

1 СТАТИСТИЧЕСКОЕ РАСПРЕДЕЛЕНИЕ ВЫБОРКИ

1.1 Содержание: генеральная и выборочная совокупность, статистический ряд, статистическое распределение случайной величины, эмпирическая функция распределения, графическое распределение статистических рядов.

1.2 Примеры решения типовых задач

1.2.1 На потоке производится замер времени (в минутах) на решение поставленной задачи. Производится выборка среди студентов, которые решали заданную задачу. Время на решение задач представлено в виде таблицы:

t_i – время на решение задачи	11	15	13	15	13	15	11	15	13	15
---------------------------------	----	----	----	----	----	----	----	----	----	----

Найти распределение частот и относительных частот. Построить полигон частот и полигон относительных частот. По данному распределению выборки найти эмпирическую функцию распределения и построить её график.

Решение. В результате опыта случайная величина принимает три значения 11, 13 и 15 минут. Два студента затратили 11 минут на решение поставленной задачи, три – 13 минут, а 5 студентов – 15 минут, то есть случайная величина T – время на решение задачи студентами, равно 11, 13 и 15, а частоты этих событий – 2, 3 и 5, соответственно. Следовательно, статистический ряд распределения частот имеет вид

t_i	11	13	15
m_i	2	3	5

Найдём объём выборки: $n = 2 + 3 + 5 = 10$. Определим относительные частоты: $\omega_1 = m_1/n = 2/10 = 0,2$, $\omega_2 = 3/10 = 0,3$, $\omega_3 = 5/10 = 0,5$. Записываем распределение относительных частот:

t_i	11	13	15
m_i/n	0,2	0,3	0,5

Построим полигон частот (рисунок 1.2.1) и полигон относительных частот (рисунок 1.2.2).

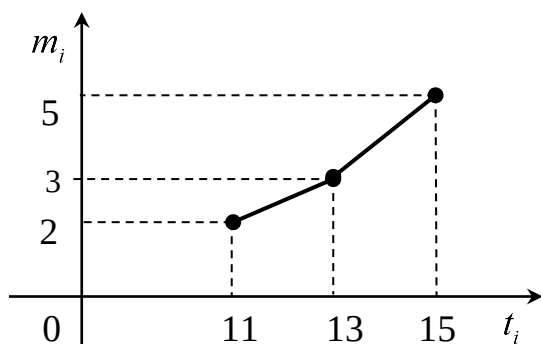


Рисунок 1.2.1 – Полигон частот случайной величины T

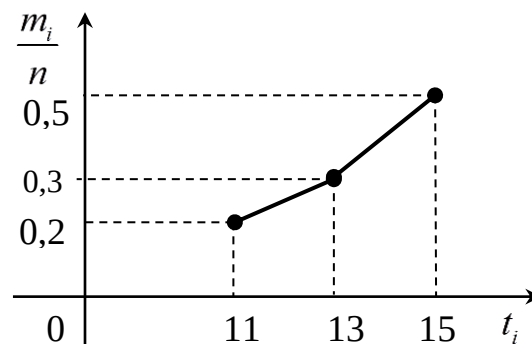


Рисунок 1.2.2 – Полигон относительных частот СВ T

Найдём эмпирическую функцию распределения по заданному распределению выборки. Наименьшая варианта равна 11, следовательно,

$$F^*(t) = 0 \text{ при } t \leq 11.$$

Значение $T < 13$, а именно $t_1 = 11$, наблюдалось два раза, следовательно,

$$F^*(t) = 2/10 = 0,2 \text{ при } 11 < t \leq 13.$$

Значение $T < 15$, а именно $t_1 = 11$ и $t_2 = 13$, наблюдалось $2+3=5$ раз, следовательно,

$$F^*(t) = 5/10 = 0,5 \text{ при } 13 < t \leq 15.$$

Так как $t = 15$ – наибольшая варианта, то

$$F^*(t) = 1 \text{ при } t > 15.$$

Запишем искомую эмпирическую функцию распределения СВ T :

$$F^*(t) = \begin{cases} 0, & t \leq 11, \\ 0,2, & 11 < t \leq 13, \\ 0,5, & 13 < t \leq 15, \\ 1, & t > 15. \end{cases}$$

График эмпирической функции распределения заданной случайной величины T изображён на рисунке 1.2.3.

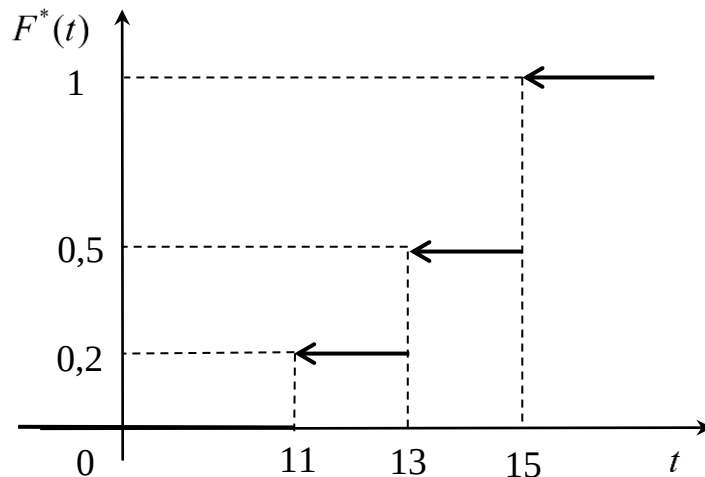


Рисунок 1.2.3 – Эмпирическая функция распределения случайной величины T

1.2.2 Пусть случайная величина ξ – температура воздуха в градусах Цельсия, измеряемая в регионе в течение суток. Из генеральной совокупности извлекается выборка объемом $n = 200$. Для заданной выборки построить сгруппированный статистический ряд, разбивая размах варьирования на шесть интервалов. Найти эмпирическую функцию распределения и построить её график. Построить гистограмму, полигон распределения частот или относительных частот, а так же кумуляту полученного статистического ряда. По графику гистограммы сделать прогноз о распределении случайной величины ξ . Значения выборки температуры приведены в виде таблицы:

7,1	5,0	9,6	1,2	9,9	4,0	4,5	12,8	8,1	4,8	5,2	8,3	7,6	8,2
8,2	8,8	5,4	5,4	11,6	5,6	8,8	7,8	5,8	4,8	8,1	9,1	6,1	6,0
3,0	9,4	7,3	8,3	5,1	9,8	6,8	5,2	7,0	5,0	9,5	6,9	11,0	8,9
9,0	9,5	5,4	1,6	10,0	2,8	7,8	8,8	10,2	8,4	3,4	1,9	9,0	4,6
1,7	12,0	9,7	7,4	8,9	5,7	12,6	7,9	8,0	8,1	5,3	6,3	7,1	10,0
5,0	6,0	6,9	9,7	6,4	7,4	5,4	8,9	2,9	10,9	7,7	12,7	5,1	8,0
11,8	7,0	3,5	2,5	11,4	10,4	8,8	10,7	11,2	4,0	4,1	4,3	4,2	6,3
5,2	4,7	6,0	5,6	8,2	3,7	3,6	7,5	5,3	7,3	6,5	8,4	7,9	12,2
7,4	7,2	8,7	4,2	6,6	7,3	13,0	12,4	8,7	6,0	8,5	5,1	6,5	9,4
5,0	4,8	5,2	3,7	10,0	6,0	8,6	8,6	6,9	7,6	6,7	8,5	6,9	6,0
6,8	8,8	7,7	9,8	7,8	7,8	7,2	4,0	10,4	5,4	8,6	5,2	3,9	7,1
9,2	4,9	8,1	2,6	6,0	6,0	10,6	6,9	7,5	10,5	6,1	7,5	10,9	5,4
6,1	6,2	5,9	1,3	3,5	8,0	10,8	7,9	3,8	8,3	10,3	13,0	8,5	3,1
7,5	7,5	5,7	6,7	5,5	10,2	8,0	10,5	5,3	7,4	12,9	10,3	5,1	7,6
3,3	3,2	7,6	1,0										

Решение. Составим вариационный ряд, расположив результаты эксперимента в порядке возрастания.

1,0	1,2	1,3	1,6	1,7	1,9	2,5	2,6	2,8	2,9	3,0	3,1	3,2	3,3
3,4	3,5	3,5	3,6	3,7	3,7	3,8	3,9	4,0	4,0	4,0	4,1	4,2	4,2
4,3	4,5	4,6	4,7	4,8	4,8	4,8	4,9	5,0	5,0	5,0	5,0	5,1	5,1
5,1	5,1	5,2	5,2	5,2	5,2	5,2	5,3	5,3	5,3	5,4	5,4	5,4	5,4
5,4	5,4	5,5	5,6	5,6	5,7	5,7	5,8	5,9	6,0	6,0	6,0	6,0	6,0
6,0	6,0	6,0	6,1	6,1	6,1	6,2	6,3	6,3	6,4	6,5	6,5	6,6	6,7
6,7	6,8	6,8	6,9	6,9	6,9	6,9	6,9	7,0	7,0	7,1	7,1	7,1	7,2
7,2	7,3	7,3	7,3	7,4	7,4	7,4	7,4	7,5	7,5	7,5	7,5	7,5	7,6
7,6	7,6	7,6	7,7	7,7	7,8	7,8	7,8	7,8	7,9	7,9	7,9	8,0	8,0
8,0	8,0	8,1	8,1	8,1	8,1	8,2	8,2	8,2	8,3	8,3	8,3	8,4	8,4
8,5	8,5	8,5	8,6	8,6	8,6	8,7	8,7	8,8	8,8	8,8	8,8	8,8	8,9
8,9	8,9	9,0	9,0	9,1	9,2	9,4	9,4	9,5	9,5	9,6	9,7	9,7	9,8
9,8	9,9	10,0	10,0	10,0	10,2	10,2	10,3	10,3	10,4	10,4	10,5	10,5	10,6
10,7	10,8	10,9	10,9	11,0	11,2	11,4	11,6	11,8	12,0	12,2	12,4	12,6	12,7
12,8	12,9	13,0	13,0										

Минимальный элемент выборки равен 1,0, а максимальный элемент равен 13,0, число интервалов разбиения $k = 6$. Находим длину частичного интервала:

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{13 - 1}{6} = 2.$$

Заданный отрезок $[1;13]$ разбиваем на шесть равных интервалов, длина которых равна двум градусам Цельсия. В каждом из интервалов разбиения вычисляем число вариантов, которые попали в данный интервал, то есть находим частоты m_i . Составим интервальный статистический ряд, при этом укажем

середины интервалов, которые необходимы для построения гистограммы, а также накопленные относительные частоты для построения эмпирической функции распределения и кумуляты.

Номер интервала i	Частичные интервалы	Середины интервалов	Частоты m_i	Относительные частоты m_i/n	Накопленные частоты $\sum_{i=1}^n m_i/n$
1	[1;3)	2	10	0,05	0,05
2	[3;5)	4	26	0,13	0,18
3	[5;7)	6	56	0,28	0,46
4	[7;9)	8	64	0,32	0,78
5	[9;11)	10	30	0,15	0,93
6	[11;13]	12	14	0,07	1,00

Составим эмпирическую функцию распределения. По определению, эмпирическая функция распределения равна нулю, если $x \leq 1$.

Если случайная величина принимает значение из интервала $(1;3]$, то $F^*(x) = 0,05$. При расчётах используем формулу 1.1.1.

Если же случайная величина попадает в интервал $(3;5]$, то эмпирическая функция распределения примет значение $F^*(x) = 0,05 + 0,13 = 0,18$.

В случае попадания случайной величины в интервал $(5;7]$ значение эмпирической функции распределения равно $F^*(x) = 0,05 + 0,13 + 0,28 = 0,46$.

Для значений случайной величины, которые лежат в интервале $(7;9]$, эмпирическая функция распределения принимает числовое значение, которое равно $F^*(x) = 0,05 + 0,13 + 0,28 + 0,32 = 0,78$.

В интервале $(9;11]$ значение эмпирической функции распределения равно $F^*(x) = 0,05 + 0,13 + 0,28 + 0,32 + 0,15 = 0,93$.

На интервале $(11;13]$ эмпирическая функция принимает значение, которое равно значению $F^*(x) = 0,05 + 0,13 + 0,28 + 0,32 + 0,15 + 0,07 = 1$.

По определению, эмпирическая функция распределения равна единице, если случайная величина принимает значение, большее 13.

Таким образом, значения эмпирической функции распределения записаны в последнем столбце интервального статистического ряда. Запишем эмпирическую функцию распределения в аналитическом виде:

$$F^*(x) = \begin{cases} 0, & \text{если } x \leq 1; \\ 0,05, & \text{если } 1 < x \leq 3; \\ 0,18, & \text{если } 3 < x \leq 5; \\ 0,46, & \text{если } 5 < x \leq 7; \\ 0,78, & \text{если } 7 < x \leq 9; \\ 0,93, & \text{если } 9 < x \leq 11; \\ 1, & \text{если } 11 < x \leq 13; \\ 1, & \text{если } x > 13. \end{cases}$$

График заданной эмпирической функции распределения случайной величины ξ – температура воздуха в градусах Цельсия, измеряемая в регионе в течение суток, изображён на рисунке 1.1.1. На графике эмпирической функции распределения по оси абсцисс откладываются значения случайной величины, а по оси ординат отмечаются накопленные относительные частоты или частоты.

При построении гистограммы частостей или относительных частот на оси абсцисс откладываем частичные интервалы наблюдаемых значений заданной случайной величины ξ . На каждом из полученных интервалов строим прямоугольник с высотой m_i/nh , где h является длиной указанного интервала. С геометрической точки зрения, площадь построенного прямоугольника будет равна частости заданного частичного интервала. Гистограмма частостей заданной случайной величины ξ – температура воздуха в регионе в течение суток, изображена на рисунке 1.1.2 в виде прямоугольников.

Если на гистограмме частостей соединить середины верхних сторон элементарных прямоугольников, то полученная линия образует полигон частостей. На рисунке 1.1.2 полигон частостей заданной случайной величины изображён штриховой линией.

В случае построения кумуляты, на оси абсцисс откладываем частичные интервалы наблюдаемых значений заданной случайной величины ξ , а по оси ординат – накопленные частоты. Изображаем точки, абсциссы которых совпадают с правым концом частичного интервала, не входящего в него, а ординаты равны накопленным относительным частотам. Соединяя полученные точки плавной линией, строим график кумуляты. График кумуляты заданной случайной величины ξ – температура воздуха в данном регионе, изображён на рисунке 1.1.4 сплошной линией. На этом же рисунке изображён график эмпирической функции распределения (линии со стрелками).

Так как гистограмма заданной случайной величины ξ – температуры воздуха в регионе, которая изображена на рисунке 1.1.2, напоминает кривую нормального распределения, то можно выдвинуть гипотезу о нормальном распределении заданной случайной величины.

1.3 Задания для решения на практическом занятии

1.3.1 Найти эмпирическую функцию распределения по заданному распределению выборки:

x_i	2	4	7	10	12
m_i	5	26	44	18	7

1.3.2 Найти эмпирическую функцию распределения по заданному распределению выборки:

x_i	5	10	15	20	25	30
m_i	72	46	38	24	12	8

1.3.3 Результаты исследования непрерывной случайной величины ξ представлены в виде интервального статистического ряда:

Частичный интервал	[1;4)	[4;7)	[7;10)	[10;13)	[13;16]
Сумма частот вариант интервала	12	18	30	24	16

Найти эмпирическую функцию распределения и кумуляту. Построить их графики.

1.3.4 Результаты исследования непрерывной случайной величины η представлены в виде интервального статистического ряда:

Частичный интервал	[2;3)	[3;4)	[4;5)	[5;6)	[6;7]
Сумма частот вариант интервала	15	14	10	6	5

Найти эмпирическую функцию распределения и кумуляту. Построить их графики.

1.3.5 Построить полигон частот и полигон относительных частот по данному распределению выборки случайной величины ξ :

x_i	1	2	4	5	7
m_i	25	13	12	6	4

По виду графиков статистического ряда предположить закон распределения случайной величины ξ .

1.3.6 Построить полигон частот и полигон относительных частот по данному распределению выборки случайной величины η :

x_i	23	25	27	28	33	35
m_i	18	22	17	23	19	21

По виду графиков статистического ряда предположить закон распределения случайной величины η .

1.3.7 Построить гистограмму частот по данному распределению выборки случайной величины ξ (объём выборки равен $n = 100$):

Частичный интервал	[1;4)	[4;7)	[7;10)	[10;13)	[13;16]
Сумма частот вариант интервала	12	18	30	24	16

По виду гистограммы частот статистического распределения выдвинуть гипотезу о законе распределения случайной величины ξ .

1.3.8 Построить гистограмму относительных частот по данному распределению выборки случайной величины η (объём выборки равен $n = 100$):

Частичный интервал	[1;4)	[4;7)	[7;10)	[10;13)	[13;16]
Сумма частот вариант интервала	12	18	30	24	16

По виду гистограммы относительных частот статистического распределения выдвинуть гипотезу о законе распределения случайной величины η .

1.3.9 Из генеральной совокупности случайной величины ξ – среднего роста человека, произведена выборка, в которой оказалось только пять значений 160, 165, 170, 175 и 180 сантиметров. Данные эксперимента представлены в виде таблицы.

160	180	165	175	165	165	170	165	175	165
170	170	160	170	180	170	175	175	165	170
165	175	170	175	170	175	165	170	170	175
175	170	175	165	175	160	180	175	175	175
170	165	180	170	170	170	175	180	160	170

Найти объём выборки и записать статистический ряд в виде распределения частот и относительных частот. Построить эмпирическую функцию распределения и кумуляту по статистическому распределению случайной величины. Построить полигон частот и полигон относительных частот статистического распределения. На основании графиков статистического ряда выдвинуть гипотезу о распределении заданной случайной величины ξ .

1.3.10 В результате опыта произведено изучение случайной величины η . Из генеральной совокупности извлечена выборка, которая представлена в виде таблицы.

4,0	2,8	8,6	5,7	7,3	2,4	10,0	4,7	9,6	3,2
12,8	14,0	2,4	10,4	3,8	9,4	2,5	7,5	2,3	13,2
2,7	8,7	4,2	3,6	4,4	6,0	7,4	6,2	11,0	2,0
8,8	2,2	11,8	4,3	6,5	5,5	4,6	3,4	6,3	8,1
5,3	4,0	6,7	8,5	3,2	8,3	2,1	8,2	4,8	5,2
7,0	6,8	3,7	2,0	5,6	4,5	11,4	3,8	7,6	4,9
6,9	5,9	2,6	6,6	11,6	2,2	5,4	5,2	3,3	6,4
3,0	7,1	11,2	4,3	8,4	3,5	8,9	12,4	8,0	2,5
13,6	4,1	7,2	3,0	2,8	10,8	3,9	4,9	5,1	7,7
2,6	3,1	5,8	9,5	10,2	4,6	6,1	2,9	3,6	5,0

Для заданной выборки построить сгруппированный статистический ряд, разбивая размах варьирования на шесть интервалов. Найти эмпирическую функцию распределения и построить её график. Построить гистограмму, полигон частот, полигон распределения частостей или относительных частот, а также кумуляту полученного статистического ряда.

1.4 Задания для контролируемой самостоятельной работы

1.4.1 Из генеральной совокупности случайной величины ξ извлечена выборка, которая представлена в виде таблицы. Для заданной выборки построить сгруппированный статистический ряд, разбивая размах варьирования на шесть интервалов. Найти эмпирическую функцию распределения и построить её график. Построить гистограмму, полигон частот, полигон распределения частот или относительных частот и кумуляту полученного статистического ряда.

1.4.1.1

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

1.4.1.2

364	330	324	341	356	347	357	361	328	344
325	358	340	372	329	366	337	370	354	360
343	348	325	350	362	336	363	345	359	324
350	333	352	327	335	363	348	356	338	372
332	368	369	334	348	347	326	331	341	339

1.4.1.3

12	1	72	40	32	58	6	33	64	11
36	50	14	3	16	44	23	54	8	68
0	26	38	15	42	5	46	21	40	19
48	13	2	30	56	17	31	7	20	52
24	10	28	66	4	34	60	45	35	9

1.4.1.4

400	140	215	180	225	110	200	320	290	150
245	240	325	230	270	220	280	210	300	295
175	255	235	130	350	170	215	160	245	200
200	310	260	265	220	340	230	285	205	265
250	190	210	330	270	275	100	240	325	375

1.4.1.5

38,0	36,5	36,0	36,7	37,2	36,5	37,6	36,1	36,0	37,5
36,8	37,7	37,4	38,2	36,2	36,8	36,6	37,1	36,4	36,5
37,2	38,1	37,0	36,3	36,8	36,3	37,3	36,7	37,8	38,2
36,4	36,9	37,8	37,5	36,4	38,3	37,7	36,1	36,6	36,3
37,6	37,3	36,6	37,1	37,9	36,2	36,3	38,4	36,8	37,9

1.4.1.6

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

1.4.1.7

98	40	135	92	68	88	86	74	98	140
43	95	64	49	106	75	55	82	115	80
85	118	46	63	90	20	72	65	76	78
60	96	30	66	73	52	82	84	35	120
100	62	94	103	80	70	130	109	82	58

1.4.1.8

1,8	11,5	4,9	6,0	2,6	7,8	10,9	6,3	5,0	3,8
7,2	2,2	6,6	8,6	7,5	9,0	11,8	0,2	9,6	12,0
4,3	10,0	8,4	2,4	0,6	10,6	3,0	9,4	0,0	5,5
2,0	4,6	1,0	10,3	5,6	2,8	0,4	3,2	6,9	9,8
8,0	8,2	12,0	5,2	8,8	4,5	9,2	11,2	3,4	1,6

1.4.1.9

105	260	345	340	95	368	275	35	455	60
690	124	55	80	245	25	169	460	187	570
115	10	133	205	20	160	70	178	40	196
5	391	230	0	151	50	520	65	290	45
225	110	15	142	414	320	30	575	65	310

1.4.1.10

14,0	6,4	2,4	7,5	6,8	2,8	7,0	9,6	7,5	7,8
6,2	10,0	6,0	9,8	11,0	7,0	3,0	7,4	5,4	12,0
2,0	4,2	6,5	13,0	4,6	9,4	7,2	11,5	7,6	7,0
4,0	8,0	10,5	4,4	2,6	4,8	13,0	5,2	9,2	5,8
8,3	3,8	8,6	6,6	6,5	10,7	5,0	3,5	2,1	8,9

1.4.1.11

225	238	235	251	248	253	241	234	256	249
250	227	223	236	240	236	221	243	233	220
232	234	228	239	221	231	254	225	242	231
226	244	247	222	230	247	235	255	223	246
238	224	252	229	237	220	244	226	245	232

1.4.1.12

36	8	0	12	22	44	4	5	6	7
8	17	54	21	30	23	15	5	34	9
27	10	20	1	41	3	50	16	17	35
9	19	29	38	13	15	24	33	26	18
18	1	11	32	2	14	3	25	8	45

1.4.1.13

175	460	490	470	505	400	310	435	150	397
455	265	465	230	420	100	420	520	389	450
400	475	341	415	357	365	205	381	445	320
325	405	125	349	245	425	373	435	535	470
250	333	410	280	450	295	430	130	440	190

1.4.1.14

65	79	51	60	90	52	78	64	53	92
72	66	88	62	63	83	55	88	69	54
86	73	67	81	76	91	84	72	56	85
57	87	80	75	54	71	93	68	79	70
86	60	74	89	82	77	55	85	86	58

1.4.1.15

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

1.4.1.16

40	130	94	48	96	115	56	84	125	88
100	92	138	76	160	152	98	128	86	38
90	10	104	142	108	80	82	115	60	98
134	72	44	106	78	52	112	70	158	120
70	102	74	64	148	110	156	190	118	68

1.4.1.17

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

1.4.1.18

200	20	150	65	160	125	80	170	280	45
25	210	15	250	70	45	190	135	40	180
100	55	110	10	120	230	130	35	10	145
195	25	175	220	5	75	30	245	300	240
50	105	60	115	165	0	260	85	95	90

1.4.1.19

7,0	2,0	7,6	15,5	9,9	11,0	14,0	10,2	9,4	11,8
10,0	13,5	12,4	10,6	5,0	8,5	11,2	17,0	11,6	15,0
5,5	10,2	18,0	7,9	10,8	1,0	8,8	11,4	10,0	9,7
12,0	7,3	4,0	12,6	19,0	12,8	17,5	9,1	4,5	16,5
14,5	12,2	10,4	13,0	8,2	16,0	10,0	6,0	10,4	3,0

1.4.1.20

196	161	151	184	193	153	166	155	204	177
157	197	185	191	183	195	158	203	167	178
160	186	198	163	172	182	202	180	179	168
169	189	162	199	164	201	173	187	154	175
187	159	170	171	200	165	181	174	191	205

1.4.1.21

10	55	35	11	41	47	39	3	66	5
22	1	12	33	15	2	50	29	30	20
1	13	8	43	26	16	28	7	19	32
14	23	0	25	9	27	17	53	5	11
7	44	24	6	60	21	4	18	35	6

1.4.1.22

75	128	154	134	104	152	143	154	166	124
125	172	120	150	158	199	155	215	149	168
170	225	174	100	137	160	179	164	112	164
116	152	131	187	155	200	162	195	162	85
150	191	95	156	183	140	108	146	175	148

1.4.1.23

51	77	56	75	15	42	23	58	48	31
29	53	41	14	47	16	65	22	69	60
45	28	13	57	32	24	17	46	21	72
70	12	27	40	25	39	37	18	33	20
11	43	55	30	59	61	44	35	19	50

1.4.1.24

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

1.4.1.25

90	54	50	35	74	50	51	78	65	38
53	59	85	63	58	48	34	59	52	52
32	70	62	56	30	80	58	51	42	66
61	55	54	72	57	55	60	36	58	80
40	69	59	45	64	53	76	57	75	56

1.4.1.26

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

1.4.1.27

120	10	240	175	180	30	90	55	330	110
50	130	5	140	15	190	170	250	105	200
60	260	55	0	10	20	210	100	270	40
25	65	220	75	150	160	35	125	45	290
360	210	70	60	80	85	280	30	230	300

1.4.1.28

25	29	15	5	16	23	14	7	22	9
10	14	26	18	12	1	19	17	14	17
20	15	11	13	27	16	24	2	24	8
18	21	15	0	19	12	28	11	16	10
4	9	16	22	23	6	13	17	3	30

1.4.1.29

810	840	930	819	882	925	852	980	958	898
960	965	900	934	822	942	890	954	833	945
870	813	970	866	938	886	950	831	915	860
905	990	844	975	910	825	904	894	990	836
839	874	816	878	848	946	828	985	856	920

1.4.1.30

5	32	18	24	12	26	7	28	9	3
40	1	34	11	1	13	38	23	2	30
16	17	10	19	48	0	14	39	46	10
7	9	31	5	20	36	22	15	29	6
8	42	6	27	4	21	36	3	8	44

2 СТАТИСТИЧЕСКАЯ ОЦЕНКА ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

2.1 Содержание: классификация точечных оценок, интервальные оценки параметров распределения, доверительные интервалы для нормального распределения.

2.2 Примеры решения типовых задач

2.2.1 Из генеральной совокупности, исследования случайной величины T – температура тела человека, измеряемая в течение определённого промежутка времени, извлечена выборка объёма $n=100$. Данные представлены в виде статистического ряда

варианта t_i	36,2	36,4	36,6	36,8	37,0
частота m_i	6	18	52	20	4

Найти выборочную среднюю температуру, выборочную и исправленную дисперсии измерения температуры.

Решение. Находим выборочную среднюю по формуле (2.1.3):

$$\bar{t}_B = \frac{\sum_{i=1}^5 m_i t_i}{n} = \frac{6 \cdot 36,2 + 18 \cdot 36,4 + 52 \cdot 36,6 + 20 \cdot 36,8 + 4 \cdot 37,0}{100} = 36,596.$$

Находим выборочную дисперсию по формуле (2.1.4):

$$D_B = \frac{\sum_{i=1}^5 m_i (t_i - \bar{t}_B)^2}{n} = \frac{6 \cdot (36,2 - 36,596)^2 + 18 \cdot (36,4 - 36,596)^2 + 52 \cdot (36,6 - 36,596)^2 + 20 \cdot (36,8 - 36,596)^2 + 4 \cdot (37,0 - 36,596)^2}{100} = 0,031184.$$

Находим исправленную дисперсию по формуле (2.1.5):

$$s^2 = \frac{n}{n-1} \cdot D_B = \frac{100}{100-1} \cdot 0,031184 = 0,0315.$$

2.2.2 Из генеральной совокупности исследуемой случайной величины ξ произведена выборка объёмом $n=100$, которые записаны в виде интервального статистического ряда

Частичный интервал	[2;4)	[4;6)	[6;8)	[8;10)	[10;12]
Сумма частот вариантов интервала	8	20	42	18	12

Найти выборочную среднюю, выборочную дисперсию и выборочное среднее квадратическое отклонение, исправленную дисперсию и исправленное среднее квадратическое отклонение наблюдаемых значений случайной величины ξ .

Решение. В начале находим середины интервалов и выбираем их в качестве вариант наблюдаемых значений случайной величины ξ : $x_1=3$, $x_2=5$, $x_3=7$, $x_4=9$, $x_5=11$.

Находим выборочную среднюю по формуле (2.1.3):

$$\bar{x}_B = \frac{\sum_{i=1}^5 m_i x_i}{n} = \frac{8 \cdot 3 + 20 \cdot 5 + 42 \cdot 7 + 18 \cdot 9 + 12 \cdot 11}{100} = 7,12.$$

Находим выборочную дисперсию по формуле (2.1.4):

$$D_B = \frac{\sum_{i=1}^5 m_i (x_i - \bar{x}_B)^2}{n} = \frac{8 \cdot (3 - 7,12)^2 + 20 \cdot (5 - 7,12)^2}{100} + \\ + \frac{42 \cdot (7 - 7,12)^2 + 18 \cdot (9 - 7,12)^2 + 12 \cdot (11 - 7,12)^2}{100} = 4,7056.$$

Находим исправленную дисперсию по формуле (2.1.5):

$$s^2 = \frac{n}{n-1} \cdot D_B = \frac{100}{100-1} \cdot 4,7056 = 4,7531.$$

Выборочное среднее квадратическое отклонением равно $\sigma_B = \sqrt{D_B} = \sqrt{4,7056} = 2,169$, а «исправленное» среднее квадратическое отклонение равно $s = \sqrt{s^2} = \sqrt{4,7531} = 2,18$.

2.2.3 Пусть случайная величина ξ имеет нормальное распределение с известным средним квадратическим отклонением $\sigma = 3$. Найти доверительный интервал для математического ожидания, если среднее выборочное наблюдаемых значений $n = 36$ прямых равноточных измерений равно $\bar{x}_B = 18$, а уровень значимости равен $\alpha = 0,05$.

Решение. Пользуясь таблицей функции Лапласа (приложение 3), определим квантиль стандартизированного нормального распределения, учитывая, что доверительная вероятность равна $P = 1 - \alpha = 1 - 0,05 = 0,95$:

$$2 \cdot \Phi(u_{\alpha/2}) = 1 - \alpha, \quad 2 \cdot \Phi(u_{\alpha/2}) = 0,95, \quad \Phi(u_{\alpha/2}) = 0,475, \quad u_{\alpha/2} = 1,96.$$

Для определения доверительного интервала для математического ожидания воспользуемся неравенством (2.1.7): $\bar{x}_B - u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} < a < \bar{x}_B + u_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$.

$$18 - 1,96 \cdot \frac{3}{\sqrt{36}} < a < 18 + 1,96 \cdot \frac{3}{\sqrt{36}} \quad \text{или} \quad 17,02 < a < 18,98.$$

Следовательно, искомый доверительный интервал имеет вид (17,02;18,98).

Раскроем смысл полученного результата: если будет произведено достаточно большое число выборок заданного объёма, то в 95 % случаев из них доверительные интервалы накроют математическое ожидание и только в 5 % случаев оцениваемое математическое ожидание может выйти за пределы доверительных интервалов.

2.2.4 Найти минимальный объём выборки, на основании которой можно было бы оценить математическое ожидание времени исполнения некоторой операции с ошибкой, не превышающей 15 часов, и надёжностью $P = 1 - \alpha = 0,9$, если предположить, что время выполнения указанной операции ξ является

случайной величиной, имеющей нормальное распределение с средним квадратическим отклонением $\sigma = 60$.

Решение. Пользуясь таблицей функции Лапласа (приложение 3), определим квантиль стандартизированного нормального распределения:

$$2 \cdot \Phi(u_{\alpha/2}) = 1 - \alpha, \quad 2 \cdot \Phi(u_{\alpha/2}) = 0,9, \quad \Phi(u_{\alpha/2}) = 0,45, \quad u_{\alpha/2} = 1,64.$$

Воспользуемся формулой, связывающей предельную погрешность ε оценки математического ожидания по выборочной средней: $\varepsilon = u_{\alpha/2} \sigma / \sqrt{n}$.

Следовательно, $n \geq \frac{u_{\alpha/2}^2 \cdot \sigma^2}{\varepsilon^2} = \frac{1,64^2 \cdot 60^2}{15^2} = 43,0336$. Таким образом, $n = 44$.

2.2.5 В результате исследования некоторой случайной величины ξ – линейный размер некоторого объекта, подчинённого нормальному закону распределения, произведена выборка объёмом $n = 5$: $x_1 = 2, x_2 = 2,1, x_3 = 2,2, x_4 = 2,1, x_5 = 2$. Оценить с помощью доверительного интервала математическое ожидание линейного размера. Доверительную вероятность $P = 1 - \alpha$ принять равной 0,99. Найти минимальное число измерений, которое необходимо выполнить, чтобы с надёжностью $1 - \alpha = 0,99$ можно было утверждать, что предельная погрешность точечной оценки исследуемой величины не превышала 0,002. Построить доверительный интервал, покрывающий среднее квадратическое отклонение σ с надёжностью $P = 1 - \alpha = 0,99$.

Решение. Найдём точечные оценки неизвестных параметров: математического ожидания a и среднего квадратического отклонения σ .

$$\hat{a} = \bar{x}_B = \frac{2 \cdot 2 + 2 \cdot 2,1 + 1 \cdot 2,2}{5} = 2,08.$$

$$\hat{\sigma} = s = \sqrt{\frac{1}{5-1} (2 \cdot (2 - 2,08)^2 + 2 \cdot (2,1 - 2,08)^2 + 1 \cdot (2,2 - 2,08)^2)} = 0,08.$$

По таблице распределения Стьюдента (приложение 5) по доверительной вероятности $P = 1 - \alpha = 0,99$ и числу степеней свободы $\nu = n - 1 = 4$ находим квантиль t -распределения: $t_{\alpha/2; n-1} = t_{0,005; 4} = 4,604$. Для определения доверительного интервала воспользуемся неравенством (2.1.8).

$$\bar{x}_B - t_{\alpha/2; n-1} \cdot \frac{s}{\sqrt{n}} < a < \bar{x}_B + t_{\alpha/2; n-1} \cdot \frac{s}{\sqrt{n}}.$$

$$2,08 - 4,604 \cdot \frac{0,08}{\sqrt{5}} < a < 2,08 + 4,604 \cdot \frac{0,08}{\sqrt{5}} \text{ или } 1,915 < a < 2,245.$$

Следовательно, искомый доверительный интервал имеет вид (1,915; 2,245). Таким образом, если будет произведено достаточно большое число выборок заданного объёма, то в 99 % случаев из них доверительные интервалы накроют математическое ожидание и только в 1 % случаев оцениваемое математическое ожидание может выйти за пределы доверительных интервалов.

Для расчёта истинного числа измерений, необходимых для определения истинного линейного размера объекта с погрешностью, не превышающей $\varepsilon = 0,02$, воспользуемся формулой предельной погрешности: $\varepsilon = t_{\alpha/2; n-1} s / \sqrt{n}$.

$$\text{Следовательно, } n \geq \frac{t_{\alpha/2; n-1}^2 \cdot s^2}{\varepsilon^2} = \frac{(4,604)^2 \cdot (0,08)^2}{(0,002)^2} \approx 33,91 \text{ измерения.}$$

Таким образом, для того чтобы предельная погрешность ε истинного значения линейного размера исследуемого объекта не превышала $\varepsilon = 0,002$, следует, помимо пяти проделанных измерений, выполнить ещё $n - 5 = 34 - 5 = 29$ измерений.

По заданной доверительной вероятности $P = 1 - \alpha = 0,99$ и числу степеней свободы $\nu = n - 1 = 5 - 1 = 4$, используя таблицу доверительных интервалов для среднего квадратического отклонения σ (приложение 7), находим нижнюю границу $\gamma_1 = 0,519$ и верхнюю границу $\gamma_2 = 4,390$. Для определения доверительного интервала для среднего квадратического отклонения воспользуемся неравенством (2.1.9): $s \cdot \gamma_1 < \sigma < s \cdot \gamma_2$.

Следовательно, искомый доверительный интервал равен:

$$0,08 \cdot 0,519 < \sigma < 0,08 \cdot 4,390 \text{ или } 0,04152 < \sigma < 0,3512.$$

Таким образом, искомый доверительный интервал имеет вид $(0,04152; 0,3512)$. То есть, если будет произведено достаточно большое число выборок заданного объёма, то в 99 % случаев из них доверительные интервалы накроют среднее квадратическое отклонение и только в 1 % случаев оцениваемое среднее квадратическое отклонение может выйти за пределы доверительных интервалов.

2.3 Задания для решения на практическом занятии

2.3.1 Из генеральной совокупности, исследования СВ ξ , извлечена выборка объёма $n = 100$. Данные представлены в виде статистического ряда:

варианта x_i	7,2	7,6	8,4	8,0	7,4
частота m_i	8	20	45	15	12

Найти выборочную среднюю, выборочную и исправленную дисперсии и соответствующие средние квадратические отклонения.

2.3.2 В итоге пяти измерений длины стержня одним прибором (без систематических ошибок) получены следующие результаты (в мм): $x_1 = 70$, $x_2 = 72$, $x_3 = 81$, $x_4 = 83$, $x_5 = 81$. Найти: а) выборочную среднюю длину стержня; б) выборочную и исправленную дисперсию ошибок прибора; в) выборочное и исправленное среднее квадратическое отклонение.

2.3.3 В итоге шести измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты: $x_1 = 2$, $x_2 = 5$, $x_3 = 8$, $x_4 = 7$, $x_5 = 5$, $x_6 = 3$. Найти: а) выборочную среднюю исследуемой

величины; б) выборочную и исправленную дисперсию; в) выборочное и исправленное среднее квадратическое отклонение.

2.3.4 Среди студентов факультета производится измерение роста. Из общего числа студентов, которые произвели измерения роста, произведена выборка в количестве 100 студентов. В результате получен интервальный статистический ряд.

Рост	Число студентов
[160;164)	9
[164;168)	15
[168;172)	20
[172;176)	29
[176;180)	17
[180;184)	7
[184;188]	3

Найти: а) выборочную среднюю роста обследованных студентов; б) выборочную и исправленную дисперсию роста обследованных студентов; в) выборочное и исправленное среднее квадратическое отклонение роста обследованных студентов.

2.3.5 В результате опыта произведено изучение случайной величины η . Из генеральной совокупности извлечена выборка, которая представлена в виде таблицы, представленной в задаче № 1.3.10. Найти: а) выборочную среднюю исследуемой величины, определённую в задаче № 1.3.10; б) выборочную и исправленную дисперсию данной случайной величины; в) выборочное и исправленное среднее квадратическое отклонение этой величины.

2.3.6 Найти доверительный интервал для оценки с надёжностью 0,98 неизвестного математического ожидания нормально распределённой случайной величины ξ генеральной совокупности, если заданы генеральное среднее квадратическое отклонение $\sigma = 4$, а выборочная средняя равна $\bar{x}_B = 20$ и объём выборки равен $n = 80$.

2.3.7 Выборка из большой партии электроламп содержит 200 ламп. Средняя продолжительность горения лампы выборки оказалась равной 800 часов. Найти с надёжностью 0,90 доверительный интервал для средней продолжительности горения лампы всей партии, если известно, что среднее квадратическое отклонение, продолжительности горения лампы равно 50 часов.

2.3.8 Станок-автомат штампует детали. По выборке объёма $n = 200$ вычислена выборочная средняя длины изготовленных деталей. Найти с надёжностью 0,99 точность ε , с которой выборочная средняя оценивает математическое ожидание линейного размера детали, если среднее квадратическое отклонение равно 10 мм.

2.3.9 Найти минимальный объём выборки, при котором с надёжностью 0,95, точность оценки математического ожидания генеральной совокупности нормально распределённой случайной величины по выборочной средней будет равна $\varepsilon = 0,4$, если среднее квадратическое отклонение равно $\sigma = 5$.

2.3.10 По данным независимых равноточных измерений некоторой физической величины найдена средняя выборочная $\bar{x}_B = 55$ и исправленное среднее квадратическое отклонение $s = 4$. Оценить истинное значение измеряемой величины при помощи доверительного интервала с надёжностью $P = 1 - \alpha = 0,95$.

2.3.11 По данным выборки объёма $n = 50$ из генеральной совокупности найдено исправленное среднее квадратическое отклонение $s = 2$ нормально распределённой случайной величины. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надёжностью 0,9.

2.3.12 Найти минимальный объём выборки, на основании которой можно было бы оценить математическое ожидание времени исполнения некоторой работы с ошибкой, не превышающей 10 часов, и надёжностью $P = 1 - \alpha = 0,99$, если предположить, что время выполнения указанной работы ξ является случайной величиной, имеющей нормальное распределение со средним квадратическим отклонением $\sigma = 80$.

2.3.13 В результате исследования некоторой случайной величины ξ , подчинённой нормальному закону распределения, произведена выборка: $x_1 = 3,5$, $x_2 = 4,1$, $x_3 = 5,7$, $x_4 = 4,1$, $x_5 = 3,5$. Оценить с помощью доверительного интервала математическое ожидание измерений. Доверительную вероятность $P = 1 - \alpha$ принять равной 0,95. Найти минимальное число измерений, которое необходимо выполнить, чтобы с надёжностью 0,95 можно было утверждать, что предельная погрешность точечной оценки исследуемой величины не превышала 0,005. Построить доверительный интервал, покрывающий среднее квадратическое отклонение σ с надёжностью $P = 1 - \alpha = 0,95$.

2.3.14 Определить численность выборки при обследовании остатков на расчётных счетах у клиентов банка, чтобы с вероятностью 0,9 предельная ошибка равнялась 15 денежным единицам, а среднее квадратическое отклонение – 350 денежным единицам.

2.3.15 При подсчёте объёмов ежедневных продаж джемперов предпринимателем в течение месяца получены следующие результаты: 1; 8; 7; 10; 13; 4; 9; 9; 10; 14; 8; 7; 9; 6; 8; 11; 3; 15; 5; 5; 11; 6; 5; 12; 6; 7; 8; 14; 11; 9. Найти: а) выборочную среднюю ежедневных продаж джемперов предпринимателем; б) выборочную и исправленную дисперсию; в) выборочное и исправленное среднее квадратическое отклонение объёмов продаж джемперов.

2.4 Задания для контролируемой самостоятельной работы

2.4.1 Из генеральной совокупности случайной величины ξ извлечена выборка, которая представлена в виде таблицы в задаче 1.4.1. Найти выборочную среднюю, выборочную дисперсию и выборочное среднее квадратическое отклонение, исправленную дисперсию и исправленное среднее квадратическое отклонение наблюдаемых значений случайной величины ξ .

2.4.2 Из генеральной совокупности нормально распределённой случайной величины ξ – месячной заработной платы рабочих предприятия, рассчитанной в условных денежных единицах, извлечена выборка: $x_1, x_2, x_3, x_4, x_5, x_6, x_7$. Требуется:

1) с помощью доверительного интервала оценить математическое ожидание случайной величины ξ – месячной заработной платы рабочих предприятия, при известном среднем квадратическом отклонении $\sigma = \sigma_0$ и уровне значимости $\alpha = \alpha_1$;

2) найти минимальный объём выборки, на основании которой можно было бы оценить математическое ожидание месячной заработной платы рабочих с ошибкой, не превышающей ε_1 условных денежных единиц месячной заработной платы при известном среднем квадратическом отклонении $\sigma = \sigma_0$ и уровне значимости $\alpha = \alpha_2$;

3) с помощью доверительного интервала оценить математическое ожидание нормально распределённой случайной величины ξ при неизвестном среднем квадратическом отклонении и уровне значимости $\alpha = \alpha_3$;

4) найти минимальный объём выборки, на основании которой можно было бы оценить математическое ожидание случайной величины ξ – месячной заработной платы рабочих с ошибкой, не превышающей ε_2 условных денежных единиц месячной заработной платы при неизвестном среднем квадратическом отклонении и уровне значимости $\alpha = \alpha_1$;

5) с помощью доверительного интервала оценить при уровне значимости $\alpha = \alpha_2$ среднее квадратическое отклонение нормально распределённой случайной величины ξ .

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	σ_0	α_1	α_2	α_3	ε_1	ε_2
2.4.1.1	3,2	3,5	3,7	4,0	3,7	3,5	3,2	0,5	0,05	0,01	0,40	0,164	0,023
2.4.1.2	3,8	4,2	4,6	5,0	4,6	4,2	3,8	0,6	0,10	0,02	0,01	0,154	0,038
2.4.1.3	4,5	5,0	5,5	6,0	5,5	5,0	4,5	1,0	0,20	0,05	0,02	0,125	0,047
2.4.1.4	5,2	5,8	6,4	7,0	6,4	5,8	5,2	1,2	0,01	0,10	0,05	0,105	0,015
2.4.1.5	5,9	6,6	7,3	8,0	7,3	6,6	5,9	1,4	0,02	0,01	0,20	0,172	0,059
2.4.1.6	6,6	7,4	8,2	9,0	8,2	7,4	6,6	1,6	0,40	0,02	0,10	0,147	0,065
2.4.1.7	7,3	8,2	9,1	10,0	9,1	8,2	7,3	1,8	0,20	0,05	0,40	0,115	0,077
2.4.1.8	7,7	8,8	9,9	11,0	9,9	8,8	7,7	2,0	0,02	0,10	0,01	0,136	0,083
2.4.1.9	5,4	7,6	9,8	12,0	9,8	7,6	5,4	4,0	0,05	0,01	0,02	0,183	0,009
2.4.1.10	3,1	6,4	9,7	13,0	9,7	6,4	3,1	5,0	0,01	0,02	0,05	0,199	0,092

2.4.1.11	0,7	0,8	0,9	1,0	0,9	0,8	0,7	0,2	0,10	0,05	0,20	0,167	0,024
2.4.1.12	1,4	1,6	1,8	2,0	1,8	1,6	1,4	0,5	0,40	0,10	0,10	0,152	0,039
2.4.1.13	2,1	2,4	2,7	3,0	2,7	2,4	2,1	0,7	0,05	0,01	0,40	0,124	0,043
2.4.1.14	1,3	1,7	2,1	2,5	2,1	1,7	1,3	0,9	0,10	0,02	0,01	0,106	0,012
2.4.1.15	2,0	2,5	3,0	3,5	3,0	2,5	2,0	1,2	0,20	0,05	0,02	0,173	0,051
2.4.1.16	3,3	3,7	4,1	4,5	4,1	3,7	3,3	0,8	0,01	0,10	0,05	0,148	0,068
2.4.1.17	4,0	4,5	5,0	5,5	5,0	4,5	4,0	1,1	0,02	0,01	0,20	0,113	0,072
2.4.1.18	4,7	5,3	5,9	6,5	5,9	5,3	4,7	1,2	0,40	0,02	0,10	0,138	0,084
2.4.1.19	5,4	6,1	6,8	7,5	6,8	6,1	5,4	1,4	0,10	0,05	0,40	0,185	0,007
2.4.1.20	6,1	6,9	7,7	8,5	7,7	6,9	6,1	1,6	0,05	0,10	0,01	0,196	0,098
2.4.1.21	6,8	7,7	8,6	9,5	8,6	7,7	6,8	1,6	0,20	0,01	0,02	0,163	0,028
2.4.1.22	7,5	8,5	9,5	10,5	9,5	8,5	7,5	2,0	0,01	0,02	0,05	0,155	0,035
2.4.1.23	6,7	8,3	9,9	11,5	9,9	8,3	6,7	3,2	0,02	0,05	0,10	0,127	0,044
2.4.1.24	4,4	7,1	9,8	12,5	9,8	7,1	4,4	3,0	0,40	0,10	0,20	0,107	0,017
2.4.1.25	2,1	5,9	9,7	13,5	9,7	5,9	2,1	4,1	0,05	0,01	0,40	0,179	0,056
2.4.1.26	0,9	1,1	1,3	1,5	1,3	1,1	0,9	0,4	0,10	0,02	0,01	0,143	0,062
2.4.1.27	1,7	2,0	2,3	2,6	2,3	2,0	1,7	0,6	0,20	0,05	0,02	0,117	0,075
2.4.1.28	2,4	2,8	3,2	3,6	3,2	2,8	2,4	0,9	0,01	0,10	0,05	0,132	0,081
2.4.1.29	1,0	1,6	2,2	2,8	2,2	1,6	1,0	0,8	0,02	0,01	0,20	0,187	0,005
2.4.1.30	1,6	2,3	3,0	3,7	3,0	2,3	1,6	1,4	0,40	0,02	0,10	0,193	0,096

3 СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ

3.1 Содержание: статистическая гипотеза, статистический критерий значимости проверки нулевой гипотезы, проверка гипотезы о нормальном распределении, статистическая проверка непараметрических гипотез по критерию согласия χ^2 Пирсона.

3.2 Примеры решения типовых задач

3.2.1 Результаты исследования 200 предприятий по стоимости основных производственных фондов (случайная величина ξ) в условных денежных единицах представлены в виде сгруппированного интервального статистического ряда:

Интервалы стоимости производственных фондов, у. е.	[19;20)	[20;21)	[21;22)	[22;23)	[23;24)	[24;25]
Частоты m_i	10	26	56	64	30	14

Проверить нулевую гипотезу о нормальном распределении стоимости производственных фондов предприятий при уровне значимости $\alpha = 0,05$.

Решение. Из условия следует, что точные параметры предполагаемого нормального распределения неизвестны, поэтому нулевая гипотеза может быть записана в виде: $H_0: F(x)$ является функцией нормального распределения с двумя параметрами, математическим ожиданием $M(x) = \hat{a} = \bar{x}_B$ и дисперсией $D(x) = \hat{\sigma}^2 = s^2$.

Для проверки предложенной нулевой гипотезы определяем значения x'_i середин интервалов и находим точечные оценки математического ожидания и среднего квадратического отклонения нормально распределённой случайной величины ξ .

$$\hat{a} = \bar{x}_B = \frac{\sum_{i=1}^6 m_i x'_i}{n} =$$

$$= \frac{10 \cdot 19,5 + 26 \cdot 20,5 + 56 \cdot 21,5 + 64 \cdot 22,5 + 30 \cdot 23,5 + 14 \cdot 24,5}{200} = 22,1 \text{ (у. е.);}$$

$$\hat{\sigma}^2 = s^2 = \frac{\sum_{i=1}^6 m_i (x'_i - \bar{x}_B)^2}{n-1} =$$

$$= \frac{10 \cdot (-2,6)^2 + 26 \cdot (-1,6)^2 + 56 \cdot (-0,6)^2 + 64 \cdot 0,4^2 + 30 \cdot 1,4^2 + 14 \cdot 2,4^2}{199} = 1,52 \text{ (y. e.)};$$

$$\hat{\sigma} = s = \sqrt{1,52} \text{ (y. e.)}.$$

Для проверки гипотезы о нормальном распределении случайной величины ξ с помощью критерия χ^2 в начале интервалы наблюдаемых значений нормируют, то есть выражают их в единицах среднего квадратического отклонения s : $u_i = (x_i - \bar{x}_B)/s$, причём наименьшее значение u_i полагают равным $(-\infty)$, наибольшее – $(+\infty)$.

Вычислим теоретические вероятности p_i попадания случайной величины ξ с математическим ожиданием 22,1 и средним квадратическим отклонением 1,52 условных денежных единиц в частичные интервалы $[x_{i-1}; x_i)$ по формуле

$$p_i = P(x_{i-1} \leq \xi < x_i) = \Phi(u_i) - \Phi(u_{i-1}); \quad (i = \overline{1;6}),$$

где $\Phi(u_i) = \frac{1}{\sqrt{2\pi}} \int_0^{u_i} e^{-\frac{t^2}{2}} dt$, значения которой находят по таблице значений функции Лапласа (приложение 3).

Например, вероятность того, что случайная величина попадёт в первый частичный интервал $(-\infty; 20)$, равна:

$$\begin{aligned} p_1 = P(-\infty < \xi < 20) &= \Phi\left(\frac{20 - 20,1}{1,233}\right) - \Phi\left(\frac{-\infty - 22,1}{1,233}\right) = \Phi(-1,70) - \Phi(-\infty) = \\ &= \Phi(+\infty) - \Phi(1,70) = 0,5 - 0,45543 = 0,04457. \end{aligned}$$

$$\begin{aligned} p_2 = P(20 \leq \xi < 21) &= \Phi\left(\frac{21 - 20,1}{1,233}\right) - \Phi\left(\frac{20 - 22,1}{1,233}\right) = \Phi(0,89) - \Phi(-1,70) = \\ &= \Phi(1,70) - \Phi(0,89) = 0,45543 - 0,31327 = 0,14216 \end{aligned}$$

и так далее. После этого вычисляются теоретические частоты нормального распределения $n_{теор} = np_i$ и наблюдаемое значение критерия $\chi^2_{набл} = \sum_{i=1}^k \frac{(m_i - np_i)^2}{np_i}$.

Вычисления, необходимые для определения наблюдаемого значения выборочной статистики χ^2 , приведём в таблице:

Интервалы изменения наблюдаемых значений СВ ξ	Частоты m_i	Нормированные интервалы $[u_{i-1}; u_i)$	p_i	np_i	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
[19; 20)	10	$(-\infty; -1,70)$	0,04457	9	1	0,11
[20; 21)	26	$[-1,70; -0,89)$	0,14216	28,4	5,76	0,20
[21; 22)	56	$[-0,89; -0,08)$	0,28918	56,2	0,04	0,0007
[22; 23)	64	$[-0,08; +0,73)$	0,29918	59,8	17,64	0,29
[23; 24)	30	$[+0,73; +1,54)$	0,17092	34,2	17,64	0,52
[24; 25]	14	$[+1,54; +\infty)$	0,06178	12,4	2,56	0,23
Σ	200		1,00000	200,0		$\chi^2_{набл} = 1,35$

В результате вычислений получили наблюдаемое значение критерия $\chi^2_{набл} = 1,35$. Найдём по таблице квантилей χ^2 -распределения по заданному уровню значимости $\alpha = 0,05$ и числу степеней свободы $\nu = k - r - 1 = 6 - 2 - 1 = 3$ критическое значение $\chi^2_{0,05;3} = 7,815$. Так как $\chi^2_{набл} = 1,35 < \chi^2_{0,05;3} = 7,815$, то нет оснований для отклонения нулевой гипотезы о нормальном распределении стоимости производственных фондов предприятий с параметрами $a = 22,1$ и $\sigma = 1,233$.

3.2.2 В цехе с 5 станками ежедневно регистрируется число вышедших из строя станков (случайная величина ξ). Всего проведено 100 наблюдений.

Получены следующие числовые данные $\left(n = \sum_{i=1}^6 m_i = 100 \right)$:

Число вышедших из строя станков	0	1	2	3	4	5
Частоты m_i	8	28	31	18	9	6

Проверить гипотезу о том, что распределение числа вышедших из строя станков (случайная величина ξ) имеет распределение Пуассона при уровне значимости $\alpha = 0,05$.

Решение. Согласно условию задачи, необходимо проверить нулевую гипотезу H_0 : $F(x)$ – функция распределения числа вышедших из строя станков

имеет вид $F(x, \mu) = \sum_{i=1}^x \frac{\mu^i \cdot e^{-\mu}}{i!}$ с параметром $\hat{\mu} = \bar{x}_B = \frac{8 \cdot 0 + 28 \cdot 1 + 31 \cdot 2 + 18 \cdot 3}{100} + \frac{9 \cdot 4 + 6 \cdot 5}{100} = 2,1$. Вычислим теоретические вероятности p_i выхода из строя x_i

станков в течение n наблюдений по формуле Пуассона (приложение 1):

$$p_i = P_n(x_i) = \frac{\hat{\mu}^{x_i} \cdot e^{-\hat{\mu}}}{x_i!} = \frac{(2,1)^{x_i} \cdot e^{-2,1}}{x_i!}; x_i = \overline{0;5}.$$

Вычисления, необходимые для определения наблюдаемого значения выборочной статистики χ^2 , приведём в таблице:

Число вышедших из строя станков	Частоты m_i	p_i	np_i	$(m_i - np_i)^2$	$\frac{(m_i - np_i)^2}{np_i}$
0	8	0,122	12,2	17,64	1,45
1	28	0,257	25,7	5,29	0,21
2	31	0,270	27,0	16,00	0,59
3	18	0,189	18,9	0,81	0,04
4	9	0,099	9,9	0,81	0,08
5	6	0,063	6,3	0,03	0,01
Σ	100	1,000	100,0		$\chi_{набл}^2 = 2,38$

В результате вычислений получили наблюдаемое значение критерия $\chi_{набл}^2 = 2,38$. Найдём по таблице квантилей χ^2 -распределения по заданному уровню значимости $\alpha = 0,05$ и числу степеней свободы $\nu = k - r - 1 = 6 - 1 - 1 = 4$ критическое значение $\chi_{0,05;4}^2 = 9,488$. Так как $\chi_{набл}^2 = 2,38 < \chi_{0,05;4}^2 = 9,488$, то нет оснований для отклонения нулевой гипотезы о том, что распределение числа вышедших из строя станков является распределением Пуассона с параметром $\mu = 2,1$.

3.3 Задания для решения на практическом занятии

3.3.1 Используя критерий χ^2 Пирсона, при уровне значимости 0,05 установить, согласуется ли гипотеза о нормальном распределении генеральной совокупности с данными выборки объёма $n = 100$.

Частичный интервал	[4;9)	[9;14)	[14;19)	[19;24)	[24;29)	[29;34)	[34;39]
Частоты m_i	5	9	14	41	15	10	6

3.3.2 Имеются данные распределения товарооборота 50 магазинов города за год. Данные товарооборота приведены в таблице, где в первом столбце указан товарооборот в условных денежных единицах, а во втором – число магазинов, товарооборот которых оказался в данном интервале.

Частичный интервал	Частоты m_i	Частичный интервал	Частоты m_i	Частичный интервал	Частоты m_i
[70;100)	1	[190;220)	8	[310;340)	2
[100;130)	2	[220;250)	8	[340;370)	2
[130;160)	5	[250;280)	6	[370;400)	1
[160;190)	8	[280;310)	5	[400;430)	2

Используя критерий χ^2 Пирсона, при уровне значимости 0,10 установить, согласуется ли гипотеза о нормальном распределении генеральной совокупности с данными выборки объема $n = 50$.

3.3.3 Из большой партии деталей было отобрано 500 штук с целью исследования закона распределения деталей по линейной длине. Результаты опытов приведены в таблице:

Частичный интервал	Частоты m_i	Частичный интервал	Частоты m_i
[19,85;19,95)	160	[20,35;20,45)	15
[19,95;20,05)	125	[20,45;20,55)	13
[20,05;20,15)	80	[20,55;20,65)	12
[20,15;20,25)	50	[20,65;20,75)	8
[20,25;20,35)	35	[20,75;20,85]	2

Используя критерий χ^2 Пирсона, при уровне значимости 0,02, установить, согласуется ли гипотеза о показательном распределении генеральной совокупности с данными выборки объема $n = 500$.

3.3.4 На телефонной станции производились наблюдения за числом неправильных соединений в минуту. Всего проведено 200 наблюдений. Получены следующие числовые данные $\left(n = \sum_{i=1}^6 m_i = 100 \right)$:

Число неверных соединений (случайная величина ξ)	0	1	2	3	4	5	6	7	8	9
Частоты m_i	8	12	20	29	31	33	27	24	10	6

Проверить гипотезу о том, что распределение числа вышедших из строя станков (случайная величина ξ) имеет распределение Пуассона при уровне значимости $\alpha = 0,1$.

3.3.5 Данные об урожайности ржи на различных участках сельскохозяйственного предприятия приведены в следующей таблице:

Урожайность, ц/га	[8;11)	[11;14)	[14;17)	[17;20)	[20;23)	[23;26)	[26;29]
Доля участка, % общей площади	3	10	24	30	20	8	5

Составить интервальный статистический ряд распределения относительных частот наблюдаемых значений непрерывной случайной величины ξ – урожайности ржи на различных участках сельхозугодий. Построить гистограмму и полигон относительных частот урожайности ржи. По видам гистограммы и полигона относительных частот случайной величины ξ сделать предварительный

выбор вида закона распределения урожайности ржи, в центнерах с гектара, на различных участках сельхозугодий. Найти эмпирическую функцию распределения и построить её график. Вычислить числовые характеристики выборки: среднюю выборочную, выборочную дисперсию, выборочное среднеквадратическое отклонение.

Найти интервальные оценки параметров нормального распределения (математического ожидания и среднего квадратического отклонения) с доверительной вероятностью $P=1-\alpha=0,98$.

Построить доверительный интервал для математического ожидания урожайности ржи, если известно среднее квадратическое отклонение $\sigma=6$ ц/га. Проверить согласие эмпирической функции распределения с модельной нормальной функцией распределения при помощи критерия χ^2 Пирсона.

3.4 Задания для контролируемой самостоятельной работы

3.4.1 При исследовании непрерывной случайной величины ξ , эксперимент продублировали n раз при одних и тех же условиях. В результате эксперимента все полученные значения случайной величины ξ попали в промежуток $[x_0; x_k]$. После обработки данных интервал наблюдаемых значений случайной величины разбит на интервалы равной длины $[x_0; x_1)$, $[x_1; x_2)$, ..., $[x_{k-1}; x_k]$. Результаты исследования случайной величины ξ представлены в виде интервального статистического ряда. В первой строке интервального ряда указываются интервалы наблюдаемых значений случайной величины ξ , а во второй – частоты m_i появления значений этой СВ ξ в указанном интервале $\left(\sum_i m_i = n\right)$.

1. Описать случайную величину ξ , которая могла бы соответствовать интервальному статистическому ряду вашего варианта, указав при этом единицы измерения значений выбранной случайной величины. При решении остальных пунктов задания необходимо указывать наименование вами выбранной случайной величины.

2. Составить интервальный статистический ряд распределения относительных частот наблюдаемых значений непрерывной случайной величины ξ . Построить гистограмму и полигон относительных частот СВ ξ . По видам гистограммы и полигона относительных частот случайной величины ξ сделать предварительный выбор вида закона распределения этой случайной величины.

3. Найти эмпирическую функцию распределения и построить её график.

4. Вычислить числовые характеристики выборки: среднюю выборочную, выборочную дисперсию (смещённую и несмещённые оценки), выборочное среднеквадратическое отклонение.

5. В зависимости от того какое предположение Вы сделали насчёт вида закона для случайной величины ξ , записать её плотность распределения вероятностей и функцию распределения.

6. Исходя из предположения, которое было выдвинуто относительно закона распределения случайной величины ξ , найти интервальные оценки параметров распределения (математическое ожидание и среднее квадратическое отклонение) с доверительной вероятностью $P=1-\alpha$, где $\alpha=0,01$, если номер варианта равен $4k$; $\alpha=0,02$, если номер варианта равен $4k+1$; $\alpha=0,05$, если номер варианта равен $4k+2$; $\alpha=0,1$, если номер варианта равен $4k+3$. По условию задачи среднее квадратическое отклонение неизвестно, поэтому для записи доверительного интервала для математического ожидания при известном среднем квадратическом отклонении принять его равным $\sigma = \frac{x_k - x_0}{2}$.

7. Проверить согласие эмпирической функции распределения с модельной нормальной функцией распределения при помощи критерия согласия χ^2 .

3.4.1.1	18 –	20 –	22 –	24 –	26 –	28 –	30 –	32 –	34 –	36 –
	20	22	24	26	28	30	32	34	36	38
	14	36	130	210	102	82	61	45	15	5

3.4.1.2	31 –	34 –	37 –	40 –	43 –	46 –	49 –	52 –	55 –	58 –
	34	37	40	43	46	49	52	55	58	61
	3	16	32	45	86	175	135	56	32	20

3.4.1.3	1 – 5	5 – 9	9 –	13 –	17 –	21 –	25 –	29 –	33 –	37 –
			13	17	21	25	29	33	37	41
	9	13	27	78	167	86	54	47	12	7

3.4.1.4	37 –	42 –	47 –	52 –	57 –	62 –	67 –	72 –	77 –
	42	47	52	57	62	67	72	77	82
	12	16	46	76	133	87	45	21	14

3.4.1.5	12 –	18 –	24 –	30 –	36 –	42 –	48 –	54 –	60 –
	18	24	30	36	42	48	54	60	66
	12	24	43	65	97	64	32	13	10

3.4.1.6	1,3 –	2,0 –	2,7 –	3,4 –	4,1 –	4,8 –	5,5 –	6,2 –	6,9 –
	2,0	2,7	3,4	4,1	4,8	5,5	6,2	6,9	7,6
	14	30	54	80	42	21	17	8	4

3.4.1.7	4,6 –	5,4 –	6,2 –	7,0 –	7,8 –	8,6 –	9,4 –	10,2 –	11,0 –
	5,4	6,2	7,0	7,8	8,6	9,4	10,2	11,0	11,8
	8	10	14	28	64	25	16	9	6

3.4.1.8	36,5 –	36,7 –	36,9 –	37,1 –	37,3 –	37,5 –	37,7 –	37,9 –	37,9 –
	36,7	36,9	37,1	37,3	37,5	37,7	37,9	37,9	38,1
	4	8	16	21	61	27	16	7	2

3.4.1.9	168 –	171 –	174 –	177 –	180 –	183 –	186 –	189 –
	171	174	177	180	183	186	189	192
	2	5	11	22	48	18	8	6

3.4.1.10	3,4 – 3,8	3,8 – 4,2	4,2 – 4,6	4,6 – 5,0	5,0 – 5,4	5,4 – 5,8	5,8 – 6,2	6,2 – 6,6		
	3	5	8	12	16	36	20	4		
3.4.1.11	65 – 66	66 – 67	67 – 68	68 – 69	69 – 70	70 – 71	71 – 72	72 – 73	73 – 74	75 – 76
	10	24	63	158	121	78	56	23	12	5
3.4.1.12	47,1 – 47,3	47,3 – 47,5	47,5 – 47,7	47,7 – 47,9	47,9 – 48,1	48,1 – 48,3	48,3 – 48,5	48,5 – 48,7	48,7 – 48,9	
	6	18	38	87	126	96	26	12	5	
3.4.1.13	12,3 – 12,7	12,7 – 13,1	13,1 – 13,5	13,5 – 13,9	13,9 – 14,3	14,3 – 14,7	14,7 – 15,1	15,1 – 15,5	15,5 – 15,9	
	4	17	35	87	129	91	27	12	3	
3.4.1.14	189 – 194	194 – 199	199 – 204	204 – 209	209 – 214	214 – 219	219 – 224	224 – 229	229 – 234	
	4	6	13	34	75	98	70	12	3	
3.4.1.15	10,1 – 11,1	11,1 – 12,1	12,1 – 13,1	13,1 – 14,1	14,1 – 15,1	15,1 – 16,1	16,1 – 17,1	17,1 – 18,1	18,1 – 19,1	
	5	14	44	73	41	18	12	6	2	
3.4.1.16	162 – 166	166 – 170	170 – 174	174 – 178	178 – 182	182 – 186	186 – 190	190 – 194		
	6	11	15	32	56	29	14	5		
3.4.1.17	2,0 – 2,3	2,3 – 2,6	2,6 – 2,9	2,9 – 3,2	3,2 – 3,5	3,5 – 3,8	3,8 – 4,1	4,1 – 4,4		
	4	6	18	26	64	21	9	4		
3.4.1.18	8,15 – 8,18	8,18 – 8,21	8,21 – 8,24	8,24 – 8,27	8,27 – 8,30	8,30 – 8,33	8,33 – 8,36			
	4	8	18	38	21	10	6			
3.4.1.19	2304 – 2314	2314 – 2324	2324 – 2334	2334 – 2344	2344 – 2354	2354 – 2364	2364 – 2374			
	3	6	20	37	18	9	5			
3.4.1.20	158,2 – 158,4	158,4 – 158,6	158,6 – 158,8	158,8 – 160,0	160,0 – 160,2	160,2 – 160,4	160,4 – 160,6			
	3	10	24	68	29	15	5			
3.4.1.21	0,12 – 0,125	0,125 – 0,13	0,13 – 0,135	0,135 – 0,14	0,14 – 0,145	0,145 – 0,15	0,15 – 0,155			
	2	8	12	38	16	10	4			
3.4.1.22	350 – 351	351 – 352	352 – 353	353 – 354	354 – 355	355 – 356				
	4	10	14	32	19	11				

3.4.1.23	48 – 52	52 – 56	56 – 60	60 – 64	64 – 68	68 – 72
	3	12	32	17	12	4

3.4.1.24	15,1 – 15,3	15,3 – 15,5	15,5 – 15,7	15,7 – 15,9	15,9 – 16,1	16,1 – 16,3
	2	5	16	35	18	4

3.4.1.25	17,0 – 17,6	17,6 – 18,2	18,2 – 18,8	19,4 – 20,0	20,6 – 21,2	21,8 – 22,4	23,0 – 23,6	23,6 – 24,2
	3	5	16	41	56	39	18	6

3.4.1.26	0,45 – 0,85	0,85 – 1,25	1,25 – 1,65	1,65 – 2,05	2,05 – 2,45	2,45 – 2,85	2,85 – 3,25	3,25 – 3,65
	3	8	18	39	17	13	9	5

3.4.1.27	20 – 25	25 – 30	30 – 35	35 – 40	40 – 45	45 – 50	50 – 55	55 – 60	60 – 65
	3	9	36	93	158	86	39	5	3

3.4.1.28	14 – 17	17 – 20	20 – 23	23 – 26	26 – 29	29 – 32	32 – 35	35 – 38	38 – 41	41 – 44
	3	5	37	127	413	148	43	13	7	4

3.4.1.29	5,15 – 5,35	5,35 – 5,55	5,55 – 5,75	5,75 – 5,95	5,95 – 6,15	6,15 – 6,35
	2	9	32	47	38	10

3.4.1.30	1700 – 1720	1720 – 1740	1740 – 1760	1760 – 1780	1780 – 1800	1800 – 1820
	34	45	98	142	86	39