачок в

точке 2, величина скачка равна $\frac{2}{9}$, следующий скачок происходит в точке 5, его величина скачка равна $\frac{1}{9}$, очередные скачки происходят в точках 8 и 12 и имеют величины соответственно $\frac{2}{9}$ и $\frac{4}{9}$. То есть

$$F_n(x) = \begin{cases} 0, & x < 2 \\ \frac{2}{9}, & 2 \leq x < 5 \\ \frac{1}{3}, & 5 \leq x < 8 \\ \frac{5}{9}, & 8 \leq x < 12 \\ 1, & x \geq 12 \end{cases}$$

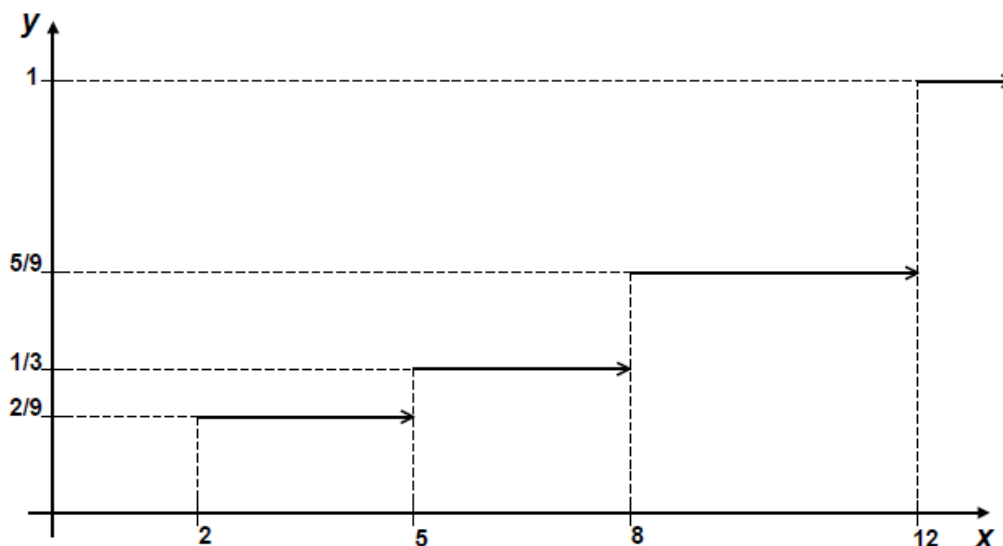


Рис. 2. График эмпирической функции распределения.

Пример. Найти эмпирическую функцию распределения вероятностей по данному распределению выборки, представленной в таблице 3.

Таблица 3. Сгруппированная выборка.

x_i	1	4	6
n_i	10	15	25

Решение. Найдем объём выборки: $n = 10 + 15 + 25 = 50$. Наименьший элемент выборки равен 1, поэтому $F_n(x) = 0$ при $x < 1$.

Значение $x < 4$, а именно $x_1 = 1$, наблюдалось 10 раз, следовательно,

$$F_n(x) = \frac{10}{50} = 0.2 \quad \text{при } 1 \leq x < 4$$

Значение $x < 6$, а именно $x_1 = 1$ и $x_2 = 4$, наблюдалось $10+15=25$ раз, следовательно,

$$F_n(x) = \frac{25}{50} = 0.5 \quad \text{при } 4 \leq x < 6$$

Так как $x = 6$ - наибольшее значение из выборки, то $F_n(x) = 1$ при $x \geq 6$. В результате эмпирическая функция распределения имеет вид:

$$F_n(x) = \begin{cases} 0 & \text{при } x < 1 \\ 0,2 & \text{при } 1 \leq x < 4 \\ 0,5 & \text{при } 4 \leq x < 6 \\ 1 & \text{при } x \geq 6 \end{cases}$$

Задача 1. Объем выборки равен $n = 50$. Найти x .

x_i	11	12	13	14	15	16
n_i	3	6	x	19	14	2

Решение.

$$\sum_i n_i = n \Rightarrow x = 50 - 3 - 6 - 19 - 14 - 2 = 6.$$

Задача 2. Объем выборки равен $n = 100$. Найти относительную частоту x и частоту n_4 элемента выборки 4.

x_i	-1	0	3	4	9	11
$\frac{n_i}{n}$	0.05	0.15	0.3	x	0.2	0.1

Решение. $\sum_i n_i = n \Rightarrow \sum_i \frac{n_i}{n} = 1 \Rightarrow x = 1 - 0.05 - 0.15 - 0.3 - 0.2 - 0.1 = 0.2$,
 $n_4 = 100 \cdot 0.2 = 20$.

Графические характеристики

Полигон — это ломаная линия, проходящая через середины верхних границ прямоугольников гистограммы (соединяющая точки $(x_i^*; \frac{n_i}{nh}$, где x_i^* — середина i -го интервала). Полигон и гистограмма являются статистическими аналогами теоретической плотности.

Кумулята — это ломаная линия, соединяющая точки $(x_i; \sum_1^{i-1} \frac{n_j}{n})$. Кумулята дает представление о графике функции распределения.

Для нахождения приближенных значений выборочных медианы, моды и квантилей по группированной выборке применяют интерполяционные формулы.

Медианным называется интервал, в котором накопленная сумма частот впервые достигает $\frac{1}{2}$.

Выборочной группированной медианой называется значение m_e^* :

$$m_e^* = x_e + \frac{n/2 - (n_1 + \dots + n_{m_e-1})}{n_{m_e}} \cdot h,$$

где n — объем выборки, h — длина интервала группировки, x_e — левая граница медианного интервала, n_i — численность i -го интервала, n_{m_e} — численность медианного интервала.

Модальным называется интервал, имеющий наибольшую численность.

Выборочной группированной модой называется значение m_0^* :

$$m_0^* = x_0 + h \cdot \frac{n_{m_0} - n_{m_0-1}}{2n_{m_0} - n_{m_0-1} - n_{m_0+1}},$$

где x_0 — левая граница модального интервала, n_{m_0} — численность модального интервала, n_{m_0-1} , n_{m_0+1} — численности интервалов слева и справа от модального.

Квантильным порядка q интервалом называется интервал, в котором сумма накопленных частот впервые достигает значения q .

Выборочной группированной квантилью называется значение x_q^* :

$$x_q^* = x_{(q)} + h \cdot \frac{nq - (n_1 + \dots + n_{(q)-1})}{n_{(q)}},$$

где $x_{(q)}$ — левая граница квантильного интервала, $n_{(q)}$ — численность квантильного интервала, $n_1, \dots, n_{(q)-1}$ — численности интервалов, предшествующих квантильному.

Пример. Вычислить выборочное среднее, выборочную дисперсию и исправленную выборочную дисперсию.

Группы (интервал)	[-30; -10]	(-10,0]	(0,15]	(15,42]
Число элементов в	28	34	25	13
Частота $\frac{n_i}{n}$	0,28	0,34	0,25	0,13

Решение. Вычислим высоты прямоугольников (ординаты гистограммы) по формуле $y_i = \frac{n_i}{nh_i}$, где h_i — длина соответствующего

интервала

$$y_1 = \frac{0,28}{20} = 0,014 \quad \text{для всех } x \in [-30; -10]$$

$$y_2 = \frac{0,34}{10} = 0,034 \quad \text{для всех } x \in (-10; 0]$$

$$y_3 = \frac{0,25}{15} = 0,016 \quad \text{для всех } x \in (0; 15]$$

$$y_4 = \frac{0,13}{27} = 0,0048 \quad \text{для всех } x \in (15; 42]$$

Гистограмма приведена на рис.4.

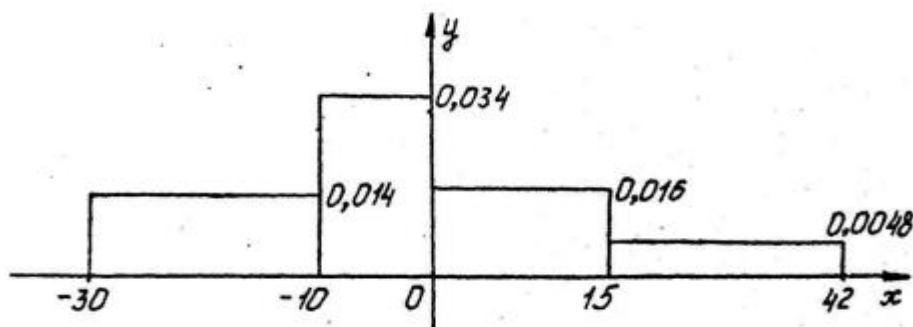


Рис.4. Гистограмма.

Распределение признака в вариационном ряду по накопленным частотам (частостям) изображается с помощью кумуляты.

Кумулята или кумулятивная кривая в отличие от полигона строится по накопленным частотам или частостям. При этом на оси абсцисс помещают значения признака, а на оси ординат — накопленные частоты или частости (рис. 3).



Рис. 3. Кумулята распределения домохозяйств по размеру

4. Рассчитаем накопленные частоты:

Накопленная частота первого интервала рассчитывается следующим образом: $0 + 4 = 4$, для второго: $4 + 12 = 16$; для третьего: $4 + 12 + 8 = 24$ и т.д.

Размер заработной платы руб в месяц X_j	Численность работников чел. f_j	Накопленные частоты S
до 5000	4	4
5000 — 7000	12	16
7000 — 10000	8	24
10000 — 15000	6	30
Итого:	30	-

При построении кумуляты накопленная частота (частость) соответствующего интервала присваивается его верхней границе:



Основные характеристики выборки:

\bar{x}_B – выборочная средняя;

D_B – выборочная дисперсия;

σ_B – выборочное среднее квадратичное отклонение;

S^2 – исправленная дисперсия.

$$\bar{x}_B = \frac{\sum_{i=1}^k x_i n_i}{n} = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \dots + x_k n_k}{n},$$

$$D_B = \frac{\sum_{i=1}^k (x_i - \bar{x}_B)^2 n_i}{n} = \frac{(x_1 - \bar{x}_B)^2 n_1 + (x_2 - \bar{x}_B)^2 n_2 + \dots + (x_k - \bar{x}_B)^2 n_k}{n},$$

$$D_B = \overline{x_B^2} - \bar{x}_B^2, \text{ где } \overline{x_B^2} = \frac{x_1^2 n_1 + x_2^2 n_2 + \dots + x_k^2 n_k}{n},$$

$$\sigma_B = \sqrt{D_B}, \quad S^2 = \frac{n}{n-1} D_B.$$

Пример. Вычислить выборочное среднее, выборочную дисперсию и исправленную выборочную дисперсию.

x_i	-5	-4	-1	0	1	2
n_i	30	20	10	30	20	10

Решение.

Объем

выборки

равен

$$n = \sum_{i=1}^6 n_i = 30 + 20 + 10 + 30 + 20 + 10 = 120.$$

$$\bar{x} = \frac{\sum_{i=1}^6 n_i x_i}{n} = -\frac{5}{4} - \frac{4}{6} - \frac{1}{12} + \frac{1}{6} + \frac{2}{12} = -\frac{5}{3},$$

$$s_0^2 = \frac{\sum_{i=1}^6 n_i x_i^2}{n} - \bar{x}^2 = \frac{25}{4} + \frac{16}{6} + \frac{1}{12} + \frac{1}{6} + \frac{4}{12} - \frac{25}{9} \approx 6.722, \quad s_1^2 = \frac{n}{n-1} s_0^2,$$

$$s_1^2 = \frac{120}{119} \cdot 6.722 \approx 6.778$$

Ответ. $\bar{x} = -\frac{5}{3}, s_0^2 \approx 6.722, s_1^2 \approx 6.778.$

Пример. Вычислить \bar{x} и s_1^2 (исправленную выборочную дисперсию) по выборке 10, 12, 14.

Решение.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$s_0^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

$$\bar{x} = \frac{\sum_{i=1}^3 x_i}{n} = \frac{10 + 12 + 14}{3} = 12,$$

$$s_1^2 = \frac{1}{2} \left((10-12)^2 + (12-12)^2 + (14-12)^2 \right) = 4.$$

Пример. Найти размах, моду и медиану вариационного ряда

10;11;11;11;11;13;15;16;16;16;16;35.

Решение. Размахом вариационного ряда называется разность между его максимальным элементом и минимальным. Модой называется элемент выборки, встречающийся с наибольшей частотой. В качестве медианы в выборке объема $2n+1$ берется значение $x_{(n+1)}$ в вариационном ряде. Если объем выборки равен $2n$, то в качестве оценки медианы берется $\frac{1}{2}(x_{(n)} + x_{(n+1)})$.

Размахом является $35-10=25$. В выборке две моды – 11 и 16. Объем выборки равен 12, поэтому медиана равна $\frac{1}{2}(13+15)=14$.

Пример. Дана выборка 14;18;22. Найти несмещенные оценки математического ожидания и дисперсии.

Решение. Несмещенными оценками математического ожидания и дисперсии являются соответственно выборочное среднее и исправленная выборочная дисперсия.

$$\bar{x} = \frac{\sum_{i=1}^3 x_i}{n} = \frac{14+18+22}{3} = 18,$$

$$s_1^2 = \frac{1}{2}((18-14)^2 + (18-18)^2 + (22-18)^2) = 16.$$

Пример.

1. Произвести группировку выборки:

87,8	104,5	90,9	92,4	74,3	71,7	75,0	92,0	76,0	82,2
86,3	85,0	75,0	91,5	105,7	112,4	80,8	74,5	86,6	95,8
100,4	109,7	52,7	96,6	87,6	101,4	103,4	90,6	88,0	79,9
91,6	84,2	108,6	77,5	89,2	82,6	90,0	86,1	80,3	92,8
103,8	84,5	90,6	113,5	101,1	113,7	94,3	90,7	70,7	93,5
96,8	111,5	103,8	106,8	80,5	91,8	82,2	86,9	100,3	100,1
89,4	84,7	93,4	101,3	118,7	99,3	105,0	92,7	96,7	82,8
100,9	81,3	96,1	84,6	86,3	83,7	84,6	80,7	102,3	104,2
89,4	90,9	89,6	66,5	120,4	100,4	86,8	70,4	91,9	98,3
111,7	90,2	87,9	81,1	88,1	103,3	85,0	69,1	82,2	101,8
80,4	77,7	79,3	96,2	94,1	87,6	104,2	81,4	81,6	115,7
84,2	93,2	112,7	86,8	79,8	89,8	88,1	110,9	109,0	84,8
82,6	89,1	88,9	97,9	78,0	87,5	68,1	107,7	95,5	88,6

◀ 1. Упорядочим выборку (получим вариационный ряд).

52,7	66,5	68,1	69,1	70,4	70,7	71,7	74,3	74,5	75,0
75,0	76,0	77,5	77,7	78,0	79,3	79,8	79,9	80,3	80,4
80,5	80,7	80,8	81,1	81,3	81,4	81,6	82,2	82,2	82,2
82,6	82,6	82,8	83,7	84,2	84,2	84,5	84,6	84,6	84,7
84,8	85,0	85,0	86,1	86,3	86,3	86,6	86,8	86,8	86,9
87,5	87,6	87,6	87,8	87,9	88,0	88,1	88,1	88,6	88,9
89,1	89,2	89,4	89,4	89,6	89,8	90,0	90,2	90,6	90,6
90,7	90,9	90,9	91,5	91,6	91,8	91,9	92,0	92,4	92,7
92,8	93,2	93,4	93,5	94,1	94,3	95,5	95,8	96,1	96,2
96,6	96,7	96,8	97,9	98,3	99,3	100,1	100,3	100,4	100,4
100,9	101,1	101,3	101,4	101,8	102,3	103,3	103,4	103,8	103,8
104,2	104,2	104,5	105,0	105,7	106,8	107,7	108,6	109,0	109,7
110,9	111,5	111,7	112,4	112,7	113,5	113,7	115,7	118,7	120,4

2. Минимальный элемент выборки равняется $x_{\min} = 52,7$, а максимальный $x_{\max} = 120,4$.

Определим сначала число интервалов k . Рекомендуется брать такое k , что $2^{k-1} \sim n$.

В данном примере $n = 130$.

$$2^7 = 128 \sim 130; \quad k - 1 = 7, \quad k = 8.$$

Выберем число интервалов $k = 8$.

3. Определим длину интервала h .

$$h = \frac{x_{\max} - x_{\min}}{k}.$$

Находим

$$h = \frac{120,4 - 52,7}{8} \approx 8,4567.$$

4. Найдем границы интервалов группировки $x_0 = x_{\min}$, $x_i = x_0 + hi$, $i = 1, \dots, k$:

$$x_0 = 52,7, \quad x_1 = 61,2, \quad x_2 = 69,6, \quad x_3 = 78,1, \quad x_4 = 86,5,$$

$$x_5 = 95,0, \quad x_6 = 103,4, \quad x_7 = 111,9, \quad x_8 = 120,4.$$

5. Составим таблицу группировки и внесем границы интервалов в столбец «Интервал»:

№	Интервал	n_i	$\frac{n_i}{n}$	$\sum_1^i \frac{n_j}{n}$
1	[52,7 – 61,2)			
2	[61,2 – 69,6)			
3	[69,6 – 78,1)			
4	[78,1 – 86,5)			
5	[86,5 – 95,0)			
6	[95,0 – 103,4)			
7	[103,4 – 111,9)			
8	[111,9 – 120,4)			

6. Подсчитаем, сколько элементов выборки попало в каждый интервал, и заполним в таблице столбец «Численность n_i »:

№	Интервал	n_i	$\frac{n_i}{n}$	$\sum_1^i \frac{n_j}{n}$
1	[52,7 – 61,2)	1		
2	[61,2 – 69,6)	3		
3	[69,6 – 78,1)	11		
4	[78,1 – 86,5)	31		
5	[86,5 – 95,0)	40		
6	[95,0 – 103,4)	22		
7	[103,4 – 111,9)	15		
8	[111,9 – 120,4)	7		

По столбцу численностей рассчитаем остальные столбцы таблицы:

№	Интервал	n_i	$\frac{n_i}{n}$	$\sum_1^i \frac{n_j}{n}$
1	[52,7 – 61,2)	1	0,008	0,008
2	[61,2 – 69,6)	3	0,023	0,031
3	[69,6 – 78,1)	11	0,085	0,115
4	[78,1 – 86,5)	31	0,238	0,354
5	[86,5 – 95,0)	40	0,308	0,662
6	[95,0 – 103,4)	22	0,169	0,831
7	[103,4 – 111,9)	15	0,115	0,946
8	[111,9 – 120,4)	7	0,054	1

2. По группированной выборке, полученной выше постройте гистограмму, полигон и кумуляту, найти выборочную медиану, выборочную моду

Покажем, как найти выборочную медиану

№	Интервал	n_i	$\frac{n_i}{n}$	$\sum_1^i \frac{n_j}{n}$
1	[52,7 – 61,2)	1	0,008	0,008
2	[61,2 – 69,6)	3	0,023	0,031
3	[69,6 – 78,1)	11	0,085	0,115
4	[78,1 – 86,5)	31	0,238	0,354
5	[86,5 – 95,0)	40	0,308	0,662
6	[95,0 – 103,4)	22	0,169	0,831
7	[103,4 – 111,9)	15	0,115	0,946
8	[111,9 – 120,4)	7	0,054	1

◀ Медианным является интервал № 5, так как в нем впервые накопленная сумма частот, равная 0,662, достигает $\frac{1}{2}$.

$$m_e^* = 86,5 + \frac{130/2 - (1 + 3 + 11 + 31)}{40} \cdot 8,4567 \approx 90,517. \blacktriangleright$$

Задача № 1

Построить гистограмму и эмпирическую функцию распределения вероятностей. Найти выборочное среднее, выборочную дисперсию, исправленную выборочную дисперсию.

1.

x_i	-11	-8	-6	-4	-3	-2
n_i	50	20	30	20	40	50

2.

x_i	-7	-3	-1	0	4	5
n_i	90	50	10	30	40	50

3.

x_i	-4	-1	1	2	4	5
n_i	40	10	20	20	20	50

4.

x_i	1	3	4	6	7	13
n_i	60	20	40	30	40	80

Задача №2

1. Произвести группировку выборки

20, 2; 19, 2; 16, 9; 19, 3; 17, 1; 17, 8; 16, 6; 16, 3; 15, 2; 18, 0; 16, 8; 20, 0;
17, 7; 16, 6; 19, 0; 17, 5; 17, 8; 20, 6; 17, 2; 18, 0; 17, 1; 18, 4; 17, 4; 15, 8;
19, 4; 17, 8; 19, 8; 19, 6; 16, 3; 20, 0; 17, 4; 19, 3; 19, 3; 16, 5; 18, 8; 17, 2;
18, 7; 18, 6; 19, 2; 16, 2; 18, 2; 17, 4.

2. По группированной выборке, полученной выше постройте гистограмму, полигон и кумуляту, найти выборочную медиану, выборочную моду