Министерство образования Республики Беларусь

Учреждение образования
«Могилевский государственный университет продовольствия»

Кафедра высшей математики

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания и задания к проведению практических занятий
для студентов технологических специальностей

Могилев 2011

УДК 519.22
ББК 22.172

Рассмотрено и рекомендовано к изданию
на заседании кафедры высшей математики
Протокол № 2 от 19 октября 2010 г.

Составитель
О.А. Шендрикова

Рецензент
кандидат физико-математических наук, доцент
С.В. Подолян

УДК 519.22
ББК 22.172

Содержание

Занятие № 1. ***Генеральная совокупность и выборка. Вариационный ряд, полигон, гистограмма. Эмпирическая функция распределения***

Математическая статистика – раздел математики, в котором изучаются методы сбора, систематизации и обработки результатов наблюдений массовых случайных явлений для выявления существующих закономерностей.

Генеральной совокупностью называется совокупность всех возможных объектов или реализаций, подлежащих изучению относительно качественного или количественного признака.

Выборкой называется совокупность объектов, отобранных случайным образом из генеральной совокупности.

Для получения хороших оценок характеристик генеральной совокупности необходимо, чтобы выборка была ***репрезентативной***. В силу закона больших чисел можно утверждать, что выборка будет репрезентативной, если ее осуществлять случайно. Выборка называется ***случайной***, если любой объект генеральной совокупности с одинаковой вероятностью может попасть в эту выборку.

Конкретные значения выборки, полученные в результате наблюдений (испытаний), называют ***реализацией выборки*** или ***вариантами выборки*** и обозначают $x_1, x_2, ..., x_n$.

Вариационным рядом называется последовательность значений $x_1, x_2, ..., x_n$ СВ *X*, записанная в возрастающем порядке.

Числа n_i, показывающие, сколько раз встречаются варианты x_i (i = 1, 2, …, n) в ряде наблюдений, называются ***частотами***. Отношение частот к объему выборки называется ***относительными частотами***, которые находятся по формуле

$$w_i = \frac{n_i}{n}. \quad (1)$$

Вариационные ряды бывают дискретными и интервальными. ***Дискретные вариационные ряды*** строят в том случае, если значения изучаемого признака отличаются друг от друга не менее чем на некоторую конечную величину (например, значения получены в результате проведения подсчета).

Если значения изучаемого признака могут отличаться на сколь угодно малую величину (например, значения, полученные в результате проведения измерений), то строят ***интервальные вариационные ряды***.

Перечень вариант и соответствующих им частот или относительных частот называется ***статистическим рядом***.

Статистический ряд распределения выборки при условии, что вариационный ряд *дискретный,* строится следующим образом. В первой строке записываются все значения выборки (варианты), во второй строке – соответствующие им значения частот или относительных частот.

В случае интервального вариационного ряда в первую строку таблицы статистического распределения записывают частичные промежутки: $[x_0, x_1)$, $[x_1, x_2)$,...,$[x_{k-1}, x_k]$, которые, как правило, берутся одинаковыми по длине $h = x_0 - x_1 = x_2 - x_1 = ...$ – шаг разбиения. Длину промежутка можно также найти по формуле $h = \frac{x_{\max} - x_{\min}}{m}$, где m находится по формуле Стерджесса $m = 1 + 3{,}222 \lg n$; $x_{\max} - x_{\min}$ – разность между наибольшим и наименьшим значениями СВ (размах вариации). За начало первого интервала рекомендуется брать $x_{нач} = x_{\min} - \frac{h}{2}$. Во вторую строчку таблицы записывают количество вариант n_i, $(i = \overline{1, k})$, попавших в каждый интервал.

Графическое изображение зависимости между величинами дает возможность представить эту зависимость наглядно (геометрически).

Полигоном частот называют ломаную линию, отрезки которой соединяют точки с координатами $(x_1, n_1), (x_2, n_2), ..., (x_k, n_k)$, а ***полигоном относительных частот*** – ломаную, отрезки которой соединяют точки с координатами (x_1, w_1), (x_2, w_2),..., (x_k, w_k).

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты равны отношению $\frac{n_i}{h}$. ***Гистограммой относительных частот*** называют фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты равны отношению $\frac{w_i}{h}$.

Эмпирической функцией распределения называется функция $F^*(x)$, определяющая для каждого значения *x* относительную частоту события $X < x$:

$$F^*(x) = \frac{n_x}{n}, \tag{2}$$

где n_x – число вариант (наблюдений), при которых наблюдалось значение признака, меньшее x; n – общее число наблюдений (объем выборки).

Свойства эмпирической функции распределения $F^*(x)$:

- значения эмпирической функции принадлежат отрезку [0;1];
- $F^*(x)$ – неубывающая функция;
- если x_1 – наименьшая варианта, то $F^*(x) = 0$, при $x \le x_1$;
- если x_k – наибольшая варианта, то $F^*(x) = 1$, при $x > x_k$.

Пример 1

Производителя продукции детского питания интересует продажа в магазине 6 видов питания для детей. Менеджер в течение 60 дней собирал данные о продаже. Каждый вид детского питания имеет номер от 1 до 6 соответственно. В результате наблюдений получили следующие данные:

3, 2, 5, 6, 1, 4, 6, 4, 6, 3, 6, 4, 2, 1, 5, 3, 1, 6, 4, 5, 4, 2, 2, 4, 2, 6, 3, 1, 5, 6, 1, 6, 6, 4, 2, 5, 4, 3, 6, 4, 1, 5, 6, 3, 2, 4, 4, 5, 2, 5, 6, 2, 3, 5, 4, 1, 2, 5, 3.

1. Что в данной задаче является генеральной совокупностью?
2. Что представляет собой выборка?
3. Постройте:
   a. вариационный ряд;
   b. статистический ряд;
4. Найдите эмпирическую функцию распределения выборки и постройте ее график.
5. Постройте интервальный статистический ряд.
6. Постройте полигон частот и гистограмму частот.
7. Проанализируйте полученные данные.

Решение

1. Генеральной совокупностью в данном наблюдении является реализация всей продукции.

2. Выборка в данном случае представляет собой результат 2-х месяцев продаж данной продукции.

3. Построим вариационный ряд. Для этого упорядочим все элементы выборки по возрастанию.

1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6.

Тогда статистический ряд будет иметь вид, представленный в таблице 1.

Таблица 1 – Статистический ряд распределения

x_i	1	2	3	4	5	6
n_i	7	10	8	12	10	13
w_i	7/60	10/60	8/60	12/60	10/60	13/60

4. Найдем эмпирическую функцию распределения (по формуле (2)) и построим ее график.

Наименьшая варианта (вид детского питания) равна 1, следовательно, $F^*(x)=0$ при $x\le 1$ (см. свойства эмпирической функции распределения). При $1<x\le 2$, значение $x_1=1$ наблюдалось 7 раз, тогда $F^*(x)=\frac{7}{60}=0{,}12$. При $2<x\le 3$ имеем 2 значения, а именно $x_1=1$, $x_2=2$, они наблюдались 7+10 = 17 раз, следовательно, $F^*(x)=\frac{17}{60}=0{,}28$, и т.д. Искомая эмпирическая функция распределения имеет вид

$$F^*(x)=\begin{cases} 0 & \text{при } x\le 1, \\ 0{,}12 & \text{при } 1<x\le 2, \\ 0{,}28 & \text{при } 2<x\le 3, \\ 0{,}42 & \text{при } 3<x\le 4, \\ 0{,}62 & \text{при } 4<x\le 5, \\ 0{,}78 & \text{при } 5<x\le 6, \\ 1 & \text{при } x>6. \end{cases}$$

График эмпирической функции будет иметь следующий вид (рисунок 1):

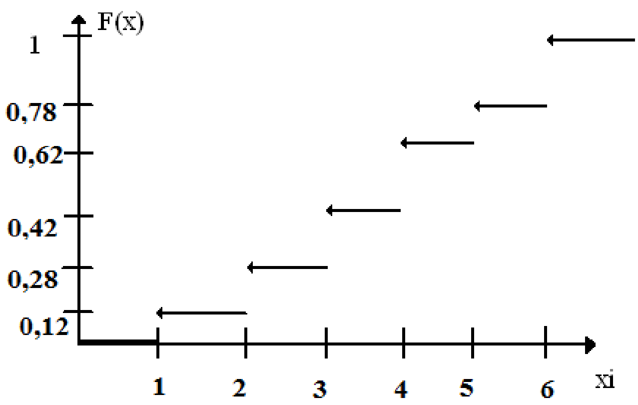

Рисунок 1 – Эмпирическая функция распределения

5. Постоим теперь интервальный статистический ряд. В нашем случае $n=60$. Значение m найдем по формуле $m=1+3{,}222\lg n=1+3{,}222\lg 60=6$. То-

гда длина интервала $h \approx 1$. За начало первого интервала возьмем $x_{нач} = x_{\min} - \frac{h}{2} = 1 - 0{,}5 \approx 0{,}5$.

Интервальный статистический ряд представлен в таблице 2.

Таблица 2 – Интервальный статистический ряд

$[x_{i-1}, x_i)$	[0,5; 1,5)	[1,5; 2,5)	[2,5; 3,5)	[3,5; 4,5)	[4,5; 5,5)	[5,5; 6,5)
n_i	7	10	8	12	10	13

6. Для построения полигона частот (рисунок 2) воспользуемся данными из таблицы 1. На оси *Ox* будем откладывать значения x_i, а на оси *Oy* – значения частот n_i. Соединим соответствующие точки (x_i, n_i) линиями. На рисунках 2 и 3 справа выполнено построение схематично (вручную), слева – построение осуществлялось с помощью Excel.

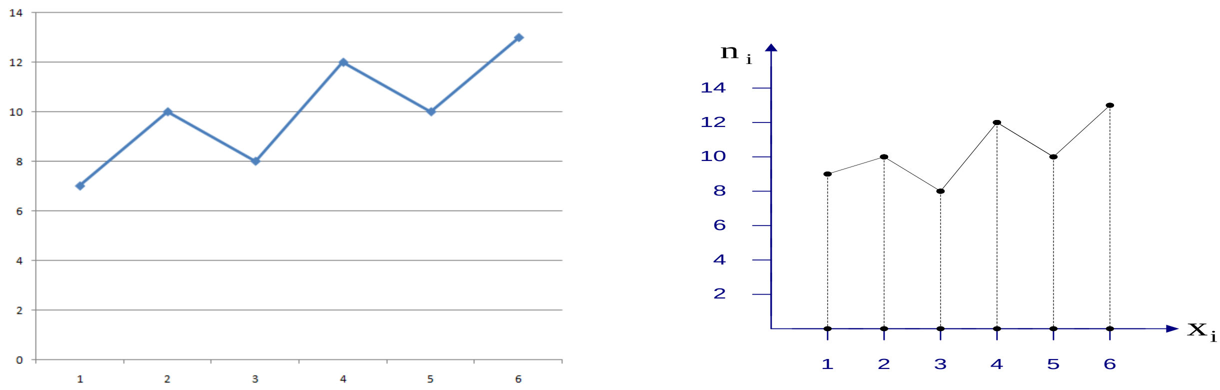

Рисунок 2 – Полигон частот

Для построения гистограммы частот (рисунок 3) воспользуемся данными из таблицы 2. На оси *Ox* будем откладывать интервалы $[x_{i-1}, x_i)$, а на оси *Oy* – значения частот n_i.

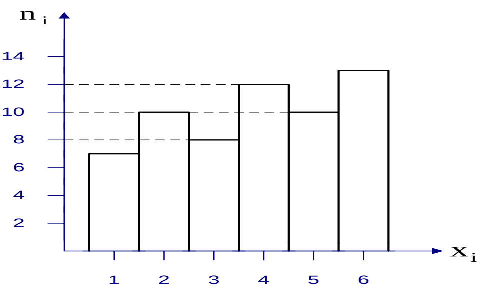

Рисунок 3 – Гистограмма частот

7. Постарайтесь самостоятельно ответить на следующие вопросы:

a) Какой вид детского питания пользуется большим спросом?

b) Какой вид детского питания пользуется наименьшим спросом?

c) Какие бы рекомендации вы дали производителю детского питания?

Задачи

1. При обследовании 50 членов семей рабочих и служащих относительно количественного признака X – количество членов семьи, получены следующие результаты: 5, 3, 2, 1, 4, 6, 3, 7, 9, 1, 3, 2, 5, 6, 8, 2, 5, 2, 3, 6, 8, 3, 4, 4, 5, 6, 5, 4, 7, 5, 6, 4, 8, 7, 4, 5. 7, 8, 6, 5, 7, 5, 6, 6, 7, 3, 4, 6, 5, 4.

Составьте вариационный ряд распределения частот количественного признака X для данной совокупности.

2. Наблюдения за процентным содержанием жира в молоке от 20 коров дали следующие результаты:

3,86 4,06 3,67 3,97 3,76 3,61 3,86 4,04 3,84 3,94 3,98 3,57 3,87 4,07 3,99 3,69 3,76 3,71 3,94 3,82

Постройте по этим данным интервальный вариационный ряд распределения процентного содержания жира в молоке.

3. Данные об урожайности ржи на различных участках колхозного поля приводятся в таблице 3.

Таблица 3 – Данные об урожайности ржи

Урожайность ржи, ц/га	9 – 12	12 – 15	15 – 18	18 – 21	21 – 24	24 – 27
Доля участка от общей посевной площади, %	6	12	33	22	19	8

Постройте гистограмму частот по данным таблицы об урожайности ржи.

4. Через каждый час измерялось напряжение тока (В) в электросети. При этом были получены соответствующие значения, представленные в таблице 4:

Таблица 4 – Результаты измерения напряжения тока в электрической сети

Напряжение	211	212	215	218	220	225	227
Частота	1	1	2	4	5	3	2

Постройте по данным результатов измерения напряжения тока полигон частот.

5. Ниже представлены оценки, данные экспертами вкусу и запаху (так называемая оценка качества) сыра «Российский»:

40, 43, 41, 43, 44, 40, 42, 42, 45, 44, 41, 45, 45, 42, 44, 45, 45, 45, 43, 44.

Постройте по данным оценок качества сыра «Российский» эмпирическую функцию распределения.

Занятие № 2. Числовые характеристики вариационного ряда. Точечные оценки параметров распределения СВ по данным выборки

Числа, характеризующие наиболее существенные черты распределения, называются ***числовыми характеристиками СВ.***

В таблице 5 даны определения основных числовых характеристик.

Таблица 5 – Числовые характеристики

Величина, обозначение	Определение	Формула
1	2	3
$\overline{x_B}$ Выборочная средняя	Среднее арифметическое всех значений выборки	$x_B = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$ или $x_B = \sum_{i=1}^{k} x_i \cdot w_i$, где $w_i = \frac{n_i}{n}$
M_o Мода	Варианта, имеющая наибольшую частоту	

Продолжение таблицы 5

1	2	3
M_e Медиана	Значение признака, приходящееся на середину вариационного ряда	Если число членов ряда нечетное, $n = 2k+1$, то $M_e^* = x_{(k+1)}$. Если четное, т.е. $n = 2k$, то $M_e^* = \dfrac{x_k + x_{k+1}}{2}$
D_B Выборочная дисперсия	Среднее арифметическое квадратов отклонений значений выборки от выборочной средней	$D_B = \dfrac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 \cdot n_i$ или $D_B = \sum_{i=1}^{k}(x_i - \bar{x}) \cdot w_i$, где $w_i = \dfrac{n_i}{n}$
σ_B Среднее квадратическое отклонение	Квадратный корень из дисперсии	$\sigma_B = \sqrt{D_B}$

Выборочная средняя в случае *интервального вариационного ряда* вычисляется по формулам, приведенным в таблице 1, только в этом случае за значение признака в каждом интервале берут середину интервала.

Значение ***моды*** *для интервального ряда* можно получить по формуле

$$M_o = x_0 + h\frac{n_2 - n_1}{(n_2 - n_1) + (n_2 - n_3)}, \tag{3}$$

где x_0 – начальное значение модального интервала, т.е. интервала, который содержит моду; h – длина модального интервала; n_1 – частота интервала, предшествующего модальному, n_2 – частота модального интервала, n_3 – частота интервала, следующего за модальным.

Для *интервального упорядоченного вариационного ряда* ***медиана*** вычисляется по формуле

$$M_e = x_n + \frac{\frac{n}{2} - S_{M_{e-1}}}{n_{M_e}} \cdot h, \quad (4)$$

где x_n – начало медианного интервала, h – ширина медианного интервала, n_{M_e} – частота медианного интервала, $S_{M_{e-1}}$ – сумма частот интервалов, предшествующих медианному, n – объем выборки.

Пример 2

При измерении длины семян полсолнечника (в мм) получены следующие результаты: 10,00; 11,45; 10,50; 11,51; 11,06; 11,14; 11,73; 11,06; 10,46; 11.19; 11,70; 11, 95; 12,27; 11,43; 10,41; 12,75; 11,70; 11,86; 10,03; 11,99; 10,72; 11,35; 12,02; 10;62; 11,11; 10,62; 12,39; 11,22; 11,21; 12,20; 12,28; 11,45; 12.70; 11,96; 12,60; 11,14; 9,88; 10,67; 9,94; 11,28; 11,60; 7,99; 11,97; 10,75; 11,00; 11,78; 11,10; 12,46; 11,30; 11,82.

Постройте по этим данным интервальное распределение выборки. Определите: выборочную среднюю, моду, медиану, дисперсию, среднее квадратическое отклонение.

Решение

Построим интервальный вариационный ряд.

Длину интервала (шаг) найдем по формуле

$$h = \frac{x_{\max} - x_{\min}}{m} = \frac{x_{\max} - x_{\min}}{1 + 3{,}222 \lg n} = \frac{12{,}75 - 7{,}99}{1 + 3{,}222 \lg 50} = 0{,}80 .$$

За начало интервала возьмем $x_{нач} = x_{\min} - \frac{h}{2} = 7{,}99 - \frac{0{,}8}{2} = 7{,}59$.

Построим таблицу, в которую занесем границы интервалов, соответствующие им частоты, а также середины интервалов (таблица 6).

Таблица 6 – Интервальное распределение выборки

Границы интервалов	Частота, N_i	Середина интервалов, x_i	Накопленная частота	$n_i x_i$
1	2	3	4	5
[7,59; 8,39)	1	7,99	1	7,99
[8,39; 9,19)	0	8,79	1	0

Продолжение таблицы 6

1	2	3	4	5
[9,19; 9,99)	2	9,59	3	19,18
[9,99; 10,79)	10	10,39	13	103,9
[10,79; 11,59)	17	11,19	30	190,23
[11,59; 12,39)	15	11,99	45	179,85
[12,39; 13,19)	5	12,79	50	63,95

Для нахождения выборочной средней воспользуемся формулой из таблицы 5:

$$x_B = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i ,$$

где x_i – середина интервала.

Имеем

$$x_B = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i = \frac{7{,}99 + 0 + 19{,}18 + 103{,}9 + 190{,}23 + 179{,}85 + 63{,}95}{50} = 11{,}30 .$$

Для того чтобы найти моду, определим интервал, соответствующей наибольшей частоте. Это интервал [10,79; 11,59). Тогда значение моды найдем по формуле (3).

Получаем $M_o = 10{,}79 + 0{,}80 \cdot \dfrac{17-10}{(17-10)+(15-17)} = 11{,}41$.

Это означает, что чаще всего значение длины семян 11,41(мм) встречается в интервале [10,79; 11,59).

Для определения медианы находим накопленные частоты, складывая последовательно частоты из таблицы 6. Далее устанавливаем накопленную частоту, большую половины общего количества элементов. В нашем случае это частота 30 (30>50). Этой частоте соответствует интервал [10,79; 11,59).Тогда медиану определяем по формуле (4): $M_e = 10{,}79 + \dfrac{50/2 - 13}{17} \cdot 0{,}8 = 11{,}35$.

Величина M_e означает, что половина длин семян распределена до значения 11,35 (т.е. меньше этой), а остальная половина – после (т.е. больше полученной величины).

Найдем дисперсию. Для этого дополним таблицу 6 вычислениями $(x_i)^2$ *и* $(x_i)^2 \cdot n_i$ (таблица 7).

Таблица 7 – Дополнения для вычисления дисперсии

Границы интервалов	Частота, n_i	Середина интервалов, x_i	Накопленная частота	$n_i x_i$	$(x_i)^2$	$n_i(x_i)^2$
[7,59; 8,39)	1	7,99	1	7,99	63,84	63,84
[8,39; 9,19)	0	8,79	1	0	77,26	0
[9,19; 9,99)	2	9,59	3	19,18	91,97	183,94
[9,99; 10,79)	10	10,39	13	103,9	107,95	1079,50
[10,79; 11,59)	17	11,19	30	190,23	125,21	2128,57
[11,59; 12,39)	15	11,99	45	179,85	143,76	2156,40
[12,39; 13,19)	5	12,79	50	63,95	163,58	817,90

Дисперсию можно найти по формуле $D_B = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 \cdot n_i$, из таблицы 5. Но т.к. выборочная средняя у нас уже найдена, то можно воспользоваться следующей формулой:

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

Получим

$$D_B = \frac{63{,}84 + 183{,}94 + 1079{,}50 + 2128{,}57 + 2156{,}40 + 817{,}90}{50} - (11{,}30)^2 =$$
$$= 128{,}603 - 127{,}69 = 0{,}913.$$

Среднее квадратическое отклонение равно $\sigma_B = \sqrt{D_B} = \sqrt{0{,}913} = 0{,}95$.

Статистическая оценка $\tilde{\theta}$ неизвестного параметра θ теоретического распределения (генеральной совокупности) – это приближенное значение, зависящее от наблюдаемых значений признака.

Таким образом, статистическая оценка (или ***статистика)*** является функцией от результатов наблюдений.

Статистика, используемая в качестве приближенного значения неизвестного параметра генеральной совокупности, называется его ***точечной оценкой.***

Статистическую оценку неизвестного параметра генеральной совокупности называют ***несмещенной***, если ее математическое ожидание равно оцениваемому параметру, т.е. $M(\widetilde{\theta}) = \theta$, в противном случае $M(\widetilde{\theta}) \neq \theta$ ее называют ***смещенной***.

Если статистическая оценка параметра сходится по вероятности к оцениваемому параметру. т.е $\forall \varepsilon > 0$ выполняется $\lim\limits_{n \to \infty} p(\widetilde{\theta}_n) \xrightarrow{\text{по вероят}} \theta$, то ее называют ***состоятельной***, т.е $\forall \varepsilon > 0$.

Несмещенная оценка $\widetilde{\theta}_n$ параметра θ называется ***эффективной***, если она при заданном объеме выборки имеет наименьшую возможную дисперсию.

Пусть $x_1,\ x_2,\ x_3, \dots, x_n$ – выборка из генеральной совокупности и $M(X_i) = M(X) = a$, $D(X_i) = D(X)$, тогда:

1) выборочная средняя $\overline{x_B} = \dfrac{1}{n}\sum\limits_{i=1}^{k} x_i \cdot n_i$ – несмещенная и состоятельная оценка математического ожидания $M(X)$;

2) исправленная выборочная дисперсия $s^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - \overline{x})^2 = \dfrac{n}{n-1} D_B$ – несмещенная и состоятельная оценка дисперсии;

3) относительная частота $w = \dfrac{m}{n}$ появления события A в n независимых испытаниях является несмещенной, состоятельной и эффективной оценкой неизвестной вероятности p этого события.

Пример 3

По выборочным данным примера 2 необходимо найти несмещенные оценки дисперсии *D(X)* и среднего квадратического отклонения $\sigma(X)$.

Решение

Исправленную выборочную дисперсию найдем по формуле $s^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i - x)^2 = \dfrac{n}{n-1} D_B;.$

$$s^2 = \frac{50}{50-1} \cdot 0{,}913 = 0{,}931$$

Тогда исправленное среднее квадратическое отклонение равно $\sigma_B = \sqrt{D_B} = \sqrt{0{,}931} = 0{,}96$.

Задачи

1. В итоге четырех измерений некоторой физической величины одним прибором (без систематических ошибок) получены следующие результаты: 8, 9, 11, 12, 8, 10, 11, 10, 11, 10, 11, 9, 11, 10, 9. Найти выборочную среднюю результатов измерений, моду, медиану, а также исправленную выборочную дисперсию.

2. В таблице 8 приведены данные о потреблении основных продуктов питания в Республике Беларусь за период с 1995 по 2003г. (на душу населения в год, кг) (выборочная совокупность):

Таблица 8 – Данные о потреблении основных продуктов питания

Продукты	1995	1996	1997	1998	1999	2000	2001	2002	2003
Мясо и мясопродукты	58	60	60	62	62	59	59	57	58
Молоко и молочные продукты	367	369	350	372	334	295	303	285	265
Яйца (шт.)	297	299	296	271	237	224	224	227	224
Рыба и рыбопродукты	7,3	8,7	8,7	6,1	6,2	9,5	15,0	16,7	15,5
Сахар	32,1	32,4	34,6	37	34,7	34,9	41,1	39,6	32,7
Растительное масло	6,5	6,8	7,0	6,6	8,3	8,7	12,1	13,5	13,2
Картофель	182	188	182	173	170	174	172	170	172
Овощи и бахчевые	83	88	88	87	89	93	98	103	107
Плоды и ягоды	38	39	37	35	24	25	31	47	44
Хлебные продукты	121	122	122	118	115	110	105	98	97

Найти числовые характеристики по выборке для каждого из приведенных продуктов.

3. Обследование качества пряжи на прочность дало результаты, представленные в таблице 9 (интервальное распределение).

Таблица 9 – Результаты обследования пряжи

Кре- пость, г	120-140	140-160	160-180	180-200	200-220	220-240	240-260	260-280
Число случаев	1	6	19	58	53	24	16	3

Найти моду и медиану этого распределения. Сделать выводы.

4. Сведения о величине удоя молока за лактационный период на молочной ферме колхоза помещены в таблице 10.

Таблица 10 – Сведения о величине удоя молока

Величина удоя, кг	Количество коров	Величина удоя, кг	Количество коров
400-600	1	1600-1800	14
600-800	3	1800-2000	12
800-1000	6	2000-2200	10
1000-1200	11	2200-2400	6
1200-1400	15	2400 и выше	2
1400-1600	20		

Найти выборочную среднюю удоя молока на данной ферме.

5. В таблице 11 предоставлены данные о времени протекания химической реакции. По этим данным необходимо найти оценки для дисперсии и среднего квадратического отклонения распределения времени протекания химической реакции.

Таблица 11 – Время протекания химической реакции

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

Занятие № 3. ***Доверительная вероятность. Доверительный интервал. Построение доверительного интервала для математического ожидания нормально распределенного признака при известном и неизвестном среднем квадратическом отклонении***

При небольшом объеме выборки следует использовать интервальные оценки.

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

Интервал $(\tilde{\theta}_1, \tilde{\theta}_2)$, накрывающий с вероятностью γ истинное значение параметра θ, называется ***доверительным интервалом***, а вероятность γ – ***надежностью оценки*** или ***доверительной вероятностью***.

В таблице 12 приведены интервальные оценки математического ожидания нормального распределения при известном и неизвестном среднем квадратическом отклонении.

Таблица 12 – Интервальные оценки

	Известно среднее квадратическое отклонение σ	***Неизвестно среднее квадратическое отклонение*** σ
Интервальное оценивание математического ожидания а	$\left(\overline{X} - t\cdot\frac{\sigma}{\sqrt{n}},\ \overline{X} + t\cdot\frac{\sigma}{\sqrt{n}}\right)$, γ – заданная доверительная вероятность, а t определяется из уравнения $\Phi_0(t) = \frac{\gamma}{2}$ по таблице функции Лапласа	$\left(\overline{X} - t_\gamma\cdot\frac{s}{\sqrt{n}},\ \overline{X} + t_\gamma\cdot\frac{s}{\sqrt{n}}\right)$, где t_γ определяется по таблице функции Стьюдента по уровню значимости $1-\gamma$ и числу степеней свободы $n-1$, а s – исправленное среднее квадратическое отклонение СВ Х

	Известно математическое ожидание a	*Неизвестно математическое ожидание* a
Интервальное оценивание среднего квадратического отклонения σ	$\left(\frac{\sqrt{n}\cdot s_0}{\chi_2}, \frac{\sqrt{n}\cdot s_0}{\chi_1}\right)$, где n – объем выборки, $s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2$, а χ_2, χ_1 определяются по таблице квантилей $\chi^2_{\alpha,n}$ распределения χ^2_n	$\left(\frac{\sqrt{n-1}\cdot s}{\chi_2}, \frac{\sqrt{n-1}\cdot s}{\chi_1}\right)$, где s – исправленное среднее квадратическое отклонение СВ X, χ_2, χ_1 – определяются по таблице $\chi^2_{\alpha,n}$ при $k = n-1$ и $\alpha = \frac{1+\gamma}{2}$ и $\alpha = \frac{1-\gamma}{2}$ соответственно

Пример 4

При измерении массы 5 пирожных типа «Эклер» определенной партии установлено, что их масса (г) равна: 68,00; 71,19; 69,64; 70,15; 70,39.

a) Определить доверительный интервал, который накрывает среднее значение массы пирожного с вероятностью 0,95 при известной дисперсии равной 2г.

b) Определить доверительный интервал, который накрывает среднее значение массы пирожного с вероятностью 0,95 при неизвестной дисперсии.

Решение

a) Найдем границы доверительного интервала среднего веса пирожного типа «Эклер», т.е. границы доверительного интервала для генеральной средней.

Найдем выборочную среднюю

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i = \frac{68{,}00 + 71{,}19 + 69{,}64 + 70{,}15 + 70{,}39}{5} = 69{,}87.$$

По условию задачи $\sigma = 2$г, $n = 5$, $\gamma = 0{,}95$.

Используем формулу из таблицы 12 для нахождения доверительного интервала при известном среднем квадратическом отклонении $\left(\overline{X} - t\cdot\frac{\sigma}{\sqrt{n}},\ \overline{X} + t\cdot\frac{\sigma}{\sqrt{n}}\right)$. Найдем t из соотношения $\Phi_0(t) = \frac{\gamma}{2}$. Имеем

$\Phi_0(t) = \frac{0{,}95}{2} = 0{,}475$. По таблице функции Лапласа [1, Приложение 2] найдем при каком t значение $\Phi_0(t) = 0{,}475$.

$\Phi_0(1{,}96) = 0{,}475$, следовательно, t = 1,96.

Найдем предельную ошибку выборки и доверительный интервал по формуле

$$\Delta = t \cdot \frac{\sigma}{\sqrt{n}}. \tag{5}$$

$\Delta = 1{,}96 \cdot \frac{2}{\sqrt{5}} = 1{,}75$.

Тогда доверительный интервал будет иметь вид

$$\overline{X} - \Delta < \overline{X} < \overline{X} + \Delta.$$

В нашем случае $69{,}87 - 1{,}75 < \overline{X} < 69{,}87 + 1{,}75$.

Значит, $(68{,}12; 71{,}62)$ – искомый доверительный интервал.

b) Теперь найдем границы доверительного интервала при неизвестной дисперсии. Воспользуемся формулой $\left(\overline{X} - t_\gamma \cdot \frac{s}{\sqrt{n}},\ \overline{X} + t_\gamma \cdot \frac{s}{\sqrt{n}}\right)$. Для начала найдем исправленное среднее квадратическое отклонение по формуле

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}. \tag{6}$$

$$s = \sqrt{\frac{(68{,}00-69{,}87)^2 + (71{,}19-69{,}87)^2 + (69{,}64-69{,}87)^2 + (70{,}15-69{,}87)^2 + (70{,}39-69{,}87)^2}{4}} =$$

$$= \sqrt{\frac{5{,}63}{4}} = 1{,}19.$$

Далее величину t_γ определим по таблице функции Стьюдента [1, Приложение 6] по заданному уровню значимости $1-\gamma = 1-0{,}95 = 0{,}05$ и числу степеней свободы $n-1 = 5-1 = 4$. Имеем $t_{0,05;4} = 2{,}78$. Тогда $\overline{X} - t_\gamma \cdot \frac{s}{\sqrt{n}} = 69{,}87 - 2{,}77 \cdot \frac{1{,}19}{\sqrt{5}} = 68{,}39$ и $\overline{X} + t_\gamma \cdot \frac{s}{\sqrt{n}} = 69{,}87 + 2{,}77 \cdot \frac{1{,}19}{\sqrt{5}} = 71{,}34$.

Итак, полученный доверительный интервал (68,39; 71,34).

Задачи

1. Для определения среднего процентного содержания сырого белка в зернах пшеницы было отобрано 626 зерен. Обследование показало, что выборочная средняя равна 16,8, а выборочная дисперсия равна 4. Постройте доверительный интервал, который бы «накрывал» с надежностью 0,95 неизвестное математическое ожидание распределения процентного содержания белка в зернах пшеницы.

2. Для оценки остаточных знаний по высшей математике были протестированы 25 студентов. Получены следующие результаты в баллах: 107, 90, 114, 88, 117, 110, 103, 120, 96, 122, 93, 100, 121, 110, 135, 85, 120, 89, 100, 126, 90, 94, 99, 116, 111. Постройте по этим данным 95%-й доверительный интервал (т.е. с надежностью γ=0,95) для оценки среднего балла тестирования всех студентов курса.

3. Для определения процента изделий второго сорта в партии производится случайная выборка объемом 1000 единиц. Определить доверительные границы для процента изделий второго сорта во всей партии, которые могут быть гарантированы с доверительной вероятностью 0,997, если в выборке оказалось 25 изделий второго сорта.

4. Для определения средней урожайности пшеницы с одного гектара на площади в 20000 га проведено выборочное обследование, результаты которого приведены в таблице13.

Таблица 13 – Урожайность пшеницы

Урожайность пшеницы, ц/га	15,0-17,5	17,5-20,0	20,0-22,5	22,5-25,0	25,0-27,5
Площадь, га	100	600	800	450	50

Какова вероятность того, что средняя урожайность пшеницы на всем массиве отличается от среднего выборочного не более чем на 15 кг?

5. Для выяснения всхожести семян из партии, содержащей 8000 ед., отобрано 500 ед., из них взошло 440 ед. Найти вероятность того, что доля всхожих семян во всей партии отличается от доли их в выборке по абсолютной величине не более чем на 0,03.

6. В совхозе имеется 20000 овец. Случайным образом было отобрано 500 голов, настриг шерсти с которых дал следующие результаты, помещенные в таблицу 14.

Таблица 14 – Результата настрига шерсти

Настриг шерсти с овцы, кг	2,5-3,5	3,5-4,5	4,5-5,5	5,5-6,5	6,5-7,5
Число овец	44	71	162	174	49

Найти границы, в которых с вероятностью 0,978 заключен средний настриг шерсти с одной овцы в совхозе.

Занятие № 4. *Статистическая проверка статистических гипотез. Критерии согласия Пирсона и Колмогорова*

Статистической называют гипотезу о виде неизвестного распределения или о параметрах известных распределений. Например, группа исследователей в области животноводства разработала новый вид кормов, полагая, что эти корма способны повысить жирность молока. Однако может быть и так, что они ошибаются. Как быть? В статистике поступают так. Выдвигают *нулевую гипотезу* H_0, которая состоит в том, что при использовании данного корма средняя жирность молока остается неизменной. Наряду с этим рассматривают гипотезу, состоящую в том, что данный вид кормов эффективен (*альтернативная гипотеза* H_1).

Цель статистической проверки гипотезы заключается в том, чтобы на основании выборочных данных принять решение о справедливости нулевой гипотезы H_0 (либо ее принять, либо ее отклонить).

Так как проверка статических гипотез осуществляется статистическими методами на основании выборочных данных, то решение неизбежно сопровождается вероятностью ошибочного заключения в сторону принятия нулевой гипотезы или альтернативной. Результаты решения относительно нулевой гипотезы можно увидеть в таблице 15.

Таблица 15 – Ошибки первого и второго рода

Статистическое решение	Фактическая оценка нулевой гипотезы	
	Верна	Неверна
Не отвергать нулевую гипотезу	Правильное решение, его вероятность равна $1-\alpha$	Ошибка второго рода, её вероятность равна β

Продолжение таблицы 15

Отвергнуть нулевую гипотезу	Ошибка первого рода, её вероятность равна α *– уровень значимости*	Правильное решение, его вероятность равна $1-\beta$ *– мощность критерия*

Проверка статистических гипотез осуществляется с помощью статистического критерия, который является функцией от результатов наблюдений. *Статистический критерий* (или просто критерий) – это специально подобранная случайная величина, которая служит для проверки нулевой гипотезы, точное или приближенное распределение которой известны.

Критерием согласия называют статистический критерий проверки гипотезы о предполагаемом законе неизвестного распределения. Он используется для проверки согласия предполагаемого вида распределения с опытными данными на основании выборки.

Существуют различные критерии согласия: Пирсона, Колмогорова, Фишера и др.

Критерий согласия Пирсона – наиболее часто употребляемый критерий для проверки гипотезы о нормальном законе распределения генеральной совокупности

Проверка гипотезы о нормальном распределении для *дискретного вариационного ряда* сводится к следующему алгоритму:

1) по формуле $\chi^2=\sum_{i=1}^{k}\frac{(n_i-np_i)^2}{np_i}$ вычисляем $\chi^2_{набл}$ – выборочное значение статистики критерия;
2) находим число степеней свободы $\kappa=l-3$ (l – число различных значений вариант);
3) выбрав уровень значимости α критерия по таблице χ^2 – распределения, находим критическую точку $\chi^2_{\alpha,k}$ [1, Приложение 5];
4) если $\chi^2_{набл}\le\chi^2_{\alpha,k}$, то гипотеза H_0 не противоречит опытным данным; если $\chi^2_{набл}>\chi^2_{\alpha,k}$, то гипотеза отвергается.

В случае *непрерывного вариационного ряда* проверка гипотезы о нормальном распределении сводится к следующему алгоритму:

1) вычисляем выборочную среднюю $\overline{x}_B$ и выборочное среднее квадратичное отклонение σ_B, причем вместо вариант x_i берем среднее арифметическое концов интервала $x_i^* = \frac{x_i + x_{i+1}}{2}$;

2) нормируем СВ X, т.е переходим к новой СВ Y: $y = \frac{x_i - \overline{x}_B}{\sigma_B}$;

3) вычисляем теоретические вероятности попадания в интервалы $(y_i; y_{i+1})$:

$$P(y_i \le y \le y_{i+1}) = \Phi(y_{i+1}) - \Phi(y_i),$$

где $\Phi(y)$ – функция Лапласа [1, Приложение 2];

4) по формуле $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$ вычисляем $\chi^2_{набл}$ – выборочное значение статистики критерия;

5) находим число степеней свободы $\kappa = l - 3$ (l – число интервалов выборки. Далее согласно п.3 п.4 предыдущего алгоритма необходимо проверить гипотезу H_0.

Критерий Колмогорова применяется только для непрерывных случайных величин. Он связывает эмпирическую функцию распределения $F^*(x)$ с функцией распределения $F(x)$ непрерывной СВ X.

Проверку гипотезы с помощью критерия Колмогорова проводят в нижеприведенном порядке:

1) располагаем результаты наблюдений по возрастанию их значений в виде интервального вариационного ряда;

2) находим эмпирическую функцию распределения $F^*(x)$;

3) вычисляем, пользуясь предполагаемой функцией $F(x)$, значения теоретической функции распределения, соответствующие наблюдаемым значениям СВ X;

4) находим для каждого x_i модуль разности между эмпирической и теоретической функциями распределения;.

5) определяем $\lambda = D\sqrt{n} = \max\left|F^*(x) - F(x)\right|\sqrt{n}$;

6) находим критические значения λ_α в зависимости от уровня значимости по таблице16.

Таблица 16 – Критерий значений λ_α

α	0,50	0,40	0,30	0,20	0,10	0,05	0,02	0,01	0,001
λ_α	0,828	0,895	0,974	1,073	1,224	1,358	1,510	1,627	1,950

Если опытное значение $\lambda \geq \lambda_\alpha$, то гипотеза о соответствии теоретического закона распределения с данными выборки отклоняется. Если $\lambda \leq \lambda_\alpha$, то гипотеза принимается.

Пример 5

Наблюдения за межремонтным интервалом Т (в месяц.) работы зерноуборочного комплекса дали следующие результаты:
0,000; 0,001; 0,003; 0,012; 0,044; 0,156; 0,534; 0,802; 0,007; 0,822; 0,873; 0,838; 0,170; 0,476; 0,322; 0,648; 0,991; 0,107; 0,726; 0,393; 0,827; 0,419; 0,071; 0,659; 0,309; 0,927; 0,778; 0,327; 0,961; 0,826; 0,308; 0,414; 0,707; 0,515; 0,729; 0,742; 0,884; 0,632; 0,846; 0,318; 0,394; 0,502; 0,471; 0,306; 0,600; 0,835; 0,678; 0,454; 0,623; 0,648.

a) Проверить при уровне значимости 0,01 с помощью критерия Колмогорова показательное распределение совокупности.

b) Проверить при уровне значимости 0,01 с помощью критерия Пирсона показательное распределение совокупности.

Решение.

а) Проверяемая гипотеза H_0 состоит в том, что случайная величина X имеет нормальное распределение.

Составим интервальный вариационный ряд. Найдем длину интервала по формуле $h = \frac{x_{\max} - x_{\min}}{m} = \frac{x_{\max} - x_{\min}}{1 + 3{,}222 \lg n} = \frac{0{,}991 - 0{,}000}{1 + 3{,}222 \lg 50} = 0{,}15$. За начальное значение первого интервала возьмем $x_{нач} = 0{,}000$. Интервальный вариационный ряд представлен в виде таблице 17, в которой в строки последовательно записаны интервалы, соответствующие им частоты, а также середины интервалов.

Таблица 17 – Интервальный вариационный ряд

$[x_{i-1}, x_i)$	[0,00; 0,15)	[0,15; 0,30)	[0,30; 0,45)	[0,45; 0,60)	[0,60; 0,75)	[0,75; 0,90)	[0,90; 1,05)
n_i	8	2	10	6	11	10	3
x_i	0,075	0,225	0,375	0,525	0,675	0,825	0,975

Найдем выборочную среднюю, дисперсию и среднее квадратическое отклонение. Имеем $\overline{x_B} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$, где x_i – середина интервалов. Тогда

$$\overline{x_B} = \frac{0{,}075\cdot 8 + 0{,}225\cdot 2 + 0{,}375\cdot 10 + 0{,}525\cdot 6 + 0{,}675\cdot 11 + 0{,}825\cdot 10 + 0{,}975\cdot 3}{50} = 0{,}526$$

Дисперсию найдем по формуле:

$$D_B = \overline{x^2} - (\bar{x}_B)^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i)^2\cdot n_i - (\bar{x}_B)^2 =$$

$$= \frac{(0{,}075)^2\cdot 8 + (0{,}225)^2\cdot 2 + (0{,}375)^2\cdot 10 + (0{,}525)^2\cdot 6 + (0{,}675)^2\cdot 11 + (0{,}825)^2\cdot 10 + (0{,}975)^2\cdot 3}{50} -$$

$$-(11{,}30)^2 = 0{,}366 - 0{,}277 = 0{,}090.$$

Тогда среднее квадратическое отклонение равно $\sigma_B = \sqrt{D_B} = \sqrt{0{,}090} = 0{,}300$.

Вычислим эмпирическую функцию распределения $F^*(x_i) = \sum_{i=1}^{7}\frac{n_i}{50}$; теоретическую функцию $F(x\ )$ по формуле $F(x) = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right)$, где $\Phi(x)$ – функция Лапласа, a – выборочная средняя; разности $\left|F^*(x_i) - F(x_i)\right|$. Результаты расчетов занесем в таблицу 18.

Таблица 18 – Результаты расчетов

X	0,00	0,15	0,30	0,45	0,60	0,75	0,90	1,05
$F^*(x_i)$	0	0,16	0,20	0,40	0,52	0,79	0,94	1
$F(x)$	0,040	0,105	0,227	0,401	0,599	0,773	0,894	0,960
$\left\|F^*(x_i) - F(x_i)\right\|$	0,040	0,050	0,027	0,001	**0,079**	0,017	0,040	0,040

$$F(0{,}00) = \frac{1}{2} + \Phi\left(\frac{0{,}000 - 0{,}526}{0{,}3}\right) = 0{,}5 + \Phi(-1{,}75) = 0{,}5 - \Phi(1{,}75) = 0{,}500 - 0{,}4599 = 0{,}040$$

(здесь и далее будем использовать свойство нечетности функции Лапласа $\Phi(-x) = -\Phi(x)$.

$$F(0{,}15) = \frac{1}{2} + \Phi\left(\frac{0{,}15 - 0{,}526}{0{,}3}\right) = 0{,}5 - \Phi(1{,}25) = 0{,}500 - 0{,}3944 = 0{,}105,$$

$$F(0{,}30)=\frac{1}{2}+\varPhi\left(\frac{0{,}30-0{,}526}{0{,}3}\right)=0{,}5-\varPhi(0{,}75)=0{,}500-0{,}2734=0{,}227\,,$$

$$F(0{,}45)=\frac{1}{2}+\varPhi\left(\frac{0{,}45-0{,}526}{0{,}3}\right)=0{,}5-\varPhi(0{,}25)=0{,}500-0{,}0987=0{,}401\,,$$

$$F(0{,}60)=\frac{1}{2}+\varPhi\left(\frac{0{,}60-0{,}526}{0{,}3}\right)=0{,}5+\varPhi(0{,}25)=0{,}500+0{,}0987=0{,}599\,,$$

$$F(0{,}75)=\frac{1}{2}+\varPhi\left(\frac{0{,}75-0{,}526}{0{,}3}\right)=0{,}5+\varPhi(0{,}75)=0{,}500+0{,}2734=0{,}773\,,$$

$$F(0{,}90)=\frac{1}{2}+\varPhi\left(\frac{0{,}90-0{,}526}{0{,}3}\right)=0{,}5+\varPhi(1{,}25)=0{,}500+0{,}3944=0{,}894\,,$$

$$F(0{,}00)=\frac{1}{2}+\varPhi\left(\frac{0{,}000-0{,}526}{0{,}3}\right)=0{,}5-\varPhi(1{,}75)=0{,}500+0{,}4599=0{,}960\,.$$

Максимальное отклонение эмпирической функции распределения от теоретической есть 0,079 (выделена жирным шрифтом в таблице 18), тогда $\lambda=D\sqrt{n}=\max\left|F^*(x)-F(x)\right|\sqrt{n}=\left|F^*(0{,}60)-F(0{,}60)\right|\cdot\sqrt{50}=0{,}560$.

Критическое значение критерия Колмогорова при $\alpha=0{,}01$ равно $\lambda_\alpha=1{,}627$ (см. таблицу 16). Так как $\lambda\le\lambda_\alpha$, т.е. 0,560<1,627, то гипотеза H_0 о соответствии теоретического закона распределения с данными выборки принимается.

b) Проверим гипотезу H_0 о том, что случайная величина X имеет нормальное распределение с помощью критерия Пирсона.

Вернемся к таблице 17. Так как частоты $n_2=2$ и $n_7=3$ меньше 5, то объединяем соответствующие этим частотам интервалы с соседними. Так как СВ X определена на $(-\infty;+\infty)$, то крайние интервалы заменяем на $(-\infty;0{,}30)$ и $[0{,}75;+\infty)$. Тогда:

$$p_1=p\{-\infty<X<0{,}30\}=\varPhi_0\left(\frac{0{,}30-0{,}526}{0{,}3}\right)-\varPhi_0(-\infty)=\frac{1}{2}-\varPhi_0(0{,}75)=0{,}227\,,$$

$$p_2=\varPhi_0\left(\frac{0{,}45-0{,}526}{0{,}3}\right)-\varPhi_0\left(\frac{0{,}30-0{,}526}{0{,}3}\right)=\varPhi_0(-0{,}25)-\varPhi_0(-0{,}75)=0{,}175\,,$$

$$p_3 = \Phi_0\left(\frac{0{,}6-0{,}526}{0{,}3}\right) - \Phi_0\left(\frac{0{,}45-0{,}526}{0{,}3}\right) = \Phi_0(0{,}25) + \Phi_0(0{,}25) = 0{,}197\ ,$$

$$p_4 = \Phi_0\left(\frac{0{,}75-0{,}526}{0{,}3}\right) - \Phi_0\left(\frac{0{,}6-0{,}526}{0{,}3}\right) = \Phi_0(0{,}75) - \Phi_0(0{,}25) = 0{,}175,$$

$$p_5 = p\{0{,}75 < X < +\infty\} = \Phi_0(+\infty) - \Phi_0\left(\frac{0{,}75-0{,}526}{0{,}3}\right) = \frac{1}{2} - \Phi_0(0{,}75) = 0{,}227\ .$$

Все вычисления занесем в таблицу 19:

Таблица 19 – Результаты вычислений

$[x_{i-1}, x_i)$	$(-\infty; 0{,}30)$	[0,30; 0,45)	[0,45; 0,60)	[0,60; 0,75)	$[0{,}75; +\infty)$
n_i	10	10	6	11	13
p_i	0,227	0,175	0,197	0,175	0,227
$n \cdot p_i$	11,35	8,75	9,85	8,75	11,5

Вычисляем

$\chi^2_{набл}$:

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = \frac{(10-11{,}35)^2}{11{,}35} + \frac{(10-8{,}75)^2}{8{,}75} + \frac{(6-9{,}85)^2}{9{,}85} +$$

$$+ \frac{(11-8{,}75)^2}{8{,}75} + \frac{(13-11{,}5)^2}{11{,}5} \approx 2{,}41$$

Находим число степеней свободы, $\kappa = l - 3 = 5 - 3 = 2$. По условию $\alpha = 0{,}01$ и $\kappa = 3$ по таблице χ^2 – распределения находим $\chi^2_{\alpha,k} = 9{,}2$. Получили, $\chi^2_{набл} < \chi^2_{\alpha,k}$. Следовательно, нет оснований отвергнуть проверяемую гипотезу H_0 и данные выборки не противоречат нормальному распределению.

Задачи

1. Предприятие, выпускающее определенную продукцию утверждает, что добиться более высокой прибыли ему помешали неравномерность поставок по месяцам года. Поставщик утверждает, что поставки были равномерными. Распределение поставок представлено в таблице 20.

Таблица 20 – Распределение поставок

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
Объем поставок	19	23	26	18	20	20	20	20	32	27	35	40

При $\alpha = 0{,}05$ определите, кто прав – предприятие или поставщик?

2. Проектный контролируемый размер шоколадной конфеты составляет $a_0 = 35$ мм. Измерения 25 случайно отобранных конфет дали следующие результаты, представленные в таблице 21.

Таблица 21 – Результаты измерений размера конфет

Контролир. разм., мм	34,8	34,9	35	35,1	35,2	35,3
Число изделий	3	4	5	4	6	3

Требуется на уровне значимости $\alpha = 0{,}05$ проверить гипотезу H_0: $a = 35$ мм, при конкурирующей гипотезе H: $a \neq 35$ мм.

Литература

1. Гмурман, В.Е. Теория вероятностей и математическая статистика. – М: Высш. Шк., 1977. – 480 с.

2. Гмурман, В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высш. Шк., 1979. – 400 с.

3. Гусак, А. А. Высшая математика: В 2 т. Т. 2. Учеб. для студентов вузов. – 5-е изд. – Минск: ТетраСистемс, 2004. – 448 с.

4. Гнеденко, Б.В. Курс теории вероятностей. – М., 1965.

5. Письменный, Д.Т. Конспект лекций по теории вероятностей и математической статистике. – М.: Айрис-прес, 2004. – 256 с.

Учебное издание

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания

Составитель
Шендрикова Ольга Александровна

Редактор *А.А. Щербакова*
Технический редактор *Т. В. Багуцкая*

Подписано в печать 09.06.2011. Формат 60х84 1/16.
Бумага офсетная. Гарнитура Таймс. Ризография.
Усл. печ. л. 1,9. Уч.-изд. л. 2.
Тираж 30 экз. Заказ 92.

Учреждение образования
«Могилевский государственный университет продовольствия».
ЛИ № 02330/013913 от 08.02.2007.
Пр-т Шмидта, 3, 212027, Могилев.

Отпечатано в учреждении образования
«Могилевский государственный университет продовольствия».
Пр-т Шмидта, 3, 212027, Могилев.