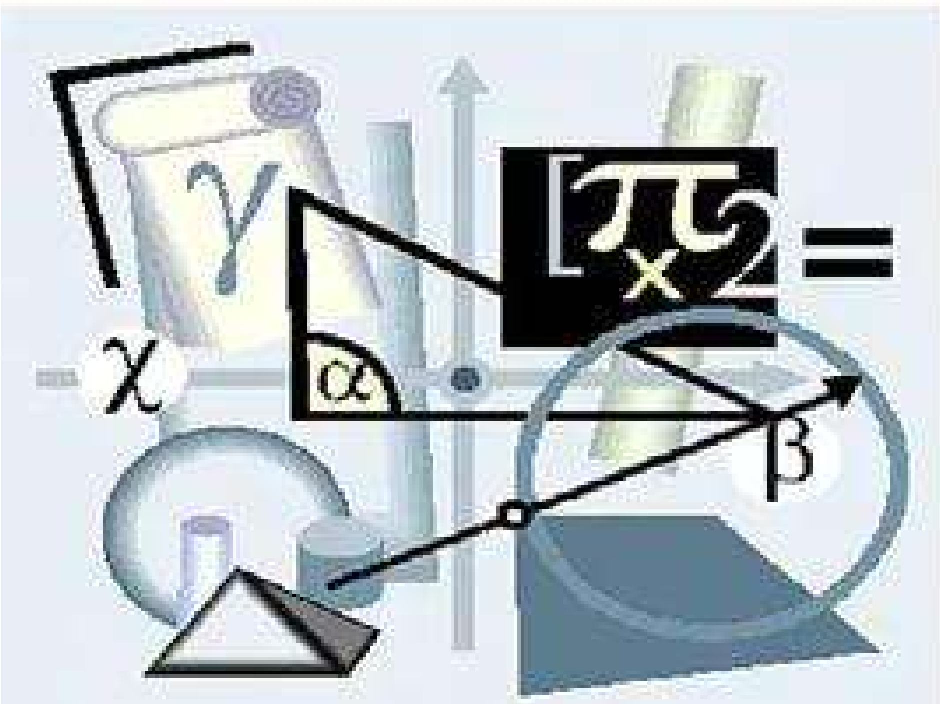
Министерство сельского хозяйства Российской Федераций
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«Приморская государственная сельскохозяйственная академия»
Институт земледелия и агротехнологий

Савельева Е.В.

Основы математической биостатистики

Учебное пособие для обучающихся специальности: 36.05.01 - «Ветеринария» и направлениям подготовки: 36.03.02 – «Зоотехния»; 36.03.01 «Ветеринарно – санитарная экспертиза»



Уссурийск, 2016

УДК 517

Составитель: Савельева Е.В., канд. техн. наук, доцент кафедры физики и высшей математики ПГСХА.

Рецензент: Н.А. Чугаева, к.б.н., доцент кафедры химий и генетики ИЖиВМ ФГБОУ ПГСХА.

Основы математической биостатистики: учебное пособие для обучающихся специальности 36.05.01 - «Ветеринария» и направлениям подготовки: – 36.03.02 – «Зоотехния»; 36.03.01 «Ветеринарно – санитарная экспертиза» /ФГБОУ ВО ПГСХА; сост. Е.В.Савельева. – Уссурийск, 2016. – 210 с.

Пособие содержит методические материалы для освоения дисциплины, задания для практических занятий и самостоятельной работы обучающихся, имеющих прикладной характер для биологических направлений.В пособии рассмотрены понятия: дискретной и непрерывной случайной величины, вариационные ряды, расчет сводных характеристики выборки, доверительные интервалы, испытания гипотез, дисперсионный анализ, линейная корреляция. Приведены варианты контрольных работ для организации самостоятельной работы студентов.

Издается по решению методического совета ФГБОУ ВО «Приморская государственная сельскохозяйственная академия».

Введение

Курс теории вероятностей и математической статистики входит в цикл фундаментальных дисциплин, изучение которых является обязательным для студентов сельскохозяйственных учебных заведений.

Одной из важнейших сфер приложения теории вероятностей и математической статистики является животноводство. Развитие современного животноводства сопровождается накоплением большого количества информации по многим вопросам генетики, селекции, продуктивности, здоровья животных, поведенческих функций и т.д. В задачу науки входят классификация, упорядочение и систематизация этих данных, их научный анализ. Подобный подход позволяет формулировать практические предложения, способствующие ускорению развития тех или иных отраслей животноводства, совершенствовать и создавать новые перспективные отрасли, прогнозировать развитие того или иного направления. В ветеринарии, дополнительно к перечисленным возможностям, использование научного анализа позволяет теоретически моделировать течение болезни или действия лечебных факторов и разрабатывать методы профилактики и лечения животных.

Основные теоретические положения математической статистики базируются на теории вероятностей. Основное отличие математической статистики от теории вероятностей в том, что в математической статистике рассматриваются не только действия над законами распределения и числовыми характеристиками, но и приближенные методы отыскания этих законов и характеристик по результатам экспериментов.

Цель данных учебно-методических указаний – помочь изучающим теорию вероятностей и математическую статистику в усвоении необходимых теоретических знаний и приобретении практических навыков для квалифицированного использования статистической информации в целях принятия правильных решений в вопросах прогнозирования.

Оглавление

РАЗДЕЛ 1. СЛУЧАЙНЫЕ СОБЫТИЯ.

Глава 1. Относительная частота появления события. Классическое определение вероятности. Элементы комбинаторики.

1.1. Необходимый теоретический минимум.

Классификация случайных событий. Относительная частота и вероятность появления события.

Предметом теорий вероятностей является изучение вероятностных закономерностей массовых однородных случайных событий. Знание закономерностей, которым подчиняются массовые случайные события, позволяет предвидеть, как эти события будут протекать.

Определение 1.1. Опыт (испытание) – это создание комплекса условий. Событие – это результат опыта.

Пример. Опыт – подбрасывание игральной кости. Событие – выпадение трех очков.

События, которые могут произойти в результате опыта, можно подразделить на три вида:

а) ***достоверное событие*** – событие, которое всегда происходит при проведении опыта;

б) ***невозможное событие*** – событие, которое в результате опыта произойти не может;

в) ***случайное событие*** – событие, которое может либо произойти, либо не произойти.

События обозначаются заглавными буквами *A,B,C,D*.

Пример. Брошена игральная кость – *испытание*;

Выпадение 4-х очков – *событие;* какое? - *случайное;*

Выпадение не больше 6-и очков - *событие;* какое? – *достоверное;*

Выпадение 10-и очков - *событие*; какое? – *невозможное.*

Классификация случайных событий.

1. События А и В называются ***совместными,*** если появление одного не исключает появление другого в результате одного опыта, в противном случае (появление одного исключает появление другого) события называются ***несовместными.***

Примеры:

1) совместными событиями являются попадания двух стрелков в мишень;
2) несовместными появление карты черной масти и карты красной масти.

2. События называют ***единственно возможными,*** если появление в результате испытания одного и только одного из них является достоверным событием.

Пример. Стрелок произвел выстрел по цели. События: А – попадание в цель, В – промах *единственно возможные* в данном испытании.

3. События называют ***равновозможными***, если есть основание считать, что ни одно из этих событий не является более возможным, чем другие.

Пример. Брошена игральная кость. Равновозможные события: А – выпала 3 и В – выпала 5; или С – выпало четное число и Д – выпало нечетное число очков.

4. Случайные события образуют полную группу, если они попарно ***несовместны,*** но в результате опыта одно из них обязательно произойдет.

Пример.

Стрелок произвел 2 выстрела.

Полная группа событий: $\{A_1, A_2, A_3\}$; где

Событие A_1 - промах;

Событие A_2 - одно попадание;

Событие A_3 - два попадания;

5. ***Противоположными*** называют *два* единственно возможных события образующих полную группу событий или противоположным событием к событию A называется событие $\bar{A}$, которое происходит только тогда, когда событие A не происходит.

Пример. Опыт – подбрасывание монеты.

A – выпадение решки; $\bar{A}$ - выпадение орла.

Определение 1.2. Относительной частотой события А в данной серии опытов называется отношение числа опытов, в которых появилось событие А, к общему количеству проведенных испытаний:

$$W(A)=\frac{m}{n}100\% \tag{1.1}$$

где n – общее число опытов, m – число появлений события A.

Пример. Посеяно 100 зерен, взошло 80. Найти частоту всхожести.

$$W(A)=\frac{m}{n}=\frac{80}{100}=0{,}8 \;\; или \;\; 80\%$$

Пример. По цели произвели 24 выстрела, причем было зарегистрировано 19 попаданий. Относительная частота поражения цели $W(A)=\frac{19}{24}$.

Определение 1.3. Статистической вероятностью случайного события A называется число, характеризующее возможность наступления события A. Приближенными значениями статистической вероятности служат частости события при большом числе испытаний

$$P^*(A)=\lim_{n\to\infty}\frac{m}{n} \tag{1.2}$$

Приближенными значениями статистической вероятности служат частости события при большом числе испытаний

$$P(A)=\lim_{N\to\infty}\frac{M}{N} \qquad \text{или} \qquad P(A)=W(A)$$

Пример. Естествоиспытатель К. Пирсон терпеливо подбрасывал монету и после каждого бросания не ленился записывать полученный результат. Проделав эту операцию 24 000 раз, он обнаружил, что герб выпадал в 12 012 случаях. Вычисляя относительную частоту выпадения герба, он получил, что практически равно 1/2, т.к. $W(A)=\frac{m}{n}=\frac{12012}{24000}=0{,}5005$

Вероятность является одним из основных понятий теории вероятностей. Существует несколько определений этого понятия. Рассмотрим определение, которое называют классическим.

Каждый из возможных результатов испытания, т.е. каждое событие, которое может наступить в испытании, назовем *элементарным исходом.*

Те элементарные исходы, при которых интересующее нас событие наступает, назовем *благоприятствующими* этому событию.

Определение 1.4. Вероятностью события А называют отношение числа благоприятствующих этому событию исходов к общему числу всех единственно возможных и равновозможных исходов испытания.

$$P(A)=\frac{M}{N} \qquad (1.3)$$

где M – число элементарных исходов, благоприятствующих событию A;

N – число всех возможных элементарных исходов испытания.

Вероятность есть число, характеризующее возможность появления события.

Свойства вероятности.

Свойство 1. Вероятность достоверного события равна единице.

Свойство 2. Вероятность невозможного события равна нулю.

Свойство 3. Вероятность случайного события есть положительное число, заключенное между нулем и единицей.

Пример. В урне содержится 6 одинаковых, тщательно перемешанных шаров, причем 2 из них – красные, 3 – синие и 1 – белый. Из урны наудачу вынимают один шар. Какова вероятность того, что извлеченный шар: а) красный; б) синий; в) белый?

Решение. а) Пусть событие A – извлекли красный шар.

Число благоприятствующих событию A исходов, $m=2$ (т.к. в урне 2 красных шара);

Число возможных исходов n=6 (т.к. всего 6 шаров) $P(A)=\frac{2}{6}=\frac{1}{3}$;

б) Пусть событие B – извлекли синий шар: $P(B)=\frac{3}{6}=\frac{1}{2}$;

в) Пусть событие C – извлекли белый шар: $P(C)=\frac{1}{6}$;

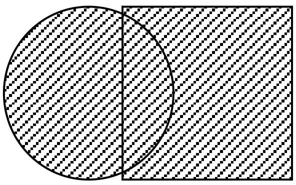
Операции над событиями

Правило суммы. Если некоторый объект *A* может быть выбран из совокупности объектов $\boldsymbol{m}$ способами, а другой объект *B* – $\boldsymbol{k}$ способами, то выбрать либо *A*, либо *B* можно $\boldsymbol{m+k}$ способами.

Пример. В урне 4 зелёных, 6 красных, 5 белых шаров. Сколькими способами можно выбрать из урны цветной шар?

Решение. Цветной шар – это шар зелёного или красного цвета. Зелёный можно извлечь 4 способами (m=4), красный – 5 способами (k=5). Тогда по *правилу суммы* цветной шар можно достать $m+k$=4+5=**9 способами.**

Пример. Событие А – попадание в круг, а событие В – попадание в квадрат. Тогда их сумма А+В заключается в попадании или в круг или в квадрат



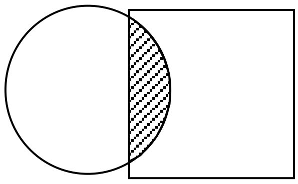
Правило произведения. Пусть из первого множества один элемент можно выбрать $\boldsymbol{m}$ способами. При этом из второго множества один элемент выбирается $\boldsymbol{n}$ способами. Тогда пара элементов из этих двух множеств может быть выбрана $\boldsymbol{m\cdot n}$ способами.

- Правило произведения обобщается и на выборку из нескольких множеств, то есть число выборок по одному элементу из *k* множеств равно $m_1\cdot m_2\cdot\ldots\cdot m_k$.

Пример. Даны два множества чисел {1; 2; 3} и {4; 5}. Первое число выбирают из первого множества, а второе число – из второго множества. Сколько различных двузначных чисел можно составить, выбирая числа указанным способом?

Решение. В данном случае m=3 и n=2 $\Rightarrow$ по *правилу произведения* $m \cdot n$=6. Т.е. указанным способом из данных множеств можно составить **всего 6 чисел**. (Действительно, это числа 14, 24, 34, 15, 25, 35. Других вариантов нет!)

Пример. Событие А – попадание в круг, а событие В – попадание в квадрат; Тогда их произведение АВ заключается в попадании в общую часть круга и квадрата.



Основные формулы комбинаторики.

При вычислении вероятностей часто приходится использовать некоторые формулы ***комбинаторики*** – науки, изучающей комбинации, которые можно составить по определенным правилам из элементов некоторого конечного множества. Определим основные такие комбинации.

Определение 1.5*. *Перестановки – это комбинации, составленные из всех n элементов данного множества и отличающиеся только порядком их расположения.

$$\boldsymbol{P_n = n! =} 1 \cdot 2 \cdot 3 \cdot 4 \cdot \ldots\ldots\ldots \cdot n \qquad (1.4)$$

Пример. Сколько различных списков (отличающихся порядком фамилий) можно составить из 7 различных фамилий?

Решение. $P_7 = 7! = 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$.

Определение 1.6. Размещения – комбинации из m элементов множества, содержащего n различных элементов, отличающиеся либо составом элементов, либо их порядком (порядок важен).

$$A_n^m = \frac{n!}{(n-m)!} \qquad (1.5.)$$

Пример. Сколько двузначных чисел с различными цифрами можно составить из цифр 1,2,3,4 ?

Решение. Здесь m=2, n=4. Используем размещения, т.к. каждая комбинация может отличаться элементами и их порядком, то по формуле (2) находим:

$$A_4^2 = \frac{4!}{(4-2)!} = 12 \quad (12,21,13,31,23,24\ldots\ldots).$$

Замечание. Если элементы в комбинации повторяются, то имеем ***размещения с повторениями,*** их число равно n^m

Определение 1.7. Сочетания – неупорядоченные наборы из *m* элементов множества, содержащего *n* различных элементов (то есть наборы, отличающиеся только составом элементов).

$$C_n^m = \frac{n!}{m!(n-m)!} \tag{1.6}$$

Пример. В отборочных соревнованиях принимают участие 10 человек, из которых в финал выходят трое. Сколько может быть различных троек финалистов?

Решение. В отличие от предыдущего примера, здесь не важен порядок финалистов, следовательно, ищем число сочетаний из 10 по 3:

$$C_{10}^3 = \frac{10!}{3!\cdot 7!} = \frac{8\cdot 9\cdot 10}{6} = 120.$$

Геометрическое определение вероятности.

Случайно из области S выбирается точка A. Необходимо определить вероятность попадания точки A в подобласть R.

Роль элементарных событий в данном эксперименте играют точки области S. Все множество точек области S образует пространство элементарных событий. Все элементарные события – равновозможны, так как все точки области S равноправны в отношении попадания туда случайной точки A. Но число этих элементарных событий бесконечно. Тогда, вероятность случайного события A определяется равенством:

$$P(A) = \frac{R}{S}$$

Пример. В прямоугольник 5×4 см^2 вписан круг радиуса 1,5 см. Какова вероятность того, что точка, случайным образом поставленная в прямоугольник, окажется внутри круга?

Решение. По определению геометрической вероятности искомая вероятность равна отношению площади круга (в который точка должна попасть) к площади прямоугольника (в которой точка ставится), т.е. = 0,353.

1.2. Примеры решения типовых задач.

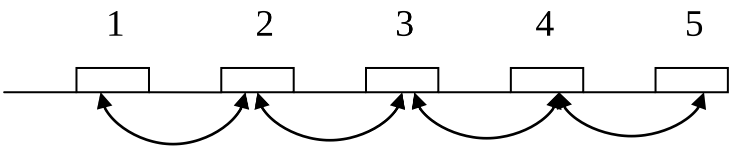
Задача 1.1. Для прохождения практики на завод приехали пять студентов, которым предложили пять различных станков, расположенных в одном ряду. На рабочие места студентов распределили произвольно. Найти вероятность того, что два друга окажутся рядом.

Решение. Рассмотрим событие: A – два студента окажутся за станками, расположенными рядом.

Общее число случаев расстановки пяти студентов по пяти станкам равно числу перестановок из пяти различных элементов.

$n = P_5 = 5! = 2 \cdot 3 \cdot 4 \cdot 5 = 120$

Для определения числа случаев благоприятствующих тому, что два студента будут работать рядом, составим следующую схему рабочих мест.



Если два друга попадут за первые два станка, то при двух различных случаях они будут работать рядом. Если один будет за первым, а другой за вторым станком или наоборот. Таким образом, каждый из четырех, показанных на схеме случаев, имеет два варианта. Тогда общее число случаев благоприятствующих тому, что эти студенты будут работать рядом, равно: $m = 4 \cdot 2 \cdot 3! = 48,$ где $P_3 = 3!$-комбинации оставшихся трех мест.

Вероятность искомого события равна: $P(A) = \frac{48}{120} = 0{,}4$

Задача 1.2. Для подготовки к экзаменам студентам предложили 30 вопросов. Студент подготовил только 20 из них. Каждый экзаменационный билет содержит три вопроса. Какова вероятность того, что данному студенту попадется набор вопросов, два из которых он знает.

Решение. Рассмотрим событие: *A*- студент знает ответ на два вопроса билета. Так как набор вопросов комплектуется из 30 и порядок вопросов не играет роли, то общее число возможных вариантов равно числу сочетаний из 30 вопросов по три вопроса, т.е.

$$n = C_{30}^{3} = \frac{30!}{3! \cdot 27!} = \frac{27! \cdot 28 \cdot 29 \cdot 30}{1 \cdot 2 \cdot 3 \cdot 27!} = \frac{28 \cdot 29 \cdot 30}{2 \cdot 3} = 4060$$

Студенту необходимо, чтобы попались любые два вопроса, которые он знает. И эти два вопроса могут сочетаться с любым вопросом, который он знает. Поэтому общее число случаев, благоприятствующих такому сочетанию, равна m= $C_{20}^{2} \cdot 10$ т.к. студент знает 20 вопросов и, следовательно, остальные 10 не знает.

$$C_{20}^{2} = \frac{20!}{2! \cdot 18!} = \frac{18! \cdot 19 \cdot 20}{1 \cdot 2 \cdot 18!} = \frac{19 \cdot 20}{2} = 190 \qquad m = 190 \cdot 10 = 1900$$

Тогда вероятность искомого события равна: $P(A) = \frac{m}{n} = \frac{1900}{4060} = 0{,}468 \approx 0{,}47$

Задача 1.3. На семи карточках написаны буквы Д,Н,Е,Т,С,Т,У. Перемешав карточки, вынимают одну за другой и кладут рядом. Какова вероятность того, что составится слово

1) «СТУДЕНТ»; 2) «НЕТ»

Решение. 1) Рассмотрим событие: *A*- составлено слово «СТУДЕНТ»

Общее число случаев равно числу перестановок из 7различных объектов.

$$n = P_5 = 7! = 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 = 5040$$

Число благоприятствующих случаев равно числу перестановок повторяющейся буквы «Т»

$$m = P_2 = 2! = 1 \cdot 2 = 2$$

Используя классическое определение вероятности, имеем $P(A) = \frac{2}{5040} = \frac{1}{2520}$

2) Рассмотрим событие: *B*- составлено слово «НЕТ».

В данном опыте число всевозможных случаев n равно числу размещений из 7 элементов по 3 элемента , т.к. комбинации могут отличаться как самими элементами, так и их порядком.

$$n= A_7^3 = \frac{7!}{(7-3)!} = \frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7}{1\cdot 2\cdot 3\cdot 4} = 5\cdot 6\cdot 7 = 210$$

Число благоприятствующих случаев равно $m=1$ (из заданного набора букв только один раз можно составить слово «НЕТ»)

В результате получаем $P(A) = \dfrac{1}{210}$

1.3. Задачи для самостоятельного решения.

1.Опыт – бросание двух монет. Является ли полной следующая группа событий: A – появление двух гербов; B – появление герба и решки?

2.Опыт – два выстрела по мишени. Являются ли несовместными события:

C – хотя бы одно попадание; D – хотя бы один промах?

3. При проверке семян овса из 200 посеянных семян взошло 160.

Какова доля всходов?

4. При транспортировке арбузов из 1000 испортилось 50. Какова относительная частота (в процентах) хороших арбузов.

5.Относительная частота появления клубней картофеля, имеющих механические повреждения при уборке равна 0,15. В корзине 350 клубней. Сколько клубней окажется поврежденными?

6. В коробке 10 карандашей, из них 3 красных, остальные чёрные. Какова вероятность, что наудачу взятый карандаш окажется: 1) красным; 2) чёрным; 3) белым.

7. Выпущено 100 лотерейных билетов, причем установлено 11 денежных призов, причём из них 8 выигрышей по 100 рублей, 2 по 500 рублей, 1 по 100 рублей. Студент купил 25 билетов, три из них принесли ему выигрыш по 100 рублей и один 500 рублей, остальные билеты оказались безвыигрышными. Найти вероятность и относительную частоту события: 1) купленный билет выигрышный; 2) на купленный билет пал выигрыш 100

рублей; 3) на купленный билет пал выигрыш 500 рублей; 4) на купленный билет пал выигрыш 1000 рублей.

6.В стаде 100 коров, из них 40 не превышают трёхлетнего возраста. Наудачу отбирается одно животное. Найти вероятность того, что возраст коровы не менее трёх лет.

7.В книге 170 страниц. Какова вероятность, что номер наудачу открытой страницы, есть число кратное 7?

8.Брошены две игральные кости. Определить вероятность того, что сумма вскрывшихся очков равна 5.

9.Какова вероятность, что номер зачетки студента: а) четный; б) делится на пять; в) оканчивается нулем?

10.Вычислите вероятность выпадения двух гербов при трехкратном бросании одной монеты.

11.Сколькими способами можно выбрать один цветок из корзины в которой имеется 12 гвоздик, 15 роз и 7 хризантем?

12.Сколько различных трехзначных чисел можно составить из цифр 0,2,3,4,5 если: а) цифры не повторяются; б) цифры могут повторяться?

13.В хозяйстве 5 участков земли, которые необходимо занять под 5 культур. Какова вероятность, что произвольное закрепление культур за участками совпадает с запланированными.

14. Набирая номер телефона абонент забыл последние две цифры и помня, что эти цифры различны, набрал наудачу. Найти вероятность того, что набраны нужные цифры.

15.В клетке 6 подопытных мышей, из них 4 белые, 2 серые. Случайно отбирают 2 мыши без возраста. Найти вероятность того, что: 1) обе мыши серые; 2) обе мыши белые.

16.Исследователь зафиксировал результаты полевого опыта с 20 делянок, 7 из них засеяно ячменем, остальные пшеницей. Результаты внёс в ЭВМ. При распечатке ведомости результаты «смешались».а) Найдите вероятность того, что при этом каждой делянке соответствует верный результат; б) Отбираются результаты для трёх делянок. Какова вероятность, что они засеяны ячменём? Пшеницей? Только одна ячменём?

17. В лаборатории работает 20 человек, из них 55% женщин; 6 сотрудников должны уехать в командировку. Какова вероятность того, что среди них женщин и мужчин будет поровну?

18.На пяти карточках написаны буквы: А, В, Г, Л, О. Тщательно перемешав карточки, вынимают их одна за другой и кладут рядом в порядке поступления. Какова вероятность того, что получится слово:

1) «ВОЛГА»; 2) «ВОЛ».

19. В отрезке *AB* длины 3 случайно появляется точка С. Определить вероятность того, что расстояние от точки С до В превосходит 1.

20*.Первый замок с «секретом» имеет на общей оси четыре диска, каждый из которых разделен на пять секторов, на которых написаны различные буквы. Второй замок имеет пять дисков, разделенных на четыре сектора. Замок можно открыть при определенной комбинации букв относительно корпуса замка. Какой замок лучше?

21*.Пять человек вошли в лифт на 1-ом этаже девятиэтажного дома. Какова вероятность, что пассажиры выйдут из лифта на нужных этажах?

22*.Сколькими способами можно расставить на книжной полке десятитомник произведении Д. Лондона, располагая их:

а) в произвольном порядке;

б) так, чтобы 1, 5 и 9 тома стояли рядом (в любом порядке);

в) так, чтобы 1,2,3 тома не стояли рядом (в любом порядке).

1.Для определения всхожести пшеницы посеяли две серии по 200 зерен. Получено соответственно 189 и 193 всходов. Какова относительная частота всхожести в каждой серии?

2. Из крупной партии яиц было отобрано наудачу 400 штук. Из них оказалось 50 негодных. Найти относительную частоту годных яиц (в %).

3.В урне 22 шара, различающихся только цветом (9 синих, 5 желтых, 8 белых). Что более вероятно: извлечение из урны желтого шара или появление трех очков при бросании игральной кости?

4.Результаты взвешивания четырёх клубней картофеля следующие: 93г, 80г, 131г, 200г. Какова вероятность, что наугад взятый клубень имеет вес не менее 100г.

5. Из города А в город В ведут 5 дорог, а из города В в город С – 3 дороги. Сколько путей, проходящих через В, ведут из А в С?

6. Из полного комплекта домино (28 шт.) выбирается наугад одна кость. Какова вероятность, что это будет 3 : 3? И какова вероятность, что следующую кость можно приставить к первой.

7. Четыре мальчика сидят на одной скамеечке. Какова вероятность, что в следующий раз они сядут в том же порядке, если любой порядок для них одинаково возможен.

8. В хозяйстве 7 комбайнов требуют ремонта, три их них зерноуборочные, четыре картофелеуборочные. Наугад выбирают два комбайна. Найти вероятность того, что: 1) все два комбайна зерноуборочные; 2) два картофелеуборочные.

9.Имеется материал 5-ти цветов. Каждый цвет пронумерован. Какова вероятность, что закройщик не помня номера цветов составит полотно флага Россий?

10.На пяти карточках написаны буквы Т, С, О, Р, П. Тщательно перемешав карточки, вынимают их одна за другой и кладут рядом в порядке поступления. Какова вероятность, что получится слово: а) «СПОРТ»; б) «СТО»?

11.В номере прибора три последние цифры неизвестны. Какова вероятность, что наугад записанный номер окажется верным, если известно, что: 1) эти цифры различны; 2) могут повторяться?

12.На первом курсе студенты слушают лекцию по восьми предметам. Первого сентября в расписании включаются четыре лекции по разным предметам. Какова вероятность того, что студент не знающий расписания, угадает все предметы, по которым будут прочитаны лекции 1 сентября?

13. В прямоугольном броневом щите размером 2х1 м имеется невидимая для противника амбразура размером 10х10 см. Определить вероятность того, что пуля, попавшая в щит, попадет в амбразуру, если попадание в любую точку щита равновозможно.

14*.Группа туристов из 12 юношей и 7 девушек выбирает по жребию 5 человек для приготовления ужина. Найти вероятность следующих событии: а) все пять человек девушки; б) все пятеро – юношей; в) 1 юноша и 4 девушки; г) 3 юноши и 2 девушки.

15*.На каждой из восьми одинаковых карточек написаны числа: 2,4,6,7,8,11,12,13. Карточки тщательно перемешаны. Наудачу берутся карточки. Найти вероятность того, что образованная из двух выбранных чисел дробь сократима?

16.

Вопросы для самопроверки.

1.Предмет теорий вероятности.

2. Дать определение и привести примеры достоверного, невозможного и случайного события, понятие полной группы событий.

3. Относительна частота, её свойства.

4. Статистическое определение вероятности.

5. Классическое определение вероятности, её свойства.

6. Перестановки: определение, формула для подсчета. Примеры.

7.Размещения: определение, формула для подсчета. Примеры.

8. Сочетания: определение, формулы для подсчета.Примеры.

5. Формула полной вероятности. Формула гипотез.

Глава 2. Действия над событиями. Основные теоремы алгебры событий. Формула полной вероятности. Формула Байеса.

2.1. Необходимый теоретический минимум.

Основные понятия суммы и произведение событий.

Определение 2.1. Суммой (объединением) двух совместных событии A и B называют событие С, состоящее в появлении или события A или события B или *A и B* вместе. ***C = или A или B или (A и B)***

Пример. A- попал в мишень 1 стрелок; B – попал в мишень 2 стрелок.

$C = A+B$ - мишень поражена.

- *Суммой двух несовместных событии* А и В называют событие С, состоящее в появлении или события А или события В , *С = или А или В.*

Пример. Из колоды карт наугад достается одна карта.

A – достать карту – пик; B - достать карту – крести; $C = A+B$ – достать карту черной масти.

- *Суммой нескольких событий,* соответственно, называется событие, заключающееся в том, что произошло хотя бы одно из этих событий.

Пример. Два стрелка делают по одному выстрелу по мишени. Если событие А – попадание первого стрелка, а событие В – второго, то сумма $A+B$ – это хотя бы одно попадание при двух выстрелах.

Определение 2.2. Произведением A*B событий *A и B* называется событие ***C***, состоящее в том, что произошло и событие A, и событие B.

C = A*B = и A и B

Пример. Студент сдает экзамен, при этом он учил 30 вопросов. В билете 3 вопроса. Рассмотрим событие С – студент сдал экзамен на отлично.

Студент сдаст успешно экзамен, если он:

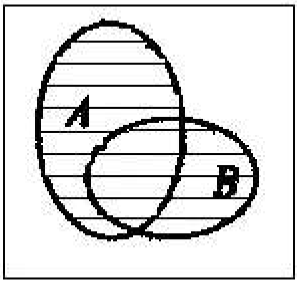
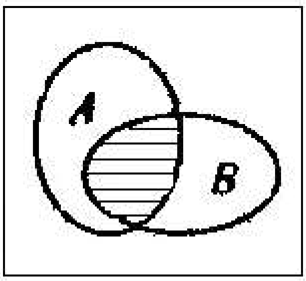
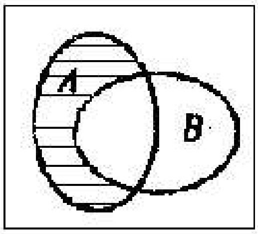
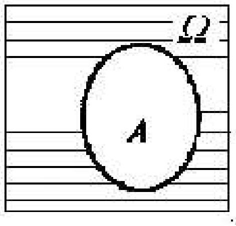
A_1 – знает 1 вопрос в билете; A_2 - знает 2 вопрос в билете; A_3 - знает 3 вопрос в билете,тогда С = и A_1 и A_2 и A_3, следовательно С – есть произведение событий: $C = A_1 * A_2 * A_3$.

Определение 2.3. Противоположным (дополнением) событием к событию A называется событие $\overline{A}$, которое происходит только тогда, когда событие A не происходит.

Операций над событиями соответствуют аналогичные операции над множествами (см. таблица 2.1).

Геометрическая интерпретация действий над событиями.

Таблица 2.1.

Определение	***Геометрическая интерпретация***	***Пример***
Суммой (объединением) событий *A* и *B* называется новое событие *C=A+B*, которое заключается в наступлении хотя бы одного из событий: *или* А *или B или A* и *B*	***A+B*** 	*A* – {награждение победителя призом} *B* – {награждение победителя денежной премией} *A+B* – {награждение победителя или призом, или премией, или и тем и другим}
Произведением (пересечением) событий *A* и *B* называется новое событие *C=A·B*, которое заключается в наступлении событий *A и B* одновременно	***A·B*** 	*A* – {награждение победителя призом} *B* – {награждение победителя премией} *A·B* – {награждение победителя одновременно и призом и премией}
Разностью событий *A* и *B* называется событие *A\B*, которое заключается в наступлении события *A* и ненаступлении события *B*	***A|B*** 	*A* – {награждение победителя призом} *B* – {награждение победителя денежной премией} *A\B* – {награждение победителя призом без выдачи премии}
Отрицанием события *A* называется событие $\overline{A}$ (не *A*), заключающееся в ненаступлении события *A*	$\overline{A}$ 	*A* – {награждение победителя призом} $\overline{A}$ – {ненаграждение победителя призом}

Зависимые и независимые события. Условная вероятность.

Для нахождения вероятности от произведения событий, рассмотрим понятия независимости и зависимости событий на следующем примере.

Пример. Из урны, содержащей 6 белых и 4 черных шара, наудачу последовательно вынимают два шара. Рассмотрим событие С, что эти шары белого цвета.

Решение.

Пусть: А – первый раз достать белый шар; В – второй раз достать белый шар;

С – достать два белых шара, С = и А и В, следовательно имеем произведение С = А*В;

Вероятность события С найдем при двух различных условиях.

1 условие (схема с возвращением): После того как первый раз достали белый шар, его вернули в урну, потом стали доставать второй шар.

При таком условий имеем:

P(A) =6/10; P(B) = 6/10

В этом случае вероятность появления события В не зависела от события А. Такие события называются независимыми

Определение 2.4. События называются ***независимыми,*** если появление одного из них не изменяет вероятности появления другого события.

2 условие (схема без возвращения): После того как первый раз достали белый шар, его не вернули в урну, потом стали доставать второй шар.

При таком условий имеем:

P(A) =6/10; P(B) = 5/9

В этом случае вероятность появления события В зависела от события А. Такие события называются зависимыми

Определение 2.5. События называются ***зависимыми,*** если появление одного из них изменяет вероятность появления другого события.

Вероятность события *В* вычисляли при условий, что событие А произошло, такая вероятность называется условной.

Определение 2.6. Условной вероятностью события B при осуществлении события A называется такая вероятность, которая вычислена при предположении, что событие B состоялось. Обозначение $P(B/A)=P_A(B)$.

Аналогично $P(A/B)=P_B(A)$- условная вероятность события A при условии, что событие B уже состоялось.

Вывод: *При случайном выборе событий по схеме с возвратом события независимые, а по схеме без возврата – зависимые.*

Основные теоремы алгебры событий.

Теорема 2.1. (сложения).

а) Если события A и B несовместные, то вероятность от суммы:

$$P(A+B)=P(A)+P(B) \tag{2.1}$$

б) Если события A и B несовместные, то вероятность от суммы:

$$P(A+B)=P(A)+P(B)-P(AB) \tag{2.2}$$

Следствия.

1. Вероятность от суммы событий, образующих полную группу равна 1.

$$P(A_1+A_2+..+A_n)=P(A_1)+P(A_2)+....+P(A_n)$$

2. Сумма вероятностей противоположных событий равна 1.

$$P(A)+P(\overline{A})=1 \tag{2.3}$$

Пример. Подбрасывается игральная кость. Какова вероятность выпадения не менее 5 очков?

Решение. С – выпадение не менее 5 очков – это сумма двух событий:

А – выпадение 5 очков;

В – выпадение 6 очков,

где А и В – несовместные события, следовательно по теореме (1) имеем:

$$P(C)=P(A)+P(B)=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$$

Теорема 2.2. (произведение).

а) Вероятность от произведения зависимых событий равна произведению вероятности первого события на условную вероятность второго.

$$P(A\cdot B)=P(A)\cdot P(B/A) \tag{2.4}$$

б) Если события A и B независимые, то

$$P(A\cdot B)=P(A)\cdot P(B) \tag{2.5}$$

Вернемся к примеру и найдем вероятность события С в случае зависимости (2 условие):

$$P(A\cdot B)=P(A)\cdot P(B/A)=\frac{6}{10}\cdot\frac{5}{9}=\frac{1}{3}$$

В этом случае имеем схему без возвращения.

В случае независимости (1 условие):

$$P(A\cdot B)=P(A)\cdot P(B)=\frac{6}{10}\cdot\frac{6}{10}=0{,}36$$

Следствия.

1. Если имеем большее число событий, то теорема 2 запишется в виде:

$$P(A\cdot B\cdot C)=P(A)\cdot P(B/A)\cdot P(C/A\cdot B)$$

2. Вероятность появления хотя бы одного события из группы событий, независимых в совокупности: $P(A)=1-P(\overline{A}_1)\cdot P(\overline{A}_2)\cdot\ldots\cdot P(\overline{A}_n)$

Формула полной вероятности. Формула Байеса.

Определение 2.7. Пусть событие A может произойти только совместно с одним из событий $H_1, H_2,\ldots, H_n$, образующих полную группу несовместных событий. Тогда события $H_1, H_2,\ldots, H_n$ называются ***гипотезами.***

Теорема 2.3. Вероятность события A, наступающего совместно с гипотезами $H_1, H_2,\ldots, H_n$, равна:

$$p(A)=\sum_{i=1}^{n}p(H_i)p(A/H_i),$$

где $p(H_i)$ – вероятность i- й гипотезы, а $p(A/H_i)$ – вероятность события A при условии реализации этой гипотезы. Данная формула носит название ***формулы полной вероятности***.

Пример. Имеются три одинаковые урны с шарами. В первой из них 3 белых и 4 черных шара, во второй – 2 белых и 5 черных, в третьей – 10 черных шаров. Из случайно выбранной урны наудачу вынут шар. Найти вероятность того, что он белый.

Решение. Будем считать гипотезами H_1, H_2 и H_3 выбор урны с соответствующим номером.

H_1- урна № 1; H_2- урна № 2 ;H_3 - урна № 3

Так как по условию задачи все гипотезы равновозможны, то $p(H_1)=p(H_2)=p(H_3)=\frac{1}{3}$. Найдем условную вероятность A при реализации каждой гипотезы: $p(A/H_1)=\frac{3}{7}$, $\quad p(A/H_2)=\frac{2}{7}$, $p(A/H_3)=0$.

Тогда $p(A)=\frac{1}{3}\cdot\frac{3}{7}+\frac{1}{3}\cdot\frac{2}{7}+\frac{1}{3}\cdot 0=\frac{5}{21}\approx 0{,}238.$

Формула Байеса (теорема гипотез).

Пусть известен результат опыта, а именно то, что произошло событие *A*.

Этот факт может изменить априорные (то есть известные до опыта) вероятности гипотез. Например, в предыдущем примере извлечение из урны белого шара говорит о том, что этой урной не могла быть третья, в которой нет белых шаров, то есть $p\,(H_3/A) = 0$.

Формулы Бейеса позволяют переоценить вероятности гипотез после того, как становится известным результат испытания (событие А наступило).

$$p(H_i\,/\,A)=\frac{p(H_i)\,p(A/\,H_i)}{p(A)}.$$

Пример. После двух выстрелов двух стрелков, вероятности попаданий которых равны 0,6 и 0,7, в мишени оказалась одна пробоина. Найти вероятность того, что попал первый стрелок.

Решение. Пусть событие *A* – одно попадание при двух выстрелах, а гипотезы:

H_1 – первый попал, а второй промахнулся,

H_2 – первый промахнулся, а второй попал,

Вероятности гипотез: $p(H_1) = 0{,}6\cdot 0{,}3 = 0{,}18$, $p(H_2) = 0{,}4\cdot 0{,}7 = 0{,}28$, тогда $p(A/H_1) = p(A/H_2) = 1$,

Следовательно, полная вероятность $p(A) = 0{,}18\cdot 1 + 0{,}28\cdot 1 = 0{,}46$.

Применяя формулу Байеса, получим: $p(H_1\,/\,A)=\frac{0{,}18\cdot 1}{0{,}46}=\frac{9}{23}\approx 0{,}391.$

2.2. Примеры решения типовых задач.

Задача 1.1. Среди 8 лотерейных билетов имеется 5 билетов с выигрышем. Наудачу покупают 3 билета. Какова вероятность, что а) все три билета выигрышные,

б) два билета из трех выигрышные.

Решение:

Для вычисления вероятности используем классическое определение вероятности: $P(A)=\dfrac{M}{N}$ Обозначим события:

а) A – все три наудачу купленные билеты выигрышные.

Общее число возможных случаев в данном опыте равно числу способов, которыми можно извлечь три билета из 8, при этом комбинации, составленные из 8 по 3 билета, есть сочетания, т.к. отличаются хотя бы одним элементом. Число сочетаний вычисляют по формуле:

$$C_n^m=\frac{n!}{m!(n-m)!},\qquad \text{,где}\quad n!=1\cdot 2\cdot 3\cdot\ldots\cdot(n-1)\cdot n$$

В нашей задаче

$$N=C_8^3=\frac{8!}{3!(8-3)!}=\frac{8!}{3!5!}=\frac{1\cdot 2\cdot 3\cdot 4\cdot 5\cdot 6\cdot 7\cdot 8}{1\cdot 2\cdot 3\cdot 1\cdot 2\cdot 3\cdot 4\cdot 5}=56$$

Число благоприятствующих случаев есть число комбинаций из 5 билетов по 3, это число сочетаний из 5 по 3 :

$$M=C_5^3=\frac{5!}{3!\cdot(5-3)!}=\frac{1\cdot 2\cdot 3\cdot 4\cdot 5}{1\cdot 2\cdot 3\cdot 1\cdot 2}=10\qquad P(A)=\frac{10}{56}=\frac{5}{28}$$

Можно предложить второй способ решения этой задачи, используя теорему умножения вероятностей для зависимых событий.

$$P(A)=P\left(A_1\right)\cdot P\left(A_2/A_1\right)\cdot P(A_3/A_1A_2),$$

где $P(A_2/A_1)$ - условная вероятность события A_2 при условии, что произошло событие A_1.

$P(A_3/A_1A_2)$- условная вероятность появления события A_3 при условии, что произошли события A_1 и A_2.

В нашей задаче:

A_1 – первый купленный билет выигрышный.

A_2 – второй купленный билет выигрышный.

A_3 – третий купленный билет выигрышный.

Тогда $A = A_1 \cdot A_2 \cdot A_3$. События A_1, A_2, A_3 зависимые.

$$P(A_1) = \frac{5}{8};\; P(A_2/A_1) = \frac{4}{7};\; P(A_3/A_1A_2) = \frac{3}{6} = \frac{1}{2} \qquad P(A) = P(A_1A_2A_3) = \frac{5}{8} \cdot \frac{4}{7} \cdot \frac{1}{2} = \frac{5}{28}$$

Как видим, ответы совпали.

б) B – два из трех наудачу купленных билетов выигрышные.

Число всех возможных случаев: $N = C_8^3 = 56$

Число благоприятствующих событию B случаев по правилу произведения: $M = C_5^2 \cdot C_3^1 = \frac{5!}{2!\cdot 3!} \cdot \frac{3!}{2!\cdot 1!} = \frac{3 \cdot 4 \cdot 5}{2} = 30$. Здесь C_5^2 есть число вариантов отбора выигрышных билетов; C_3^1 - число вариантов отбора невыигрышных билетов, тогда: $P(B) = \frac{30}{56} = \frac{15}{28}$.

Задача 1.2. Два студента независимо один от другого ищут книгу в магазинах. Вероятность того, что книга будет найдена первым студентом, равна 0,6, а вероятность того, что вторым – 0,8. Найти вероятность того, что:

1) оба студента найдут нужную книгу;
2) ни один студент не найдет нужную книгу;
3) только один студент найдет нужную книгу;
4) хотя бы один из студентов найдет нужную книгу.

Решение:

1) Обозначим события:

A – оба студента нашли нужную книгу,

A_1 – первый студент нашел книгу,

A_2 - второй студент нашел книгу.

Тогда событие A заключается в том, что и первый студент нашел книгу, и второй студент нашел книгу, т.е. имеем произведение событий: $A=A_1 \cdot A_2$, по условию задачи A_1 и A_2независимы, поэтому используем формулу умножения вероятностей для независимых событии: $P(A) = P(A_1) \cdot P(A_2)$

В нашей задаче: $P(A_1) = 0{,}6;\quad P(A_2) = 0{,}8,$ поэтому $P(A_1) = 0{,}6 \cdot 0{,}8 = 0{,}48$

2) B - ни один из студентов не нашел нужной книги.

Событие В состоит в том, что оба студента не нашли нужной книги.

Обозначим: $\overline{A}_1$ - первый студент не нашел книгу,

$\overline{A}_2$ - второй студент не нашел книгу.

Тогда $B = \overline{A}_1 \cdot \overline{A}_2$, получим $P(B) = P(\overline{A}_1) \cdot P(\overline{A}_2)$.

Событие $\overline{A}_1$ - противоположное по отношению к событию A_1, поэтому:

$P(\overline{A}_1) + P(A_1) = 1 \;\Rightarrow\; P(\overline{A}_1) = 1 - P(A_1) \;\Rightarrow 1 - 0{,}6 = 0{,}4$

Аналогично: $P(\overline{A}_1) = 1 - P(A_1) \;\Rightarrow 1 - 0{,}8 = 0{,}2$

Тогда $P(B) = 0{,}4 \cdot 0{,}2 = 0{,}08$

3) C - только один из студентов найдет нужную книгу.

I	II
+ и –	
или	
– и +	

а)

Событие C состоит в следующем: первый студент найдет книгу и одновременно второй студент не найдет книгу или первый студент не найдет книгу и второй студент одновременно найдет нужную книгу, т.е. это событие можно представить следующей схемой (а), отсюда событие C представим следующим образом: $C = A_1 \cdot \overline{A}_2 + \overline{A}_1 \cdot A_2$.

Используя теорему умножения для независимых событий и теорему сложения вероятностей для несовместных событий:

$$P(A + B) = P(A) + P(B)$$

получим:

$P(C) = P(A_1) \cdot P(\overline{A}_2) + P(\overline{A}_1) \cdot P(A_2)$

$P(C) = 0{,}6 \cdot 0{,}2 + 0{,}4 \cdot 0{,}8 = 0{,}12 + 0{,}32 = 0{,}44$

I	II
+ и +	
или	
– и +	
или	
+ и -	

б)

4) D - хотя бы один из студентов найдет нужную книгу.

Вероятность этого события можно найти различными способами.

I способ. Событие D состоит в том, что нужную книгу найдут оба студента или один из них. Это событие можно представить в виде схемы б).

Отсюда: $D = A + C \Rightarrow P(D) = P(A) + P(C) \Rightarrow P(D) = 0{,}48 + 0{,}44 = 0{,}92$

II способ. Событие D противоположно событию B, т.е. $D = \overline{B}$,

отсюда: $P(D) = P(\overline{B}) = 1 - P(B) = 1 - 0{,}08 = 0{,}92$.

III способ. Используем теорему сложения вероятностей для совместных событий: $P(A + B) = P(A) + P(B) - P(AB)$

В нашем случае:

$$P(A + B) = P(A) + P(B) - P(AB) = 0{,}6 + 0{,}8 - 0{,}6 \cdot 0{,}8 = 1{,}4 - 0{,}48 = 0{,}92$$

Как видим, ответы совпали при решении задачи различными способами.

Задача 2.3. На склад поступает продукция трех ферм, причем продукция первой фермы составляет 20 %, второй 46 %, третьей 34 %. Известно также, что средний процент бракованного мяса для первой фермы равен 3%, второй – 2%, третьей – 1%. Найти вероятность того, что:

1) наудачу взятое мясо будет бракованное;

2) бракованное мясо получено с первой фермы.

Решение:

1) Наудачу взятая туша мяса может быть привезена или с первой фермы (гипотеза H_1), или со второй (гипотеза H_2), или с третьей (гипотеза H_3). События несовместны и составляют полную группу. Вероятности гипотез H_n даны в условии задачи: P(H_1) = 0,20, P(H_2) = 0,46, P(H_3) = 0,34. Известны и условные вероятности. Вероятность того, что наудачу взятая туша мяса будет бракованной (событие А) при условии, что мясо привезено с первой фермы (гипотеза H_1): $P(A/H_1) = 0{,}03$·

Аналогично $P(A/H_2) = 0{,}02;$

$$P(A/H_3) = 0{,}01$$

Событие А (наудачу взятая туша мяса будет бракованной) может произойти только вместе с одним из несовместных событий H_n, поэтому полную вероятность события А определим по формуле полной вероятности:

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) + P(H_3)P(A/H_3) =$$
$$= 0{,}2 \cdot 0{,}03 + 0{,}46 \cdot 0{,}02 + 0{,}34 \cdot 0{,}01 = 0{,}0186$$

2) Известно, что событие *A* уже произошло, требуется найти послеопытную вероятность гипотезы H_1. По формуле Бейеса:

$$P(H_1/A) = \frac{P(H_1) \cdot P(A/H_1)}{P(A)} = \frac{0{,}2 \cdot 0{,}03}{0{,}0186} = 0{,}322$$

2.3. Задачи для самостоятельного решения.

1. Некоторая популяция растений, в количестве 100 штук , состоит из особей трёх типов, помеченных АА, Аа, аа. Численность каждого типа составляет соответственно 20, 65, 15. Из популяции выбирают одно растение. Найдите вероятность, что выбранное растение принадлежит к типу АА или аа.

2. На ферме работают два транспортера для раздачи кормов. Вероятность выхода из строя каждого из них соответственно равна 0,2; 0,3. Какова вероятность, что произойдёт поломка, если одновременный выход из строя этих транспортеров исключён.

3. Производится бомбометание по трём складам боеприпасов, причём сбрасывается одна бомба. Вероятность попадания в 1-ый склад 0,1; во второй – 0,05; в третий – 0,03. При попадании в один из складов взрываются все три. Найти вероятность, что склады будут взорваны.

4. В корзине 12 плодов. Из них 4 плода поражены болезнью в скрытой форме. Из корзины последовательно извлекают два плода. Найти вероятность того, что: 1) оба плода здоровы, если первый плод после осмотра не вернули в корзину; 2) оба плода здоровы, если первый после осмотра вернули в корзину.

5. Многолетними наблюдениями установлено, что в некотором районе в сентябре 10 дней дождливые хозяйство должно в течение первых трёх дней сентября выполнить определенную работу. Чему равна вероятность того, что первые три дня будет без дождя.

6. Событие, что зерно прорастет, зависит от следующих факторов: 1) посевной материал отвечает стандарту; 2) подготовка почвы к посеву выполнена с соблюдением правил агротехники; 3) содержание влаги, необходимое для роста, оптимальное. Вероятности выполнения этих событий соответственно равны 0,8; 0,9; 0,7. Найти вероятность того, что зерно прорастет.

7. Коэффициент использования рабочего времени (относительное время) у двух комбайнов соответственно равны 0,8 и 0,6. Учитывая, что остановки в работе каждого комбайна случайны и независимы, определить относительное время: 1) совместной работы комбайнов; 2) работы только одного комбайна.

8.Механик обслуживает 3 прибора. Вероятность безотказной работы этих приборов в течение часа соответственно равны 0,7 ; 0,8 ; 0,9.Найти вероятность того, что в течении часа: 1) все три будут исправны; 2) откажут все три прибора; 3) только два будут исправны; 4) хотя бы один будет исправен ?

9. Бой стеклопосуды может произойти при погрузке или транспортировке. Вероятность боя при погрузке равна 0,15,а при транспортировке – 0,23.Какова вероятность боя стеклопосуды?

10. Вероятность того, что студент А правильно решил задачу равна 0,75. Для студента В вероятность верно решить задачу – 0,8. Какова вероятность того, что задача будет решена верно, если оба студента будут решать ее независимо друг от друга? (Решить разными способами).

11.В системе находятся пять клапанов. Вероятность отказа для любого клапана равна 0,3.Какова вероятность, что хотя бы один клапан откажет?

12*. Три охотника стреляют по зайцу с вероятностью попадания 0,3; 0,5; 0,7. Какова вероятность, что заяц будет убит двумя пулями.

13*. Три электрические цепи (рис.2.1), (рис.2.2) составлены из элементов 1,2,3. Выход из строя этих элементов равен 0,1. Какова вероятность выхода из строя: а) первой цепи; б) второй цепи. Какова вероятность того, что эти цепи работают.

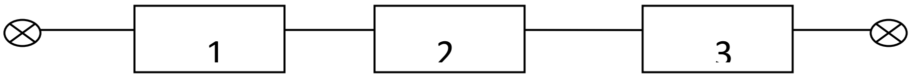
а) 

Рисунок. 2.1

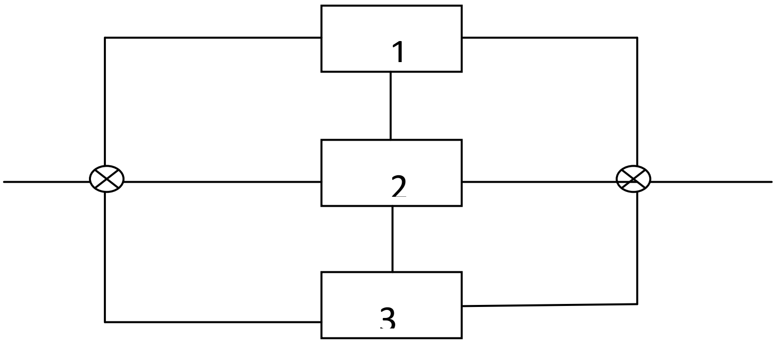
б) 

Рисунок. 2.2

14*. Четыре охотника договорились стрелять по дичи в определенной последовательности. Следующий охотник производит выстрел лишь в случае промаха предыдущего. Вероятность попадания в цель каждым из охотников одинакова и равна 0,8. Найти вероятность того, что будет произведено: а) один, б) два, в) три, г) четыре выстрела.

15. В трёх корзинах находится картофель. В первой 90% неповрежденных клубней, во второй 85% и в третьей – 80%. Из наугад выбранной корзины берут один клубень. Какова вероятность, что клубень не повреждён.

16. В первой коробке 20 цыплят, из них 4 петушка, во второй коробке 15 цыплят, из них 3 петушка. Из второй коробки переложили в первую одного цыплёнка. После этого из первой коробки достают наудачу одного цыплёнка. Какова вероятность, что это петушок.

17. Для проверки качества зерна пшеницы было установлено, что зёрна могут быть разбиты на три группы. К зёрнам 1 группы принадлежат 96% зёрен; ко 2-й группе - 2%; к 3-й группе – 2%. Вероятность, что зерно прорастёт: для 1 группы равно 0,5; для 2 группы – 0,7; для 3 группы – 0,8. Определить вероятность того, что: 1) взятое наугад зерно прорастёт; 2) прорастёт зерно из 2 группы.

18. При исследовании жирности молока коров все стадо было разбито на три группы. В первой группе оказалось 70%, во второй 23% и в третьей 7% всех коров. Вероятность того, что молоко, полученное от отдельной коровы, имеет не менее 4% жирности, равна 0,6; 0,35 и 0,1 для каждой группы коров соответственно. 1.Определить вероятность того, что для взятой наудачу коровы, жирность молока составит не менее 4%.; 2.Взятая наудачу корова дает молоко жирностью не менее 4%. Найти вероятность того, что эта корова из первой группы.

19.Азотное удобрение поступает на склад хозяйства из пункта 1 и пункта 2, причём , из 1 –го пункта в 2 раза больше, чем из 2-го. Вероятность того, что удобрение из первого пункта удовлетворяет стандарту 0,9, а соответствующая вероятность для второго пункта 0,7. Случайным образом взятое для пробы на складе хозяйства удобрение оказалось стандартным. Найти вероятность того, что взятое удобрение из 2 –го пункта.

---

1. В хозяйстве на каждые выращенные 100 арбузов – 30 арбузов с весом более 3,5 кг, 25 арбузов – от 3 до 3,2 кг, остальные менее 3 кг. Определить вероятность того, что вес наудачу выбранного арбуза свыше 3 кг.

2. Вероятность завести двигатель у трактора при первой попытке 0,35; при второй 0,4. Какова вероятность, что двигатель заведён.

3. В стаде животных из 24 голов одной породы, 4 животных не получили прививку. Наудачу последовательно, без возвращения отбирают трёх животных. Найти вероятность, что эти трое животных будут без прививок.

4. Покупатель приобрёл пылесос, полотёр, вентилятор. Вероятность того, что пылесос не выйдет из строя в течение гарантийного срока равна 0,95; для полотёра 0,9., для вентилятора 0,8. Найти вероятность того, что: 1) все приборы выдержат гарантийный срок; 2) ни один прибор не выдержит гарантийный срок; 3) хотя бы один прибор выдержит гарантийный срок; 4) только один прибор выдержит гарантийный срок. Какие из этих событий образуют полную группу.

5. Два стрелка стреляют по мишени, вероятность попадания первого стрелка 0,6, второго - 0,7. Какова вероятность того, что мишень будет поражена.

6.В хозяйстве два одинаковых стада оленей. В первом стаде 10% животных выбраковано. Во втором - 20%. Наудачу отловлено одно животное. Какова вероятность, что: 1) это животное выбраковано; 2) выбракованное животное из первого стада.

7.На склад поступило 1000 т картофеля. Из них 200 т доставлено хозяйством №1, 360 т хозяйством №2 и 500 т хозяйством №3. Вероятность того, что картофель окажется нестандартным для хозяйства №1 составляет 0,05, для хозяйства №2 - 0,08, для хозяйства №3 - 0,15. Какова вероятность, что наугад взятая партия картофеля окажется стандартной.

Вопросы для самопроверки.

1. Что такое сумма и произведение двух событий, нескольких событий. Приведите примеры.
2. Сформулируйте теорему сложения вероятностей несовместных событий и её следствия.
3. Геометрическая интерпретация действий над событиями.
4. Условная и безусловная вероятность. Зависимые и независимые события. События, независимые в совокупности. Определения. Примеры. Схемы случайного выбора с возвращением и без возвращения.
5. Докажите теорему умножения вероятностей зависимых событий. Сформулируйте её следствия.
6. Вероятность появления хотя бы одного из n событий.
7. Формула полной вероятности и формула гипотез (Байеса).

Глава 3. Повторные независимые испытания.

3.1. Необходимый теоретический минимум.

Схема и формула Бернулли.

Пусть проводится конечное число n последовательных независимых испытаний, в каждом из которых может произойти определенное событие: либо *успех* (событие А), либо противоположное событие-*неудача* (событие А).Такая последовательность испытаний называется *схемой* Бернулли*,* если вероятности положительного исхода в каждом испытании одинаковы.

Определение 3.1. Испытания называются ***повторными независимыми испытаниями,*** если условия их проведения не меняются, вероятность появления события *A* в каждом испытании одна и та же и не зависит от исходов других испытаний.

Обозначим: n-число всех проведенных испытаний;

$p=P(A)$-вероятность появления события *A* в каждом отдельном испытании;

$q=1-p$ – вероятность не появления события *A* в каждом отдельном испытании; m-число появления события *A*(чистота события *A*);

$P_n(m)$(или Pm,n)-вероятность того , что событие *A* появится ровно m раз при n повторных независимых испытаниях.

Вероятность $P_n(m)$ вычисляется по формуле Бернулли:

$$P_n(m)=C_n^m\cdot p^m .q^{n-m} \tag{3.1}$$

Применяют формулу Бернулли обычно при $n \leq 10$, т.к. при больших значениях n получаем произведение очень больших $(n!)$ и очень малых (p^m, q^n) чисел, что плохо с вычислительной точки зрения.

Частные случаи формулы Бернулли:

1. Вероятность того что событие *A* не появится ни разу в n независимых повторных испытаниях $(m = 0)$ находится по формуле:

$$P_n(0)= q^n \tag{3.2}$$

2. Вероятность появления события *А* в каждом из n повторных независимых испытаний $(m = n)$ находится по формуле:

$$P_n(n) = p^n \tag{3.3}$$

Асимптотические формулы.

Как уже было отмечено, при $n>10$ применять формулу Бернулли неудобно, поэтому в этом случае используют *асимптотические*, то есть приближённые формулы, а именно:

I.Формула Пуассона – даёт хорошее приближение для формулы Бернулли при **достаточно большом *n* и малом *p*,** т.е. при **=*n·p*<10:**

$$P_n(m) \approx \frac{a^m}{m!} \cdot e^{-a}. \tag{3.4}$$

• При проведении расчётов можно пользоваться тем, что формула (3.3) затабулирована (см. *таблицу 1* Приложения).

Если задан процесс, проистекающий во времени или пространстве, причём вероятность отдельного события за малый промежуток времени пропорциональна длине этого промежутка, то говорят, что имеет место пуассоновский поток событий, для которого $\boldsymbol{a=a_1 \cdot t_2/t_1}$. Здесь *а* – среднее количество ситуаций за срок, t_1 – общая единица времени, t_2 – единица времени, в которую проводится исследование.

II. Локальная формула Муавра-Лапласа – применяется при ***достаточно большом* n *и не слишком малых* p *и* q,** *т.е. при* **np≥10:**

$$P_n(m) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x), \text{где } \varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}},\ x = \frac{m - np}{\sqrt{npq}} \tag{3.5}$$

• Функция $\varphi(x)$ - затабулирована (см. таблицу 2 приложения), при этом следует учитывать, что:

а) $\varphi(-x) = \varphi(x)$ - функция четная, таблица составлена для $x \ge 0$;

б) $\varphi(x) = 0$ при $x \ge 4$.

III.Интегральная формула Муавра-Лапласа – используется при тех же условиях, что и локальная формула Муавра-Лапласа, но с её помощью находят вероятность того, что при *n* повторных независимых испытаниях событие *А* появится не менее m_1 и не более m_2 раз:

$$P_n(m_1 \le m \le m_2) \approx \Phi(x_2) - \Phi(x_1), \quad \text{где } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} \cdot dt, \qquad (3.6)$$

$$x_1 = \frac{m_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{m_2 - np}{\sqrt{npq}} \quad .$$

- Функция *Φ(x)* называется интегральной функцией Лапласа, ее значения затабулированы (см. таблицу 3 приложения), при пользовании таблицей следует учесть, что:

а) $\Phi(-x)=-\Phi(x)$ – функция нечетная;
б) $\Phi(x)=0{,}5$ при $x \ge 5$.

Наивероятнейшая частота

При проведении серии n независимых повторных испытаний частоты события *A* различны: одни частоты более вероятны, другие – менее. Существует частота, обладающая наибольшей вероятностью (иногда две частоты с одинаковыми наибольшими вероятностями). Такая частота называется *наивероятнейшей частотой* (или *наивероятнейшим числом*), обозначается m_0 и находится из двойного неравенства

$$np - q \le m_0 \le np + p \quad . \qquad (3.7)$$

- 1)m_0 выражается в *целых числах*;
  2)если левая и правая границы неравенства (3.7) – целые числа, то существуют *две наивероятнейшие частоты*;
  3)если границы неравенства (3.7) – дробные, то *наивероятнейшая частота одна* (ей соответствует целое число, лежащее между левой и правой границами неравенства).

3.2. Примеры решения типовых задач.

Задача 3.1. Каждый пятый клиент банка приходит в банк брать проценты с вклада. Сейчас в банке ожидают своей очереди обслуживания шесть человек.

1. Найти вероятность того, что из них будут брать проценты:

а) только два человека; **б)** ни один человек;

в) все шесть человек; **г)** от двух до четырёх человек;

д) хотя бы один человек.

2. Определить наивероятнейшее число очередников, берущих проценты.

Решение. Имеем повторение испытаний по *схеме Бернулли*, где событие *A – человек в очереди будет брать проценты* $\Rightarrow$ *p = P(A)* = 1/5 $\Rightarrow$ *q=1–p*=1 –1/5=4/5. Здесь ***n*=6<10**, поэтому решаем по *формуле Бернулли*.

1. а) Найдём вероятность события *B– 2 очередника из 6 будут брать проценты.* Здесь *m*=2 $\Rightarrow$ по *формуле Бернулли* (3.1) имеем:

$$P(B)=P_6(2)=C_6^2\cdot p^2\cdot q^2=\frac{6!}{2!\cdot(6-2)!}\cdot\left(\frac{1}{5}\right)^2\cdot\left(\frac{4}{5}\right)^2=768/3125\approx 0{,}25.$$

б) *C – ни один из очередников не будет брать проценты* $\Rightarrow$ *m*=0 $\Rightarrow$ по (3.2):

$$P(C)=P_6(0)=q^6=\left(\frac{4}{5}\right)^6=\frac{4096}{15625}\approx 0{,}262.$$

в) *D – все очередники будут брать проценты* $\Rightarrow$ *m=n*=6 $\Rightarrow$ по формуле (3.3):

$$P(D)=P_6(6)=\left(\frac{1}{5}\right)^6=\frac{1}{15625}\approx 0{,}00006$$

г) *E – от двух до четырёх очередников будут брать проценты* $\Rightarrow 2\le m\le 4$ $\Rightarrow$ *m*=2, или *m*=3, *или m*=4 $\Rightarrow$ по теореме *сложения для несовместных событий*: $P(E)=P(2\le m\le 4)=P_6(2)+P_6(3)+P_6(4)$.

Найдём $P_6(i)$ при $i=\overline{2,4}$ по *формуле Бернулли* (3.1):

$P_6(2)=768/3125$; $P_6(3)=256/3125$; $P_6(4)=48/3125$.

В итоге получим:

P(E)=768/3125+256/3125+48/3125=1072/3125≈0,343.

д) *F – хотя бы один очередник берёт проценты* $\Rightarrow$ *m*=1, <u>или</u> *m*=2, <u>или</u> *m*=3, или *m*=4, или *m*=5, или *m*=6. Можно *P(F)* найти аналогично решению предыдущего пункта этой задачи, но гораздо удобнее перейти к *противоположному событию:*

$\overline{F}$ – *ни один очередник не берёт проценты* $\Rightarrow$

$$\overline{F}=D\Rightarrow P(\overline{F})=P(D)=\frac{4096}{15625}\Rightarrow P(F)=1-4096/15625=$$

= 11529/15625 ≈ 0,74.

2. Требуемую наивероятнейшую частоту определим, исходя из неравенства (3.7), где $np - q$=6·(1/5)–4/5=0,4; $np + p$=6·(1/5) + 1/5=1,4. Тогда получаем: $0,4 \le m_0 \le 1,4$. Откуда следует, что $m_0 = 1$.

Задача 3.2. На телефонной станции неправильное соединение происходит с вероятностью 1/200. Найти вероятность того, что среди 200 соединений произойдёт:

1) точно одно неправильное соединение;

2) меньше чем три неправильных соединений;

3) больше чем два неправильных соединений.

Решение.

Имеем повторение испытаний по *схеме Бернулли*, где событие A – *соединение неправильное*; $p=P(A)=1/200$ (по условию). Так как $\boldsymbol{n=200 > 10}$ и $\boldsymbol{a=np=200\cdot1/200=1 < 10}$, то будем использовать *формулу Пуассона* (3.4).

1) Имеем событие B – *точно 1 неправильное соединение* $\Rightarrow m=1 \Rightarrow$

$$P(B)=P_{200}(1)\approx \frac{1^1}{1!}\cdot e^{-1} \approx 0{,}3679 .$$

$P_{200}(1)$ можно определить по таблицам (см. таблицу 1 Приложения), беря значение ячейки, стоящей на пересечении столбца m=1 и строки a=1.

2) C – *меньше трёх неправильных соединений* $\Rightarrow m=0$, или $m=1$, или $m=2 \Rightarrow$ $P(C)=P_{200}(m < 3) = P_{200}(0)+P_{200}(1)+P_{200}(2) \approx 0{,}3679+0{,}3939+0{,}1839 = 0{,}9197$.

3) D – *больше двух неправильных соединений*. Перейдём к *противоположному событию*, так как в этом случае нужно вычислить меньше слагаемых:

$P(D)=P_{200}(m > 2)=1 - P_{200}(m \le 2) = 1 - P_{200}(m < 3) \approx 1 - 0{,}9197=0{,}0803$.

Задача 3.3. Вероятность того, что посетитель, зашедший в кафе, сделает заказ, равна 0,8. Определить вероятность того, что из 100 зашедших в кафе: **1)** 75 человек сделают заказ; **2)** от 65 до 75 сделают заказ; **3)** не менее 75 сделают заказ.

Решение. Имеем повторение по *схеме Бернулли*, где событие

A – посетитель сделал заказ;

$p=P(A)=0{,}8$; $q=1-p=0{,}2$. Так как $\boldsymbol{n=100>10}$ и $\boldsymbol{np=100\cdot 0{,}8=80>10}$, то применяем формулу *Муавра-Лапласа.*

1) B – 75 человек сделают заказ $\Rightarrow$ $m=75$ $\Rightarrow$ по локальной формуле Муавра- Лапласа (3.5):

$$P(B)=P_{100}(75)\approx\frac{1}{\sqrt{100\cdot 0{,}8\cdot 0{,}2}}\cdot\varphi(x)=\frac{\varphi(x)}{4},\text{ где } x=\frac{75-80}{4}=-1{,}25.$$

Используя *чётность* функции $\varphi(x)$ и *таблицу 2* Приложения, находим: $\varphi(x)=\varphi(-1{,}25)=\varphi(1{,}25)=0{,}1826\Rightarrow P(B)\approx 0{,}1826/4=0{,}04565$.

2) C – от 65 до 75 человек сделают заказ $\Rightarrow$ $65\le m\le 75$ $\Rightarrow$ по интегральной теореме Муавра-Лапласа (3.4), где $m_1=65$, $m_2=75$, находим:

$$x_1=\frac{65-80}{4}=-3{,}75;\qquad x_2=\frac{75-80}{4}=-1{,}25\quad\Rightarrow$$

$$\Rightarrow P(C)=P_{100}(65\le m\le 75)\approx \Phi(x_2)-\Phi(x_1)=\Phi(-1{,}25)-\Phi(-3{,}75)=$$
$$=-\Phi(1{,}25)+\Phi(3{,}75)=-0{,}3944+0{,}4999=0{,}1055.$$

Значения функции Лапласа $\Phi(-1{,}25)$ и $\Phi(-3{,}75)$ нашли по *таблице 3* Приложения с учётом *нечётности* функции $\Phi(x)$.

3) D *– не менее 75 человек сделают заказ* $\Rightarrow m\ge 75\Rightarrow 75\le m\le 100\Rightarrow m_1=75$, $m_2=100\Rightarrow x_1=-1{,}25$; $x_2=\dfrac{100-80}{4}=5\Rightarrow$

$\Rightarrow \Phi(x_1)=\Phi(-1{,}25)=-\Phi(1{,}25)=-0{,}3944$; $\Phi(x_2)=\Phi(5)=0{,}5\Rightarrow$ подставим в формулу (3.6):

$$P(D)=P_{100}(m\ge 75)=P_{100}(75\le m\le 100)\approx \Phi(x_2)-\Phi(x_1)=$$
$$=\Phi(5)-\Phi(-1{,}25)==0{,}5+0{,}3944=0{,}8944.$$

Задача 3.4. Монета брошена 10 раз. Найти наивероятнейшую частоту выпадения герба.

Решение: Проводимые испытания являются повторными независимыми, так как вероятность выпадения герба при каждом подбрасывании одна и та же

$= 0{,}5$. Число проведенных испытаний $n = 10$. Найдем наивероятнейшую частоту выпадения герба из условия.

$p(n+1) - 1 \le k_0 \le p(n+1),$

$0{,}5(10+1) - 1 \le k_0 \le 0{,}5(10+1),$

$4{,}5 \le k_0 \le 5{,}5$

$k_0 = 5$

Итак, наивероятнейшая частота выпадения герба при 10 подбрасываний монеты равна 5.

3.3. Задачи для самостоятельного решения.

1. В магазине для студенческого общежития приобрели 4 телевизора. Для каждого из них вероятность выхода из строя в течение гарантийного срока 0,15. Какова вероятность, что: 1) два телевизора выйдут из строя; 2) хотя бы один телевизор выйдет из строя.

2.Склад может обеспечить удобрением пять хозяйств. Вероятность получения заявки складом на удобрение от каждого хозяйства одна и та же, равна 0,8. Найти вероятность того, что: 1) заявку на удобрение дадут три хозяйства; 2) не менее трёх хозяйств; 3) менее трёх хозяйств. Какое число хозяйств, давших заявку, наиболее вероятно.

3.В среднем на 1 га площади посева встречается 20% сорняков. Какова вероятность, что из наугад взятых 400 стеблей окажется 60 стеблей сорняков.

4.В телятнике 100 телят. Вероятность того, что каждый теленок даёт привес равна 0,6. Определить вероятность того, что этот привес дадут 70 телят.

5.В лаборатории из партии семян, имеющих всхожесть 90%, высеяно 600 семян. Найти вероятность того, что число семян, давших всходы, не менее 520 и не более 570, если принять, что каждое посеянное зерно взойдёт с одной и той же вероятностью 0,9.

6.На конном заводе на каждые 10 выращенных жеребят приходится 6 жеребят, отвечающих стандарту. Найти вероятность того, что из 600 выращенных жеребят, от 360 до 380 будут соответственно - стандарту.

7.Среди семян ржи 0,4% семян сорняков. Какова вероятность при случайном отборе из 500 семян обнаружить 5 семян сорняков.

8.В результате изучения работы картофелекопалки установлено, что вероятность механического повреждения каждого клубня равна 0,01. Найти вероятность того, что: 1) среди 800 клубней поврежденных нет; 2) среди 800 клубней поврежденных более двух.

9.Из партии, в которой доля первосортных деталей равна 0,8, отобрано 60 деталей. Определить: 1) вероятность того, что деталей 1-го сорта среди отобранных точно 49, 2) вероятность того, что первосортных деталей среди отобранных не менее 40, но не более 48, 3) наивероятнейшее число первосортных деталей в отобранной партии.

10. Всхожесть семян оценивается вероятностью 0,75. Найти вероятность того, что из 500 посеянных семян не взойдет: 1) точно 130, 2) не более 130. Найти наивероятнейшее число взошедших семян.

---

1. Из пяти яиц в среднем получается 4 живых цыплёнка. Какова вероятность того, что из 4-х яиц получится 3 живых цыплёнка.

2. В семье 6 детей. Считая вероятность рождения мальчика и девочки равными 0,5, опредеблить вероятность того, что в данной семье : 1) четыре мальчика; 2) не более четырёх мальчиков. Найти наивероятнейшее число рождения мальчиков.

3. При сборе грибов вероятность попадания белого гриба и подосиновика равна 0,2. Какова вероятность того, что в наугад взятой корзине из 100 грибов, окажется 30 белых и подосиновиков.

4. В хозяйстве на каждые выращенных 100 огурцов -30 огурцов с весом более 0,3 кг. Найти вероятность того, что из 5000 выращенных огурцов от 1450 до 1560 огурцов будет весом более 0,3 кг.

5. Вероятность нарушения герметичности банки консервов равна 0,0005. Найти вероятность того, что среди 2000 банок: 1) две окажутся с нарушением герметичности; 2) более трёх окажутся с нарушением герметичности.

6. Доля изделий высшего сорта на данном предприятии составляет 31%. Найти наивероятнейшее число изделий высшего сорта в случайно отобранной партии из 75 изделий, вероятность этого числа и вероятность того, что изделий высшего сорта будет более 30.

7. Игральная кость брошена 100 раз. Найти вероятность того, что: 1) 5 очков выпадут 50 раз, 2) 6 очков выпадут не более 50 раз.

Вопросы для самопроверки.

1. Дать определение повторных независимых испытаний.
2. Записать соотношения между вероятностями p и q.
3. Записать и пояснить формулу Бернулли.
4. Записать и пояснить условия применения формулы Пуассона.
5. Сформулировать условия локальной и интегральной теорем Муавра-Лапласа и записать соответствующие формулы.
6. Дать понятие наивероятнейшего числа.

Контрольная работа «Случайные события».

1. Варианты заданий для студентов очной формы обучения (назначаются преподавателем).

№ варианта	Зад. 1.	Зад. 2.	Зад. 3.	Зад. 4.	Зад. 5.	Зад.6	Зад.7
1	1	11	21	31	41	60	61
2	2	12	22	32	42	59	62
3	3	13	23	33	43	58	63
4	4	14	24	34	44	57	64
5	5	15	25	35	45	56	65

6	6	16	26	36	46	55	66
7	7	17	27	37	47	54	67
8	8	18	28	38	48	53	68
9	1	13	25	31	49	52	69
10	7	15	21	33	50	51	70
11	3	17	23	34	41	60	62
12	4	12	26	32	42	59	64
13	5	11	22	37	43	58	66
14	6	16	24	35	44	57	65
15	7	18	25	32	45	56	61
16	2	14	21	33	46	55	63
17	3	12	23	33	47	54	67
18	8	17	28	38	48	53	68
19	9	19	29	39	49	52	70
20	10	20	30	40	50	51	69

2. Варианты заданий для студентов заочной формы обучения (номер варианта соответствует последней цифре шифра зачетной книжки).

№ варианта	**Зад. 1.**	**Зад. 2.**	**Зад. 3.**	**Зад. 4.**	**Зад.5**
1	11	21	31	41	51
2	12	22	32	42	62
3	13	23	33	43	53
4	14	24	34	44	64
5	15	25	35	45	55
6	16	26	36	46	66
7	17	27	37	47	57
8	18	28	38	48	68
9	19	29	39	49	59
0	20	30	40	50	60

Задания для контрольной работы.

1.Для определения всхожести семян было отобрано и высеяно в лабораторных условиях 1000 штук. 980 семян дали нормальные всходы. Найдите относительную частоту недоброкачественных семян.

2. Относительная частота появления неисправной лампы в партии равна 0,02. Отбирается 200 ламп. Сколько ламп окажется неисправными?

3.Относительная частота испорченных яблок (кг) при транспортировке равна 0,03. Транспортируется 150 кг яблок. Сколько кг. яблок окажется испорченными?

4. Из партии калькуляторов наудачу было отобрано 300 штук. Из них оказалось 20 неисправных. Найти относительную частоту годных калькуляторов.

5. При испытании нового антибиотика на кроликах, больных пневмонией, получены следующие результаты . Из больных принимающих антибиотики выжило 65, погибло 25. Какова относительная частота погибших кроликов.

6.В обувном магазине за день было продано 150 пар детской обуви, 100 пар женской обуви, 80 пар мужской обуви. Какова относительная частота продажи женской обуви.

7. При измерении общей длины 100 растений льна, получены следующие данные: 20 растений с длиной от 50 до 80 см, остальные с длиной от 81 до 110 см. Какова относительная частота растений с длиной больше 80 см?

8. При стрельбе из винтовки относительная частота попадания в цель равна 0,75. Найти число попаданий, если всего произведено 140 выстрелов.

10.Относительная частота занятий , пропущенных студентом в некоторый период была 0,04. Сколько было пропусков, если за этот период студент присутствовал 168 раз.

11. В читальном зале имеется 12 учебников по физике, из которых 6 в мягком переплете. Библиотекарь взял три учебника. Найти вероятность того, что

1. три учебника окажутся в мягком переплете (решить двумя способами).
2.* два в мягком и один в твердом переплете.

12. В ящике 40 деталей, из них 5 с дефектом. Последовательно без возврата достают три детали. Какова вероятность того, что:

1. они без дефекта (решить двумя способами;

2*.две без дефекта и одна с дефектом.

13. В компьютерном классе 12 компьютеров, из них исправны 9. Какова вероятность, что 1.три студента вошедшие в класс выберут исправные компьютеры (два способа); 2*. два выберут исправные компьютеры, один неисправный?

14. В некотором районе города находится 10 магазинов 6 продовольственных и 4 непродовольственных. Случайным образом для приватизации были отобраны три магазина. Найти вероятность того, что:

1.все отобранные магазины окажутся непродовольственными (два способа);

2*. два продовольственных и один непродовольственный.

15. В ящике лежат 20 теннисных мячей. Из них 12 новых и 8 игравших. Наугад извлекают три мяча для игры. Найти вероятность того, что: 1.три мяча будут игравшими (два способа); 2*. два игравших и один новый?

16. В группе 25 студентов. Из них 5 отличников. Преподаватель наугад вызывает четырех студентов. Найти вероятность того, что: 1. эти четыре студента будут отличниками (два способа); 2*. только два отличника.

17. В ящике 22 детали. Среди них четыре оказались с браком. 1.Что более вероятно: извлечение из ящика двух деталей с браком или появление трёх очков при бросании игральной кости. 2. Найти вероятность того, что из трех наугад извлеченных деталей две с браком, одна без (два способа)?

18. В урне 20 шаров, различающихся только цветом: 13 красных, остальные зелёные. 1. Что более вероятно: извлечение из урны двух зеленых шаров или появление герба при однократном бросании монеты. 2. Найти вероятность того, что из четырех наугад извлеченных шаров два зеленого цвета, два красных (два способа)?

19. Из колоды в 36 карт наугад выбрали 5 карт. Какова вероятность того, что: 1. все три пиковой масти; 2. два туза, три короля (два способа); 3. две дамы, два короля и один туз?

20. В ящике 20 фруктов. Из них 2 апельсина, 12 яблок и 6 груш. Какова вероятность того, что из пяти наугад вытащенных фруктов окажется:

1. пять груш; 2. три яблока и две груши (два способа); 3. один апельсин, два яблока и две груши.

21. В двух отсеках находится посевной материал. Семена первого отсека имеют всхожесть 80% , второго – 85%. Отбирается по одному зерну из каждого отсека. Найдите вероятность того, что: а) оба зерна не взойдут; б) хотя бы одно взойдёт.

22. На ферме работают два транспортера для раздачи кормов. Вероятность выхода из строя каждого из них соответственно равна 0,25 и 0,2. Какова вероятность, что: 1) оба транспортера выйдут из строя; 2) произойдет поломка хотя бы одного из транспортеров?

23. Три ветеринара сдают экзамен комиссии на повышение квалификации. Вероятность того, что первый из них сдаст экзамен равна 0,9,второй – 0,8, третий– 0,7.Найти вероятность того, что экзамен будет сдан:

1) только одним ветеринаром; 2) всеми тремя; в) хотя бы одним.

24. Из трех орудий произвели залп по цели. Вероятность попадания в цель при одном выстреле из первого орудия равна 0,7; для второго орудия – 0,9; для третьего – 0,85. Найти вероятность того, что 1) все три орудия попадут в цель; 2) только один снаряд попадет в цель; 3) ни один не попадет в цель; 3) хотя бы один попадет в цель.

25. Эксперт оценивает качественный уровень трех видов изделий по потребительским признакам. Вероятность того, что изделию первого вида будет присвоен знак качества, равна 0,9; изделию второго вида – 0,85; изделию третьего вида – 0,8. Найти вероятность того, что знак качества будет присвоен: 1) всем трем изделиям: 2) только одному изделию; 3) по крайней мере одному изделию.

26. Устройство состоит из трех элементов, работающих независимо. Вероятность поломки первого, второго и третьего элементов соответственно равны 0,2; 0,12 и 0,3. Найти вероятность того, что: 1) все элементы будут работать безотказно; 2) все окажутся неисправными; 3) поломается хотя бы один элемент.

27.Для сигнализации об аварии установлены три независимо работающих устройства. Вероятность того, что при аварии сработает первое устройство, равна 0,92; второе – 0,98 и третье – 0,9. Найти вероятность того, что: 1) при аварии сработают все три устройства; 2) ни одно не сработает; 3) сработает хотя бы одно устройство.

28.В поле работают три бригады. Вероятность выполнить норму для каждой бригады соответственно равна 0,8; 0,9; 0,6. Определить вероятность, что: 1) три бригады норму не выполнят; 2) выполнит только одна бригада; 3) хотя бы одна выполнит.

29. Хозяйством послана машина за удобрением на три базы. Вероятность наличия нужного удобрения на 1 базе – 0,9; на 11 – 0,8; на 111 – 0,6. Найти вероятность того, что: 1) ни на одной базе не окажется нужного удобрения;
2) только на одной базе окажется нужное удобрение; 3) хотя бы на одной базе не окажется нужного удобрения.

30. В студии телевидения четыре телевизионные камеры. Для каждой камеры вероятность того, что она выключена в данный момент равна 0,3.Найти вероятность того, что в данный момент времени включены:1) все четыре камеры; 2) хотя бы одна камера.

31. В трёх ящиках находятся детали. В первом 90% деталей без брака, во втором 75% и в третьем – 80%. Из наугад выбранного ящика берут одну деталь. Какова вероятность, что: 1) деталь без брака; 2) небракованная деталь из 2 ящика?

32. Перед туристами на схеме две дороги, ведущие в определенный населенный пункт. Если идти по первой дороге, то вероятность попасть в нужный пункт без опоздания равна 0,8, а если по второй, то соответствующая

вероятность равна 0,6. Туристы наудачу выбирают одну из дорог и прибывают без опоздания. Какова вероятность, что: 1) туристы прибудут без опоздания; 2) туристами была выбрана первая дорога.

33. Для проверки качества саженцев кедров было установлено, что они могут быть разбиты на три группы. К 1 группе принадлежат 86% саженцев; ко 2-й группе - 10%; к 3-й группе – 4%. Вероятность того, что саженец приживется: для 1 группы равно 0,7; для 2 группы – 0,4; для 3 группы – 0,8. Определить вероятность того, что: 1) взятый наугад саженец приживется; 2) приживется саженец из 2 группы.

34. Рабочий работает на двух станках, на которых обрабатываются однотипные детали . Вероятность, что деталь будет обрабатываться на первом станке равна 65%, на втором 35%, при этом вероятность брака для первого – 0,09, для второго – 0,05. Определить вероятность того, что 1) наугад взятая деталь будет бракованной; 2) бракованной будет деталь с первого станка?

35. Для посева пшеницы был заготовлен посевной материал, содержащии небольшое количество примесей зерен 2 и 3 сортов. 90% всего зерна было 1 сорта, 7% - 2-ого, 3% - 3-его. Вероятность того, что из зерна пшеницы вырастет полновесный колос для семян 1 сорта равна 0,9; 2 сорта – 0,8; 3 сорта – 0,6. Определить вероятность того что: 1. из взятого наудачу зерна вырастет полновесный колос; 2). этот колос вырастет из зерна 1 сорта?

36.В сборочный цех предприятия поступают детали с трех автоматов. Первый автомат дает 3% брака, второй – 1%, третий – 2%. При этом с каждого автомата в цех поступило соответственно 500, 200 и 300 изделии. Определите вероятность того что: 1) на сборку попала бракованная деталь; 2) бракованная деталь из третьего цеха.

37. На вспашке работают 6 тракторов, 2 новых и 4 после капитального ремонта. Вероятность поломки для нового трактора равна 0,1, а для трактора после ремонта – 0,3. Какова вероятность, что 1) к концу смены один из тракторов выйдет из строя; 2) из строя выйдет новый трактор?

38. Детали поступают на склад из цеха №1 и цеха №2, причем, из 1-ого цеха в 2 раза больше, чем из 2-ого. Вероятность того, что деталь из первого цеха удовлетворяет стандарту, равна 0,95, а соответствующая вероятность для второго цеха равна 0,9. Определить вероятность, что 1) взятая деталь удовлетворяет стандарту; 2) удовлетворяет стандарту деталь из первого цеха.

39. Птицефабрика имеет три конвейера по сбору яиц. Первый и второй конвейеры движутся с одинаковой скоростью, а третий в два раза медленнее. Количество собранных яиц пропорционально скорости движения. Первый конвейер дает 20% боя, второй – 15%, третий – 10%. Все яйца поступают на один пункт сортировки. Какова вероятность, что: 1) наугад взятое яйцо при сортировке имеет бой; 2) разбитое яйцо с третьего конвейера.

40. В комплекте 20 экзаменационных билетов. Студент Иванов подготовил к экзамену 15 из них. По жребию ему досталась вторая очередь. Какова вероятность, что: 1) Иванов получит подготовленный билет, если неизвестно какой билет достался первому студенту; 2) вытянут подготовленный билет, при условии, что первому студенту достался билет, который Иванов не знает?

41. Вероятность попадания в цель при одном выстреле равна 0,8. Найти вероятность того, что из пяти выстрелов попаданий будет: а) четыре; б) менее двух.

42. Прибор состоит из шести узлов. Вероятность безотказной работы для каждого узла равна 0,7. Найти вероятность того, что из шести узлов выйдут из строя: а) два; б) менее двух узлов.

43. Студенты высадили семь деревьев. Вероятность, что каждое дерево приживется, равна 0,6. Найти вероятность того, что из семи деревьев приживутся: а) пять; б) не менее шести деревьев.

44. В поле работает восемь тракторов. Вероятность бесперебойной работы каждого трактора за смену постоянна и равна 0,8. Найти вероятность того, что за смену из 8 тракторов поломаются: а) четыре; б) не более двух.

45. Всхожесть семян некоторой культуры составляет 90%. Найти вероятность того, что на опытном участке из шести посеянных семян взойдут: а) четыре; б) более четырех. Найти наивероятнейшее число взошедших семян.

46. Вероятность выигрыша в лотерее на 1 билет равна 0,6. Куплено 13 билетов. Найти вероятность того, что число выигрышных билетов окажется: а) ровно 3; б) не более 3. Найти наивероятнейшее число выигрышных билетов.

47. Вероятность попадания в цель при одном выстреле равна 0,7. Найти вероятность того, что из пяти выстрелов попаданий будет: а) четыре; б) менее двух. Найти наивероятнейшее число попадании.

48. Из каждых 100 шкурок норки 25% высшего сорта. Проверяется на качество 8 шкурок. Какова вероятность, что: 1) 6 из них высшего сорта; 2) не менее четырех высшего сорт? Найти наивероятнейшее число шкурок высшего сорта из 8 отобранных.

49.Монета подброшена 7 раз. Найти вероятность того, что: 1) герб при этом появился 5 раз; 2) не менее 6 раз. Сколько раз необходимо подбросить монету, чтобы наивероятнейшее число выпадений герба было равно 10 раз?

50. Проводится игровая лотерея, в которой всего 100 билетов. Выигрывают билеты с номерами кратными 3, или 5. Найти вероятность того, что из взятых четырех билетов: 1) выиграют два; 2) хотя бы один выигрышный. Найти наивероятнейшее число выигрышных билетов среди четырех отобранных.

51. Вероятность появления бракованной детали равна 0,006. Найти вероятность того, что из 500 случайно отобранных деталей бракованных окажется: а) три; б) не более двух.

52. Книга издана тиражом 1000 экземпляров. Вероятность того, что в книге имеется дефект брошюровки, равна 0,002. Найти вероятность того, что тираж содержит неправильно сброшюрованных книг: а) ровно пять; б) не менее двух.

53. Система содержит 100 элементов. Вероятность, с которой может быть неисправный элемент системы равна 0,04. Определить вероятность того, что

за определенный срок неисправны будут: а) четыре элемента; б) более одного элемента.

54. Известно, что на поле у 2% кустов картофеля, стебли поражены фитофторой. Найти вероятность того, что из 300 кустов картофеля фитофторой будут поражены: а) четыре куста: б) менее двух кустов.

55. Вероятность выживания бактерий после радиоактивного облучения равна 0,005. Найти вероятность того, что после облучения из 400 бактерий останется: а) пять бактерий; б) более двух бактерий.

56. Вероятность сбоя в работе телефонной станции при каждом вызове равна 0,04. Определить вероятность того, что среди 900 поступивших вызовов имеется: а) 8 сбоев; б) не более 4 сбоев?

57. Вероятность соблюдения расписания движения поездов на перегоне в течение недели равна 0,98. Найти вероятность того, что число нарушении графика движения поездов за 50 недель будет: а) ровно 2; б) больше двух?

58.Среди семян пшеницы 0,06% сорняков. Какова вероятность, что при отборе 1000 семян обнаружат: 1) три сорняка; 2) не менее трех?

59. При хранении посадочного материала среди семян пшеницы 0,1% заражаются вредителями. Определите вероятность того, что при выборке 500 семян: 1) окажется зараженными ровно пять; 2) хотя бы одно будет зараженное.

60. Телефонная станция обслуживает 800 абонентов. Вероятность того, что любой абонент не позвонит на телефонный компьютер в течение часа равна 0,99. Какова вероятность, что в течение часа позвонят: 1) 5 абонентов; 2) хотя бы один абонент?

61. На склад магазина поступают изделия, из которых 80% оказываются высшего сорта. Найти вероятность того, что из 100 взятых наудачу изделии высшего сорта окажется: а) ровно 85 ; б) не менее 75 и не более 85?

62. Принимая одинаково вероятным рождение мальчика и девочки, найти вероятность того, что среди 450 новорожденных будет: а) 230 мальчиков; б) не менее 230 и не более 400 мальчиков?

63. Приживаемость деревьев данной породы составляет 90 %. Найти вероятность того, что из 300 саженцев приживутся: а) 200 деревьев; б) от 200 до 270 деревьев?

64. В инкубаторе содержится 500 птенцов. Вероятность выживания каждого до трех дней равна 0,8. Какова вероятность, что этот срок выживут: 1) 300 птенцов; 2) от 400 до 450 птенцов?

65. Завод выпускает приборы, среди которых в среднем 90% без дефекта. Найти вероятность того, что в партии из 400 приборов дефектных окажется: а) 100 приборов; б) не более 50 и не более 100?

66. Вероятность того, что деталь не прошла проверку ОТК, равна 0,2. Найти вероятность того, что среди 400 деталей непроверенных окажется: а) 70 деталей; б) от 70 до 100 деталей?

67. В телятнике содержится 120 телят. Вероятность того, что каждый теленок к определенному сроку даст необходимый привес равна 0,6. Определить вероятность того, что привес дадут: 1) 100 телят; 2) от 70 до 80 телят.

68. Средний процент нарушений работы кинескопа телевизора в течение гарантийного срока 10%. Вычислить вероятность того, что из 40 телевизоров гарантийный срок выдержат: а) 30 телевизоров; б) не менее 30 и не более 36 телевизоров?

69. Принимая одинаково вероятным рождение мальчика и девочки, найти вероятность того, что среди 470 новорожденных будет: а) 230 мальчиков; б) не менее 230 и не более 400 мальчиков?

70. Игральную кость бросают 700 раз. Какова вероятность того, что число очков, кратное трем, выпадет: а) ровно 274; б) не меньше 260 и не больше 274 раз?

РАЗДЕЛ 2. СЛУЧАЙНАЯ ВЕЛИЧИНА.

Глава 4. Дискретная случайная величина. Числовые характеристики дискретной случайной величины.

4.1. Необходимый теоретический минимум.

Виды случайной величины. Закон распределения. Биномиальное распределение.

Определение 4.1. Случайной величиной называется величина, которая в результате опыта может принять то или иное значение, заранее до опыта неизвестное.

- Те значения, которые случайная величина может принять в результате некоторого опыта, образуют *множество её возможных значений*.
- Случайные величины обозначают заглавными греческими буквами $(X, Y, Z \ldots)$, а их соответствующие значения – строчными буквами (x_1 $x_2,, \ldots, x_n$; y_1 , y_2,,; y_n ; z_1, z_2 , z_n).

Виды случайной величины:

1) ***дискретная*** -принимает отдельные изолированные значения (число студентов на лекции, число зерен в колосе и т.д.);
2) ***непрерывная*** -принимает значения из некоторого бесконечного или конечного интервала (время, рост и т.д.)

Примеры.

1. X – число выпавших очков при подбрасывании игральной кости (дискретная);
2. Y – глубина задела семени при посеве (непрерывная);
3. Z– число погрешностей при измерений (дискретная);
4. H – масса животного (непрерывная);

Чтобы задать случайную величину, нужно задать ее закон распределения.

Определение 4.2. Законом распределения случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.

- Закон распределения случайной величины может быть задан *таблично, графически, аналитически* (в виде формулы $P(X=x_i)=f(x_i)$).

Дискретная случайная величина (ДСВ) обычно задаётся с помощью *таблицы*, в первой строке которой содержится перечень её возможных значений, а во второй – соответствующие вероятности:

X	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

Заметим, что событие, заключающееся в том, что случайная величина примет одно из своих возможных значений, является достоверным, поэтому

$$\sum_{i=1}^{n(\infty)} p_i = 1.$$

Для *графического* представления дискретной величины следует построить точки $M_1(x_1,\ p_1),\ M_1(x_2,\ p_2),\ \ldots,\ M_n(x_n,\ p_n)$ в декартовой системе координат XOP (по горизонтальной оси откладываются значения случайной величины, по вертикальной – её вероятности). Затем эти точки соединяют отрезками прямых. Полученную фигуру называют *многоугольником* или *полигоном распределения* (рис.4.1).

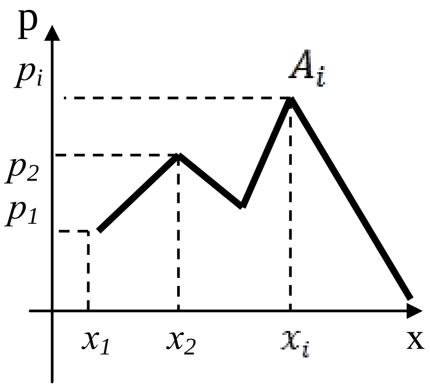


Рисунок 4.1.

Примером аналитического задания ДСВ является *биномиальный закон распределения*.

Определение 4.3. Биномиальным законом распределения дискретной случайной величины называется закон распределения *числа появления события A в* n *независимых повторных испытаниях (*случайная величина *X).*

Вероятность возможных значений $X=m$ вычисляется *по формуле Бернулли:*

$P_n(m)=C_n^m \cdot p^m \cdot q^{n-m}$.

- Если число испытаний велико ($n\to\infty$), а вероятность появления события в одном испытании очень мала ($p\to 0$), то для вычисления вероятности того, что $X=m$, используют приближённую *формулу Пуассона*:

$$P_n(m)=\frac{a^m}{m!}e^{-a}, \quad a=n\cdot p$$

В этом случае говорят, что случайная величина распределена по *закону Пуассона.*

Пример. Вероятность допустить ошибку при измерении равна 0,1. Производится два измерения. Составить закон распределения возможного числа ошибок.

Решение. Случайная величина X-число допустимых при измерении ошибок, подчиняется биномиальному закону.

По условию задачи $n=2$; $p=0,1$; $q=1-p=0,9$

При двух измерениях число допустимых ошибок может принимать значения: 0,1,2. Ряд распределения имеет вид:

x	0	1	2
p	p_1	p_2	p_3

Найдем соответствующие вероятности по формуле Бернулли

$P_n(m)=C_n^m \cdot p^m \cdot q^{n-m}$ где $C_n^m=\frac{n!}{m!\cdot(n-m)!}$ число сочетаний.

$p_1=P_2(0)=C_2^0\cdot 0,1^0\cdot 0,9^2=1\cdot 1\cdot 0,9^2=0,81$

$p_2=P_2(1)=C_2^1\cdot 0,1^1\cdot 0,9^1=2\cdot 1\cdot 0,9^2=0,18$

$p_3=P_2(2)=C_2^2\cdot 0,1^2\cdot 0,9^0=1\cdot 0,1^2\cdot 1 \;=0,01$

x	0	1	2
p	0,81	0,18	0,01

проверка:

$$\sum_{i=1}^{n} P_i = 1$$

Числовые характеристики дискретной случайной величины.

Случайная величина считается заданной, если известен её закон распределения. Закон распределения содержит всю необходимую информацию о случайной величине. Но не всегда на практике построить такой закон. Для определения изменчивости и сравнения различных величин используются такие характеристики: математическое ожидание *M(x*), дисперсия *D(x),* среднее квадратическое отклонение σ(x).

Дискретная случайная величина характеризуется рядом *числовых характеристик*, основные из которых приведены ниже.

Математическое ожидание.

Определение 4.4. Математическим ожиданием дискретной случайной величины называется сумма произведений всех возможных значений случайной величины на соответствующую вероятность:

$$M(X)=\sum_{i=1}^{n} x_i \cdot p_i = x_1p_1 + x_2p_2 + \ldots + x_np_n \quad . \qquad (4.1)$$

Замечание. Из определения следует, что математическое ожидание дискретной случайной величины, есть неслучайная величина (переменная), а постоянное величина (число).

Пример . Задан закон распределения потерь урожая при уборке риса *X* (%). Найти математическое ожидание случайной величины *X*.

X	2	3	5
p	0,3	0,1	0,6

Решение. Искомое математическое ожидание равно сумме произведении всех возможных значении случайной величины на их вероятности. По формуле (4.1) имеем: $M(X) = 2\cdot 0{,}3 + 3\cdot 0{,}1 + 5\cdot 0{,}6 = 3{,}9$

Как видно из примера, математическое ожидание 3,9 является приблизительно средним значением случайной величины, находится в середине интервала от 2 до 5. То, что оно является средним значением, в этом и есть вероятностный смысл математического ожидания.

- Математическое ожидание является *характеристикой среднего значения* случайной величины.

Математическое ожидание обладает следующими *свойствами:*

1.*M(C)=C* (*C* – постоянная);

2. *M(C·X)=C·M(X);*

3.$M(X_1+X_2+\cdots+X_n)=M(X_1)+M(X_2)+\cdots+M(X_n)$;

4.$M(X+C)=M(X)+C$;

5. $M(X\text{-}C)=M(X)\text{-}C \Rightarrow M(X)=M(X\text{-}C)+C$;

6.$M(X_1 \cdot X_2 \cdot \ldots X_n)=M(X_1)\cdot M(X_2)\cdot \ldots M\cdot(X_n)$, где $X_1, X_2, \ldots, X_n$ – независимые случайные величины.

7.$M[X\text{-}M(X)]=0$, где $[X\text{-}M(X)]$ – отклонение случайной величины от её математического ожидания.

Мода и медиана.

Определение 4.5. Модой M_0 дискретной случайной величины называется её наиболее вероятное значение (рис.4.2).

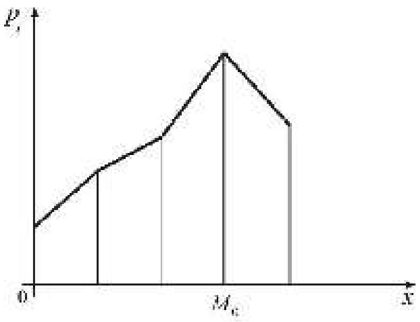


Рисунок 4.2

Определение 4.6. Медианой $M_Д$ (или M_e) случайной величины X называется такое её значение, относительно которого равновероятно получение большего или меньшего значения случайной величины, т.е. $P(X<M_Д)=P(X>M_Д)$.

Математическое ожидание, мода и медиана – это *характеристики положения* случайной величины.

Дисперсия.

Определение 4.7. Дисперсией дискретной случайной величины называется математическое ожидание квадрата отклонения случайной величины от её математического ожидания:

$$D(X)=M[X-M(X)]^2 . \qquad (4.2)$$

Дисперсию также можно вычислять по следующей формуле (упрощенной):

$$D(X)=M(X^2)-[M(X)]^2 . \qquad (4.3)$$

Дисперсия является характеристикой *рассеивания* возможных значений случайной величины вокруг математического ожидания.

Свойства дисперсии:

1.$D(C)=0$ (C – постоянная);

2. $D(C\cdot X)=C^2\cdot D(X)$;

3.$D(X_1+X_2+\cdots+X_n)=D(X_1)+D(X_2)+\cdots+D(X_n)$ ($X_1;X_2;\cdots X_n$– независимые);

4. $D(X-Y)=D(X+Y)$;

5. $D(X\pm C)=D(X)$.

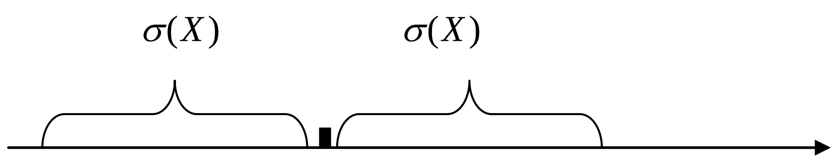
Дисперсия имеет размерность квадрата случайной величины, но существует необходимость иметь показатель рассеяния случайной величины той же размерности, что и размерность случайной величины. Для этого ввели характеристику, которая называется средним квадратическим отклонением.

Определение 4.8. Средним квадратическим отклонением случайной величины X называется квадратный корень из дисперсии.

$$\sigma(X)=\sqrt{D(X)} \tag{4.4}$$

Среднее квадратическое отклонение имеет размерность случайной величины, является показателем линейного отклонения значений случайной величины от математического ожидания.

Например, если X выражается в линейных метрах, то $\sigma(X)$ будет выражаться также в линейных метрах, а $D(X)$ - в квадратных метрах.



Определение 4.9. Коэффициентом вариации называется отношение среднего квадратического отклонения случайной величины к её математическому ожиданию, вычисленное в процентах:

$$v(X)=\frac{\sigma(X)}{M(X)}\cdot 100\% \ . \tag{4.5}$$

- Дисперсия, среднее квадратическое отклонение, коэффициент вариации являются характеристиками рассеивания случайной величины.

- Если случайная величина имеет биномиальное распределение, то её числовые характеристики можно определить по формулам:

$$M(X)=np;\ D(X)=npq;\ \sigma(X)=\sqrt{npq}$$

Для случайной величины, распределённой по закону Пуассона, имеем:

$$M(X)=np;\ D(X)=np;\ \sigma(X)=\sqrt{np}\ .$$

4.2. Примеры решения типовых задач.

Задача 4.1. Доля плодов, зараженных болезнью в скрытой форме, составляет 30%. Случайным образом отбирается 4 плода. Требуется:

1) Составить закон распределения дискретной случайной величины *X*, числа зараженных плодов среди отобранных;
2) Построить многоугольник распределения;
3) Найти числовые характеристики: *M(X)*, M_0, *D(X)*, $\sigma(X)$

Решение.

1) Возможные значения случайной величины *X*, равной числу зараженных плодов: 0,1,2,3,4. Соответствующие вероятности отдельных значении вычислим по формуле Бернулли: $P_n(m)=\frac{n!}{m!(n-m)!}\cdot p^m\cdot q^{n-m}$, где p=0,3;

q=1-p= 0,7; n=4.

$P(X=0)=P_4(0)=q^4=0{,}7^4=0{,}2401$ - вероятность того, что среди 4 – х отобранных плодов ни одного зараженного;

$P(X=1)=P_4(1)=\frac{4!}{1!3!}\cdot 0{,}3^1\cdot 0{,}7^3=0{,}4116$- вероятность того, что среди отобранных один плод будет зараженным;

$P(X=2)=P_4(2)=\frac{4!}{2!2!}\cdot 0{,}3^2\cdot 0{,}7^2=0{,}2646$ - вероятность того, что среди отобранных плодов два окажутся зараженными;

$P(X=3)=P_4(3)=\frac{4!}{3!1!}\cdot 0{,}3^3\cdot 0{,}7^1=0{,}0756$- вероятность того, что среди отобранных плодов три окажутся зараженными;

$P\,(X=4)=P_4(4)=p^4=0{,}3^4=0{,}0081$ - вероятность того, что среди отобранных четырех плодов два окажутся зараженными.

Так как число значений случайной величины X конечно и невелико, то закон распределения дискретной случайной величины запишем в виде таблицы:

X	*0*	*1*	*2*	*3*	*4*
P	*0,2401*	*0,4116*	*0,2646*	*0,0756*	*0,0081*

Выполним проверку - сумма полученных вероятностей должна равняться 1.

Проверка: P = 0,2401 + 0,4116 + 0,2646 + 0,0756 + 0,0081 = 1

2) Для построения многоугольника распределения или полигона распределения в прямоугольной системе координат по горизонтальной оси откладываем значения случайной величины, а по вертикальной оси значения вероятности. По осям берем разный масштаб: по оси вероятностей – 0,1; по оси значений случайной величины – 1.

Полученные точки соединяем прямыми линиями, полученная ломаная является

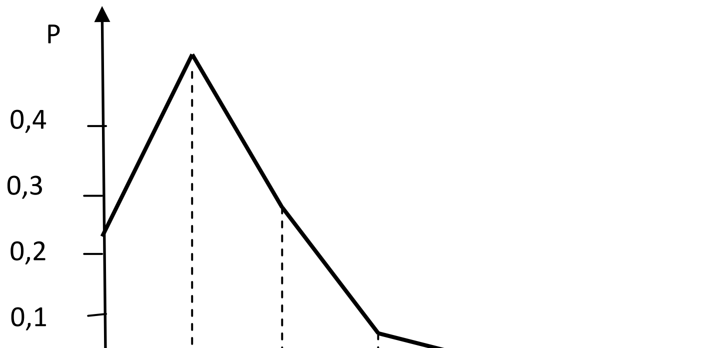


Рисунок 4.3.

многоугольником распределения (рис. 4.3).

3) Найдем числовые характеристики, применяя формулы вычисления характеристик для биномиального распределения:

$$M(X)=np=4\cdot 0{,}7=2{,}8$$

$$D(X)=4\cdot 0{,}7\cdot 0{,}3=0{,}84$$

$$\sigma(X)=\sqrt{0{,}84}\approx 0{,}91$$

Задача 4.2. Дан ряд распределения случайной величины:

X	1	2	5	8	10
p	0,1	0,2	0,4	0,25	0,05

Найти математическое ожидание, моду, медиану, дисперсию, среднее квадратическое отклонение этой случайной величины. Построить многоугольник распределения.

Решение. 1) Найдём *математическое ожидание* по формуле (4.1):

$$M(X)=\sum_{i=1}^{5} p_i = 1\cdot 0{,}1+2\cdot 0{,}2+5\cdot 0{,}4+8\cdot 0{,}25+10\cdot 0{,}05=5 \Rightarrow \boldsymbol{M(X)=5.}$$

2) *Мода* – это значение случайной величины, которому отвечает наибольшая вероятность $\Rightarrow$ $\boldsymbol{M_0=5}$, так как $P(X=5)=0{,}4$ – наибольшее значение вероятности (в заданном ряде распределения).

3) *Дисперсию* вычислим двумя способами:

Первый способ – по определению дисперсии (формула (4.2)). Предварительно составим ряд распределения случайной величины $[X-M(X)]^2$ (при этом вероятности значений случайной величины X остаются неизменными).

$\boldsymbol{[X\text{-}M(X)]^2}$	$(1\text{-}5)^2=16$	$(2\text{-}5)^2=9$	$(5\text{-}5)^2=0$	$(8\text{-}5)^2=9$	$(10\text{-}5)^2=25$
P	0,1	0,2	0,4	0,25	0,05

Тогда $D(X)=M[X-M(X)]^2=16\cdot 0{,}1+6\cdot 0{,}2+0\cdot 0{,}4+9\cdot 0{,}25+25\cdot 0{,}05=6{,}9.$

Второй способ – по вычислительной формуле (4.3). Здесь нужно составить закон распределения случайной величины X^2, а затем найти математическое ожидание случайной величины X^2:

$\boldsymbol{X^2}$	$1^2=1$	$2^2=4$	$5^2=25$	$8^2=64$	$10^2=100$
p	0,1	0,2	0,4	0,25	0,05

$M(X^2) = 1{\cdot}0{,}1 + 4{\cdot}0{,}2 + 25{\cdot}0{,}4 + 64{\cdot}0{,}25 + 100{\cdot}0{,}05 = 31{,}9$

$\Rightarrow D(X)=31{,}9 - 5^2 = 6{,}9.$

При нахождении дисперсии по формуле (4.3) вычислительных действий производится меньше, поэтому меньше вероятность арифметической ошибки, т.е. определение дисперсии по формуле (4.3) предпочтительнее, особенно когда $M(X)$ не является целым числом.

4)Среднее квадратическое отклонение случайной величины X найдём по формуле (4.4): $\sigma = \sqrt{D(X)} = \sqrt{6{,}9} \approx 2{,}63$.

5) Построим многоугольник распределения (рис.4.4) и отметим на графике найденные числовые характеристики.

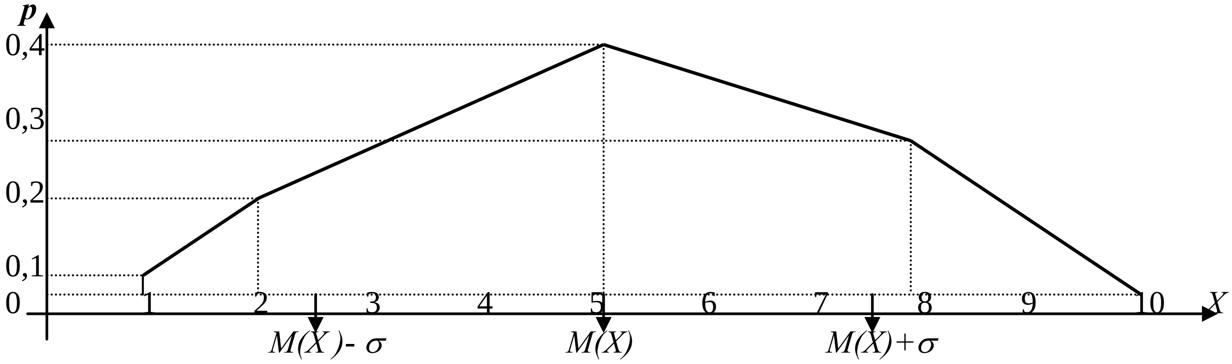


Рисунок 4.4.

Получили, что $M(X)=M_0=M_e=5$, т.е. величины математического ожидания, моды, медианы совпали. Такое распределение называют *одномодальным, симметрическим.*

Вычисления $M(X)$, $D(X)$ удобно выполнять с помощью следующей таблицы подсчётов:

№ п/п	***x***	***p***	$x_i{\cdot}p_i$	x_i^2	$x_i^2{\cdot}p_i$
1	1	0,1	$1{\cdot}0{,}1=0{,}1$	$1^2=1$	$1{\cdot}0{,}1=0{,}1$
2	2	0,2	$2{\cdot}0{,}2=0{,}4$	$2^2=4$	$4{\cdot}0{,}2=0{,}8$
3	5	0,4	2,0	25	10,0
4	8	0,25	2,0	64	16,0
5	10	0,05	0,5	100	5,0

Σ	1	5	31,9

сумма $M(X)$ $M(X^2)$

Задача 4.3. Две независимые случайные величины заданы таблицами распределения

x	-5	10	15
p	0,2	0,7	0,1

y	5	10
p	0,6	0,4

Найти математическое ожидание случайной величины Z=X-2У двумя способами: 1) составив ряд распределения случайной величины Z; 2) использую свойства математического ожидания.

Решение. 1) Составим ряд распределения случайной величины X-2У.Для этого сначала составим ряд распределения величины -2У

-2y	-10	-20
p	0,6	0,4

Составим ряд распределения случайной величины X-2У

x-2y	-15	-25	0	-10	5	-5
p	0,12	0,08	0,42	0,28	0,06	0,04

Выполним проверку: $\sum_{i=1}^{n} \mathrm{P}_i = 0{,}12 + 0{,}08 + 0{,}42 + 0{,}28 + 0{,}06 + 0{,}04 = 1$

Математическое ожидание

$\mathrm{M}(\mathrm{x}-2\mathrm{y}) = -15 \cdot 0{,}12 - 25 \cdot 0{,}08 + 0 \cdot 0{,}42 - 10 \cdot 0{,}28 + 5 \cdot 0{,}06 - 5 \cdot 0{,}04 = -6{,}5$ $\mathrm{M}(\mathrm{x}-2\mathrm{y}) = -6{,}5$

2) Используем свойства математического ожидания

$\mathrm{M}(\mathrm{x}-2\mathrm{y}) = \mathrm{M}(\mathrm{x}) + \mathrm{M}(-2\mathrm{y}) = \mathrm{M}(\mathrm{x}) - 2\mathrm{M}(\mathrm{y})$

Найдем математическое ожидание каждой случайной величины

M(x)=-5·0,2+10·0,7+15·0,1=7,5 M(y)=5·0,6+10·0,4=7

$\mathrm{M}(\mathrm{x}-2\mathrm{y}) = 7{,}5 - 2 \cdot 7 = -6{,}5$. Результаты вычислений одинаковы.

4.3. Задачи для самостоятельного решения.

1. Возможные значения случайной величины X таковы: x_1=2, x_2=5.x_3=8. Известно, *что P(x=2) = 0,4; P(x=5) = 0.15*. Найдите *P(x=8)* -?

2. Случайная величина *X* имеет распределение:

x	*1*	*2*	*3*	*4*	*5*
P	*1/4*	*1/8*	*1/4*	*1/8*	*1/4*

Найти: 1) $P(x\leq3)$; 2) $P(2\leq x<5)$; 3) $P(1\leq x<4)$; 4) $P(x>2)$.

3. Задан закон распределения дискретной случайной величины. Найти: 1) *M(x)*; 2) *Д(x),* используя определение дисперсии и формулу для нахождения дисперсии; 3) *σ(x); 4)υ(x)*

x	*5*	*8*	*10*	*15*
P	*0.1*	*0.2*	*0.5*	*0.2*

4.Число вызовов, поступающих в ветеринарные клиники двух районов в течение недели, имеют соответственно законы распределения:

x	*0*	*1*	*2*
P	*0,8*	*0,15*	*0,05*

y	*0*	*1*	*2*
P	*0,82*	*0,1*	*0,08*

В какую ветеринарную клинику поступает больше вызовов?

5.Вероятность того, что клубень поврежден при переработке равно 0,15. Случайным образом отбирается 7 клубней.

1) Найдите закон распределения случайной величины *X*-числа поврежденных клубней среди отобранных;

2) постройте многоугольник распределения;

3) найдите вероятность события:

а) А- в выборке не менее одного поврежденного клубня;

б) В – от двух до пяти (включительно) поврежденных клубней;

4) найдите наивероятнейшее число поврежденных клубней среди 7 отобранных;

5) найдите числовые характеристики данного распределения. Покажите на графике *M(x,) σ(x).*

6.Определите среднее число солнечных дней на протяжении недели, если для данной местности вероятность того, что каждый день будет солнечным составляет 0,6.

7. Доля зараженного вредителями зерна в скрытой форме составляет 0,004.

а) Составить закон распределения числа зараженных зерен среди 500 отобранных;

б) найдите вероятность того, что более трех зерен, среди отобранных, будут зараженными;

в) найдите *M(x), Д(x), σ(x).*

8*. Потери урожая (в%) при уборке риса подчиняются следующему закону распределения:

x	*5*	*10*	*15*	*20*	x_5
P	*0,05*	*0,3*	*0,4*	*0,2*	p_5

Найти неизвестное числовое значение и соответствующую ему вероятность, если известно, что *M(x)* = 15,5.

9.При однократном обследовании, у двух животных берут анализ крови на наличие болезни. Вероятность того, что анализ крови будет положительным для первого животного, равна 0,1, для второго 0,2. Рассматриваются две случайные величины: *x* – число отрицательных показаний для первого животного; *y* - число отрицательных показаний для второго животного.

Постройте ряды распределения случайных величин *x, y* и *(x + y).* Найдите их числовые характеристики.

10. Найдите математическое ожидание и дисперсию случайной величины Z = 3x – *1/2y* - 3, если известно, что *M(x)* = 2; *M(y)* = 8; *Д(x)* = 0,4; *Д(y)* - 0,6.

11*. Вероятность госпитализации пациента при эпидемии гриппа равна 0,002. Сколько пациентов поликлиника направит на госпитализацию, если математическое ожидание числа госпитализированных равно 0.01.

1.Опыт состоит из трех независимых бросаний монеты. Составить ряд распределения числа появления герба.

2.Задан закон распределения дискретной случайной величины *у*. Найти *М(у), Д(у), σ(у), υ(у).* Построить полигон распределения, показать на графике *М(у),σ(у)*

у	*-2*	*-1*	*1*	*2*
Р	*0,1*	*0,2*	*0,5*	*0,2*

3.Вероятность того, что наудачу взятое пальто из изготовленной партии на фабрике окажется первосортным, равна 0,8. Отбираются первые попавшие 5 пальто.

1)Составьте закон распределения числа первосортных пальто среди пяти отобранных;

2) Найдите вероятность того, что более 3-х пальто из пяти окажутся первосортными;
3) Постройте многоугольник распределения.
4) Найдите *М(х), Д(х),* проверить справедливость формул для нахождения *М(х), Д{х)* случайной величины, распределенной по биномиальному закону распределения.

4. Вероятность выживания бактерий после радиоактивного облучения равна 0,005. Облучению подвергаются 400 бактерий.

   1) Найдите закон распределения случайной величины *х,* равной числу бактерий, выживших после радиоактивного облучения;
   2) найдите вероятности того, что после облучения останется не менее 3-х бактерий;
   3) найдите числовые характеристики.

5. В одном из опытов по сортировке ржи было подсчитано количество зерен в колосьях. Результаты объединены в таблицу:

x_i	30	40	50	60	70	80
n_i	10	10	20	30	20	10

Найдите *М(х); Д(х); σ(х),* приняв относительные частоты за вероятности.

6. Найдите математическое ожидание и дисперсию случайной величины $Z = 1/3y - 2x + 2$ если известно, что $M(x) = 3$; $Д(x) = 0,6$; $M(y) = 9$; $Д(y) = 0,9$.

7.Число растений на 1 дм^2 для двух злаков характеризуется следующими таблицами:

x	*20*	*35*
P	*0,7*	*0,3*

y	*50*	*30*
P	*0,6*	*0,4*

8.Составьте закон распределения числа растений на 1 дм^2 для двух злаков, *(x+y)*. Найдите *М(х), Д(х)* полученного распределения. Найдите *М(х),Д(х)* полученного распределения непосредственно используя свойства *М(х) и Д(х).*

Вопросы для самопроверки.

1. Сформулируйте определение случайной величины.
2. Какие случайные величины называются дискретными? Приведите примеры.
3. Дайте определение закона распределения случайной величины.
4. Пояснить биномиальный закон распределения случайной величины.
5. Распределение Пуассона.
6. Перечислить числовые характеристики дискретной случайной величины и объяснить их вероятный смысл.
7. Дайте определение математического ожидания *М(х)* дискретной случайной величины. Вероятностный смысл *М(х), Д(х)*.
8. Дайте определение дисперсии дискретной случайной величины. Вероятностный смысл *Д(х)*.

9. Упрощенная формула для вычисления дисперсии.

10. Приведите свойства $M(x)$ и $Д(x)$.

11. Математическое ожидание и дисперсия числа появления события в *«n»* повторных независимых испытаниях.

12. Понятие среднего квадратического отклонения и коэффициента вариации.

Глава 5. Непрерывная случайная величина. Интегральная и дифференциальная функции распределения

5.1. Необходимый теоретический минимум

Интегральная функция распределения, её свойства

Одним из наиболее удобных и универсальных способов задания закона распределения непрерывной случайной величины (НСВ) Х является функция распределения.

Определение 5.1. Интегральной функцией распределения случайной величины X называется функция, определённая для каждого действительного значения x , равная вероятности того, что случайная величина X примет значение, меньшее x, то есть $\boldsymbol{F(x)=P(X<x)}$ или $\boldsymbol{F(x)=P(-\infty < X < x)}$.

Геометрически функция распределения интерпретируется как вероятность того, что случайная точка X попадет левее данной точки x (рис. 5.1).

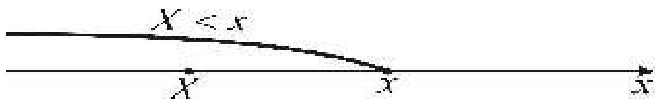


Рисунок. 5.1.

Свойства интегральной функций.

1. $0 \le F(x) \le 1$ (это свойство следует непосредственно из определения).

2. Интегральная функция распределения $F(x)$ есть неубывающая функция, то есть для $x_2 > x_1 \Rightarrow F(x_2) \ge F(x_1)$.

Следствие 1. Вероятность того, что непрерывная случайная величина X примет значение, заключённое в интервале ($a; b$), равна приращению интегральной функции распределения на этом интервале:

$$P(a<X<b)=F(b)-F(a). \qquad (5.1)$$

Следствие 2. Вероятность того, что непрерывная случайная величина X примет одно определённое значение, например, x_0, равна нулю: $P(X=x_0)=0$.

Объединяя *следствие 1* и *следствие 2*, можно записать:

$$P(a<X<b)=P(a\le X<b)=P(a< X \le b) = P(a < X < b) = P(a \le X \le b).$$

3. Если все возможные значения случайной величины X принадлежат интервалу ($a;b$), то: а) $F(x)=0$ при $x\le a$ (иначе: $F(a)=0$);

б) $F(x)=1$ при $x \ge b$ (иначе $F(b)=1$).

Следствие. Если значения случайной величины принадлежат интервалу (-∞; ∞), то: а) $\lim\limits_{x\to-\infty} F(x)=0$ (иначе $F(-\infty)=0$);

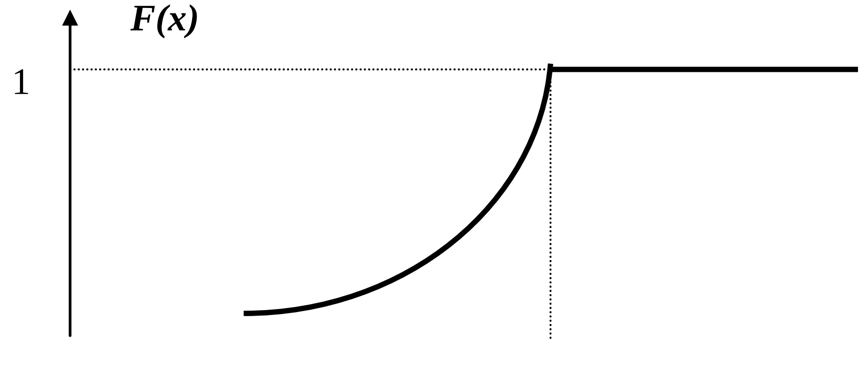
б) $\lim\limits_{x\to+\infty} F(x)=1$ (иначе $F(\infty)=1$).

4. Функция $F(x)$ в точке x_0 непрерывна слева, то есть

$$\lim_{x\to x_0-0} F(x)=F(x_0).$$

График функции распределения $F(x)$ для *непрерывной* случайной величины (НСВ) имеет вид, представленный на рис.5.1, и в общем случае задаётся следующим образом:

$$F(x)=\begin{cases} 0 \quad \text{при } x\le a; \\ \varphi(x) \quad \text{при } a<x\le b; \\ 1 \text{ при } x>b. \end{cases}$$



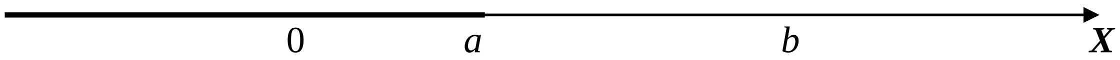


Рисунок 5.1.

- *Интегральная функция распределения* существует для *любой* случайной величины, как для непрерывной, так и для дискретной.
- При известном законе распределения интегральная функция распределения дискретной случайной величины имеет вид:

$$F(x)=P(X<x)=\sum_{(x_i<x)} p_i=\sum_{(x_i<x)} P(X=x_i)\ . \tag{5.2}$$

- В формуле (5.2) ($x_i<x$) означает, что суммирование ведётся по всем индексам i, для которых это неравенство выполняется.
- *Графиком* функции распределения *дискретной* случайной величины является *ступенчатая фигура*, сохраняющая постоянное значение на каждом интервале не содержащем точек x_i и терпящая в этих точках скачок, равный p_i (рис.5.5).

Дифференциальная функция распределения, её свойства

Определение 5.2. Плотностью (дифференциальной функцией) распределения непрерывной случайной величины называется первая производная от функции распределения: $\boldsymbol{f(x)=F'(x)}$.

Плотность распределения существует только для непрерывной случайной величины.

Свойства плотности распределения:

1. Плотность распределения $f(x)$ неотрицательна, то есть $f(x)\geq 0$.

2. Вероятность того, что непрерывная случайная величина примет значение, принадлежащее интервалу $(a;b)$ равна:

$$P(a<X<b)=\int_a^b f(x)dx\ . \tag{5.3}$$

3. Интегральная функция распределения непрерывной случайной величины при известной плотности распределения вычисляется по формуле:

$$F(x)=\int_{-\infty}^{x} f(x)dx. \tag{5.4}$$

4. Несобственный интеграл от плотности распределения в пределах от $-\infty$ до $+\infty$ равен единице:

$$\int_{-\infty}^{\infty} f(x)dx = 1. \qquad (5.5)$$

Следствие. Если все значения НСВ сосредоточены на интервале$(a;\, b)$, то:

$$\int_{a}^{b} f(x)dx = 1. \qquad (5.6)$$

- *График плотности распределения* f(x) называется *кривой распределения*. Причём площадь под кривой распределения равна 1 (по свойству (4)).

Числовые характеристики непрерывных случайных величин

1.Математическое ожидание M(X):

$$M(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx. \qquad (5.7)$$

Если все значения НСВ принадлежат интервалу *(a;b),* то формула (5.7) примет следующий вид:

$$M(X) = \int_{a}^{b} x \cdot f(x)dx. \qquad (5.8)$$

2.*Мода $M_0(X)$ непрерывной случайной величины* – это то её возможное значение, которому соответствует локальный максимум плотности распределения (рис.5.2)

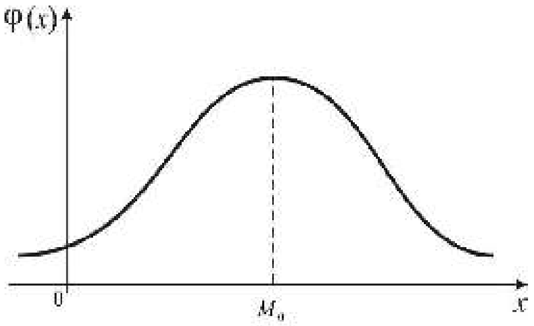


Рисунок 5.2.

3.*Медиана $M_e(X)$ непрерывной случайной величины* – это то её возможное значение, которое определяется равенством: $P(X<M_e)=P(X>M_e)$. (Геометрически медиана есть точка, в которой ордината *f(x)* делит пополам площадь, ограниченную кривой распределения, рис. 5.3))

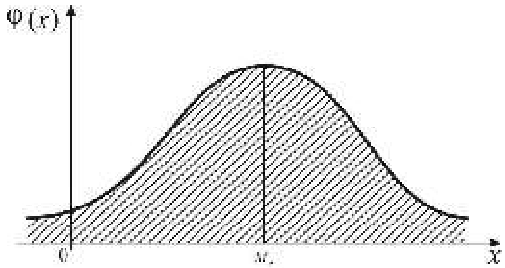


Рисунок 5.3.

4.*Дисперсия D(X):*

$$D(X) = \int_{-\infty}^{\infty} [x - M(X)]^2 \cdot f(x)dx. \qquad (5.9)$$

Если все значения НСВ принадлежат интервалу *(a;b),* то формула (5.9) примет следующий вид:

$$D(X) = \int_{a}^{b} [x - M(X)]^2 \cdot f(x)dx. \qquad (5.10)$$

Дисперсию также можно вычислять (соответственно для $x \in (-\infty;+\infty)$ и $x \in (a;\ b)$) по упрощенным формулам:

$$D(X) = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx - [M(X)]^2. \qquad (5.11)$$

$$D(X) = \int_{a}^{b} x^2 \cdot f(x)dx - [M(X)]^2. \qquad (5.12)$$

- Все свойства математического ожидания и дисперсии, перечисленные выше для дискретных случайных величин, справедливы и для непрерывных случайных величин.

5.*Среднее квадратическое отклонение* *σ(X) непрерывной случайной* величины определяется также, как и для дискретной случайной величины, то есть: $\sigma(X) = \sqrt{D(x)}$.

Представим числовые характеристики дискретных и непрерывных случайных величин в виде таблицы 5.1:

Таблица 5.1.

Название числовой характеристики	Для дискретной случайной величины	Для непрерывной случайной величины

Математическое ожидание	$M(X)=\sum_{i=1}^{n}x_i p_i$	$M(X)=\int_{-\infty}^{+\infty}x\varphi(x)dx$
Дисперсия	$D(X)=M(X-M(X))^2$	$D(X)=\int_{-\infty}^{+\infty}(x-M(X))^2\varphi(x)dx$
	Упрощенная $D(X)=M\ (X^2)-M^2(X)$	Упрощенная $D(X)=\int_{-\infty}^{+\infty}x^2\varphi(x)dx-M^2(X)$
Среднеквадратическое отклонение	$\sigma(X)=\sqrt{D(X)}$	

5.2. Примеры решения типовых задач.

Задание 5.1. Составить интегральную функцию распределения дискретной случайной величины X, заданной таблицей. Построить график.

x	3	5	7	9
p	0,2	0,3	0,4	0,1

Решение.

По определению $F(x)=P(X<x)$. Вычислим значения функции $F(x)$ на интервалах $(-\infty;3],(3;5],(5;7],(7;9],(9;+\infty)$

1) при $x\in(-\infty;3] \Longrightarrow F(x)=0$,так как случайная величина X не принимает ни одного значения меньшего 3.

2) при $x\in(3;5] \Longrightarrow F(x)=P(X<x)=P(x=3)=0{,}2$,так как случайная величина X принимает одно числовое значение $x=3$, меньшее x.

3) при $x\in(5;7] \Longrightarrow F(x)=P(X<x)=P(x=3)+P(x=5)=0{,}2+0{,}3=0{,}5$,так как случайная величина X принимает два значения $x=3$ и $x=5$,меньшее x.

4) при $x\in(7;9] \Longrightarrow F(x)=P(X<x)=P(x=3)+P(x=5)+P(x=7)=0{,}2+0{,}3+0{,}4=0{,}9$

5) при $x\in(9;+\infty) \Longrightarrow F(x)=P(X<x)=P(x=3)+P(x=5)+P(x=7)+P(x=9)=$ $=0{,}2+0{,}3+0{,}4+0{,}1=1$

Итак, интегральная функция распределения аналитически может быть записана в виде:

$$F(x)=\begin{cases} 0 & \text{при} \quad x\epsilon(-\infty;3] \\ 0{,}2 & \text{при} \quad x\epsilon(3;5] \\ 0{,}5 & \text{при} \quad x\epsilon(5;7] \\ 0{,}9 & \text{при} \quad x\epsilon(7;9] \\ 1 & \text{при} \quad x\epsilon(9;+\infty) \end{cases}$$

Строим график: по оси абсцисс откладываем значения случайной величины и отмечаем интервалы, а по оси ординат – значения $F(x)$, соответствующие каждому интервалу (рис. 5.5).

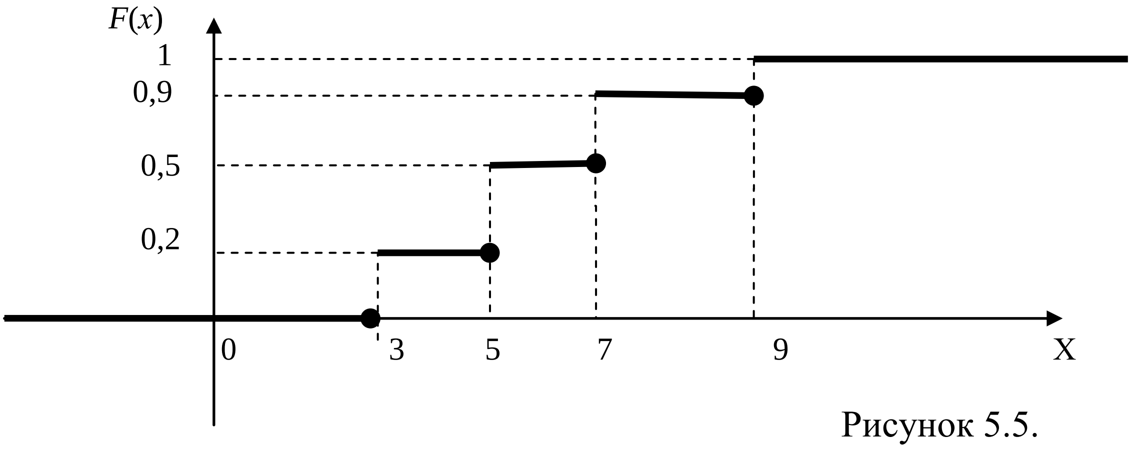


Рисунок 5.5.

Задача 5.2. Случайная величина X задана интегральной функцией распределения *F(x).* Найти:

1) дифференциальную функцию распределения *f(x);*

2) математическое ожидание *M(X);*

3) дисперсию *D(X);*

4) вероятность попадания случайной величины *X* в интервал *(2;-5);*

5) постройте графики функций *F(x)* и *f(x).*

$$F(x)=\begin{cases} 0 & ,x<1 \\ \frac{1}{8}x^2-\frac{1}{8} & ,1\leq \text{x} \leq 3 \\ 1 & ,x<3 \end{cases}$$

Решение:

1) Найдем дифференциальную функцию, плотность распределения f(x) по

определению: $f(x) = F'(x)$

$$f(x) = \begin{cases} 0, & x < 1; \\ \dfrac{1}{4}x, & 1 \le x \le 3; \\ 0, & x > 3. \end{cases}$$

2) Математическое ожидание непрерывной случайной величины находим по формуле: $M(X) = \int\limits_{-\infty}^{\infty} x \cdot f(x)dx$

$$M(X) = \int\limits_{-\infty}^{\infty} x \cdot f(x)dx = \int\limits_{-\infty}^{1} x \cdot 0dx + \int\limits_{1}^{3} x \cdot \frac{1}{4}xdx + \int\limits_{3}^{+\infty} x \cdot 0dx = \frac{1}{4}\int\limits_{1}^{3} x^2dx = \frac{1}{4} \cdot \frac{x^3}{3}\Bigg|_1^3 = \frac{1}{12}[27-1] = \frac{26}{12} = \frac{13}{6}$$

$\underline{M(X) = 2{,}167}$

3) Дисперсию найдем по формуле: $D(X) = M(X^2) - [M(X)]^2$

$$M(X^2) = \int\limits_{-\infty}^{\infty} x^2 f(x)dx = \int\limits_{-\infty}^{1} x^2 \cdot 0dx - \int\limits_{1}^{3} x^2 \cdot \frac{1}{4}xdx + \int\limits_{3}^{\infty} x^2dx = \frac{1}{4}\left[\frac{x^4}{4}\right]_1^3 = \frac{1}{16}[81-1] = 5$$

$D(X) = 5 - \dfrac{169}{36} = 0{,}306$ $\qquad$ $D(X) = 0{,}306$

4) Вероятность попадания в интервал $(\alpha;\beta)$ находим по

формуле $P(\alpha < X < \beta) = F(\beta) - F(\alpha)$

$$P(2 < X < 5) = F(5) - F(2) = 1 - \frac{1}{8}(4-1) = 1 - \frac{3}{8} = \frac{5}{8} = 0.625.$$

5) Строим график функции *F(x)*, при этом учитываем, что $0 \le F(x) \le 1$ и на отрезке $[1;3]$ графиком (рис. 5.6) является парабола $y = \frac{1}{8}(x^2 - 1)$

График плотности *f(x)* (рис.5.7) при $x < 1$ и $x > 3$ совпадает с осью *OX* , на отрезке $[1;3]$ с прямой $y = \dfrac{1}{4}x$

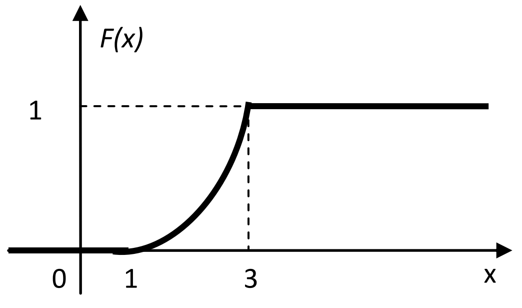


Рисунок 5.6.

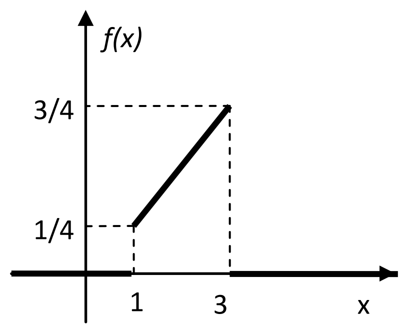


Рисунок 5.7.

5.3. Задачи для самостоятельного решения.

1. Случайная величина задана таблицей:

x	*-3*	*- 1*	*0*	*2*	*5*
P	*0,2*	*0,1*	*0,4*	*0,25*	*0,05*

Составьте интегральную функцию распределения и постройте ее график, найдите $P(-3 \leq x \leq 2)$?

2. Игральная кость бросается три раза. Найдите *М(х)* и составьте интегральную функцию распределения числа выпадений шестерки.

3. Случайная величина *х* задана интегральной функцией; задан интервал *(а; в)*

$$F(x) = \begin{cases} 0, -\infty < x \leq 0 & a = 0 \\ 0{,}2x, 0 < x \leq 5 & в = 3 \\ 1, \ 5 < x < +\infty & \end{cases}$$

1) Постройте график этой функции.

2) Найдите вероятность того, что случайная величина примет значение из данного интервала.

3) Найдите дифференциальную функцию распределения и постройте ее график.

4) Найдите *М(х)* и *Д(х).*

5*.Дана дифференциальная функция распределения случайной величины

$$f(x) = \begin{cases} 0, при -\infty < x < 0 \\ {}^{2}/_{9x}, при\ 0 < x < 3 \\ 0, при\ 3 \leq x < +\infty \end{cases}$$

Требуется: 1) найти интегральную функцию распределения;

2) построить графики дифференциальной и интегральной функции;

3) найти математическое ожидание и дисперсию этой случайной величины.

---

1. Случайная величина задана таблицей:

y	*1*	*3*	*6*	*8*
P	*0,3*	*0,5*	*0,1*	*0,4*

Составьте интегральную функцию распределения и постройте ее график. $P(1 \leq x \leq 6)$-?

2.Случайная величина задана функцией распределения:

$$F(x) = \begin{cases} 0, при -\infty < x < 2 \\ x^2 - 2, при\, 2 \leq x \leq 3 \\ 1, при\, x > 3 \end{cases}$$

1) Постройте график $F(x)$;

2) Вычислите вероятность попадания случайной величины х в интервале [1; 2,5] и [2,5; 3,5].

3) Найдите дифференциальную функцию $f(x)$ и постройте ее график.

4) Найдите $M(x)$, $Д(x)$, $\sigma\{x)$

3*. Случайная величина х задана функцией распределения:

$$f(x) = \begin{cases} 0, при -\infty < x < 1 \\ (a+1)x, при\, 1 < x \leq 3 \\ 0, при\, x > 3 \end{cases}$$

Найдите: 1) коэффициент *a*; 2)$M(x)$ и $Д(x)$.

4*. Дана дифференциальная функция распределения случайной величины:

$$f\ (x) = \begin{cases} 0, при -\infty < x \leq 0 \\ x/8\, , при\, 0 < x \leq 4 \\ 0, при\, 4 < x < +\infty \end{cases}$$

1) Найдите интегральную функцию для этого распределения.
2) Постройте график дифференциальной и интегральной функции.
3) Вероятность попадания случайной величины в интервалы: а) (-2; 3); б) (3; 8).
4) Найдите математическое ожидание этой случайной величины.

Вопросы для самопроверки.

1. Дайте определение непрерывной случайной величины, привести примеры.

2. Что такое интегральная функция распределения случайной величины, каковы ее свойства? График *F(x) - ?*
3. Определение и основные свойства дифференциальной функции, распределение непрерывной случайной величины.
4. Вероятностный смысл плотности распределения.
5. Формулы для вычисления числовых характеристик непрерывной случайной величины.

Глава 6. Нормальный закон распределения. Закон больших чисел.

6.1. Необходимый теоретический минимум

Нормальный закон распределения. Кривая Гаусса.

Закон нормального распределения вероятностей непрерывной случайной величины занимает особое место среди различных теоретических законов, т. к. является основным во многих практических исследованиях. Им описывается большинство случайных явлений, связанных с производственными процессами.

К случайным явлениям, подчиняющимся нормальному закону распределения, относятся ошибки измерений производственных параметров, распределение технологических погрешностей изготовления, рост и вес большинства биологических объектов и др.

Определение 6.1. Непрерывная случайная величина X имеет ***нормальное распределение вероятностей*** с параметрами a и, если ее плотность распределения имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}, \qquad (6.1)$$

где a - математическое ожидание случайной величины;

σ-среднее квадратичное отклонение нормального распределения.

- Асимметрия, эксцесс, мода и медиана нормального распределения соответственно равны:

$$A_S = 0; \quad E_K = 0; \quad M_0 = a; \quad M_e = a.$$

Функция распределения нормальной случайной величины задается формулой:
$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt. \qquad (6.2)$$

Графиком дифференциальной функции нормального распределения (6.1) является *кривая Гаусса* (рис. 6.1), которая наглядно показывает распределение плотности вероятности случайной величины *X*.

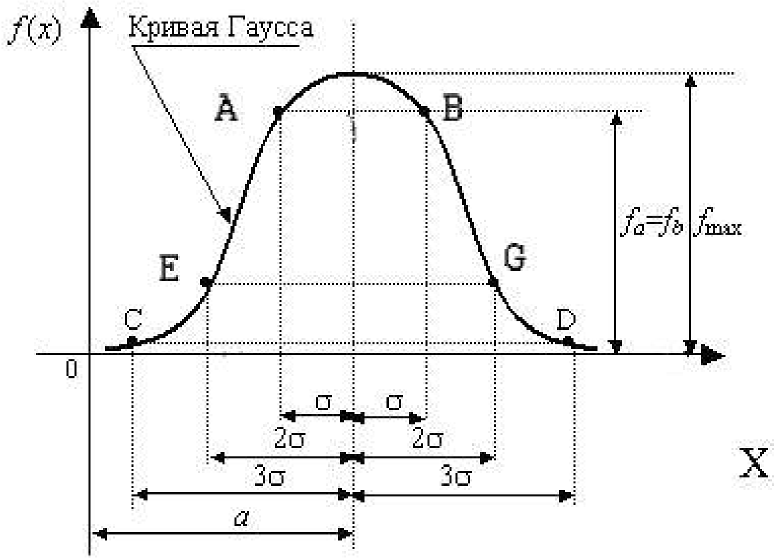


Рисунок.6.1.

Свойства нормальной кривой (кривой Гаусса):

1. Кривая симметрична относительно прямой $x = a$;

2. Нормальная кривая расположена над осью X, т. е. при всех значениях X функция f(x) всегда положительна;

3. Ось ох является горизонтальной асимптотой графика: $\lim\limits_{|x|\to\infty} f(x) = 0$

4. При x = a функция f(x) имеет максимум равный: $f_{\max} = \frac{1}{\sigma\sqrt{2\pi}} \approx \frac{0,4}{\sigma}$

5. В точках *A и B* при $x=a-\sigma$ и $x=a+\sigma$ кривая имеет точки перегиба, ординаты которых равны: $f_A=f_B=\dfrac{1}{\sigma\sqrt{2\pi e}}\approx 0{,}6\cdot f_{\max}\approx\dfrac{0{,}24}{\sigma}$

Влияние параметров а и σ на форму кривой Гаусса.

Изменение величины параметра *а* (математического ожидания случайной величины) не изменяет форму нормальной кривой, а приводит лишь к ее смещению вдоль оси X: вправо, если *а возрастает*, и влево, если *а убывает*.

При *а=0* нормальная кривая симметрична относительно оси ординат.

Изменение величины параметра σ(среднего квадратичного отклонения) изменяет форму нормальной кривой:

с *возрастанием* σ ординаты нормальной кривой убывают, кривая *растягивается* вдоль оси X и прижимается к ней;

с *убыванием* σ ординаты нормальной кривой увеличиваются, кривая *сжимается* вдоль оси X и становится более "островершинной".

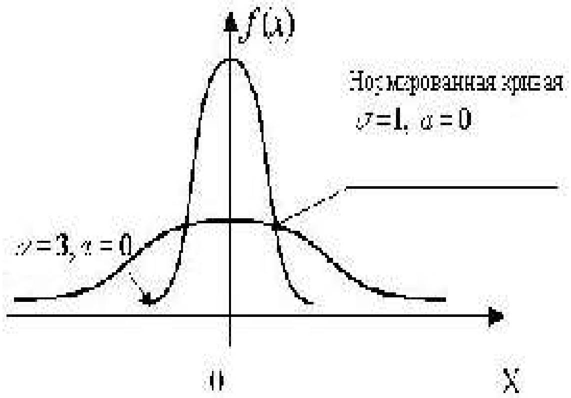
При этом, при любых значениях α и σ площадь ограниченная нормальной кривой и осью X, остается равной единице (т. е. вероятность того, что случайная величина, распределенная нормально, примет значение ограниченное на оси X нормальной кривой, равна 1).

Определение 6.2. Нормальное распределение с произвольными параметрами α и σ, т. е. описываемое дифференциальной функцией

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}$$

называется ***общим нормальным распределением***, кратко записывается так: *X~N(a;σ).*

Определение 6.3. Нормальное распределение называется ***стандартным*** (или **нормированным**), если его параметры равны *а*=0, σ =1, то есть



$X \sim N(0;1)$,(рис. 6.2).

- *Плотность вероятности* стандартного нормального распределения обычно обозначается $\varphi(x)$ и определяется формулой:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$

Значения функции $\varphi(x)$ затабулированы
(см. таблицу 2 Приложения).

- Функция распределения стандартного нормального распределения имеет вид:

Рисунок 6.2.

- $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ и выражается через *функцию Лапласа* $\Phi(x)$, которая также затабулирована (см. таблицу 3 Приложения).

От произвольного нормального распределения $X \sim N(a;\ \sigma)$ можно перейти к стандартному $X \sim N(0;1)$, воспользовавшись следующими формулами:

1) формула для вычисления значений плотности общего нормального распределения через плотность стандартного нормального распределения:

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x-a}{\sigma}\right); \qquad (6.3)$$

2) формула для вычисления значений интегральной функции нормального распределения через функцию Гаусса:

$$F(x) = 0{,}5 + \Phi\left(\frac{x-a}{\sigma}\right). \qquad (6.4)$$

Пример. Дано $X \sim N(160;8)$. Вычислить значения дифференциальной и интегральной функций заданного нормального распределения в точке x=130.

*Решение.*1)Найдем сначала «новый» аргумент: $\frac{x-a}{\sigma} = \frac{130-160}{8} = -3{,}75.$

Тогда по формуле (6.3) получаем:

$$f(130) = \frac{1}{8}\varphi(-3{,}75) = \frac{1}{8}\varphi(3{,}75) = \frac{1}{8} \cdot 0{,}0004 = 0{,}00005.$$

- Здесь использовали свойство четности функции $\varphi(x)$ и таблицу 2 Приложения.

2)Значение интегральной функции $F(130)$ найдем по формуле (6.4):

$F(130)=0{,}5+\Phi(-3{,}75)= 0{,}5 - \Phi(3{,}75)=0{,}5 - 0{,}4999=0{,}0001$.

- Здесь использовали свойство нечетности функции $\Phi(x)$ и таблицу 2 Приложения.

Если требуемая *точность* ответа задачи предполагает вычисление вероятности случайной величины с заданной точностью, а в таблице значений функции $Ф(x)$ нет нужного значения x, можно использовать формулу линейной интерполяции:

$$Ф(x) = Ф(x_1) + \frac{x - x_1}{x_2 - x_1} \cdot \left[Ф(x_2) - Ф(x_1)\right] \quad , \quad где\; x \in (x_1; x_2). \qquad (6.5)$$

Пример. Найти значение функции Лапласа в точке x=3,33 с точностью до четырёх знаков после запятой.

Решение. В таблице 3 Приложения нет аргумента $x=3,33$. Поэтому находим те значения, между которыми расположено x=3,33 => x_1=3,20 и x_2=3,40 => $Ф(x_1)=Ф(3,20)=0,49931$; $Ф(x_2)=Ф(3,40)=0,49966$. Подставим найденные значения в формулу (6.5) и проведём соответствующие вычисления:

$$Ф(3,33) = Ф(3,20) + \frac{3,33-3,20}{3,40-3,20}\left[Ф(3,40) - Ф(3,20)\right] = 0,4931 + \frac{0,13}{0,20}(0,49966 - 0,49954) \approx 0,4995$$

Основные формулы нормального распределения. Правило трех сигм.

Случайные величины с нормальным распределением обычно используют при решении двух типов задач:

1. Найти вероятность того, что случайная величина $X \sim N(a;\sigma)$ принимает значение в интервале $(c;\, d)$. Находят эту вероятность по формуле:

$$P(c < X < d) = Ф\left(\frac{d-a}{\sigma}\right) - Ф\left(\frac{c-a}{\sigma}\right). \qquad (6.6)$$

Пример. Случайная величина X распределена по нормальному закону. Математическое ожидание и среднее квадратичное отклонение этой случайной величины равны a=30 и $\sigma = 10$. Найти вероятность того, что X примет значение в интервале (10, 50).

Решение: По условию: $c = 10; d = 50; a = 30; \sigma = 10$. Тогда

$$P(10 < X < 50) = Ф\left(\frac{50-30}{10}\right) - Ф\left(\frac{10-30}{10}\right) = Ф(2) - Ф(-2) = 2 \cdot Ф(2)$$

Пользуясь готовыми таблицами Лапласа (см. приложение 3), имеем:
$P(10 < X < 50) = 2 \cdot 0,4772 = 0,9544$.

2. Найти вероятность того, что случайная величина $X \sim N(a;\sigma)$ отличается от своего среднего значения a (центра рассеивания) по абсолютной величине не больше, чем на ε - ***заданное отклонение.***

Для этого применяют формулу:

$$P(|x-a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right). \qquad (6.7)$$

Преобразуя формулу (6.7) получим вероятность попадания нормально распределенной величины в интервал $(a-\varepsilon < x \pi a+\varepsilon)$, который называют ***доверительным***, а вероятность попадания $\gamma = P(a-\varepsilon < x \pi a+\varepsilon)$ ***доверительной вероятностью или надежностью.***

Правило «трех сигм».

При подстановке в формулу (6.7) $\varepsilon = 3\sigma$ получаем:

$$P(|x-a| < 3\sigma) = 2\Phi\left(\frac{3\sigma}{\sigma}\right) = 2\Phi(3) = 2 \cdot 0,49895 = 0,9973 \approx 99,7\%.$$

Данный результат называют правилом «трех сигм».

Правило «трех сигм» гласит:

практически достоверно (на 99,7%), что значения нормально распределенной случайной величины отличаются от своего среднего значения a не более, чем на 3σ,

или:

практически все значения нормально распределенной случайной величины X находятся в интервале $(a-3\sigma;\ a+3\sigma)$ (см. рис.6.1).

Интервал $(a-3\sigma;\ a+3\sigma)$ называют ***диапазоном*** изменения нормально распределенной случайной величиной.

Закон больших чисел.

Закон больших чисел утверждает, что при очень большом числе случайных явлений средний их результат перестаёт быть случайным и может быть предсказан с большой степенью определённости. Например, невозможно предсказать какую конкретную сумму снимет или положит очередной клиент банка, но суммарное значение этих операций за день остаётся примерно одним и тем же в обычные будничные дни.

При решении задач используют теоремы, выражающие закон больших чисел в виде следующих неравенств:

1. ***Неравенство Маркова:***

$$P(X \geq \varepsilon) \leq \frac{M(X)}{\varepsilon} \qquad (6.8)$$

или

$$P(X < \varepsilon) \geq 1 - \frac{M(X)}{\varepsilon}. \qquad (6.9)$$

2. ***Неравенство Чебышева:***

$$P(|X - M(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2} \qquad (6.10)$$

или

$$P(|X - M(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}. \qquad (6.11)$$

• Здесь *M(X), D(X)* - соответственно математическое ожидание и дисперсия случайной величины *X*

Пример. В результате систематических проводимых наблюдений было установлено, что среднее значение потребности фермерского хозяйства в некоторой продукции 25 единиц. Оценить вероятность того, что суточная потребность в этом продукте превысит 30 единиц.

Решение. Пусть X- *потребность хозяйства в некотором продукте.* Так как среднее значение – это математическое ожидание, то *M(X)=25*. Требуемую вероятность оценим, используя неравенство Маркова в виде (6.8), где ε=*30*. Тогда получим:

$$P(X>30)<25/30 \;\Rightarrow\; \qquad P(X>30)<5/6.$$

3. ***Формула, выражающая теорему Чебышева:***

$$P(|\overline{X}_n - \overline{a}_n| \geq \varepsilon) \leq \frac{\sigma^2}{n \cdot \varepsilon^2} \qquad (6.12) \qquad \text{или} \qquad P(|\overline{X}_n - \overline{a}_n| < \varepsilon) \geq 1 - \frac{\sigma^2}{n \cdot \varepsilon^2}. \qquad (6.13)$$

• Здесь $X_1, X_2, \ldots, Xn$ – независимые случайные величины; $M(X_i)=a_i$ – математическое ожидание; $D(X_i)=\sigma_i^2 \le \sigma^2$ – ограниченные в совокупности дисперсии;

$\overline{X}_n = \dfrac{X_1+X_2+\ldots+X_n}{n} = \dfrac{1}{n}\cdot\sum_{i=1}^{n} X_i$ –средняя арифметическая случайных величин;

$\overline{a}_n = \dfrac{a_1+a_2+\ldots+a_n}{n} = \dfrac{1}{n}\cdot\sum_{i=1}^{n} a_i$. – средняя арифметическая математических ожиданий случайных величин.

Следствие 1. Если случайные величины имеют одинаковые математические ожидания и одинаковые дисперсии, то есть $M(X_i)=a$ $D(X_i)=\sigma^2$, то:

$$P\left(\left|\overline{X}_n - a\right| \ge \varepsilon\right) \le \frac{\sigma^2}{n\cdot\varepsilon^2} \quad (6.14) \qquad \text{или} \qquad P\left(\left|\overline{X}_n - a\right| < \varepsilon\right) \ge 1-\frac{\sigma^2}{n\cdot\varepsilon^2}. \quad (6.14)$$

Следствие 2. В применении к выборочному методу формула (6.13)

принимает вид:
$$P\left(\left|\overline{X}_в < \overline{X}_г\right| < \varepsilon\right) \le 1-\frac{\sigma^2}{n\cdot\varepsilon^2}, \qquad (6.15)$$

где $\overline{X}_в = \dfrac{1}{n}\cdot\sum_{i=1}^{n} x_i$ – выборочная средняя;

$\overline{X}_г = \dfrac{1}{n}\cdot\sum_{i=1}^{n} M(X_i)$ – генеральная средняя.

4. ***Формула, выражающая теорему Бернулли:***

$$P\left(\left|\frac{m}{n}-p\right| \ge \varepsilon\right) \le \frac{pq}{n\cdot\varepsilon^2} \quad (6.16) \text{ или } \quad P\left(\left|\frac{m}{n}-p<\varepsilon\right|\right) \ge 1-\frac{pq}{n\cdot\varepsilon^2} \quad (6.17)$$

- Здесь p – вероятность появления события A в каждом из n повторных независимых испытаниях; q=1 – p – вероятность непоявления события A в повторных испытаниях; m/n – относительная частота появления события A.
- Иногда теорему Бернулли удобнее применять в следующей форме:

$$P\left(\left|m-np\right| \ge \varepsilon\right) \le \frac{npq}{\varepsilon^2} \quad (6.18) \text{ или } \quad P\left(\left|m-p \le \varepsilon\right|\right) \ge 1-\frac{npq}{\varepsilon^2}. \quad (6.19)$$

Здесь m – абсолютная частота появления события, np – математическое ожидание, npq – дисперсия числа появления события A в n испытаниях.

Закон больших чисел имеет важное практическое значение. Именно, на этом законе основано утверждение, что среднее арифметическое результатов измерений считается наиболее точным, наиболее близким к истинному значению измеряемой величины. Закон больших чисел широко используется в статистике, на нем основан выборочный метод, рассмотренный в следующем разделе.

6.2. Примеры решения типовых задач.

Задача 6.1. Средний живой вес коров равен 415 кг, а среднее квадратическое отклонение составляет 21 кг. Считая, что живой вес коров подчиняется нормальному закону распределения, требуется::

1) записать дифференциальную функцию распределения;
2) вычислить процент коров с весом от 350 до 450 кг; проиллюстрировать графически;
3) найти вероятность, с которой можно гарантировать, что предельное отклонение живого веса коров от их среднего значения будет: а) не более 15,75 кг; б) более 15,75 кг;
4) определить величину предельного отклонения в ту или другую сторону живого веса от среднего значения с вероятностью 0,9901.
5) найти диапазон изменения веса коров.

Решение. Случайная величина X – *живой вес коров;* по условию $X \sim N(415;21)$, откуда следует, что $a=415$, $\sigma=21$.

1) Подставим параметры данного нормального распределения $a=415$, $\sigma=21$ в формулу (6.1). Получим искомую плотность распределения вероятностей:

$$f(x)=\frac{1}{21\sqrt{2\pi}}exp\left(-\frac{(x-415)^2}{2\cdot 21^2}\right) \Rightarrow \quad f(x)=\frac{1}{21\sqrt{2\pi}}exp\left(-\frac{(x-415)^2}{882}\right).$$

2)Для нахождения процента коров, имеющих вес от 350 до 450 кг, воспользуемся формулой (6.6) при $c=350$, $d=450$. Получаем:

$$P(350 < X < 450)=\Phi\left(\frac{450-415}{21}\right)-\Phi\left(\frac{350-415}{21}\right)=$$

$$=\Phi(1,67)+\Phi(3,09)=0,4525+0,4986\approx 0,951=95,1\%.$$

Графическая иллюстрация найденной вероятности приведена на рис.6.3.

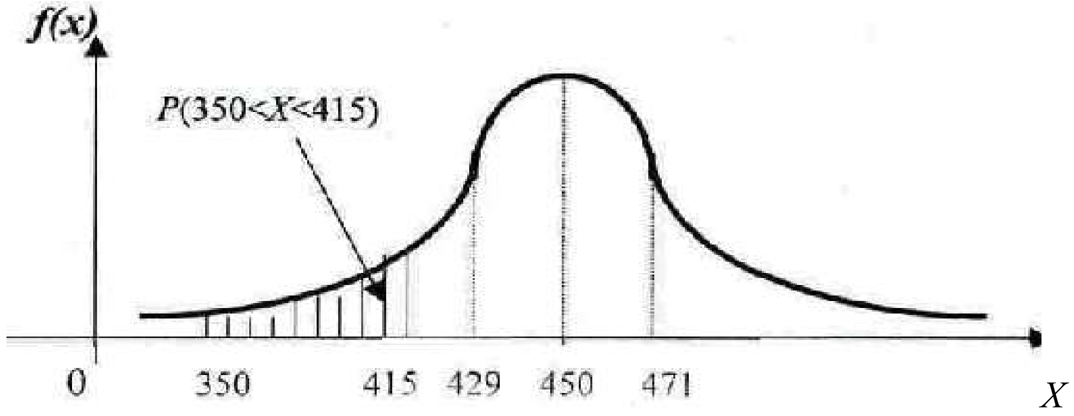


Рисунок 6.3

3)а) Найдём по формуле (6.7) (при ε=15,75) вероятность, с которой можно гарантировать, что предельное отклонение живого веса коров от их среднего значения будет не более 15,75 , то есть:

$$P(|x-a| \leq 15{,}75) = P(|x-a| < 15{,}75).$$

Получим:

$$P(|x-a| \leq 15{,}75) = 2\Phi\left(\frac{15{,}75}{21}\right) = 2\Phi(0{,}75) = 2\cdot 0{,}2734 = 0{,}5468 \approx 0{,}547$$

б) Для вычисления вероятности отклонения X от a более чем на 15,75, то есть $P(|x - a| > 15{,}75)$, перейдём к рассмотрению вероятности противоположного события:

$P(|x-a| > 15{,}75) = 1 - P(|x-a| \leq 15{,}75) = 1 - 0{,}5468 = 0{,}4532 \approx 0{,}453$.

4)В данном случае известно, что $P(|x-a| < \varepsilon) = 0{,}9901$, а найти требуется ε.. Снова воспользуемся формулой (6.7):

$$P(|x-a| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{21}\right) = 0{,}9901 \Rightarrow \Phi\left(\frac{\varepsilon}{21}\right) = \frac{0{,}9901}{2} = 0{,}4502.$$

По таблице 3 Приложения находим неизвестное аргумента значение (ε/21) по известному значению функции (0,4502):

$$\varepsilon / 21 = 2{,}58 \Rightarrow \varepsilon = 2{,}58\cdot 21 = 54{,}18.$$

Можно также написать *доверительный интервал,* в который с заданной вероятностью *p=0,990*1 (или *надёжностью)* попадут значения рассматриваемой случайной величины *X:*

$$-54{,}18 < x - 415 < 54{,}18 \quad \Rightarrow \quad 415 - 54{,}18 < x < 415 + 54{,}18 .$$

$360{,}82 < x < 469{,}18$ - доверительный интервал с надёжностью 0,9901.

5)Диапазон изменения есть интервал ($a - 3\sigma;\ a + 3\sigma$), подставим данные задачи: (415–3*21; 415+3*21), получим вес коров лежит в диапазоне от 352 кг до 478 кг.

Задача 6.2. Случайная величина *X* задана рядом распределения:

X	2	4	6	8	10	12
р	0,7	0,2	0,04	0,03	0,02	0,01

Требуется: 1) найти вероятность того, что $|X - M(X)| < 2;$;

2) оценить эту же вероятность, пользуясь неравенством Чебышева.

Решение. 1) Вычислим математическое ожидание случайной величины *X*: $M(X) = 2\cdot 0{,}7+4\cdot 0{,}2+6\cdot 0{,}04+8\cdot 0{,}03+10\cdot 0{,}02+12\cdot 0{,}01 = 3$.

Тогда получаем:

$$P(|X - M(X)| < 2) = P(|X - 3| < 2) = P(1 < X < 5) = P(X = 2) + P(X = 4) = 0{,}7 + 0{,}2 = 0{,}9.$$

2) Оценим эту же вероятность, пользуясь неравенством Чебышева в форме (6.11) при $\varepsilon = 2$, $M(X)=3$. Получим: $P(|X - 3| < 2) \geq 1 - \frac{D(X)}{2^2}$.

Для нахождения дисперсии *D(X)* воспользуемся ее определением:

$$D(X) = M[X - M(X)]^2 = M[X - 3]^2.$$

Закон распределения случайной величины $[X - M(X)]^2 = [X - 3]^2$ имеет вид:

$[X - M(X)]^2$	1	1	9	25	49	81
p	0,7	0,2	0,04	0,03	0,02	0,01

Тогда $D(X) = 1\cdot 0{,}7+1\cdot 0{,}2+9\cdot 0{,}04+25\cdot 0{,}03+49\cdot 0{,}02+81\cdot 0{,}01=3{,}8$.

В результате получим:

$$P(|X - 3| < 2) \geq 1 - \frac{3{,}8}{4}, \quad \text{то есть} \quad P(|X - 3| < 2) \geq 0{,}05.$$

- Заметим, что полученные ответы хотя и согласуется, но разница большая так как неравенство Чебышева дает хорошую оценку при большом числе испытаний.

Задача 6.3. На поле прямоугольной формы посеяно 2000 рядов культуры. Для определения средней урожайности собрали початки в каждом десятом ряду и на основании этих данных вычислили выборочную среднюю урожайность. Дисперсия урожайности на каждом обследованном участке не превышает 20. Оценить вероятность того, что выборочная средняя урожайность будет по абсолютной величине отличаться на всем поле не более чем на: а) 1 ц/га; б) 0,5 ц/га. Сделать заключение о влиянии предельной ошибки на оцениваемую вероятность.

Решение. При решении задачи используем формулу (6.15), выражающую *теорему Чебышева в применении к выборочному методу,* где ε – предельная ошибка: $P\left(\left|\overline{X}_в - \overline{X}_г\right| < \varepsilon\right) \le 1 - \dfrac{\sigma^2}{n \cdot \varepsilon^2}$; $D(X) \le \sigma^2 = 20$; $n = 2000/10 = 200.$

1) При $\varepsilon = 1ц/га$ имеем: $P\left(\left|\overline{X}_в - \overline{X}_г\right| < 1\right) \le 1 - \dfrac{20}{200 \cdot 1^2} = 0{,}9.$

После преобразований получаем: $P\left(\left|\overline{X}_в - \overline{X}_г\right| < 1\right) \le 0{,}9$ (*).

2) При $\varepsilon = 0{,}5ц/га$ имеем: $P\left(\left|\overline{X}_в - \overline{X}_г\right| < 0{,}5\right) \le 1 - \dfrac{20}{200 \cdot 0{,}5^2} = 0{,}6.$

После преобразований получаем: $P\left(\left|\overline{X}_в - \overline{X}_г\right| < 0{,}5\right) \le 0{,}6$ (**).

- Анализируя ответы (*) и (**) можно сделать вывод, что чем меньше предельная ошибка ε, тем с меньшей вероятностью можно утверждать, что абсолютная величина отклонения выборочной средней от генеральной средней не превысит эту ошибку.

6.3. Задачи для самостоятельного решения.

1. Составьте дифференциальную функцию для нормально распределенной случайной величины, если заданы:

а) *M(x)*=4; *σ(x)*=0,2; б) *M(x)*=3; *Д(x)*=1/16; в) *M(x)*=0; *σ(x)*=1

Постройте график этой функции в последнем случае.

2. Случайная величина *x* имеет плотность распределения:

$$f(x) = \frac{2}{3\sqrt{2\pi}} e^{\frac{2(x-30)^2}{3}}$$

Найдите математическое ожидание и дисперсию.

3. Урожай картофеля (ц/га) есть нормально распределенная случайная величина со средним значением 100 ц/га и дисперсией 225. Определите вероятность того, что урожай, с наудачу взятого га, примет значение в промежутке от 90 ц/га до 110 ц/га.

4. Процент содержания примесей в партии муки - случайная величина, распределенная по нормальному закону, где $M(x)$ = 16 %, $\sigma(x)$ = 4 %. Определите вероятность того, что в наудачу взятой пробе примесей будет от 10 % до 25 %.

5. Нормально распределенная случайная величина x имеет дисперсию равную 0,16. Найдите вероятность того, что значения этой случайной величины отклоняются от ее математического ожидания на величину меньшую 0,2.

6. Средняя глубина посева семян составляет 4 см; отдельные отклонения от этого значения распределены нормально со средним квадратическим отклонением 0,6 см. Определите:

1) вероятность того, что значения этой случайной величины отклоняются от ее математического ожидания по абсолютной величине на величину большую 0,4; 2) долю семян, посеянных на глубину менее 3 см.

7. Вероятность того, что нормально распределенная случайная величина отклоняется от своего математического ожидания по абсолютной величине меньше, чем на 5,6, равна 0,7272. Найдите среднее квадратическое отклонение.

8. Средний вес одного яблока равен 130 г, отклонение в весе характеризуется средним квадратическим отклонением 25г. Отбирается подряд без выбора - любое яблоко. Определите:

1) вероятность того, что вес выбранного яблока окажется не менее 150 г;

2) величину, которую не превзойдет вес яблока с вероятностью 0,98.

9. Случайная величина x - масса одного зерна - распределена нормально. Математическое ожидание массы одного зерна равна 0,18 г, среднее квадратическое отклонение 0,05 г. Хорошие всходы дают зерна, масса которых больше чем 0,15 г.

Найдите: а) процент семян, которые дадут хорошие всходы; б) доверительный интервал, в котором с вероятностью 0,9545 будет заключен вес зерна; в) весь диапазон в котором будет заключен вес любого зерна.

10*. Методом проб установлено, что потери зерна при уборке составляют 3 г на 1м^2; среднее квадратическое отклонение равно 1 г. Найдите: 1) вероятность того, что на 1 га потери составят не менее чем 29,8 кг; 2) величину, которую с вероятностью 0,99 не превышает потери на 1 га. Считать, что x (потери зерна) есть нормально распределенная случайная величина.

---

1. Составьте дифференциальную функцию для нормально распределенной случайной величины, если даны ее параметры:

а) $M(x) = -0{,}5$; $Д(x)=4$

б) $M(x)= 2$; $\sigma(x) = 3/5$

в) $M(X)=1$; $\sigma(x)= 0{,}5$.

2. Урожайность яровой пшеницы (ц/га) является нормально распределенной случайной величиной с плотностью распределения:

$$f(x) = \frac{2}{3\sqrt{2\pi}} e^{\frac{2(x-30)^2}{3}}$$

Найдите *М(х), Д(х)* урожайности. Постройте график *f(x).*

3. Вес яиц в некоторой партии имеет нормальное распределение с параметрами $a=60$ г и $\sigma=6г$. В заготовку принимаются яйца от 51 г до 66 г. Определите процент яиц, идущих в заготовку.

4. Размер плода - случайная величина, распределенная нормально; математическое ожидание равно 5 см; среднее квадратическое отклонение 0,8 см. Определите: 1) вероятность того, что размер любого наудачу выбранного плода, будет заключен в интервале от 4,5 до 5,5 см; 2) вероятность того, что размер любого плода отклоняется от среднего размера по абсолютной величине на величину, меньшую 6,5 см; 3) процент плодов, имеющих размер, свыше 4,5 см.

5. Высота стебля кукурузы в определенный срок является случайной величиной, распределенной по нормальному закону, где $M(x) = 175$ см; $\sigma(x) = 6$ см. В каких границах в соответствии с правилом «трех сигм» можно гарантировать высоту стебля кукурузы.

6*.Средняя масса плодов в одном ящике равна 10 кг, а среднее квадратическое отклонение в массе плодов одного ящика 1,5 кг.

Найдите: 1) вероятность того, что в 100 ящиках масса плодов окажется не менее 970 кг; 2) наибольшее значение, которое с вероятностью 0,95 не превзойдет масса 100 ящиков. Учитывать, что масса плодов в одном ящике - нормально распределенная случайная величина.

Типовой расчет «Случайная величина».

1. Варианты заданий для студентов очной формы обучения (назначаются преподавателем).

№ варианта	Зад. 1.	Зад. 2.	Зад. 3.	Зад. 4.
1	1	11	21	31
2	2	12	22	32
3	3	13	23	33
4	4	14	24	34
5	5	15	25	35
6	6	16	26	36
7	7	17	27	37
8	8	18	28	38
9	1	13	25	31
10	7	15	21	33
11	3	17	23	34
12	4	12	26	32
13	5	11	22	37
14	6	16	24	35
15	7	18	25	32
16	2	14	21	33
17	3	12	23	33
18	8	17	28	38
19	9	19	29	39
20	10	20	30	40

2. Варианты заданий для студентов заочной формы обучения (номер варианта соответствует последней цифре шифра зачетной книжки).

№ варианта	**Зад. 1.**	**Зад. 2.**
1	11	31
2	12	32
3	13	33
4	14	34
5	15	35
6	16	36
7	17	37
8	18	38
9	19	39
0	20	40

Задание 1 (биномиальный закон распределения - типовая задача 5.1)

№ 1-10. Предположим, что производится обработка стада животных дезинфицирующим составом против заболевания А, вероятность события = заболевание ликвидировано равно ***p***. Из стада после обработки отбирается ***n*** животных.

Требуется:

1) составить закон распределения числа здоровых животных среди ***n*** отобранных; 2) построить многоугольник распределения; 3) найти вероятности событий: а) *A* - среди ***n*** животных будет не более k_1 здоровых;б) *B* - не менее k_2 здоровых; в) *C* - от k_1 до k_2 (включительно) здоровых; 4) сколько здоровых животных вероятнее всего будет среди ***n*** отобранных; 5) найти числовые характеристики *M(x), Д(x), σ(x), υ(x).*

№	n	*p*	k_1	k_2
1	5	0,8	2	4
2	4	0,9	2	3
3	6	0,7	3	5

4	5	0,6	1	3
5	5	0,75	2	4
6	4	0,8	1	3
7	6	0,9	3	6
8	4	0,95	2	3
9	5	0,8	1	2
10	6	0,65	3	5

Задание 2.(числовые характеристики дискретной случайной величины - типовая задача 5.2, интегральная функция распределения дискретной случайной величины – типовая задача 5.1).

Задание №11 – 20. Закон распределения дискретной случайной величины X задан в виде таблицы.

I. Найти: 1) математическое ожидание $M(X)$;2) дисперсию $D(X)$ (два способа); 3) среднее квадратическое отклонение $\sigma(X)$;4) коэффициент вариации; 5) построить многоугольник распределения/

II. Составить интегральную функцию распределения. Построить график.

11.	20	25	30	35	40		16.	10	20	30	40	50
	0.2	0.3	0.2	0.1	0.2			0.1	0.3	0.2	0.1	0.3
12.	15	25	35	45	55		17.	10	30	50	70	90
	0.2	0.3	0.1	0.2	0.2			0.3	0.1	0.2	0.1	0.3
13.	20	30	40	50	60		18.	15	35	55	75	95
	0.1	0.4	0.1	0.3	0.1			0.1	0.3	0.1	0.4	0.1
14.	10	35	50	75	90		19.	5	25	50	75	80
	0.2	0.3	0.1	0.3	0.1			0.1	0.4	0.1	0.3	0.1

15.	40	50	60	70	80		20.	35	50	75	80	90
	0.1	0.2	0.1	0.5	0.1			0.1	0.5	0.1	0.2	0.1

Задание 3.(непрерывная случайная величина – типовая задача 5.2).

№21-30. Непрерывная случайная величина *X* задана функцией распределения *F(x)*. Найти 1) плотность распределения *f(x);* 2)математическое ожидание; 3) дисперсию и среднее квадратичное отклонение; 4)вероятность попадания случайной величины в интервал $(\alpha;\beta)$;5) построить графики функции распределите *F (x)* и плотности распределения *f(x)*.

21	$F(x)=\begin{cases} 0 \ \ при \ x\le 1 \\ \frac{1}{4}(x-1) \ при \ 1<x\le 5 \\ 1 \ \ при \ x>5 \end{cases}$ (0;2)	26	$F(x)=\begin{cases} 0 \ \ при \ x\le 0 \\ x^2/9 \ при \ 0<x\le 3 \\ 1 \ \ при \ x>3 \end{cases}$ (2;4)
22	$F(x)=\begin{cases} 0 \quad при \ x<0 \\ \frac{1}{4}x \ при \ 0\le x<4 \\ 1 \quad при \ x\ge 4 \end{cases}$ (3;5)	27	$F(x)=\begin{cases} 0 \quad при \ x<3 \\ (x-3)^2 \ при \ 3\le x<4 \\ 1 \quad при \ x\ge 4 \end{cases}$ (1;3,5)
23	$F(x)=\begin{cases} 0 \ \ при \ x\le 0 \\ x^2/64 \ при \ 0<x\le 8 \\ 1 \ \ при \ x>8 \end{cases}$ (–2;4)	28	$F(x)=\begin{cases} 0 \ \ при \ x\le 0 \\ x^2/36 \ при \ 0<x\le 6 \\ 1 \ \ при \ x>6 \end{cases}$ (–3;3)
24	$F(x)=\begin{cases} 0 \quad при \ x<2 \\ x-2 \ при \ 2\le x<3 \\ 1 \quad при \ x\ge 3 \end{cases}$ (2,5;4)	29	$F(x)=\begin{cases} 0 \quad при \ x<0 \\ 0.5x \ при \ 0\le x<2 \\ 1 \quad при \ x\ge 2 \end{cases}$ (1;–1)

25	$F(x)=\begin{cases} 0 \ при\ x\le -1 \\ \frac{1}{3}(x+1)\ при\ -1<x\le 2 \\ 1\ при\ x>2 \end{cases}$ (0;2)	30	$F(x)=\begin{cases} 0\ при\ x\le 0 \\ x^2/25\ при\ 0<x\le 5 \\ 1\ при\ x>5 \end{cases}$ (4;7)

Задание 3.(нормальный закон распределения – типовая задача 6.1).

№ 31 – 35. Жирность молока коров на фермерских хозяйствах края (в процентах) есть нормально распределенная величина с математическим ожиданием, равным ***a%*** и ***σ %***. Требуется:

а) найти дифференциальную функцию распределения, построить ее график;

б) определить процент коров в крае, у которых жирность молока заключена в пределах от ***c %*** до ***d %;***

в) определить, с какой вероятностью можно гарантировать, что предельное отклонение жирности молока каждой коровы от математического ожидания будет: 1) не более ***ε%;*** 2) более ***ε%;***

г) найти какое предельное отклонение в ту или другую сторону жирности молока от математического ожидания можно гарантировать с вероятностью ***δ*** ;

д) определить весь диапазон изменения жирности молока по всем хозяйствам края.

№ вар.	*a%*	*σ %*	*c%*	*d%*	*ε%*	*δ*
31	3,8	0,15	3,7	4,2	0,5	0,99
32	4	0,13	3,8	4,5	0,4	0,95
33	3,5	0,17	3,2	3,7	0,5	0,99
34	3,7	0,11	3,5	4	0,3	0,68
35	3,2	0,14	3,1	3,4	0,2	0,95

№ 36 – 40. На молочной ферме испытывали эффективность кормов. При скармливании смеси №1 получили средний удой ***a*** л, ***σ*** л. Считая удой коров случайной величиной подчиняющейся нормальному закону распределения требуется:

а) найти дифференциальную функцию распределения, построить ее график;

б) определить процент коров, у которых надои заключены в пределах от ***c*** до ***d*** литров;

в) определить, с какой вероятностью можно гарантировать, что предельное отклонение удоев на ферме от математического ожидания будет:

1) не более ε л; 2) более ε л;

г) найти какое предельное отклонение в ту или другую сторону от среднего значения можно гарантировать с вероятностью δ;

д) определить весь диапазон изменения удоев на ферме.

№ вар.	$a_л$	$\sigma_л$	$c_л$	$d_л$	$\varepsilon_л$	δ
36	11	4,4	3	6	4	0,99
37	12	5	4	7	5	0,95
38	10	3,8	4	7	3,5	0,98
39	13	5,2	5	7	5	0,68
40	11	4,6	3	6	4	0.95

РАЗДЕЛ 3. МАТЕМАТИЧЕСКАЯ СТАТИСТИКА.

Глава 7. Исследование вариационных рядов.

7.1. Необходимый теоретический минимум.

7.1. Предмет и задачи математической статистики.

Математической статистикой называется наука, занимающаяся разработкой методов получения, описания и обработки опытных данных с целью изучения закономерностей случайных массовых явлений.

Исходным материалом для статистического исследования реального явления служит набор результатов наблюдений над ними или же результатов специально поставленных испытаний. Вопросов, которые при этом возникают, очень много. Укажем некоторые из них.

1. Оценка значения неизвестной вероятности случайного события.
2. Определение неизвестной функции распределения.

Задача ставится так: в результате n независимых испытаний над случайной величиной X получены следующие ее значения: $x_1, x_2, \ldots, x_j, \ldots, x_n$. Требуется определить, хотя бы и приближенно, неизвестную функцию распределения $F(x)$ величины X.

3. Определение неизвестных параметров распределения.

Общая задача ставится так: случайная величина X имеет функцию распределения определенного вида, зависящую от k параметров, значения которых неизвестны. На основании последовательных наблюдений величины X нужно найти значения этих параметров. В частности, в случае нормальной функции распределения неизвестными являются параметры a и σ.

4. Проверка статистических гипотез.

На основании некоторых соображений можно считать, что функция распределения случайной величины X есть $F(x)$; спрашивается, совместимы ли наблюдаемые значения с гипотезой, что X действительно имеет распределение $F(x)$. В качестве другого примера статистической гипотезы можно привести проверку однородности статистического материала (например, оценить правдоподобность гипотезы о равенстве двух неизвестных функций распределения).

5. Оценка зависимости.

Производится последовательность наблюдений сразу двух случайных величин X и Y. Требуется определить наличие функциональной или корреляционной связи между ними.

7.2. Понятие вариационного ряда. Виды вариационных рядов. Геометрическое изображение.

Определение 7.1. Генеральной совокупностью называется совокупность всех возможных однородных предметов или явлений, над которыми проводятся наблюдения, или совокупность всех возможных наблюдений, проводимых над некоторой случайной величиной в одинаковых условиях.

Определение 7.2. Выборочной совокупностью (выборкой) называется совокупность предметов или явлений, отобранная из соответствующей генеральной совокупности.

Определение 7.3. Объемом совокупности (генеральной или выборочной) называется общее число ее элементов.

- Если N и n – соответственно объемы генеральной и выборочной совокупности, то, очевидно, что $N>n$, при n < 30 выборка называется ***малой***, при n > 30 ***большой.***
- *Выборочный метод* заключается в том, что из генеральной совокупности берется выборка значительно меньшего объема и определяются характеристики выборки, которые принимаются в качестве приближенных значений соответствующих характеристик генеральной совокупности.

Выборки бывают:

а) ***повторные*** – если отобранный объект (перед отбором следующего) возвращается в генеральную совокупность;

б) ***бесповторные*** – если отобранный объект не возвращается в генеральную совокупность.

Пусть из генеральной совокупности извлечена выборка X объемом *n*. Случайный *выбор элемента* рассматривается как независимое наблюдение над величиной ξ, имеющей некоторое распределение вероятностей. Если те значения, которые приняла случайная величина ξ в *n* наблюдениях, записать не в порядке получения, а в порядке возрастания (то есть ***ранжируя***), то получим упорядоченную выборку $x_1, x_2, \ldots, x_n$,, называемую ***вариационным рядом***.

Определение 7.4. Вариантой называется значение x_i случайной величины, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных.

Определение 7.5. Размахом R вариационного ряда называется разность между его наибольшей $x_{\max}$ и наименьшей $x_{\min}$ вариантами:

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

• Если изучается дискретная случайная величина, то при достаточно большом объеме выборки в ней могут содержаться повторяющиеся значения. Для каждого такого значения можно подсчитать, сколько раз оно встретилось в ряде наблюдений.

Определение 6.6. Частотой (весом) варианты называется численность отдельной группы сгруппированного ряда наблюдаемых данных.

- Если i – индекс варианты, то f_i – число измеренных значений i-ой варианты.

Определение 6.7. Относительной частотой варианты x_i называется отношение f_i числа повторения x_i к объему выборки n:

$$\omega_i = \frac{f_i}{n}. \tag{7.1}$$

• Общая сумма частот вариационного ряда равна объему данной совокупности, т. е.

$$\sum_{i=1}^{k} f_i = n,$$

где k – число групп, n – общее число наблюдений, или объем совокупности.

• Общая сумма частностей равна единице

$$\sum \frac{f_i}{n} = 1 \text{ или } \sum \frac{f_i}{n} \cdot 100 = 100\%$$

Данные наблюдений, среди которых много повторяющихся изображают в виде таблицы (таб. 7.1). Такая таблица называется ***таблицей статистического распределения.***

Таблица 7.1

Значения x_i	x_1	x_2	…	x_k
Частоты m_i	f_1	f_2	…	f_k
Относительные частоты	$\omega_1 = \frac{f_1}{n}.$	$\omega_2 = \frac{f_i}{n}.$		$\tilde{\omega}_k = \frac{f_i}{n}.$

Пример. Выборка оценок десяти студентов следующая:

5; 4; 3; 3; 4; 3; 3; 5; 4; 3; 3.

Составить статистический ряд, сделать контроль.

Решение. Расположив эти значения в порядке неубывания, получим следующий ряд: 3; 3; 3; 3; 3; 4; 4; 4; 5; 5. Значения 3; 4; 5, принятые случайной величиной в процессе наблюдения (k =3), являются вариантами, а их количества 5; 3; 2 – весы (частоты) соответствующих вариант. В итоге получим вариационный ряд:

Значения x_i	3	4	5
Частоты f_i	5	3	2
Относительные частоты	5/10	3/10	2/10

Контроль: $\sum_{i=1}^{3} f_i = 5 + 3 + 2 = 10.$

Определение 7.8. ***Дискретным вариационным рядом*** распределения (***распределением частот*** или ***относительных частот***) называется ранжированная совокупность вариант x_i с соответствующими им частотами или относительными частотами.

Алгоритм составления дискретного вариационного ряда:

– найти минимальное ($x_{\min}$) и максимальное ($x_{\max}$) значения выборки;

– в первый столбец таблицы записать варианты значений случайной величины (генеральной совокупности), начиная с $x_{\min}$ и заканчивая $x_{\max}$;

– просмотреть по одному все элементы выборки в протоколе наблюдений, и подсчитать количество значений, соответствующих каждой варианте (то есть f_i);

– подсчитать количество элементов выборки (ее объем).

• Если изучается случайная величина, имеющая непрерывное распределение вероятностей, то возможные значения заполняют целый интервал или всю числовую ось. В этом случае, скорее всего, вариационный ряд не будет содержать повторяющихся значений. То же самое может иметь место, если наблюдение производится над дискретной случайной величиной, число возможных значений которой очень велико.

Определение 7.9. ***Интервальным вариационным рядом (интервальным распределением частот*** или ***относительных частот)*** называется упорядоченная последовательность интервалов варьирования случайной величины с соответствующими частотами или относительными частотами попаданий в каждый из них значений случайной величины.

Алгоритм составления интервального вариационного ряда:

– найти минимальное ($x_{\min}$) и максимальное ($x_{\max}$) значения выборки;

– найти объем n выборки;

– определить оптимальное число интервалов по *формуле Стерджеса*:

$$k = 1 + 3{,}322 \cdot \lg n. \qquad (7.2)$$

– найти длину интервала по формуле:

$$h = \frac{x_{\max} - x_{\min}}{k}. \qquad (7.3)$$

– заполнить первый столбец таблицы интервалами исследуемой совокупности. За начало первого интервала рекомендуется брать величину:

$$x_{нач} = x_{\min} \qquad (7.4)$$

– просмотреть по одному все элементы выборки в протоколе наблюдений, и подсчитать количество значений, соответствующих каждому интервалу (то есть f_i).

Определение 7.10. Накопленной частотой $S_x^{нак}$ в точке x называют суммарную частоту членов статистической совокупности со значениями

• Каждая генеральная совокупность имеет функцию распределения *F(x)* , которая обычно неизвестна. По выборке можно найти эмпирическую функцию распределения $F^*(x)$, для которой (на основании закона больших чисел) вместо вероятностей p_i берутся относительные частоты p_i/m.

Определение 7.11. Эмпирической функцией распределения (выборочной функцией распределения) называется функция $F^*(x)$, определяющая для каждого значения x относительную частоту события $X<x$:

$$F^*(x) = \tilde{P}(X < x) = S_x^{нак}. \qquad (7.5)$$

- Значениями эмпирической функции распределения являются накопленные частости.
- Эмпирическая функция распределения выборки служит для оценки теоретической функции распределения *F(x)* генеральной совокупности.
- Выборочную функцию распределения можно задавать таблично или графически.

Графическое представление статистического ряда.

Для наглядного представления вариационного ряда (или статистического распределения) пользуются графическим изображением вариационных рядов (полигоном, гистограммой, кумулятой).

Интервальный ряд изображается столбиковой диаграммой, в которой основания столбиков, расположенные по оси абсцисс, – это интервалы значений варьирующего признака, а высоты столбиков – частоты, соответствующие масштабу по оси ординат. Диаграмма такого типа называется ***гистограммой.***

Если имеется дискретный ряд распределения или используются середины интервалов, то графическое изображение такого ряда называется ***полигоном,*** которое получается соединением прямыми точек с координатами X_i и f_i.

Если по оси абсцисс откладывать значения классов, а по оси ординат – накопленные частоты с последующим соединением точек прямыми линиями, получается график, называемый ***кумулятой.*** Накопленные частоты находят последовательным суммированием, или кумуляцией частот в направлении от первого класса до конца вариационного ряда.

Пример. Имеются данные о яйценоскости 50 кур-несушек за 1 год, содержащихся на птицеферме (табл. 7.3).

Т а б л и ц а 7.3.

Яйценоскость кур-несушек

№ курицы-несушки	Яйценос кость, шт.	№ курицы-несушки	Яйценос кость, шт.	№ курицы-несушки	Яйценос кость, шт.	№ курицы-несушки	Яйценос кость, шт.	№ курицы-несушки	Яйценос кость, шт.
1	212	11	221	21	227	31	231	41	237
2	214	12	222	22	228	32	232	42	238
3	217	13	223	23	228	33	232	43	238
4	217	14	223	24	228	34	232	44	238
5	217	15	226	25	229	35	232	45	239
6	218	16	226	26	229	36	234	46	240
7	219	17	226	27	229	37	235	47	241

8	220	18	227	28	231	38	235	48	241
9	221	19	227	29	231	39	235	49	242
10	221	20	227	30	231	40	237	50	245

Требуется построить интервальный ряд распределения и отобразить его графически в виде гистограммы, полигона и кумуляты.

Решение.

Видно, что признак варьирует от 212 до 245 яиц, полученных от несушки за 1 год.

В нашем примере по формуле Стерждесса определим число групп:

$k = 1 + 3{,}322 lg 50 = 6{,}643 \approx 7$.

Рассчитаем длину (размах) интервала по формуле:

$$h = \frac{245 - 212}{7} = 4{,}71 \approx 5.$$

Построим интервальный ряд с 7 группами и интервалом 5 шт. яиц (табл. 7.4). Для построения графиков в таблице рассчитаем середину интервалов и накопленную частоту.

Т а б л и ц а 7.4.

Интервальный ряд распределения яйценоскости

i	Группа кур-несушек по величине яйценоскости X_i	Число кур-несушек f_i	Середина интервала x_i'	Накопленная частота $S_{нак}$
1	212-217	5	214,5	5
2	217-222	7	219,5	12
3	222-227	9	224,5	21
4	227-232	14	229,5	35
5	232-237	6	234,5	41
6	237-242	8	239,5	49
7	242-247	1	244,5	50
Итого		50	-	-

Построим гистограмму распределения яйценоскости (рис. 7.1).

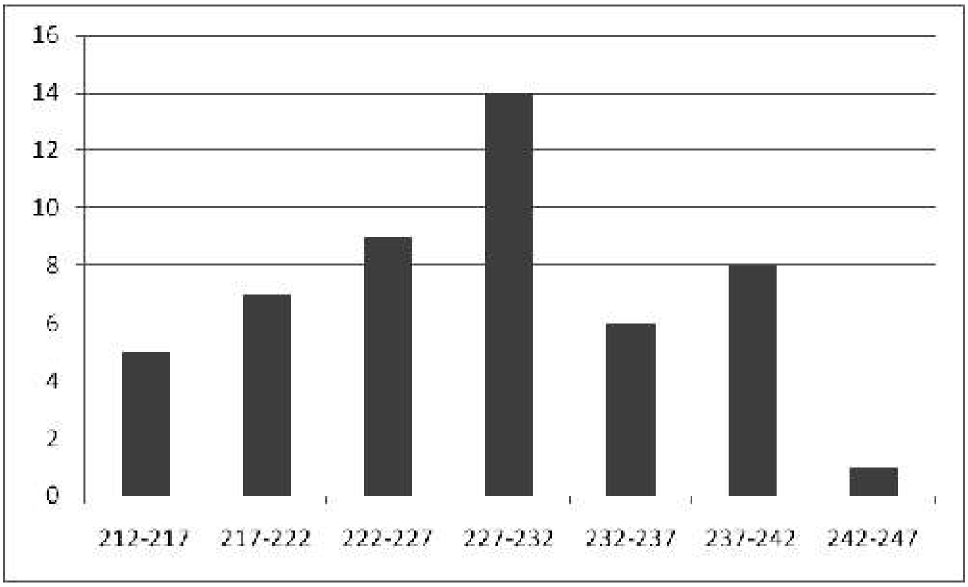


Рисунок 7.1. Гистограмма распределения яйценоскости

Данные гистограммы показывают характерную для многих признаков форму распределения: чаще встречаются значения средних интервалов признака, реже – крайние (малые и большие) значения признака. Форма этого распределения близка к нормальному закону распределения, которое образуется, если на варьирующую переменную влияет большое число факторов, ни один из которых не имеет преобладающего значения. Обратите внимание на то, что один из интервалов имеет наибольшее значение вариант анализируемого признака, такой интервал называется *модальным.*

Если середины верхушек столбиков гистограммы обозначить точками ,а затем соединить линиями, получится полигон распределения (рис.7.1.). Полигон и кумулята распределения яйценоскости имеют вид (рис. 7.2. и 7.3).

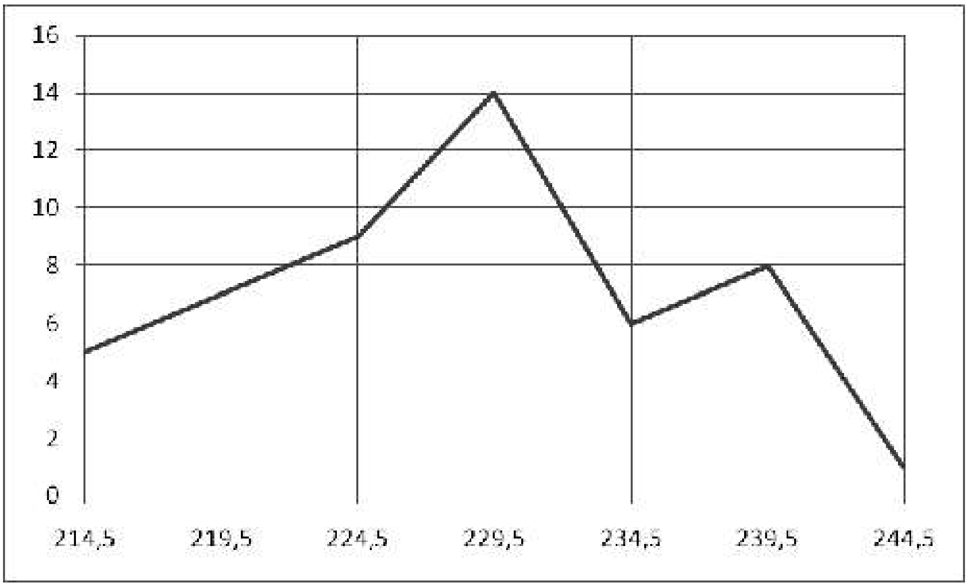


Рисунок 7.2. Полигон распределения яйценоскости

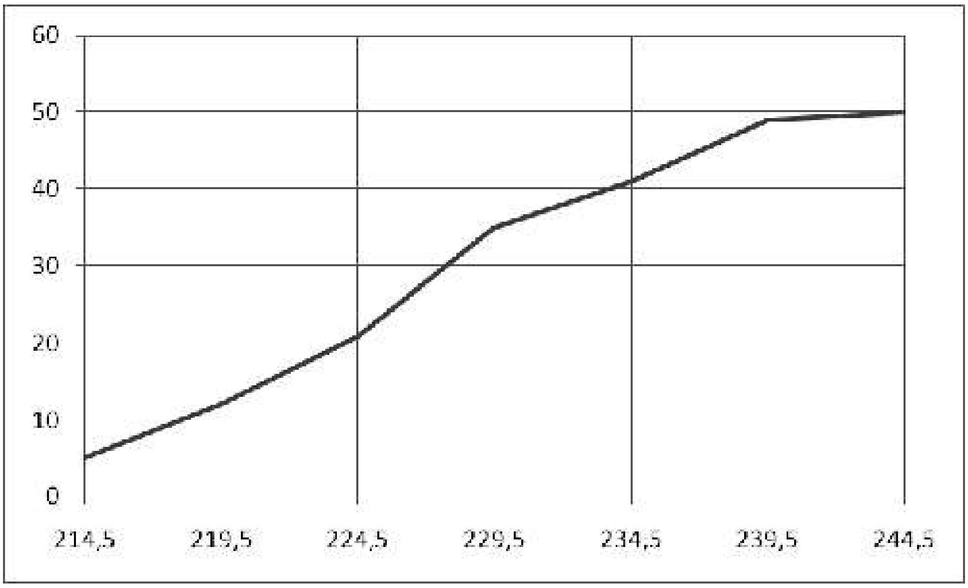


Рисунок 7.3. Кумулята распределения яйценоскости

7.3.Числовые характеристики выборочной совокупности.

Для того, чтобы количественно охарактеризовать самые существенные свойства распределения, а также для того, чтобы можно было сравнить разные распределения, вычисляют средние показатели - выборочные числовые характеристики.

7.3.1. Характеристики уровня вариационных рядов.

В статистике используются различные величины в зависимости от того, какие цели при анализе материала ставит исследователь. Понятием средней величины пользуемся в тех случаях, когда требуется определить средний надой по стаду, средний привес, средний прирост стада, средние клинические показатели деятельности сердца, лёгких, среднего состава крови и во многих других случаях.

1) Выборочная средняя (средняя арифметическая).

Средняя арифметическая является наиболее распространенной среди средних величин. Ее применяют в тех случаях, когда даны отдельные объекты с индивидуальными значениями признаков, выраженными абсолютными показателями. Среднюю арифметическую определяют как отношение суммы индивидуальных значений признаков к их количеству.

Различают среднюю арифметическую простую и взвешенную. Среднюю арифметическую простую применяют в случае, если индивидуальные значения признака в совокупности встречаются по одному разу (*несгруппированный ряд*), а взвешенную – если индивидуальные значения признака представлены несколькими объектами (*сгруппированный ряд*).

- средняя арифметическая (выборочная средняя) простая:

$$\bar{x}_в = \frac{\sum x_i}{n}, \qquad (7.6)$$

где $\bar{x}_в$ – выборочная средняя; x_i –варианты; n–число вариант.

- средняя арифметическая взвешенная:

$$\bar{x}_в = \frac{\sum x_i f_i}{n}, \qquad (7.7)$$

где f–частота вариант.

- Если ряд интервальный, то при вычислении среднего арифметического в роли x_i в формулах (7.6) и (7.7) представителем каждого интервала берется его середина.
- Если среднее арифметическое находится для генеральной совокупности, то его называют генеральной средней и обозначают $a = \bar{x}_г$.

Рассмотрим методику вычисления простой выборочной средней.

Пример. Имеются данные по 8 коровам об их удое за год (табл. 7.5).

Т а б л и ц а 7.5.

Удой коровы

№ коровы	Удой коровы за год, кг
	x
1	3811
2	3600
3	3480
4	3758
5	4059
6	4222
7	4160
8	3971
Итого	$\sum x = 31061$

Требуется определить средний удой на одну корову за год.

Так как даны индивидуальные значения удоя молока по каждой корове, то средний удой определяется по формуле средней арифметической простой:

$$\bar{x}_в = \frac{\sum x_i}{n} = \frac{31061}{8} = 3883 \text{ кг.}$$

Таким образом, среднегодовой удой от коровы за год составляет 3883 кг.

Пример. Имеются данные о продуктивности норок (табл. 7.6). Требуется определить средний приплод норок.

Т а б л и ц а 7.6.

Приплод норок

Приплод на самку, гол.	Численность норок, гол.	Накопленные частоты
x	f	s
1	9	9
2	38	47
3	81	128
4	164	292
5	75	367
6	47	414
7	17	431
Итого	$\sum f = 431$	×

Т.к. значение вариант повторяются, то имеем дискретный сгруппированный ряд, тогда среднее значение определяем по формуле (7.7):

$$\bar{x}_в = \frac{\sum x_i f_i}{n} = \frac{9\cdot 1 + 2\cdot 38 + 3\cdot 81 + 4\cdot 164 + 5\cdot 75 + 6\cdot 47 + 7\cdot 17}{431} \approx 4$$

2) Мода и медиана

Средние величины, описанные выше, являются обобщающими характеристиками совокупности по тому или иному признаку. Вспомогательными характеристиками являются, так называемые, структурные средние, к которым относятся мода, квартили, децили, медиана и др. Наиболее употребляемыми являются мода и медиана.

Мода – это величина, которая встречается в совокупности наиболее часто, то есть признак с наибольшей частотой. Этот показатель используется в тех случаях, когда требуется охарактеризовать наиболее часто встречающуюся величину признака (наиболее распространенный размер животноводческих ферм на сельскохозяйственных предприятиях, преобладающие цены на сельскохозяйственную продукцию и т. п.).

а) В дискретном вариационном ряду модой является признак с наибольшей частотой в примере

б) В интервальном вариационном ряду моду находят по формуле:

$$Mo = x_{Mo} + h_{Mo}\frac{f_{Mo} - f_{Mo-1}}{2f_{Mo} - f_{Mo-1} - f_{Mo+1}}, \qquad (7.9)$$

где Mo – мода;

x_{Mo}–нижняя граница модального интервала;

h_{Mo} – величина модального интервала;

f_{Mo}– частота модального интервала;

f_{Mo-1}– частота интервала, предшествующего модальному;

f_{Mo+1}– частота интервала, следующего за модальным.

Модальным интервалом является интервал с наибольшей частотой.

Медианой называется величина, делящая численность упорядоченного вариационного ряда (расположенного в порядке возрастания или убывания признака) на две равные части. Медиана характеризует количественную границу значений изменяющегося признака, которыми обладает половина единиц совокупности. Например, если медианное значение удоя коровы составляет 4735 кг, то это означает, что половина коров имеет удой молока ниже 4735 кг и половина коров выше.

а) Формула расчета медианы для дискретного ряда:

$$M_e = \begin{cases} \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right), & \textit{если } n - \textit{четное}, \\ x_{\frac{n+1}{2}}, & \textit{если } n - \textit{нечетное} \end{cases} \quad (7.10)$$

Формула расчета медианы в интервальном вариационном ряду:

$$Me = x_{Me} + h_{Me} \frac{\frac{\sum f}{2} - s_{Me-1}}{f_{Me}}, \quad (7.11)$$

где Me – медиана;

x_{Me}– нижняя граница медианного интервала;

h_{Me}– величина медианного интервала;

$\sum f$ – сумма частот;

s_{Me-1} – сумма частот, накопленных в интервалах, предшествующих медианному;

f_{Me}– частота медианного интервала.

Медианным интервалом является интервал, накопленная частота которого равна или превышает половину суммы частот.

Рассмотрим методику расчета средней взвешенной моды и медианы для дискретного сгруппированного ряда.

Пример. Определим моду и медиану для данных в таблице (7.6).

Самую большую частоту – 164 имеют норки с приплодом в 4 головы, следовательно, мода равна 4.

Медианой будет признак с номером $\frac{431}{2} + 0{,}5 = 216$. Из накопленных частот видно, что медианой будет норка, имеющая приплод в 4 головы.

Пример. Имеются данные о среднесуточных приростах живой массы у молодняка крупного рогатого скота (табл. 6.7).

Т а б л и ц а 7.7.

Среднесуточный прирост живой массы ремонтных телок на откорме

Группы скота по среднесуточному приросту живой массы, г	Поголовье скота, гол.	Накопленные частоты
	f	s
500–525	21	21
525–550	36	57
550–575	98	155
575–600	143	298
600–625	136	434
625–650	100	534
650–675	86	620
675–700	10	630
Итого	$\sum f = 630$	×

Требуется определить моду и медиану.

Моду и медиану рассчитывают по формулам (7.9), (7.11) для интервального вариационного ряда.

Для нахождения моды необходимо определить модальный интервал. Таким будет интервал 575–600 с наибольшей частотой 143. Отсюда мода равна:

$$Mo = x_{Mo} + h_{Mo}\frac{f_{Mo} - f_{Mo-1}}{2f_{Mo} - f_{Mo-1} - f_{Mo+1}} = 575 + 25\cdot\frac{143-98}{2\cdot143-98-136} = 597 \text{г.}$$

Для нахождения медианы надо определить медианный интервал. Половина суммы частот равна 315 (630:2). Следовательно, согласно накопленным частотам медианным интервалом будет 600 – 625 (315 < 434).

Медиана равна: $$Me = x_{Me} + h_{Me}\frac{\frac{\sum f}{2} - s_{Me-1}}{f_{Me}} = 600 + 25\cdot\frac{\frac{630}{2} - 298}{143} = 603 \text{ г.}$$

7.3.2. Показатели колеблемости вариационных рядов.

Для измерения вариации применяют различные показатели, из которых основными являются размах вариации (лимит), среднее линейное отклонение, дисперсия, среднее квадратическое отклонение, коэффициент вариации.

1) *Размах вариации* определяется как разница между наибольшим и наименьшим значениями признака:

$$R = x_{\max} - x_{\min},$$

где $x_{\min}$, x_{max}– минимальное и максимальное значение признака. Чем больше разность между x_{max} и $x_{\min}$, тем, в общем, сильнее изменчивость признака.

2) Среднее линейное отклонение представляет собой среднюю арифметическую из абсолютных отклонений отдельных вариант от средней арифметической:

- простое $L = \dfrac{\sum |x - \bar{x}_в|}{n}$;

- взвешенное $L = \dfrac{\sum |x - \bar{x}_в| f}{\sum f}$, где L– среднее линейное отклонение; $\bar{x}$ – средняя арифметическая; x–варианты;n– число вариант;f– частоты.

3) Выборочную дисперсию $D_в$ рассчитывают как среднюю арифметическую квадратов отклонений вариант от их выборочной средней:

а) простая (для несгруппированного ряда):

$$D_в = \frac{\sum (x_i - \bar{x}_в)^2}{n} \qquad (7.12)$$

б) взвешенная (для сгруппированного ряда):

$$D_в = \frac{\sum (x_i - \bar{x}_в)^2 f_i}{n} \qquad (7.13)$$

в) упрощенная формула:

$$D_в = \frac{1}{n}\sum_{i=1}^{4} x_i^2 \cdot f_i - \bar{x}_в^2. \qquad (7.14)$$

- Для облегчения вычисления дисперсий используют следующие свойства:

1. Дисперсия не изменится, если все значения признака увеличить или уменьшить на постоянное число.
2. При умножений признака на постоянное число дисперсия умножается на квадрат этого числа.

4) Среднее квадратическое отклонение σ равно корню квадратному из дисперсии:

- простое: $$\sigma = \sqrt{D_в} = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2}{n}} \qquad (7.15)$$

-взвешенное: $$\sigma = \sqrt{D_в} = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2 f_i}{n}} \qquad (7.16)$$

5) Исправленная дисперсия и среднее квадратическое отклонение. Число степеней свободы.

Выборочная дисперсия имеет систематическую ошибку, приводящую к уменьшению дисперсий. Чтобы это устранить, вводят поправку, умножая $D_в$ на $\frac{n}{n-1}$. В результате получают исправленную дисперсию: $S^2 = \frac{n}{n-1} D_в$. При малых выборках S ощутимо отличается от $D_в$, например при $n=2$ имеем $S^2 = 2D_в$.

Подставим $S^2 = \frac{n}{n-1} D_в$ в (7.12), (7.13) получим формулы для вычисления исправленной дисперсий:

$$S^2 = \frac{\sum (x_i - \bar{x}_в)^2}{n-1} \qquad (7.14) \qquad\qquad S^2 = \frac{\sum (x_i - \bar{x}_в)^2 f_i}{n-1} \qquad (7.15)$$

Формулы для вычисления исправленного среднего квадратического отклонения:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2}{n-1}} \qquad (7.16) \qquad\qquad S = \sqrt{S^2} = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2 f_i}{n-1}} \qquad (7.17)$$

Величина $\nu = n - 1$ в знаменателях формул (7.14 – 17) называется *числом степеней свободы.* Уясним его значение.

Если известен ряд, состоящий из n членов, то для него общей характеристикой является средняя арифметическая. Как может быть определено каждое *отдельное значение* ряда? Очевидно, что конкретное значение признака можно узнать, если известны а) средняя арифметическая и б) остальные наблюдения, т.е. $n-1$. Иначе говоря, определение одного значения в данной совокупности зависит от остальных значений. Так, например, если известно, что 2 кролика в сумме весят 6 кг, а один из них весит 2,5 кг, то вес второго уже точно определен весом первого, т.е. имеется лишь одна степень свободы (2-1 =1). Если 3 кролика весят 5 кг, то вес одного всегда точно определяется весом двух других, между которыми уже возможна вариация, т.е. в этом случае имеется 2 степени свободы вариаций (3-1=2).

В общем виде при численности членов совокупности n число степеней свободы ν будет равно $\nu = n-1$. При большом n разница между n и $(n-1)$, но при малых выборках $(n<30)$ разница будет значительной. Так если $n=6$, а сумма квадратов отклонений равна 60, то средний квадрат отклонений от среднего будет равен не 60:6=10, а 60:5=12.

5) *Коэффициент вариации* представляет собой процентное отношение среднего квадратического отклонения к средней арифметической величине:

$$V = \frac{\sigma_в}{\bar{x}_в} 100\%, \qquad (7.18)$$

Коэффициент вариаций является относительным показателем изменчивости. Изменчивость принято считать незначительной если *V<10%,* средней если *10%<V<20%* , значительной если *V>20%*

Рассмотрим методику расчета показателей вариации для простого дискретного ряда.

Пример. Имеются данные о поголовье бычков, поступивших на мясокомбинат (табл. 7.8).

Необходимо определить показатели колеблемости поголовья бычков по всей совокупности.

Т а б л и ц а 7.8.

Поголовье бычков, поступивших на мясокомбинат

№ партии	Поголовье бычков, гол.	Поголовье бычков	
		абсолютные отклонения	квадрат отклонений
	x	$\lvert x-\bar{x}_в \rvert$	$(x-\bar{x}_в)^2$
1	19	8	64
2	32	5	25
3	31	4	16
4	25	2	4
5	30	3	9
6	32	5	25
7	27	0	0
8	18	9	81
9	26	1	1
10	30	3	9
Итого	$\sum x_i = 270$	$\sum \lvert x-\bar{x}_в \rvert = 40$	$\sum (x-\bar{x}_в)^2 = 234$

Колеблемость поголовья бычков определяется с помощью средней арифметической простой.

Среднее поголовье бычков в партии:

$$\bar{x}_в = \frac{\sum x_i}{n} = \frac{270}{10} = 27 \text{ гол.}$$

Размах вариации поголовья бычков:

$$R = x_{\max} - x_{\min} = 32 - 18 = 14 \text{гол.}$$

Среднее линейное отклонение поголовья бычков от средней:

$$L = \frac{\sum \lvert x_i - \bar{x}_в \rvert}{n} = \frac{40}{10} = 4 \text{ гол.}$$

В задаче имеем малую выборку $n = 10 < 30$, следовательно дисперсию и среднее квадратическое отклонение поголовья бычков будет вычислять по исправленным формулам для дисперсий и среднего квадратического отклонения (7.14), (7.16) :

$$S^2 = \frac{\sum (x-\bar{x}_в)^2}{n-1} = \frac{234}{9} = 26 \qquad S = \sqrt{26} \approx 5 \text{гол.}$$

Коэффициент вариации поголовья бычков:

$$V = \frac{S}{\bar{x}_в}100 = \frac{5}{27} \cdot 100\% = 3\%.$$

Среднее квадратическое отклонение показывает, что поголовье бычков по данной совокупности колеблется в пределах ±5 гол.,а коэффициент вариации равен ±3 % по отношению к среднему уровню, что указывает на незначительную изменчивость.

7.4. Упрощенный подсчет числовых характеристики (метод условной средней)

Для упрощенного подсчета числовых характеристик используют построение вариационного ряда и условную среднюю.

Для упрощения счета взвешенного выборочного среднего и выборочной дисперсий применяют следующие формулы:

$$\bar{x}_в = A + \mathrm{h} \cdot \bar{u}, \qquad (7.19)$$

$$S^2 = \mathrm{h}^2\left(\overline{u^2} - (\bar{u})^2\right) \qquad (7.20)$$

где условное среднее (близкое к ожидаемому среднему, или варианта с максимальной частотой);

$u_i = \frac{x_i - A}{h}$ - условное отклонение средних значений интервалов от выбранной средней условной (условные варианты);

$\bar{u} = \frac{\sum u_i f_i}{n}$ - среднее взвешенное условных вариант;

$\overline{u^2} = \frac{\sum u_i^2 f_i}{n}$ - среднее взвешенное от квадрата условных вариант.

Покажем на примере интервального ряда вычисление числовых характеристик прямым и упрощенным способом.

Пример. Рассмотрим построенный ранее интервальный вариационный ряд распределения яйценоскости кур (стр. 106, таб.7.4)

Интервальный ряд распределения яйценоскости

i	Группа кур-несушек по величине яйценоскости X_i	Число кур-несушек f_i	Середина интервала $x_i^{,}$	Накопленная частота $S_{нак}$
1	212-217	5	214,5	5
2	217-222	7	219,5	12
3	222-227	9	224,5	21
4	227-232	14	229,5	35
5	232-237	6	234,5	41
6	237-242	8	239,5	49
7	242-247	1	244,5	50
Итого		50	-	-

1) Найдем числовые характеристики прямым способом.

Объем выборки равен $n=50$.

Составим расчетную таблицу 1:

1. Средняя взвешенная (итоговую сумму 4 –ого столбца разделим на объем выборки =50).

№	середина интервала $x_i^{,}$	частота f_i	$x_i\, f_i$	$x_i - x_в$	$(x_i - x_в)^2$	$(x_i - x_в)^2\, f_i$
1	**2**	**3**	**4**	**5**	**6**	**7**
1	214,5	5	1072,5	-13,7	187,7	938,4
2	219,5	7	1536,5	-8,7	75,7	529,8
3	224,5	9	2020,5	-3,7	13,7	123,2
4	229,5	14	3213,0	1,3	1,7	23,7
5	234,5	6	1407,0	6,3	39,7	238,1
6	239,5	8	1916,0	11,3	127,7	1021,5
7	244,5	1	244,5	16,3	265,7	265,7
Сумма:			**11410**			**3140,5**

$$\overline{x}_в = \frac{\sum_{i=1}^{n} x_i\, f_i}{n};\qquad \overline{x}_в = \frac{11410}{50} \approx 228{,}2$$

2. Выборочную дисперсию (итоговую сумму 7-ого столбца разделим на объем выборки =50).

Т.к. выборка достаточно большая n=50>30, вычислим дисперсию по формуле, не учитывая исправлений:

$$S^2 = \frac{\sum (x_i - \bar{x}_в)^2 f_i}{n} \qquad S^2 = \frac{\sum 3140{,}5}{50} = 62{,}81$$

3. Среднее квадратическое отконение

$$S_в = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2 \cdot f_i}{n}}; \qquad S_в = \sqrt{62{,}81} \approx 7{,}9$$

4. Коэффициент вариаций

$$V_в = \frac{S_в}{\bar{x}_в} \cdot 100\%; \qquad V_в = \frac{7{,}9}{228{,}2} \cdot 100\% \approx 3{,}47\%$$

Коэффициент вариаций меньше 10%, следовательно, выборка слабо изменчива, однородна.

2) Найдем числовые характеристики упрощенным способом.

Составим расчетную таблицу 2:

№	частота f_i	середина интервала x_i'	u_i	$u_i f_i$	$u_i^2 f_i$
1	**2**	**3**	**4**	**5**	**6**
1	5	214,5	-3	-15	45
2	7	219,5	-2	-14	28
3	9	224,5	-1	-9	9
4	14	229,5	0	0	0
5	6	234,5	1	6	6
6	8	239,5	2	16	32
7	1	244,5	3	3	9
Итого:	**50**	**1606,5**	**0**	**-13**	**129**

1) Вычислим условные варианты по формуле:

$$u_i = \frac{x_i - A}{h}$$

За условную среднюю примем значение варианты с наибольшей частотой *A=229, h=5* (длина интервала), получим:

$$u_1 = \frac{x_1 - A}{h} = \frac{214{,}5 - 229{,}5}{5} = -3 \qquad u_2 = \frac{219{,}5 - 229{,}5}{5} = -2$$

$$u_3 = \frac{224{,}5 - 229{,}5}{5} = -1 \qquad u_4 = \frac{229{,}5 - 229{,}5}{5} = 0$$

Аналогично,получим:

$$u_5 = 1; \qquad u_6 = 2; \qquad u_7 = 1$$

Заполним 4 – ый столбец таблицы.

2)Вычислим среднее взвешенное условных вариант и среднее взвешенное от квадрата условных вариант (итоговые суммы 5-ого и 6-ого столбцов разделим на объем выборки =50,).

$$\overline{u} = \frac{\sum u_i f_i}{n} = \frac{-13}{50} \approx -0{,}26 \qquad \overline{u}^2 = \frac{\sum u_i^2 f_i}{n} = \frac{129}{50} = 2{,}58$$

3)Вычислим числовые характеристики по упрощенным формулам (7.19) и (7.20):

Средняя выборочная:

$$\overline{x}_в = A + \mathrm{h} \cdot \overline{u} = 229{,}5 + 5 \cdot (-0{,}26) = 228{,}2$$

Выборочная дисперсия:

$$S^2 = \mathrm{h}^2 \left(\overline{u}^2 - (\overline{u})^2 \right) = 25 \cdot (2{,}58 - (-0{,}26)^2) = 64 \cdot 3{,}3 = 62{,}81$$

Выборочное среднее квадратическое отклонение:

$$S = \sqrt{\mathrm{S}^2} = \sqrt{62{,}81} \approx 7{,}9$$

Значения числовых характеристик при первом и втором способе вычисления совпали.

III. Показатели распределения

1) ***Коэффициентом асимметрии ("скошенности")*** *A* случайной величины *X* называется величина

$$A = \frac{\mu_3}{\sigma_X^3}$$

Если A > 0, то кривая распределения более полога справа от $M_0(X)$;

если A < 0, то кривая распределения более полога слева от $M_0(X)$.

Коэффициент асимметрии для выборки рассчитывается по формуле:

$$A_s = \frac{n}{(n-1)(n-2)} \cdot \frac{\mu_3}{s^3} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x - \bar{x}}{s} \right)^3,$$

где A_s– коэффициент асимметрии;

$\mu_3 = \frac{\sum (X - \bar{X}_в)^3}{n}$ – момент третьего порядка;

$s = \sqrt{\frac{\sum (x - \bar{x}_в)^2}{n-1}}$ – выборочное среднее квадратическое отклонение;

n – число вариант.

2) *Коэффициентом эксцесса ("островершинности")* $\boldsymbol{\varepsilon}$ случайной величины X называется величина

$$E = \frac{\mu_4}{\sigma_X^4} - 3$$

Величина E характеризует островершинность или плосковершинность распределения.

Для нормального закона распределения: *A = 0 и ε = 0;* остальные распределения сравниваются с нормальным:

Если $\boldsymbol{\varepsilon} > 0$ – более островершинные,

а распределения "плосковершинные" имеют $\varepsilon < 0$

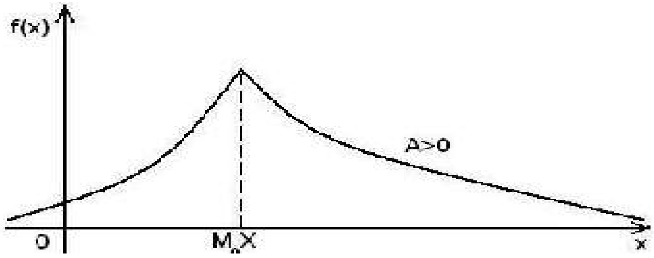
Показатель эксцесса для выборки рассчитывается по формуле:

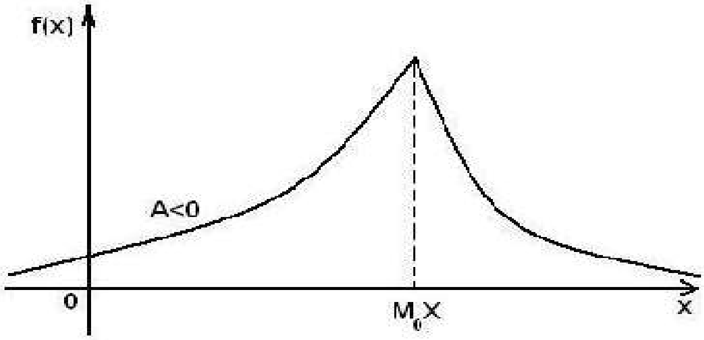
$$\varepsilon_k = \left[\frac{n(n-1)}{(n-1)(n-2)(n-3)} \cdot \frac{\mu_4}{s^4} \right] - \frac{3(n-1)^2}{(n-2)(n-3)} \text{ или}$$

$$\varepsilon_k = \left[\frac{n(n-1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x - \bar{x}}{s} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)},$$

где ε_k– эксцесс;

$\mu_3 = \frac{\sum (X - \bar{X})^4}{n}$ – момент четвертого порядка.





Пример. Имеются выборочные данные о сервис - периоде по 25 коровам (табл. 7.9).

Т а б л и ц а 7.9.

Сервис-период коров

№ п/п	Сервис-период, дн.	Квадрат сервис-периода	Нормированное отклонение от средней	
			в третьей степени	в четвертой степени
	x	x^2	$\left(\frac{x-\bar{x}_в}{s}\right)^3$	$\left(\frac{x-\bar{x}_в}{s}\right)^4$
1	77	5929	-1,4272	1,6069
2	86	7396	-0,0302	0,0094
3	99	9801	0,6477	0,5605
4	68	4624	-7,3064	14,1777
5	83	6889	-0,1980	0,1154
6	83	6889	-0,1980	0,1154
7	98	9604	0,4650	0,3602
8	83	6889	-0,1980	0,1154
9	89	7921	-0,0001	0,0000
10	82	6724	-0,3053	0,2056
11	77	5929	-1,4272	1,6069
12	91	8281	0,0028	0,0004
13	83	6889	-0,1980	0,1154
14	85	7225	-0,0649	0,0261

15	92	8464	0,0124	0,0029
16	106	11236	3,3668	5,0460
17	106	11236	3,3668	5,0460
18	88	7744	-0,0022	0,0003
19	83	6889	-0,1980	0,1154
20	109	11881	5,5480	9,8216
21	86	7396	-0,0302	0,0094
22	107	11449	4,0143	6,3798
23	80	6400	-0,6237	0,5328
24	87	7569	-0,0108	0,0024
25	108	11664	4,7399	7,9620
Итого	$\sum x = 2236$	$\sum x^2 = 202918$	$\sum\left(\frac{x-\bar{x}}{s}\right)^3 = 9{,}6454$	$\sum\left(\frac{x-\bar{x}}{s}\right)^4 = 53{,}934$

Требуется рассчитать основные статистические показатели распределения, характеризующие данную выборочную совокупность.

Средняя продолжительность сервис-периода коров:

$$\bar{x}_в = \frac{\sum x}{n} = \frac{2236}{25} = 89{,}44 \text{ дн.}$$

Выборочная дисперсия исправленная (n<30) сервис-периода коров:

$$s^2 = \frac{\sum (x-\bar{x}_в)^2}{n-1} = \frac{n\sum x^2 - \left(\sum x\right)^2}{n(n-1)} = \frac{25\cdot 202918 - 2236^2}{25\cdot(25-1)} = 122{,}09.$$

Выборочное среднее квадратическое отклонение сервис-периода коров:

$$s = \sqrt{s^2} = \sqrt{122{,}09} = 11{,}049 \text{ дн.}$$

Асимметрия сервис-периода коров:

$$A_s = \frac{n}{(n-1)(n-2)}\sum\left(\frac{x-\bar{x}_в}{s}\right)^3 = \frac{25}{(25-1)\cdot(25-2)}\cdot 9{,}9454 = 0{,}45.$$

Эксцесс сервис-периода коров:

$$\varepsilon_k = \left[\frac{n(n-1)}{(n-1)(n-2)(n-3)}\sum\left(\frac{x-\bar{x}_в}{s}\right)^4\right] - \frac{3(n-1)^2}{(n-2)(n-3)} =$$

$$= \left[\frac{25\cdot(250-1)}{(25-1)\cdot(25-2)\cdot(25-3)}\cdot 53{,}934\right] - \frac{3\cdot(25-1)^2}{(25-2)\cdot(25-3)} = -0{,}529.$$

Полученные значения асимметрии и эксцесса показывают, что данное распределение имеет правостороннюю асимметрию и более высокую вариацию, чем при нормальном распределении.

Точность и надежность оценки. Доверительная вероятность, доверительный интервал.

Как уже было сказано выше, точечной называют оценку, которая определяется одним числом. Все оценки, рассмотренные выше, - точечные. При выборке малого объёма точечная оценка может значительно отличаться от оцениваемого параметра, т. е. приводить к грубым ошибкам. По этой причине при небольшом объёме выборки следует пользоваться интервальными оценками.

Интервальной называют оценку, которая определяется двумя числами – концами интервала. Интервальные оценки позволяют установить точность и надёжность оценок (смысл этих понятий выясняется ниже).

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 <u>точность</u> 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.

Надёжностью (доверительной вероятностью) оценки Θ по Θ^* называют вероятность γ, с которой осуществляется неравенство $\left|\Theta-\Theta^*\right|<\delta$. Обычно надёжность оценки задаётся наперёд, причём в качестве

γ берут число, близкое к единице. Наиболее часто задают надёжность, 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 $\left|\Theta-\Theta^{*}\right|<\delta$ равносильным ему двойным неравенством $-\delta<\Theta-\Theta^{*}<\delta$, или $\Theta^{*}-\delta<\Theta<\Theta^{*}+\delta$, имеем $P\left[\Theta^{*}-\delta<\Theta<\Theta^{*}+\delta\right]=\gamma$.

Это соотношение следует понимать так: вероятность того, что интервал $(\Theta^{*}-\delta,\Theta^{*}+\delta)$ заключает в себе (покрывает) неизвестный параметр Θ, равна γ. Доверительным называют интервал $(\Theta^{*}-\delta,\Theta^{*}+\delta)$, который покрывает неизвестный параметр с заданной надёжностью γ.

- Интервал $(\Theta^{*}-\delta,\Theta^{*}+\delta)$ имеет случайные концы (их называют доверительными границами). Действительно, в разных выборках получаются различные значения Θ^{*}. Следовательно, от выборки к выборке будут изменяться и концы доверительного интервала, т.е. доверительные границы сами являются случайными величинами – функциями от $x_1, x_2, ..., x_n$.

Так как случайной величиной является не оцениваемый параметр Θ, а доверительный интервал, то более правильно говорить не о вероятности попадания Θ в доверительный интервал, а о вероятности того, что доверительный интервал покроет Θ.

Доверительные интервалы для оценки математического ожидания нормального распределения при известном σ.

Пусть количественный признак X генеральной совокупности распределён нормально, причём среднее квадратическое отклонение σ этого распределения известно. Требуется оценить неизвестное математическое ожидание a по выборочной средней $\bar{x}$. Поставим своей задачей найти доверительные интервалы, покрывающие параметр a с надёжностью γ.

Примем без доказательства, что если случайная величина X распределена нормально, то выборочная средняя $\bar{X}$, найденная по независимым наблюдениям, также распределена нормально. Параметры распределения $\bar{X}$ таковы:

$$M(\overline{X}) = a, \ \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}.$$

Приняв во внимание, что по условию нам задана вероятность γ, получаем следующую формулу (чтобы получить рабочую формулу, выборочную среднюю вновь обозначим через $\overline{x}$)

$$P\left(\overline{x}_в - \frac{t\sigma}{\sqrt{n}} < a < \overline{x}_в + \frac{t\sigma}{\sqrt{n}}\right) = 2\Phi_0(t) = \gamma. \qquad (7.17)$$

Смысл полученного соотношения таков: с надёжностью γ можно утверждать, что ***доверительный интервал*** $\left(\overline{x}_в - \frac{t\sigma}{\sqrt{n}}, \overline{x}_в + \frac{t\sigma}{\sqrt{n}}\right)$ покрывает неизвестный параметр a; точность оценки $\delta = \frac{t\sigma}{\sqrt{n}}$.

Укажем ещё, что число t определяется из равенства $2\Phi_0(t) = \gamma$, или $\Phi_0(t) = \frac{\gamma}{2}$; по таблице функции Лапласа находят аргумент t, которому соответствует значение функции Лапласа, равное $\frac{\gamma}{2}$.

Поясним смысл, который имеет заданная надёжность. Надёжность γ=0,95 указывает, что если произведено достаточно большое число выборок, то 95% из них определяет такие доверительные интервалы, в которых параметр действительно заключён; лишь в 5 % случаев он может выйти за границы доверительного интервала.

Доверительные интервалы для оценки математического ожидания нормального распределения при неизвестном σ

Пусть количественный признак X генеральной совокупности распределён нормально, причём среднее квадратическое отклонение σ неизвестно. Требуется оценить неизвестное математическое ожидание a с помощью доверительных интервалов. Разумеется, невозможно воспользоваться результатами предыдущего параграфа, в котором σ предполагалось известным.

Оказывается, что по данным выборки можно построить случайную величину $T = \frac{\overline{X} - a}{S/\sqrt{n}}$, которая имеет распределение Стьюдента с k = n-1 степенями свободы; здесь $\overline{X}$ - выборочная средняя, S – «исправленное» среднее квадратическое отклонение, n – объём выборки.

Пользуясь распределением Стьюдента, находим:

$$P\left(\overline{X}_в - \frac{t_\gamma S}{\sqrt{n}} < a < \overline{X}_в + \frac{t_\gamma S}{\sqrt{n}}\right) = \gamma .$$

Значит, доверительный интервал $\left(\overline{x}_в - \frac{t_\gamma s}{\sqrt{n}}, \overline{x}_в + \frac{t_\gamma s}{\sqrt{n}}\right)$ покрывает неизвестный параметр a с надёжностью γ. По заданным n и γ в таблицах Стьюдента можно найти соответственное t_γ.

Пример. Случайная величина *X* – вес полугодовалого поросенка в хозяйстве (то есть в генеральной совокупности) - распределена нормально. По выборке объёма n = 16 найдены выборочная средняя $\overline{x}_в$=20,2 кг и «исправленное» среднее квадратическое отклонение *S=0,8* кг. Оценить неизвестное математическое ожидание при помощи доверительного интервала с надёжностью 0,95.

Решение. Найдём t_γ. Пользуясь таблицей, по γ=0,95 и n = 16 находим t_γ=2,13.

Найдём доверительные границы:

$$\overline{x} - \frac{t_\gamma s}{\sqrt{n}} = 20{,}2 - \frac{2{,}13 \cdot 0{,}8}{\sqrt{16}} = 19{,}774 кг,$$

$$\overline{x} + \frac{t_\gamma s}{\sqrt{n}} = 20{,}2 + \frac{2{,}13 \cdot 0{,}8}{\sqrt{16}} = 20.626 кг.$$

Итак, с надёжностью 0,95 неизвестный параметр a заключён в доверительном интервале 19,774<a<20,626 (кг).

7.2. Примеры решение типовых задач

Задача 7.1. По данному распределению выборки найти и построить эмпирическую функцию распределения.

Значения x_i	1	4	5	8
Частоты f_i	10	5	20	15

Решение. Найдём эмпирическую функцию распределения.

а) Наименьшая варианта равна единице, следовательно, $F^*(x)=0$ при $x \le 1$.

б) Значение $x<4$, а именно $x_1=1$, наблюдалось 10 раз, значит $F^*(x)=\frac{10}{50}=0{,}2$ при $1<x\le 4$.

в) Значения $x<5$, а именно $x_1=1$ и $x_2=4$, наблюдались 10+5=15 раз, поэтому $F^*(x)=\frac{15}{50}=0{,}3$ при $4<x\le 5$.

г) Значения $x<8$, а именно $x_1=1$, $x_2=4$ и $x_3=5$ наблюдались 10+5+20=35 раз, поэтому $F^*(x)=\frac{35}{50}=0{,}7$ при $5<X\le 8$.

д) Так как x =8 – наибольшая варианта, то $F^*(x)=1$ при x>8.

Итак, аналитическое выражение эмпирической функции распределения:

$$F^*(x)=\begin{cases}0 & при \quad x\le 1;\\ 0{,}2 & при \quad x\le 4;\\ 0{,}3 & при \quad 4<x\le 5;\\ 0{,}7 & при \quad 5<x\le 8;\\ 1 & при \quad x>8.\end{cases}$$

Задача 7.2. Из генеральной совокупности извлечена выборка объема $n=60$:

x_i	3	4	6	26
m_i	8	38	12	2

1) Требуется найти числовые характеристики выборки:

1.выборочную среднюю; 2.моду; 3.медиану; 4.выборочную дисперсию.

2) Определить несмещенную оценку:

1.генеральной средней; 2.генеральной дисперсии.

Решение.

1) Определим числовые характеристики.

1. Выборочную среднюю найдем по формуле (7.2) для взвешенной средней (так как частоты вариантов признака различны). Все необходимые расчеты приведены в таблице 7.10.

Таблица 7.10.

i	x_i	m_i	$x_i \cdot m_i$	$x_i^2 \cdot f_i$	$S_н$
1	**2**	**3**	**4**	**5**	**6**
1	3	8	$3 \cdot 8 = 24$	$3^2 \cdot 8 = 72$	8
2	4	38	$4 \cdot 38 = 152$	$4^2 \cdot 38 = 608$	$8 + 38 = 46$
3	6	12	72	432	$46 + 12 = 58$
4	26	2	52	1352	$58 + 2 = 60$
$\sum$	–	60	$300 \rightarrow \sum_{i=1}^{4} x_i f_i$	$2464 \rightarrow \sum_{i=1}^{4} x_i^2 \cdot f_i$	–

Итак, имеем: $$\bar{x}_в = \bar{x} = \frac{1}{60} \cdot \sum_{i=1}^{4} x_i \cdot f_i = \frac{1}{60} \cdot 300 = 5.$$

(Требуемая сумма произведений $x_i \cdot f_i$ – есть сумма элементов четвертого столбца табл. 6.10). Значит, $\bar{x}_в = 5$.

2. Заданный вариационный ряд дискретный, для которого наибольшая частота $f_2 = 38$ отвечает значению признака $x_2 = 4$. Поэтому мода $M_0 = 4$.

3. Для нахождения медианы выборки строим ряд накопленных частот или кумулятивный ряд (шестой столбец таблицы 6.10). По формуле (6.10) при $n = 60$ получаем:

$$M_e = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) = \frac{1}{2}\left(x_{\frac{60}{2}} + x_{\frac{60}{2}+1}\right) = \frac{1}{2}\left(x_{30} + x_{31}\right).$$

Здесь, x_{30} и x_{31} – соответственно 30-ый и 31-ый члены вариационного ряда. Из кумулятивного ряда видно, что $x_{30} = x_{31} = 4$.

4. Выборочную дисперсию найдем по формуле (6.13) для дисперсии взвешенной: $D_в = \frac{1}{60}\sum_{i=1}^{4} x_i^2 \cdot f_i - \bar{x}_в^2$. Здесь $\sum_{i=1}^{4} x_i^2 \cdot f_i = 2464$ (данная сумма найдена в пятом столбце табл. 6.10), а $\bar{x}_в^{\,2} = 5^2 = 25.$ Итак, получаем: $D_в = \frac{1}{60} \cdot 2464 - 25 \approx 16{,}067.$

2) Определим несмещенные оценки:

1. Несмещенной оценкой генеральной средней является выборочная средняя, то есть $\bar{x}_в = 5.$

2. Несмещенной оценкой генеральной дисперсии является дисперсия исправленная, которую вычислим по формуле (2.14):

$$S^2 = \frac{n}{n-1} \cdot D_в = \frac{60}{60-1} \cdot 16{,}067 \approx 16{,}339.$$

Задача 7.2. Из генеральной совокупности сельхозпредприятий Приморского края сделана выборка из 100 хозяйств. Для исследуемого признака X – величина удоя за лактационный период на одну голову (в кг) – данные сведены в следующую таблицу:

Величина удоя, кг	476-1006	1006-1536	1536-2066	2066-2596	2596-3126	3126-3656	3656-4186	4186-4716
Количество коров	23	30	24	14	3	3	2	1

Необходимо найти:

1. среднюю величину удоя по данной выборке;
2. среднее линейное отклонение удоя;
3. среднее квадратическое выборочное отклонение удоя;
4. исправленную выборочную дисперсию;
5. выборочный коэффициент вариации.

Решение. Здесь имеем интервальный сгруппированный ряд. Взяв в качестве представителя каждого интервала его середину, требуемые выборочные характеристики будем находить так же, как и для дискретного ряда. Все необходимые расчеты проведены в таблице: 7.13:

1. Средняя величина удоя (по формуле (6.5)) составит:

$$\overline{x_в} = \frac{1}{100} \cdot \sum_{i=1}^{8} x_i \cdot f_i = \frac{1}{100} \cdot 162080 = 1620{,}8(кг).$$

(Требуемая сумма найдена в пятом столбце табл. 7.13).

2. Среднее линейное отклонение удоя найдем по формуле (7.6):

$$L = \frac{1}{100} \cdot \sum_{i=1}^{8} |x_i - \overline{x}_в| \cdot f_i = \frac{1}{100} \cdot 61458{,}8 = 614{,}588(кг).$$

(Сумма $\sum_{i=1}^{8} |x_i - \overline{x}_в| \cdot f_i$ вычислена в столбце 7 табл. 7.13).

3. Среднее квадратическое выборочное отклонение удоя вычислим по формуле (7.16):

$$\sigma = \sqrt{\frac{1}{100} \sum_{i=1}^{8} x_i^2 \cdot m_i - \overline{x}^2} = \sqrt{\frac{1}{100} \cdot 324620860 - (1620{,}8)^2} = \sqrt{619215{,}96} \approx 786{,}9(кг)$$

Таблица 7.13.

i	Интервалы	Середина интервала x_i	f_i	$x_i f_i$	$\lvert x_i - \overline{x}_в \rvert$	$\lvert x_i - \overline{x}_в \rvert \cdot f_i$	$x_i^2 \cdot f_i$
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	476-1006	741	23	$741 \cdot 23 = 17043$	879,8	20235,5	$741^2 \cdot 23 = ?$
2	1006-1536	1271	30	$1271 \cdot 30 = 3813$	349,8	10494	$1271^2 \cdot 30 = ?$
3	1536-2066	1801	24	$1801 \cdot ? = 43224$	180,2	?	$1801^2 \cdot 24 = ?$
4	2066-2596	2331	14	$? \cdot 14 = 32634$	?	9942,8	$2331^2 \cdot ? = ?$
5	2596-3126	2861	3	?	1240,2	3720,6	$2861^2 \cdot ? = ?$
6	3126-3656	3391	3	?	1770,2	5310,6	?
7	3656-4186	3921	2	?	?	?	?

8	4186-4716	4451	1	?	2830,2	2830,2	?
	$\sum$	–	100	162080	–	61458,8	324620860

4. Исправленная выборочная дисперсия (по формуле (6.14)):

$$S^2 = \frac{100}{100-1} \cdot D_в = \frac{100}{99} \cdot \sigma_в{}^2 = \frac{100}{99} \cdot 61921{,}96 \Rightarrow S^2 \approx 62547{,}67.$$

5. Выборочный коэффициент вариации (по формуле (6.19)) равен:

$$V = \frac{\sigma_в}{\bar{x}_в} \cdot 100(\%) = \frac{786{,}9}{1620{,}8} \cdot 100(\%) \approx 48{,}6\%.$$

Вывод: выборка неоднородная.

Задача 7.3.На ферме испытывалось влияние витаминов на прибавку в весе телят. Для этой цели было осмотрено 20 телят одного возраста. Средний вес их оказался равным 340 кг, а исправленное среднеквадратическое отклонение 15 кг. Найти доверительный интервал для генеральной средней с надёжностью 0,95.

Решение.

В соответствии с условием задачи $\bar{X}_в = 340, \; S = 15, \; \gamma = 0{,}95, \; n = 20.$

Объём выборки n<30, поэтому при расчёте используем распределение Стьюдента. Находим значение $t_{\gamma,n-1}$. В нашем случае $\gamma = 0{,}95$ и число степеней свободы k = 20 –1=19. Тогда по *табл*. 4 *приложения* 2 имеем

$t_{\gamma,n-1} = 2{,}09$.

Вычислим теперь предельную ошибку выборки (по формуле (7.8)):

$$\varepsilon = \frac{2{,}09 \cdot 15}{\sqrt{20}} = 7{,}01 \approx 7.$$

Получим доверительный интервал:

$340 - 7 \le X_{ген.} \le 340 + 7$ или $333 \le X_{ген.} \le 347$.

7.3. Задачи для самостоятельного решения.

1.Найти эмпирическую функцию распределения по следующему интервальному вариационному ряду:

i	1	2	3	4	5	6	7	8	9
$x_i < X \le x_{i+1}$	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16	16-18
m_i	6	4	2	18	29	11	10	17	3

2.Дано распределение признака X – удоя коров на молочной ферме за лактационный период (в ц); n = 100 (коров):

x_i	4-6	6-8	8-10	10-12	12-14	14-16	16-18	18-20	20-22	22-24	24-26
f_i	1	3	6	11	15	20	14	12	10	6	2

Требуется построить:

1)гистограмму; 2) кумуляту; 3) эмпирическую функцию распределения.

3.Продукция (в млн руб.) за отчетный период у 25 сельхозпредприятий составила: 4,2; 6,5; 7,3; 10,5; 6,0; 9,2; 6,8; 3,3; 10,2; 11,5; 11,9; 14,1; 11,2; 14,2; 17,7; 17,6; 18,5; 19,2; 7,3; 3,2; 4,1; 3,7; 4,0; 5,1; 6,5.

Построить интервальный вариационный ряд, взяв за интервалы 13,2 – 14,2; 14,2 – 15,2 и т.д. Начертить гистограмму.

4.Найти выборочную среднюю по данному распределению выборки (методом условных вариант):

x_i	3250	3270	3280
f_i	2	5	3

5.Дано распределение выборки объема $\text{n} = 100$:

x_i	340	360	375	380
m_i	20	m_2	18	12

Необходимо найти: 1. величину m_2; 2. моду M_0; 3. выборочную дисперсию $D_в$; 4. исправленную дисперсию S^2; 5. коэффициент вариации

6. Учет удоев на ферме в июне месяце дал следующее распределение:

Удой молока (в литрах)	5-7	7-9	9-11	11-13	13-15	15-17	17-19
Число коров	5	8	15	32	97	10	3

1. Определить средний удой в день, выборочную дисперсию и среднее квадратическое отклонение.
2. Найти моду и медиану (графически и аналитически).

7. На овцеводческой ферме из стада произведена выборка 36 овец для взвешивания. Их средний вес оказался равным 50кг. Предположив распределение веса нормальным и определив несмещённую оценку выборочной дисперсии $S^2=16$, найти доверительный интервал для оценки среднего веса во всём стаде с надёжностью: а)0,8; б)0,9; в)0.95. Сделайте выводы о зависимости величины интервала и надёжности оценки.

---

1. Результаты выборочного обследования 100 рабочих крупного предприятия, проведенного с целью определения времени, затрачиваемого на обработку детали, приведены в таблице:

Время обработки (мин.)	3,6-4,2	4,2-4,8	4,8-5,4	5,4-6,0	6,0-6,6
Число рабочих	14	33	35	12	6

Требуется найти: 1.среднее время обработки детали; 2. среднее линейное отклонение времени обработки детали; 3. среднее квадратическое отклонение; 4. исправленную выборочную дисперсию; 5. выборочный коэффициент вариации.

2**.** Выборочное обследование 100 промышленных предприятий РФ по среднегодовой численности рабочих приведено в таблице:

Предприятия со среднегодовой численностью рабочих	Число предприятий
до 100	35
100-200	19
200-500	22
500-1000	11
1000-3000	7

3000-10000	3
более 10000	m_7

1.Найти величину m_7. 2. Определить несмещенную оценку:

а) генеральной средней; б) генеральной дисперсии.

3. Найти выборочную среднюю по данному распределению выборки объема $n = 20$ (методом условных вариант):

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

4. Исправленная дисперсия для выборки объема $n = 21$ составила 6,3. Найти выборочную дисперсию.

5. На молочной ферме испытывали эффективность различных кормов. При кормлении коров смесью №1 было получено распределение надоев «А», при кормлении смесью №2 – распределение «В». На основе полученных распределений выяснить, какая смесь кормов лучше для удойности коров.

Надои молока (л)	Количество коров	
	в распределении «А»	в распределении «В»
0-4	3	0
4-8	8	5
8-12	21	9
12-16	11	14
16-20	6	13
20-24	1	8
24-28	0	1
$\sum$	50	50

6. Найти исправленную выборочную дисперсию по данному распределению выборки:

x_i	23,5	26,1	28,2	30,4
m_i	2	3	4	1

Указание. Перейти к условным вариантам $u_i = 10x_i - 268$.

13. При обследовании годового надоя коров случайным образом отобрали 307 коров, данные по ним сгруппировали и составили таблицу:

Надои (в кг)	3000-3400	3400-3800	3800-4200	4200-4600	4600-5000
Число коров	43	71	102	64	27

1. Найти выборочное среднее, дисперсию (выборочную и исправленную), средне квадратическое отклонение.
2. Определить моду и медиану (аналитически и графически).

к

Вопросы для самопроверки.

1. Понятие генеральной и выборочной совокупности. Повторная и бесповторная выборки. Репрезентативная выборка.

2. Понятие варианты и частоты (веса) варианты. Относительная частота варианты. Построение таблицы статистического распределения.

3. Дискретный и интервальный вариационные ряды: понятие и алгоритмы построения.

4. Накопленная частота и накопленная относительная частота. Эмпирическая функция распределения.

5. Полигон и гистограмма (частот и относительных частот). Кумулятивная кривая.

6. Выборочное среднее (простое и взвешенное). Мода, ее нахождение для дискретного и интервального вариационных рядов. Медиана, её вычисление для дискретного и интервального вариационных рядов.

7. Среднее линейное отклонение (простое и взвешенное). Дисперсия (простая и взвешенная).Среднее квадратическое отклонение (простое и взвешенное).

8. В чём отличие точечной оценки числовой характеристики от интервальной?

9. Определение доверительного интервала. Определение доверительной вероятности и уровня значимости.

10.Как определяется предельная ошибка выборки при нахождении доверительного интервала для генеральной средней?

11. Нахождение доверительного интервала для генеральной средней нормально распределённой генеральной совокупности при известном значении σ и неизвестном σ (n>30). Укажите какие таблицы Приложения 3 используются.

12. Построение доверительного интервала для $X_{ген.}$ по малой выборке (n<30).Укажите использованные таблицы Приложения 3.

13.Как связаны величина доверительного интервала и доверительная вероятность?

14**.**Доверительный интервал для генеральной дисперсии и среднего квадратического отклонения укажите два способа построения интервала. Какие таблицы Приложения 3 при этом используются?

15**.**Как построить доверительный интервал для генеральной доли по малой выборке?

Глава 8. Статистические гипотезы.

8.1. Необходимый теоретический минимум.

Основные понятия. Виды статистических гипотез

Ранее нами было показано, что выборочные характеристики являются оценками генеральных параметров, которые, как правило, остаются неизвестными. Там же описаны точечные и интервальные способы оценки неизвестных параметров по значениям выборочных характеристик.

Ниже будут обсуждаться *сравнительные оценки* генеральных параметров по разности, наблюдаемой между сравниваемыми выборками. Это важно, так как ни одно исследование не обходится без сравнений. Сравнивать приходится данные опыта с контролем, урожайность одной культуры с урожайностью другой, продуктивность одной группы животных с продуктивностью другой и т. д.

Вопрос о достоверности выборочной разности с ее ошибкой приходится решать исходя из той или иной гипотезы, т. е. предположения или допущения относительно параметров сравниваемых групп, которое выражено в терминах вероятности и может быть проверено по выборочным характеристикам.

Часто необходимо знать закон распределения генеральной совокупности. Если закон распределения неизвестен, но имеются основания предположить, что он имеет определённый вид (назовём его А), выдвигают гипотезу: генеральная совокупность распределена по закону А. Таким образом, в этой гипотезе речь идёт о виде предполагаемого распределения.

Возможен случай, когда закон распределения известен, а его параметры неизвестны. Если есть основания предположить, что неизвестный параметр Θ равен определённому значению Θ_0, выдвигают гипотезу: $\Theta = \Theta_0$. Таким образом, в этой гипотезе речь идёт о предполагаемой величине параметра одного известного распределения.

Возможны и другие гипотезы: о равенстве параметров двух или нескольких распределений о независимости выборок и многие другие.

Определение 8.1. Статистической называют гипотезу о виде неизвестного распределения, или о параметрах известных распределений.

Например, статистическими являются гипотезы:

1) генеральная совокупность распределена по закону Пуассона;
2) дисперсии двух нормальных совокупностей равны между собой.

В первой гипотезе сделано предположение о виде неизвестного распределения, во второй – о параметрах двух известных распределений.

Гипотеза «на Марсе есть жизнь» не является статистической, поскольку в ней не идёт речь ни о виде, ни о параметрах распределения.

Определение 8.2. Проверкой статистических гипотез называется сопоставление высказанной гипотезы относительно генеральной совокупности с имеющимися выборочными данными, сопровождаемое количественной оценкой степени достоверности получаемого вывода.

Статистические гипотезы подразделяют на две группы:

I. *Нулевая (основная)* гипотеза – это выдвинутая гипотеза H_0 (т.е. проверяемое утверждение);

II. *Альтернативная (конкурирующая)* гипотеза – это гипотеза H_1, отрицающая или исключающая основную гипотезу.

Например, если нулевая гипотеза состоит в предположении, что математическое ожидание a нормального распределения равно 10, то конкурирующая гипотеза, в частности, может состоять в предположении, что $a \neq 10$. Коротко это записывают так: H_0: a=10; H_1: $a \neq 10$.

Различают гипотезы, которые содержат только одно и более одного предположений.

Простой называют гипотезу, содержащую только одно предположение. Например, если λ - параметр показательного распределения, то гипотеза $H_0 : \lambda = 5$ – простая. Гипотеза H_0: математическое ожидание нормального распределения равно 3 (σ известно) – простая.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез. Например, сложная гипотеза H: $\lambda > 5$ состоит из бесчисленного множества простых вида $H_i : \lambda = b_i$, где b_i - любое число, большее 5. Гипотеза H_0: математическое ожидание нормального распределения равно 3 (σ неизвестно) – сложная.

- Гипотезу проверяют на основании выборки, полученной из генеральной совокупности. Из-за случайности выборки в результате проверки могут возникнуть ошибки и приниматься неправильные решения.

Ошибки первого и второго рода.

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость её проверки. Поскольку проверку производят статистическими методами, её называют статистической. В итоге статистической проверки гипотезы в двух случаях может быть принято неправильное решение, т. е. могут быть допущены ошибки двух родов.

Виды ошибок, возникающих при проверке статистических гипотез:

1. *Ошибка первого рода* – имеет место тогда, когда отвергается правильная гипотеза H_0;
2. *Ошибка второго рода* – имеет место тогда, когда принимается неправильная гипотеза H_0.

Ошибка первого рода состоит в том, что будет отвергнута правильная гипотеза.

Ошибка второго рода состоит в том, что будет принята неправильная гипотеза.

Правильное решение может быть принято также в двух случаях:

1.гипотеза принимается, причём и в действительности она правильная;

2.гипотеза отвергается, причём и в действительности она неверна.

Вероятность совершить ошибку первого рода ($\alpha = P(H_1 / H_0)$) принято обозначать через α; её называют *уровнем значимости.* Наиболее часто уровень значимости принимают равным 0,05 или 0,01. Если, например, принят уровень значимости, равный 0,05, то это означает, что в пяти случаях из ста имеется риск допустить ошибку первого рода (отвергнуть правильную гипотезу).

Вероятность ошибки второго рода обозначают β, т.е. $\beta = P(H_0 / H_1)$.

Вероятность принять верную гипотезу равна $P(H_0 / H_0) = 1 - \alpha$. Вероятность отвергнуть неверную гипотезу H_0 (*мощность критерия*) равна $P(H_1 / H_1) = 1 - \beta$.

Подчеркнём, что последствия этих ошибок могут оказаться весьма различными. Например, если отвергнуто правильное решение «продолжать строительство жилого дома», то эта ошибка первого рода повлечёт материальный ущерб: если же принято неправильное решение «продолжать строительство», несмотря на опасность обвала стройки, то эта ошибка второго рода может повлечь гибель людей. Можно привести примеры, когда ошибка первого рода влечёт более тяжёлые последствия, чем ошибка второго рода.

Статистический критерий проверки нулевой гипотезы. Наблюдаемое значение критерия

Для проверки нулевой гипотезы используют специально подобранную случайную величину, точное или приближённое распределение которой известно. Обозначим эту величину в целях общности через K.

Определение 8.3. Статистическим критерием (или просто критерием) называют случайную величину K, которая служит для проверки нулевой гипотезы H_0.

- Выбор критерия для проверки статистических гипотез может быть выполнен на основании различных принципов. Чаще всего для этого используют *принцип отношения правдоподобия,* который позволяет построить критерий, наиболее мощный среди всех возможных критериев.

Суть принципа правдоподобия: выбрать такой критерий (статистику) *К* с известной функцией плотности $f(k)$ при условии справедливости основной гипотезы H_0, чтобы при заданном уровне значимости α можно найти такую критическую точку $k_{кр}$ распределения $f(k)$, которая разделит область значений критерия на две области: *допустимую* (где результаты выборочного наблюдения выглядят наиболее правдоподобно) и *критическую* (где результаты выборочного наблюдения выглядят менее правдоподобно в отношении гипотезы H_0).

Например, если проверяют гипотезу о равенстве дисперсий двух нормальных генеральных совокупностей, то в качестве критерия K принимают отношение исправленных выборочных дисперсий: $F = \frac{s_1^2}{s_2^2}$.

Эта величина случайная, потому что в различных опытах дисперсии принимают различные, наперёд неизвестные значения, и распределена по закону Фишера-Снедекора.

Определение 8.4. Наблюдаемым значением $K_{набл}$ называется значение статистики, вычисленное по выборке значений.

Для проверки гипотезы по данным выборок вычисляют частные значения входящих в критерий величин и таким образом получают частное (наблюдаемое) значение критерия.

Наблюдаемым значением $K_{набл}$ называют значение критерия, вычисленное по выборкам. Например, если по двум выборкам найдены исправленные выборочные дисперсии $s_1^2 = 20$ и $s_2^2 = 5$, то наблюдаемое значение критерия $F_{набл} = \dfrac{s_1^2}{s_2^2} = \dfrac{20}{5} = 4$.

После выбора определённого критерия множество всех его возможных значений разбивают на два непересекающихся подмножества: одно из них содержит значения критерия, при которых нулевая гипотеза отвергается, а другая – при которых она принимается.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

Областью принятия гипотезы (областью допустимых значений) называют совокупность значений критерия, при которых гипотезу принимают.

Основной принцип проверки статистических гипотез можно сформулировать так: *если наблюдаемое значение критерия принадлежит критической области – гипотезу отвергают, если наблюдаемое значение критерия принадлежит области принятия гипотезы – гипотезу принимают.*

Поскольку критерий K - одномерная случайная величина, все её возможные значения принадлежат некоторому интервалу. Поэтому критическая область и область принятия гипотезы также являются интервалами и, следовательно, существуют точки, которые их разделяют.

Критическими точками (границами) $k_{кр}$ называют точки, отделяющие критическую область от области принятия гипотезы.

Различают одностороннюю (правостороннюю или левостороннюю) и двустороннюю критические области.

Правосторонней называют критическую область, определяемую неравенством $K > k_{кр}$, где $k_{кр}$ - положительное число.

Левосторонней называют критическую область, определяемую неравенством $K < k_{кр}$, где $k_{кр}$ - отрицательное число.

Односторонней называют правостороннюю или левостороннюю критическую область.

Двусторонней называют критическую область, определяемую неравенствами $K < k_1, K > k_2$, где $k_2 > k_1$.

В частности, если критические точки симметричны относительно нуля, двусторонняя критическая область определяется неравенствами
(в предположении, что $k_{кр}$>0):

$K < -k_{кр}, K > k_{кр}$, или равносильным неравенством $|K| > k_{кр}$.

Алгоритм проверки статистической гипотезы:

1. Определить основную H_0 и альтернативную H_1 гипотезы.
2. Задать уровень значимости α.
3. Выбрать критерий.
4. Найти границы критической области (по альтернативной гипотезе H_1, по уровню значимости α – с помощью таблиц).
5. Вычислить наблюдаемое значение $K_{набл}$ статистики.
6. Сравнить значение статистики с критической областью.
7. Принять решение: а) если значение статистики не принадлежит критической области, то гипотеза H_0 не отвергается; б) если значение статистики принадлежит критической области, то гипотеза H_0 отвергается, а гипотеза H_1 принимается.

• Проверка каждого типа статистических гипотез выполняется с помощьюсоответствующего критерия, например:

– проверка гипотезы о *виде закона распределения* случайной величины осуществляется на основании *критерия согласия Пирсона* χ^2;

– проверка гипотезы о *равенстве неизвестных значений дисперсии* двух генеральных совокупностей – с помощью *критерия Фишера F*;

– проверка гипотез о *неизвестных значениях параметров* генеральных совокупностей – посредством *критерия Z нормально распределенной случайной величины* и *t –критерия Стьюдента*.

Основные виды проверки гипотез

1) Проверка гипотезы о виде закона распределения. Критерий χ2 как критерий согласия

Критерий χ^2 как критерий согласия используют при проверке принадлежности эмпирического распределения к теоретическому, например, к нормальному, биноминальному, распределению Пуассона и т. п.

В этом случае значение критерия χ^2 определяют, исходя из частот (f) эмпирического распределения и частот (f_o) теоретического распределения:

$$\chi^2_{факт} = \sum \frac{(f - f_o)^2}{f_o} \qquad (8.1)$$

При этом возможны случаи, когда теоретические частоты заранее известны и когда неизвестны. Во втором случае теоретические частоты определяют на основе теоретического распределения исходя из численности выборки.

При проверке гипотезы о соответствии эмпирического распределения теоретическому сравнивают фактическое значение критерия $\chi^2_{факт}$ с табличным $\chi^2_{табл}$. Если $\chi^2_{факт}$ меньше $\chi^2_{табл}$, следовательно, эмпирическое распределение соответствует теоретическому. В противном случае эмпирическое распределение не соответствует теоретическому, распределение частот в нем носит другой характер.

Рассмотрим методику применения критерия χ^2 как критерия согласия.

Пример. В результате учета яйценоскости 50 кур-несушек, содержащихся на птицеферме, был построен интервальный вариационный ряд (табл. 8.1). Средняя арифметическая ряда равна 228,8, а выборочное среднее квадратическое отклонение – 7,95.

Требуется установить соответствие данного распределения нормальному с уровнем вероятности 0,95.

Решение.

Проверка гипотезы о соответствии теоретическому распределению предполагает расчет теоретических частот этого распределения.

Для нормального распределения порядок расчета этих частот следующий:

1) по эмпирическим данным рассчитывают среднюю арифметическую ряда $\bar{x}$ и среднее квадратическое отклонение s;

2) находят нормированное отклонение t каждого эмпирического значения от средней арифметической: $t = \frac{x - \bar{x}_в}{s}$;

3) по формуле или с помощью таблиц интеграла вероятностей Лапласа находят значение плотности нормального распределения $\varphi(t)$:

$\varphi(t) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{t^2}{2}}$, где s – выборочное среднее квадратическое отклонение;

Таблица 8.1.

Распределение поголовья

№ п/п	Группа кур-несушек по величине яйценоскости		Фактическое распределение поголовья (эмпирические частоты)	Середина интервала	Нормированное отклонение	Плотность нормального распределения	Теоретическое распределение поголовья (теоретические частоты)	Взвешенные квадраты разностей
	$x_{\min}$	$x_{\max}$	f	x'	$t=\frac{x'-\bar{x}}{s}$	$\varphi=\frac{1}{s\sqrt{2\pi}}e^{-\frac{t^2}{2}}$	$f_0=nh\varphi(t)$	$\frac{(f-f_o)^2}{f_o}$
1	212	217	5	214,5	-1,8214	0,0096	2,39	2,86
2	217	222	7	219,5	-1,1924	0,0246	6,16	0,11
3	222	227	9	224,5	-0,5635	0,0428	10,70	0,27
4	227	232	14	229,5	0,0654	0,0501	12,52	0,18
5	232	237	6	234,5	0,6943	0,0394	9,86	1,51
6	237	242	8	239,5	1,3233	0,0209	5,23	1,47
7	242	247	1	244,5	1,9522	0,0075	1,87	0,40
Итого	50	х	х	х	48,72	χ^2 = 6,80		

4) вычисляют теоретические частоты f_0 по формуле:

$$f_0 = nh\varphi(t),$$

где n– число вариант (сумма частот);

h – величина интервала.

Фактическое значение критерия $\chi^2_{набл}$ равно 6,8. Табличное значение критерия $\chi^2_{кр}$ при заданном уровне значимости $\alpha = 0{,}05$ и $\nu = (k-1)(l-1) = (2-1)\cdot(7-1) = 6$ степенях свободы вариации равно 12,592.

Поскольку наблюдаемое значение критерия меньше критического, то нулевая гипотеза о соответствии эмпирического распределения теоретическому принимается. Распределение яйценоскости кур-несушек соответствует нормальному распределению.

2) Оценка существенности разности средних независимых и сопряженных выборок по t – критерию.

Дисперсия является одним из основных показателей вариации (рассеяния). К сравнению дисперсий прибегают всякий раз, когда требуется сопоставить совокупности по степени их однородности, сравнить качество измерительную инструментов или вычислительных методов.

Проверка статистических гипотез относительно дисперсий имеет важное значение и для принятия экономических решений. Например, сравнительная экономическая оценка двух сортов сельскохозяйственной культуры. Основными биологическими признаками сортов, принимаемыми здесь в расчет, являются средняя многолетняя урожайность, качество зерна и вариация урожайности в связи с действием различных факторов. При равенстве первых показателей последний становится главным в принятии решения об использовании сортов. Для случая различной вариации урожайности возможны такие варианты решения: а) предпочтение отдается сорту, имеющему меньшую колеблемость урожайности по годам, поскольку при этом требуется меньше собственной техники и рабочей силы для уборки урожая без потерь; б) предпочтение отдается сорту, имеющему большую вариации урожайности, поскольку выяснено, что она связана с действием факторов, уровни которых планируется регулировать; в) выбор - на оптимальном сочетанием посевов двух сортов, учитывавшим особенности каждого из них и поставленную экономическую цель.

Методы измерения вариации, оценивания действия факторов опираются на расчет и сравнения дисперсий. Так, сопоставление вариации признака в двух совокупностях достигается проверкой гипотезы о равенстве дисперсий. Для этого применяется F -критерий Фишера - Снедекора. Наблюдаемое значение критерия вычисляется по формуле:

$$F_{набл} = \frac{S_1^2}{S_2^2}, \qquad (8.2)$$

где $S_1^2 \geq S_2^2$. Критическое значение критерия находят из таблицы F – распределения по заданному уровню значимости и числу степеней свободы ν_1 и ν_2 (соответственно для большой и малой дисперсий, $\nu_1 = n_1 - 1,\ \nu_2 = n_2 - 1)$.

Если $F_{набл} > F_{кр}$, то гипотеза о равенстве дисперсий отвергается, в противном случае, т.е. когда $F_{набл} \leq F_{кр}$, гипотеза не отвергается. Использование F -критерия предполагает, что обе совокупности имеют нормальное распределение.

Пример. В результате испытаний двух сортов гороха на трех сортоучастках за 5 лет получены следующие сведения: средняя урожайность сорта А составила 19,8 , сорта Б - 20,6 ц/га; дисперсия урожайности - по сорту А 9,98, по сорту Б - 25,12. Требуется оценить, существенно ли различив в вариации урожайности сортов.

Решение.

Поскольку средние значения урожайности примерно равны, для ответа на поставленный вопрос достаточно проверить гипотезу о равенстве дисперсий. По исходным данным $F_{набл} = 25{,}12 / 9{,}98 = 2{,}51$. Критическое значение критерия при $\alpha = 0{,}05$, $\nu_1 = 15 - 1 = 14$ и $\nu_2 = 15 - 1 = 14 (n_1 = n_2 = 15 = 15 \cdot 3)$ равно 2,48 (получено интерполяцией данных приложения 5).

Следовательно $F_{набл} > F_{кр}$, и гипотеза о равенстве дисперсий отвергается с рискам ошибиться в 5 случаях из 100. Есть основания полагать, что урожайность сорта Б менее стабильна в связи о различными погодными и почвенными условиями.

Окончательный вывод об экономической целесообразности использования сортов может быть сделан после анализа влияния факторов урожайности. Если влияние будет доказано, а факторы управляемы, то, скорее всего, преимущество будет иметь сорт Б, как более интенсивный. Если сколько-нибудь существенное влияние факторов отсутствует, либо они не управляемы, то предпочтение имеет сорт А, поскольку для получения единицы продукции он требует меньше ресурсов в расчете на своевременную уборку максимального урожая.

В зависимости от содержания исходной информации нулевая гипотеза о равенстве средник значений двух совокупностей может состоять в предположении, что последние являются выборками из одной генеральной совокупности или выборками из двух генеральных совокупностей. В первом случае предполагается, что разность между средними соизмерима с ошибками выборочного наблюдения, а во втором, кроме того, что генеральные средние равны. Первому варианту нулевой гипотезы предпочтение обычно отдается, когда дисперсии двух выборок равны, второму - при неравенстве дисперсий.

При условии нормального распределения изучаемого признака проверку гипотезы относительно средних проводят с помощью t -критерия Стьюдента. Для обоих вариантов постановки гипотезы наблюдаемое значение t -критерия может быть вычислено по общей формуле:

$$t_{набл} = \frac{|\bar{d}|}{\sqrt{S^2}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}, \qquad (8.3)$$

в которой:

- для первого $|d| = |\tilde{X}_1 - \tilde{X}_2|; S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$,

где S^2 - уточненная оценка дисперсии генеральной совокупности, полученная усреднением двух выборочных оценок S_1^2 и S_2^2;

- для второго варианта $|d| = |\tilde{X} - \tilde{Y}|; S^2 = \frac{(n_1 - 1)S_x^2 + (n_2 - 1)S_y^2}{n_1 + n_2 - 2}$,

где S^2 - средняя оценка генеральных дисперсий.

Если объемы выборок равны, т.е. $n_1 = n_2 = n$, то формула (8.3) приобретает вид

$$t_{набл} = \frac{|\bar{d}|}{\sqrt{2S^2}} \cdot \sqrt{n}, \qquad (8.4)$$

где $2S^2 = S_1^2 + S_2^2$ или $2S^2 = S_x^2 + S_y^2$.

Табличное значение критерия находят по заданному уровню значимости α и числу степеней свободы, которое для первого варианта определяется по формуле:

$$\nu = n_1 + n_2 - 2, \tag{8.5}$$

а для второго –

$$\nu = (n_1 + n_2 - 2) \cdot \left(0{,}5 + \frac{S_x^2 \cdot S_y^2}{S_x^4 + S_y^4} \right), \tag{8.6}$$

Выражение (8.5) по сравнению с (8.6) уменьшает число степеней свободы почти вдвое, если выборочные дисперсии различаются существенно. Это вполне справедливо, поскольку S^2 в такой ситуации зависит главным образом от большей выборочной дисперсии, и поэтому на среднюю ошибку разности средних:

$$S_d = \sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \tag{8.7}$$

влияет преимущественно вариация признака только в одной совокупности.

Критерии достоверности выборочных показателей

При помощи стандартной ошибки выборки определяют достоверность полученной выборочной средней.

Критерий достоверности $t = \dfrac{\bar{x}}{s_{\bar{x}}}$.

По величине t судят о достоверности какого-либо статистического параметра, основываясь на связи этой величины с уровнем вероятности **(P).** Для больших выборок (n>175) (табл. 8.2).

Если вычисленное значение t будет меньше 2,0 , то выборочный параметр недостоверен, т.е. он не характеризует генеральную совокупность и выводы, полученные при этом, не могут быть распространены на генеральную совокупность.

Таблица 8.2.

Уровень вероятности

	Уровень вероятности Р	Уровень значимости или существенности р(α)
t=2,0	Р=0,95	0,05
t=2,6	Р=0,99	0,01
t=3,3	Р=0,999	0,001

Бывает, что недостоверность вызвана небольшим числом наблюдений, поэтому следует увеличить объём выборки и заново ставить эксперимент.

При малом числе наблюдений (n< 30) на величину t влияет объём выборки. Чем меньше выборка, тем менее точно величина t позволяет судить о достоверности выборочной средней. Эту особенность значения изучил английский учённый Госсет (Стъюдент), который разработал таблицы значений t в зависимости от числа наблюдений. Чем меньше число наблюдений, тем больше значение t для одного и того же уровня вероятности.

Значения t даны при различном числе *степеней свободы.* Под числом степеней свободы понимается количество вариантов, которые могут принимать произвольные значения, не меняя величины средней (k= n-1).

Ошибка разности между средними арифметическими двух вариационных рядов.

В экспериментальной и практической работе большое значение имеет определение разности между средними показателями двух сравниваемых групп животных и установление достоверности этой разности.

Например, часто требуется определить, достоверна ли прибавка удоя при введении в рацион коров какого-либо кормового компонента.

В ветеринарных исследованиях очень важно бывает доказать, что примененная доза или новый лекарственный препарат достоверно уменьшает долю заболевших животных по сравнению с долей больных животных в контрольной группе, не подвергавшихся лечению.

Определение достоверности разности между средними арифметическими или долями двух вариационных рядов можно определить с помощью ошибки разности m_d

Формула ошибки разности следующая: $s_d = \sqrt{s_{\bar{x}_1}{}^2 + s_{\bar{x}_2}{}^2}$,

где $s_{\bar{x}_1}{}^2$ - квадрат ошибки средней арифметической 1-го вариационного ряда,

$s_{\bar{x}_2}{}^2$ - квадрат ошибки средней арифметической 2-го вариационного ряда,

s_d - искомая ошибка разности.

Эта формула используется в тех случаях, когда варианты одной выборки некоррелированы с вариантами другой выборки.

Критерий достоверности разности между средними арифметическими вычисляют по следующей формуле

$$t_d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_{\bar{x}_1}{}^2 + s_{\bar{x}_2}{}^2}} \geq t_{st} \qquad (8.8)$$

Найденный критерий t_d сравнивают с табличным значением критерия Стъюдента при числе степеней свободы $k=n_1+n_2-2$, где n_1 и n_2 — объемы выборок.

Если $t_d<t_{st}$ - разность считается недостоверной. Это значит, что не получено никакого определённого ответа о разности между соответствующими генеральными параметрами. Если получена благоприятная по смыслу исследования разность между двумя выборочными средними, но эта разность оказалась недостоверной, то это значит, что между соответствующими генеральными средними могут быть любые соотношения, а какие именно неизвестно, но это не может служить доказательством отсутствия разницы между генеральными средними.

Если в выборочном исследовании оказалась достоверная разница между выборочными показателями, то такая же разница по знаку будет и между соответствующими генеральными параметрами. Основной вывод может быть перенесён на генеральную совокупность.

Для коррелированных выборок используется следующая формула разности между средними: $s_d = \sqrt{s_{\bar{x}_1}{}^2 + s_{\bar{x}_2}^2 - 2r \cdot s_{\bar{x}_1} \cdot s_{\bar{x}_2}}$,

где s_1 и s_2 – ошибки средних арифметических по каждой выборке;

r – коэффициент корреляции между вариантами обеих выборок при попарном их сопоставлении.

Пример. При определении доли влияния системы содержания на конверсию корма на 1 кг прироста живой массы бройлеров различных кроссов было отмечено, что лучший генетический потенциал по данному показателю по всем кроссам выявлен у молодняка, выращенного в реконструированных птичниках с регулируемым микроклиматом (см. табл. 8.3).

Определим достоверность различий в показателях по конверсии корма на 1 кг прироста живой массы бройлеров при разных системах содержания.

Решение.

Сравнение проведём с данными при содержании бройлеров в реконструированных птичниках с регулируемым микроклиматом.

Вычислим критерий достоверности разности между средними арифметическими для кросса «Смена-4»:

$$t_d = \frac{2{,}2 - 2{,}06}{\sqrt{0{,}02^2 + 0{,}02^2}} = \frac{0{,}14}{0{,}028} = 5$$

$t_d > t = 3{,}3 \Rightarrow$ различие достоверно при уровне значимости P<0,001

Для кросса «Кобб-500» $t_d = \dfrac{2{,}07 - 1{,}9}{\sqrt{0{,}05^2 + 0{,}02^2}} = \dfrac{0{,}17}{0{,}054} = 3{,}1$

$t_d > t = 2{,}6 \Rightarrow$ различие достоверно при уровне значимости P<0,01

Для кросса «Арбор-Айкрес» $t_d = \dfrac{2{,}02 - 1{,}95}{\sqrt{0{,}03^2 + 0{,}02^2}} = \dfrac{0{,}07}{0{,}036} = 1{,}94$

$t_d < t = 2 \Rightarrow$различие недостоверно.

Таблица 8.3.

Конверсия корма на 1 кг прироста живой массы высокопродуктивных бройлеров при разных системах содержания

Кросс	Система содержания								
	Напольная на глубокой подстилке			Клеточная (в батареях типа БКМ-3Б)			Напольная в реконструированных птичниках		
	$\bar{x} \pm s_{\bar{x}}$	σ	C_V %	$\bar{x} \pm s_{\bar{x}}$	σ	C_V %	$\bar{x} \pm s_{\bar{x}}$	σ	C_V %
СМЕНА-4	2,17 $\pm$0,01***	0,11	5,0	2,2 $\pm$0,02 ***	0,07	3,2	2,06 $\pm$0,02	0,10	4,8
ИЗА	1,99$\pm$ 0,01***	0,10	4,3	2,05 $\pm$0,02 ***	0,10	3,4	1,90 $\pm$0,02	0,06	3,2
РОСС-308	2,07$\pm$0,02	0,12	4,8	2,10 $\pm$0,03	0,08	3,8	2,03 $\pm$0,03	0,09	4,4
КОББ-500	1,97$\pm$ 0,02*	0,10	5,1	2,07 $\pm$0,05 **	0,15	7,2	1,90 $\pm$0,02	0,07	3,7
АРБОР АЙКРЕС	2,02$\pm$ 0,02*	0,13	4,9	2,02 $\pm$0,03	0,10	4,9	1,95 $\pm$0,02	0,03	1,5

*/ Достоверно выше по сравнению с напольным содержанием в реконструированных птичниках с регулируемым микроклиматом при Р<0,05;
**/ при Р<0,01;
***/ при Р<0,001

8.2. Примеры решения типовых задач.

Задача 8.1. По приведенным ниже данным сравнить средние удои коров, получавших различные рационы. Для проверки нулевой гипотез о равенстве средних принять уровень значимости $\alpha = 0{,}05$.

Рацион	Поголовье коров, получавших рацион, гол	Среднесуточный удой молока в пересчете на базисную жирность, кг/гол	Среднеквадратическое отклонение в молочной продуктивности коров, кг/гол
№1	10	16,2	3,8
№2	8	17,7	4,2

Решение.

Поскольку выборочные дисперсии различаются несущественно: $F_{набл} = 4{,}2^2 / 3{,}8^2 = 1{,}22 < F_{0,05} = 3{,}29$, то есть основания предполагать, что выборки со средним $\tilde{X}_1 = 16{,}2$ и $\tilde{X}_2 = 17{,}7$ получены из одной генеральной совокупности, а расхождение между $\tilde{X}_1$ и $\tilde{X}_2$ случайно. Для проверки $H_0 : \left|\tilde{X}_1 - \tilde{X}_2\right| = 0$ уточним оценку генеральной дисперсии $S^2 = \dfrac{(10-1)\cdot 3{,}8^2 + (8-1)\cdot 4{,}2^2}{10+8-2} = 15{,}84.$

Подставим полученное значение S^2 и другие величины в (5.11): $t_{набл} = \dfrac{|16{,}2 - 17{,}7|}{\sqrt{15{,}84}} \cdot \sqrt{\dfrac{10\cdot 8}{10+8}} = 0{,}79.$

Далее по формуле (5.13) определяем число степеней свободы $\nu = 10 + 8 - 2 = 16.$ $\alpha = 0{,}05$ и $\nu = 16$ соответствует $t_{кр.} = 2{,}12$ (приложение 3). $t_{набл}$ меньше, чем $t_{кр.}$, поэтому нулевая гипотеза не отвергается.

Задача 8.2. Используя данные (табл. 8.4) , оценить влияние обеспеченности рационов протеином на изменение живой кассы свиноматок в начале лактации. С этой целью сравнить средние значения потерь массы по вариантам опыта, применяя t- критерий Стьюдента. Уровень значимости принять равным 0,05.

Решение.

Учитывая, что $n_1 = n_2 = n_3 = 5$, наблюдаемой значение критерия для каждой из трех пар средних можно рассчитать по формуле (8.8).

Таблица 8.4.

Изменение живой массы свиноматок в начале лактации при разном уровне сырого протеина в рационе.

Показатели	Содержание сырого протеина в рационе, %		
	9,3	11,8	14,3
Количество свиноматок, гол.	5	5	5
Потери живой массы в первые 20 дней лактации, кг/гол.	14,8	7,5	1,0
Среднеквадратическое отклонение потерь живой массы, кг/гол.	1,5	0,8	0,3

Для $\tilde{X}_1$ и $\tilde{X}_2$: $\quad t_{набл} = \dfrac{|14{,}8 - 7{,}5|}{\sqrt{1{,}5^2 + 0{,}8^2}} \cdot \sqrt{5} = 9{,}60.$

Для $\tilde{X}_1$ и $\tilde{X}_3$: $\quad t_{набл} = \dfrac{|14{,}8 - 1{,}0|}{\sqrt{1{,}5^2 + 0{,}3^2}} \cdot \sqrt{5} = 20{,}2.$

Для $\tilde{X}_2$ и $\tilde{X}_3$: $\quad t_{набл} = \dfrac{|7{,}5 - 1{,}0|}{\sqrt{0{,}8^2 + 0{,}3^2}} \cdot \sqrt{5} = 17{,}0.$

Сравнение дисперсий показывает, что только между G и G_2^2 различие несущественно ($F_{набл} = 1{,}5^2 / 0{,}8^2 = 3{,}52 < F_{0,05} = 6{,}39$). Для G_1^2 и G_3^2 $F_{набл} = 1{,}5^2 / 0{,}3^2 = 25{,}0$, для G_2^2 и G_3^2 $F_{набл} = 0{,}8^2 / 0{,}3^2 = 7{,}11$ при $F_{0,05} = 6{,}39$. Поэтому для первой пары средних число степеней свободы определяется по формуле(5.13), а для двух других - по (5.14):

$$\nu_{1-2} = 5 + 5 - 2 = 8,$$

$$\nu_{1-3} = (5 + 5 - 2)\left(0{,}5 + \frac{1{,}5^2 \cdot 0{,}3^2}{1{,}5^4 + 0{,}3^4}\right) = 4{,}32,$$

$$\nu_{2-3} = (5 + 5 - 2)\left(0{,}5 + \frac{0{,}8^2 \cdot 0{,}3^2}{0{,}8^4 + 0{,}3^4}\right) = 5{,}10.$$

Соответственно критические знамения *t*-критерия оказываются различными: $t_{0,05}(8) = 2{,}31, t_{0,05}(4{,}32) = 2{,}70, t_{0,05}(5{,}10) = 3{,}35.$. Сравнение наблюдаемых значений *t* - критерия с табличными дает основание отвергнуть все три нулевые гипотезы о равенстве средних. Следовательно, изменение обеспеченности рационов протеином существенно влияет на массу свиноматок в начале лактации.

7.3. Задания для самостоятельного решения.

1. Для нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 6{,}24$ извлечена выборка объема $n = 100$ и по ней найдена выборочная средняя $\bar{x} = 68{,}21$. Требуется при уровне значимости $\alpha = 0{,}01$ проверить основную гипотезу $H_0 : a = 67$ при альтернативной $H_1 : a \neq 26$.

2. В таблице приведена динамика среднесуточных привесов крупного рогатого скота (в граммах) в период 2010-2015 гг. (на начало года) в сельскохозяйственных организациях Приморского края:

Год	2000	2001	2002	2003	2004
Среднесуточный привес КРС (граммов)	168	113	265	269	274

В средствах массовой информации было заявлено, что средний привес КРС в период 2000-2010 гг. составит 220 граммов.

1) На основе имеющейся информации при уровне значимости $\alpha = 0{,}01$ проверьте истинность этого утверждения.

2) Изменится ли полученный в пункте 1 вывод, если уровень значимости принять 0,05. Объясните результаты.

3. В опыте провели оценку качества посева ячменя по глубине посева при вспашке на 20 см (колонка №1) и при фрезеровании на 8 - 10 см (колонка №2). Результаты замеров оказались следующими:

№1	№2
4,0	5,4
5,3	4,9
5,6	5,3
4,4	5,4
2,5	4,7
5,3	5,3
7,7	5,2

3,4	5,3
4,5	4,3
8,2	5,5

Требуется вычислить статистические характеристики для каждой выборки. Определить существенность разности между средними и сделать вывод.

---

1. Поставщик удобрений утверждает, что применение новой партии удобрений обеспечивает урожайность пшеницы в 60 ц/га. Удобрения внесли на площади в 37 га и получили урожай 55 ц/га при «исправленном» среднем квадратичном отклонении 3 ц/га. При 5 %-м уровне значимости оценить справедливость утверждения поставщика.

2. Процент гибели свекловичной тли при обработке инсектицидом (колонка № и без обработки (колонка №2) составил:

№1	№2
95	76
67	67
70	68
69	75
77	69
79	65
70	56
83	67
92	65
90	53

Требуется вычислить статистические характеристики для каждой выборки. Определить существенность разности между средними и сделать вывод.

Вопросы для самопроверки.

1. Понятие и виды статистических гипотез.

2.Виды ошибок, возникающие при проверке гипотез.

2.Уровень значимости; мощность критерия.

3.Понятие статистического критерия, наблюдаемое значение критерия.

4.Допустимая и критическая области, критическая точка.

6.Общий вид алгоритма проверки статистических гипотез.

7. Методика проверки гипотезы о законе распределения.

8. Каким образом проверить равенство двух дисперсий, средних?

Глава 9. Дисперсионный анализ.

9.1. Необходимый теоретический минимум.

Основные понятия.

Рассмотренные в предыдущей главе методы сравнения дисперсий и средних в ходе дисперсионного анализе используются комплексно, причем одновременно для неограниченного числа совокупностей. Основная цель дисперсионного анализа - оценка влияния факторов на результативный признак. Первоначально дисперсионный анализ был предложен Р.Фишером (l925) для обработки результатов агрономических опытов. В дальнейшем этот метод нашел применение во всех областях наук, где требуется математическая обработка результатов эксперимента.

Определение 9.1. Дисперсионный анализ – это статистический метод анализа результатов наблюдений, зависящих от различных, одновременно действующих факторов, выбор наиболее важных факторов и оценка их влияния. Дисперсионный анализ находит применение в различных областях науки и техники.

Известно, что многие признаки и свойства живых организмов находятся под влиянием различных факторов: наследственности, условий среды, внутренних факторов организма, искусственного отбора. Степень и направленность воздействия различных факторов неодинаковы, поэтому важно определить долю влияния отдельных факторов на изменчивость признака. Для решения подобной задачи используют метод дисперсионного анализа, разработанный Р.Фишером.

Сущность дисперсионного анализа состоит в установлении роли отдельных факторов в изменчивости признака.

В зависимости от количества изучаемых факторов различают ***однофакторный и многофакторный дисперсионный анализ***. Рассмотрим подробнее метод однофакторного дисперсионного анализа.

Главная идея дисперсионного анализа заключается в расчленении общего объема вариации результативного признаке по источникам ее образования. Постановка эксперимента и последующая обработка его результатов направлены на то, чтобы реализовать эту идею применительно к изучаемым зависимостям. Самая простая однофакторная схема дисперсионного анализа предполагает, что

$$C_{общ} = C_{факт} + C_{ост}, \tag{9.1}$$

где $C_{общ}$ - общий объем вариации результативного признака;

$C_{факт}$ - объем вариации, обусловленный действием изучаемого фактора;

$C_{ост}$- остаточный объем вариации, вырванный действием случайных факторов.

Однофакторный дисперсионный анализ.

Предположим, что имеется K выборок с объемами n_1, $n_2,...,n_k$, $N = n_1 + n_2 + ... + n_k$, и наблюдения можно представить в виде $x_{ij} = a_j + \varepsilon_{ij}$, где i - номер наблюдения в выборке; j- номер выборки; a_j- групповые математические ожидания; ε_{ij} - случайные ошибки с $M\varepsilon_{ij}$=0, о которых предполагается, что они независимы и одинаково расположены.

Подобная ситуация возникает, когда существует некий фактор, принимающий K различных значений (называемых уровнями), и каждая группа объектов, чьи признаки мы примеряем, подвергается воздействию определенного уровня этого фактора.

Определение 9.2. Методы математической статистики, изучающие воздействие одного фактора на объекты и их признаки, называют в совокупности ***однофакторным анализом.***

Предполагается, что ошибки нормально распределены: $\varepsilon_{ij} \in N\left(0, \sigma^2\right)$. Тогда можно изучать влияние фактора, вычисляя дисперсии некоторых величин. Совокупность этих методов называют *однофакторным дисперсионным анализом.*

Основной гипотезой, нуждающейся в проверке, является гипотеза о равенстве групповых средних $H_0 : a_1 = a_2 = ... = a_k$. Иными словами, проверяют гипотезу о том, что фактор вообще не влияет на наблюдения. В случае нормальных ошибок ее можно проверить, вычислив две разные оценки дисперсии.

Рассмотрим группу экспериментальных животных, подвергнутых ультрафиолетовому облучению. В процессе эксперимента измерялась температура тела животных. Результаты измерений были занесены в таблицу:

Таблица 9.1

Температура тела животных

№ испытания	Уровень фактора А (мощность ультрафиолетового облучения)		
	А1	А2	А3
1	37,4	37,8	38,0
2	37,3	37,9	37,9
3	37,0	37,5	38,4
4	36,9	37,4	38,3
$\bar{x}_j$	37,15	37,65	38,15

Физический фактор А (ультрафиолетовое излучение) имеет $m = 3$ постоянных уровней (3 различных мощности облучения). На всех уровнях распределения случайной величины Х (температуры тела животного) предполагается нормальным, а дисперсии одинаковыми, хотя и неизвестными.

В данном эксперименте число проведенных наблюдений при действии каждого из уровней фактора одинаково.

Все значения величины X, наблюдаемые при каждом фиксированном уровне фактора $A_{j,}$ составляют группу, и в последней строке таблицы представлены соответствующие выборочные групповые средние, вычисленные по формуле $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$.

Здесь n – число испытаний, j – номер столбца, i - номер строки, в которой расположено данное значение случайной величины. Общая средняя арифметическая всех nm наблюдений находится как

$$\bar{x} = \frac{1}{m}\sum_{j=1}^{m} \bar{x}_j \ . \tag{9.1}$$

Введем следующие понятия:

Факторная сумма квадратов отклонений групповых средних от общей средней $\bar{x}$, которая характеризует рассеивание «между группами» (т.е. рассеивание за счет исследуемого фактора):

$$C_{\text{факт}} = n\sum_{j=1}^{m}(\bar{x}_j - \bar{x})^2 \tag{9.2}$$

Остаточная сумма квадратов отклонений наблюдаемых значений группы от своей групповой средней $\bar{x}_j$, которая характеризует рассеивание «внутри групп» (за счет случайных причин):

$$C_{\text{ост}} = \sum_{i=1}^{n}\left(x_{i1} - \bar{x}_1\right)^2 + \sum_{i=1}^{n}\left(x_{i2} - \bar{x}_2\right)^2 + \ldots + \sum_{i=1}^{n}\left(x_{il} - \bar{x}_l\right)^2 \tag{9.3}$$

Общая сумма квадратов отклонений наблюдаемых значений от общей средней $\bar{x}$:

$$C_{\text{общ}} = \sum_{j=1}^{m}\sum_{i=1}^{n}\left(x_{ij} - \bar{x}\right)^2 , \tag{9.4}$$

Можно доказать следующее равенство: $C_{\text{общ}} = C_{\text{факт}} + C_{\text{ост}}$.

С помощью дисперсий: $S_{\text{общ},}$, $S_{\text{факт}}, S_{\text{ост}}$ производится оценка общей, факторной и остаточной дисперсий:

$$S^2_{\text{общ}} = \frac{1}{mn-1} C_{\text{общ}}, \quad S^2_{\text{факт}} = \frac{1}{m-1} C_{\text{факт}}, \quad S^2_{\text{ост}} = \frac{1}{m(n-1)} C_{\text{ост}}. \tag{9.5}$$

В основе однофакторного дисперсионного анализа лежит тесная связь между различием в групповых средних $\bar{x}_j$ и соотношением между двумя видами дисперсий – факторной, которая характеризует влияние фактора А на величину Х, и остаточной, которая характеризует влияние случайных причин. Сравнивая факторную дисперсию с остаточной по величине их отношения судят, насколько сильно проявляется влияние фактора.

Для сравнения двух дисперсий используют ***показатель критерия Фишера***

$$F_{\text{эксп}} = S^2_{\text{факт}} / S^2_{\text{ост}}. \tag{9.6}$$

При этом при заданном уровне значимости проверяют *нулевую гипотезу о равенстве факторной и остаточной дисперсии* (изучаемый фактор не вызывает изменчивости признака) при конкурирующей гипотезе об их неравенстве (изучаемый фактор вызывает изменчивость признака).

По таблице критических значений распределения Фишера-Снедекора при уровне значимости, равном половине заданного уровня α, находят критическое значение $F_{kp}(\alpha/2; k_1; k_2)$. Здесь $k_1 = m-1;\ k_2 = m(n-1)$.

Если $F_{\text{эксп}} < F_{kp}$, нулевую гипотезу считают согласующейся с результатами наблюдений.

Если $F_{\text{эксп}} > F_{kp}$, то эту гипотезу отвергают в пользу конкурирующей.

- Если окажется, что $S^2_{\text{факт}} < S^2_{\text{ост}}$, следует сделать вывод об отсутствии влияния фактора А на Х.
- Если проверка покажет значимость различий между $S^2_{\text{факт}}$ и $S^2_{\text{ост}}$, следует сделать вывод о существенном влиянии фактора А на Х.

9.2. Примеры решения типовых задач.

Задача 9.1. Имеются данные о настриге шерсти овец в зависимости от их живой массы (табл. 9.2). Требуется определить достоверность разницы в настриге шерсти овец в зависимости от их живой массы с уровнем вероятности суждения 0,05.

Решение.

Для расчета показателей вариации настриг шерсти овец возведем в квадрат (табл. 9.3).

Т а б л и ц а 9.2

Настриг шерсти овец, кг

Живая масса овец, кг	Овцы								Сумма	Численность групп	Средний квадрат суммы
	1	2	3	4	5	6	7	8			
	x_{ij}								$\sum_{j=1}^{n_i} x_{ij}$	n_i	$\frac{\left(\sum_{j=1}^{n_i} x_{ij}\right)^2}{n_i}$
До 52,4	5,7	6,3	5,9	6,5	6,1	–	–	–	30,5	5	186,05
52,5–54,6	5,7	6,3	6,9	6,8	6,4	6,1	6,3	6,7	51,2	8	327,68
Свыше 54,6	7,2	7,1	6,9	6,3	7,0	6,8	7,1	–	48,4	7	334,65
Итого	×	×	×	×	×	×	×	×	$\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}$ = 130,1	$\sum_{i=1}^{k} n_i =$ = N =20	$\sum_{i=1}^{k}\left(\frac{\left(\sum_{j=1}^{n_i} x_{ij}\right)^2}{n_i}\right) =$ = 848,38

Т а б л и ц а 9.3

Квадрат настрига шерсти овец

Живая масса овец, кг	Овцы								Сумма квадратов
	1		3	4	5	6	7	8	
	x_{ij}^2								$\sum_{j=1}^{n_i} x_{ij}^2$
До 52,4	32,46	39,69	34,81	42,25	37,21	–	–	–	186,45
52,5–54,6	32,49	39,69	47,61	46,24	40,96	37,21	39,69	44,89	328,78
Свыше 54,6	51,84	50,41	47,61	39,69	49,00	46,24	50,41	–	335,2
Итого	×	×	×	×	×	×	×	×	$\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 =$ = 850,43

Показатели вариации будут равны:

общая вариация:

$$w_o = \sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}^2 - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\right)^2}{N} = 850,43 - \frac{130,1^2}{20} = 4,1295;$$

групповая вариация:

$$w_{\text{ãð}} = \sum_{i=1}^{k}\left(\frac{\left(\sum_{j=1}^{n_i} x_{ij}\right)^2}{n_i}\right) - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}\right)^2}{N} = 848,38 - \frac{130,1^2}{20} = 2,0809;$$

остаточная вариация:

$$w_{\text{îñò}} = w_{\text{î}} - w_{\text{ãð}} = 4,1295 - 2,0809 = 2,0486.$$

Рассчитаем число степеней свободы вариации:

общей: $\nu_o = N - 1 = 20 - 1 = 19$;

групповой: $\nu_{гр} = k - 1 = 3 - 1 = 2$;

остаточной вариации: $\nu_{\text{îñò}} = N - k = 20 - 3 = 17$.

Отсюда дисперсии будут равны:

общая: $s_{\text{î}}^2 = \dfrac{w_{\text{î}}}{\nu_{\text{î}}} = \dfrac{4,1295}{19} = 0,2173$;

групповая: $s_{\text{ãð}}^2 = \dfrac{w_{\text{ãð}}}{\nu_{\text{ãð}}} = \dfrac{2,0809}{2} = 1,0405$;

остаточная: $s_{\text{îñò}}^2 = \dfrac{w_{\text{îñò}}}{\nu_{\text{îñò}}} = \dfrac{2,0486}{17} = 0,1205$,

Фактическое значение F-критерия для групповой и остаточной дисперсий: $F_{\text{ôàêò}} = \dfrac{s_{\text{ãð}}^2}{s_{\text{îñò}}^2} = \dfrac{2,0809}{0,1205} = 8,63$,

Табличное значение F-критерия при уровне значимости 0,05, 2 степенях свободы для групповой дисперсии и 17 степенях свободы для остаточной дисперсии равно 3,59 (таблица «ЗначениеF-критерия Фишера при уровне значимости 0,05».

Результаты дисперсионного анализа представлены в табл.9.3.

Таблица 9.3

Однофакторный дисперсионный анализ

Источники вариации	Вариация (сумма квадратов отклонений)	Степень свободы вариации	Дисперсия	Отношение дисперсий	
				фактическое	табличное
Групповая	4,1295	19	0,2173	8,63	3,59
Остаточная	2,0809	2	1,0405	1	×
Общая	2,0486	17	0,1205	×	×

Данные таблицы показывают, что фактическое отношение дисперсий больше табличного, следовательно, разница в среднем настриге шерсти по группам овец с различной живой массой достоверна. Живая масса овец оказывает влияние на их настриг шерсти.

9.3. Задачи для самостоятельного решения.

1.Способом случайного отбора получены сведения по 18 хозяйствам области о дозах минеральных удобрений и урожайности сахарной свеклы. Данные сгруппированы по факторному признаку (табл. 9.4). Требуется, используя метод дисперсионного анализа, оценить влияние различных доз минеральных удобрений на урожайность сахарной свеклы.

Таблица 9.4.

Группировка 18 хозяйств области по количеству внесенных минеральных удобрений на 1 га посева сахарной свеклы

Группы хозяйств по количеству внесенных минеральных удобрений на 1 га, ц действующего вещества	Число хозяйств	Урожайность сахарной свеклы, ц/га
1,0-2,9	10	301,320,319,306,294,333,327,325,308,292
3,0-4,9	4	332,346,378,354
5,0 и более	4	380,369,406,352

2.Провести дисперсионный анализ данных полевого опыта (табл. 9.5), определить $НСР_{05}$ и сгруппировать сорта по отношению к стандарту.

Таблица 9.5.

Урожай озимой пшеницы, ц/га

Варианты сорта	Повторение, X				Суммы, V	Средние, $\bar{x}$
	I	II	III	IV		
1.Мироновская 808	47,8	46,9	45,4	44,1	184,2	46,0
2.Альбидум 114	53,7	50,3	50,6	48,0	202,6	50,6
3.Саратовская 10	46,7	42,0	43,4	40,7	172,8	43,2
4.Безостая 1	48,0	47,0	45,9	45,7	186,6	46,6
5.Одесская 51	41,8	40,0	43,0	41,6	166,4	41,6
Суммы Р	238,0	226,2	228,3	220,1	$\sum X = 912{,}6$	$\bar{x} = 45{,}6$

Вопросы для самопроверки.

1.Сущность дисперсионного анализа.

2.Модель однофакторного анализа.

3.Проверка нулевой гипотезы при дисперсионном анализе.

4. Как определить общую, факторную и остаточные суммы?

5. Как определить обобщенную ошибку среднего, ошибку разности?

6. Для чего производится статистическая обработка результатов исследования?

7. Как сделать вывод о существенности различий между средними по вариантам?

Глава 10. Элементы корреляционного анализа.

10.1. Необходимый теоретический минимум.

Статистическая и корреляционная зависимости случайных величин. Уравнения регрессии.

Определение 10.1. Зависимость между случайными величинами *X* и *Y*, состоящая в том, что каждому значению одной величины соответствует распределение другой величины, называется ***статистической зависимостью***.

- К переменным величинам, связанным статистической зависимостью относятся, например: удой и расход кормов; урожайность и количество внесённых удобрений, рост человека и его масса, объём дневной продукции и её себестоимость по различным предприятиям.

Чтобы изучить статистическую зависимость, нужно знать условное математическое ожидание случайной величины. На практике часто используют связь между изменениями одной случайной величины *X* и изменениями математического ожидания (средней) другой величины *Y*, то есть $M(Y/X=x)=f(x)$ или $M(X/Y=y)=\varphi(y)$.

Определение 10.2. Зависимость между двумя переменными величинами состоящая в том, что при изменении одной величины изменяется среднее значение (математическое ожидание) другой величины называется ***корреляционной зависимостью*** (от англ. correlation - согласование, связь, взаимосвязь, соотношение, взаимозависимость).

В математической статистике имеем дело с оценками числовых характеристик, поэтому в качестве оценки условного математического ожидания принимают условное среднее. *Условным средним* $\overline{y}_x$ называется среднее арифметическое наблюдаемых значений *Y*, соответствующих $X=x$.

Например, если при X=3, величина Y принимает значения $y_1=2$, $y_2=7$, $y_3=6$, то $\bar{y}_x = \dfrac{2+7+6}{3} = 5$. Условные средние являются функциями соответственно от x и y.

Выборочным уравнением регрессии Y на X называется зависимость вида:

$$\bar{y}_x = f^*(x) \qquad (10.1)$$

Соответственно, ***выборочным уравнением*** регрессии X на *Y* *называется зависимость вида*

$$\bar{x}_y = \varphi^*(y) \qquad (10.2)$$

Графики соответствующих функций $f^*(x)$, $\varphi^*(y)$ называются ***выборочными линиями регрессии.***

Статистическую связь между переменными можно изучать методами корреляционного и регрессионного анализа.

Основной задачей *регрессионного* анализа является установление формы зависимости между переменными и изучение этой зависимости.

Основная задача *корреляционного* анализа – выявление связи между переменными и оценка её тесноты. Для решения этих задач используют соответствующий математический аппарат. Рассмотрим задачу в общем виде.

Пусть имеется два ряда наблюдаемых зависимых между собой величин *X* и *Y*. Если x_i и y_i встречаются по одному разу, то их значения записывают в виде таблицы 10.1.

Таблица 10.1

X_i	x_1	x_2	x_3	…	x_k	…	x_n
Y_i	y_1	y_2	y_3	…	y_k	…	y_n

Если каждому значению y_i отвечает несколько значений *x*, а каждому x_j – несколько значений *y*, то эти данные записывают в виде таблицы 10.2 .

Таблица 10.2

Y / X	x_1	x_2	$\cdots$	x_k	$\sum\limits_{j=1}^{k}$
y_1	n_{11}	n_{12}	$\cdots$	n_{1k}	$\sum n_{1j} = l_1$
y_2	n_{21}	n_{22}	$\cdots$	n_{2k}	$\sum n_{2j} = l_2$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
y_m	n_{m1}	n_{m2}	$\cdots$	n_{mk}	$\sum n_{mj} = l_m$
$\sum\limits_{i=1}^{m}$	$\sum n_{i1} = n_1$	$\sum n_{i2} = n_2$	$\cdots$	$\sum n_{ik} = n_k$	$n = \sum\limits_{i=1}^{m} l_i = \sum\limits_{j=1}^{k} n_j$

- В таблице 10.2. n_{ij} – частота, показывающая число повторений парные значений x_i и y_j . Если парные значения не появляются ни разу, то в соответствующей клетке ставят прочерк.

Таблица 10.2, в которой результаты наблюдений записаны в порядке возрастания с указанием частоты n_{ij}, называется ***корреляционной таблицей.***

Корреляционная таблица может быть составлена как для дискретных так и для непрерывных признаков. В последнем случае строят интервальный ряд для значений *X, Y* затем переходят к дискретному ряду, заменяя соответствующий интервал $(x_{i-1};x_i)$ его серединой $x_i* = \dfrac{x_{i-1} + x_i}{2}$. Также поступают со случайной величиной *Y*.

Графически, если взаимосвязь двух признаков задана таблицей 10.1, то в системе координат строят точки $(x_i;y_i)$ и по расположению точек судят о форме зависимости (линейная, показательная и т.д.). Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определённой линии, выражающей форму связи (см. рис. 10.1).

- Данные корреляционной таблицы также можно изобразить графически. Для этого каждую пару значений $(x_i;y_j)$ изображают точкой в системе координат; частоту n_{ij}, с которой эта пара встречается в таблице, изображают в виде числа, помещённого возле соответствующей точки.
- Графическое изображение корреляционной таблицы называется ***полем корреляции***. На таком графике наглядно можно проследить общую тенденцию в изменении *Y* с изменением *X* (см. рис. 10.3).

Отыскание параметров выборочного уравнения линейной регрессии

Пусть известны результаты опыта, целью которого является исследование зависимости между двумя величинами X и Y (например, величина прибыли и объём инвестиций, изменение по месяцам курса доллара и т. д.). Предположим, что различные значения x признака X и соответствующие значения y признака Y наблюдаются по одному разу (см. табл. 10.1). Строим точки $(x_i;y_i)$ на координатной плоскости.

Предположим, что точки расположены приближённо вдоль некоторой прямой, уравнение которой ищем в виде:

$$y=a+bx\,. \qquad (10.1)$$

Среди всех прямых линий $y=a+bx$ ищут наиболее близкую к данной системе точек, причём близость будем измерять суммой квадратов отклонений: $$S=\sum_{i=1}^{n}\left[y_i-\left(a+bx_i\right)\right]^2.$$

Из всех прямых выбирают ту, для которой сумма сумма квадратов отклонений S минимальна.

Так как минимизируется сумма квадратов отклонений экспериментальных и теоретических значений, то предложенный метод называют *методом наименьших квадратов*. А прямую построенную по этому методу, называют *МНК-прямой*. Для определения параметров уравнения прямой линии y на x , получают систему *нормальных уравнений*.

Система нормальных уравнений для отыскания неизвестных параметров прямой линии регрессии имеет вид:

$$\begin{cases}\left(\sum_{i=1}^{n}x_i^{\,2}\right)\cdot b+\left(\sum_{i=1}^{n}x_i\right)\cdot a=\sum_{i=1}^{n}x_i\cdot y_i\\ \left(\sum_{i=1}^{n}x_i\right)\cdot b+na=\sum_{i=1}^{n}y_i\end{cases}. \qquad (10.2)$$

Решая систему (10.2) получим формулы для определения неизвестных параметров регрессии a и b (или *МНК– оценок*):

$$b=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\quad;\qquad a=\bar{y}-b\bar{x}\ . \tag{10.3}$$

Преобразуя формулу (10.3) для нахождения параметра b, получим:

$$b=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\overline{x^2}-\left(\bar{x}\right)^2}\quad;\qquad a=\bar{y}-b\bar{x}, \tag{10.3(a)}$$

где $\bar{x}=\frac{\sum x}{n};\quad \bar{y}=\frac{\sum y}{n};\quad \overline{xy}=\frac{\sum xy}{n};\quad \overline{x^2}=\frac{\sum x^2}{n}$.

Уравнение регрессии *Y* по *X* можно также записать в виде:

$$y=b\left(x-\bar{x}\right)+\bar{y}\ . \tag{10.4}$$

Угловой коэффициент b в уравнении регрессии называется *выборочным коэффициентом регрессии Y по X* . Обозначим его b_{yx}, тогда уравнение регрессии можно записать в виде:

$$y=b_{yx}\left(x-\bar{x}\right)+\bar{y}\quad \text{или}\quad y-\bar{y}=b_{yx}\left(x-\bar{x}\right). \tag{10.5}$$

Коэффициент регрессии *Y* по *X* показывает на сколько единиц в *среднем* изменяется переменная *Y* при увеличении переменной *X* на одну единицу.

Из уравнений (10.5) видно, что прямая регрессии проходит через точку $(\bar{x};\bar{y})$, которая является центром тяжести системы данных точек.

Аналогично можно получить уравнение прямой регрессии *X* по *Y* :

$$x-\bar{x}=b_{xy}\left(y-\bar{y}\right)\ . \tag{10.6}$$

В формуле (10.6) b_{xy} – *выборочный коэффициент регрессии X по Y:*

$$b_{xy}=\frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\overline{y^2}-\left(\bar{y}\right)^2}\ . \tag{10.7}$$

Формулы для нахождения коэффициентов регрессии можно переписать следующим образом:

$$b_{yx} = \frac{\mu}{S_x^{\ 2}}\ ; \qquad b_{xy} = \frac{\mu}{S_y^{\ 2}}; \tag{10.8}$$

где:

$\mu = \overline{xy} - \overline{x} \cdot \overline{y}$ – *выборочная ковариация,*

$S_x^{\ 2} = \overline{x^2} - \left(\overline{x}\right)^2$ – *выборочная дисперсия* по переменной *X,*

$S_y^{\ 2} = \overline{y^2} - \left(\overline{y}\right)^2$ – *выборочная дисперсия* по переменной *Y*.

В уравнениях регрессий b_{yx} и $\dfrac{1}{b_{xy}}$ – угловые коэффициенты линий регрессий, которые пересекаются в точке $(\bar{x}; \bar{y})$.

Предположим, что зависимость между *X* и *Y* задана корреляционной таблицей 10.2. Тогда в формулах для нахождения параметров уравнений регрессий появляются *веса*, задающие число одинаковых слагаемых:

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{n}; \quad \overline{y} = \frac{\sum_{j=1}^{m} y_j \cdot l_j}{n};$$

$$\overline{xy} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} x_i \cdot y_j \cdot n_{ij}}{n}; \quad \overline{x^2} = \frac{\sum_{i=1}^{k} x_i^{\ 2} \cdot n_i}{n}. \tag{10.9}$$

Эмпирический коэффициент корреляции и его свойства

Предположим, что наблюдения проводят над системой двух случайных величин *X* и *Y*, связанных линейной зависимостью. Для оценки тесноты связи между ними по результатам выборки вычисляют статистику, называемую *коэффициентом корреляции*.

Коэффициент корреляции вычисляют по формулам:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}; \tag{10.10}$$

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{S_x S_y} \quad \text{или} \quad r = \frac{\mu}{S_x S_y}. \qquad (10.11)$$

В формулах (10.11) S_x, S_y – выборочные средние квадратические отклонения соответственно по переменным X и Y, определяемые по формулам:

$$S_x = \sqrt{S_x^{\ 2}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} = \sqrt{\overline{x^2} - (\overline{x})^2};$$

$$S_y = \sqrt{S_y^{\ 2}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}} = \sqrt{\overline{y^2} - (\overline{y})^2}.$$

Если известны коэффициенты регрессии b_{yx} и b_{xy}, то коэффициент корреляции вычисляется по формуле:

$$r = \pm\sqrt{b_{yx} \cdot b_{xy}}\,. \qquad (10.12)$$

Коэффициент корреляции есть ***среднее геометрическое*** коэффициентов регрессии и имеет их знак.

- Если выборочные данные заданы в виде корреляционной таблицы, то при вычислении коэффициента корреляции учитывают кратность значений ((см.10.9)).

Свойства коэффициента корреляции:

1.Значения коэффициента корреляции изменяются на отрезке [-1;1]: $-1 \le r \le 1$.

2. Чем больше r , тем теснее связь между изучаемыми признаками.

- В зависимости от того насколько $|r|$ близок к 1, различают связь *слабую, умеренную, заметную, достаточно тесную, тесную и весьма тесную.*

3. Если $|r| = 1$, то корреляционная связь становится функциональной.

4. Если $r = 0$, то между изучаемыми признаками нет линейной корреляционной зависимости, но не исключается существование какого либо другого вида корреляционной зависимости (параболической, гиперболической, показательной и т.д.).

5. Если $r > 0$, то связь между признаками *X, Y* прямая; если $r < 0$, то эта связь обратная.

- Выборочный коэффициент корреляции r является оценкой коэффициента корреляции ρ генеральной совокупности и поэтому также служит для измерения линейной связи между количественными признаками.

- Если выборка имеет достаточно большой объём $(n \geq 50)$ и репрезентативна, то интервальная оценка коэффициента корреляции имеет вид:

$$r - t_{1-\alpha}\frac{\sqrt{1-r^2}}{\sqrt{n-2}} \leq \rho \leq r + t_{1-\alpha}\frac{\sqrt{1-r^2}}{\sqrt{n-2}}. \qquad (10.13)$$

- В двойном неравенстве (10.13) α – это уровень значимости, $\Phi(t_{1-\alpha}) = 1 - \alpha = \gamma$ – доверительная вероятность. Значение $t_{1-\alpha}$ находят по таблице значений интегральной функции Лапласа (*таблица* 3 *Приложения* 3).

Коэффициент корреляции, возведенный в квадрат, называется ***коэффициентом детерминации r²***. Он показывает долю изменений, которые вызваны факторным признаком. Коэффициент детерминации r^2 является прямым способом выражения зависимости одного признака от другого. Если известно, что У находится в причинной связи с Х, то r^2 – это доля вариации У, обусловленная влиянием Х.

Для изучения корреляционных связей большое значение имеет коэффициент регрессии ρ, который показывает, насколько в среднем изменяется признак (Х), если коррелирующий с ним признак (У) изменяется на определенную величину.

Корреляционные зависимости наблюдаются между очень многими признаками организмов – морфологическими, физиологическими, а также между различными биологическими процессами. Различают положительную и отрицательную корреляции. При положительной корреляции с увеличением одного признака увеличивается и другой. Например, с увеличением живой массы коров первотёлок возрастает и удой; чем выше процент жира в молоке, тем выше и процент белка в нём. При отрицательной корреляции с увеличением удоя у коров снижается жирность молока; куры с высокой яйценоскостью имеют более мелкие яйца.

В зоотехнической и ветеринарной практике изучение корреляционной зависимости имеет большое значение. Так, например, для животновода очень важно знать, какова связь между средним удоем за лактацию и процентом жира в молоке, иначе говоря, дают ли более высокоудойные коровы молоко с повышенным содержанием жира или, наоборот, с пониженным и насколько часто встречаются исключения из той или другой зависимости. Или другой пример: из-за отрицательной корреляционной зависимости между высокой молочностью и способностью к откорму невозможно выведение породы, сочетающую высокую молочную продуктивность и высокие мясные качества; между устойчивостью к эймериозу у кур и массой тела существует положительная корреляция, поэтому, чем более упитанные куры, тем менее они предрасположены к заболеванию.

Надежность зависимости

Определяют две черты зависимости между переменными: величину зависимости и надежность зависимости.

Надежность зависимости – менее наглядное понятие, чем величина зависимости, однако чрезвычайно важна. Оно непосредственно связано с репрезентативностью той определенной выборки, на основе которой строятся выводы. Другими словами, надежность говорит, насколько вероятно, что зависимость подобная найденной, будет вновь обнаружена (подтвердится) на данных другой выборки, извлеченной из той же самой популяции. Если исследование удовлетворяет некоторым специальным критериям, то надежность найденных зависимостей между переменными выборки можно количественно оценить и представить с помощью стандартной статистической меры (называемой p-уровень, или статистический уровень значимости).

Статистическая значимость результата представляет собой оцененную меру уверенности в его правильности. Уровень значимости или p-уровень, - это показатель, находящийся в убывающей зависимости от надежности результата. Более высокий p-уровень соответствует более низкому уровню доверия к найденной в выборке зависимости между переменными. Именно *p*-уровень представляет собой вероятность ошибки, связанной с распространением наблюдаемого результата на всю популяцию. *Чем слабее зависимость между переменными, тем большего объема требуется выборка, чтобы значимо ее обнаружить.* Другими словами, если зависимость между переменными почти отсутствует, объем выборки, необходимый для ее значимого обнаружения, почти равен объему всей популяции, которой предполагается бесконечным.

Так как надежность изучения связей в значительной степени зависит от количества сопоставляемых данных, необходимо измерять существенность полученного уравнения регрессии и индекса (коэффициента) корреляции. Показатели корреляции, исчисленные для ограниченной по объему совокупности, могут быть искажены действием случайных факторов.

Существенность индекса (коэффициента) корреляции, а, следовательно, всего уравнения регрессии, может быть оценена с помощью дисперсионного анализа (F-критерия Фишера). При этом сравнивают факторную и остаточную дисперсии с учетом числа степеней свободы вариации. F-критерий в данном случае рассчитывают по формуле:

$F_{факт} = \frac{s^2_{y_x}}{s^2_{\varepsilon}}$, где $s^2_{y_x} = \frac{\sum(\bar{y}_x - y)^2}{n-1}$ – выборочная факторная дисперсия;

$s^2_{\varepsilon} = \frac{\sum(y - \bar{y}_x)^2}{n-k}$ – выборочная остаточная дисперсия;

n– численность выборочной совокупности;

k – число параметров в уравнении регрессии.

Значение F-критерия можно получить также, используя значения индекса или коэффициента корреляции:

$$F_{факт} = \frac{I^2(n-k)}{(1-I^2)(k-1)}; \quad F_{факт} = \frac{r^2(n-k)}{(1-r^2)(k-1)}.$$

Полученное значение F-критерия сравнивают с табличным значением. При этом для факторной дисперсии число степеней свободы вариации составляет $\nu_{y_x} = k - 1$, а для остаточной дисперсии $\nu_\varepsilon = n - k$. *Если фактическое значение F-критерия больше табличного, следовательно, связь между признаками достоверна и уравнение регрессии в полной мере отражает эту связь. Если фактическое значение F-критерия меньше табличного, то можно сделать вывод, что связь между признаками носит случайный характер.*

Для оценки значимости индекса (коэффициента) корреляции и уравнения регрессии также используют *t*-критерий Стьюдента, который для больших выборок рассчитывают по формулам:

$$t_{факт} = \frac{I\sqrt{n-k}}{1-I^2};\ t_{факт} = \frac{r\sqrt{n-k}}{1-r^2}.$$

Для малых выборок формулы имеют вид:

$$t_{факт} = \frac{I\sqrt{n-k}}{\sqrt{1-I^2}};\ t_{факт} = \frac{r\sqrt{n-k}}{\sqrt{1-r^2}}.$$

Также, как при дисперсионном анализе, фактическое значение *t*-критерия сравнивают с табличным с учетом числа степеней свободы вариации $\nu = n - k$. *Если фактическое значение t-критерия больше табличного, то связь достоверна, если меньше, то связь несущественна.*

Рассмотрим методику корреляционного анализа для парной корреляции.

Пример. По выборочным данным получены сведения о среднегодовом удое коров и расходе кормов на голову (табл. 10.3).

Требуется определить зависимость удоя молока от уровня кормления коров. Анализ данных показывает, что с увеличением расхода кормов среднегодовой удой молока от коровы повышается. Это подтверждает и график (рис. 10.1).

Расположение точек на графике показывает, что связь между признаками имеет прямолинейный характер и поэтому может быть выражена уравнением прямой линии:

$$y_x = a_0 + a_1 x.$$

Для определения неизвестных параметров уравнения a_0 и a_1 необходимо решить систему нормальных уравнений:

$$\begin{cases} \sum y = na_0 + a_1 \sum x \\ \sum xy = a_0 \sum x + a_1 \sum x^2 \end{cases}$$

Т а б л и ц а 10.3

Среднегодовой удой коровы и расход кормов на корову

№ п/п	Среднегодовой удой коровы, кг	Расход кормов на корову, корм. ед.	Произведение вариант	Квадрат среднегодового удоя	Квадрат расхода кормов	Ожидаемый среднегодовой удой коровы, ц
	y	x	xy	y^2	x^2	y_x
1	3003	3431	10303293	9018009	11771761	3358
2	3225	3417	11019825	10400625	11675889	3339
3	3231	3526	11392506	10439361	12432676	3486
4	3622	3532	12792904	13118884	12475024	3494
5	3781	3421	12934801	14295961	11703241	3344
6	3205	3588	11499540	10272025	12873744	3570
7	3537	3571	12630627	12510369	12752041	3547
8	3859	3690	14239710	14891881	13616100	3708
9	3671	3585	13160535	13476241	12852225	3566
10	3793	3636	13791348	14386849	13220496	3635
11	3884	3714	14425176	15085456	13793796	3740
12	3706	3635	13471310	13734436	13213225	3633
13	3628	3687	13376436	13162384	13593969	3703
14	4049	3919	15868031	16394401	15358561	4017
15	4260	3977	16942020	18147600	15816529	4095
16	4113	3958	16279254	16916769	15665764	4069
17	4435	4264	18910840	19669225	18181696	4482
18	4356	4145	18055620	18974736	17181025	4322
19	4278	4174	17856372	18301284	17422276	4361

20	4522	4352	19679744	20448484	18939904	4601
Итого	$\sum y =$ = 76158	$\sum x =$ = 75222	$\sum xy =$ = 288629892	$\sum y^2 =$ = 293644980	$\sum x^2 =$ = 284539942	×

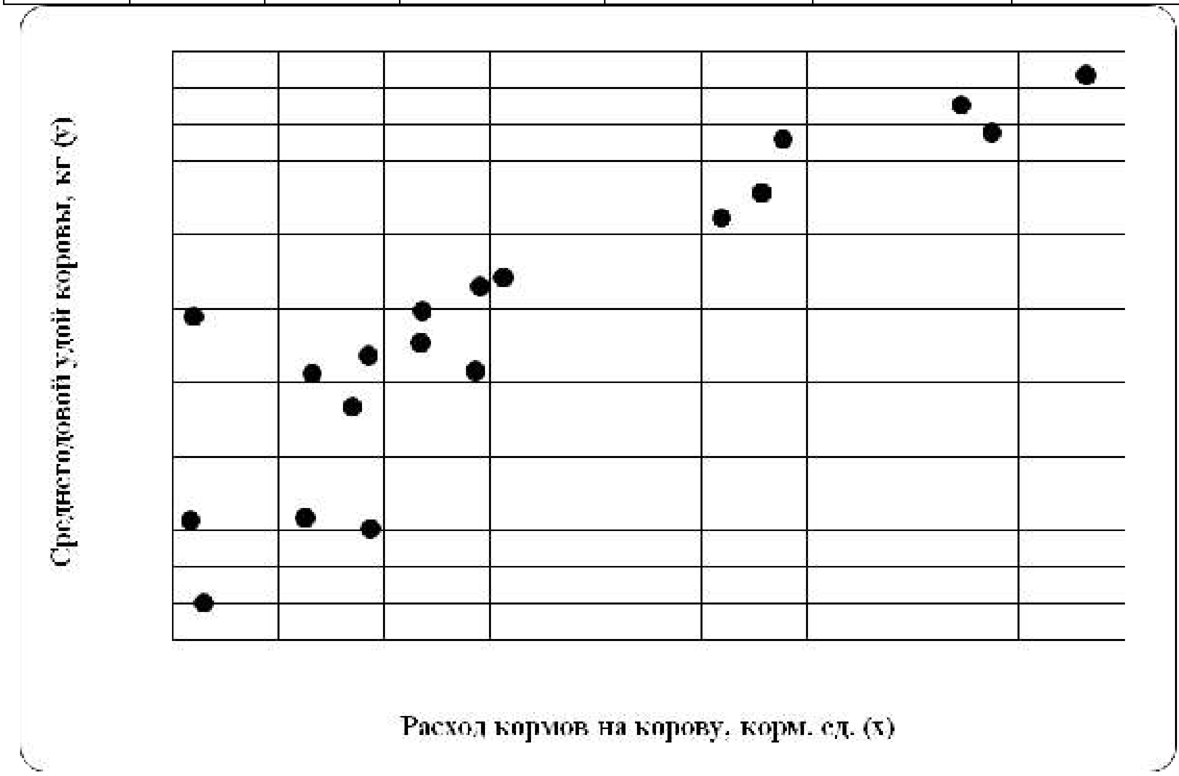


Рисунок 10.1 Зависимость удоя коров от расхода кормов

Значения $\sum y$, $\sum x$, $\sum xy$ и $\sum x^2$ определим по данным наблюдений (табл. 14) и подставим в уравнения:

$$\begin{cases} 76158 = 20a_0 + 75222a_1 \\ 288629892 = 75222a_0 + 284539942a_1 \end{cases}$$

После решения системы уравнений получим значения параметров:

$a_0 = -1274$; $a_1 = 1{,}35$.

Уравнение регрессии имеет вид:

$y_x = -1274 + 1{,}35x$.

Величина a_0 = −1274 в уравнении регрессии не имеет смысла. Коэффициент регрессии a_1 = 1,35 характеризует изменение продуктивности коров по данной совокупности в зависимости от уровня кормления. При увеличении или уменьшении расхода кормов на 1 корм. ед. среднегодовой удой коровы, соответственно увеличивается или уменьшается в среднем на 1,35 кг.

Полученное уравнение регрессии, кроме оценки влияния уровня кормления на продуктивность коров, позволяет прогнозировать ее в зависимости от величины данного фактора. При этом, уровень кормления должен находиться в пределах его изменения в исходной выборочной совокупности. Ожидаемый удой молока в зависимости от расхода кормов представлен в последней графе табл. 14.

Для оценки тесноты связи рассчитаем коэффициент корреляции:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - \left(\sum x\right)^2\right)\left(n\sum y^2 - \left(\sum y\right)^2\right)}} =$$

$$r = \frac{20 \cdot 288629892 - 75222 \cdot 76158}{\sqrt{\left(20 \cdot 284539942 - 75222^2\right) \cdot \left(20 \cdot 293644980 - 76158^2\right)}} = 0{,}902 \cdot$$

Его значение близко к единице, поэтому можно утверждать, что полученное уравнение регрессии достаточно хорошо описывает исследуемую зависимость. Коэффициент детерминации r^2 = 0,813 показывает, что 81,3 % колеблемости в среднегодовом удое коровы объясняется уровнем кормления.

Оценим достоверность коэффициента корреляции с помощью *F*-критерия и *t*-критерия Стьюдента.

Фактическое значение *F*-критерия равно:

$$F_{факт} = \frac{r^2(n-k)}{(1-r^2)(k-1)} = \frac{0{,}902^2 \cdot (20-2)}{(1-0{,}902^2) \cdot (2-1)} = 78{,}57 \cdot$$

Табличное значение *F*-критерия при уровне значимости 0,05 и при $v_1 = k - 1 = 2 - 1 = 1$ и $v_2 = n - k = 20 - 2 = 18$ степенях свободы вариации составляет 4,41 (таблица «Значение*F*-критерия Фишера при уровне значимости 0,05»).

Фактическое значение критерия выше табличного, поэтому с вероятностью 0,95 можно утверждать, что связь между признаками достоверна, и уравнение регрессии в полной мере отражает эту связь.

Фактическое значение t-критерия равно:

$$t_{факт} = \frac{r\sqrt{n-k}}{\sqrt{1-r^2}} = \frac{0{,}902\cdot\sqrt{20-2}}{\sqrt{1-0{,}902^2}} = 8{,}86\cdot$$

Табличное значение t-критерия при уровне значимости 0,05 и при $\nu = n - k = 20 - 2 = 18$ степенях свободы вариации составляет 2,1009 (таблица «Значениеt-критерия Стьюдента при уровне значимости 0,10, 0,05 и 0,01»).

Фактическое значение критерия выше табличного, следовательно, вывод о достоверности связи между признаками подтверждается.

10.2. Решение типовых задач.

Задача 10.1. Изучается зависимость себестоимости единицы изделия (*Y*, тыс. руб.) от величины выпуска продукции (*X*, шт.) по группам предприятий за отчётный период. Экономист обследовал пять предприятий и получил следующие данные:

X	2	3	4	5	6
Y	1,9	1,7	1,8	1,6	1,4

Найти линейную регрессию *Y* на *X* и выборочный коэффициент корреляции.

Решение.

Построим данные точки ($x; y$) в системе координат (рис. 10.2).

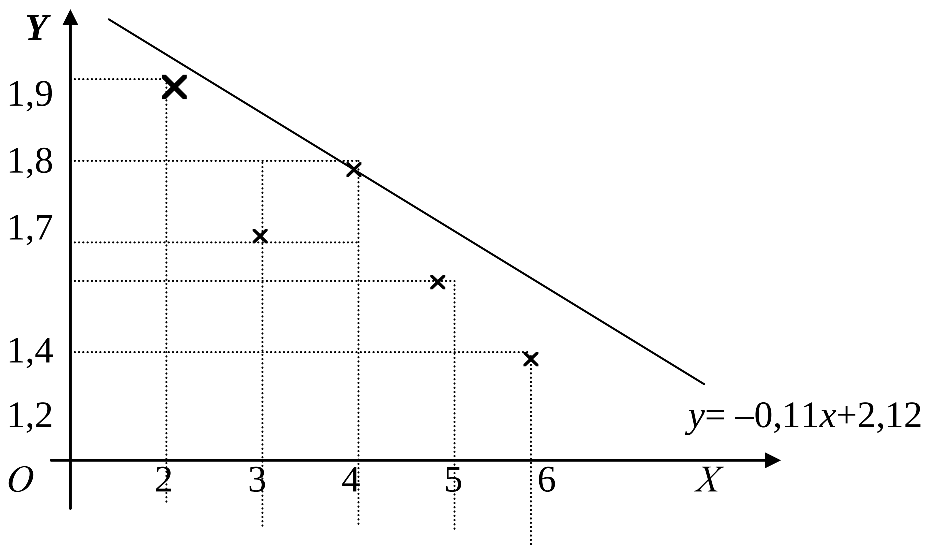


Рисунок 10.2.

Уравнение регрессии ищем в виде: $y = a + bx$, параметры b и a можно найти по одной из формул ((10.5), (10.6), (10.6*a)*). Так как данные при расчётах не требуют округления, то будем использовать формулу (10.6):

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}; \qquad a = \bar{y} - b\bar{x}.$$

Обозначим: $\Delta x_i = x_i - \bar{x}; \quad \Delta y_i = y_i - \bar{y}$, тогда имеем:

$$b = \frac{\sum_{i=1}^{n}\Delta x_i \Delta y_i}{\sum_{i=1}^{n}\Delta x_i^{\,2}}.$$

Найдём среднюю выборочную по *X* и по *Y*. Для этого используем суммы, записанные в последней строке первого и второго столбцов:

Заполняем таблицу с четвёртого по восьмой столбец. В столбцах 6,7,8 подсчитываем суммы. Суммы, полученные по столбцам 6 и 8, подставляем в формулу для нахождения b:

$$b = \frac{-1{,}10}{10} = -0{,}11; \qquad a = 1{,}68 - (-0{,}11)\cdot 4 = 2{,}12.$$

$$\bar{x} = \frac{\sum x}{n} = \frac{20}{5} = 4; \quad \bar{y} = \frac{\sum y}{n} = \frac{8{,}4}{5} = 1{,}68.$$

Таблица 10.3

Для вычислений удобно составить расчетную таблицу:

№	x_i	y_i	Δx_i	Δy_i	Δx_i^2	Δy_i^2	$\Delta x_i \Delta y_i$	$\hat{y}_i$	Δ_i	Δ_i^2
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	2	1,9	-2	0,22	4	0,0484	-0,44	1,9	0	0
2	3	1,7	-1	0,02	1	0,0004	-0,02	1,79	-0,09	0,0081
3	4	1,8	0	0,12	0	0,0144	0	1,68	0,12	0,0144
4	5	1,6	1	-0,08	1	0,0064	-0,08	1,57	0,03	0,0009
5	6	1,4	2	-0,28	4	0,0784	-0,56	1,46	-0,06	0,0036
$\sum$	20	8,4	-	-	10	0,1480	-1,10	-	-	0,0270

Тогда искомое уравнение регрессии: <u>*y=2,12 - 0,11x*</u>, где

$b_{yx} = -0{,}11$ – коэффициент регрессии *Y* по *X*.

Теперь по уравнению регрессии рассчитаем теоретические значения $\hat{y}_i$ и заполним столбец 9. Например: $y_1 = 2{,}12 - 0{,}11 \cdot 2 = 1{,}9$.

Далее найдём отклонения наблюдаемых значений от теоретических: $\Delta_i = y_i - \hat{y}_i$. Затем определим квадраты этих отклонений (то есть заполним столбцы 10 и 11 расчетной таблицы). Заметим, что в последнем столбце получили сумму квадратов отклонений, равную Δ_i^2 =0,027. Так как в сравнении с данными число мало, следовательно выполнено условие МНК.

Коэффициент корреляции вычислим по формуле (5.13):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n}\Delta x_i \cdot \Delta y_i}{\sqrt{\sum_{i=1}^{n}\Delta x_i^{\,2} \cdot \sum_{i=1}^{n}\Delta y_i^{\,2}}}.$$

Итак, получаем: $r = \dfrac{-1{,}1}{\sqrt{10 \cdot 0{,}148}} = -0{,}904$.

Вывод: связь между признаками весьма тесная, обратная, т.е. чем больше продукции выпущено, тем меньше её себестоимость.

Задача 10.2. В таблице приведены данные о количестве внесённых удобрений *X* (ц/га) и урожайность *Y*(ц/га) на 100га пахотной земли:

X /Y	10	12	14	16	18	20	n_i
10	9	4	1				14
30	1	10	9	3			23
50		2	6	14	6		28
70			1	10	18	6	35
n_j	10	16	17	27	24	6	100

Требуется: 1.Найти коэффициент корреляции, сделать вывод о тесноте связи. 2.Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3. Определить величину среднего урожая, если внести по 80ц удобрений на гектар.4. Найти, сколько надо внести удобрений, чтобы получить урожай 15ц с гектара.

Решение.

1.Для нахождения коэффициента корреляции используем формулу (5.14):

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{S_x \cdot S_y} \quad \text{или} \quad r = \frac{\mu}{S_x \cdot S_y}.$$

Найдём средние выборочные значения и средние квадратические отклонения переменных X и Y:

$$\bar{x} = \frac{1}{100}(10 \cdot 14 + 30 \cdot 23 + 50 \cdot 28 + 70 \cdot 35) = 46{,}80;$$

$$\bar{y} = \frac{1}{100}(10 \cdot 10 + 12 \cdot 16 + 14 \cdot 17 + 16 \cdot 27 + 18 \cdot 24 + 20 \cdot 6) = 15{,}14;$$

$$S_x{}^2 = \overline{x^2} - \left(\overline{x}\right)^2; \quad S_x = \sqrt{S_x{}^2}; \quad S_y{}^2 = \overline{y^2} - \left(\overline{y}\right)^2; \quad S_y = \sqrt{S_y{}^2}.$$

$$S_x{}^2 = \frac{1}{100}\left(10^2 \cdot 14 + 30^2 \cdot 23 + 50^2 \cdot 28 + 70^2 \cdot 35\right) - (46{,}80)^2 = 445{,}76;$$

Таким образом: $S_x = \sqrt{445{,}76} \approx 21{,}113$.

$$S_y{}^2 = \frac{1}{100}\left(10^2 \cdot 10 + 12^2 \cdot 16 + 14^2 \cdot 17 + 16^2 \cdot 27 + 18^2 \cdot 24 + 20^2 \cdot 6\right) -$$

$-(15{,}14)^2 = 8{,}0204$. Значит: $S_y = \sqrt{8{,}0204} \approx 2{,}832.$

Чтобы найти ковариацию μ вычислим величину $\overline{xy}$.

$$\overline{xy} = \frac{1}{100}(10 \cdot 10 \cdot 9 + 10 \cdot 30 \cdot 1 + 10 \cdot 12 \cdot 4 + 30 \cdot 12 \cdot 10 + 50 \cdot 12 \cdot 2 + 14 \cdot 10 \cdot 1 + 30 \cdot 14 \cdot 9 +$$
$$+ ? \cdot 50 \cdot 6 + 70 \cdot 14 \cdot 1 + 16 \cdot ? \cdot 3 + 50 \cdot 16 \cdot 14 \cdot + ? \cdot ? \cdot 10 + 18 \cdot 50 \cdot 6 +$$
$$+ 18 \cdot ? \cdot 18 + ? \cdot 70 \cdot ? \;) = 759.$$

Тогда получаем, что $\mu = \overline{xy} - \overline{x} \cdot \overline{y} = 759 - 46{,}80 \cdot 15{,}14 = 50{,}448.$

Следовательно, коэффициент вариации $r = \dfrac{50{,}448}{21{,}113 \cdot 2{,}832} \approx 0{,}844$

Вывод: связь между признаками тесная, прямая.

2.Найдём коэффициенты регрессии, используя формулы (5.11):

$$b_{yx} = \frac{\mu}{S_x^{\ 2}} = \frac{50{,}448}{445{,}76} \approx 0{,}113; \qquad b_{xy} = \frac{\mu}{S_y^{\ 2}} = \frac{50{,}448}{8{,}0204} \approx 6{,}290.$$

Составим теперь уравнения регрессий:

а) уравнение регрессии Y на X:

$$y - \bar{y} = b_{yx}(x - \bar{x}) \Rightarrow y - 15{,}14 = 0{,}113(x - 46{,}8) \Rightarrow$$
$$\Rightarrow \ y = 0{,}113x + 9{,}852.$$

б) уравнение регрессии X на Y :

$$x - \bar{x} = b_{xy}(y - \bar{y}) \Rightarrow x - 46{,}80 = 6{,}290(y - 15{,}14) \Rightarrow$$

$\Rightarrow x = 6{,}290y - 48{,}431.$ Графики найденных регрессий приведены на рис. 10.3.

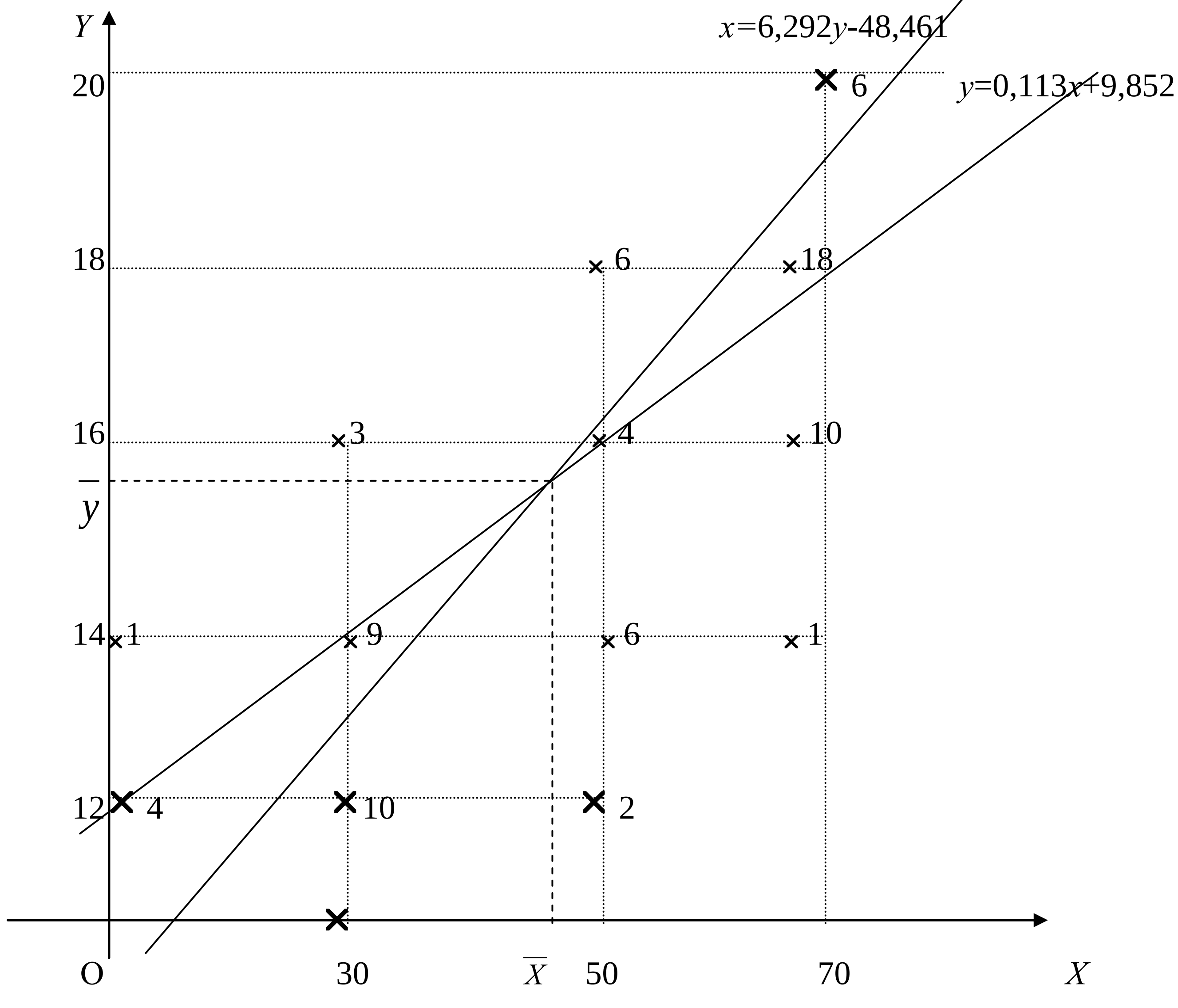


Рисунок 10.3.

3.Чтобы найти урожайность, используя уравнение регрессии Y на X, найдём значение y при x=80: $y = 0{,}113 \cdot 80 + 9{,}852 = 18{,}892 \approx 19\ (ц/га)$.

Значит, если будет внесено на 1 гектар 80 центнеров удобрений, то получим урожай примерно 19 центнеров с гектара.

4.Требуемое количество удобрений найдем, используя уравнение регрессии X на Y: $x = 6{,}290 \cdot 15 - 48{,}431 = 45{,}919 \approx 46$. Для достижения урожайности в 15 ц/га необходимо внести 46 ц/га удобрений.

10.3. Задачи для самостоятельного решения.

1.По исходным данным графически определить характер зависимости. Найти уравнение прямой регрессии Y на X и построить эту прямую.

X	1	1,5	3	4,5	5
Y	1,25	1,4	1,5	1,75	2,25

2.Опрос случайно выбранных 10 студентов вуза позволяет выявить зависимость между средним баллом по результатам предыдущей сессии и числом часов в неделю, затраченных студентом на самостоятельную подготовку. Соответствующие данные представлены в таблице:

Средний балл	4,6	4,3	3,8	3,8	4,2	4,3	3,8	4,0	3,1	3,9
Число часов	25	22	9	15	15	30	20	30	10	17

Постройте график исходных данных и определите по нему характер зависимости. Составьте выборочное уравнение регрессии. Рассчитайте выборочный коэффициент корреляции. Дайте интерпретацию полученных результатов.

3. В таблице приведены идеальные данные о росте и весе людей среднего возраста, сохранивших «спортивный» вес:

Рост (x_i)	178	166	172	168	176
Вес (y_i)	76	64	70	66	74

Вычислив коэффициент корреляции, убедится, что он равен 1.

4. Распределение 50 гастрономических магазинов области по уровню издержек обращения X (%) и годовому объёму товарооборота Y(млн.руб.) представлено в таблице:

X/Y	0,5-2,0	2,0-3,5	3,5-5,0	5,0-6,5	6,5-8	n_x
4-6	-	-	-	3	2	5
6-8	-	4	8	8	1	21
8-10	2	5	5	2	-	14
10-12	3	1	5	-	-	9
12-14	1	-	-	-	-	1
n_y	6	10	18	13	3	50

1)Построить корреляционное поле; 2)составить уравнение регрессии Y по X и построить линию регрессии; 3) найти коэффициент корреляции.

---

1. С целью анализа взаимного влияния зарплаты и текучести рабочей силы на пяти однотипных фирмах с одинаковым числом работников, проведены измерения уровня месячной зарплаты *X* и числа уволившихся за год рабочих *Y:*

X	100	150	200	250	300
Y	60	35	20	20	15

1) Построить график исходных данных и определите по нему характер зависимости. 2) Найти уравнение регрессии Y на X и выборочный коэффициент корреляции. 3) Построить найденную линию регрессии.

2. На основе выборочных данных о деловой активности некоторых коммерческих структур, оценивается теснота связи между прибылью Y(тыс. руб.) и затратами X (коп.) на 1 рубль произведённой продукции. Соответствующие данные приведены в таблице:

X	96	77	77	89	81
Y	220	1070	1000	600	790

Для данной выборки вычислены следующие величины:

$\bar{x} = 84; \ S_x = 7{,}43; \quad \bar{y} = 736; \ S_y = 306; \quad r_{xy} = -0{,}986$.

3. Имеются выборочные данные о глубине вспашки полей под озимые культуры X(см) и их урожайности Y(ц/га):

X	10	15	20	25	30
Y	5	10	16	20	24

Вычислить эмпирический коэффициент корреляции и на его основе сделайте вывод о наличии и виде зависимости между признаками. Составить уравнение этой зависимости. Сделать прогноз урожайности пшеницы при глубине вспашки 22 см.

4. Зависимость между суточной выработкой продукции Y (тонн) и величиной основных фондов X(млн руб.) для совокупности 50 выбранных предприятий задана уравнениями регрессии: $y_x=0{,}676x-4{,}79;$ $x_y=0{,}810y+16{,}70.$ Найти коэффициент корреляции и доверительный интервал для коэффициента корреляции генеральной совокупности с надёжностью $\gamma = 0{,}95$.

5.Дана корреляционная таблица, в которой указаны: X – удои от коровы за одну лактацию, Y – содержание жира в молоке. Найдите: а)коэффициент корреляции; б) уравнение прямой линии регрессии Y на X.

Y/X	2000-2400	2400-2800	2800-3200	3200-3600	3600-4000	4000-4400	4400-4800	4800-5200	n_y
3,6-3,7	-	1	-	-	-	-	-	-	1
3,7-3,8	1	2	1	-	-	-	-	-	4
3,8-3,9	1	1	3	-	-	-	-	-	5
3,9-4,0	-	-	5	2	2	-	-	-	9
4,0-4,1	-	1	1	-	1	2	-	1	6
4,1-4,2	-	1	-	-	-	1	1	-	3
4,2-4,3	-	-	-	1	-	-	-	-	1
4,3-4,4	-	-	-	-	1	-	-	-	1
n_x	2	6	10	3	4	3	1	1	30

Вопросы для самопроверки.

1.Поятие статистической и корреляционных зависимостей.

3.В чём состоит различие между функциональной и корреляционной зависимостью?

4.Основные задачи регрессионного и корреляционного анализа.

5.Что такое корреляционное поле, как оно строится?

6.Опишите, как составляется корреляционная таблица.

7.Что называется эмпирической линией регрессии?

8.В каком виде ищут уравнение линии регрессии?

9. В чём состоит сущность метода наименьших квадратов для определения параметров линии регрессии?

10. Нахождение коэффициентов регрессии, что они показывают?

11.Определение коэффициент корреляции, его свойства.

Типовой расчет «Статистический анализ вариационных рядов»

По 25 овцам имеются данные о настриге шерсти и длины волоса шерсти (табл. 1 и 2).

Т а б л и ц а 1

Настриг шерсти, кг (у)

Номер овцы	Предпоследняя цифра номера зачетной книжки									
	0	1	2	3	4	5	6	7	8	9
1	3,9	4,2	4,0	4,1	4,2	4,2	4,1	4,1	4,2	4,2
2	5,0	4,9	5,2	5,1	5,1	5,1	4,9	5,0	5,1	4,8
3	4,6	4,8	4,5	4,7	4,8	4,6	4,8	4,8	4,7	4,9
4	4,9	4,8	4,8	4,8	4,7	4,6	4,6	4,6	4,8	4,7
5	5,2	5,4	5,1	5,3	5,4	5,1	5,4	5,3	5,2	5,3
6	4,8	4,6	4,8	4,8	4,7	4,5	4,8	4,7	4,5	4,7
7	4,5	4,6	4,7	4,8	4,5	4,7	4,5	4,6	4,8	4,6
8	4,2	4,5	4,1	4,1	4,4	4,4	4,2	4,2	4,3	4,5
9	5,1	5,2	5,2	5,2	5,2	5,4	5,3	5,4	5,4	5,2
10	5,4	5,3	5,1	5,3	5,5	5,1	5,4	5,3	5,2	5,4

11	4,9	4,7	4,7	4,6	4,8	4,6	4,7	4,8	4,9	4,6
12	4,6	4,8	4,8	4,8	4,5	4,6	4,8	4,9	4,5	4,8
13	4,6	4,7	4,9	4,8	4,7	4,8	4,7	4,9	4,9	4,7
14	4,5	4,5	4,3	4,1	4,2	4,1	4,5	4,1	4,2	4,5
15	4,8	4,8	4,8	4,9	4,6	4,6	4,6	4,7	4,9	4,8
16	4,9	5,1	5,0	5,0	4,8	4,9	4,9	5,0	4,9	4,8
17	5,0	4,9	4,9	5,0	4,9	5,0	4,8	5,0	4,8	5,0
18	4,4	4,4	4,2	4,6	4,6	4,4	4,5	4,4	4,4	4,2
19	4,8	5,2	4,9	5,1	5,1	5,2	5,1	5,0	5,1	5,2
20	3,9	4,2	4,0	4,1	3,9	3,9	4,0	4,1	4,0	3,9
21	5,1	5,1	4,9	5,0	5,1	4,9	5,1	5,0	5,0	5,2
22	4,4	4,5	4,4	4,2	4,5	4,3	4,1	4,5	4,2	4,4
23	4,5	4,6	4,6	4,8	4,9	4,6	4,7	4,8	4,8	4,7
24	4,5	4,3	4,3	4,5	4,3	4,5	4,6	4,4	4,1	4,4
25	4,1	4,2	4,6	4,3	4,2	4,1	4,5	4,3	4,3	4,5

Т а б л и ц а 2

Длина волоса шерсти, см (х)

Номер овцы	Последняя цифра номера зачетной книжки									
	0	1	2	3	4	5	6	7	8	9
1	11,3	11,9	12,1	12,1	11,9	13,4	12,9	13,4	13,7	12,4
2	20,8	21,0	21,2	20,6	20,4	22,1	22,4	19,8	21,3	22,3
3	17,9	18,9	17,6	18,2	17,1	18,8	16,9	17,0	18,8	19,5
4	18,7	17,7	18,5	18,4	18,3	18,9	17,0	17,4	17,5	16,8
5	22,5	22,9	23,0	24,7	22,6	24,5	23,5	24,8	23,4	24,2
6	19,0	17,2	17,9	19,2	19,1	17,6	17,3	18,8	18,1	17,4
7	17,1	19,2	18,0	16,8	16,8	18,8	19,1	17,6	18,3	17,3
8	14,0	15,6	16,4	14,0	15,5	16,4	15,8	15,8	15,0	14,5
9	21,8	20,2	20,4	19,6	20,9	20,0	22,0	22,1	20,8	21,3
10	22,8	24,7	22,9	22,7	24,4	24,6	22,8	23,9	24,4	22,7
11	18,7	19,6	19,2	16,9	18,7	17,7	16,9	18,9	19,0	16,8
12	19,8	22,1	20,1	19,8	22,0	19,5	19,5	22,3	19,8	20,6
13	17,5	18,8	18,6	17,3	18,0	18,3	16,7	17,4	18,2	17,4
14	14,5	16,5	16,1	14,1	15,0	15,7	15,6	14,1	14,9	14,9
15	18,5	18,1	17,7	17,0	18,7	18,6	18,1	19,2	17,3	16,9
16	21,7	21,4	21,5	21,6	20,3	22,3	22,0	22,2	22,2	22,5
17	19,7	22,3	21,2	22,5	20,6	19,6	20,7	21,4	22,2	20,8

18	16,7	19,2	16,7	19,6	19,0	17,6	16,7	17,5	17,7	18,0
19	22,2	20,9	22,2	21,4	21,3	19,8	21,3	19,9	21,9	20,0
20	11,8	13,8	11,5	12,8	13,2	11,5	13,4	12,7	12,1	13,0
21	20,4	20,5	20,2	20,8	22,3	20,0	20,5	20,9	21,5	21,3
22	14,1	16,5	15,4	15,9	15,0	15,1	15,4	14,2	15,5	15,4
23	19,5	19,9	20,9	20,4	22,3	21,1	22,4	20,6	20,7	21,2
24	15,3	15,7	16,1	16,0	16,0	15,4	16,8	14,2	16,7	14,4
25	13,3	12,4	11,4	12,6	13,9	11,5	13,7	13,0	11,9	13,8

Выполните следующие задания:

1. Постройте интервальные ряды распределения настрига и длины волоса шерсти, отобразите их графически в виде гистограмм, полигонов и кумулят.

2. Для анализа рядов распределения рассчитайте средние величины (среднюю арифметическую, моду, медиану), выборочные показатели вариации (дисперсию, среднее квадратическое отклонение, коэффициент вариации) и показатели распределения (коэффициенты асимметрии и эксцесса).

3. С помощью выборочного метода рассчитайте для каждого ряда распределения предельные ошибки выборочной средней, найдите доверительные пределы генеральной средней при уровне вероятности суждения 0,95

4. Проверьте гипотезу соответствия рядов распределения настрига и длины волоса шерсти нормальному закону распределения.

5. Используя данные интервального ряда распределения настрига шерсти с помощью дисперсионного анализа рассчитайте достоверность разницы в настриге шерсти в зависимости от длины волоса шерсти.

6. С помощью корреляционного анализа определите влияние длины волоса шерсти на настриг шерсти. Для этого постройте линейное уравнение регрессии, рассчитайте коэффициент корреляции и оцените его достоверность с помощью *t*-критерия Стьюдента и *F*-критерия Фишера.

По каждому заданию сделайте соответствующие выводы.

План работы

1. Вариационные ряды распределения. Графическое представление данных.
2. Статистические оценки параметров распределения.
3. Интервальные оценки. Доверительные интервалы. Ошибки выборочной средней.
4. Статистические гипотезы
5. Дисперсионный анализ
6. Корреляционный анализ

Образец оформления

Таблица 3.

Первичные данные по настригу (кг) и длине волоса шерсти (см) овец

№	**Длина волоса шерсти, см (X)**	**Настриг шерсти, кг (Y)**
1	11,9	4,2
2	20,4	5,1
3	17,1	4,8
4	18,3	4,7
5	22,6	5,4
6	19,1	4,7
7	16,8	4,5
8	15,5	4,4
9	20,9	5,2
10	24,4	5,5
11	18,7	4,8
12	22,0	4,5
13	18,0	4,7
14	15,0	4,2
15	18,7	4,6
16	20,3	4,8
17	20,6	4,9
18	19,0	4,6
19	21,3	5,1

20	13,2	3,9
21	22,3	5,1
22	15,0	4,5
23	22,3	4,9
24	16,0	4,3
25	13,9	4,2

I. *Интервальные ряды распределения. Графическое представлен данных*

1. Интервальный ряд распределения длины волоса шерсти.

   Минимальное значение – 11,9 см;

   Максимальное значение – 24,4 см;

   Размах вариации – 10,7 см; **R = Xmax-Xmin**;

   Число групп (классов) – 6; **k=1+3,322 lg*n***,

   Длина интервала – 2,1см; **h=R/k**

Таблица 2.

Интервальный ряд распределения длины волоса шерсти

Номер интервала	**Группа овец**		**Число овец (абсол. част. интервала) f_i**	**Середина интервала X'_i**	**Накопленная частота f_i'**
	нижняя граница	**верхняя граница**			
1	11,9	14	3	12,95	3
2	14	16,1	4	15,05	7
3	16,1	18,2	3	17,15	10
4	18,2	20,3	6	19,25	16
5	20,3	22,4	7	21,35	23
6	22,4	24,5	2	23,45	25
итого			25		

2. Интервальный ряд распределения настрига шерсти.

   Минимальное значение – 3,9 кг;

   Максимальное значение – 5,5 кг;

   Размах вариации – 1,6 кг; **R = Xmax-Xmin;**

Число групп – 6; **k=1+3,322 lg*n*,**

Длина интервала – 0,3; **h=R/k**

Таблица 3.

Интервальный ряд распределения настрига шерсти

Номер интервала	**Группа овец**		**Число овец (абсол. част. интервала) fi**	**Середина интервала Y'_i**	**Накопленная частота f_i'**
	нижняя граница	**верхняя граница**			
1	3,9	4,2	4	4,05	4
2	4,2	4,5	5	4,35	9
3	4,5	4,8	8	4,65	17
4	4,8	5,1	5	4,95	22
5	5,1	5,4	1	5,25	23
6	5,4	5,7	2	5,55	25
итого			25		

Графическое представление данных

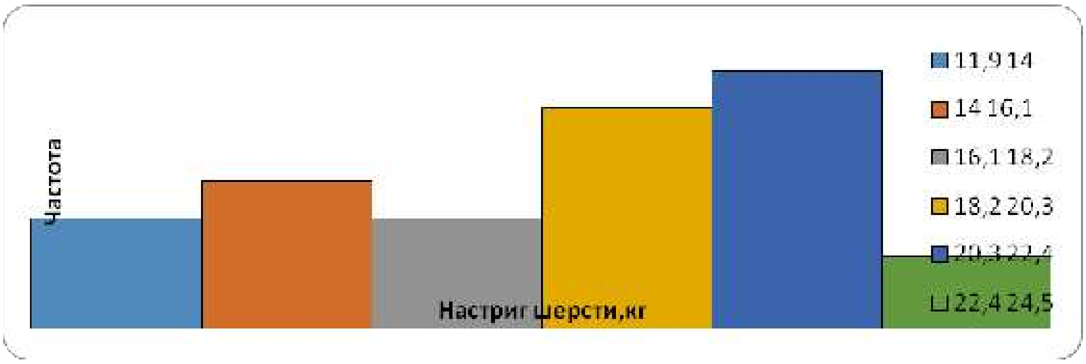


Рисунок 1. Гистограмма распределения настрига шерсти (кг)

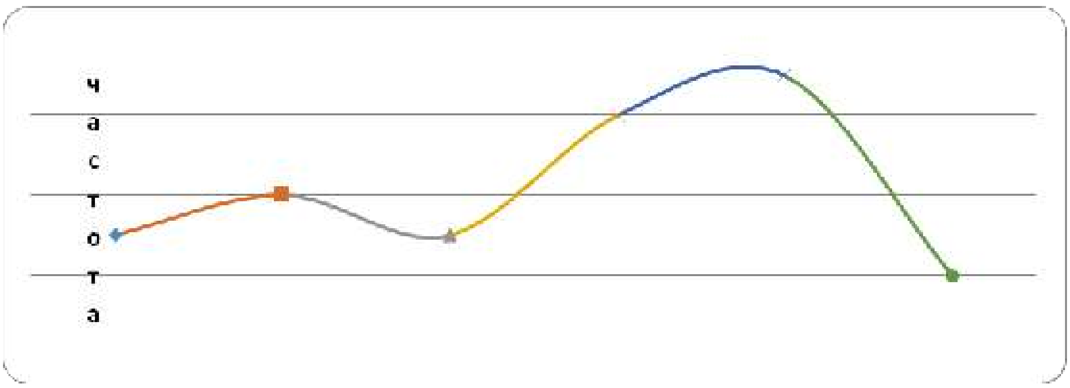


Рисунок 2. Полигон распределения настрига шерсти (кг)

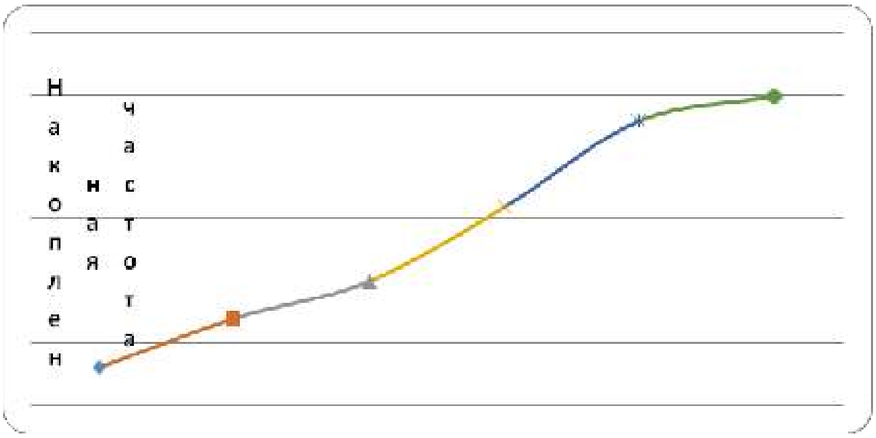


Рисунок 3 Кумулята распределения настрига шерсти (кг)

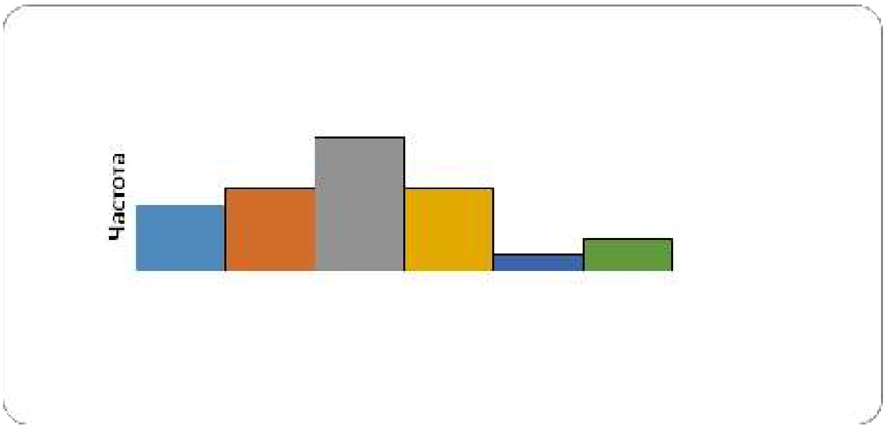


Рисунок 4 Гистограмма распределения длины волоса шерсти (см)

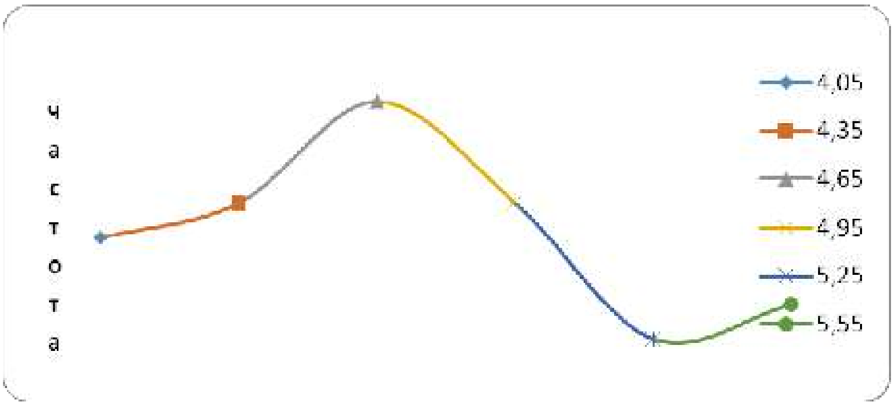
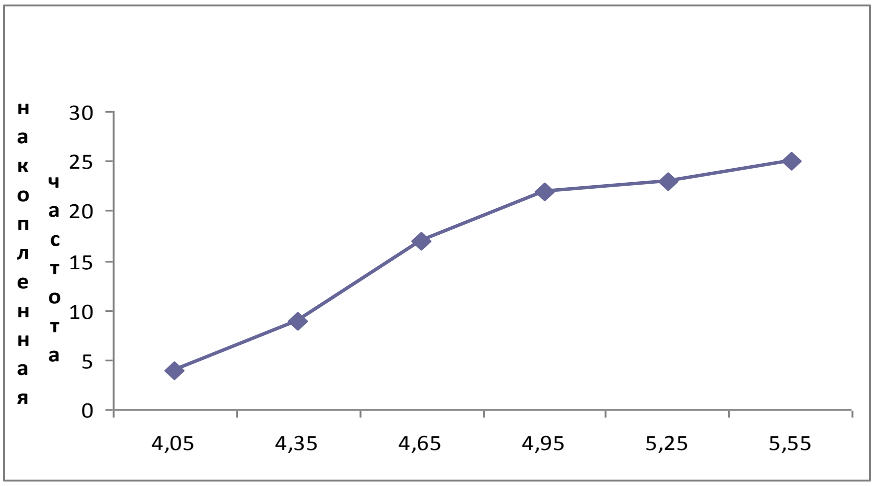


Рисунок 5 Полигон распределения длины волоса шерсти



(см)

Рисунок 6. Кумулята распределения длины волоса шерсти (см)

II. Статистические оценки параметров распределения. Точечные оценки

Таблица 4.

Выборочные средние показатели, показатели вариации и показатели распределения

Показатели	Настриг шерсти	Длина волоса
Среднее арифметическое	4,70	18,53
Дисперсия	0,15	10,24
Среднее квадратичное отклонение	0,39	3,26
Коэффициент вариации	8,40	17,62
Асимметрия	0,10	-0,27
Эксцесс	-0,29	-0,69
Мода	4,65	20,65
Медиана	4,63	19,07

III. Интервальные оценки. Доверительные интервалы. Ошибки выборочной средней.

1. Абсолютная (стандартная) ошибка выборочной средней

$S_{\bar{x}} = \frac{S}{\sqrt{n}}$ $S_{\bar{x}} = \frac{3{,}26}{\sqrt{25}} = 0{,}652$ $S_{\bar{Y}} = \frac{0{,}39}{\sqrt{25}} = 0{,}079$

2. Предельная ошибка выборочной средней $\varepsilon = t \cdot S_{\bar{x}}$,

где t=2,06≈ 2,1 при γ =95%

$\varepsilon_{\bar{x}} = 0{,}652 \cdot 2{,}1 = 1{,}37$ $\varepsilon_{\bar{y}} = 0{,}079 \cdot 2{,}1 = 0{,}166$

3. Доверительный интервал для генеральной средней.

$$\left(\bar{x} - \frac{t_\gamma \cdot S}{\sqrt{n}}; \bar{x} + \frac{t_\gamma \cdot S}{\sqrt{n}}\right)$$

Доверительный интервал для генеральной средней длины волоса шерсти, см: (18,532-1,37 ; 18,532+1,37)

(17,2 ; 19,9) - *это доверительный интервал, который с вероятностью 95% содержит неизвестную среднюю длину волоса шерсти генеральной совокупности овец данной породы.*

Доверительный интервал для генеральной средней настрига шерсти, кг: (4,704-0,166 ; 4,704+0,166)

(4,5 ; 4,9) - *это доверительный интервал, который с вероятностью 95% содержит неизвестную среднюю настрига шерсти генеральной совокупности овец данной породы.*

IV. Статистические гипотезы. Проверка гипотезы о соответствии рядов распределения настрига и длины волоса шерсти нормальному закону распределения

Нулевая гипотеза H_0: *Распределение признака «длина волоса шерсти / настриг шерсти» соответствует нормальному закону (α=0,05).*

Для проверки гипотезы H_0 выберем критерий согласия Пирсона - χ^2

$$\chi^2 = \Sigma \frac{(f - f_0)}{f_0}$$

Табличное значение для χ^2 =11,07 ($\nu = (k-1)(l-1) = (2-1)\cdot(6-1) = 5$;$\alpha$=0,05)

$$f_0 = nh\varphi(t)$$

$$\varphi(t) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$$

Для признака X:

Таблица 5.

Средняя арифметическая	18,53
Среднее квадратичное отклонение	3,26
Уровень значимости	0,05
Степени свободы вариации	5
Фактический уровень значимости	0,48
Фактическое значение хи-квадрат	4,49
Табличное значение хи-квадрат	11,07

$\chi^2_{\text{фактическое}}=4{,}49$

Вывод: Поскольку фактическое значение критерия меньше табличного ($\chi^2_{\text{фактическое}}<\chi^2_{\text{теоретическое}}$), то нулевая гипотеза о соответствии эмпирического распределения теоретическому принимается, т.е. распределение признака «длина волоса шерсти» соответствует нормальному закону (α=0,05).

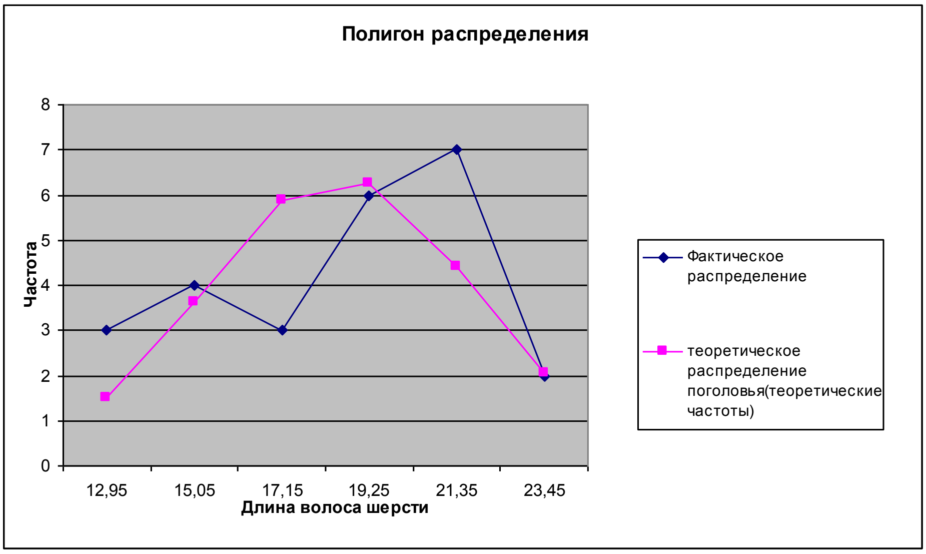


Рисунок 7. Полигон распределения длины волоса шерсти.

Для признака У:

Таблица 6.

Средняя арифметическая	4,70
Среднее квадратичное отклонение	0,39
Уровень значимости	0,05
Степени свободы вариации	5
Фактический уровень значимости	0,33
Фактическое значение хи-квадрат	5,78
Табличное значение хи-квадрат	11,07

$\chi^2_{\text{фактическое}}=5{,}78$

Вывод: Поскольку фактическое значение критерия меньше табличного ($\chi^2_{\text{фактическое}}<\chi^2_{\text{теоретическое}}$), то нулевая гипотеза о соответствии эмпирического распределения теоретическому принимается, т.е. распределение признака «настриг шерсти» соответствует нормальному закону (α=0,05).

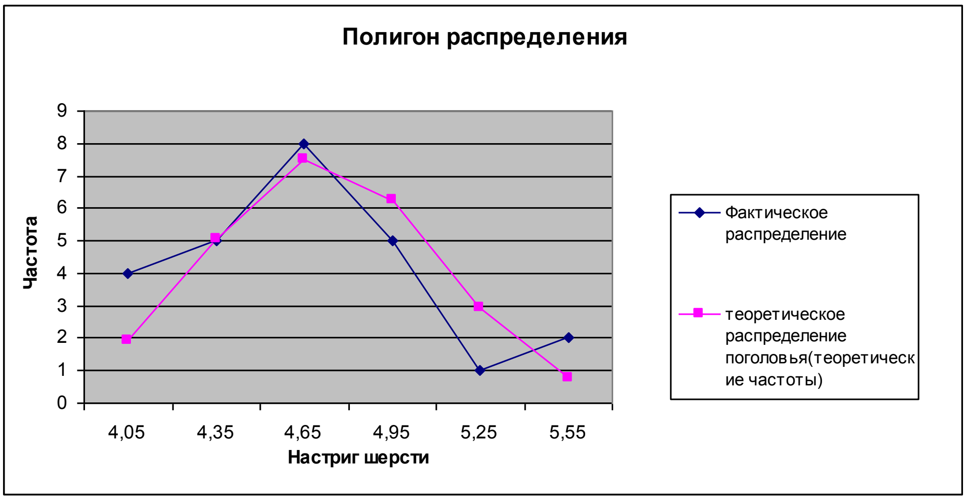


Рисунок 8. Полигон распределения настрига шерсти.

IV. Дисперсионный анализ.

Основной гипотезой, нуждающейся в проверке, является гипотеза о равенстве групповых средних $H_0: a_1 = a_2 = \ldots = a_k$. Иными словами, проверяют гипотезу о том, что фактор (длина волоса шерсти) вообще не влияет на наблюдения (настриг шерсти)

Таблица 7.

Длина волоса шерсти		Настриг шерсти овец							
Интервал		1	2	3	4	5	6	7	8
11,9	14	4,2	3,9	4,2					
14	16,1	4,2	4,5	4,4	4,3				
16,1	18,2	4,5	4,8	4,7					
18,2	20,3	4,7	4,8	4,6	4,6	4,7			
20,3	22,4	4,8	5,1	4,9	5,2	5,1	4,5	5,1	4,9
22,4	24,5	5,4	5,5						

Таблица 8.

Однофакторный дисперсионный анализ

Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	3,199933	5	0,639987	22,1220376	2,56218E-07	2,740057541
Внутри групп	0,549667	19	0,02893			
Итого	3,7496	24				

$S^2_{межгр.}=3,2$

$S^2_{внутригр.}=0,55$

$S^2_{общ.}=3,75$

$F_{набл.}=22,1$ $F_{кр.}=2,74$

Данные таблицы показывают, что фактическое отношение дисперсий больше табличного, следовательно, разница в среднем настриге шерсти по группам овец с различной длиной волоса шерсти достоверна при уровне значимости 0,05. Длина волоса шерсти овец оказывает влияние на их настриг шерсти. Другими словами, предположение о том, что длина волоса не влияет на вариацию настрига не имеет места.

VI. Корреляционный анализ. Регрессия. Уравнение линии регрессии.

R (или r) – коэффициент корреляции. Устанавливает, есть ли связь между признаками, и насколько она тесная $-1 \le R \le 1$

Если же модуль коэффициента корреляции ~1,то связь близка к линейной.

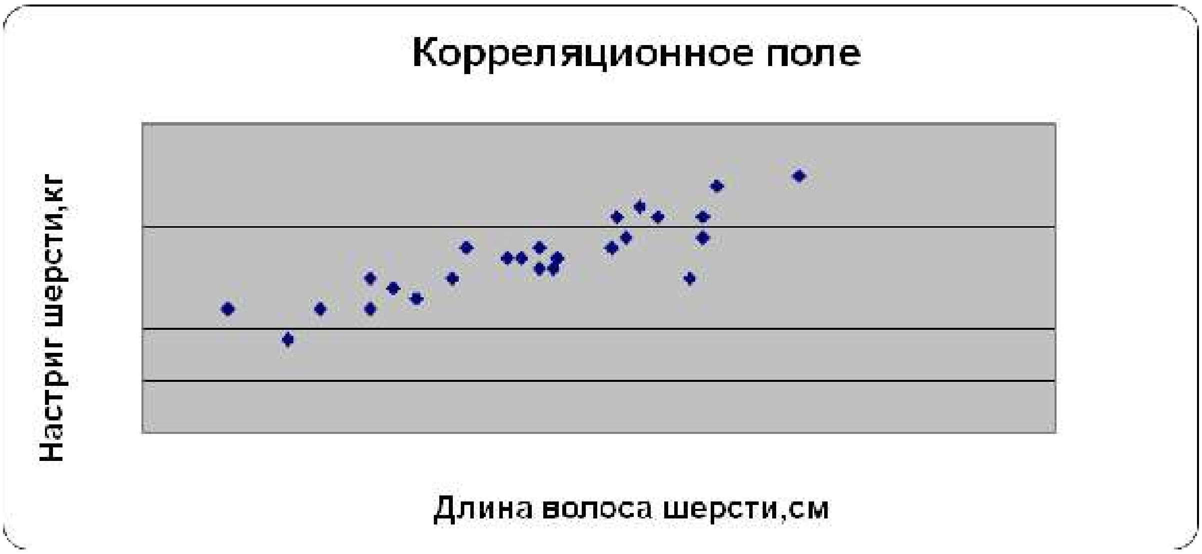


Рисунок 9. Корреляционное поле.

Уравнение прямолинейной регрессии:

$$Y = b(y/x) \cdot X + B$$

Таблица 9.

Регрессионная статистика	
Множественный R	0,877540876
R-квадрат	0,770077989
Нормированный R-квадрат	0,76008138
Стандартная ошибка	0,193606006
Наблюдения	25

Таблица 10.

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	2,887484429	2,887484429	77,03391994	8,4375E-09
Остаток	23	0,862115571	0,037483286		
Итого	24	3,7496			

Таблица 11. Параметры линии регрессии

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	2,73573373	0,227573729	12,02130727	2,13329E-11
Длина волоса шерсти, см (x)	0,106209058	0,01210098	8,776896942	8,4375E-09

R=0,876

Вывод: связь между признаками тесная и близка к линейной.

R^2 (коэффициент детерминации) =0,77=77%

Вывод: вариация настрига шерсти обусловлена на 77% влиянием длины волоса шерсти. Остальные 23% вариации настрига обусловлены неучтенными факторами.

Для того, чтобы составить уравнение регрессии необходимо найти параметры B (У – пересечение) и [Ъ(У / Х) .

Y-пересечение	2,7
Длина волоса шерсти, см (x)	0,106

Тогда уравнение регрессии будет иметь вид: $Y = 0,106 \cdot X + 2,7$

Вывод: При увеличении длины волоса шерсти на 1 см., настриг шерсти в среднем увеличивается на 106 г.

Литература.

Основная литература.

1 .Практикум по статистике: учеб. пособие / под ред. А.П. Зинченко. – М.: КолосС, 2005. – 393 с.МСХ РФ

2.Яковлев, В.Б. Статистика. Расчеты в Microsoft Excel: учеб. пособие / В.Б. Яковлев. – М.: КолосС, 2005. – 352 с. МСХ РФ

3. Кузнецова, Т.А. Высшая математика [Электронный ресурс]: учеб. пособие / Т.А. Кузнецова, Е.С. Мироненко, С.А. Розанова. – Электрон. текст. дан. - М.: Физматлит, 2009. – 168 с. - Режим доступа: www.e.lanbook.com.

4. Михеев, В.И. Высшая математика. Краткий курс [Электронный ресурс]: учеб. пособие / В.И. Михеев. – Электрон. текст. дан. - М.: Физматлит, 2007. – 200 с.- Режим доступа: www.e.lanbook.com.

Дополнительная литература.

1.Высшая математика: учебник для студ. вузов/ И.И. Баврин, В.Л. Матросов. - М.: ВЛАДОС, 2007. - 520 с. Дополнительная литература Информатика. Базовый курс: учеб. пособие для студ. вузов, (бакалавров и специалистов)/ Ред. С.В. Симонович. - 3-е изд.. - СПб.: Питер, 2011. - 637 с.: ил. - (Учебник для вузов. Стандарт третьего поколения)

2.Гмурман, В.Е. Теория вероятностей и математическая статистика: учеб. пособие / В.Е. Гмурман. – 12-е изд., перераб. - М.: Высшее образование, Юрайт, 2009. – 479 с. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для студентов вузов/ В.Е. 3.Гмурман; В.Е. Гурман. - 11-е изд.. - М.: Юрайт, 2011. - 403 с.

Сдвижков, О.А. Математика в Excel: учеб. пособие / О.А. Сдвижков. – М.: СОЛОН-Пресс, 2007. – 192 с.

4.Балдин, К..В. Краткий курс высшей математики: учебник / К.В.Балдин.- М.: Дашков и Ко, 2009.- 512 с. Практикум по высшей математике: учебное пособие, Ч.1 / Л.И.Дюженкова, О.Ю.Дюженкова, Г.А. Михалин.- 2-е изд.- М.: БИНОМ, 2012.- 449.

Приложения

Таблица 1. Значения функции $p(m) = \frac{a^m}{m!} \cdot e^{-a}$

a/m	0	1	2	3	4	5	6	7
0,1	0,90484	09048	00452	00015	00000	00000	00000	00000
0,2	81873	16375	01638	00109	00006	00000	00000	00000
0,3	74082	22225	03334	00333	00025	00002	00000	00000
0,4	67032	26813	05363	00715	00072	00006	00000	00000
0,5	69653	30327	07582	01264	00158	00016	00001	00000
0,6	54881	32929	09879	01976	00296	00036	00004	00000
0,7	49659	34761	12166	02839	00497	00070	00008	00001
0,8	44933	35946	14379	03834	00767	00123	00016	00002
0,9	40657	36591	16466	04940	01112	00200	00030	00004
1	36788	36788	18394	06131	01533	00307	00051	00007
2	13534	27067	27067	18045	09022	03609	01203	00344
3	04979	14936	22404	22404	16803	10082	05041	02160
4	01832	07326	14653	19537	19537	15629	10420	05954
5	00674	03369	08422	14037	17547	17547	14622	10445
6	00248	01487	04462	08924	13385	16062	16062	13768
7	00091	00638	02234	05123	09123	12772	14900	14900

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

x	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	3824	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	2956	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	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	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	0097	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0036	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

Таблица 3 Значения функции $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\cdot\int\limits_0^x e^{-\frac{t^2}{2}}dt$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0040	0080	0120	0160	0199	0239	0279	0319	0359
0,1	0398	0438	0478	0517	0557	0596	0636	0675	0714	0753
0,2	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
0,3	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
0,4	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
0,5	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
0,6	2257	2291	2324	2357	2389	2422	2454	2486	2517	2549
0,7	2580	2611	2642	2673	2708	2734	2764	2794	2823	2852
0,8	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
0,9	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
1,0	3413	3438	3461	3485	3508	3531	3554	3577	3599	3621
1,1	3643	3665	3696	3708	3729	3749	3770	3790	3810	3830
1,2	3849	3869	3883	3907	3925	3944	3962	3980	3997	4015
1,3	4032	4049	4066	4082	4099	4115	4131	4147	4162	4177
1,4	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
1,5	4332	4345	4357	4370	4382	4394	4406	4418	4429	4441
1,6	4452	4463	4474	4484	4495	4505	4515	4525	4535	4545
1,7	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
1,8	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
1,9	4713	4719	4726	4732	4738	4744	4750	4756	4761	4767
2,0	4772	4778	4783	4788	4793	4798	4803	4808	4812	4817
2,1	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
2,2	4861	4864	4868	4871	4875	4878	4881	4884	4887	4890
2,3	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
2,4	4918	4920	4922	4925	4927	4929	4931	4932	4034	4936
2,5	4938	4940	4941	4943	4945	4946	4948	4949	4951	4951
2,6	4953	4955	4956	4067	4959	4960	4961	4962	4963	4964
2,7	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
2,8	4974	4975	4976	4977	4977	4978	4979	4979	4980	4981
2,9	4981	4982	4982	4983	4984	4984	4985	4985	4986	4986

x		x		x		x	
3,0	0,49865	3,5	0,49977	4,0	0,499968	4,5	0,4999966
3,1	0,49903	3,6	0,49984	4,1	0,499979	4,6	0,4999979
3,2	0,49931	3,7	0,49989	4,2	0,499987	4,7	0,4999987
3,3	0,49952	3,8	0,49993	4,3	0,499991	4,8	0,4999992
3,4	0,49966	3,9	0,49995	4,4	0,499995	4,9	0,4999995

Таблица 4

Значения *критерия Стьюдента* при различных уровнях значимости *(P)*

Число степеней Свободы, *V*	У зовни значимости *(P)*		
	0,05	0,01	0,001
1	12,7	63,66	-
2	4,30	9,93	31,60
3	3,18	5,84	12,94
4	2,78	4,60	8,61
5	2,57	4,03	6,86
6	2,45	3,71	5,96
7	2,37	3,50	5,41
8	2,31	3,36	5,04
9	2,26	3,25	4,78
10	2,23	3,17	4,59
11	2,20	3,11	4,44
12	2,18	3,06	4,32
13	2,16	3,01	4,22
14	2,15	2,98	4,14
15	2,13	2,95	4,07
16	2,12	2,92	4,02
17	2,11	2,90	3,97
18	2,10	2,88	3,92
19	2,09	2,86	3,88
20	2,09	2,85	3,85
21	2,08	2,83	3,82
22	2,07	2,82	3,79
23	2,07	2,81	3,77
24	2,06	2,80	3,75
25	2,06	2,79	3,73
26	2,06	2,78	3,71
27	2,05	2,77	3,69
28	2,05	2,76	3,67
29	2,05	2,76	3,66
30	2,04	2,75	3,65
ТО	1,96	2,58	3,29

Таблица 5

Теоретические значения распределения χ^2 при разных степенях свободы (V) и критических уровнях значимости

Число степеней свободы, V	Уровень значимости			Число степеней свободы, V	Уровень значимости		
	P=0,05	P=0,01	P=0,001		P=0,05	P=0,01	P=0,001
1	3,84	6,63	10,8	**18**	28,87	34,81	42,3
2	5,99	9,21	13,8	**19**	30,14	36,19	43,8
3	7,81	11,34	16,3	**20**	31,41	37,57	45,3
4	9,49	13,28	18,5	**21**	32,67	38,93	
5	11,07	12,83	20,5	**22**	33,92	40,29	
6	12,59	16,81	22,5	**23**	35,17	41,64	
7	14,07	18,48	24,3	**24**	36,42	42,98	
8	15,51	20,09	26,1	**25**	37,65	44,31	
9	16,92	21,67	27,9	**26**	38,89	45,64	
10	18,31	23,21	29,6	**27**	40,11	46,96	
11	19,68	24,72	31,3	**28**	41,34	48,28	
12	21,03	26,22	32,9	**29**	42,56	49,59	
13	22,36	27,69	34,5	**30**	43,77	50,89	
14	23,68	29,14	36,1				
15	25,00	30,58	37,7	**50**	67,50	76,15	
16	26,30	32,00	39,3	**80**	101,88	112,3	
17	27,59	33,41	40,8	**100**	124,34	135,8	

Таблица 6. Таблица F – критерия Фишера.

Таблиц F- критерия (критерия Фишера)

f_2	f_1										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**15**
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	245.95
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.43
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.53
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.46
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.35
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.31
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.27
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.23
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20

Савельева Екатерина Владимировна

Основы математической биостатистики: учебное пособие для обучающихся специальности 36.05.01 - «Ветеринария» и направлениям подготовки: – 36.03.02 – «Зоотехния»; 36.03.01 «Ветеринарно – санитарная экспертиза»

Подписано в печать ____________2016г.

Формат 60×84 1/16. Бумага офсетная.

Печать RISOGRAPH TR 1510. Уч.-изд.л. 12,6

Тираж 30 экз. Заказ №

ФГБОУ ВО

«Приморская государственная сельскохозяйственная академия».

692510. г. Уссурийск, пр. Блюхера,44.

Участок оперативной полиграфии ФГБОУ ВО ПГСХА

692500. г. Уссурийск, ул. Раздольная, 8.