15. Основы математической статистики

Пусть необходимо изучить совокупность однородных объектов относительно некоторого признака, характеризующего эти объекты. Данная совокупность объектов называется *генеральной совокупностью.* Если число объектов (*объем*) генеральной совокупности достаточно большое, то из нее случайным образом выбирается для изучения ограниченное число объектов, называемое *выборочной совокупностью* или *выборкой*. Выборка представляет собой ряд чисел, а если их расположить упорядоченно (в порядке возрастания), то получится *вариационный ряд.*

Основными статистическими характеристиками выборки являются:

размах выборки – разность между максимальным и минимальным значениями элементов выборки;

мода ряда чисел – наиболее часто встречающееся число в данном ряду;

медиана вариационного ряда чисел – число, находящееся посередине в вариационном ряде (если количество чисел нечетное) или среднее арифметическое двух чисел, записанных посередине, в вариационном ряде (если количество чисел четное).

Задача 15.1. Найти среднее арифметическое значение, моду, медиану и размах вариационного ряда 6, 7, 7, 7, 8, 9, 9, 10, 12.

Решение

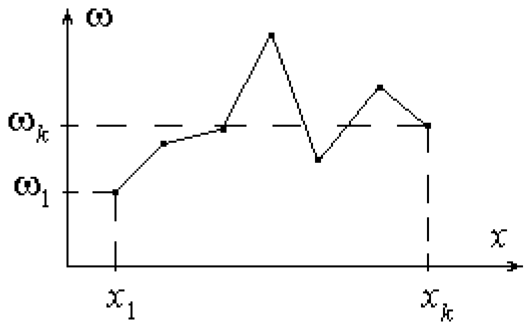
Количество чисел – 9. Тогда среднее арифметическое значение $\frac{1}{9}(6+7+7+7+8+9+9+10+12) = \frac{25}{3}$, мода – 7, медиана – 8, размах 12–6=6.

Пусть из генеральной совокупности отобрана выборка объема n, причем значение x_1 наблюдалось n_1 раз, значение x_2 – n_2 раз,…, значение x_k – n_k раз ($x_1 < x_2 < ... < x_k$). Результаты наблюдений записываем в таблицу, называемую *статистическим рядом (распределением) выборки*

x_i	x_1	x_2	$\dots$	x_k
n_i	n_1	n_2	$\dots$	n_k
ω_i	ω_1	ω_2	$\dots$	ω_k

В таблице x_i называются *вариантами,* n_i *– частотами наблюдений,* $\omega_i = \frac{n_i}{n}$ *– относительными частотами наблюдений,* k *– числом разрядов.*

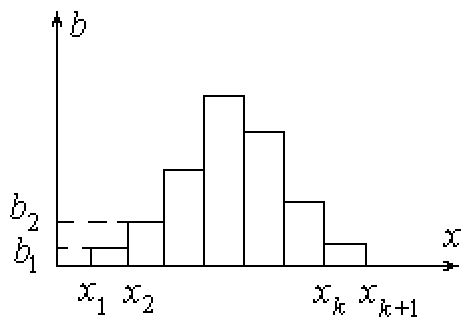
Очевидно, что $\sum_{i=1}^{k} n_i = n$, $\sum_{i=1}^{k} \omega_i = 1$. Для иллюстрации статистического ряда строим *полигон относительных частот* – ломаную, последовательно соединяющую точки (x_i, ω_i).



В случае изучения непрерывно распределенного признака объектов выборки строим *интервальный статистический ряд (распределение) выборки.* Весь диапазон наблюдений разбиваем на *частичные интервалы* (x_i, x_{i+1}) (желательно одинаковой длины h) и находим n_i – число наблюдений, попавших в соответствующий интервал.

(x_i, x_{i+1})	(x_1, x_2)	(x_2, x_3)	$\dots$	(x_k, x_{k+1})
n_i	n_1	n_2	$\dots$	n_k
ω_i	ω_1	ω_2	$\dots$	ω_k

В таблице n_i называют *частотами интервалов,* $\omega_i = \frac{n_i}{n}$ – *относительными частотами интервалов,* k – *числом разрядов* (число интервалов). Для иллюстрации статистического ряда строим *гистограмму* (примерный вид на рисунке справа) – ступенчатую фигуру, состоящую



из прямоугольников с основаниями (x_i, x_{i+1}) и высотами $b_i = \frac{\omega_i}{h}$.

Обычно число k берут не произвольно, а согласуют с объемом выборки, чтобы гистограмма помогала определять вид распределения. Для вычисления k пользуются формулой Стерджесса $k \approx 1 + 3{,}322 \cdot \lg(n)$.

Так, например, при объеме выборки $n = 100$ получаем $k \approx 7{,}644$, тогда число интервалов берут 7 или 8.

Задача 15.2. Построить статистический ряд распределения для выборки $\{6, 7, 7, 7, 7, 8, 8, 9, 10, 10\}$.

Решение

Объем выборки $n = 10$. Получаем статистический ряд

x_i	6	7	8	9	10
n_i	1	4	2	1	2
ω_i	0,1	0,4	0,2	0,1	0,2

Чтобы изучить генеральную совокупность относительно некоторого признака, необходимо знать распределение, которое имеет этот признак, следовательно, нужно оценить параметры этого закона распределения.

Для выборки объема n оценкой математического ожидания случайной величины является *выборочная средняя* (она равна среднему арифметическому значений выборки) $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Используя данные статистического ряда выборки, выборочную среднюю можно рассчитывать по формуле $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} x_i \omega_i$, где k – число разрядов. Эта оценка является *несмещенной* (несмещенность – свойство, характеризующее зависимость оценки от случайного состава выборки). Оценкой дисперсии случайной величины служит *выборочная дисперсия*

$$D_в = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^{\,2}\right) - \overline{x}^2 = \sum_{i=1}^{k} x_i^{\,2}\omega_i - \overline{x}^2 .$$

Тогда *выборочное среднеквадратическое отклонение* $\sigma_в = \sqrt{D_в}$ является оценкой среднеквадратического отклонения. Выборочная дисперсия не является несмещенной оценкой, поэтому для небольшого объема выборки ($n \le 30$) используют другую оценку $s^2 = \frac{n}{n-1} D_в$, которая называется *исправленной выборочной дисперсией* и является несмещенной оценкой дисперсии. Соответственно, $s = \sqrt{s^2}$ – *исправленное выборочное среднеквадратическое отклонение*. В случае исследования непрерывно распределенного признака в приведенных формулах вместо наблюдаемых значений x_i обычно берут середины частичных интервалов.

Задача 15.3. Проведено четыре измерения некоторой случайной величины (в граммах): 42,6; 43,2; 42,4; 43,0. Найти оценки математического ожидания, и дисперсии.

Решение

Количество значений – 4. Находим выборочную среднюю:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{4}\left(42,6 + 43,2 + 42,4 + 43,0\right) = 42,8.$$

Эта оценка является несмещенной. Находим выборочную дисперсию:

$$D_{\text{в}} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^{\,2}\right) - \overline{x}^2 = \frac{1}{4}\left(42,6^2 + 43,2^2 + 42,4^2 + 43,0^2\right) - 42,8^2 = 8,78.$$

Эта оценка дисперсии не является несмещенной. Находим исправленную выборочную дисперсию: $s^2 = \dfrac{n}{n-1} D_{\text{в}} = \dfrac{4}{3}\cdot 8,78 = 11,71$.

В математической статистике важную роль играют оценки (эмпирические значения) начальных и центральных моментов. Для непрерывно распределенного признака *эмпирические начальные моменты первого, второго, третьего и четвертого порядков* определяются по формулам

$$M_1^* = \sum_{i=1}^{k} x_i\,\omega_i,\ M_2^* = \sum_{i=1}^{k} (x_i)^2\omega_i,\ M_3^* = \sum_{i=1}^{k} (x_i)^3\omega_i,\ M_4^* = \sum_{i=1}^{k} (x_i)^4\omega_i.$$

Из этих моментов можно получить *эмпирические центральные моменты второго, третьего и четвертого порядков*:

$$m_2^* = M_2^* - \left(M_1^*\right)^2,\ m_3^* = M_3^* - 3M_2^*M_1^* + 2\left(M_1^*\right)^3,$$

$$m_4^* = M_4^* - 4M_3^*M_1^* + 6M_2^*\left(M_1^*\right)^2 - 3\left(M_1^*\right)^4.$$

Заметим, что $\overline{x} = M_1^*$ и $D_{\text{в}} = m_2^*$.

Моменты более высокого порядка используются редко, а по введенным моментам вычисляются такие характеристики, как

– *асимметрия эмпирического распределения* $As^* = \dfrac{m_3^*}{\sigma_{\text{в}}^3}$;

– *эксцесс эмпирического распределения* $Ex^* = \dfrac{m_4^*}{\sigma_{\text{в}}^4} - 3$;

– *выборочный коэффициент вариации* $V^* = \dfrac{\sigma_{\text{в}}}{\overline{x}}$.

Для нормального закона распределения оценки асимметрии и эксцесса должны быть близки к нулю, для равномерного закона оценка асимметрии должна быть близка к нулю, для показательного закона оценка коэффициента вариации должна быть близка к единице.

Для оценки параметров предполагаемого закона распределения используется *метод моментов*. Суть его заключается в приравнивании теоретических моментов рассматриваемого распределения к соответствующим эмпирическим моментам того же порядка. Если закон распределения зависит от одного параметра, то для его оценки достаточно одного уравнения: $M(X) = \overline{x}$. Например, для показательного распределения $M(X) = \dfrac{1}{\lambda}$, тогда из уравнения следует: $\lambda = \dfrac{1}{\overline{x}}$.

В случае, когда закон распределения зависит от двух параметров, необходимо два уравнения: $\begin{cases} M(X) = \overline{x} \\ D(X) = D_{\text{в}} \end{cases}$. Так, для нормального закона параметры m, σ находятся из системы уравнений $\begin{cases} m = \overline{x} \\ \sigma = \sigma_{\text{в}} \end{cases}$. Для равномерного закона параметры a, b можно определить из системы $\begin{cases} \dfrac{a+b}{2} = \overline{x} \\ \dfrac{(b-a)^2}{12} = D_{\text{в}} \end{cases}$, откуда получаем $a = \overline{x} - \sqrt{3D_{\text{в}}}$, $b = \overline{x} + \sqrt{3D_{\text{в}}}$ (можно также принять во внимание размах выборки).

Чтобы принять для изучаемого признака подходящий закон распределения, по виду гистограммы и оценкам параметров выдвигаем предположение (*гипотезу*) о теоретическом законе распределения всей генеральной совокупности. Для проверки правильности выдвинутой гипотезы используется *критерий Пирсона*. Принято использовать случайную величину, характеризующую степень расхождения теоретического и эмпирического распределений, в виде $U = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$, где p_i – вероятность попадания в i-й частичный интервал случайной величины, распределенной по предполагаемому теоретическому закону. Рассматриваемая величина U – случайная, так как в различных опытах принимает неизвестные заранее значения, и, если выполняется выдвинутая гипотеза, имеет *распределение* χ^2 (*хи-квадрат*), которое зависит только от *числа степеней свободы* $r = k - 1 - l$, где l – число параметров предполагаемого теоретического распределения, k – количество интервалов.

По данным статистического ряда вычисляем наблюдаемое значение $U = \chi^2_{\text{набл}}$, где вероятности p_i находятся по формуле $p_i = F(x_{i+1}) - F(x_i)$. Здесь $F(x)$ – функция распределения для предполагаемого теоретического закона распределения. Вероятности p_i также можно приближенно вычислять по формуле $p_i \approx f(x_i^*) \cdot h$, где $f(x)$ – плотность распределения для предполагаемого теоретического закона распределения, x_i^* – середина i-го частичного интервала, h – его длина.

Критическое значение $\chi^2_{\text{крит}} = \chi^2_{\text{крит}}(r, \alpha)$ находим из специальных таблиц (Приложение 2) для числа степеней свободы r и *уровня значимости* α (допустимой вероятности отвергнуть правильную гипотезу). Если $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то выдвинутая гипотеза принимается, значит, можно принять при заданном уровне значимости теоретическое распределение исследуемого признака. Если $\chi^2_{\text{набл}} > \chi^2_{\text{крит}}$, то гипотеза отвергается, следовательно, генеральная совокупность имеет иное распределение или иные параметры.

Задача 15.4. (о выборе вида теоретического распределения). Требуется по данным, представленным в выборочной совокупности 1, выдвинуть гипотезы о теоретическом законе распределения и проверить согласованность выборочных и теоретических распределений по критерию Пирсона при уровне значимости $\alpha = 0,05$.

Выборка 1									
14,22	13,14	16,78	12,64	11,91	14,43	19,11	19,58	19,04	19,02
12,45	13,85	16,36	8,65	6,60	16,56	13,21	9,25	19,31	12,25

14,58	20,95	16,34	16,22	17,38	11,67	11,68	20,05	11,07	10,69
12,64	23,65	20,54	23,97	16,64	21,18	11,03	17,85	21,68	12,31
8,93	16,90	12,78	15,32	23,10	22,03	22,87	15,21	9,64	22,45
3,18	17,64	17,54	20,12	15,35	10,23	11,21	13,94	12,40	19,21
19,63	22,22	18,32	13,24	21,85	14,01	7,89	14,21	24,56	13,26
16,0	17,85	5,23	19,63	24,01	11,44	21,54	15,36	12,45	6,89
26,38	16,65	11,57	7,63	18,66	16,16	20,05	14,27	23,69	16,61
17,85	14,25	15,65	14,42	20,03	19,95	23,65	16,23	13,87	12,51

Решение

1. Построим статистический закон распределения и гистограмму.

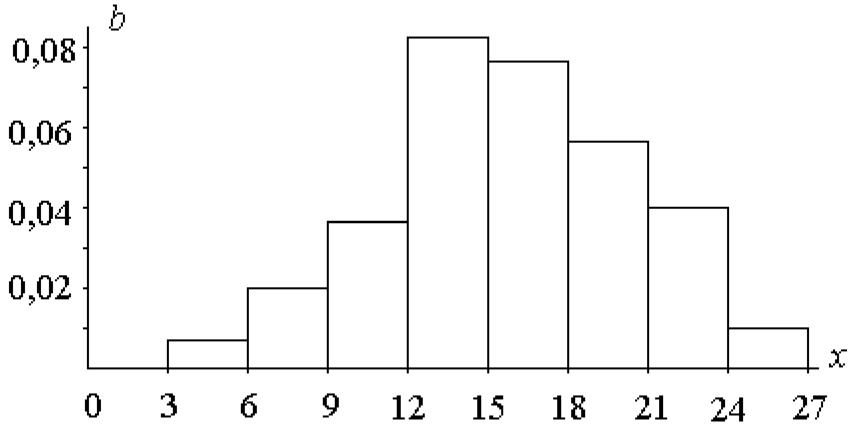
Найдем наименьшее и наибольшее значения в выборке объема $n=100$: $x_{\min}=3,18$ и $x_{\max}=26,38$. Округляем эти значения до ближайших целых чисел так, чтобы все статистические значения входили в интервал. Получаем интервал $(3;27)$. Выберем число частичных интервалов $k=8$, длина каждого интервала будет $h=(27-3)/8=3$. Строим интервальный статистический ряд.

интервал	**3-6**	**6-9**	**9-12**	**12-15**	**15-18**	**18-21**	**21-24**	**24-27**
n_i	2	6	11	25	23	17	13	3
ω_i	0,02	0,06	0,11	0,25	0,23	0,17	0,13	0,03
b_i	0,007	0,020	0,037	0,083	0,077	0,057	0,043	0,010
x_i^*	4,5	7,5	10,5	13,5	16,5	19,5	22,5	25,5

Здесь $\omega_i=\dfrac{n_i}{n}$ – относительные частоты, $b_i=\dfrac{\omega_i}{h}$ – плотность относительной частоты (высота i-го прямоугольника гистограммы), а x_i^* – середины частичных интервалов.

Опишем, каким образом заполняется вторая строка. В выборке берем первое значение $14,22$. Оно лежит в промежутке $(12;15)$. Записываем первое наблюдение в соответствующую ячейку второй строки. Затем рассматриваем следующее значение выборки – $13,14$. Оно тоже лежит в промежутке $(12;15)$. Записываем второе наблюдение в соответствующую ячейку второй строки. И так далее. После обработки всей выборки складываем количество наблюдений в каждой ячейке второй строки и получаем частоты n_i. Если некоторое значение попало на границу интервалов, то принято записывать по полнаблюдения в соседние интервалы. Исключение составляют первая левая граница и последняя правая – наблюдение полностью записывается в интервал.

Гистограмма относительных частот имеет следующий вид



2. Вычисляем параметры статистического распределения.

Для этого вычисляем эмпирические начальные моменты:

$$M_1^* = \sum_{i=1}^{8} x_i^* \omega_i = 15,87, \; M_2^* = \sum_{i=1}^{8} (x_i^*)^2 \omega_i = 274,$$

$$M_3^* = \sum_{i=1}^{8} (x_i^*)^3 \omega_i = 5042, \; M_4^* = \sum_{i=1}^{8} (x_i^*)^4 \omega_i = 97469.$$

Вычисляем эмпирические центральные моменты:

$$m_2^* = M_2^* - \left(M_1^*\right)^2 = 22,14, \; m_3^* = M_3^* - 3M_2^* M_1^* + 2\left(M_1^*\right)^3 = -9,20,$$

$$m_4^* = M_4^* - 4M_3^* M_1^* + 6M_2^* \left(M_1^*\right)^2 - 3\left(M_1^*\right)^4 = 1160.$$

Тогда выборочная средняя $\overline{x} = M_1^* = 15,87$, выборочная дисперсия $D_в = m_2^* = 22,14$, выборочное среднеквадратическое отклонение $\sigma_в = \sqrt{D_в} = 4,71$, асимметрия $As^* = \dfrac{m_3^*}{\sigma_в^3} = -0,09$, эксцесс $Ex^* = \dfrac{m_4^*}{\sigma_в^4} - 3 = -0,64$, коэффициент вариации $V^* = \dfrac{\sigma_в}{\overline{x}} = 0,3$.

3. Выдвигаем гипотезу о виде теоретического закона распределения.

Так как асимметрия и эксцесс близки к нулю, а также по виду гистограммы, выдвигаем гипотезу: изучаемый признак в генеральной совокупности имеет нормальное распределение. Запишем плотность нормального распределения $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$, где параметры m и σ определяем, используя метод моментов: $m = \overline{x} = 15{,}87$ и $\sigma = \sigma_{в} = 4{,}71$.

4. Проверим гипотезу о нормальном распределении статистических данных с параметрами $m = 15{,}87$ и $\sigma = 4{,}71$ по критерию Пирсона. Находим наблюдаемое значение $\chi^2_{набл}$. Для этого удобно построить следующую расчетную таблицу.

1	2	3	4	5	6	7	8	9	10
x_i	x_{i+1}	n_i	z_i	z_{i+1}	$\Phi(z_i)$	$\Phi(z_{i+1})$	p_i	$n \cdot p_i$	$\frac{(n_i - np_i)^2}{np_i}$
3	6	2	-2,73	-2,10	-0,4968	-0,4821	0,0147	1,47	0,191
6	9	6	-2,10	-1,46	-0,4821	-0,4279	0,0542	5,42	0,062
9	12	11	-1,46	-0,82	-0,4279	-0,2939	0,1340	13,40	0,430
12	15	25	-0,82	-0,18	-0,2939	-0,0714	0,2225	22,25	0,340
15	18	23	-0,18	0,45	-0,0714	0,1736	0,2450	24,50	0,092
18	21	17	0,45	1,09	0,1736	0,3621	0,1885	18,85	0,182
21	24	13	1,09	1,73	0,3621	0,4582	0,0961	9,61	0,196
24	27	3	1,73	2,36	0,4582	0,4909	0,0327	3,27	0,022
Сумма Σ									2,52

В данной таблице в первом столбце записаны левые границы частичных интервалов, во втором столбце – правые границы частичных интервалов, в третьем столбце – относительные частоты. Четвертый и пятый столбцы необходимы для нахождения значений функции распределения на границах, и значения в них вычисляются соответственно по формулам $z_i = \frac{x_i - \overline{x}}{\sigma_в}$ и $z_{i+1} = \frac{x_{i+1} - \overline{x}}{\sigma_в}$. Значения функции распределения на границах (шестой и седьмой столбцы) находятся из специальной таблицы (Приложение 1). Тогда вероятности попадания в каждый частичный интервал (восьмой столбец) находятся как $p_i = F(x_{i+1}) - F(x_i) = \Phi(z_{i+1}) - \Phi(z_i)$. Девятый столбец – вспомогательный (в нашей выборке $n = 100$) и, наконец, складывая вычисленные значения в десятом столбце, получаем наблюдаемое значение

$$\chi^2_{\text{набл}} = \sum_{i=1}^{8} \frac{(n_i - np_i)^2}{np_i} = 2,52.$$

По таблице критических значений распределения χ^2 (Приложение 2) для уровня значимости $\alpha = 0,05$ и числа степеней свободы $r = 8 - 3 = 5$ находим $\chi^2_{\text{крит}} = 11,1$. Так как $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то гипотеза о нормальном законе распределения генеральной совокупности принимается.

Задача 15.5. (о выборе вида теоретического распределения). Требуется по данным, представленным в выборочной совокупности 2, выдвинуть гипотезы о теоретическом законе распределения и проверить согласованность выборочных и теоретических распределений по критерию Пирсона при уровне значимости $\alpha = 0,05$.

Выборка 2

17,32	13,65	16,95	30,32	9,65	8,12	26,27	20,36	31,21	24,98
10,25	9,05	16,68	9,65	12,21	30,54	23,87	27,15	30,69	25,65
22,11	8,95	29,64	24,10	9,63	10,21	12,81	27,92	29,37	22,33
29,63	11,54	19,63	27,81	16,60	14,12	17,58	14,35	14,65	21,05
11,96	29,68	18,65	31,26	19,54	56,32	21,84	12,50	20,58	29,64
31,25	8,65	15,34	19,54	20,36	24,67	25,52	28,95	24,35	17,13
8,56	14,32	31,58	15,34	16,85	27,54	21,38	10,25	19,69	12,67
17,58	16,35	27,24	26,35	31,58	25,62	12,35	9,64	11,45	12,38
28,64	21,50	31,79	16,58	8,39	10,25	31,20	26,35	8,57	16,35
11,3	16,25	14,47	26,65	13,38	11,08	24,56	25,97	27,45	15,64

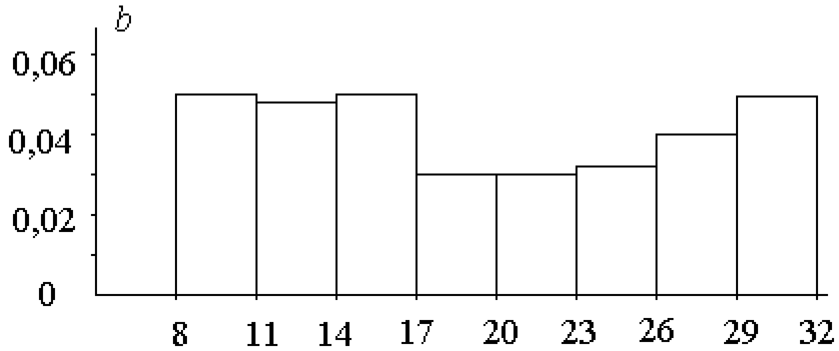
Решение

1. Построим статистический закон распределения и гистограмму.

Найдем наименьшее и наибольшее значения в выборке объема $n = 100$: $x_{\min} = 8,12$ и $x_{\max} = 31,79$. Округляем эти значения до ближайших целых чисел так, чтобы все статистические значения входили в интервал. Получаем интервал $(8;32)$. Вновь выберем число частичных интервалов $k = 8$, длина каждого интервала будет $h = \dfrac{32-8}{8} = 3$. Строим интервальный статистический ряд аналогично предыдущей выборке.

Интервал	8-11	11-14	14-17	17-20	20-23	23-26	26-29	29-32
n_i	15	14	15	9	9	10	13	15
ω_i	0,15	0,14	0,15	0,09	0,09	0,10	0,13	0,15
b_i	0,050	0,047	0,050	0,030	0,030	0,033	0,043	0,050
x_i^*	9,5	12,5	15,5	18,5	21,5	24,5	27,5	30,5

Гистограмма относительных частот имеет следующий вид.



2. Вычисляем параметры статистического распределения.

Для этого вычисляем эмпирические начальные моменты:

$$M_1^* = \sum_{i=1}^{8} x_i^* \omega_i = 19,73, \; M_2^* = \sum_{i=1}^{8} (x_i^*)^2 \omega_i = 442,$$

$$M_3^* = \sum_{i=1}^{8} (x_i^*)^3 \omega_i = 10855, \; M_4^* = \sum_{i=1}^{8} (x_i^*)^4 \omega_i = 283254.$$

Вычисляем эмпирические центральные моменты:

$$m_2^* = M_2^* - \left(M_1^*\right)^2 = 52,73, \; m_3^* = M_3^* - 3M_2^* M_1^* + 2\left(M_1^*\right)^3 = 53,73,$$

$$m_4^* = M_4^* - 4M_3^* M_1^* + 6M_2^* \left(M_1^*\right)^2 - 3\left(M_1^*\right)^4 = 4329.$$

Тогда выборочная средняя $\overline{x} = M_1^* = 19,73$, выборочная дисперсия $D_{\text{в}} = m_2^* = 52,73$, выборочное среднеквадратическое отклонение $\sigma_{\text{в}} = \sqrt{D_{\text{в}}} = 7,26$, асимметрия $As^* = \dfrac{m_3^*}{\sigma_{\text{в}}^3} = 0,14$, эксцесс $Ex^* = \dfrac{m_4^*}{\sigma_{\text{в}}^4} - 3 = -1,44$, коэффициент вариации $V^* = \dfrac{\sigma_{\text{в}}}{\overline{x}} = 0,37$.

3. Выдвигаем гипотезу о виде теоретического закона распределения.

Так как асимметрия близка к нулю, а также по виду гистограммы, выдвигаем гипотезу о равномерном распределении. Запишем плотность равномерного распределения $f(x)=\dfrac{1}{b-a}$, где параметры a и b определяем, используя метод моментов: $a=\overline{x}-\sqrt{3}\sigma_{\text{в}}=7{,}16$ и $b=\overline{x}+\sqrt{3}\sigma_{\text{в}}=32{,}30$.

4. Проверим гипотезу о равномерном распределении статистических данных с параметрами $a=7{,}16$ и $b=32{,}30$ по критерию Пирсона. Находим наблюдаемое значение $\chi^2_{\text{набл}}$ по расчетной таблице.

1	2	3	4	5	6	7
x_i	x_{i+1}	n_i	$f(x)$	p_i	$n\cdot p_i$	$\dfrac{(n_i-np_i)^2}{np_i}$
8	11	15	0,0396	0,1187	11,87	0,827
11	14	14	0,0396	0,1187	11,87	0,384
14	17	15	0,0396	0,1187	11,87	0,827
17	20	9	0,0396	0,1187	11,87	0,692
20	23	9	0,0396	0,1187	11,87	0,692
23	26	10	0,0396	0,1187	11,87	0,063
26	29	13	0,0396	0,1187	11,87	0,002
29	32	15	0,0396	0,1187	11,87	0,827
					Сумма Σ	4,32

В данной таблице в первом столбце записаны левые границы частичных интервалов, во втором столбце – правые границы частичных интервалов, в третьем столбце – относительные частоты. В четвертом столбце записывается значение плотности распределения $f(x)=\dfrac{1}{b-a}=\dfrac{1}{32{,}30-7{,}16}=\dfrac{1}{25{,}14}$. В пятом столбце определяются вероятности попадания в каждый частичный интервал по формуле $p_i=F\left(x_{i+1}\right)-F\left(x_i\right)=h\cdot f\left(x\right)$. Шестой столбец – вспомогательный (в нашей выборке $n=100$) и, наконец, складывая вычисленные значения в седьмом столбце, получаем наблюдаемое значение

$$\chi^2_{\text{набл}}=\sum_{i=1}^{8}\frac{\left(n_i-np_i\right)^2}{np_i}=4{,}32.$$

По таблице критических значений распределения χ^2 (Приложение 2) для уровня значимости $\alpha = 0,05$ и числа степеней свободы $r = 8 - 3 = 5$ находим $\chi^2_{\text{крит}} = 11,1$. Так как $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то гипотеза о равномерном законе распределения генеральной совокупности принимается.

Задача 15.6. (о выборе вида теоретического распределения). Требуется по данным, представленным в выборочной совокупности 3, выдвинуть гипотезы о теоретическом законе распределения и проверить согласованность выборочных и теоретических распределений по критерию Пирсона при уровне значимости $\alpha = 0,05$.

	Выборка 3								
7,12	0,25	22,56	8,69	3,45	3,68	7,54	8,65	7,12	10,20
2,45	1,02	4,57	37,12	3,01	8,35	8,65	9,31	16,25	3,20
5,24	1,24	4,98	6,54	3,75	3,25	12,50	9,89	11,25	17,54
0,59	1,32	9,65	0,15	35,66	3,15	32,94	10,57	11,23	20,20
0,87	4,05	2,68	1,23	3,96	14,52	24,10	10,46	11,28	21,50
0,17	4,05	2,96	18,34	6,46	1,06	6,56	13,46	5,02	3,96
12,35	4,02	2,85	7,14	14,23	2,01	19,63	0,12	15,68	26,35
6,05	4,44	2,68	9,68	14,52	1,98	6,58	13,65	5,32	4,91
1,68	39,64	2,56	5,91	12,60	2,52	0,38	17,64	16,39	34,52
1,58	4,31	2,10	7,85	12,34	14,56	19,62	0,67	13,63	8,95

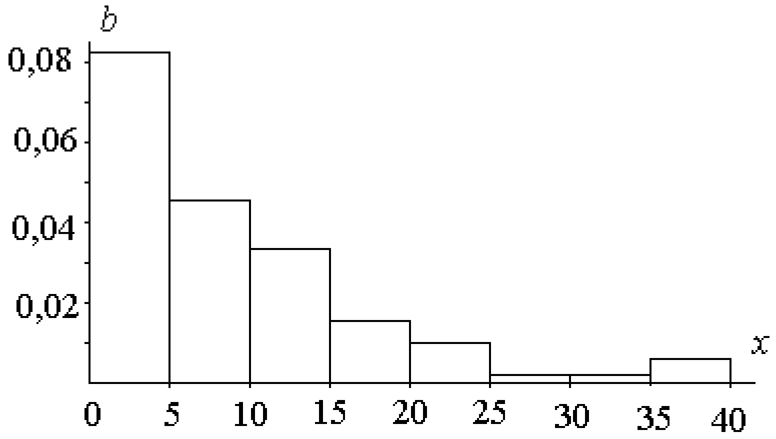
Решение

1. Построим статистический закон распределения и гистограмму.

Найдем наименьшее и наибольшее значения в выборке объема $n = 100$: $x_{\min} = 0,12$ и $x_{\max} = 39,64$. Округляем эти значения до ближайших целых чисел так, чтобы все статистические значения входили в интервал. Получаем интервал $(0;40)$. Опять удобно выбрать число частичных интервалов $k = 8$, длина каждого интервала будет $h = \dfrac{40-0}{8} = 5$. Строим интервальный статистический ряд, как описано в предыдущих случаях.

Интервал	**0-5**	**5-10**	**10-15**	**15-20**	**20-25**	**25-30**	**30-35**	**35-40**
n_i	42	23	17	8	5	1	1	3
ω_i	0,42	0,23	0,17	0,08	0,05	0,01	0,01	0,03
b_i	0,084	0,046	0,034	0,016	0,010	0,002	0,002	0,006
x_i^*	2,5	7,5	12,5	17,5	22,5	27,5	32,5	37,5

Гистограмма относительных частот имеет следующий вид:



2. Вычисляем параметры статистического распределения.

Для этого вычисляем эмпирические начальные моменты:

$$M_1^* = \sum_{i=1}^{8} x_i^* \omega_i = 9{,}18, \; M_2^* = \sum_{i=1}^{8} (x_i^*)^2 \omega_i = 152,$$

$$M_3^* = \sum_{i=1}^{8} (x_i^*)^3 \omega_i = 3567, \; M_4^* = \sum_{i=1}^{8} (x_i^*)^4 \omega_i = 101414.$$

Вычисляем эмпирические центральные моменты:

$$m_2^* = M_2^* - \left(M_1^*\right)^2 = 67{,}73, \; m_3^* = M_3^* - 3M_2^* M_1^* + 2\left(M_1^*\right)^3 = 928,$$

$$m_4^* = M_4^* - 4M_3^* M_1^* + 6M_2^* \left(M_1^*\right)^2 - 3\left(M_1^*\right)^4 = 25985.$$

Тогда выборочная средняя $\overline{x} = M_1^* = 9{,}18$, выборочная дисперсия $D_в = m_2^* = 67{,}73$, выборочное среднеквадратическое отклонение $\sigma_в = \sqrt{D_в} = 8{,}23$, асимметрия $As^* = \dfrac{m_3^*}{\sigma_в^3} = 1{,}66$, эксцесс $Ex^* = \dfrac{m_4^*}{\sigma_в^4} - 3 = 2{,}66$, коэффициент вариации $V^* = \dfrac{\sigma_в}{\overline{x}} = 0{,}90$.

3. Выдвигаем гипотезу о виде теоретического закона распределения.

Так как коэффициент вариации близок к единице, а также по виду гистограммы выдвигаем, гипотезу о показательном распределении. Запишем плотность показательного распределения $f(x) = \lambda e^{-\lambda x}$, где параметр λ определяем, используя метод моментов: $\lambda = \dfrac{1}{\overline{x}} = 0{,}11$.

4. Проверим гипотезу о показательном распределении статистических данных с параметром $\lambda = 0{,}11$ по критерию Пирсона. Находим наблюдаемое значение $\chi^2_{\text{набл}}$. Для этого удобно построить следующую расчетную таблицу.

1	2	3	4	5	6	7	8	9	10
x_i	x_{i+1}	n_i	z_i	z_{i+1}	e^{-z_i}	$e^{-z_{i+1}}$	p_i	$n \cdot p_i$	$\frac{(n_i - np_i)^2}{np_i}$
0	5	42	0	0,55	1	0,5769	0,4231	42,31	0,002
5	10	23	0,55	1,10	0,5769	0,3329	0,2440	24,40	0,080
10	15	17	1,10	1,65	0,3329	0,1920	0,1409	14,09	0,601
15	20	8	1,65	2,20	0,1920	0,1108	0,0812	8,12	0,002
20	25	5	2,20	2,75	0,1108	0,0639	0,0469	4,69	0,020
25	30	1	2,75	3,30	0,0639	0,0369	0,0270	2,70	1,070
30	35	1	3,30	3,85	0,0369	0,0213	0,0156	1,56	0,201
35	40	3	3,85	4,40	0,0213	0,0123	0,0090	0,90	4,900
Сумма Σ									6,88

В данной таблице в первом столбце записаны левые границы частичных интервалов, во втором столбце – правые границы частичных интервалов, в третьем столбце – относительные частоты. Четвертый и пятый столбцы, значения в которых вычисляются соответственно по формулам $z_i = \lambda x_i$ и $z_{i+1} = \lambda x_{i+1}$, а также шестой и седьмой столбцы необходимы для нахождения вероятности попадания в каждый частичный интервал (восьмой столбец) по формуле $p_i = F(x_{i+1}) - F(x_i) = e^{-z_i} - e^{-z_{i+1}}$. Девятый столбец – вспомогательный (в нашей выборке $n = 100$) и, наконец, складывая вычисленные значения в десятом столбце, получаем наблюдаемое значение

$$\chi^2_{\text{набл}} = \sum_{i=1}^{8} \frac{(n_i - np_i)^2}{np_i} = 6{,}88.$$

По таблице критических значений распределения χ^2 (Приложение 2) для уровня значимости $\alpha = 0{,}05$ и числа степеней свободы $r = 8 - 2 = 6$ находим $\chi^2_{\text{крит}} = 12{,}6$. Так как $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то гипотеза о показательном законе распределения генеральной совокупности принимается.