МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение

высшего образования

«Курганский государственный университет»

Кафедра алгебры, геометрии и методики преподавания математики

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

(Часть 1)

Методические указания для практических занятий и самостоятельной работы для студентов направления 01.03.01 «Математика»

Курган 2016

Кафедра алгебры, геометрии и методики преподавания математики

Направление подготовки: 01.03.01 «Математика»

Дисциплина: «Матесатическая статистика».

Составил: ст. преподаватель Е.А. Лукерьянова.

Утверждены на заседании кафедры «25» февраля 2016 г.

Рекомендованы методическим советом университета «17» декабря 2015 г.

ВВЕДЕНИЕ

Целью изучения дисциплины «Математическая статистика» является получение фундаментального образования, способствующего развитию личности.

Задачами дисциплины являются: изучение основ математической статистики; овладение методами и приемами решения задач математической статистики; овладение методами сбора, обработки и анализа статистических данных; формирование навыков проведения сплошного и выборочного наблюдения, умения выделять конкретное вероятностно-статистическое содержание в прикладных задачах учебной и профессиональной деятельности.

Краткое содержание дисциплины:
Выборочный метод; статистическая оценка параметров распределения; методы расчета сводных характеристик выборки; элементы теории корреляции; статистическая проверка статистических гипотез; однофакторный дисперсионный анализ.

Особая роль в подготовке студентов к профессиональной деятельности принадлежит самостоятельной работе, организуемой в процессе обучения. Настоящие методические указания предназначены для организации самостоятельной работы по изучению курса.

Методические указания включают в себя планы занятий, задания для рубежного контроля, таблицы, список литературы. Планы занятий содержат вопросы для повторения, теоретическую справку, образцы решения задач, задачи различного уровня сложности для решения в аудитории, самостоятельного решения, решения дома, таблицы значений функции $\varphi(x)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}$, $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{x^2}{2}}\,dz$, таблица значений $t_y=t(\gamma,n)$, $q_\gamma=q(\gamma,n)$ критические точки распределения x^2.

Цель данного пособия – оказать помощь студентам в организации самостоятельной работы по курсу «Математическая статистика», связанной с подготовкой к практическим занятиям, текущему и рубежному контролю, подготовкой к сдаче зачета по дисциплине «Математическая статистика».

Тема 1. Статистическое распределение выборки
Теоретическая справка

Математическая статистика – раздел математики, в котором изучаются методы сбора, систематизации и обработки результатов наблюдений массовых случайных явлений для выявления существующих закономерностей.
Предметом математической статистики является изучение случайных величин по результатам наблюдений.

Совокупность всех подлежащих изучению объектов или возможных результатов всех мыслимых наблюдений, производимых в неизменных условиях над одним объектом, называется генеральной совокупностью.
Итак, генеральная совокупность – это случайная величина, заданная на пространстве элементарных событий Ω с выделенным в нем классом S подмножеств событий, для которых указаны их вероятности.

Выборочной совокупностью называется совокупность случайно отобранных из генеральной совокупности объектов.

Выборка – это последовательность $X_1, X_2, \ldots, X_n$ независимых одинаково распределенных случайных величин, распределение каждой из которых совпадает с распределением генеральной случайной величины.

Число объектов в генеральной совокупности называется объемом выборки. Обозначается N. Число объектов в выборочной совокупности называется объемом выборки. Обозначается n.

Пусть для изучения количественного (дискретного или непрерывного) признака X из генеральной совокупности извлечена выборка $x_1, x_2, \ldots, x_к$ объема n. Наблюдавшиеся значения x_i признака X называют *вариантами,* а последовательность вариант, записанных в возрастающем порядке, - *вариационным рядом*.

Статистическое распределение выборки называют перечнем вариант x_i вариационного ряда и соответствующих им частот n_i (сумма всех частот выборки n) или относительных частот w_i (сумма всех относительных частот равна единице).

Статистическое распределение выборки можно задать также в виде последовательности интегралов и соответствующих им частот (в качестве частоты интервала принимают сумму частот вариант, попавших в этот интервал).

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $\mathrm{F}^*(\mathrm{x})$, определяющую для каждого значения x относительную частоту события X<x:
$\mathrm{F}^*(\mathrm{x}) = \mathrm{n_x}/\mathrm{n}$, где $\mathrm{n_x}$ - число вариант, меньших x; n – объем выборки;

Эмпирическая функция обладает следующими свойствами.

Свойство 1. *Значения эмпирической функции принадлежат отрезку* $[0; 1]$.

Свойство 2. $\mathrm{F}^*(\mathrm{x})$ - *неубывающая функция.*

Свойство 3. *Если* x_1 *- наименьшая варианта, а* $x_к$ *- наибольшая, то* $F^*(x) = 0$ *при* $x \leq x_1$ *и* $F^*(x) = 1$ *при* $x \geq x_k$.

Полигоном частот называют ломаную, отрезки которой соединяют точки (x_1, n_1), (x_2, n_2),…, (x_k, n_k), где x_i - варианты выборки, n_i - соответствующие им частоты.

Полигоном относительных частот называют ломаную, отрезки которой соединяют точки (x_1, w_1), (x_2, w_2),…, (x_k, w_k), где x_i - варианты выборки, w_i - соответствующие им относительные частоты.

При непрерывном распределении признака весь интервал, в котором заключены все наблюдаемые значения признака, разбивают на ряд частичных интервалов длины h и находят частоту n_i - сумму частот вариант, попавших в i-интервал.

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины h, а высоты равны отношению $n_i/$ h (плотность частоты).

Площадь частичного i - го прямоугольника равна $h(n_i/$ h)= n_i - сумме частот вариант, попавших в i - й интервал. Площадь гистограммы частот равна сумме всех частот, т.е. объему выборки n.

Гистограммой относительных частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины h, а высоты равны отношению $w_i/$ h (плотность относительной частоты). Площадь гистограммы относительных частот равна сумме всех относительных частот, т.е. единице.

Образцы решения типовых задач

1 Из партии однотипных высокоомных сопротивлений отобрано для контроля 10 штук. Измерения дали следующие отклонения от номинала (в килоомах).

Таблица 1– Отклонения от номинала (в килоомах)

№ сопротивления	1	2	3	4	5	6	7	8	9	10
Отклонение	1	3	-2	2	4	2	3	3	-2	4

Построить вариационный ряд, статистическое распределение частот, эмпирическую функцию распределения признака.

Решение:

Вариационный ряд: -2; -2; 1; 2; 2; 3; 3; 4; 4; 5.

Таблица 2 – Статистическое распределение частот

x_i	-2	1	2	3	4	5
n_i	2	1	2	2	2	1

$n = 10.$

Для построения эмпирической функции распределения и ее графика найдем относительные частоты появления вариант.

$W_1 = \frac{2}{10}$; $W_2 = \frac{1}{10}$; $W_3 = \frac{2}{10}$; $W_4 = 0{,}2$; $W_5 = 0{,}2$; $W_6 = \frac{1}{10}$.

$$F^*(x) = W(X < x).$$

Пусть $x \leq -2$, тогда $F^*(x) = W(X < -2) = 0$.

Пусть $-2 < x \leq 1$, тогда $F^*(x) = W(X < 1) = W(X = -2) = 0{,}2$.

Пусть $1 < x \leq 2$,

тогда $F^*(x) = W(X < 2) = W(X = -2) + W(x = 1) = \frac{2}{10} + \frac{1}{10} = \frac{3}{10}$.

Пусть $2 < x \leq 3$,тогда

$F^*(x) = W(X < 3) = W(X = -2) + W(x = 1) + W(x = 2) = \frac{2}{10} + \frac{1}{10} + \frac{2}{10} = \frac{5}{10}$.

Пусть $3 < x \leq 4$,тогда

$$F^*(x) = W(X < 4) = W(X = -2) + W(x = 1) + W(x = 2) + W(x + 3) = \\ = \frac{2}{10} + \frac{1}{10} + \frac{2}{10} + \frac{2}{10} = \frac{7}{10}$$

Пусть $4 < x \leq 5$,тогда $F^*(x) = W(X < 5) = \frac{9}{10}$.

Пусть $x > 5$,тогда $F^*(x) = 1$.

$$F^*(x) = f(x) = \begin{cases} 0, & x \leq -2, \\ 0{,}2, & -2 < x \leq 1, \\ 0{,}3, & 1 < x \leq 2, \\ 0{,}5, & 2 < x \leq 3, \\ 0{,}7, & 3 < x \leq 4, \\ 0{,}9, & 4 < x \leq 5, \\ 1, & x > 5 \end{cases}$$

Вопросы для повторения

1 Что изучает математическая статистика, предмет математической статистики.

2 Основные задачи математической статистики.

3 Генеральная и выборочная совокупности.

4 Объем генеральной совокупности, объем выборочной совокупности.

5 Повторные и бесповторные выборки.

6 Репрезентативность выборки.

7 Вариационный ряд.

8 Статистическое распределение выборки.

9 Эмпирическая функция распределения и ее свойства.

10 Полигон и гистограмма.

11 Кумулятивная кривая.

Задачи для решения в аудитории

1 На каждую сотню деталей, изготовляемых цехом, в среднем бывают две неудовлетворяющие стандарту (брак). Было произведено 10 партий по 100 изделий в каждой, отклонение числа обнаруженных бракованных изделий от среднего приведено в таблице 3.

Таблица 3 — Отклонение числа обнаруженных бракованных изделий от среднего

№ партии	1	2	3	4	5	6	7	8	9	10
Отклонение от среднего	-1	0	1	1	-1	1	0	-2	2	1

Построить вариационный ряд, статистическое распределение выборки, эмпирическую функцию распределения, полигон частот.

2 Дано распределение коров по проценту жирности молока (таблица 4).

Таблица 4 — Распределение коров по проценту жирности молока

Жирность молока, %	3,45-3,55	3,55-3,65	3,65-3,75	3,75-3,85	3,85-3,95	3,95-4,05	4,05-4,15	4,15-4,25	4,25-4,35
Число коров	1	1	3	4	7	5	2	1	1

Построить гистограмму и полигон частот.

3 Известны данные о посевных площадях картофеля (тыс. гектаров) по районам Курганской области:

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

Построить статистическое распределение частот, начертить полигон частот.

4 Выборка задана в виде распределения частот (таблица 5).

Таблица 5 — Выборка и ее частота

x_i	2	5	7
n_i	1	3	6

Найти распределения относительных частот.

5 Найти эмпирическую функцию по данному распределению выборки. Построить график этой функции.

Таблица 6 — Распределение выборки

x_i	2	5	7	8
n_i	1	3	2	4

6 Построить гистограмму частот по данному распределению выборки.

Таблица 7— Распределение выборки

Номер интервала, i	Частичный интервал	Сумма частот, вариант интервала
1	2 – 7	5
2	7 – 12	10
3	12 – 17	25
4	17 – 22	6
5	22 – 27	4

7 В 1889 – 1890 г. г. был измерен рост 1000 взрослых мужчин (рабочих московских фабрик). Результаты измерений указаны в таблице 8.

Таблица 8 — Рост 1000 взрослых мужчин

Рост(см)	Число мужчин	Рост(см)	Число мужчин
143 –146	1	167 –170	170
146 –149	2	170 –173	120
149 –152	8	173 –176	64
152 –155	26	176 –179	28
155 –158	65	179 –182	10
158 –161	120	182 –185	3
161 –164	181	185 –188	1
164 –167	201	Всего	1000

Построить полигон, гистограмму частот.

8 На 10 мая 1982 г. по районам Пермской области было посеяно яровых культур (в процентах к площади): 61, 58, 55, 32, 41, 45, 36, 43, 43, 51, 50, 34, 27, 22, 22, 16, 29, 22, 17, 20, 27, 13, 45, 15, 13, 6, 9, 15, 4, 15, 20, 13, 3, 4, 7, 6, 6, 11.

Построить интервальный и дискретный вариационные ряды. Взять равные интервалы длинною h = 10. Начертить полигон и гистограмму относительных частот.

9 Имеются следующие данные о жирности молока в колхозе (таблица 9).

Таблица 9 — Соотношение жирности и числа коров

Жирность, %	3,5	3,6	3,7	3,8	3,9	4	4,1	4,2
Число коров	8	11	28	43	71	52	29	8

Преобразовать статистическое распределение в интервальный ряд с равными интервалами так, чтобы приведенные в таблицу данные были центрами интервалов, построить гистограмму частот.

10 Уровень рентабельности предприятий лёгкой промышленности характеризуется следующими данными (таблица10).

Таблица 10 — Уровень рентабельности предприятий лёгкой промышленности

Уровень рентабельности, %	до5	5 – 10	10 – 15	15 – 20	20 – 25	30 – 35	35 и более
Центральное значение	2,5	7,5	12,5	17,5	22,5	27,5	Условно принимают 32,5

Постройте кумулятивную кривую.

Задачи для решения дома

1 Построить полигон частот и относительных частот по данным следующего статистического распределения. Распределение деталей по сортам (таблица 11).

Таблица 11 — Распределение деталей по сортам

Сорт	1	2	3	4 брак	Итого
Количество деталей	38	7	3	2	50

2 Работа пресс-автомата, изготавливающего шарики для подшипников, до регулировки пресса и после нее характеризуется следующими результатами проверки смещений полушарий изготовленных шаров (при допуске 0,125 мм). Данные проверок помещены в таблице 12. Построить полигон, гистограмму частот.

Таблица 12 — Смещение полушарий изготовленных шаров.

Смещение полушарий (мм)	Количество шариков	Смещение полушарий (мм)	Количество шариков
0,025 – 0,035	23	0,095 – 0,105	202
0,035 – 0,045	25	0,105 – 0,115	353
0,045 – 0,055	52	0,115 – 0,125	397
0,055 – 0,065	91		
0,065 – 0,075	98	Всего проверено	1550
0,075 – 0,085	111		
0,085 – 0,095	198		

3 Испытывалась чувствительность второго канала 40 телевизоров. Данные испытаний указаны в таблице, где в первой строке даны интервалы чувствительности в микровольтах, во второй – средние точки этих интервалов, в третьей – число телевизоров n_i,чувствительность которых оказалась в данном интервале.

Таблица 13 — Чувствительность телевизоров

Интервал	75 – 125	125–175	175–225	225–275	275–325	325–375	375–425
f_{co}	100	150	200	250	300	350	400
n_i	0	1	5	9	6	8	6
Интервал	425–475	475–525	525–575	575–625	625–675	675–725	
f_{co}	450	500	550	600	650	700	
n_i	2	2	0	0	1	0	

Построить эмпирическую функцию распределения, гистограмму частот.

4 Имеются следующие данные о производстве мяса на 100 га сельхозугодий в области за год (таблица 14).

Таблица 14 — Производство мяса в районах

Производство мяса (ц, живой вес)	17,5	22,5	27,5	32,5	37,5	42,5	47,5
Количество районов	3	4	5	7	6	5	2

Преобразовать данный статистический ряд в интервальный с равными интервалами так, чтобы приведенные в таблице значения вариант были центрами интервалов, построить гистограмму относительных частот (взять 6 интервалов длиною h = 5).

Тема 2. Статистические оценки параметров распределения. Точечные оценки

Теоретическая справка

Пусть требуется изучить количественный признак генеральной совокупности. Допустим, что из теоретических соображений удалось установить, какое именно распределение имеет признак. Возникает задача оценки параметров, которыми определяется это распределение.

Статистической оценкой неизвестного параметра теоретического распределения называют функцию от наблюдаемых случайных величин. $Q^* = f(x_1, x_2, ..., x_n)$, где $x_1, x_2, ..., x_n$ – выборка.

Для того чтобы статистические оценки давали «хорошие» приближения оцениваемых параметров, они должны удовлетворять определенным требованиям: оценка должна быть несмещенной, эффективной и состоятельной.

Несмещенной называют статистическую оценку Q^*, математическое ожидание которой равно оцениваемому параметру Q при любом объеме выборки, т. е. $M(Q^*) = Q$.

Смещенной называют оценку, математическое ожидание которой не равно оцениваемому параметру.

Эффективной называют статистическую оценку, которая при заданном объеме выборки n имеет наименьшую возможную дисперсию.

При рассмотрении выборок большого объема (n велико) к статистическим оценкам предъявляется требование состоятельности.

Состоятельной называют статистическую оценку, которая при $n \to \infty$ стремится по вероятности к оцениваемому параметру. Например, если дисперсия несмещенной оценки при $n \to \infty$ стремится к нулю, то такая оценка оказывается состоятельной.

Пусть изучается генеральная совокупность относительно количественного признака X.

Генеральной средней называют среднее арифметическое значений признака генеральной совокупности.

Если все значения признака различны, то $X_Г = \frac{x_1 + x_2 + \cdots + x_N}{N}$.

Если значения признака имеют частоты $N_1, N_2, ..., N_k$, где $N_1, + N_2 + ... + N_k = N$, то $X_Г = \frac{X_1 N_1 + X_2 N_2 + \cdots X_K N_K}{N}$.

Пусть для изучения генеральной совокупности относительно количественного признака X извлечена выборка объема n.

Выборочной средней называют среднее арифметическое значение признака выборочной совокупности.

Если все значения признака выборки различны, то $X_В = \frac{X_1 + X_2 + \cdots + X_n}{n}$.

если же все значения имеют частоты $n_1, n_2, ..., n_k$, то $X_В = \frac{X_1 n_1 + X_2 n_2 + \cdots X_K n_K}{n}$.

Выборочная средняя является несмещенной и состоятельной оценкой генеральной средней.

Если выборка представлена интервальным вариационным рядом, то за x_i принимают середины частичных интервалов.

Для того чтобы охарактеризовать рассеяние значений количественного признака X генеральной совокупности вокруг своего среднего значения, вводят сводную характеристику – генеральную дисперсию.

Генеральной дисперсией Dг называют среднее арифметическое квадратов отклонений значений признака генеральной совокупности от их среднего значения.
Если все значения признака генеральной совокупности объема N различны, то $D_{Г} = \frac{\sum_{i=1}^{N}\left(x_i - \overline{x_{Г}}\right)^2}{N}$. Если же значения признака имеют соответственно частоты $N_1, N_2, ..., N_k$, где $N_1, +N_2 + ... + N_k = N$, то $D_{Г} = \frac{\sum_{i=1}^{N} N_i\left(x_i - \overline{x_{Г}}\right)^2}{N}$.

Кроме дисперсии для характеристики рассеяния значений признака генеральной совокупности вокруг своего среднего значения пользуются сводной характеристикой – средним квадратическим отклонением.

Генеральным средним квадратическим отклонением (стандартом) называют квадратный корень из генеральной дисперсии: $\sigma_{г}=\sqrt{D_{г}}$.

Для того чтобы наблюдать рассеяние количественного признака значений выборки вокруг своего среднего значения, вводят сводную характеристику – выборочную дисперсию.

Выборочной дисперсией называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения.
Если значения признака выборки имеют частоты $n_1, n_2, ..., n_k$, то
$D_B = \frac{\sum_{i=1}^{k}(x_i - \bar{x}_B)^2 \cdot n_i}{n}$.
Для характеристики рассеивания значений признака выборки вокруг своего среднего значения пользуются сводной характеристикой – средним квадратическим отклонением. Выборочным средним квадратическим отклонением называют квадратный корень из выборочной дисперсии: $\sigma_{в}=\sqrt{D_{в}}$.
Вычисление дисперсии (выборочной или генеральной) можно упростить, используя следующую теорему.
Теорема. Дисперсия равна среднему квадратов значений признака минус квадрат общей средней: $D = \overline{x^2} - \left[\overline{x}\right]^2$.
Если выборка представлена интервальным вариационным рядом, то за x_i принимают середины частичных интервалов.

Выборочная дисперсия является смещенной оценкой генеральной дисперсии, т.е. математическое ожидание выборочной дисперсии не равно оцениваемой генеральной дисперсии, а равно М(D $_{г}) = \frac{n-1}{n} D_{в}$. Для исправления выборочной дисперсии достаточно умножить ее на дробь $\frac{n}{n-1}$, получим исправленную дисперсию, которую обычно обозначают через S^2. Исправленная дисперсия является несмещенной оценкой. В качестве оценки генеральной дисперсии принимают исправленную дисперсию. Для оценки среднего квадратического генеральной совокупности используют исправленное среднее квадратическое отклонение $S=\sqrt{S^2}$. Формулы для вычисления выборочной дисперсии и исправленной дисперсии отличаются только знаменателями. При достаточно больших n выборочная и исправленная дисперсии мало отличаются, поэтому на практике исправленной дисперсией пользуются, если $n<30$. Кроме выборочной средней и выборочной дисперсии применяются и другие характеристики вариационного ряда.

Модой M_0 называют варианту, которая имеет наибольшую частоту.

Например:

x_i	1	4	7	9
n_i	5	1	20	6

$M_0 = 7$

Медианой m_0 называют варианту, которая делит вариационный ряд на две части, равные по числу вариант.

Если $n = 2k+1$, то $m_e = x_{k+1}$.

Если $n = 2k,$ то $m_e = \frac{x_k+x_{k+1}}{2}$.

Например: для вариационного ряда 2, 3, 5, 6, 7 медиана равна 5, а для вариационного ряда 2, 3, 5, 6, 7, 9 медиана равна 5,5.

Размахом варьирования R называют разность между наибольшей и наименьшей вариантами.

$$\mathrm{R} = x_{max} - x_{min}.$$

Так, для вариационного ряда 1,3, 4, 5, 6, 10 размах варьирования равен 9.

Средним абсолютным отклонением Θ называют среднее арифметическое абсолютных отклонений.

$$\Theta = \frac{\sum_{i=1}^{k} n_i |x_i - \bar{x}_B|}{n}.$$

Так, если

x_i	1	3	6	16
n_i	4	10	5	1

То $\bar{x}_B = \frac{4\cdot1+10\cdot3+5\cdot6+1\cdot16}{20}$=4, $\Theta = \frac{4\cdot|1-4|+10\cdot|3-4|+5\cdot|6-4|+1\cdot|16-4|}{20}$=2,2.

<u>Коэффициентом вариации V</u>называют выраженное в процентах отношение выборочного среднего квадратического отклонения к выборочной средней.

$$V = \frac{\sigma_B}{\bar{x}_B} \cdot 100\%.$$

Коэффициент вариации используется для сравнения рассеяния двух вариационных рядов: тот из рядов имеет большее рассеяние, у которого коэффициент вариации больше.

Для предыдущего примера V=71,25% , так как $\bar{x}_\text{В} = 4, \sigma_\text{В} = \sqrt{8{,}1} \approx 2{,}85.$

$D_B = \frac{\sum_{i=1}^{4}(x_i - \bar{x}_B)^2 \cdot n_i}{n}, D_B = \frac{-12+10+20+144}{20} = 8{,}1.$

$\sigma_B = \sqrt{D_B}, v = \frac{2{,}85}{4} \cdot 100\% \approx 71{,}25\%.$

При построении интервального вариационного ряда необходимо помнить:

1) число интервалов k можно определить по формулам:
$$k \approx 1 + 3{,}3 \cdot \ln n,$$
$$k \approx 2 \cdot \ln n,$$
$$k \approx log_2 n + 1;$$
2) длина интервала $h \approx \frac{2}{k}$, при этом h и k подбирают так, чтобы $h \cdot k = R$;
3) за начало первого интервала можно взять x_{min} или $x_{min} - \frac{h}{2}$ (x_{min}принимают за середину первого интервала).

Образцы решения типовых задач

2 По данному распределению выборки дать характеристику распределения признака, вычислить для этого:

1)$\bar{x}$, 2) медиану, 3) дисперсию, среднее квадратическое отклонение, размах варьирования, среднее абсолютное отклонение, коэффициент вариации, 4) моду.

Таблица 15− Распределению выборки

Частичный интервал	1−5	5−9	9−13	13−17	17−21
Сумма частот вариант частичного интервала	10	20	50	12	8

Решение:

Составим вспомогательную таблицу центров интервалов.

Таблица 16 – Центры интервалов

Середина интервала	3	7	11	15	19
Число интервалов	10	20	50	12	8

1) $\bar{x}_\beta = \frac{3\cdot 10+7\cdot 20+11\cdot 50+15\cdot 12+19\cdot 8}{100} = \frac{1052}{100} = 10{,}52$; $\bar{x}_\beta = \frac{\sum_{i=1}^{k} n_i x_i}{n}$;

2) вычислим выборочную дисперсию

$D_B = \frac{\sum_{i=1}^{k} n_i (x_i - x_B)^2}{n}$;

$D_B = \frac{1}{100}(10\cdot (3-10{,}52)^2 + 20\cdot (7-10{,}52)^2 + 50\cdot (11-10{,}52)^2 + 12\cdot (15-10{,}52)^2 + +8\cdot (19-10{,}52)^2) = 16{,}4096$;

Дисперсию можно вычислить по формуле: $D_B = \overline{x^2} - (\bar{x}_B)^2$.

$D_B = \frac{1}{100}(9\cdot 10 + 49\cdot 20 + 121\cdot 50 + 225\cdot 12 + 361\cdot 8) - (10{,}52)^2 = 16{,}4096$;

3) $\sigma_B = \sqrt{D_B}$, $\sigma_B = \sqrt{16?\,4096 \approx 4{,}051;}$

4) R= $x_{max} - x_{min}$, R= $19-3 = 16$;

5) $\Theta = \frac{\sum_{i=1}^{k} n_i |x_i - \bar{x}_B|}{n}$;

$\Theta = \frac{1}{100}(10\cdot|3-10.52| + 20\cdot|7-10.52| + 50\cdot|11-10.52| + 12\cdot |15-10.52| + 8\cdot|19-10.52|) = 3.092$;

6) $= \frac{\sigma_В}{\bar{x}_в}\cdot 100\%$, $V = \frac{4{,}051}{10{,}52}\cdot 100\% = 38{,}5076\%$, $V \approx 38{,}51\%$;

7) $M_0 = 11$;

8) $m_e = 11$.

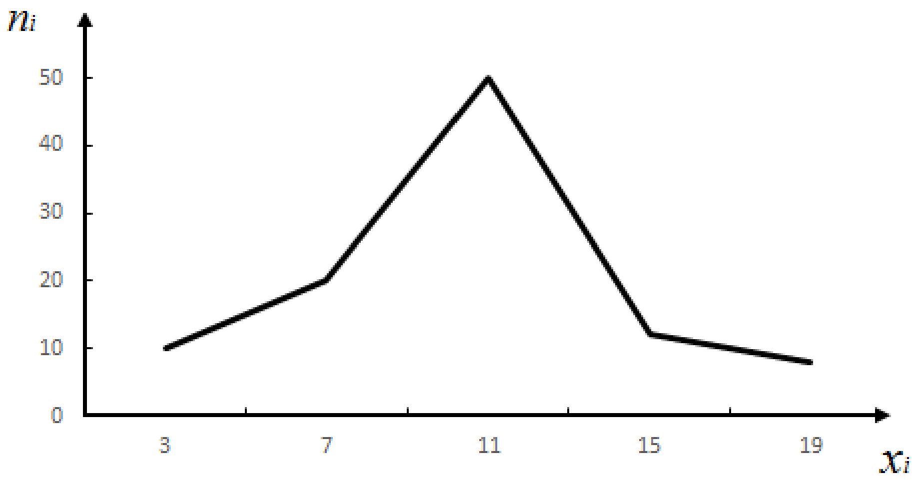

Рисунок 1 - Полигон

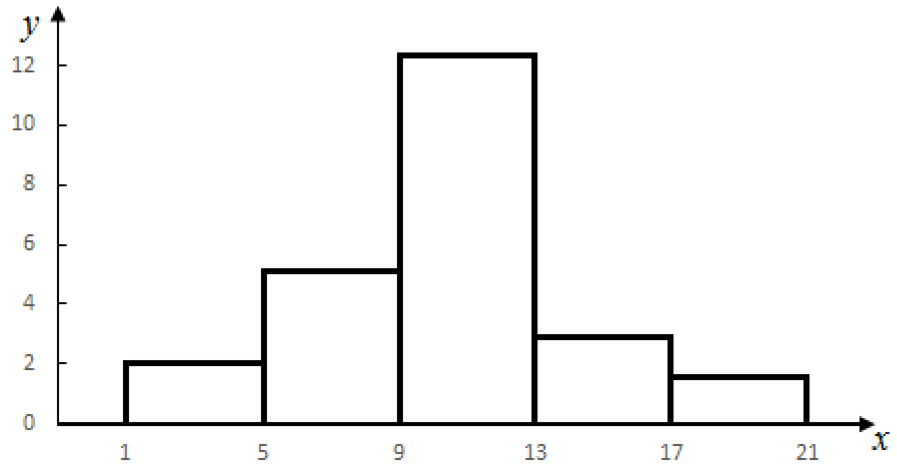

Рисунок 2 – Гистонрамма

Вопросы для повторения

1 Понятие статистической оценки параметров распределения.

2 Требования, предъявляемые к оценке параметров распределения.

3 Понятие точечной оценки.

4 Генеральная и выборочная средние.

5 Генеральная и выборочная дисперсия.

6 Оценка генеральной средней по выборочной среде.

7 Оценка генеральной дисперсии.

8 Другие характеристики вариационного ряда (мода, медиана, размах варьирования, среднее абсолютное отклонение, коэффициент вариации).

Задачи для решения в аудитории

1 Из генеральной совокупности извлечена выборка объема n = 60.

Таблица 17 — Выборка объема n = 60

x_i	1	3	6	26
n_i	8	40	10	2

Найти выборочную среднюю.

2 Найти выборочную среднюю по данному вариационному ряду:

x_i 1250, 1270, 1271, 1275, 1278.

3 Задано распределение первоначальных вариант выборки объема n:

$x_i x_1 x_2 \ \dots \ x_k$

$n_i n_1 n_2 \ \dots \ n_k$.

Доказать, что $\overline{x_B} = c + \frac{\sum_{i=1}^{k} n_i u_i}{n}$, где $u_i = x_i - c$.

4 Найти выборочную среднюю по данному распределению выборки объема n = 20.

Таблица 18 — Выборка объема n = 20

x_i	2560	2600	2620	2650	2700
n_i	2	3	10	4	1

5 Из генеральной совокупности извлечена выборка объема n = 50.

Таблица 19 — Выборка объема n = 50

x_i	2	5	7	10
n_i	16	12	8	14

Найти несмещенную оценку генеральной средней.

6 По выборке объема n = 41 найдена смещенная оценка $D_B = 3$ генеральной дисперсии. Найти несмещенную оценку дисперсии генеральной совокупности.

7 В итоге измерений длины стержня одним прибором (без систематических ошибок) получены следующие результаты (в мм): 92; 94; 103; 105; 106. Найти:

а) выборочную среднюю длину стержня;

б) выборочную и исправленную дисперсии ошибок прибора.

8 Ниже приведены результаты измерения роста (в см) случайно отобранных 100 студентов.

Таблица 20 — Рост студентов

Рост	154 – 158	158 – 162	162 – 166	166 – 170	170 – 174	174 – 178	178 – 182
Число студентов	10	14	26	28	12	8	2

Найти выборочную среднюю и выборочную дисперсию роста обследованных студентов.

9 Найти выборочную дисперсию по данному распределению выборки объема n = 10.

Таблица 21 — Выборка объема n = 10

x_i	186	192	194
n_i	2	5	3

Указание: Применить формулу $D_B = \frac{\sum_{i=1}^{k} n_1 u_1^2}{n} - \left(\frac{\sum_{i=1}^{k} n_i u_1}{n}\right)^2, u_i = x_i - c$.

10 Найти исправленную выборочную дисперсию по данному распределению выборки n = 10

Таблица 22 — Распределение выборки n = 10

x_i	102	104	108
n_i	2	3	5

11 Через час измерялось напряжение тока в электросети. Результаты измерений в вольтах представлены в таблице 23.

Таблица 23 — Изменение напряжения тока (x_i) по часам (i)

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	222	219	224	220	218	217	221	220	215	218	223	225
i	13	14	15	16	17	18	19	20	21	22	23	24
x_i	220	226	221	216	211	219	220	221	222	218	221	219

Найти оценки для математического ожидания и дисперсии результатов измерений.

12 Известно распределение рабочих механического цеха колхоза по тарифным разрядам (таблица 24).

Таблица 24 — Распределение рабочих по тарифным разрядам

Тарифные разряды	1	2	3	4	5	6
Количество рабочих	3	2	9	20	16	3

Дать характеристику распределения признака, вычислить для этого:

1) $\bar{x}$, 2) медиану, 3) дисперсию, среднее квадратическое отклонение, размах варьирования, среднее абсолютное отклонение, коэффициент вариации, 4) моду.

Построить полигон частот.

13 Дано распределение коров по проценту жирности молока (таблица 25).

Таблица 25 — Распределение коров по проценту жирности молока

Жирность молока, %	3,45 – 3,55	3,55 – 3,65	3,65 – 3,75	3,75 – 3,85	3,85–3,95	3,95 – 4,05	4,05 – 4,15	4,15 – 4,25	4,25–4,35
Число коров	1	1	3	4	7	5	2	1	1

Произвести анализ статистических данных таблицы 25. Построить гистограмму и полигон частот.

14 В ОТК были измерены диаметры 200 валиков из партии, изготовленной одним станком-автоматом. Отклонения измеренных диаметров от номинала даны в таблице 26 (в микронах).

Таблица 26 — Отклонения диаметров валиков от номинала

Границы уклонений	-20 – -15	-15 – -10	-10– -5	-5 – 0	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
Середины интервалов	-17,5	-12,5	-7,5	-2,5	2,5	7,5	12,5	17,5	22,5	27,5
Число валиков	7	11	15	24	49	41	26	17	7	3

Провести анализ статистических данных.

15 При сверлении отверстий одним и тем же сверлом и последующем измерении диаметров получены следующие данные (в мм):

40,25 40,35 40,45 40,35 40,37 40,35 40,44 40,35

40,33 40,41 40,35 40,30 40,28 40,30 40,4 40,36

40,29 40,33 40,31 40,33 40,41 40,4 40,33 40,37

40,35 40,34 40,32 40,34 40,28 40,46 40,39 40,37

40,29 40,39 40,37 40,44 40,27 40,38.

Найти выборочное среднее значение диаметра и среднее квадратическое отклонение, разбив данные измерения на 7 интервалов.

Задачи для решения дома

1 Известны данные об урожайности на 10 опытных участках одинакового размера (таблица 27).

Таблица 27 — Урожайность на участках

Участки	1	2	3	4	5	6	7	8	9	10
Урожайность ц/га	16,5	16,2	18,9	20,1	22,3	19,3	10,1	16	16,7	12,8

Дать характеристику распределения признака.

2 О ходе заготовки сочных кормов в хозяйствах области поступили следующие данные (таблица 28).

Таблица 28 — Заготовка кормов в хозяйствах

Заготовлено центнеров на условную голову скота	0,75	1,25	1,75	2,25	2,75
Количество хозяйств	8	10	14	9	2

Преобразовать данный вариационный ряд в интервальный с равными интервалами так, чтобы приведенные значения были центрами частот. Дать характеристику распределения признака.

3 Имеются данные о размерах основных фонтов (в млн руб.) 30 предприятий:

0,42; 0,24; 0,49; 0,67; 0,45; 0,27; 0,39; 0,21; 0,58; 0,4; 0,28; 0,73; 0,44; 0,66; 1,21; 0,62; 0,7; 0,81; 1,07; 0,68; 0,94; 0,76; 0,63; 0,88; 0,65; 1,14; 0,46; 1,38; 0,72; 0,91.

Построить интервальный вариационный ряд (h=0,2), приняв начало первого интервала равным 0,2 (млн руб.). Вычислить $\bar{x}_B$ двумя способами: по исходным данным и по данным построенного интервального ряда.

Тема 3. Методы точечных оценок

Вопросы для повторения

1 Дать определение обычного эмпирического момента.

2 Сформулировать определение начального эмпирического момента.

3 Что называется центральным эмпирическим моментом?

4 В чем состоит метод моментов точечной оценки?

5 Кем был предложен метод наибольшего правдоподобия?

6 Что называют функцией правдоподобия дискретной случайной величины X?

7 Что называют оценкой наибольшего правдоподобия?

8 Что называют логарифмической функцией правдоподобия?

9 Алгоритм нахождения точки максимума функции ln(L) аргумента Θ.

10 Достоинства и недостатки метода наибольшего правдоподобия.

11 Что называют функцией правдоподобия непрерывной случайной величины X?

Теоретическая справка

Метод моментов

Метод моментов точечной оценки неизвестных параметров заданного распределения состоит в приравнивании теоретических моментов рассматриваемого распределения соответствующим эмпирическим моментам того же порядка.

А. Оценка одного параметра. Пусть задан вид плотности распределения $f\,(x,\theta)$, определяемой одним неизвестным параметров θ. Требуется найти точечную оценку параметра θ.

Для оценки одного параметра достаточно иметь одно уравнение относительно этого параметра. Следуя методу моментов, приравниваем, например, начальный теоретический момент первого порядка начальному эмпирическому моменту первого порядка: $v_1 = M_1$. Учитывая, что $v_1 = M(X)$, $M_1 = \bar{x}_B (*)$, получим $\mathrm{M(X)} = \bar{x}_B$.

Б. Оценка двух параметров. Пусть задан вид плотности распределения $f(x;\ \theta_1, \theta_2)$, определяемой неизвестными параметрами θ_1 и θ_2. Для отыскания двух параметров необходимы два уравнения относительно этих параметров. Следуя методу моментов, приравниваем, например, начальный теоретический момент первого порядка начальному эмпирическому моменту первого порядка и центральный теоретический момент второго порядка центральному эмпирическому моменту второго порядка: $v_1 = M_1$, $\mu_2 = m_2$. Учитывая, что $v_1 = M(X)$, $\mu_2 = D(X)$, $M_1 = \bar{x}_B, m_2 = D_B$, получим $\left.\begin{matrix} M(X) = \bar{x}_B, \\ D(X) = D_B. \end{matrix}\right\} (**)$

Математическое ожидание и дисперсия есть функции от θ_1 и θ_2, поэтому (**) можно рассматривать как систему двух уравнений с двумя неизвестными θ_1 и θ_2. Решив эту систему относительно неизвестных параметров, тем самым получим их точечные оценки θ_1^* и θ_2^*. Эти оценки являются функциями от вариант выборки:

$\theta_1^* = \psi_1(x_1, x_2, \dots, x_n)$,

$\theta_2^* = \psi_2(x_1, x_2, \dots, x_n)$.

Метод наибольшего правдоподобия

А. Дискретные случайные величины. Пусть X – дискретная случайная величина, которая в результаты n испытаний приняла значения $x_1, x_2, \dots, x_n$. Допустим, что вид закона распределения величины X задан, но неизвестен параметр Θ, которым определяется этот закон. Требуется найти его точечную оценку.

Обозначим вероятность того, что в результате испытания величина X примет значение $x_i (i = 1, 2, \dots, n)$ через $p(x_i; \Theta)$.

Функцией правдоподобия дискретной случайной величины X называют функцию аргумента Θ:

$L\ (x_1, x_2, \dots, x_n; \Theta) = p(x_1; \Theta)\ p(x_2; \Theta) \dots p(x_n; \Theta)$, где $x_1, x_2, \dots, x_n$ – фиксированные числа.

В качестве точечной оценки параметра Θ принимают такое значение $\Theta^* = \Theta^*(x_1, x_2, \dots, x_n)$, при котором функция правдоподобия достигает максимума. Оценку Θ^* называют оценкой наибольшего правдоподобия.

Функции L и ln(L) достигают максимума при одном и том же значении Θ, поэтому вместо отыскания максимума функции L ищут (что удобнее) максимум функции ln(L).

Логарифмической функцией правдоподобия называют функцию ln(L). Как известно, точку максимума функции ln(L) аргумента Θ можно искать, например, так:

1) найти производную $\frac{d\ln(L)}{d\Theta}$;
2) приравнять производную нулю и найти критическую точку – корень полученного уравнения (его называют уравнением правдоподобия);
3) найти вторую производную $\frac{d^2\ln(L)}{d\Theta^2}$; если вторая производная при $\Theta = \Theta^*$ отрицательна, то Θ^* - точка максимума.

Найденную точку максимума Θ^* принимают в качестве оценки наибольшего правдоподобия параметра Θ.

Б. Непрерывные случайные величины. Пусть X – непрерывная случайная величина, которая в результате n испытаний приняла значения $x_1, x_2, \ldots, x_n$.

Допустим, что вид плотности распределения f(x) задан, но не известен параметр Θ, которым определяется эта функция.

Функцией правдоподобия непрерывной случайной величины X называют функцию аргумента Θ:

$L(x_1, x_2, \ldots, x_n;\ \Theta) = f(x_1;\ \Theta) f(x_2;\ \Theta) \ldots f(x_n;\ \Theta)$, где $x_1, x_2, \ldots, x_n$ – фиксированные числа.

Оценку наибольшего правдоподобия неизвестного параметра распределения непрерывной случайной величины ищут так же, как и в случае дискретной величины.

Образцы решения типовых задач

1 Найти методом наибольшего правдоподобия оценку параметра p биноминального распределения $p_n(k) = C_n^k p^k (1-p)^{n-k}$,

если в n_1 независимых испытаниях событие A появилось $x_1 = m_1$ раз и в n_2 независимых испытаниях событие A появилось $x_2 = m_2$ раз.

Решение. Составим функцию правдоподобия, учитывая, что $\Theta = p$:

$L = P_{n_1}(m_1) P_{n_2}(m_2) = C_{n_1}^{m_1} C_{n_2}^{m_2} p^{m_1+m_2} (1-p)^{[(n_1+n_2)-(m_1+m_2)]}$.

Найдем логарифмическую функцию правдоподобия:

$\ln(L) = \ln\left(C_{n_1}^{m_1} C_{n_2}^{m_2}\right) + (m_1+m_2)\ln(p) + [(n_1+n_2)-(m_1+m_2)]\ln(1-p)$.

Найдем первую производную *p:*

$$\frac{d\ln(L)}{d\,p} = \frac{m_1+m_2}{p} - \frac{(n_1+n_2)-(m_1+m_2)}{1-p}.$$

Напишем уравнение правдоподобия, для чего приравняем первую производную нулю:

$$\frac{m_1+m_2}{p}-\frac{(n_1+n_2)-(m_1+m_2)}{1-p}=0.$$

Найдем критическую точку, для чего решим полученное уравнение относительно *p:*

$$p=(m_1+m_2)/(n_1+n_2).$$

Найдем вторую производную по *p:*

$$\frac{d^2\ln(L)}{dp^2}=-\frac{m_1+m_2}{p^2}+\frac{(n_1+n_2)-(m_1+m_2)}{(1-p)^2}.$$

Легко убедится, что при $p=(m_1+m_2)/(n_1+n_2)$ вторая производная отрицательна; следовательно, $p=(m_1+m_2)/(n_1+n_2)$ - точка максимума, и значит, ее надо принять в качестве оценки наибольшего правдоподобия неизвестной вероятности p биноминального распределения:

$$p^*=(m_1+m_2)/(n_1+n_2).$$

2 Найти методом моментов по выборке $x_1, x_2, \dots, x_n$ точечную оценку неизвестного параметра λ показательного распределения, плотность .

Решение. Приравниваем начальный теоретический момент первого порядка начальному эмпирическому моменту первого порядка: $v_1=M_1$. Учитывая, что $v_1=M(X), \mathrm{M}_1=\bar{x}_B$, получим М(Х) $=\bar{x}_B$. Приняв во внимание, что математическое ожидание показательного распределения равно $\frac{1}{\lambda}$, имеем $\frac{1}{\lambda}=\bar{x}_B$. Отсюда $\lambda=\frac{1}{\bar{x}_B}$. Итак, искомая точечная оценка параметра λ показательного распределения равна велечине, обратной выборочной средней $\lambda^*=\frac{1}{\bar{x}_B}$.

Задачи для решения в аудитории

1 Случайная величина Х распределена по закону Пуассона $P_m(x_i)=\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}$, где m- число испытаний, произведенных в одном опыте; x_i – число появлений событий в i-м опыте. Найти методом моментов по выборке $x_1, x_2, \dots, x_n$ точечную оценку неизвестного параметра λ, определяющего распределение Пуассона.

2 Случайная величина Х (число семян сорняков в пробе зерна) распределена по закону Пуассона. Ниже приведено распределение семян сорняков в $n=1000$ пробах зерна (в первой строке указано количество x_i сорняков в одной пробе; во второй строке указана частота n_i – число проб, содержащих x_i семян сорняков):

x_i	*0*	*1*	*2*	*3*	*4*	*5*	*6*
n_i	*405*	*366*	*175*	*40*	*8*	*4*	*2*

Найти методом моментов точечную оценку неизвестного параметра распределения Пуассона.

3 Найти методом моментов по выборке $x_1, x_2, \dots, x_n$ точечную оценку неизвестного параметра λ закона Пуассона показательного распределения, плотность которого $f(x) = \lambda e^{-\lambda x}(x \geq 0)$.

4 Случайная величина X распределена по «двойному» закону Пуассона: $P(X = x_i) = \frac{1}{2} \cdot \frac{\lambda_1^{x_i} \cdot e^{-\lambda_1}}{x_i!} + \frac{1}{2} \cdot \frac{\lambda_2^{x_i} \cdot e^{-\lambda_2}}{x_i!}$. Ниже приведено эмпирическое распределение числа появлений событий в $n = 327$ испытаниях (в первой строке указано число n_iпоявлений события ; во второй строке приведена частота n_i −количество испытаний, в которых появилось x_i событий):

x_i	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
n_i	*28*	*47*	*81*	*67*	*53*	*24*	*13*	*8*	*3*	*2*	*1*

Найти методом моментов точечные оценки неизвестных параметров λ_1 и λ_2 «двойного распределения» Пуассона.

5 Случайная величина X (число появлений события A в m независимых испытаниях) подчинена биномиальному закону распределения с неизвестным параметром p. Ниже приведено эмпирическое распределение числа появлений события A в 1000 испытаний (в первой строке указано число x_i появлений события в одном опыте из m=10 испытаний, во второй строке приведена частота n_i – число опытов, в которых наблюдалось x_i появлений события A):

x_i	0	1	2	3	4	5	6	7
n_i	2	3	10	22	26	20	12	5

Найти методом наибольшего правдоподобия точечную оценку неизвестного параметра p биноминального распределения.

6 Найти методом наибольшего правдоподобия по выборке $x_1, x_2, .., x_n$ точечную оценку неизвестного параметра λ показательного распределения, плотность которого $f(x) = \lambda e^{-\lambda x}$ $(x \geq 0)$.

7 Устройство состоит из элементов, время безотказной работы которых подчинено гамма-распределению. Испытания пяти элементов дали следующие наработки (время работы элемента в часах до отказа): 50, 75, 125, 250, 300. Найти методом наибольшего правдоподобия точечную оценку одного неизвестного параметра β гамма – распределения, если второй параметр этого распределения $\alpha = 1{,}12$.

8 Найти методом наибольшего правдоподобия по выборке $x_1, x_2, .., x_n$ точечную оценку параметра σ (параметр a известен) распределения Кэптейна

$f(x) = \frac{g'(x)}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)}$.

Задачи для самостоятельного решения

1 Случайная величина X (число нестандартных изделий в партии изделий) распределена по закону Пуассона. Ниже приведено распределение нестандартных изделий в $n = 200$ партиях (в первой строке указано количество x_i нестандартных изделий в одной партии; во второй строке указана частота $n_i -$ число партий, содержащих x_i нестандартных изделий):

x_i	0	1	2	3	4
n_i	132	43	20	3	2

Найти методом моментов точечную оценку неизвестного параметра распределения Пуассона.

2 Найти методом моментов по выборке $x_1, x_2, \dots, x_n$ точечные оценки параметров a и b равномерного распределения, плотность которого

$f(x) = 1/(b-a)(b > a)$.

3 Найти методом моментов по выборке $x_1, x_2, \dots, x_n$ точечные оценки параметров λ_1 и λ_2 «двойного распределения» Пуассона.

$P(X = x_i) = \frac{1}{2} \cdot \frac{\lambda_1^{x_i} \cdot e^{-\lambda_1}}{x_i!} + \frac{1}{2} \cdot \frac{\lambda_2^{x_i} \cdot e^{-\lambda_2}}{x_i!}$, где $x_i -$ число появлений событий в n_i испытаниях, λ_1 и λ_2 $-$положительные числа, причем $\lambda_1 > \lambda_2$.

4 Найти методом наибольшего правдоподобия по выборке $x_1, x_2, \dots, x_n$ точечную оценку параметра β гамма – распределения (параметр α известен), плотность которого

$f(x) = \frac{1}{\beta^{\alpha+1}\Gamma(\alpha+1)} x^{\alpha} e^{\frac{-x}{\beta}} (\alpha > -1, \beta > 0, x \geq 0)$.

5 Найти методом наибольшего правдоподобия по выборке $x_1, x_2, .., x_n$ точечную оценку параметра σ (параметр a известен) распределения Кэптейна, плотность которого $f(x) = \frac{g'(x)}{\sigma\sqrt{2\pi}} e^{-[g(x)-a]^2/(2\sigma^2)}$, где g(x) –дифференцируемая функция.

Задачи для решения дома

1 Случайная величина X (число появлений события A в m независимых испытаниях) подчинена биномиальному закону распределения с неизвестным параметром p. Ниже приведено эмпирическое распределение числа появлений события в 10 опытах по пять испытаний в каждом (в первой строке указано число x_i появлений события A в одном опыте; во второй строке указана частота $n_i -$ количество опытов, в которых наблюдалось x_i появлений в A):

x_i	0	1	2	3	4
n_i	5	2	1	1	1

Найти методом моментов точечную оценку параметра p биномиального распределения.

2 Найти методом моментов по выборке $x_1, x_2, \dots, x_n$ точечную оценку параметра p биномиального распределения $P_m(x_i) = C_m^{x_i} p^{x_i}(1-p)^{m-x_i}$, где $x_i -$число появлений события в одном опыте $(i = 1,2, \dots, n)$, $m -$количеств испытаний в одном опыте.

3 Устройство состоит из элементов, время безотказной работы которых подчиненно гамма – распределению. Испытания пяти элементов дали следующие наработки (время работы элемента в часах до отказа): 50, 75, 125, 250, 300. Найти методом моментов точечные оценки неизвестных параметров α и β, которыми определяется гамма – распределение.

4 Случайная величина X (число появлений события A в m независимых испытаниях) подчинена закону распределения Пуассона с неизвестным параметром λ:

$$P_m(X = x_i) = \lambda^{x_i} \cdot e^{-\lambda} / x_i!,$$

где m – число испытаний в одно опыте, $x_i -$ число появленийсобытия в i – м опыте (i=1, 2, … , n).

Найти методом наибольшего правдоподобия по выборке $x_1, x_2, .., x_n$ точечную оценку неизвестного параметра λ распределения Пуассона.

5 Случайная величина X (время безотказной работы элемента) имеет показательное распределение $f(x) = \lambda e^{-\lambda x}$ ($x \geq 0$). Ниже приведено эмпирическое распределение среднего времени работы 1000 элементов (в первой строке указано среднее время x_i безотказной работы одного элемента в часах; во второй строке указана частота n_i – количество элементов, проработавших в среднем x_i часов):

x_i	5	15	25	35	45	55	65
n_i	365	245	150	100	70	45	25

Найти методом наибольшего правдоподобия точечную оценку неизвестного параметра λ показательного распределения.

6 Найти методом наибольшего правдоподобия по выборке $x_1, x_2, .., x_n$ точечную оценку параметра p геометрического распределения:

$$P(X = x_i) = (1-p)^{x_i - 1} \cdot p,$$

где x_i – число испытаний, произведенных до появления события; p – вероятность появления события в одном испытании.

7 Найти методом наибольшего правдоподобия по выборке $x_1, x_2, .., x_n$ точечные оценки параметров a и σ нормального распределения, плотность которого $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)}$.

Тема 4. Методы расчета свободных характеристик выборки

Вопросы для повторения

1 Равностоящие варианты. Условные варианты.

2 Обычные, начальные и центральные эмпирические моменты.

3 Связь между центральными и обычными моментами.

4 Условные эмпирические моменты. Отыскание центральных моментов по условным.

5 Метод произведений вычисления выборочных средней и дисперсии.

6 Сведение первоначальных вариант к равностоящим.

7 Эмпирические и выравнивающие частоты. Построение нормальной кривой по опытным данным.

8 Оценка отклонения эмпирического распределения от нормального. Асимметрия и эксцесс.

Образцы решения задач

Найти методом произведений асимметрию и эксцесс по заданному распределению выборки объема $n = 100$ (таблица 29).

Таблица 29 – Распределению выборки объема $n = 100$

x_i	*11*	*13*	*15*	*17*	*19*	*21*
n_i	*5*	*15*	*50*	*16*	*10*	*4*

Решение

Асимметрия эмпирического распределения определяется равенством:

$$a_s = \frac{m_3}{\sigma_B^2},$$

где m_2 -центральный эмпирический момент третьего порядка;

σ_B- выборочное среднее квадратическое отклонение.

$$m_3 = \frac{\sum_{i=1}^{k} n_i \cdot (x_i - \bar{x}_B)^3}{n}.$$

Эксцесс эмпирического распределения определяется равенством:

$e_k = \frac{m_4}{\sigma_B^4} - 3,$

где m_4-центральный эмпирический момент четвертого порядка.

$$m_4 = \frac{\sum_{i=1}^{k} n_i \cdot (x_i - \bar{x}_B)^4}{n}.$$

В случае равноотстоящих вариант с шагом h удобно вычислить m_3 и m_4 по формулам:

$$m_3 = (M_3^* - 3M_1^* M_2^* + 2(M_1^*)^3)h^3;$$

$m_4 = (M_4^* - 4M_1^* M_3^* + 6(M_1^*)^2 M_2^* - 3(M_1^*)^4)h^4,$

где, $M_k^* = \frac{\sum_{i=1}^{k} n_i u_i^k}{n}$; $u_i = \frac{x_i - c}{h}$.

Воспользуемся методом произведений. Составим расчетную таблицу (с =15).

Таблица 30 – Расчетная таблица (с =15)

x_i	n_i	u_i	$n_i u_i$	$n_i u_i^2$	$n_i u_i^3$	$n_i u_i^4$	$n_i (u_i + 1)^4$
11	5	-2	-10	20	-40	80	5
13	15	-1	-15	15	-15	15	0
15	50	0	0	0	0	0	50
17	16	1	16	16	16	16	256
19	10	2	20	40	80	160	810
21	4	3	12	36	108	324	1045
	100		23	127	149	595	2145

Контроль

$\sum_{i=1}^{k} n_i (u_i + 1)^4 = \sum_{i=1}^{k} n_i u_i^4 + 4\sum_{i=1}^{k} n_i u_i^3 + 6\sum_{i=1}^{k} n_i u_i^2 + 4\sum_{i=1}^{k} n_i u_i + n$;

$\sum_{i=1}^{k} n_i u_i^4 + 4\sum_{i=1}^{k} n_i u_i^3 + 6\sum_{i=1}^{k} n_i u_i^2 + 4\sum_{i=1}^{k} n_i u_i + n = 595 + 4 \cdot 149 + 6 \cdot 127 + 4 \cdot 23 + 100 = 2145;$

$M_1^* = \frac{23}{100} = 0{,}23$; $M_2^* = \frac{127}{100} = 1{,}27$; $M_3^* = \frac{149}{100} = 1{,}49$; $M_4^* = \frac{595}{100} = 5{,}95$;

где $M_k^* = \frac{\sum_{i=1}^{k} n_i u_i^k}{n}$.

$$m_3 = (1{,}49 - 3 \cdot 1{,}27 + 2 \cdot (0{,}23)^3) \cdot 8 = (1{,}49 - 0{,}8763 + 0{,}024334) \cdot 8 = 5{,}104272;$$

$m_4 = (5{,}95 \ -4 \cdot 0{,}23 \cdot 1{,}27 + 2 \cdot (0{,}23)^2 \cdot 1{,}27 - 3 \cdot (0{,}23)^4) \cdot 16 = (5{,}95 - 1{,}3708 + 0{,}403098 - 0{,}00839523) \cdot 16 = 79{,}58244432;$

$\sigma_B = \sqrt{D_B}$, $\quad D_B = (M_2^* - (M_1^*)^2)h^2;$

$$D_B = (1{,}27 - (0{,}23)^2) \cdot 4 = (1{,}27 - 0{,}0529) \cdot 4 = 4{,}8684\,; \quad \sigma_B \approx 2{,}21;$$

$$a_{3=} \frac{5{,}104272}{4{,}8684} \approx 1{,}05;$$

$$e_k = \frac{m_4}{\sigma_B^4} - 3\,; e_k = \frac{79,}{\sigma_B^4} - 3 \ \approx \frac{79{,}6}{23{,}7} - 3 = 0{,}36.$$

Ответ: $a_3 = 1{,}05; e_k = 0{,}36$.

Задачи для решения в аудитории

1 Найти методом произведений выборочную среднюю и выборочную дисперсию по заданному распределению выборки объема *n = 100.*

Таблица 31 — Распределение выборки

x_i	12	14	16	18	20	22
n_i	5	15	50	16	10	4

2 Найти методом произведений выборочную среднюю и выборочную дисперсию по заданному распределению выборки объема *n* = 100.

Таблица 32 — Распределение выборки объема *n = 100*

x_i	2	3	7	9	11	12,5	16	18	23	26
n_i	3	5	10	6	10	4	12	13	8	29

Указание: при вычислении выборочной дисперсии для уменьшения ошибки, вызванной группировкой (особенно при малом числе интервалов), делают поправку Шеппарда

$D'_B = D_B - \frac{1}{12} \cdot h^2.$

3 Найти методом сумм $\bar{x}_B$ и D_B по заданному распределению выборки объема *n = 100*.

Таблица 33 — Распределение выборки объема

x_i	48	52	56	60	64	68	72	76	80	84
n_i	2	4	6	8	12	30	18	8	7	5

4 Найти методом произведений асимметрию и эксцесс по заданному распределению выборки объема *n = 100.*

Таблица 34 — Распределение выборки объема *n = 100*

x_i	12	14	16	18	20	22
n_i	5	15	50	16	10	4

5 Найти методом сумм асимметрию и эксцесс по заданному распределению выборки объема *n = 100*.

Таблица 35 — Распределение выборки объема *n = 100*

x_i	48	52	56	60	64	68	72	76	80	84
n_i	2	4	6	8	12	30	18	8	7	5

6 Построить нормальную кривую по данному распределению

Таблица 36 — Статистический ряд

x_i	15	20	25	30	35	40	45	50	55
n_i	6	13	38	74	106	85	30	10	4

Материал для решения задачи №3

Пусть выборка задана в виде распределения равноотстоящих вариант и соответствующих им частот. В этом случае выборочную среднюю и дисперсию можно вычислить по формулам:

$$\bar{x}_B = M_1^* \cdot h + c,\ D_B = (M_2^* - (M_1^*)^2) \cdot h^2.$$

При использовании метода сумм условные моменты первого и второго порядка находят по формулам:

$M_1^* = \frac{d_1}{n}$, $M_2^* = \frac{S_1 + 2 \cdot S_2}{n}$,где $d_1 = a_1 - b_1, S_1 = a_1 + b_1, S_2 = a_2 + b_2$.

Как практически вычислить эти числа, будет указано на занятии.

Задачи для решения дома

1 Найти методом произведений $\bar{x}_B$ и D_B по заданному распределению выборки.

Таблица 37 — Статистический ряд

x_i	18,6	19	19,4	19,8	20,2	20,6
n_i	4	6	30	40	18	2

2 В ОТК были измерены диаметры 200 валиков из партии, изготовленной одним станком-автоматом. Отклонения измеренных диаметров от номинала даны в таблице 38 (в микронах).

Таблица 38 — Отклонения диаметров от номинала

Границы уклонений	-20 – -15	-15 – -10	-10 – -5	-5 – 0	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30
Середины интервалов	-17,5	-12,5	-7,5	-2,5	2,5	7,5	12,5	17,5	22,5	27,5
Число валиков	7	11	15	24	49	41	26	17	7	3

Вычислить методом произведений $\bar{x}_B$ и D_B.

3 Двести однотипных деталей были подвергнуты контрольному измерению. Результаты измерений приведены в таблице, где в первой строке указаны середины интервалов (шириной 0,1 мм), на которые разбит весь размах выборки, а во второй – количество результатов измерений, попавших в данный интервал.

Таблица 39 — Результаты измерений однотипных деталей

x_i	3,7	3,8	3,9	4	4,1	4,2	4,3	4,4
x_i	1	22	40	79	27	26	4	1

Найти асимметрию и эксцесс.

Тема 5. Статистические оценки параметров распределения

Интервальные оценки

Вопросы для повторения

1 Понятие статистической оценки параметров распределения.

2 Понятие точечной оценки, интервальной оценки.

3 Доверительная вероятность (надежность) оценки.

4 Доверительный интервал.

5 Оценка математического ожидания α нормально распределённого признака при известном σ.

6 Оценка математического ожидания α нормально распределенного признака при известном σ.

7 Оценка истинного значения измеряемой величины.

8 Доверительный интервал для оценки среднего квадратического отклонения σ нормального распределения.

9 Оценки точности измерений.

10 Оценка вероятности по относительной частоте.

Теоретическая справка

Интервальной называют оценку, которая определяется двумя числами – концами интервала. Интервальные оценки позволяют установить точность и надежность оценок.

Пусть найденная по данным выборки статистическая характеристика Θ^* служит оценкой неизвестного параметра Θ. Будем считать Θ постоянным числом (Θ может быть и случайной величиной). Ясно, что Θ^* тем точнее определяет параметр Θ, чем меньше абсолютная величина разности $\left|\Theta - \Theta^*\right|$. Другими словами, если $\delta > 0$ и $\left|\Theta - \Theta^*\right| < \delta$, то чем меньше δ, тем оценка точнее.

Таким образом, положительное число δ характеризует точность оценки.

Однако статистические методы не позволяют категорически утверждать, что оценка Θ^* удовлетворяет неравенству $\left|\Theta - \Theta^*\right| < \delta$, можно лишь говорить о вероятности γ, с которой это неравенство осуществляется.

Надежностью (доверительной вероятностью) оценки называют вероятность γ, с которой осуществляется неравенство $\left|\Theta - \Theta^*\right| < \delta$.

Обычно надежность оценки задается наперед, причем в качестве γ берут число, близкое к единице. Наиболее часто задают надежность, равную 0,95; 0,99 и 0,999.

Пусть вероятность того, что $\left|\Theta - \Theta^*\right| < \delta$, равна γ: $P\left(\left|\Theta - \Theta^*\right| < \delta\right) = \gamma$.

Заменив неравенство равносильным ему двойным неравенством, получим:

$P\left[\Theta^* - \delta < \Theta < \Theta^* + \delta\right] = \gamma$.

Это соотношение следует понимать так: вероятность того, что интервал $\Theta^* - \delta < \Theta < \Theta^* + \delta$ заключает в себе (покрывает) неизвестный параметр Θ, равна γ. Интервал $\left(\Theta^* - \delta, \Theta^* + \delta\right)$ называется доверительным интервалом, который покрывает неизвестный параметр с надежностью γ.

Доверительный интервал для оценки математического ожидания при известном σ.

Пусть количественный признак генеральной совокупности распределен нормально. Известно среднее квадратическое отклонение этого распределения – σ. Требуется оценить математическое ожидание *a* по выборочной средней. Найдем доверительный интервал, покрывающий *a* с надежностью γ. Выборочную среднюю $\bar{x}$ будем рассматривать как случайную величину $\bar{X}$ ($\bar{x}$ изменяется от выборки к выборке), выборочные значения признака $x_1, x_2, ..., x_n$ – как одинаково распределенные независимые СВ с математическим ожиданием каждой *a* и средним квадратическим отклонением γ. Примем без доказательства, что если величина *X* распределена нормально, то и выборочная средняя тоже распределена нормально с параметрами $M(\bar{X}) = a,\ \sigma(\bar{X}) = \dfrac{\sigma}{\sqrt{n}}$.

Потребуем, чтобы выполнялось равенство $P\left(\left|\bar{X} - a\right| < \delta\right) = \gamma$. Пользуясь формулой $P\left(\left|X - a\right| < \delta\right) = 2\Phi(\delta/\sigma)$, заменив X на $\bar{X}$ и σ на $\sigma(\bar{X}) = \sigma/\sqrt{n}$, получим $P\left(\left|\bar{X} - a\right| < t\delta/\sqrt{n}\right) = 2\Phi(\delta\sqrt{n}/\sigma) = 2\Phi(t)$, где $t = \delta\sqrt{n}/\sigma$.

Найдя из предыдущего равенства $\delta = t\sigma/\sqrt{n}$, получим окончательную формулу: $P\left(\bar{x} - \dfrac{t\sigma}{\sqrt{n}} < a < \bar{x} + \dfrac{t\sigma}{\sqrt{n}}\right) = 2\Phi(t) = \gamma.$

Число *t* определяется из равенства $\Phi(t) = \dfrac{\gamma}{2}$ по таблице функции Лапласа.

Доверительный интервал для оценки математического ожидания при неизвестном σ

По данным выборки можно построить случайную величину (ее возможные значения будем обозначать через *t)*:

$T = \dfrac{\bar{X} - a}{S/\sqrt{n}}$, которая имеет распределение Стьюдента.

Плотность распределения Стьюдента $S(t,n) = B_n\left[1 + \dfrac{t^2}{n-1}\right]^{-n/2}$, где $B_n = \dfrac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\Gamma((n-1)/2)}$.

Поскольку $S(t,n)$ – четная функция, можно записать двойное неравенство:

$$P\left(\bar{X} - t_\gamma S/\sqrt{n} < a < \bar{X} - t_\gamma S/\sqrt{n}\right) = \gamma.$$

Доверительный интервал $\left(\bar{x} - t_\gamma s/\sqrt{n},\ \bar{x} - t_\gamma s/\sqrt{n}\right) = \gamma.$ $\bar{x}, s$ вычисляются по выборке; t_γ – табличная величина.

Доверительный интервал для оценки среднего квадратического отклонения σ нормального распределения

Требуется оценить неизвестное генеральное среднее квадратическое отклонение σ по исправленной дисперсии *s*, т.е. найти доверительные интервалы, покрывающие параметр σ с заданной надежностью γ.

Потребуем выполнения соотношения $P\left(s-\delta<\sigma<s+\delta\right)=\gamma.$

Преобразуем двойное неравенство $s-\delta<\sigma<s+\delta$ в равносильное неравенство $s(1-\delta/s)<\sigma<s(1+\delta/s).$ Положив $\dfrac{\delta}{s}=q$, получим доверительный интервал: $s(1-q)<\sigma<s(1+q).$

Для отыскания q пользуются таблицей значений $q=q(\gamma,n).$

Интервальная оценка вероятности (биноминального распределения) по относительной частоте

Сначала вычислим точечную оценку.

В качестве точечной оценки неизвестной вероятности p появления события A принимают относительную частоту $W=\dfrac{m}{n},$ где m – число появлений события A, n – число испытаний. Найдем дисперсию оценки, приняв во внимание, что $D(m)=npq: D(W)=D\left[m/n\right]=D(m)/n^2=npq/n^2=pq/n.$

q – вероятность не появления события А. Среднее квадратическое отклонение $\sigma_W=\sqrt{pq/n}$. Интервальная оценка. Для биноминального распределения формулу $P\left(\left|X-a\right|<\delta\right)=2\Phi(\delta/\sigma)$ можно записать в виде: $P\left(\left|W-p\right|<\delta\right)=2\Phi(\delta/\sigma_W)$

Потребуем выполнения этого соотношения с надежностью γ.

$P\left(\left|W-p\right|<\delta\right)=2\Phi(\delta/\sigma_W)=\gamma.$ Подставляя $\sigma_W=\sqrt{pq/n}$, получим

$P\left(\left|W-p\right|<\delta\right)=2\Phi(\delta\sqrt{n}/\sqrt{pq})=2\Phi(t)=\gamma,$ где $t=\sqrt{n}/\sqrt{pq}.$ Отсюда $\delta=t\sqrt{pq/n}$ и, следовательно, $P\left(\left|W-p\right|<t\sqrt{pq/n}\right)=2\Phi(t)=\gamma.$

Случайную величину W заменим неслучайной наблюдаемой относительной частотой w и подставим $1-p$ вместо q. $|w-p|<t\sqrt{p(1-p)/n}$.

Решим это неравенство относительно p . Возведем обе части неравенства в квадрат. Получим

$\left[\left(t^2/n\right)+1\right]p^2-2\left[w+\left(t^2/n\right)\right]p+w^2<0.$ Найдем корни трехчлена. Меньший корень

$$p_1=\frac{n}{t^2+n}\left[w+\frac{t^2}{2n}-t\sqrt{\frac{w(1-w)}{n}+\left(\frac{t}{2n}\right)^2}\right].$$

Больший корень $p_1=\dfrac{n}{t^2+n}\left[w+\dfrac{t^2}{2n}+t\sqrt{\dfrac{w(1-w)}{n}+\left(\dfrac{t}{2n}\right)^2}\right].$

Искомый доверительный интервал $p_1<p<p_2.$

Образцы решения задач

1 Изучается вопрос об урожайности овощей в хозяйствах области. Результаты приведены в таблице 40.

Таблица 40 – Результаты урожайности овощей в хозяйствах области

Урожайность, ц/га	25	75	125	175	225	275	325
Количество хозяйств	11	23	26	17	20	2	1

Найти граничные значения урожайности по области с надежностью 0, 97.

Решение

Граничные значения средней урожайности по области являются границами доверительного интервала $\left(\bar{x}_{\text{В}} \cdot \frac{t \cdot \sigma}{\sqrt{n}}, \bar{x}_B + \frac{t \cdot \sigma}{\sqrt{n}}\right)$.

$$\bar{x}_B = \frac{25 \cdot 11 + 75 \cdot 23 + 125 \cdot 26 + 175 \cdot 17 + 225 \cdot 20 + 275 \cdot 2 + 325}{100}$$
$$= \frac{275 + 1725 + 3250 + 2975 + 4500 + 550 + 325}{100} = 136$$

$\sigma = \sqrt{\bar{x}^2 - (\bar{x})^2}$;

$$\overline{x^2} = \frac{25^2 \cdot 11 + 75^2 \cdot 23 + 125^2 \cdot 26 + 175^2 \cdot 17 + 225^2 \cdot 20 + 275^2 \cdot 2 + 325^2}{100}$$
$$== \frac{6875 + 129375 + 406250 + 520625 + 1012500 + 105625}{100} =$$
$$= 21812{,}5;$$

$\sigma = \sqrt{21812{,}5 - 136^2} \approx 57{,}6.$

Найдем t как аргумент $\Phi(t) = \frac{0{,}97}{2}, \Phi(t) = 0{,}485,\ t = 2,17$(приложение А)

$$\delta = \frac{t \cdot \sigma}{\sqrt{n}}, \delta = \frac{2{,}17 \cdot 27{,}6}{\sqrt{100}} = \frac{2{,}17 \cdot 27{,}6}{10} = 12{,}4992, \qquad \delta = 12{,}5.$$

Границы доверительного интервала

$\bar{x}_{\text{В}} - \delta = 136 - 12{,}5 = 123{,}5;\quad \bar{x}_B + \delta = 136 + 12{,}5 = 148{,}5.$

Округляя значения, получим $a = 136 \pm 13$, т.е. (123,149) – доверительный интервал, в котором находится средняя урожайность овощей по области с вероятностью 97%.

2 Из генеральной совокупности извлечена выборка объема 10.

Таблица 41 – Выборка объема n= 10.

x_i	-1	1	2	3	4	5
n_i	1	2	2	2	1	2

Оценить с надежностью 0,95 математическое ожидание a нормально распределенного признака генеральной совокупности по выборочной средней при помощи доверительного интервала.

Решение

Искомый доверительный интервал имеет вид

$\left(\bar{x}_{\text{В}-t_1} \cdot \frac{s}{\sqrt{n}} < a < \bar{x}_{B+t_1} \cdot \frac{s}{\sqrt{n}}\right).$

Найдем t_1. Пользуясь таблицей (приложение А), учитывая $\gamma = 0{,}95, n = 10,$ находим $t_1 = 2{,}26.$

$\delta = \frac{t_1 \cdot s}{\sqrt{n}},\ \delta = \frac{2{,}26\cdot 2}{\sqrt{10}} = \frac{4{,}52}{\sqrt{10}} = 1{,}4125.$

Найдем выборочную среднюю и исправленное среднее квадратическое отклонение по формулам: $\bar{x}_\text{В}=\frac{\sum_{i=1}^{k} n_i\, x_i}{n},\ s = \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n-1}};$

$$\bar{x}_\text{В} = \frac{-1 \cdot 1 + 1 \cdot 2 + 2 \cdot 2 + 3 \cdot 2 + 4 \cdot 1 + 5 \cdot 2}{10} = 2{,}5;$$

$s = \sqrt{\frac{12{,}25+4{,}5+0{,}5+0{,}5+2{,}25+12{,}5}{9}} = \sqrt{\frac{32{,}5}{9}} \approx 2.$

$\bar{x}_{B-\delta} = 2{,}5 - 1{,}4125 = 1{,}0875;\ \ \bar{x}_{B-\delta} = 2{,}5 + 1{,}4125 = 3{,}9125.$

Округляя результаты, получим доверительный интервал 1,1$< a < 3{,}9$, покрывающий неизвестное математическое ожидание с надежностью 0,95.

3 По данным выборки объема $n = 15$ из генеральной совокупности найдено исправленное среднее квадратическое отклонение $s = 1$ нормально распределенного количественного признака. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью *0,999*.

Решение

Задача сводится к отысканию доверительного интервала

$$s\,(1-q) < \sigma < s\,(1+q), \text{если } q < 1 \text{ или } 0 < \sigma < s(1+q), \text{если } q > 1.$$

По данным $\gamma = 0{,}999$ и $n = 15,$ по таблице (приложение А) найдем $q = 1{,}15.$

Так как $q < 1$, то подставив $s = 1$, $q = 1{,}15$ в соотношение $0 < \sigma < s(1 + q)$, получим искомый доверительный интервал $0 < \sigma < 2{,}15$.

Задачи для решения в аудитории

1 Найти доверительный интервал для оценки с надежностью 0,95 неизвестного математического ожидания α нормально распределенного признака x генеральной совокупности, если даны генеральное среднее квадратическое отклонение $\sigma = 5$, выборочная средняя $\bar{x}_B = 14$ и объем выборки n =25.

2 Станок-автомат штампует валики. По выборке объема $n = 100$ вычислена выборочная средняя диаметров изготовленных валиков. Найти с надежностью 0,95 точность σ, с которой выборочная средняя оценивает математическое ожидание диаметров изготовляемых валиков, зная, что их среднее квадратическое отклонение $\sigma = 2$ мм.

3 Найти минимальный объем выборки, при котором с надежностью 0,975 точность оценки математического ожидания α генеральной совокупности по выборочной средней будет равна $\sigma = 0{,}3$, если известно среднее квадратическое отклонение $\sigma = 1{,}2$ нормально распределенной генеральной совокупности.

4 Из генеральной совокупности извлечена выборка объема $n = 10$.

Таблица 42 — Выборка объема $n = 10$

Варианта x_i	-2	1	2	3	4	5
Частота n_i	2	1	2	2	2	1

Оценить с надежностью 0,95 математическое ожидание α нормально распределенного признака генеральной совокупности по выборочной средней при помощи доверительного интервала.

5 По данным 9 независимых равноточных измерений некоторой физической величины найдены $\bar{x}_B = 30{,}1$ и S = 6. Оценить истинное значение измеряемой величины с помощью доверительного интервала с надежностью $\gamma = 0{,}99$.

6 По данным выборки объема n из генеральной совокупности нормально определенного количественного признака найдено исправленное среднее квадратическое отклонение s. Найти доверительный интервал, покрывающий генеральное среднее квадратическое отклонение σ с надежностью 0,999, если:

a) n = 10, s = 5,1;

b) n = 5, s = 14.

7 Произведено 12 измерений одним приором (без систематической ошибки) некоторой физической величины, причем исправленное среднее квадратическое отклонение s случайных ошибок измерений оказалось равным 0,06. Найти точность прибора с надежностью 0,99.

8 При 100-кратном повторении опыта событие А наступило 62 раза с надежностью 0,95. Оценить неизвестную вероятность события А.

9 Выборочным путем установлено, что в партии из 200 деталей 160 стандартных деталей. Найти относительную частоту появления стандартной детали и надежность утверждения, что полученная частота является оценкой вероятности с относительной погрешностью не более 6%.

Задачи для решения дома

1 36 районов области получили картофелеуборочные комбайны в следующем количестве: 2, 3, 4, 1, 5, 4, 6, 4, 7, 4, 5, 3, 2, 4, 5, 7, 6, 2, 4, 3, 4, 6, 4, 3, 4, 4, 6, 4, 4, 6, 3, 4, 4, 5, 6, 4. Найти среднее число комбайнов, отправленных в район, доверительный интервал для среднего значения с надежностью 0,95.

2 Дано: s = 1,5; $\bar{x}_B = 16{,}8, n = 12, \gamma = 0{,}95$. Найти доверительный интервал для оценки неизвестного математического ожидания с заданной надежностью.

3 К началу сенокоса в районе было отремонтировано 76% косилок и 70% тракторных грабель. Каков должен быть объем выборки косилок и объем выборки грабель для определения процента их исправности, чтобы с вероятностью 0,98 погрешность определения не превышала бы 10%?

4 Произведено 10 измерений одним прибором (без систематической ошибки) некоторой физической величины, причем исправленное среднее квадратическое отклонение случайных ошибок измерений оказалось равным 0,8. Найти точность прибора с надежностью 0,95.

Приложение А
Статистические таблицы

Таблица А.1 - Значения функции $\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Таблица А.2 - Значения функции $\Phi(x)=\frac{1}{\sqrt{2\pi}}\cdot\int_0^x e^{-\frac{z^2}{2}}dz$

	Φ(*x*)	*x*	Φ(*x*)	*x*	Φ(*x*)	*x*	Φ(*x*)
0,00							
0,01	0,0000	0,32	0,1255	0,64	0,2389	0,96	0,3315
0,02	0,0040	0,33	0,1293	0,65	0,2422	0,97	0,3340
0,03	0,0080	0,34	0,1331	0,66	0,2454	0,98	0,3365
0,04	0,0120	0,35	0,1368	0,67	0,2486	0,99	0,3389
0,05	0,0160	0,36	0,1406	0,68	0,2517	1,00	0,3413
0,06	0,0199	0,37	0,1443	0,69	0,2549	1,01	0,3438
0,07	0,0239	0,38	0,1480	0,70	0,2580	1,02	0,3461
0,08	0,0279	0,39	0,1517	0,71	0,2611	1,03	0,3485
0,09	0,0319	0,40	0,1554	0,72	0,2642	1,04	0,3508
0,10	0,0359	0,41	0,1591	0,73	0,2673	1,05	0,3531
0,11	0,0398	0,42	0,1628	0,74	0,2703	1,06	0,3554
0,12	0,0438	0,43	0,1664	0,75	0,2734	1,07	0,3577
0,13	0,0478	0,44	0,1700	0,76	0,2764	1,08	0,3599
0,14	0,0517	0,45	0,1736	0,77	0,2794	1,09	0,3621
0,15	0,0557	0,46	0,1772	0,78	0,2823	1,10	0,3643
0,16	0,0596	0,47	0,1808	0,79	0,2852	1,11	0,3665
0,17	0,0636	0,48	0,1844	0,80	0,2881	1,12	0,3686
0,18	0,0675	0,49	0,1879	0,81	0,2910	1,13	0,3708
0,19	0,0714	0,50	0,1915	0,82	0,2939	1,14	0,3729
0,20	0,0753	0,51	0,1950	0,83	0,2967	1,15	0,3749
0,21	0,0793	0,52	0,1985	0,84	0,2995	1,16	0,3770
0,22	0,0832	0,53	0,2019	0,85	0,3023	1,17	0,3790
0,23	0,0871	0,54	0,2054	0,86	0,3051	1,18	0,3810
0,24	0,0910	0,55	0,2088	0,87	0,3078	1,19	0,3830
0,25	0,0948	0,56	0,2123	0,88	0,3106	1,20	0,3849
0,26	0,0987	0,57	0,2157	0,89	0,3133	1,21	0,3869
0,27	0,1026	0,58	0,2190	0,90	0,3159	1,22	0,3883
0,28	0,1064	0,59	0,2224	0,91	0,3186	1,23	0,3907
0,29	0,1103	0,60	0,2257	0,92	0,3212	1,24	0,3925
0,30	0,1141	0,61	0,2291	0,93	0,3238	1,25	0,3944
0,31	0,1179	0,62	0,2324	0,94	0,3264		
	0,1217	0,63	0,2357	0,95	0,3289		

Продолжение таблицы А.2

1,26	0,3962	1,59	*0,4441*	1,92	0,4726	*2,50*	0,4938
1,27	0,3980	1,60	*0,4452*	1,93	0,4732	*2,52*	0,4941
1,28	0,3997	1,61	*0,4463*	1,94	0,4738	*2,54*	0,4945
1,29	0,4015	1,62	*0,4474*	1,95	0,4744	*2,56*	0,4948
1,30	0,4032	1,63	*0,4484*	1,96	0,4750	*2,58*	0,4951
1,31	0,4049	1,64	*0,4495*	1,97	0,4756	*2,60*	0,4953
1,32	0,4066	1,65	*0,4505*	1,98	0,4761	*2,62*	0,4956
1,33	0,4082	1,66	*0,4515*	1,99	0,4767	*2,64*	0,4959
1,34	0,4099	1,67	*0,4525*	2,00	0,4772	*2,66*	0,4961
1,35	0,4115	1,68	*0,4535*	2,02	0,4783	*2,68*	0,4963
1,36	0,4131	1,69	*0,4545*	2,04	0,47,93	*2,70*	0,4965
1,37	0,4147	1,70	*0,4554*	2,06	0,4803	*2,72*	0,4967
1,38	0,4162	1,71	*0,4564*	2,08	0,4812	*2,74*	0,4969
1,39	0,4177	1,72	*0,4573*	2,10	0,4821	*2,76*	0,4971
1,40	0,4192	1,73	*0,4582*	2,12	0,4830	*2,78*	0,4973
1,41	0,4207	1,74	*0,4591*	2,14	0,4838	*2,80*	0,4974
1,42	0,4222	1,75	*0,4599*	2,16	0,4846	*2,82*	0,4976
1,43	0,4236	1,76	*0,4608*	2,18	0,4854	*2,84*	0,4977
1,44	0,4251	1,77	*0,4616*	2,20	0,4861	*2,86*	0,4979
1,45	0,4265	1,78	*0,4625*	2,22	0,4868	*2,88*	0,4980
1,46	0,4279	1,79	*0,4633*	2,24	0,4875	*2,90*	0,4981
1,47	0,4292	1,80	*0,4641*	2,26	0,4881	*2,92*	0,4982
1,48	0,4306	1,81	*0,4649*	2,28	0,4887	*2,94*	0,4984
1,49	0,4319	1,82	*0,4656*	2,30	0,4893	*2,96*	0,4985
1,50	0,4332	1,83	*0,4664*	2,32	0,4898	*2,98*	0,4986
1,51	0,4345	1,84	*0,4671*	2,34	0,4904	*3,00*	0,49865
1,52	0,4357	1,85	*0,4678*	2,36	0,4909	*3,20*	0,49931
1,53	0,4370	1,86	*0,4686*	2,38	0,4913	*3,40*	0,49966
1,54	0,4382	1,87	*0,4693*	2,40	0,4918	*3,60*	0,499841
1,55	0,4394	1,88	*0,4699*	2,42	0,4922	*3,80*	0,499928
1,56	0,4406	1,89	*0,4706*	2,44	0,4927	*4,00*	0,499968
1,57	0,4418	1,90	*0,4713*	2,46	0,4931	*4,50*	0,499997
1,58	0,4429	1,91	*0,4719*	2,48	0,4934	*5,00*	0,499997

Таблица А.3 – Значения $t_\gamma = t(\gamma, n)$

n \ γ	0,95	0,99	0,999
5	2,78	4,60	8,61
6	2,57	4,03	6,86
7	2,45	3,71	5,96
8	2,37	3,50	5,41
9	2,31	2,36	5,04
10	2,26	3,25	4,78
11	2,23	3,17	4,59
12	2,20	3,11	4,44
13	2,18	3,06	4,32
14	2,16	3,01	4,22
15	2,15	2,98	4,14
16	2,13	2,95	4,07
17	2,12	2,92	4,02
18	2,11	2,90	3,97
19	2,10	2,88	3,92

n \ γ	0,95	0,99	0,999
20	2,093	2,861	3,883
25	2,064	2,797	3,745
30	2,045	2,756	3,659
35	2,032	2,720	3,600
40	2,023	2,708	3,558
45	2,016	2,692	3,527
50	2,009	2,679	3,502
60	2,001	2,662	3,464
70	1,996	2,649	3,439
80	1,001	2,640	3,418
90	1,987	2,633	3,403
100	1,984	2,627	3,392
120	1,980	2,617	3,374
∞	1,960	2,576	3,291

Таблица А.4 - Значений $q = q(\gamma, n)$

n \ γ	*0,95*	*0,99*	*0,999*
5	1,37	2,67	5,64
6	1,09	2,01	3,88
7	0,92	1,62	2,98
8	0,80	1,38	2,42
9	0,71	1,20	2,06
10	0,65	1,08	1,80
11	0,59	0,98	1,60
12	0,55	0,90	1,45
13	0,52	0,83	1,33
14	0,48	0,78	1,23
15	0,46	0,73	1,15
16	0,44	0,70	1,07
17	0,42	0,66	1,01
18	0,40	0,63	0,96
19	0,39	0,60	0,92

n \ γ	*0,95*	*0,99*	*0,999*
20	0,37	0,58	0,88
25	0,32	0,49	0,73
30	0,28	0,43	0,63
35	0,26	0,38	0,56
40	0,24	0,35	0,50
45	0,22	0,32	0,46
50	0,21	0,30	0,43
60	0,188	0,269	0,38
70	0,174	0,245	0,34
80	0,161	0,226	0,31
90	0,151	0,211	0,29
100	0,143	0,198	0,27
150	0,115	0,160	0,211
200	0,099	0,136	0,185
250	0,089	0,120	0,162

Таблица А.5 - Критические точки распределения χ^2

Число степеней свободы k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,89
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Список литературы

Учебные пособия

1 Гнеденко, Б.В. Курс теории вероятностей [Текст] / Б. В Гнеденко.- 8-е изд., испр. и доп.— М. : Едиториал УРСС, 2005. – 406 с.

2 Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е Гмурман. - 9-е изд., стер. — М . : Высшая школа, 2003. — 479 с.

3 Вентцель, Е.С. Теория вероятностей [Текст] / Е.С Вентцель. - 4-е изд., стереотип. - М. : Наука, Физматгиз, 1969. - 576 с.

4 Крамер, Г. Математические методы статистики [Текст] / Г. Крамер. - М. : Наука, 1975.

5 Кремер, Н. М. Теория вероятностей и математическая статистика [Текст] / Н. М. Кремер. – М. : Наука, 2000.

6 Севастьянов, Б.А. Курс теории вероятностей и математической статистики [Текст] / Б. Севастьянов. - М. : Наука. Гл. ред. физ.-мат. лит., 1982. — 256 с.

7 Павловский, З. Введение в математическую статистику [Текст] / З. Павловский – 5-е изд., стер. - М. : 2011.— 220 с.

8 Уилкс, С. Математическая статистика [Текст] / С. Уилкс – М. : Наука. Главная редакция физико-математической литературы, 1967. 632 с.

Задачники

9 Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике [Текст] / В.Е Гмурман. – 9-е изд., стер .— М. : Высшая школа, 2004. — 404 с.

10 Емельянов, Т.В Задачник по теории вероятностей и математической статистике [Текст] / сост. Т.В Емельянов., В.П Скитович. – Л. : Издательство ленинградского университета, 1967. – 329 с.

11 Лозинский, С.И. Сборник задач по теории вероятностей и математической статистике [Текст] / С.И Лозинский. – М. : Статистика, 1967. – 126 с.

Лукерьянова Елена Александровна

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

(Часть 1)

Методические указания для практических занятий и самостоятельной работы для студентов направления 01.03.01 «Математика»

Редактор Н.Л. Борисова

Подписано в печать	Формат 60 х 84 1/16	Бумага 65 г/м^2
Печать трафаретная	Усл. печ. л 3,0	Уч.-изд. л. 3,0
Заказ	Тираж 25	Не для продажи

РИЦ Курганского государственного университета.
640000, г. Курган, ул. Советская, 63/4.
Курганский государственный университет.