Часть 2. Элементы математической статистики

Замечательно, что науке, начинавшейся с рассмотрения азартных игр, суждено было стать важнейшим объектом человеческого знания.

Лаплас

Вероятность – это важнейшее понятие в современной науке особенно потому, что никто совершенно не представляет, что оно означает.

Бертран Рассел

Главная цель расчёта – не цифры, а понимание.

2.1. Генеральная и выборочная совокупность

Æ***Генеральной совокупностью*** называется множество объектов произвольной природы, обладающих признаками, доступными для наблюдения и количественного измерения.

ÆОбъекты, входящие в генеральную совокупность, называются её ***элементами****,* а их общее число – её ***объёмом***.

Предположим, из генеральной совокупности случайным образом извлекаем элементы, значения некоторого признака для них записываем как $x_1, x_2, x_3, \dots, x_k$. Эти значения называются *наблюдениями*, их набор – *выборкой*. Количество наблюдений каждого из признаков обозначим $n_1, n_2, n_3, \dots, n_k$ и назовём частотами. Число наблюдений n называем *объёмом выборки*: $n = n_1, +n_2 + n_3 + \cdots + n_k$.

Æ***Основная задача*** математической статистики – сделать *научно обоснованные выводы о распределении* одной или более неизвестных случайных величин или их *взаимосвязи* между собой.

Æ***Выборочным методом*** называется метод решения этой задачи посредством анализа выборки, полученной в результате многократных наблюдений.

Для того чтобы характеристики случайной величины, полученные выборочным методом, были объективны, необходимо, чтобы выборка была репрезентативной, т. е. достаточно хорошо представляла исследуемую величину. В силу закона больших чисел можно утверждать, что выборка будет репрезентативной, если её осуществлять случайно, т. е. все объекты генеральной совокупности имеют одинаковую вероятность попасть в выборку.

Æ*Выборка* называется ***повторной***, если отобранный объект перед выбором следующего возвращается в генеральную совокупность. *Выборка* называется ***бесповторной***, если отобранный объект в генеральную совокупность не возвращается. На практике обычно имеют дело с бесповторными выборками.

Æ*Всякая случайная величина* X имеет определённую функцию распределения и другие числовые характеристики, которые называются *теоретическими*, в отличие от *выборочных*, которые определяются по наблюдениям.

ÆРяд наблюдений, ***упорядоченных по возрастанию***, называется ***вариационным рядом***. Его члены обозначаются $x_1, x_2, x_3, \dots, x_k$ и называются вариантами.

Наименьшее и наибольшее значения вариант обозначаются x_{min} и x_{max}, их называются ***крайними членами вариационного ряда***. Число $R = x_{max} - x_{min}$ называется ***размахом выборки***.

ÆЧисло $\omega_i = \frac{n_i}{n}$ $(i = 1, \dots, k)$ называется ***относительной частотой***.

Вариационный ряд можно записать как

x_i	x_1	x_2	x_3	…	x_k
n_i	n_1	n_2	n_3	…	n_k

В случае непрерывной случайной величины на практике часто применяют ***группировку***.

Интервал наблюдаемых значений разбивают на k частичных интервалов $[c_0, c_1)$, $[c_1, c_2), \dots, [c_{k-1}, c_k)$ одинаковой длины h.

Затем подсчитывают числа попаданий наблюдений в эти интервалы, которые принимают за частоты n_i.

В качестве новых значений вариант x_i обычно берут середины интервалов $x_i = \frac{c_{i-1} + c_i}{2}$ $(i = 1, \dots, k)$.

Þ *Примечание*. Группировку можно применять и в случае дискретной случайной величины, если шаг, с которым меняются её значения, слишком мал.

ÆРекомендуемое число интервалов вычисляют по формуле Стерджеса $k = 1 + log_2 n$.

Длину частичных интервалов вычисляют как $h = \frac{R}{k} = \frac{x_{max} - x_{min}}{k}$. Полагаем, что весь отрезок наблюдений, его часто называют интервалом наблюдений, имеет вид $[x_{min}; x_{max}]$.

Þ *Примечание*. Группировка связана с потерей части полезной информации, заключённой в выборке. Однако она имеет и свои преимущества. Оценим величину экономии, например, выполнено 1000 наблюдений некоторого признака. Рекомендуемое число интервалов: $k = 1 + log_2 n = 1 + log_2 10^3 = 1 + 3log_2 10 = 1 + 3 \cdot 3{,}3 = 1 + 9{,}9 \approx 10$. Отсюда видно, что требуется обработать $2 \cdot k = 20$ числа вместо 1000.

Æ*Набор вариант* x_i (или частичных интервалов) и их *относительных частот* ω_i называется ***статистическим рядом***.

Статистический ряд для дискретной случайной величины:

Варианта x_i	x_1	x_2	x_3	…	x_k
Частота n_i	n_1	n_2	n_3	…	n_k
Относительная частота $\omega_i = \frac{n_i}{n}$	$\omega_1 = \frac{n_1}{n}$	$\omega_2 = \frac{n_2}{n}$	$\omega_3 = \frac{n_3}{n}$	…	$\omega_k = \frac{n_k}{n}$

Статистический ряд для непрерывной случайной величины:

Интервалы	$[c_0, c_1)$	$[c_1, c_2)$	$[c_2, c_3), \ldots,$		$[c_{k-1}, c_k)$
Середина интервала x_i	x_1	x_2	x_3	…	x_k
Частота n_i	n_1	n_2	n_3	…	n_k
Относительная частота $\omega_i = \frac{n_i}{n}$	$\omega_1 = \frac{n_1}{n}$	$\omega_2 = \frac{n_2}{n}$	$\omega_3 = \frac{n_3}{n}$	…	$\omega_k = \frac{n_k}{n}$

Графически статистические ряды можно представить в виде *полигона*, *гистограммы* или *графика накопленных частот*.

Æ***Полигон частот*** – это ломаная линия, отрезки которой соединяют точки (x_1, n_1), (x_2, n_2), …, (x_k, n_k).

Æ***Полигон относительных частот*** – это ломаная линия, отрезки которой соединяют точки (x_1, ω_1), (x_2, ω_2), …, (x_k, ω_k).

Þ *Примечание*. Полигоны обычно служат для изображения выборки в случае дискретных случайных величин.

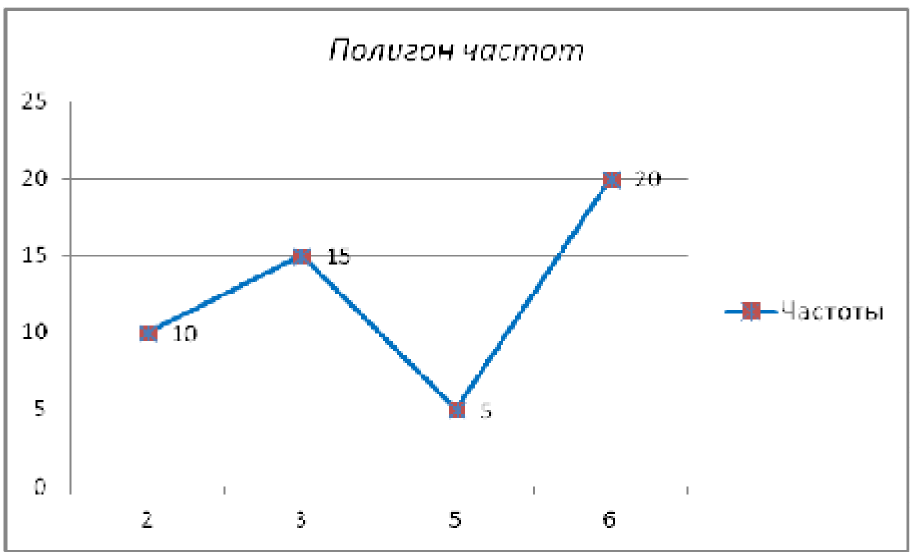
? ***Упражнение 1.*** Постройте полигоны частот и относительных частот, график накопленных частот, запишите эмпирическую функцию распределения по данному распределению выборки:

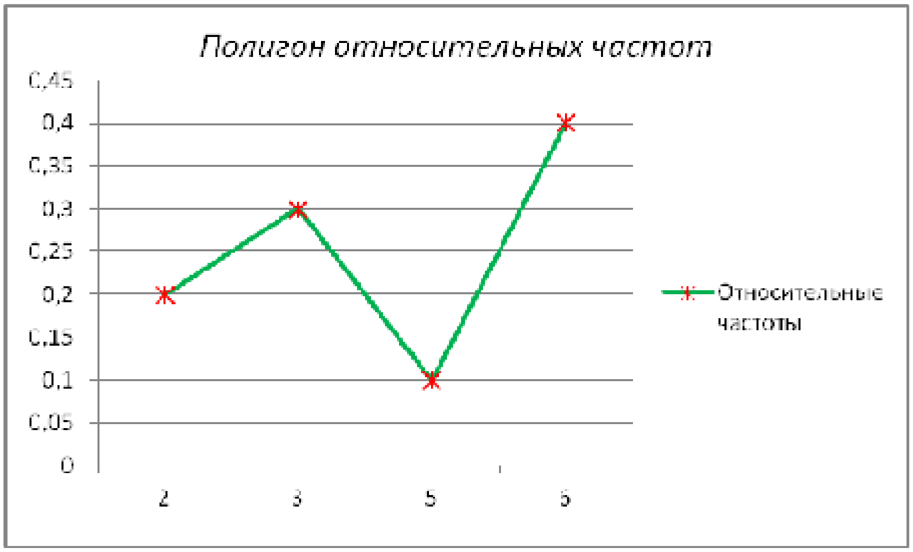
	2	3	5	6
	10	15	5	20

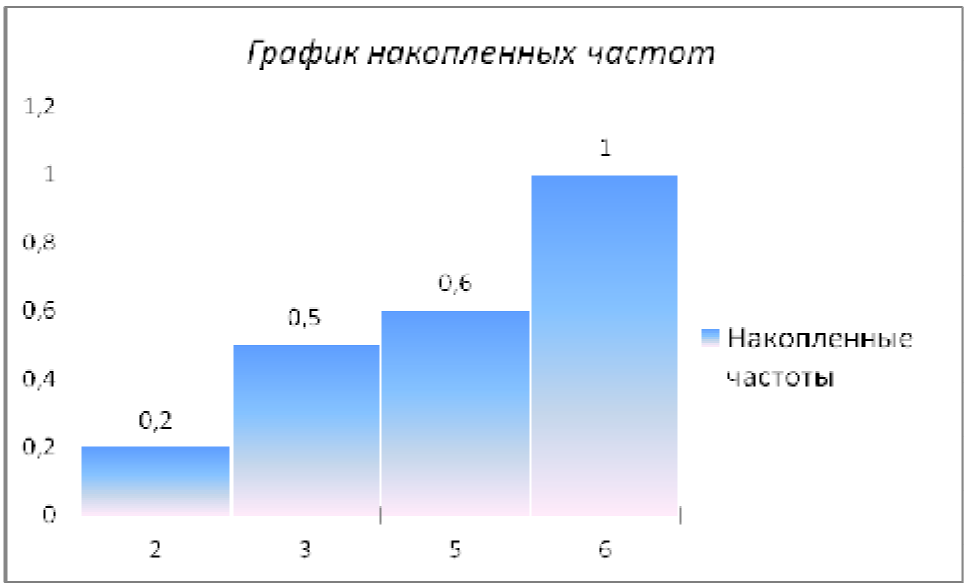
Решение

Найдём объём выборки и дополним таблицу относительными частоты, т. е. построим статистический ряд.

Варианты	Частоты	Относительные частоты	Накопленные частоты
2	10	0,2	0,2
3	15	0,3	0,5
5	5	0,1	0,6
6	20	0,4	1
Объём выборки:	50		







Запишем эмпирическую функцию распределения, используя накопленные частоты:

Æ***Гистограммой частот*** называется *ступенчатая фигура*, состоящая из прямоугольников, основанием которых служат частичные интервалы длиною h, а высоты равны n_i/h.

Величина n_i/h называется *плотностью частоты*.

Æ***Гистограммой относительных частот*** называется *ступенчатая фигура*, состоящая из прямоугольников, основанием которых служат частичные интервалы длиною h, а высоты равны ω_i/h.

Величина ω_i/h называется *плотностью относительной частоты*.

Þ *Примечание*. Гистограмма обычно служит для изображения выборки в случае непрерывных случайных величин. Очевидно, площадь *гистограммы относительных частот равна единице*. Поэтому гистограмму относительных частот можно рассматривать как график эмпирической (выборочной) плотности распределения, в этом и заключается практическая польза гистограммы относительных частот.

Æ***Графиком накопленных частот*** называется *ступенчатая фигура*, состоящая из прямоугольников, основанием которых служат частичные интервалы длиною h, а высоты равны накопленным частотам. Заметим, что график накопленных частот имеет вид ступенчатой «лестницы» (от 0 до 1).

Þ *Примечание*. График накопленных частот и эмпирическая функция распределения на практике используются для приближения теоретической функции распределения.

? ***Упражнение 2***. Постройте гистограммы частот и относительных частот, график накопленных частот, запишите эмпирическую функцию распределения по данному распределению выборки:

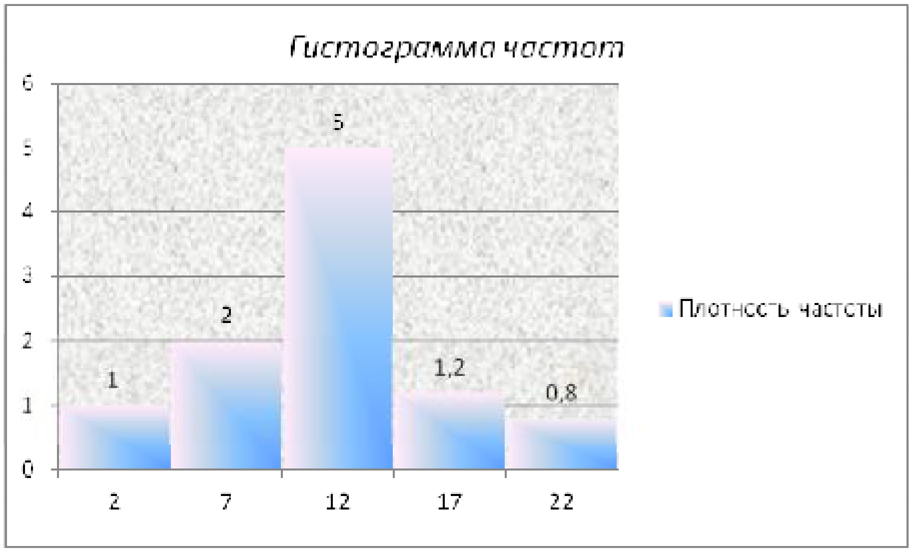
Частичный интервал	[2,7)	[7,12)	[12,17)	[17,22)	[22,27)
Число наблюдений, попавших в интервал, n_i	5	10	25	6	4

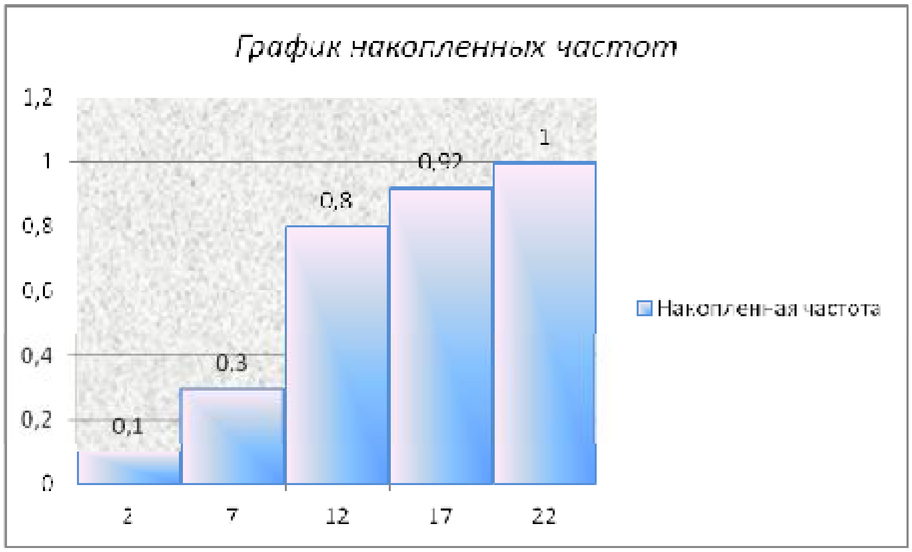
Решение

Найдём объём выборки и дополним таблицу относительными частоты, т. е. построим статистический ряд.

Частичный интервал	Частота	Относительная частота	Накопленная частота	Плотность частоты	Плотность относительной частоты	Начало интервала
[2; 7)	5	0,1	0,1	1	0,02	2
[7; 12)	10	0,2	0,3	2	0,04	7
[12; 17)	25	0,5	0,8	5	0,1	12

[17; 22)	6	0,12	0,92	1,2	0,024	17
[22; 27)	4	0,08	1	0,8	0,016	22
Объём выбор-ки:	50	Площадь гистограммы частот:		50		
Длина интервала:	5					





Используя накопленные частоты, запишем эмпирическую функцию распределения:

2.2. Выборочные характеристики и точечные оценки

Æ***Выборочными характеристиками*** называются функции от наблюдений, приближённо оценивающие соответствующие *числовые характеристики* случайной величины.

Оценки параметров генеральной совокупности делятся на два класса: *точечные* и *интервальные*.

Точечные оценки выражаются одним числом (точкой на числовой оси), находятся такие оценки по данным выборки и используются в дальнейшем вместо оцениваемого параметра. Точечная оценка, как функция от выборки, является случайной величиной и меняется от выборки к выборке при повторном эксперименте.

Интервальные оценки определяются двумя числами – концами интервала, который накрывает оцениваемый параметр.

В отличие от точечных оценок, которые не дают представления о том, как далеко от них может находиться оцениваемый параметр, интервальные оценки позволяют установить точность и надёжность оценок.

Если объём выборки $n > 30$, то при построении доверительного интервала для математического ожидания можно пользоваться нормальным законом распределения. В случае неизвестной дисперсии для определения ширины интервала используют несмещённую оценку дисперсии $\hat{S}^2$ и соответствующее выборочное среднее квадратическое отклонение $\hat{\sigma} = \sqrt{\hat{S}^2}$ [2].

К точечным оценкам предъявляют требования, которым они должны удовлетворять, чтобы хоть в каком-то смысле быть «доброкачественными». Это несмещённость, эффективность и состоятельность [1].

В качестве *точечных оценок* математического ожидания, дисперсии и среднего квадратического отклонения используют выборочные характеристики соответственно выборочное среднее, выборочная дисперсия и выборочное среднее квадратическое отклонение:

1) выборочное среднее: $\bar{X} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$;
2) выборочная смещённая (неисправленная) дисперсия:

$$\widehat{D} = \overline{X^2} - (\bar{X})^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i)^2 \cdot n_i - \left(\frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i\right)^2, \text{ или}$$

$$\widehat{D} = \frac{1}{n}\sum_{i=1}^{k} n_i \cdot \left(x_i - \bar{X}\right)^2;$$

3) выборочная несмещённая (исправленная) дисперсия:
   $\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \bar{X})^2 \cdot n_i$, или $\hat{S}^2 = \frac{n}{n-1} \cdot \widehat{D}$;
4) смещённое выборочное среднее квадратическое отклонение:
   $\hat{\sigma} = \sqrt{\widehat{D}}$;
5) несмещённое выборочное среднее квадратическое отклонение:
   $\hat{\sigma} = \sqrt{\hat{S}^2}$.

Þ *Примечание*. Выборочное среднее $\bar{X}$ есть несмещённая, эффективная и состоятельная точечная оценка математического ожидания, в то время как, выборочная дисперсия $\widehat{D}$ является смещённой точечной оценкой дисперсии. В этом случае вводят исправленную (несмещённую) точечную оценку $\hat{S}^2$ дисперсии $D(X)$.

В качестве других используемых на практике выборочных характеристик можно назвать ***выборочную моду*** M_0 и ***выборочную медиану*** M_e.

Для наблюдений *дискретной случайной* величины*:*

– *выборочная мода* M_0 равна значению варианты с наибольшей частотой n_{max};

– *выборочная медиана* M_e равна значению варианты, стоящей в середине вариационного ряда, если число наблюдаемых вариант есть нечётное число;

– если число наблюдаемых вариант есть чётное число, тогда *выборочная медиана* M_e равна *полусумме двух соседних значений вариант*, стоящих в середине вариационного ряда.

Медиана M_e есть серединный элемент.

В том случае, когда наблюдения проводятся для *непрерывной случайной* величины, то мода M_0 и медиана M_e определяются по следующим правилам.

Что касается моды M_0, то сначала определяется модальный интервал, т. е. интервал с наибольшей частотой (или относительной частотой). Затем мода M_0 вычисляется по формуле

$$M_0 = x_0 + h \cdot \frac{n_i - n_{i-1}}{2 \cdot n_i - n_{i-1} - n_{i+1}}, \qquad (1)$$

здесь x_0 – начало модального интервала, имеющего максимальную частоту, n_i – частота модального интервала, h – длина модального интервала, n_{i-1} и n_{i+1} – частоты соответственно предшествующего и последующего за модальным интервалом

Медианой называют такое число M_e, когда 50% вариант выборки меньше или равна этого значения, а 50% – больше или равна его.

В этом случае медиану определяем по алгоритму:

1) найдите медианный интервал $[x_{i-1};\ x_i)$. Это такой интервал, для которого накопленная частота $\omega_{x_i}^{\text{нак}} \leq 0.5$, в этом случае медиана $M_e \in [x_{i-1};\ x_i)$;

2) вычислите медиану M_e по одной из формул:

$$M_e = x_{i-1} + h \cdot \frac{0.5 \cdot n - n_{x_{i-1}}^{\text{нак}}}{n_i}, \qquad (2)$$

или

$$M_e = x_{i-1} + h \cdot \frac{0.5 - \omega_{x_{i-1}}^{\text{нак}}}{\omega_{x_i}^{\text{нак}} - \omega_{x_{i-1}}^{\text{нак}}}, \qquad (3)$$

здесь x_{i-1} – начало медианного интервала, h – ширина медианного интервала, n – объём выборки, $n_{x_{i-1}}^{\text{нак}}$ – накопленная частота интервала, предшествующего медианному интервалу, n_i – частота медианного интервала, $\omega_{x_i}^{\text{нак}}$ – накопленная относительная частота медианного интервала, $\omega_{x_{i-1}}^{\text{нак}}$ – накопленная относительная частота интервала, предшествующего медианному интервалу.

? ***Упражнение 3.*** Найдите выборочное среднее, смещённую и несмещённую выборочные дисперсии, смещённое и несмещённое выборочные средние квадратические отклонения, моду M_0 и медиану M_e по данному распределению выборки:

x_i	2	3	5	6
n_i	10	15	5	20

Решение

Исследования выполнялись для дискретной случайной величины. Вычисления выполним в MS Excel.

Варианты	Частоты	Относительные частоты	Накопленные частоты	Произведение варианты на частоту	Произведение квадрата варианты на частоту
2	10	0,2	0,2	20	40
3	15	0,3	0,5	45	135
5	5	0,1	0,6	25	125
6	20	0,4	1	120	720
Объём выборки:	50			210	1020
Выборочное среднее:				4,2	20,4
Смещённая выборочная дисперсия:					2,76
Несмещённая выборочная дисперсия:					2,82
Смещённая выборочная:					1,66
Несмещённая выборочная дисперсия:					1,68
Мода M0 = 3, наибольшая частота равна 15.					
Медиана Me = (3+5)/2 = 4 - среднее арифметическое двух серединных элементов.					

Þ *Примечание*. В данном примере чётное количество вариант, а именно 4, это 2, 3, 5, и 6. Поэтому медиану определяем как среднее значение дух серединных элементов: $M_\text{e} = \frac{3+5}{2} = 4$.

Несмещённая выборочная дисперсия вычисляется по формуле:

$\hat{S}^2 = \frac{50}{50-1} \cdot 2{,}76 \approx 2{,}82.$

? ***Упражнение 4.*** Найдите выборочное среднее, смещённую и несмещённую выборочные дисперсии, смещённое и несмещённое выборочные средние квадратические отклонения по данному распределению выборки:

Частичный интервал	[2,7)	[7,12)	[12,17)	[17,22)	[22,27)
Число наблюдений, попавших в интервал, n_i	5	10	25	6	4

Решение

Вычисления выполним в MS Excel. В данном примере наблюдения выполнялись над непрерывной случайной величиной.

Частичный интервал	Частота	Относительная частота	Накопленная частота	Начало интервала	Конец интервала	Середина интервала	Произведение среднего значения варианты на частоту
[2; 7)	5	0,1	0,1	2	7	4,5	22,50
[7; 12)	10	0,2	0,3	7	12	9,5	95,00
[12; 17)	25	0,5	0,8	12	17	14,5	362,50
[17; 22)	6	0,12	0,92	17	22	19,5	117,00
[22; 27)	4	0,08	1	22	27	24,5	98,00
Объём выборки:	50						695,00
Длина интервала:	5				Выборочное среднее:		13,90
Мода M0:	14,21						
Медиана Me:	17,00						

Þ *Примечание*. В данном примере максимальная частота $n_2 = 25$, [12; 17) – модальный интервал, мода M_0 вычисляем по формуле (1): $M_0 = 12 + 5 \cdot \frac{25-10}{2 \cdot 25-10-6} = 14{,}21.$

Интервал [7; 12) является медианным интервалом, поскольку накопленная частота $n_{x_i}^{\text{нак}} = n_7^{\text{нак}} = 0{,}3 < 0{,}5.$ Медиану можно вычислить по одной из формул (2) или (3):

$$M_e = x_{i-1} + h \cdot \frac{0.5 \cdot n - n_{x_{i-1}}^{\text{нак}}}{n_i} = 7 + 5 \cdot \frac{0,5 \cdot 50 - 5}{10} = 17,$$

$$M_e = x_{i-1} + h \cdot \frac{0.5 - \omega_{x_{i-1}}^{\text{нак}}}{\omega_{x_i}^{\text{нак}} - \omega_{x_{i-1}}^{\text{нак}}} = 7 + 5 \cdot \frac{0,5 - 0,1}{0,3 - 0,1} = 17.$$

2.3. Примеры решения задач

Æ Задача 1. Первичная обработка данных

Анализируется выборка из 100 малых предприятий региона. Цель обследования – измерение коэффициента соотношения заёмных и собственных средств на каждом предприятии. Результаты представлены в виде табл. 1.

Таблица 1

Коэффициенты соотношений заёмных и собственных средств на предприятий

5.56	5.45	5.48	5.45	5.39	5.37	5.46	5.59	5.61	5.31
5.46	5.61	5.11	5.41	5.31	5.57	5.33	5.11	5.54	5.43
5.34	5.53	5.46	5.41	5.48	5.39	5.11	5.42	5.48	5.49
5.36	5.40	5.45	5.49	5.68	5.51	5.50	5.68	5.21	5.38
5.58	5.47	5.46	5.19	5.60	5.63	5.48	5.27	5.22	5.37
5.33	5.49	5.50	5.54	5.40	5.58	5.42	5.29	5.05	5.79
5.79	5.65	5.70	5.71	5.85	5.44	5.47	5.48	5.47	5.55
5.67	5.71	5.73	5.05	5.35	5.72	5.49	5.61	5.57	5.69
5.54	5.39	5.32	5.21	5.73	5.59	5.38	5.25	5.26	5.81
5.27	5.64	5.20	5.23	5.33	5.37	5.24	5.55	5.60	5.51

По приведённым данным выполните:

Задание 1. Постройте статистический ряд.

Задание 2. Вычислите относительные частоты и накопленные частоты.

Задание 3. Представьте графически статистический ряд в виде полигона или гистограммы.

Задание 4. Постройте график эмпирической функции распределения (график накопленных частот).

Задание 5. Составьте эмпирическую функцию распределения.

Задание 6. Вычислите точечные оценки параметров законов распределения:

1) выборочное среднее;
2) выборочную смещённую (неисправленную) дисперсию и выборочную несмещённую (исправленную) дисперсию;
3) выборочное неисправленное среднее квадратическое отклонение и выборочные исправленное среднее квадратическое отклонение;
4) выборочную моду;
5) выборочную медиану.

Задание 7. Найдите доверительный интервал для неизвестного математического ожидания нормально распределённой генеральной совокупности при условии, что дисперсия неизвестна, если доверительная вероятность задана как $\gamma = 0.9 + 0.01 \cdot i$, где i – последняя цифра шифра зачётной книжки.

Þ *Примечание*. Если объём выборки $n > 30$, то при построении доверительного интервала для неизвестного математического ожидания можно пользоваться нормальным распределением, подставляя в формулу для ширины интервала вместо неизвестного значения среднего квадратического отклонения его оценку, определяемую по данной выборке.

Решение

Задание 1. Постройте статистический ряд.

Шаг 1. Внесём данные выборки в MS Excel, упорядочим их в порядке возрастания, найдём наименьшее $x_{min} = 5.05$ и наибольшее $x_{max} = 5.85$ значения вариант x_i, $i = 1, \dots, 100$. Вычислим размах выборки $r = x_{max} - x_{min} = 5.85 - 5.05 = 0.8$ и k – число интервалов, на которые разбивается диапазон $[x_{min};\ x_{max}]$: $k = 1 + log_2 100 = 7.62$, $k \approx 8$. Определим длину интервала $h = \frac{x_{max} - x_{min}}{k} = \frac{0.8}{8} = 0.1$.

Шаг 2. Построим статистический ряд (табл. 2).

Таблица 2

Статистический ряд

Интервал	5.05 - 5.15	5.15 -5.25	5.25-5.35	5.35-5.45	5.45-5.55	5.55- 5.65	5.65-5.75	5.75- 5.85
Частота	**5**	**8**	**12**	**20**	**26**	**15**	**10**	**4**

n_i								

<u>Задание 2</u>. Вычислите относительные частоты и накопленные частоты.

Построим сгруппированный ряд наблюдений, найдём середины интервалов, вычислим частоты, относительные и накопленные частоты (табл. 3).

Таблица 3

Сгруппированный ряд наблюдений

Номер интер-вала	Интервал	Середина интервала x_i	Часто-та n_i	Относи-тельная частота ω_i	Накоплен-ная часто-та ω_i^c	$f_i(x) = h \cdot \omega_i$
1	[5.05; 5.15)	5.1	5	0.05	0.05	0.5
2	[5.15; 5.25)	5.2	8	0.08	0.13	0.8
3	[5.25; 5.35)	5.3	12	0.12	0.25	1.2
4	[5.35; 5.45)	5.4	20	0.20	0.45	2.0
5	[5.45; 5.55)	5.5	26	0.26	0.71	2.6
6	[5.55; 5.65)	5.6	15	0.15	0.86	1.5
7	[5.65; 5.75)	5.7	10	0.10	0.96	1.0
8	[5.75; 5.85)	5.8	4	0.04	1.00	0.4

<u>Задание 3</u>. Представьте графически статистический ряд в виде полигона или гистограммы.

Выполним вспомогательные вычисления и построение гистограммы частот в MS Excel (рис. 1).

Левая граница	Правая граница	Середина интервала	Частота	Относительная частота	Накопленная частота	Значение функции

5,05	5,15	5,1	5	0,05	0,05	0,5
5,15	5,25	5,2	8	0,08	0,13	0,8
5,25	5,35	5,3	12	0,12	0,25	1,2
5,35	5,45	5,4	20	0,2	0,45	2
5,45	5,55	5,5	26	0,26	0,71	2,6
5,55	5,65	5,6	15	0,15	0,86	1,5
5,65	5,75	5,7	10	0,1	0,96	1
5,75	5,85	5,8	4	0,04	1	0,4

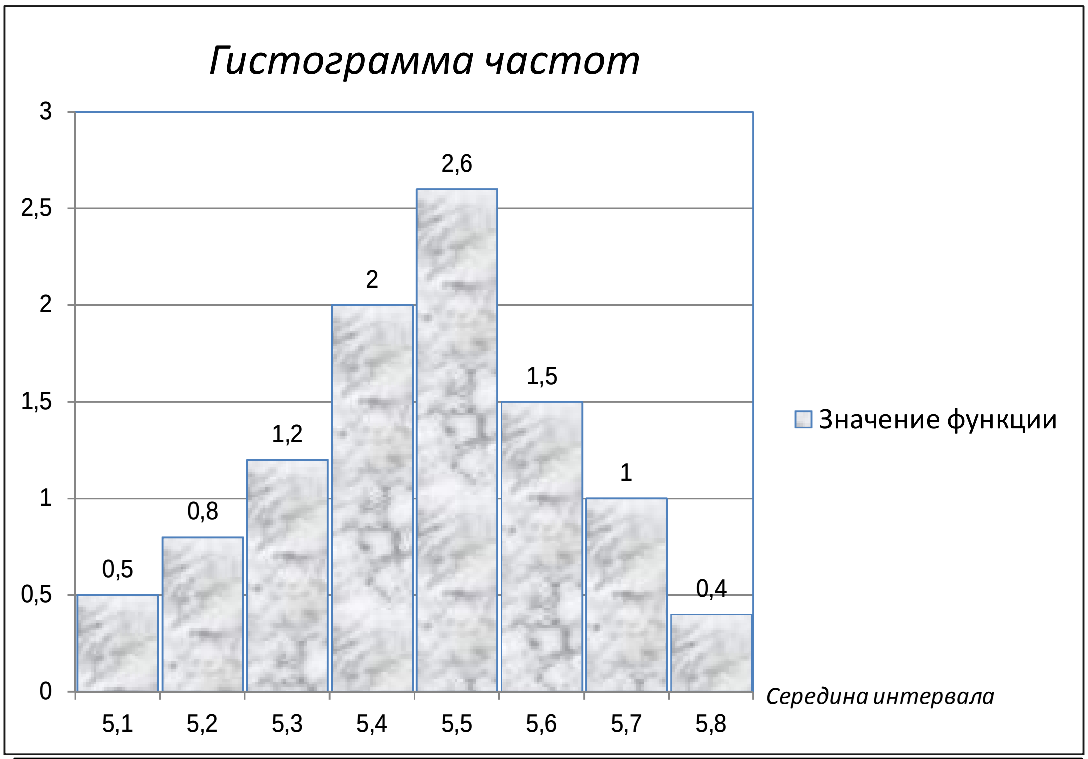


Рис.1. Гистограмма относительных частот

Задание 4. Постройте график эмпирической функции распределения (график накопленных частот).

Используем вычисленные накопленные частоты, полученные на предыдущем шаге, и построим график накопленных частот в MS Excel (рис. 2).

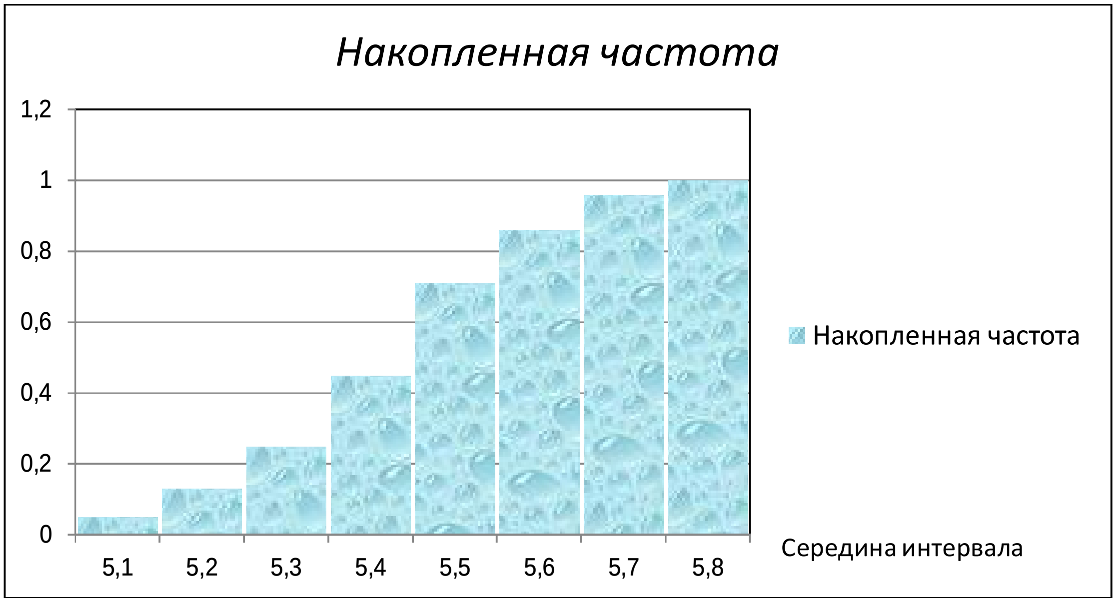


Рис. 2. График эмпирической функции распределения

Задание 5. Составьте эмпирическую функцию распределения.

Запишем эмпирическую функцию распределения, используя данные накопленных относительных частот:

Задание 6. Вычислим точечные оценки параметров законов распределения с помощью MS Excel:

1) выборочное среднее;
2) выборочную смещённую (неисправленную) дисперсию и выборочную несмещённую (исправленную) дисперсию;

3) выборочное смещённое (неисправленное) среднее квадратическое отклонение и выборочные несмещённое (исправленное) среднее квадратическое отклонение;
4) выборочную моду;
5) выборочную медиану.

Левая граница	Правая граница	Середина интервала	Частота	Произведение середины интервала на частоту	Произведение квадрата середины интервала на частоту	Относительные частоты	Накопленные частоты
5,05	5,15	5,1	5	25,5	130,05	0,05	0,05
5,15	5,25	5,2	8	41,6	216,32	0,08	0,13
5,25	5,35	5,3	12	63,6	337,08	0,12	0,25
5,35	5,45	5,4	20	108	583,2	0,2	0,45
5,45	5,55	5,5	26	143	786,5	0,26	0,71
5,55	5,65	5,6	15	84	470,4	0,15	0,86
5,65	5,75	5,7	10	57	324,9	0,1	0,96
5,75	5,85	5,8	4	23,2	134,56	0,04	1
Объём выборки:			100	545,9	2983,01		
Выборочное среднее:				5,459	29,8301	"среднее квадратов"	
Смещённая оценка дисперсии:					0,029		
Несмещённая оценка дисперсии :					0,030		
Выборочное среднее квадратическое (смещённое):						0,1715	
Выборочное среднее квадратическое (несмещённое):						0,1724	
Мода M0 :			5,485				
Медиана Me:			5,369				

Задание 7. Найдите доверительный интервал для неизвестного математического ожидания нормально распределённой генеральной совокупности при условии, что дисперсия неизвестна, если доверительная вероятность задана как $\gamma = 0.9 + 0.01 \cdot i$, где i – последняя цифра шифра зачётной книжки.

Предположим $i = 5$, тогда доверительная вероятность задана как $\gamma = 0{,}95$. Ранее вычислены выборочное среднее $\overline{x} = 5{,}459$ и исправленное среднее квадратическое отклонение $\hat{s} = 0{,}1724$.

По условию $\gamma = 0{,}95 = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\hat{s}}\right) \to \Phi\left(\frac{\varepsilon\sqrt{n}}{\hat{s}}\right) = \Phi(t_\gamma) = 0{,}457$. По таблице значений функции Лапласа (Приложение 1) находим значение аргумента $t_\gamma = 1{,}960$. Объём выборки $n = 100$. Тогда $\varepsilon = 1{,}960 * \frac{\hat{s}}{\sqrt{n}} = 1{,}960 * \frac{0{,}1724}{10} = 0{,}0337 \approx 0{,}034$. Таким образом, интервал $(\overline{x} - 0{,}034;\ \overline{x} - 0{,}034) = (5{,}425;\ 5{,}493)$ покрывает неизвестный параметр такой, как математическое ожидание $M(X)$, с надёжностью $\gamma = 0{,}95$.

Þ *Примечание*. Если в таблице значений функции Лапласа нет точного значения $\frac{\gamma}{2}$, следует взять среднее арифметическое двух ближайших соседних значений, одно меньшее, другое – большее.

$1 - \gamma = 0{,}05$. $n - 1 = 99$ – число степеней свободы. По таблице значений t-распределения

Æ Задача 2. Первичная обработка данных

Выполнены экологические статистические исследования за определённый промежуток времени, например, несанкционированный подъезд машин к закрытой территории мусорных отходов. Получили выборку, данные которой приведены в табл. 2.

Таблица 2

Несанкционированный подъезд машин

2	4	2	4	3	3	3	2	0	6	1	2	3	2	2
4	3	3	5	1	0	2	4	3	2	2	3	3	1	3
3	3	1	1	2	3	1	4	3	1	7	4	3	4	2
3	2	3	3	1	4	3	1	4	5	3	4	2	4	5
3	6	4	1	3	2	4	1	3	1	0	0	4	6	4
7	4	1	3											

По приведённым данным выполните:

Задание 1. Постройте статистический ряд.

Задание 2. Вычислите относительные частоты и накопленные частоты.

Задание 3. Представьте графически статистический ряд в виде полигона или гистограммы.

Задание 4. Постройте график эмпирической функции распределения (график накопленных частот).

Задание 5. Составьте эмпирическую функцию распределения.

Задание 6. Вычислите точечные оценки параметров законов распределения:

1) выборочное среднее;
2) выборочную смещённую (неисправленную) дисперсию и выборочную несмещённую (исправленную) дисперсию;
3) выборочное неисправленное среднее квадратическое отклонение и выборочные исправленное среднее квадратическое отклонение;
4) выборочную моду;
5) выборочную медиану.

Задание 7. Найдите доверительный интервал для неизвестного математического ожидания нормально распределённой генеральной совокупности при условии, что дисперсия неизвестна, если доверительная вероятность задана как $\gamma = 0.9 + 0.01 \cdot i$, где i – последняя цифра шифра зачётной книжки.

Решение

Задание 1. Постройте статистический ряд.

Внесём данные выборки в MS Excel, упорядочим их в порядке возрастания, найдём наименьшее $x_{min} = 0$ и наибольшее $x_{max} = 7$ значения вариант x_i, $i = 1, \ldots, 79$. Объём выборки $n = 79$. Вычислим размах выборки $r = x_{max} - x_{min} + 1 = 8$. Поскольку размах довольно мал, поэтому составим вариационный ряд по значениям.

x_i	0	1	2	3	4	5	6	7
n_i	4	13	14	24	16	3	3	2

Задание 2. Вычислите относительные частоты и накопленные частоты.

Вычисления выполним в MS Excel.

Значение вариант	Частота	Относительная частота	Накопленная частота
0	4	0,0506	0,0506

1	13	0,1646	0,2152
2	14	0,1772	0,3924
3	24	0,3038	0,6962
4	16	0,2025	0,8987
5	3	0,0380	0,9367
6	3	0,0380	0,9747
7	2	0,0253	1,0000
Объём выборки:	79		

Задание 3. Представьте графически статистический ряд в виде полигона или гистограммы.

Рис.3. Полигон относительных частот

Задание 4. Постройте график эмпирической функции распределения (график накопленных частот).

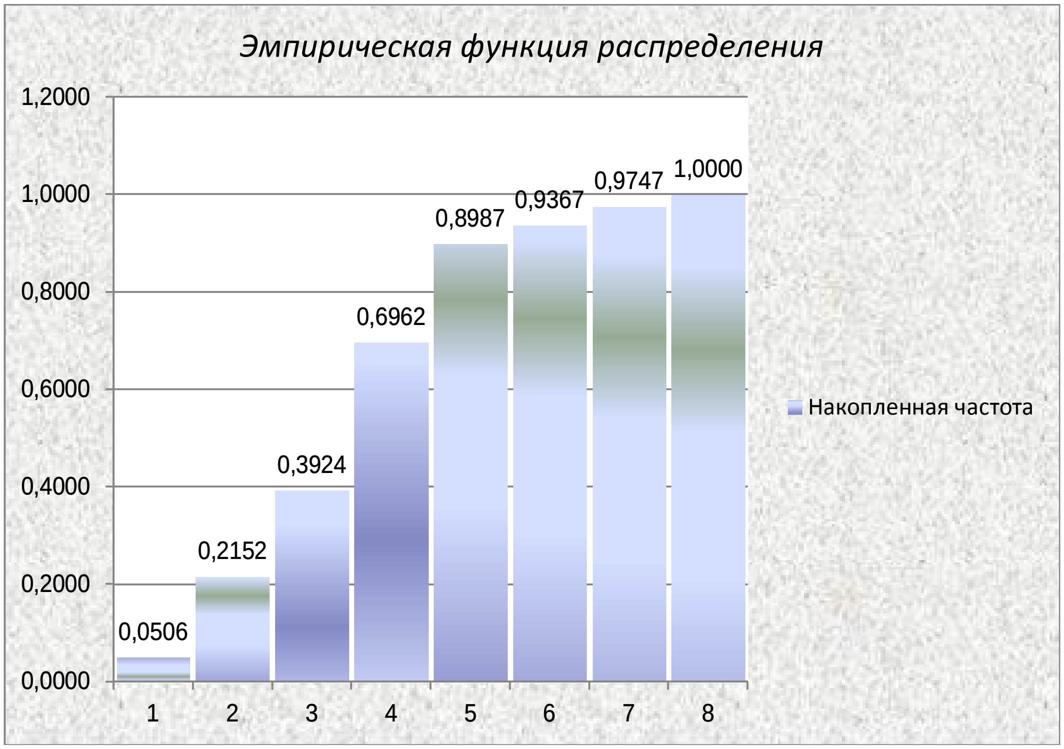


Рис. 4. График эмпирической функции распределения

Задание 5. Составьте эмпирическую функцию распределения.

Запишем эмпирическую функцию распределения, используя данные накопленных относительных частот:

$$F(x) = \begin{cases} 0, & \text{если } x \le 0, \\ 0.0506, & \text{если } 0 < x \le 1, \\ 0{,}2152, & \text{если } 1 < x \le 2, \\ 0{,}3924, & \text{если } 2 < x \le 3, \\ 0{,}6962, & \text{если } 3 < x \le 4, \\ 0{,}8987, & \text{если } 4 < x \le 5, \\ 0{,}9367, & \text{если } 5 < x \le 6, \\ 0{,}9747, & \text{если } 6 < x \le 7, \\ 1, & \text{если } x > 7 \end{cases}$$

Задание 6. Вычислите точечные оценки параметров законов распределения:

1) выборочное среднее;

2) выборочную смещённую (неисправленную) дисперсию и выборочную несмещённую (исправленную) дисперсию;
3) выборочное неисправленное среднее квадратическое отклонение и выборочные исправленное среднее квадратическое отклонение;
4) выборочную моду;
5) выборочную медиану.

Вычисления выполним в MS Excel.

Значение вариант	Частота	Произведение значения варианты на частоту	Произведение квадрата значения варианты на частоту
0	4	0	0
1	13	13	13
2	14	28	56
3	24	72	216
4	16	64	256
5	3	15	75
6	3	18	108
7	2	14	98
Объём выборки:	79	224	822
Выборочное среднее :		2,8354	10,4051
Смещённая выборочная дисперсия:			2,3653
Несмещённая выборочная дисперсия:			2,3957
Смещённая выборочное среднее квадратическое:			1,5380
Несмещённая выборочное среднее квадратическое:			1,5478
Мода (наибольшая частота):		3	
Медиана (элемент, стоящий на 40 - м месте, т. е серединный элемент):			3

Задание 7. Найдите доверительный интервал для неизвестного математического ожидания нормально распределённой генеральной совокупности при условии, что дисперсия неизвестна, если доверительная вероятность задана как $\gamma = 0.9 + 0.01 \cdot i$, где i – последняя цифра шифра зачётной книжки.

Предположим $i = 5$, тогда доверительная вероятность задана как $\gamma = 0{,}95$. Ранее вычислены выборочное среднее $\overline{x} = 2{,}8354$ и исправленное среднее квадратическое отклонение $\hat{s} = 1{,}5478$.

По условию $\gamma = 0{,}95 = 2\Phi\left(\frac{\varepsilon\sqrt{n}}{\hat{s}}\right) \to \Phi\left(\frac{\varepsilon\sqrt{n}}{\hat{s}}\right) = \Phi(t_\gamma) = 0{,}457$. По таблице значений функции Лапласа (Приложение 1) находим значение аргумента $t_\gamma = 1{,}960$. Объём выборки $n = 100$. Тогда $\varepsilon = 1{,}960 * \frac{\hat{s}}{\sqrt{n}} = 1{,}960 * \frac{1{,}5478}{\sqrt{79}} = 1{,}960 * \frac{1{,}5478}{8{,}8882} \approx 0{,}3413$. Таким образом, интервал $(\bar{x} - 0{,}3413;\ \bar{x} - 0{,}3413) = (2{,}4941;\ 3{,}1767)$ покрывает неизвестный параметр такой, как математическое ожидание $M(X)$, с надёжностью $\gamma = 0{,}95$.

2.4. ВАРИАНТЫ ЗАДАНИЙ

! *В задачах с номерами 0, 1, 2, 3, 4, 5 и 6 по данным своего варианта выполните следующие задания:*

Задание 1. Постройте статистический ряд.

Задание 2. Вычислите относительные частоты и накопленные частоты.

Задание 3. Представьте графически статистический ряд в виде полигона или гистограммы.

Задание 4. Составьте эмпирическую функцию распределения.

Задание 5. Постройте график эмпирической функции распределения.

Задание 6. Вычислите точечные оценки параметров законов распределения:

1) выборочное среднее;
2) выборочную смещённую (неисправленную) дисперсию и выборочную несмещённую (исправленную) дисперсию;
3) выборочное неисправленное среднее квадратическое отклонение и выборочные исправленное среднее квадратическое отклонение;
4) выборочную моду;
5) выборочную медиану.

Задание 7. Найдите доверительный интервал для неизвестного математического ожидания нормально распределённой генеральной совокупности при условии, что дисперсия неизвестна, если доверительная вероятность задана как $\gamma = 0.9 + 0.01 \cdot i$, где i – последняя цифра шифра зачётной книжки.

P ***Вариант 0***

В автопарке проводилось исследование продолжительности автомобильных рейсов (сутки). Результаты дорожной ведомости приведены в табл. 0.

Таблица 0

4	2	7	10	4	3	6	2	8	10
2	8	5	6	3	4	8	6	5	1
5	7	4	3	4	7	3	5	9	2
3	5	10	4	3	8	5	9	4	4
4	7	4	2	1	8	10	3	4	6
1	3	6	2	4	7	5	5	0	2
3	7	7	9	10	2	5	8	3	6
8	10	2	4	9	3	6	6	2	7
9	3	5	2	1	6	3	5	7	2
10	5	7	3	6	3	2	7	1	9

P ***Вариант 1***

Для изучения распределения заработной платы работников некоторой отрасли за определённый промежуток времени обследовано 100 человек. Результаты представлены в табл. 1.

Таблица 1

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	17	16	14	13	11	9	7	5	4	2	2

Здесь x_i – заработная плата в тыс. руб., n_i – число человек.

P ***Вариант 2***

В течение квартала на фондовой бирже выполнен сбор данных по количеству сделок для 100 инвесторов. Результаты представлены в виде табл. 2.

Таблица 2

x_i	0	1	2	3	4	5	6	7	8	9	10
n_i	17	16	14	13	11	9	7	5	4	2	2

Р ***Вариант 3***

Проведено исследование посещаемости интернет-сайта. Несколько часов подряд регистрируется число посетителей, посетивших сайт в течение данного часа. Результаты исследования приведены в табл. 3.

Таблица 3

x_i	0	1	2	3	4	5	6
n_i	57	203	383	525	532	408	273
x_i	7	8	9	10	11	12	14
n_i	139	45	27	10	4	1	1

Здесь x_i – число посетителей, n_i – время (ч).

Р ***Вариант 4***

В течение трёх месяцев проводились измерения барометрического давления воздуха (мм рт. ст.). Результаты приведены в табл. 4.

Таблица 4

x_i	745	746	747	748	749	750	751	752	753	754
n_i	2	3	7	10	11	6	8	10	6	5
x_i	755	756	757	758	759	760	761	762	763	764
n_i	8	9	8	12	13	11	8	6	4	3

Р ***Вариант 5***

Проведено исследование продолжительности работы лампочек (ч/10). Результаты исследования приведены в таб. 5.

Таблица 5

51	56	69	31	56	49	51	53	74	51
63	48	53	51	64	50	59	84	55	82
55	72	70	54	51	77	98	62	73	55

P ***Вариант 6***

На некотором участке дороги проведены измерения скорости автомобилей, км/ч. Результаты измерения даны в табл. 6.

Таблица 6

41	41	29	15	41	43	42	34	41	30
23	48	50	36	35	46	28	46	50	41
55	27	43	53	48	47	34	35	29	42
30	35	38	41	36	38	45	59	44	43

P ***Вариант 7***

У пятидесяти новорождённых измеряли массу тела с точностью до 10 г. Результаты измерений приведены в таблице 7:

Таблица 7

3,70	3,85	3,70	3,78	3,60	4,45	4,20	3,87	3,30	3,76
3,75	4,03	3,75	4,18	3,80	4,75	3,25	4,10	3,55	3,35
3,38	3,30	4,15	3,95	3,50	3,88	3,71	3,15	4,15	3,80
4,22	3,75	3,58	3,55	4,08	4,03	3,24	4,05	3,56	3,05
3,58	3,98	3,88	3,78	4,05	3,40	3,80	3,06	4,38	4,20

! *В задачах с номерами 8, и 9 по данным своего варианта выполните следующие задания:*

Задание 1. Вычислите относительные частоты и накопленные частоты.

Задание 2. Представьте графически статистический ряд в виде полигона или гистограммы.

Задание 3. Составьте эмпирическую функцию распределения.

Задание 4. Постройте график эмпирической функции распределения.

Задание 5. Вычислите точечные оценки параметров законов распределения:

6) выборочное среднее;
7) выборочную смещённую (неисправленную) дисперсию и выборочную несмещённую (исправленную) дисперсию;
8) выборочное неисправленное среднее квадратическое отклонение и выборочные исправленное среднее квадратическое отклонение;
9) выборочную моду;
10) выборочную медиану.

Задание 6. Найдите доверительный интервал для неизвестного математического ожидания нормально распределённой генеральной совокупности при условии, что дисперсия неизвестна, если доверительная вероятность задана как $\gamma = 0.9 + 0.01 \cdot i$, где i – последняя цифра шифра зачётной книжки.

Р *Вариант 8*

Исследовался средний месячный доход (тыс. руб.) жителя определённого региона по выборке из 1000 жителей. Полученный результат приведён в табл. 7.

Таблица 8

x_i	<10	[10; 15)	[15; 20)	[20; 25)	[25; 30)	>30
n_i	58	96	239	328	147	132

Указание. В качестве верхней границы последнего интервала используйте 35000 тыс. руб.

Р *Вариант 9*

Проведено исследование на прочность 200 образцов бетона на сжатие. Результаты представлены в виде статистического ряда в табл. 9.

Таблица 9

x_i	[19; 20)	[20; 21)	[21; 22)	[22; 23)	[23; 24)	[24; 25)
n_i	10	26	58	64	28	14

Здесь x_i – частичный интервал предела прочности (МПа), n_i – число наблюдений, попавших в интервал.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Гмурман В. Е. Теория вероятностей и математическая статистика / В. Е. Гмурман. – М. : Высшая школа, 2000. – 479 с.

2. Кононенко Э. Д. Интервальные оценки (методические указания к проведению практических занятий) / Э. Д. Кононенко, Л. В. Марченко.–Хабаровск: ДВГУПС, 2010. – 32 с.

3. Фадеева Л. Н. Теория вероятностей и математическая статистика / Л. Н. Фадеева, А. В. Лебедев; под ред. д-ра экон. наук, проф. Л. Н. Фадеевой. – М. : Рид Групп, 2011. – 496 с.

4. Марченко Л.В. Метод наименьших квадратов (методические указания к проведению практического занятия) / Л.В. – Марченко. Хабаровск: ДВГУПС, 2002. – 46с.

1. Данко П.Е., Попов А.Г. Высшая математика в упражнениях и задачах / П.Е. Данко, А.Г. Попов,Т. Я. Кожевникова. . – М. : Высшая школа, 2010. – 4013 с.

ОГЛАВЛЕНИЕ