МЕЖГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Маркетинг и менеджмент»

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические рекомендации к практическим занятиям для студентов специальности 1-28 01 02 «Электронный маркетинг» очной и заочной форм обучения

Могилев 2021

УДК 519. 2
ББК 22. 17
Т33

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Маркетинг и менеджмент» «5» мая 2021 г., протокол № 11

Составитель канд. пед. наук, доц. Т. С. Старовойтова

Рецензент канд. физ.-мат. наук, доц. В. А. Ливинская

В методических рекомендациях представлены материалы для практических занятий по дисциплине «Теория вероятностей и математическая статистика». Основные вопросы теории проиллюстрированы примерами решения задач. По каждому вопросу выделены задачи для самостоятельного решения.

Учебно-методическое издание

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Ответственный за выпуск	А. В. Александров
Корректор	Т. А. Рыжикова
Компьютерная верстка	Н. П. Полевничая

Подписано в печать . Формат 60×84/16. Бумага офсетная. Гарнитура Таймс. Печать трафаретная. Усл. печ. л. . Уч.-изд. л. . Тираж 56 экз. Заказ №

Издатель и полиграфическое исполнение:
Межгосударственное образовательное учреждение высшего образования
«Белорусско-Российский университет».
Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/156 от 07.03.2019.
Пр-т Мира, 43, 212022, г. Могилев.

Содержание

1 Вопросы по программе курса

1 Случайные события. Определения случайного события и классической вероятности. Действия над событиями, свойства классической вероятности. Аксиоматическое построение теории вероятностей. Определения условной вероятности и независимых событий. Теорема умножения.

2 Случайные события. Формула полной вероятности. Формула Байеса. Схема независимых испытаний Бернулли. Формула Бернулли. Теорема Пуассона. Локальная и интегральная предельные теоремы Муавра – Лапласа. Теорема Бернулли.

3 Случайные величины и их функции распределения. Дискретные и непрерывные случайные величины. Функция распределения вероятностей и ее свойства. Плотность распределения вероятностей и ее свойства.

4 Числовые характеристики случайных величин. Математическое ожидание, дисперсия, среднее квадратическое отклонение, моменты, ковариация, коэффициент корреляции, их свойства. Мода, медиана, квантиль.

5 Основные распределения случайных величин: биномиальное, Пуассона, геометрическое, нормальное, показательное, равномерное.

6 Многомерные случайные величины. Функция распределения и плотность распределения двумерной случайной величины. Дискретные и непрерывные двумерные случайные величины.

7 Условный закон распределения. Независимость случайных величин. Проверка условий независимости случайных величин.

8 Закон больших чисел. Центральная предельная теорема. Неравенство Чебышева. Теорема Чебышева.

9 Выборка. Графическое изображение выборки. Количественная, порядковая, номинальная шкалы измерения. Вариационный ряд. Полигон и гистограмма. Эмпирическая функция распределения, ее свойства, график.

10 Точечное оценивание. Требования к точечным оценкам. Несмещенные оценки математического ожидания и дисперсии. Построение оценок параметров с помощью метода моментов и метода наибольшего правдоподобия.

11 Интервальное оценивание. Доверительные интервалы для оценки параметров нормального распределения. Минимальный объем выборки для получения оценок заданной надежности и точности.

12 Проверка статистических гипотез. Уровень значимости и мощность критерия. Ошибки первого и второго рода. Критерии χ^2 – квадрат и Колмогорова – Смирнова проверки гипотез о виде распределения.

13 Критерии однородности. Критерий Стьюдента сравнения двух средних значений, критерий Фишера сравнения двух дисперсий. Однофакторный и двухфакторный дисперсионный анализ.

14 Элементы регрессионного и корреляционного анализа. Парная линейная и нелинейная регрессия. Парный коэффициент корреляции, его свойства. Проверка гипотезы о его достоверности.

2 Случайные события

2.1 Элементы комбинаторики

Многие классические задачи теории вероятностей решаются с использованием комбинаторики. Это раздел математики, в котором изучаются различные соединения (комбинации) элементов конечных множеств. Они составляются из некоторого числа *различных* элементов, принадлежащих одному и тому же множеству. Результат выбора k элементов из множества, содержащего n элементов, называется *выборкой* k элементов из n .

Правило суммы. Если элемент A можно выбрать m способами, а элемент B можно выбрать n способами, то выбор элемента A или B можно осуществить $m+n$ способами.

Правило произведения. Если элемент A можно выбрать m способами, а элемент B можно выбрать n способами, то выбор пары элементов A и B можно осуществить $m \cdot n$ способами.

Перестановками называются комбинации, состоящие из одних и тех же n различных элементов и отличающиеся только порядком их следования. Число всех возможных перестановок из n элементов вычисляется по формуле $P_n = n!$

Размещениями называются комбинации, составленные из n различных элементов по k элементов, отличающиеся либо порядком следования элементов, либо их составом.

Число всех возможных размещений из n элементов по k вычисляется по формуле $A_n^k = \dfrac{n!}{(n-k)!} = n(n-1)(n-2)\ldots(n-(k-1))$.

Символ $n!$ читается «эн факториал», при этом полагается, что $0!=1$ и $1!=1$.

Сочетаниями называются комбинации, составленные из n различных элементов по k элементов, отличающиеся хотя бы одним элементом.

Число всех возможных сочетаний из n элементов по k вычисляется по формуле $C_n^k = \dfrac{n!}{k!(n-k)!}$. Используют также формулу $C_n^k = C_n^{n-k}$.

Если при выборе k элементов из n элементов некоторые элементы возвращаются обратно (повторяются) и упорядочиваются, то такие соединения называются ***размещениями с повторениями***: $\overline{A}_n^k = n^k$.

Если при выборе k элементов из n элементов некоторые элементы повторяются без последующего упорядочивания, то такие соединения называются ***сочетаниями с повторениями***: $\overline{C}_n^k = C_{n+k-1}^k = \dfrac{(n+k-1)!}{k!(n-1)!}$

Если множество из n элементов содержит k различных элементов, при этом первый элемент повторяется n_1 раз, второй – n_2 раз, …, k-й – n_k раз, причем $n_1 + n_2 + \ldots + n_k = n,$ то перестановки элементов этого множества называются ***перестановками с повторениями.*** Их подсчет проводится по формуле $P_n\left(n_1, n_2, \ldots, n_k\right) = \dfrac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$

Случайным событием, связанным с некоторым опытом, называется всякое событие, которое при осуществлении этого опыта либо происходит, либо не происходит. Событие, которое обязательно происходит в данном опыте, называется ***достоверным*** *с*обытием (обозначается Ω).

Событие, которое никогда не происходит в данном опыте, называется **невозможным** (обозначается символом $\varnothing$).

Элементарным исходом (или ***элементарным событием***) называется любой простейший (т. е. неделимый в рамках опыта) исход опыта.

Несколько событий образуют ***полную группу***, если в результате опыта обязательно появится только одно из них.

События A и B называются ***равносильными*** или ***равными***, если событие A происходит тогда, и только тогда, когда происходит событие B (обозначается $A = B$).

Объединением **(*суммой*)** двух событий A и B называется событие $A \cup B$ (или $A + B$), состоящее в появлении хотя бы одного из событий A или B.

Пересечением **(*произведением*)** двух событий A и B называется событие $A \cap B$ (или $A \cdot B$), состоящее в одновременном появлении двух событий A и B.

Разностью двух событий A и B называется событие $A \setminus B$ (или $A - B$), состоящее в том, что происходит событие A, но не происходит событие B.

Событием, противоположным событию A, называется событие $\overline{A}$, состоящее в том, что событие A не произошло.

События A и B называются ***несовместными***, если их одновременное появление невозможно, т. е. если $A \cap B = \varnothing$.

События $A_1, A_2, \ldots, A_n$ называются ***попарно несовместными***, если в одном и том же опыте никакие два из них не могут произойти вместе.

Пример 1 – Сколькими способами можно выбрать старосту и его заместителя в группе из 26 человек?

Решение

Так как выбор старосты и его заместителя проводится по определенным критериям, то в этой комбинации важен состав (порядок). Найдем число размещений из 26 элементов по 2: $A_{26}^2 = \dfrac{26!}{(26-2)!} = \dfrac{26!}{24!} = 25 \cdot 26 = 650.$

Пример 2 **–** Сколько четырехзначных чисел можно составить из цифр 1, 3, 5, 8, если цифры в числе не повторяются?

Решение

Так как при записи числа будут использоваться все цифры, найдем число перестановок из четырех элементов: $P_4 = 4 \cdot 3 \cdot 2 \cdot 1 = 4! = 24.$

Пример 3 – Сколькими способами можно взять 7 книг из 12 находящихся на полке?

Решение

Так как порядок выбора книг не важен, то число сочетаний из 12 по 7 равно $C_{12}^{7} = \dfrac{12!}{7!(12-7)!} = \dfrac{12!}{7! \cdot 5!} = \dfrac{8 \cdot 9 \cdot 10 \cdot 11 \cdot 12}{2 \cdot 3 \cdot 4 \cdot 5} = 792.$

2.2 Классическая вероятность. Геометрическая вероятность

Вероятностью события A называется математическая оценка возможности появления этого события в результате опыта.

Классическое определение вероятности. Вероятность события A равна отношению числа m благоприятствующих событию A исходов опыта к общему числу n попарно несовместных исходов опыта, образующих полную группу событий: $P(A) = \dfrac{m}{n}.$

Исход опыта является благоприятствующим событию A, если появление в результате опыта этого исхода влечет за собой появление события A.

Свойства вероятности.

1 Вероятность случайного события A есть положительное число, заключенное между нулем и единицей, т. е. $0 \le P(A) \le 1.$

2 Вероятность достоверного события равна единице, т. е. $P(\Omega) = 1.$

3 Вероятность невозможного события равна нулю, т. е. $P(\varnothing) = 0.$

4 Вероятность противоположного события $P(\overline{A}) = 1 - P(A).$

Пример 4 – Монета бросается дважды. Какова вероятность того, что хотя бы один раз выпадет герб?

Решение

Элементарными исходами в данном опыте будут следующие события:

U_1 – оба раза выпал герб;

U_2 – герб появился только при первом бросании;

U_3 – герб появился только при втором бросании;

U_4 – герб не появился ни разу.

Благоприятствовать событию A (появление герба хотя бы один раз) будут U_1, U_2, U_3. Следовательно, $P(A)=\frac{3}{4}$.

Геометрическая вероятность. Классическое определение вероятности не может быть применимо при бесконечном числе исходов. Пусть точка X выбирается случайным образом в области Ω. Тогда вероятность попадания точки в область A определяется по формуле $P(A)=\frac{\mu_A}{\mu_\Omega}$, где μ_A – мера области A; μ_Ω – мера области Ω.

Мерой является длина соответствующих отрезков (в случае пространства R^1), площадь областей (для пространства R^2), объём соответствующих тел (для пространства R^3).

Геометрическая вероятность обладает всеми свойствами, присущими классическому определению вероятности.

Пример 5 – Два друга условились встретиться в определённом месте между 11 и 12 часами дня. Друг, пришедший первым, ждёт второго в течение $1/4$ ч, после чего уходит. Найти вероятность того, что встреча состоится, если каждый из друзей выбирает наудачу момент своего прихода в течение указанного часа.

Решение

Пусть x – время (в часах) прихода первого друга, а y – второго. Рассмотрим области Ω и A: $A=\left\{(x,y)\in R^2 : 0\le x\le 1,\ 0\le y\le 1,\ |x-y|\le\frac{1}{4}\right\}$ и $\Omega=\left\{(x,y)\in R^2 : 0\le x\le 1,\ 0\le y\le 1\right\}$ (рисунок 1).

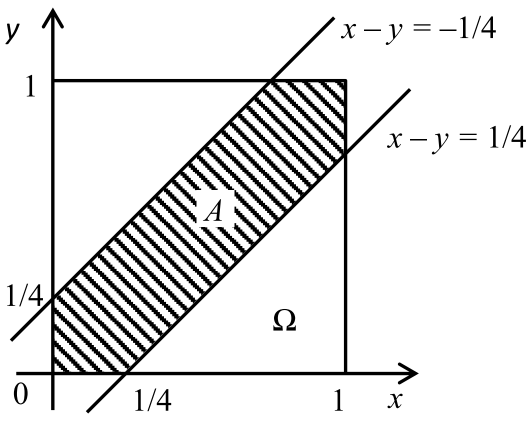

Рисунок 1

$$S_\Omega = 1\cdot 1 = 1\,\text{ед.}^2;\ S_A = 1-2\cdot\frac{1}{2}\cdot\left(\frac{3}{4}\right)^2=\frac{7}{16}\ \text{ед.}^2;\ P(A)=\frac{S_A}{S_\Omega}=\frac{7}{16}\approx 0{,}44.$$

2.3 Теоремы сложения и умножения вероятностей

Теорема 1 (теорема сложения несовместных событий). *Вероятность суммы двух несовместных событий* A и B $(A \cdot B = \varnothing)$ *равна сумме вероятностей этих событий*: $P(A+B) = P(A) + P(B)$.

Теорема 2 (теорема сложения произвольных событий). *Вероятность суммы двух произвольных событий равна сумме вероятностей этих событий без вероятности их произведения*: $P(A+B) = P(A) + P(B) - P(A \cdot B)$.

Следствие. *Сумма вероятностей противоположных событий равна единице, т. е.* $P(A) + P(\overline{A}) = 1$.

Используется также следующая формулировка: вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления.

Событие A называется ***независимым*** от события B, если вероятность события A не зависит от того, произошло событие B или нет. Событие A называется ***зависимым*** от события B, если вероятность события A меняется в зависимости от того, произошло событие B или нет.

Вероятность события B, вычисляемая при условии, что имеет место событие A, называется ***условной вероятностью*** события B: $P_A(B) = P(B|\mathrm{A}) = \dfrac{P(A \cdot B)}{P(A)}$.

Теорема 3 (умножение вероятностей независимых событий). *Вероятность произведения двух независимых событий* A и B *равна произведению вероятностей этих событий*: $P(A \cdot B) = P(A) \cdot P(B)$.

Теорема 4 (умножение вероятностей зависимых событий). *Вероятность совместного появления двух зависимых событий равна произведению вероятности одного из них на условную вероятность другого, вычисленную при условии, что первое событие уже наступило*: $P(A \cdot B) = P(A) \cdot P(B|A)$.

Пример 6 – Из партии изделий проверяется половина и признается годной вся партия, если среди проверенных изделий бракованных не более одного. Какова вероятность того, что партия из 20 изделий, в которой два бракованных, будет признана годной?

Решение

Рассмотрим события: A – среди проверяемых изделий бракованных не окажется; B – среди проверяемых изделий одно бракованное. Партия будет признана годной, если произойдет событие $A+B$. Эти события несовместны, и, следовательно, справедлива формула $P(A+B) = P(A) + P(B)$.

Подсчитаем вероятность события A. Из 20 изделий можно отобрать для проверки 10 изделий C_{20}^{10} способами. Из небракованных изделий 10 изделий отбираются C_{18}^{10} способами. Поэтому $P(A)=\dfrac{C_{18}^{10}}{C_{20}^{10}}=\dfrac{9}{38}$.

Для события B количество благоприятных исходов равно $C_{18}^{9}\cdot C_{2}^{1}$ (число способов отбора девяти небракованных изделий и одного бракованного), поэтому $P(B)=\dfrac{C_{18}^{9}\cdot C_{2}^{1}}{C_{20}^{10}}=\dfrac{20}{38}$.

По теореме сложения двух несовместных событий $P(A+B)=\dfrac{29}{38}$.

***Пример** 7* – Каждая буква слова КРУГОЗОР написана на отдельной карточке. Из этих букв составляется наудачу слово, состоящее из четырех букв. Какова вероятность того, что получится слово УЗОР?

Решение

Пусть событие A – получится слово УЗОР, A_1, A_2, A_3, A_4 – события, состоящие в последовательном извлечении букв У, З, О, Р. Эти события зависимые. Согласно формуле умножения вероятностей, получим

$$P(A)=P(A_1\cdot A_2\cdot A_3\cdot A_4)=$$
$$=P(A_1)\cdot P(A_2\mid A_1)\cdot P(A_3\mid A_1\cdot A_2)\cdot P(A_4\mid A_1\cdot A_2\cdot A_3)=\frac{1}{8}\cdot\frac{1}{7}\cdot\frac{2}{6}\cdot\frac{2}{5}=\frac{1}{420}.$$

2.4 Полная вероятность. Формула Байеса

Теорема 5. *Полная вероятность события A, которое может произойти вместе с одним из событий $H_1, H_2, \ldots, H_n$, равна сумме парных произведений вероятностей каждого из этих событий на соответствующие им условные вероятности наступления события A:* $P(A)=\sum_{i=1}^{n}P(H_i)\cdot P(A\mid H_i)$.

Теорема 6. *Вероятность гипотезы после испытания равна произведению вероятности гипотезы до испытания на соответствующую ей условную вероятность события, которое произошло при испытании, деленному на полную вероятность этого события*:

$$P(H_i\mid A)=\frac{P(H_i)\cdot P(A\mid H_i)}{\sum_{i=1}^{n}P(H_i)\cdot P(A\mid H_i)}\ \text{(формула Байеса).}$$

Пример 8 – В сборочный цех завода поступает 40 % деталей из первого цеха и 60 % из второго цеха. В первом цехе производится 90 % стандартных изделий, а во втором – 95 %. Найти вероятность того, что наудачу взятая сборщиком деталь окажется стандартной.

Решение

Взятие детали можно разбить на два этапа. Первый этап – это выбор цеха. На этом этапе имеются две гипотезы: H_1 – деталь изготовлена первым цехом, H_2 – деталь изготовлена вторым цехом. События H_1 и H_2 образуют полную группу; $P(H_1)=0,4$ и $P(H_2)=0,6.$

Второй этап – взятие детали. Событие A – взятая наудачу деталь стандартная. Числа 0,90 и 0,95 являются условными вероятностями события A при условии гипотез H_1 и H_2 соответственно, т. е. $P(A\,|\,H_1)=0,90$ и $P(A\,|\,H_2)=0,95.$

По формуле полной вероятности находим вероятность события A:

$$P(A)=\sum_{i=1}^{2}P(H_i)\cdot P(A\,|\,H_i)=0,4\cdot 0,90+0,6\cdot 0,95=0,93.$$

Пример 9 – Два стрелка стреляют по мишени. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0,6, а для второго – 0,8. Найти вероятность того, что при одном залпе в мишень попадет только один из стрелков.

Решение

Пусть попадание в цель первым стрелком – событие A, вторым – событие B, промах первого стрелка – событие $\overline{A},$ промах второго – событие $\overline{B}.$

Вероятность того, что в мишень попадет первый стрелок, а второй – нет, равна $P(A)\cdot P(\overline{B})=0,6\cdot 0,2=0,12.$

Вероятность того, что в мишень попадет второй стрелок, а первый – нет, равна $P(\overline{A})\cdot P(B)=0,4\cdot 0,8=0,32.$

Вероятность попадания в цель только одним стрелком равна $P=0,12+0,32=0,44.$

Пример 10 – Трое рабочих изготовили партию деталей, при этом первый рабочий изготовил 25 % всех деталей и в его продукции 5 % брака, второй – 35 % и брак составляет 4 %, третий – 40 % всех деталей с браком в 2 %. Случайно выбранная для контроля деталь оказалась бракованной. Какова вероятность того, что она изготовлена вторым рабочим?

Решение

Обозначим через H_1, H_2 и H_3 гипотезы, заключающиеся в том, что случайно выбранная для контроля деталь изготовлена соответственно первым, вторым и третьим рабочим. Тогда $P(H_1)=0,25$, $P(H_2)=0,35$, $P(H_3)=0,40$.

Событие A – выбранная для контроля деталь оказывается бракованной. Вероятности события A при условии, что имеют место гипотезы H_1, H_2 и H_3, даны в условии задачи: $P(A\,|\,H_1)=0,05$, $P(A\,|\,H_2)=0,04$, $P(A\,|\,H_3)=0,02$.

По формуле Байеса находим

$$P(H_2\,|\,A)=\frac{P(H_2)\cdot P(A\,|\,H_2)}{P(H_1)\cdot P(A\,|\,H_1)+P(H_2)\cdot P(A\,|\,H_2)+P(H_3)\cdot P(A\,|\,H_3)}=$$

$$=\frac{0,35\cdot 0,04}{0,25\cdot 0,05+0,35\cdot 0,04+0,40\cdot 0,02}=\frac{28}{69}.$$

2.5 Схема независимых повторных испытаний

Рассмотрим конечную последовательность n независимых испытаний, в результате каждого из которых может произойти событие A с вероятностью $P(A)=p$ или же противоположное ему событие $\overline{A}$ с вероятностью $P(\overline{A})=q=1-p$. Считается, что вероятность p события A в каждом испытании одна и та же. По условию результат любого испытания не зависит от его порядкового номера и от того, какие исходы были в предыдущих испытаниях. Такую последовательность испытаний принято называть ***схемой испытаний Бернулли*** (или *схемой независимых повторных испытаний*).

Вероятность того, что в результате n испытаний событие A произойдет ровно k раз, обозначается $P_n(k)$.

При нахождении нужной вероятности необходимо учитывать заданные значения n и p.

1 Форма Бернулли применяется в случае, если задано число испытаний n и оно не больше 10; тогда $P_n(k)$ определяется по ***формуле Бернулли*** $P_n(k)=C_n^k\cdot p^k q^{n-k}$ или $P_n(k)=\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$.

2 Если число испытаний n велико, а вероятность p появления события A в каждом испытании очень мала и $\lambda<10$, то для вычисления $P_n(k)$ используют приближенную ***формулу Пуассона*** $P_n(k)\approx\frac{\lambda^k\cdot e^{-\lambda}}{k!}$, где $\lambda=np$ – среднее число появления события A.

Имеются специальные таблицы для соответствующих значений функции Пуассона $e^{-\lambda}\dfrac{\lambda^k}{k!}$.

3 Если требуется определить вероятность массовых (n велико) и нередких (p не мало) событий $(np>10)$, тогда используется ***локальная теорема Муавра – Лапласа***. По этой теореме вероятность $P_n(k)\approx\dfrac{1}{\sqrt{npq}}\cdot\varphi(x)$ – формула ***Муавра – Лапласа***. Здесь $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ – функция Гаусса; $x=\dfrac{k-np}{\sqrt{npq}}$.

Существуют табличные данные, соответствующие положительным значениям функции $\varphi(x)$(таблица значений функции $\varphi(x)$). Так как функция $\varphi(x)$ четная, то $\varphi(-x)=\varphi(x)$ и для значений $x>4$ $\varphi(x)\approx 0$.

4 Если требуется определить вероятность того, что в серии из n независимых испытаний число наступлений некоторого события A будет не менее k_1 раз и не более k_2 раз, то для вычисления применяется ***интегральная теорема Муавра – Лапласа***. По этой теореме вероятность $P_n(k_1;k_2)\approx\Phi(x_2)-\Phi(x_1)$, где $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{t^2}{2}}dt,\ x_1=\dfrac{k_1-np}{\sqrt{npq}},\ x_2=\dfrac{k_2-np}{\sqrt{npq}}$.

Функция $\Phi(x)$ называется функцией Лапласа. Она нечетная, т. е. $\Phi(-x)=-\Phi(x)$. Имеются таблицы для соответствующих положительных значений этой функции и для значений $x>5$ $\Phi(x)\approx 0{,}5$.

Пример 11 – По мишени производится пять выстрелов, причем вероятность попадания при каждом выстреле равна 0,8. Какова вероятность того, что мишень была поражена не менее трех раз?

Решение

Вероятность поражения мишени не менее трех раз складывается из вероятности пяти, четырех и трех поражений. Так как поражения мишени – независимые события, то применим формулу Бернулли:

$$P_5(5)=C_5^5\cdot 0{,}8^5(1-0{,}8)^{5-5}=0{,}8^5=0{,}32768;$$

$$P_5(4)=C_5^4\cdot 0{,}8^4(1-0{,}8)^{5-4}=\frac{5!}{4!\cdot 1!}0{,}8^4\cdot 0{,}2^1=0{,}8^4=0{,}4096;$$

$$P_5(3)=C_5^3\cdot 0{,}8^3(1-0{,}8)^{5-3}=\frac{5!}{3!\cdot 2!}0{,}8^3\cdot 0{,}2^2=0{,}4\cdot 0{,}8^3=0{,}2048.$$

Вероятность того, что мишень была поражена не менее трех раз при пяти выстрелах, равна $P = 0,32768 + 0,4096 + 0,2048 = 0,94208$.

Пример 12 – Фирма рассылает по почте своим клиентам 600 писем с проспектами новой продукции. Вероятность того, что при пересылке письмо потеряется, равна 0,001. Найти вероятность того, что при пересылке потеряется хотя бы одно письмо с проспектами.

Решение

Количество посланных писем $n = 600$ велико, а вероятность того, что при пересылке письмо потеряется, мала $\left(p = 0,001\right)$, и $\lambda = np = 0,6 < 10$. Используем формулу Пуассона $P_n\left(k\right) \approx \frac{\lambda^k}{k!}e^{-\lambda}$.

Вероятность того, что при пересылке потеряется хотя бы одно письмо, равна $P_{500}\left(k \geq 1\right) = 1 - P_{500}\left(0\right) = 1 - \frac{0,5^0}{0!}e^{-0,5} = 1 - \frac{1}{\sqrt{e}} = 1 - 0,60653 \approx 0,40$.

Пример 13 – Вероятность попадания в мишень при одном выстреле для данного стрелка равна 0,7. Найти вероятность того, что при 200 выстрелах мишень будет поражена 160 раз.

Решение

Так как $n = 200$, $p = 0,7$, то $\lambda = np = 200 \cdot 0,7 = 140 > 10$. Будем использовать локальную теорему Муавра – Лапласа. Вероятность $q = 1 - 0,7 = 0,3$. Тогда $\sqrt{npq} = \sqrt{200 \cdot 0,7 \cdot 0,3} \approx 6,48$, $\varphi\left(x\right) = \varphi\left(\frac{160 - 200 \cdot 0,7}{\sqrt{200 \cdot 0,7 \cdot 0,3}}\right) \approx \varphi\left(3,09\right) \approx 0,0034$. Вероятность $P_{200}\left(160\right) \approx 0,0005$.

Пример 14 – Игральную кость подбрасывают 600 раз. Какова вероятность того, что при этом выпадет от 90 до 120 «шестерок»?

Решение

Вероятность выпадения «шестерки» при одном бросании $p = \frac{1}{6}$. Вероятность противоположного события $q = 1 - \frac{1}{6} = \frac{5}{6}$. Для вычисления искомой вероятности используем интегральную формулу Муавра – Лапласа:

$$P_{600}(90;120)\approx\Phi\left(\frac{120-600\cdot\frac{1}{6}}{\sqrt{600\cdot\frac{1}{6}\cdot\frac{5}{6}}}\right)-\Phi\left(\frac{90-600\cdot\frac{1}{6}}{\sqrt{600\cdot\frac{1}{6}\cdot\frac{5}{6}}}\right)=\Phi(2,19)-\Phi(-1,1)=$$

$$=0,48574+0,36433=0,85007.$$

Задачи для самостоятельного решения

1 Сколькими способами можно рассадить пять человек за столом?

2 Сколько различных буквосочетаний можно получить перестановкой карточек с буквами М, И, С, С, И, С, И, П, И.

3 Сколькими способами можно расставить 15 томов на книжной полке, если выбирать их из 30, имеющихся в наличии?

4 Имеется по одному билету в театр, в цирк и на концерт. Сколькими способами их можно распределить между четырьмя студентами (если каждый студент может получить сколько угодно билетов)?

5 В олимпиаде по программированию может участвовать команда из трех студентов группы. Сколько возможностей составить команду, если в группе двадцать студентов?

6 Имеется пять видов календарей с цветами. Сколькими способами из них можно выбрать семь календарей?

7 Сколькими способами можно составить трехцветный флаг из полос разной ширины, если имеются материи из 8 тканей?

8 В кондитерской имеются пирожные трех видов. Сколькими способами можно заказать набор, состоящий из пяти пирожных?

9 На световом табло в один ряд располагаются шесть лампочек. Сколько различных сигналов можно получить, имея две зеленые и четыре красные лампочки? Все лампочки должны гореть.

10 В лифт восьмиэтажного дома вошли пять пассажиров. Сколькими способами могут выйти пассажиры на каждом этаже, начиная со второго?

11 Два игрока играют в шахматы. Событие A: выиграл первый игрок; событие B: выиграл второй игрок. Что означают события $\overline{A}\cdot\overline{B}$, $\overline{B}\setminus A$, $\overline{A}\setminus B$?

12 Набирая номер телефона, абонент забыл две последние (различные) цифры и набрал их наугад. Какова вероятность того, что номер набран правильно?

13 Среди 100 электроламп пять испорченных. Какова вероятность того, что выбранные наудачу три лампы окажутся исправными?

14 В урне находится 15 белых, пять красных и 10 чёрных шаров. Наугад извлекается один шар. Найти вероятность того, что он будет белым, красным, чёрным.

15 Вероятности попадания в цель каждого из четырех независимых друг от друга стрелков соответственно равны 0,9; 0,8; 0,6; 0,7. Стрелки выстрелили в

цель одновременно. Найти вероятность того, что произойдет хотя бы одно попадание; не более двух попаданий.

16 У сборщика имеется четыре детали первого типа и пять деталей второго типа. Сначала он взял одну деталь, а затем вторую. Какова вероятность того, что первая из взятых деталей есть деталь первого типа, а вторая – второго?

17 В цехе работают три станка-автомата, которые штампуют однотипные детали. Производительности станков относятся как 2:3:5. Вероятность брака для первого станка равна 0,05, для второго – 0,03, для третьего – 0,02. Изготовленные детали складывают в один ящик. Найти вероятность того, что взятая наугад из ящика деталь будет бракованной; наугад взятая бракованная деталь изготовлена на втором станке.

18 Студент знает не все экзаменационные билеты. Что для него выгоднее: отвечать первым или вторым?

19 В приборе четыре лампы. Вероятность выхода из строя в течение года для каждой лампы равна 0,2. Какова вероятность того, что в течение года придется заменить не менее половины всех ламп?

20 Телефонная станция обслуживает 1000 абонентов. Вероятность того, что в течение минуты какому-либо абоненту понадобится соединение, составляет 0,0007. Найти вероятность того, что за минуту на телефонную станцию поступит не менее трех вызовов.

21 Какова вероятность того, что в столбике из ста наугад отобранных монет число монет, расположенных гербом вверх, будет от 45 до 55?

22 Вероятность поражения мишени стрелком при одном выстреле равна 0,8. Найти вероятность того, что при 100 выстрелах мишень будет поражена ровно 75 раз.

23 Найти вероятность того, что сумма двух наугад взятых положительных правильных дробей не больше единицы, а их произведение не больше 3/16.

24 На отрезке, длина которого l, наугад поставлены две точки, в результате чего этот отрезок оказывается разделённым на три части. Найти вероятность того, что из трёх получившихся частей можно составить треугольник.

25 Два парохода должны подойти к одному и тому же причалу. Время прихода обоих пароходов независимо и равновозможно в течение суток. Определить вероятность того, что одному из пароходов придётся ожидать освобождения причала, если время стоянки первого – 1 ч, второго – 2 ч.

3 Дискретные случайные величины

Случайной называют величину, которая в результате испытания примет одно, и только одно числовое значение, зависящее от случайных факторов, и заранее непредсказуемое. Случайные величины, как правило, обозначают буквами $X, Y, \ldots$, а их значения – соответствующими маленькими буквами с подстрочными индексами, например, $x_1, x_2, \ldots, x_n$.

Для полной характеристики случайной величины необходимо знать те значения, которые она может принимать, и вероятность, с которой принимается то или иное значение.

Дискретная (прерывная) случайная величина – величина, принимающая отдельно взятые, изолированные значения. Количество этих значений конечно либо бесконечно, но счётно.

Непрерывная случайная величина – величина, принимающая все числовые значения из некоторого конечного или бесконечного промежутка.

Если для дискретной случайной величины X известны все значения $x_1, x_2, \ldots, x_n$, которые она может принимать, и все вероятности $p_1, p_2, \ldots, p_n$, с которыми эти значения принимаются, то указанное соответствие между возможными значениями величины X и их вероятностями называется ***законом распределения*** дискретной случайной величины X.

Закон распределения может быть задан аналитически, в виде таблицы (в первой строке записываются все значения x_k случайной величины, а во второй (под ними) – соответствующие значения вероятности $p_k = P\left(X = x_k\right)$, причем $\sum\limits_k p_k = 1$) или графически.

Таблица соответствия значений случайной величины и их вероятностей называется ***рядом распределения***. Графическое представление этой таблицы называется ***многоугольником*** (***полигоном***) ***распределения***.

Одним из наиболее удобных и универсальных способов задания закона распределения является функция распределения.

Функцией распределения случайной величины X называется функция $F_X\left(x\right)$ или $F\left(x\right)$, определяющая вероятность того, что случайная величина X в результате испытания примет значение, меньшее $x: F\left(x\right) = P\left(X < x\right)$.

Функция распределения дискретной случайной величины имеет вид $F\left(x\right) = \sum\limits_{x_i < x} p_i$.

Функция распределения обладает следующими ***свойствами***:

– $0 \le F\left(x\right) \le 1$;

– $F\left(x\right)$ – неубывающая функция, т. е. $F\left(x_2\right) \ge F\left(x_1\right)$, если $x_2 > x_1$;

– $F\left(-\infty\right) = 0, F\left(+\infty\right) = 1$;

– $F\left(x\right)$ непрерывна слева в любой точке x, т.е. $F\left(x - 0\right) = F\left(x\right),\ x \in R$;

– $P\left(a \le X \le b\right) = F\left(b\right) - F\left(a\right)$.

Рассмотрим *числовые характеристики* дискретной случайной величины.

Математическим ожиданием дискретной случайной величины называется сумма произведений всех возможных значений случайной величины на их вероятности: $m_x = M\left(x\right) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum\limits_{i=1}^{n} x_i p_i$.

Дисперсией (***рассеиванием***) дискретной случайной величины называется математическое ожидание квадрата отклонения случайной величины от ее математического ожидания: $D(X)=M\left(X-M(X)\right)^2$. Иногда дисперсию удобно вычислять по формуле $D(X)=M\left(X^2\right)-\left(M(X)\right)^2=\sum_{i=1}^{n}x_i^{\,2}p_i-\left(M(X)\right)^2$.

Средним квадратическим отклонением случайной величины называется квадратный корень из дисперсии: $\sigma_x=\sigma(X)=\sqrt{D(X)}$.

Математическое ожидание и дисперсия являются частными случаями более общих понятий – моментов случайной величины.

$\alpha_k=M\left(X^k\right)$ – ***начальный момент порядка*** k. В частности, $\alpha_1=M(X)$. Для дискретной случайной величины $\alpha_k=\sum_i x_i^k p_i$.

$\mu_k=M\left(\left(X-M(X)\right)^k\right)$ – ***центральный момент порядка*** k, $\mu_2=D(X)$. Для дискретной случайной величины $\mu_k=\sum_i\left(x_i-M(X)\right)^k\cdot p_i$.

Коэффициентом асимметрии называется величина $A=\dfrac{\mu_3}{\sigma_x^3}$.

Коэффициентом эксцесса называется величина $E=\dfrac{\mu_4}{\sigma_x^4}-3$.

Модой M_0 дискретной случайной величины называется такое ее значение, которое имеет наибольшую вероятность.

Пример 1 – В урне четыре белых и три черных шара. Из нее последовательно вынимают шары до первого появления белого шара. Построить ряд и многоугольник распределения дискретной случайной величины X – числа извлеченных шаров. Найти функцию распределения $F(x)$.

Решение

Значения случайной величины: $x_1=1$ (первый шар белый), $x_2=2$ (первый шар черный, второй – белый), $x_3=3$ (первый и второй шары черные, третий – белый), $x_4=4$ (первый, второй и третий шары черные, четвертый – белый). Соответствующие им вероятности $p_1=P(X=1)=\dfrac{4}{7}$; $p_2=P(X=2)=\dfrac{3}{7}\cdot\dfrac{4}{6}=\dfrac{2}{7}$; $p_3=P(X=3)=\dfrac{3}{7}\cdot\dfrac{2}{6}\cdot\dfrac{4}{5}=\dfrac{4}{35}$; $p_4=P(X=4)=\dfrac{3}{7}\cdot\dfrac{2}{6}\cdot\dfrac{1}{5}\cdot\dfrac{4}{4}=\dfrac{1}{35}$.

Получаем ряд распределения, представленный таблицей 1.

Таблица 1

x_i	1	2	3	4
p_i	$\frac{4}{7}$	$\frac{2}{7}$	$\frac{4}{35}$	$\frac{1}{35}$

Контроль: $\left(\sum_{i=1}^{4} p_i = \frac{4}{7} + \frac{2}{7} + \frac{4}{35} + \frac{1}{35} = 1\right)$.

Если точки с координатами $(x_i; p_i)$ изобразить на координатной плоскости и затем соединить их последовательно отрезками, то получится многоугольник (полигон) распределения (рисунок 2).

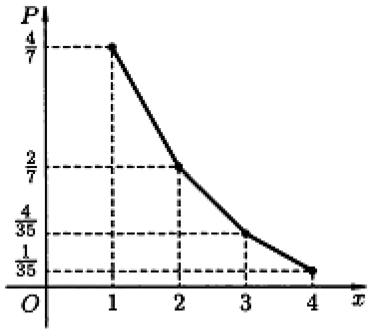

Рисунок 2

Найдем $F(x)$, используя формулу $F(x) = \sum_{x_i < x} p_i$. Получим

$$F(x) = \begin{cases} 0, & x \le 1; \\ \frac{4}{7}, & 1 < x \le 2; \\ \frac{4}{7} + \frac{2}{7} = \frac{6}{7}, & 2 < x \le 3; \\ \frac{4}{7} + \frac{2}{7} + \frac{4}{35} = \frac{34}{35}, & 3 < x \le 4; \\ \frac{4}{7} + \frac{2}{7} + \frac{4}{35} + \frac{1}{35} = 1, & x > 4. \end{cases}$$

Пример 2 – Дискретная случайная величина задана рядом распределения (таблица 2). Найти функцию распределения $F(x)$ и построить ее график.

Таблица 2

x_i	–2	1	2	3
p_i	0,08	0,40	0,32	0,20

По определению функции распределения находим соответствующие вероятности (значения функции $F(x)$): если $x \le -2$, то $F(x) = P(X < x) = 0$; если $-2 < x \le 1$, то $F(x) = P(X = -2) = 0{,}08$; если $1 < x \le 2$, то $F(x) = P(X = -2) + P(X = 1) = 0{,}08 + 0{,}40 = 0{,}48$; если $2 < x \le 3$, то $F(x) = P(X = -2) + P(X = 1) + P(X = 2) = 0{,}08 + 0{,}40 + 0{,}32 = 0{,}80$; если $x > 3$, то $F(x) = P(X = -2) + P(X = 1) + P(X = 2) + P(X = 3) = 0{,}08 + 0{,}40 = 0{,}32 + 0{,}20 = 1$. Функция распределения имеет вид:

$$F(x) = \begin{cases} 0, & x \le -2; \\ 0{,}08, & -2 < x \le 1; \\ 0{,}48, & 1 < x \le 2; \\ 0{,}80, & 2 < x \le 3; \\ 1, & x > 3. \end{cases}$$

График функции $F(x)$ представлен на рисунке 3.

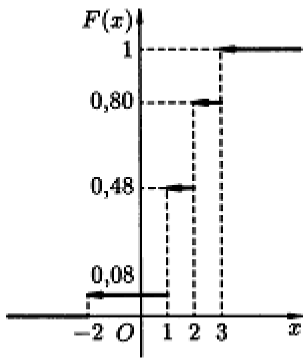

Рисунок 3

Пример 3 – В лотерее имеется 100 билетов, из них выигрышных: 1 по 50 р., 5 по 20 р., 15 по 10 р., 10 по 2 р. Найти числовые характеристики случайной величины X – суммы выигрыша на один билет.

Решение

Возможные значения величины X: $x_1 = 50$; $x_2 = 20$; $x_3 = 10$; $x_4 = 2$; $x_5 = 0$. Соответствующие им вероятности: $p_1 = \frac{1}{100} = 0{,}01$; $p_2 = \frac{5}{100} = 0{,}05$; $p_3 = \frac{15}{100} = 0{,}15$; $p_4 = \frac{10}{100} = 0{,}1$; $p_5 = \frac{100 - (1 + 5 + 15 + 10)}{100} = 0{,}69$. Ряд распределения имеет следующий вид (таблица 3).

Таблица 3

x_i	50	20	10	2	0
p_i	0,01	0,05	0,15	0,1	0,69

$$M(X)=50\cdot 0{,}01+20\cdot 0{,}05+10\cdot 0{,}15+2\cdot 0{,}1+0\cdot 0{,}69=3{,}2;$$

$$D(x)=\sum_{i=1}^{5}x_i^2p_i-\left(M(X)\right)^2=50^2\cdot 0{,}01+20^2\cdot 0{,}05+10^2\cdot 0{,}15+2^2\cdot 0{,}1-(3{,}2)^2=$$

$$=25+20+15+0{,}4-(3{,}2)^2=60{,}4-10{,}24=50{,}16;$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{50{,}16}\approx 7{,}08.$$

Задачи для самостоятельного решения

1 Наудачу отобраны четыре детали. Написать закон распределения дискретной случайной величины X – числа нестандартных деталей среди четырёх отобранных (табличный, графический, аналитический).

2 В урне пять белых и 25 черных шаров. Наугад вынули один шар. Случайная величина X – число вынутых белых шаров. Найти $M(X)$ и $D(X)$. Построить ряд распределения и функцию распределения случайной величины X.

3 Испытывается устройство, состоящее из четырех приборов, работающих независимо. Вероятности отказа каждого из приборов соответственно равны $p_1=0{,}3$, $p_2=0{,}4$, $p_3=0{,}5$, $p_4=0{,}6$. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение числа отказавших приборов.

4 Охотник, имеющий пять патронов, стреляет в цель до первого попадания (или пока не израсходует все патроны). Найти математическое ожидание и дисперсию случайной величины X – числа израсходованных патронов. Построить график функции распределения этой случайной величины, если вероятность попадания при каждом выстреле равна 0,4.

5 Два стрелка стреляют по одной мишени, делая независимо друг от друга по два выстрела. Вероятность попадания в мишень для первого стрелка – 0,4, для второго – 0,5. Пусть случайная величина X – числа попаданий в мишень для первого стрелка, Y – число попаданий вторым стрелком. Построить ряд распределения случайной величины $Z=X-Y$ – общего числа попадания – и найти её числовые характеристики $M(Z)$ и $D(Z)$.

6 Прибор состоит из трех независимо работающих элементов. Вероятность отказа каждого элемента в отдельном опыте равна 0,1. Для случайной величины X – числа отказавших элементов в одном опыте – найти ряд распределения в виде таблицы; функцию распределения $F(x)$ (и построить ее график); числовые характеристики $M(X), D(X), \sigma(X)$.

7 В группе из шести изделий имеется одно бракованное. Для его обнаружения выбирают наугад одно изделие за другим и каждое вынутое проверяют до тех пор, пока не найдут бракованное. Найти числовые характеристики $M(X), D(X), \sigma(X)$.

8 Имеется шесть ключей, из которых только один подходит к замку. Составить ряд распределения случайной величины X – числа попыток при открывании замка, если ключ, не подошедший к замку, в последующих опробованиях не участвует. Найти числовые характеристики $M(X)$ и $D(X)$.

9 Найти функцию распределения для числа выпадений герба при двух подбрасываниях монеты и построить ее график.

10 Случайная величина X может принимать два возможных значения: x_1 с вероятностью $p_1 = 0{,}3$ и x_2 с вероятностью $p_2 = 0{,}7$. Найти x_1 и x_2 $(x_2 > x_1)$, если математическое ожидание $M(X) = 2{,}7$ и дисперсия $D(X) = 0{,}21$.

11 Дан перечень возможных значений дискретной случайной величины X: $x_1 = 1$; $x_2 = 2$; $x_3 = 3$. Известно, что $M(X) = 2{,}3$ и $M(X^2) = 5{,}9$. Найти вероятности, соответствующие возможным значениям X .

12 Независимые дискретные величины X и Y таковы, что $M(X) = 2$, $M(Y) = -3$, $D(X) = 2$, $D(Y) = 9$. Найти $M(Z)$ и $D(Z)$, если $Z = 5X - 3Y + 2$.

4 Непрерывные случайные величины

Случайная величина X называется ***непрерывной случайной*** величиной, если:

– ее возможные значения сплошь заполняют промежуток (или несколько промежутков) числовой оси;

– существует такая неотрицательная функция $f(x)$, называемая ***плотностью распределения вероятности***, что вероятность попадания случайной величины X в промежуток $[a;b]$ равна определенному интегралу от $f(x)$ по этому промежутку, т. е. $P(a \le X \le b) = \int\limits_a^b f(x)dx$.

Свойства плотности распределения:

1) $f(x) \ge 0,\ x \in (-\infty; +\infty)$;

2) $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$;

3) $f(-\infty) = 0,\ f(+\infty) = 1$.

Для непрерывной случайной величины ***функция распределения*** $F(x)$ определяется из равенства $F(x) = \int\limits_{-\infty}^{x} f(t)dt$. Это означает, что функция распределения $F(x)$ непрерывной случайной величины является первообразной для ее плотности распределения вероятности $f(x)$, т. е. $f(x)=F'(x)$.

Рассмотрим числовые характеристики непрерывной случайной величины.

$M(X)=\int\limits_{-\infty}^{+\infty} x f(x)dx$ – ***математическое ожидание***.

$D(X)=\int\limits_{-\infty}^{+\infty} \left(x-M(X)\right)^2 f(x)dx$ – ***дисперсия***.

$\gamma_k = M\left(X^k\right)=\int\limits_{-\infty}^{+\infty} x^k f(x)dx$ – ***начальный момент порядка*** k.

$\mu_k = \int\limits_{-\infty}^{+\infty} \left(x-M(X)\right)^k f(x)dx$ – ***центральный момент порядка*** k.

$\sigma(X)=\sqrt{D(X)}$ – ***среднее квадратическое отклонение***.

Медианой M_e случайной величины X называется такое ее значение x_0, для которого выполняется равенство $P(X<x_0)=P(X>x_0)$. Вычисляется медиана из условия $F(x)=\frac{1}{2}$.

Мода M_0 непрерывной случайной величины с плотностью вероятности $f(x)$ есть точка x_0 максимума плотности вероятности $f(x)$.

Пример – Плотность распределения вероятности случайной величины X задана формулой

$$f(x)=\begin{cases} a\left(x^2+2x\right), \ 0\le x\le 4; \\ 0, \ x<0, \ x>4. \end{cases}$$

Найти коэффициент *a*, функцию распределения $F(x)$, $M(X)$ и $D(X)$, вероятность попадания величины X вне интервала $(2{,}5;3)$.

Решение

Найдем коэффициент $a: \int\limits_{-\infty}^{+\infty} f(x)dx = \int\limits_{-\infty}^{o} 0\cdot dx + \int\limits_{0}^{4} a\cdot\left(x^2+2x\right)dx + \int\limits_{4}^{+\infty} 0\cdot dx =$
$=a\int\limits_{0}^{4}\left(x^2+2x\right)dx = a\left(\frac{x^3}{3}+x^2\right)\Bigg|_0^4 = a\left(\frac{64}{3}+16\right)=a\cdot\frac{112}{3}$. Из равенства $a\cdot\frac{112}{3}=1$

получим $a=\dfrac{3}{112}$.

Так как $F(x)=\int\limits_{-\infty}^{x} f(t)dt$, то для $0\le x\le 4$ получим

$$F(x)=\int\limits_{-\infty}^{x} f(t)dt=\int\limits_{-\infty}^{0} 0\cdot dt+\int\limits_{0}^{x}\frac{3}{112}\left(t^2+2t\right)dt=\frac{3}{112}\left(\frac{t^3}{3}+t^2\right)\Bigg|_0^x=\frac{3}{112}\left(\frac{x^3}{3}+x^2\right).$$

Тогда

$$F(x)=\begin{cases}0, & x<0;\\ \dfrac{3}{112}\left(\dfrac{x^3}{3}+x^2\right), & 0\le x\le 4;\\ 1, & x>4.\end{cases}$$

$$M(X)=\int\limits_{-\infty}^{o} 0\cdot dx+\frac{3}{112}\int\limits_{0}^{4} x\cdot\left(x^2+2x\right)dx+\int\limits_{4}^{+\infty} 0\cdot dx=\frac{3}{112}\int\limits_{0}^{4}\left(x^3+2x^2\right)dx=$$

$$=\frac{3}{112}\left(\frac{x^4}{4}+\frac{2x^3}{3}\right)\Bigg|_0^4=\frac{3}{112}\left(\frac{128}{3}+64\right)=\frac{320}{112}\approx 2{,}86.$$

$$D(X)=M\left(X^2\right)-\left(M(X)\right)^2=\int\limits_{-\infty}^{o} 0\cdot dx+\frac{3}{112}\int\limits_{0}^{4} x^2\cdot\left(x^2+2x\right)dx+\int\limits_{4}^{+\infty} 0\cdot dx-(2{,}86)^2=$$

$$=\frac{3}{112}\left(\frac{x^5}{5}+\frac{x^4}{2}\right)\Bigg|_0^4-(2{,}86)^2=\frac{998{,}4}{112}-(2{,}86)^2\approx 8{,}91-(2{,}86)^2=0{,}7304.$$

$$P(2{,}5<X<3)=F(3)-F(2{,}5)=\frac{3}{112}(9+9)-\frac{3}{112}\left(6{,}25+\frac{15{,}625}{3}\right)\approx 0{,}18.$$

Вероятность того, что случайная величина попадет вне указанного интервала, находится по формуле вероятностей противоположных событий. Тогда получим $P\left(X\in(2{,}5;3)\right)=1-P(2{,}5<X<3)=1-0{,}18=0{,}82$.

Задачи для самостоятельного решения

1 Непрерывная случайная величина X задана плотностью распределения вероятности

$$f(x)=\begin{cases}0, & |x|\ge\dfrac{\pi}{2};\\ \dfrac{\cos x}{2}, & |x|<\dfrac{\pi}{2}.\end{cases}$$

Найти функцию распределения и построить графики плотности распределения вероятности и функции распределения.

2 Непрерывная случайная величина X имеет плотность распределения вероятности

$$f(x)=\begin{cases}a(x+1), & -1\le x\le 1;\\ 0, & x<-1,\ x>1.\end{cases}$$

Определить коэффициент a, функцию распределения $F(x)$, найти числовые характеристики случайной величины X $\left(M(X), D(X), \sigma(X)\right)$, вычислить вероятности $P(0\le X<1)$, $P\left(-2<X<\dfrac{1}{4}\right)$, $P\left(\dfrac{1}{2}<X<2\right)$. Построить графики функций $F(x)$ и $f(x)$.

3 Непрерывная случайная величина X задана функцией распределения

$$F(x)=\begin{cases}0, & x\le 0;\\ ax^2, & 0<x\le 14;\\ 1, & x>14.\end{cases}$$

Определить функцию распределения $f(x)$, найти коэффициент a, вычислить вероятность $P(7\le X<14)$, найти числовые характеристики $M(X)$, $D(X)$, $\sigma(X)$.

5 Основные распределения случайных величин

Дискретная случайная величина имеет **биномиальное распределение**, если она принимает значения $0, 1, 2, 3, \dots, n$ с вероятностями:

$$P_n(k)=C_n^k p^k q^{n-k};$$

$$M(X)\ \ M(X)=np;\ D(X)=npq.$$

Случайная величина X, распределенная по биномиальному закону, является числом успехов с вероятностью p в схеме Бернулли проведения n независимых опытов.

Распределение Пуассона. Если число испытаний n достаточно велико, а вероятность появления события A в каждом испытании мала $(p \le 0{,}1)$, то

$$P_n(k) = \frac{\lambda^k e^{-\lambda}}{k!},\ \lambda = np,\ k = 0, 1, 2, \ldots, n;$$

$$M(X) = D(X) = \lambda.$$

Непрерывная случайная величина X имеет **равномерное распределение** на отрезке $[a;b]$, если на этом отрезке плотность распределения случайной величины постоянна, а вне его равна нулю:

$$f(x) = \begin{cases} 0, & x \in [a;b]; \\ \dfrac{1}{b-a}, & x \notin [a;b]. \end{cases}$$

$$F(x) = \begin{cases} 0, & x < a; \\ \dfrac{x-a}{b-a}, & a \le x \le b; \\ 1, & x > b; \end{cases}\ M(X) = \frac{a+b}{2};\ D(X) = \frac{(b-a)^2}{12};\ \sigma(X) = \frac{b-a}{2\sqrt{3}}.$$

$$P(\alpha \le x \le \beta) = \frac{\beta - \alpha}{b-a},\ a \le \alpha \le \beta \le b.$$

Непрерывная случайная величина X называется распределенной по **показательному (экспоненциальному)** закону, если ее плотность вероятности

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0; \\ 0, & x < 0, \end{cases}$$

где λ – положительное число.

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \ge 0; \\ 0, & x < 0, \end{cases}\ M(X) = \sigma(X) = \frac{1}{2};\ D(X) = \frac{1}{\lambda^2};\ m_x = \frac{1}{\lambda}.$$

$$P(\alpha \le X \le \beta) = e^{-\lambda\alpha} - e^{-\lambda\beta}.$$

Непрерывная случайная величина X называется распределенной по **нормальному закону**, если ее плотность вероятности задается функцией

$$f(x)=\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}.$$

Параметры m_x и σ_x являются соответственно математическим ожиданием и средним квадратическим отклонением случайной величины X.

Функция распределения случайной величины, распределенной по нормальному закону, имеет вид:

$$F(X)=\frac{1}{\sigma_x\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}dx.$$

Вероятность попадания нормальной случайной величины X в интервал $(\alpha;\beta)$ $P(\alpha\le X\le\beta)=\Phi\left(\frac{\beta-m_x}{\sigma}\right)-\Phi\left(\frac{\alpha-m_x}{\sigma}\right)$, где $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{\frac{t^2}{2}}dt$ или $\Phi^*(x)=\frac{1}{\sqrt{2\pi}}\int\limits_{0}^{x}e^{\frac{t^2}{2}}dt$ – функция Лапласа, значения которой приведены в [1–5].

$P(|X-m_x|<\varepsilon)=2\Phi\left(\frac{\varepsilon}{\sigma}\right)$ $(\varepsilon>0)$ – вероятность отклонения случайной величины от среднего.

$P(|X-m_x|<3\sigma)=2\Phi(3)=0{,}9973\approx 1-$ **правило «трех сигм»**: практически достоверно, что значения нормально распределенной случайной величины принадлежат промежутку $(m_x-3\sigma;m_x+3\sigma)$.

Пример 1 – Производится три независимых выстрела по цели. Вероятности попадания при разных выстрелах одинаковы и равны $p=0{,}9$. Найти $M(X)$ и $D(X)$, где X – число попаданий в цель.

Решение

Случайная величина X имеет биномиальное распределение. Так как $n=3$, $p=0{,}9$, то $q=1-0{,}9=0{,}1$; тогда $M(X)=np=2{,}7$, $D(X)=npq=0{,}27$.

Пример 2 – Вероятность попадания в цель при одном выстреле равна 0,01. Какова вероятность того, что число попаданий при 200 выстрелах составит не менее 5 и не более 10?

Решение

Случайная величина X – число попаданий. Вероятность $p = 0{,}01$ очень мала, а число выстрелов $n = 200$ достаточно велико. Искомую вероятность найдем по формуле Пуассона $P_n(k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$. Находим $\lambda = np = 200 \cdot 0{,}01 = 2$, $e^{-\lambda} = e^{-2} \approx 0{,}135$. По теореме сложения вероятностей найдем $P(5 \le X \le 10)$: $P(5 \le X \le 10) = P(X = 5) + P(X = 6) + \ldots + P(X = 10)$. Тогда

$$P(5 \le X \le 10) = 0{,}135\left(\frac{2^5}{5!} + \frac{2^6}{6!} + \frac{2^7}{7!} + \frac{2^8}{8!} + \frac{2^9}{9!} + \frac{2^{10}}{10!}\right) \approx 0{,}053.$$

Пример 3 – Поезда метрополитена идут регулярно с интервалом две минуты. В случайный момент времени, никак не связанный с расписанием поездов, на платформу выходит пассажир. Найти вероятность того, что ему придется ждать не более полминуты.

Решение

Данная случайная величина X имеет равномерное распределение на отрезке $[0;2]$ (событие может наступить в любой момент времени на протяжении 2 мин). Функция распределения случайной величины имеет вид:

$$F(x) = \begin{cases} 0, & x \le 0; \\ \dfrac{x}{2}, & 0 < x \le 2; \\ 1, & x > 2. \end{cases}$$

Тогда искомая плотность $P\left(0 < X < \dfrac{1}{2}\right) = F\left(\dfrac{1}{2}\right) - F(0) = \dfrac{1}{4}$.

Пример 4 – Время безотказной работы прибора распределено по показательному закону $f(t) = \lambda e^{-\lambda t}$, $t \ge 0$. Определить вероятность того, что прибор проработает безотказно 100 ч, если среднее время безотказной работы таких приборов составляет 50 ч.

Решение

Из условия следует, что $\dfrac{1}{\lambda} = M(X) = 50$, значит, $\lambda = 0{,}02$. Найдем $P(X \ge 100) = 1 - F(100) = e^{-2} = 0{,}13534$. Вероятность безотказной работы прибора в течение 100 ч равна 0,14.

Пример 5 – Математическое ожидание нормально распределенной случайной величины X равно 10, а ее среднее квадратическое отклонение равно 2. Найти вероятность того, что в результате испытания случайная величина X примет значение, заключенное в интервале $(12;14)$.

Решение

Для данной случайной величины X $m_x = 10$ и $\sigma = 2$.

Используя формулу $P(\alpha \le X \le \beta) = \Phi\left(\dfrac{\beta - m_x}{\sigma}\right) - \Phi\left(\dfrac{\alpha - m_x}{\sigma}\right)$, найдем

$$P(12 \le X \le 14) = \Phi\left(\frac{14-10}{2}\right) - \Phi\left(\frac{12-10}{2}\right) = \Phi(2) - \Phi(1).$$

По таблице значений функции $\Phi(x)$ находим значения $\Phi(2) = 0{,}9772$, $\Phi(1) = 0{,}8413$. Искомая вероятность равна $0{,}9772 - 0{,}8413 = 0{,}1359$.

Задачи для самостоятельного решения

1 Производится 20 независимых опытов, в каждом из которых событие A появляется с вероятностью 0,4. Для случайной величины X – числа появлений события A в 20 опытах – найти математическое ожидание, дисперсию, среднее квадратическое отклонение, вероятность попадания в промежуток от 3 до 8.

2 Случайная величина X распределена по закону Пуассона с математическим ожиданием, равным 3. Найти вероятность того, что величина X примет значение меньшее, чем ее математическое ожидание. Определить вероятность того, что величина X примет положительно значение.

3 Случайная величина X распределена равномерно в интервале $(0;2)$. Записать плотность распределения вероятности, функцию распределения, найти $M(X)$, $D(X)$, вероятности $P(1 < X < 15)$, $P(-3 < X < 0{,}4)$.

4 Записать функцию распределения случайной величины X, распределенной по показательному закону $f(x) = \begin{cases} 3e^{-3x}, x \ge 0; \\ 0, x < 0. \end{cases}$ Найти для нее $M(X)$, $D(X)$, $\sigma(X)$, $P(0{,}13 < X < 0{,}7)$.

5 Для нормально распределенной случайной величины X математическое ожидание и среднее квадратическое отклонение равны соответственно 11 и 4. Записать плотность вероятности и функцию распределения этой величины, найти вероятность того, что X примет значение, принадлежащее интервалу

$(13;23)$, и вероятность того, что абсолютная величина отклонения $X - m_x$ окажется меньше 6.

6 Завод отправил на базу 5000 радиоламп. Вероятность того, что в пути лампа будет повреждена, равна 0,0002. Найти вероятность того, что на базу прибудут не более трех поврежденных ламп. Найти среднее число поврежденных ламп.

7 За определенный период времени среднее число ошибочных соединений, приходящееся на одного телефонного абонента, равно 8. Какова вероятность того, что для данного абонента число ошибочных соединений будет больше 4?

8 Телефонная станция получает в среднем 20 вызовов в минуту. Какова вероятность того, что за 6 с она получит ровно 3 вызова?

9 При измерении детали получаются случайные ошибки, подчиненные нормальному закону с параметром $\sigma = 10$ мм. Производится три независимых измерения детали. Найти вероятность того, что ошибка хотя бы одного измерения не превосходит по модулю 2 мм.

6 Генеральная совокупность и выборка. Вариационный ряд, полигон, гистограмма. Эмпирическая функция распределения, ее свойства и график

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

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

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

Для получения хороших оценок характеристик генеральной совокупности необходимо, чтобы выборка была ***репрезентативной*** (или ***представительной***), т. е. достаточно полно представлять изучаемые признаки генеральной совокупности. В силу закона больших чисел можно утверждать, что ***выборка будет репрезентативной, если ее осуществлять случайно***, т. е. все объекты генеральной совокупности должны иметь равные вероятности попасть в выборку.

Конкретные значения выборки $x_1, x_2, x_3, \ldots, x_k$, полученные в результате наблюдений (испытаний), называют ***реализацией выборки*** или ***вариантами выборки***.

Если значение x_1 встречается в выборке n_1 раз, x_2 – n_2 раз, …, x_k – n_k раз, то $n_1 + n_2 + \ldots + n_k = n$ – ***объем выборки***.

Операция расположения значений случайной величины (признака) по неубыванию называется ***ранжированием*** статистических данных. Полученная таким образом последовательность значений $x_1, x_2, x_3, \ldots, x_n$ случайной величины X называется ***вариационным рядом***.

Числа n_i, показывающие, сколько раз встречаются варианты x_i $(i = 1, 2, 3, \ldots, n)$ в ряде наблюдений, называются ***частотами***. Отношение частот к объему выборки называют ***частостями*** или ***относительными частотами***, которые находятся по формуле $w_i = \dfrac{n_i}{n}$.

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

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

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

В случае дискретного вариационного ряда статистический ряд распределения выборки строится следующим образом: в первой строке таблицы записываются все значения выборки (варианты), во второй строке – соответствующие им значения частот или относительных частот.

Для интервального вариационного ряда в первую строку таблицы статистического распределения записывают частичные промежутки $[x_0, x_1), [x_1, x_2), [x_2, x_3), \ldots, [x_{k-1}, x_k]$, которые, как правило, берутся одинаковыми по длине $h = x_0 - x_1 = x_2 - x_1 = \ldots$ Число h называется *шагом разбиения*.

Длину h частичного промежутка можно найти по формуле Стерджеса $h = \dfrac{x_{\max} - x_{\min}}{1 + \log_2 n}$. Число интервалов $m = 1 + \log_2 n$. При этом $\log_2 n \approx 3{,}322 \lg n$.

Разность между наибольшим и наименьшим значениями случайной величины $x_{\max} - x_{\min}$ называется ***размахом вариации***. За начало первого интервала рекомендуется брать $x_{нач} = x_{\min} - \dfrac{h}{2}$. Во вторую строку таблицы записывают количество вариантов $n_i \left(i = \overline{1, k}\right)$.

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

Полигоном частот называют ломаную линию, отрезки которой соединяют точки с координатами $(x_1, n_1), (x_2, n_2), \ldots, (x_k, n_k)$, а ***полигоном относительных частот*** – ломаную линию, отрезки которой соединяют точки с координатами $(x_1, w_1), (x_2, w_2), \ldots, (x_k, w_k)$. Полигон, как правило, служит для изображения дискретного статистического ряда.

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых являются частичные интервалы длиною h, а высоты равны отношению $\frac{n_i}{h}$; ***гистограммой относительных частот*** **–** ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты равны отношению $\frac{w_i}{h}$. Гистограмма более употребительна для интервальных статистических рядов.

Площадь гистограммы частот равна объему выборки, а площадь гистограммы частостей равна единице.

Одним из способов обработки вариационного ряда является построение ***эмпирической (статистической) функции распределения*** $F^*(x)$, определяющей для каждого значения x относительную частоту события $X < x$: $F^*(x) = \frac{n_x}{n}$, где n – объем выборки, n_x – число наблюдений, меньших x.

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

1) $0 \le F^*(x) \le 1$;

2) $F^*(x)$ – неубывающая функция;

3) $F^*(x) = 0$, если $x \le x_1$, где x_1 – наименьшая варианта;

$F^*(x) = 1$, если $x > x_k$, где x_k – наибольшая варианта.

Если X – дискретная случайная величина, то $F^*(x)$ ступенчатая, её скачки соответствуют наблюдаемым значениям и равны относительным частотам этих значений:

$$F^*(x) = \begin{cases} 0, & x \le x_1; \\ w_1, & x_1 < x \le x_2; \\ w_1 + w_2, & x_2 < x \le x_3; \\ \ldots & \\ \sum_{i=1}^{k} w_i = 1, & x > x_k. \end{cases}$$

Если X – непрерывная случайная величина, то в качестве x берут границы интервалов, записанных в интервальном ряду:

$$F^*(x)=\begin{cases}0, & x \le x_1;\\ w_1, & x = x_2;\\ w_1+w_2, & x = x_3;\\ \dots & \\ \sum\limits_{i=1}^{k-1} w_i, & x = x_k;\\ 1, & x \ge x_{k+1}.\end{cases}$$

Пример 1 – Построить статистический ряд распределения абсолютных и относительных частот выборки и полигон для распределения числа покупателей X , обратившихся в кассу супермаркета за один час. Наблюдения проводились в течение 30 ч (15 дней с 9 до 10 и с 10 до 11 часов). Получены следующие результаты:

80, 90, 100, 110, 90, 60, 100, 110, 80, 60, 70, 100, 70, 100, 80, 90, 60, 100, 100, 110, 80, 90, 80, 110, 70, 80, 90, 80, 100, 100.

Составить эмпирическую функцию распределения $F^*(x)$ и построить ее график.

Решение

Проранжировав статистические данные (т. е. исходный ряд), получим *вариационный ряд* (60, 60, 60, 70, 70, 70; 80, 80, 80, 80, 80, 80, 80, 90, 90, 90, 90, 90, 100, 100, 100, 100, 100, 100, 100, 100, 110, 110, 110, 110). Получили шесть различных значений случайной величины (шесть вариант).

Подсчитав частоту (n_i) и относительную частоту (w_i) значений каждой варианты $x_2 = 1$, $x_3 = 2$, $x_4 = 3$, $x_5 = 4$, $x_6 = 5$ (число обращений в кассу), получим статистическое распределение выборки (*дискретный статистический ряд*). Объём выборки $n = 10$, относительная частота $p_i^* = w_i = \dfrac{n_i}{n}$.

Результаты вычислений запишем в таблицу 4.

Таблица 4

x_i	60	70	80	90	100	110	
n_i	3	3	7	5	8	4	$n = \sum\limits_i n_i = 30$
w_i	$\frac{3}{30}$	$\frac{3}{30}$	$\frac{7}{30}$	$\frac{5}{30}$	$\frac{8}{30}$	$\frac{4}{30}$	$\sum\limits_i w_i = 1$

Полигон относительных частот представлен на рисунке 4.

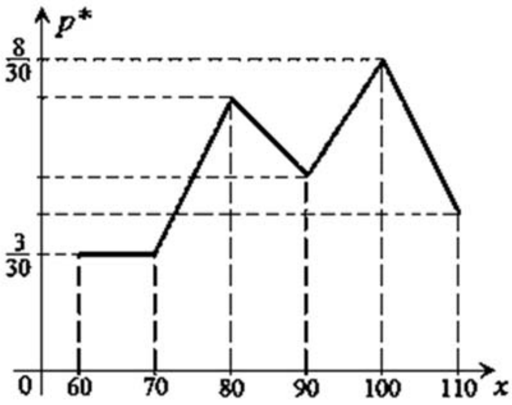

Рисунок 4

Для построения эмпирической функции распределения найдем значения $F^*(x) = \dfrac{n_x}{n}$. Имеем $F^*(x) = \dfrac{0}{30} = 0$ при $x \le 60$ (наблюдений меньше 60 нет); $F^*(x) = \dfrac{3}{30}$ при $60 < x \le 70$; $F^*(x) = \dfrac{3}{30} + \dfrac{3}{30} = \dfrac{6}{30}$ при $70 < x \le 80$ и т. д. Окончательно получим

$$F^*(x) = \begin{cases} 0, & x \le 60; \\ \dfrac{3}{30}, & 60 < x \le 70; \\ \dfrac{3}{30} + \dfrac{3}{30} = \dfrac{6}{30}, & 70 < x \le 80; \\ \dfrac{6}{30} + \dfrac{7}{30} = \dfrac{13}{30}, & 80 < x \le 90; \\ \dfrac{13}{30} + \dfrac{5}{30} = \dfrac{18}{30}, & 90 < x \le 100; \\ \dfrac{18}{30} + \dfrac{8}{30} = \dfrac{26}{30}, & 100 < x \le 110; \\ \dfrac{26}{30} + \dfrac{4}{30} = 1, & x > 110. \end{cases}$$

График эмпирической функции распределения приведен на рисунке 5.

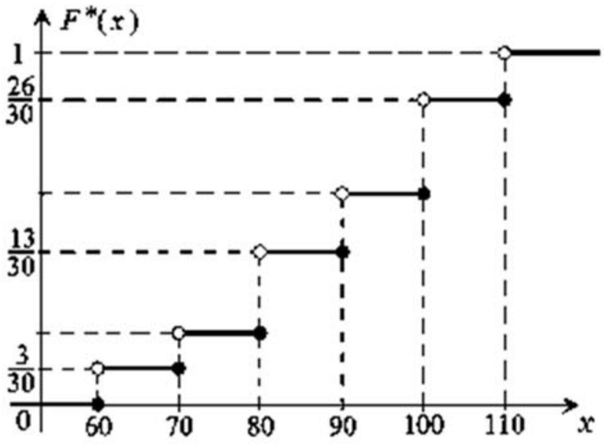

Рисунок 5

Пример 2 – 105 наудачу отобранным студентам измерили рост (с точностью до 1 см) и получили результаты, представленные в таблице 5.

Таблица 5

155	170	185	180	188	152	173	178	178	168	185	172	170	183	175
173	170	183	175	180	175	193	178	183	180	197	178	181	187	168
174	179	184	183	178	180	178	163	166	178	175	182	190	167	170
178	183	170	178	181	173	168	185	175	170	155	169	186	179	189
156	174	179	179	169	186	174	171	184	175	193	178	184	180	196
175	181	188	168	179	178	183	184	178	181	177	163	166	178	175
183	190	167	170	178	183	170	178	182	173	168	186	176	171	188

Составить интервальный статистический ряд распределения роста студентов. Построить гистограмму относительных частот. Составить эмпирическую функцию распределения и построить ее график для интервального статистического ряда.

Решение

Проранжировав данные результаты, получаем, что $x_{\min} = 152$, $x_{\max} = 197$. Величина X – рост студентов с указанной точностью – является непрерывной случайной величиной.

По формуле Стерджеса (для $n = 105$) найдем длину частичного интервала $h = \dfrac{197 - 152}{1 + \log_2 105} = \dfrac{45}{1 + 3{,}322\lg 105} = \dfrac{45}{7{,}714} \approx 5{,}8$.

Принимая $h = 6$, найдем значение $x_{нач} = 152 - \dfrac{6}{2} = 149.$

Определим количество интервалов m: $m = 1 + \log_2 105 \approx 8.$

Исходные данные разбиваем на восемь интервалов: [149,155), [155,161), [161,167), [167,173), [173,179), [179,185), [185,191), [191,197).

Подсчитав число студентов n_i, попавших в каждый из полученных промежутков, получим интервальный статистический ряд (таблица 6).

Таблица 6

Рост	149…155	155…161	161…167	167…173	173…179	179…185	185…191	191…197
n_i	3	1	6	22	33	26	10	4
w_i	0,03	0,01	0,06	0.21	0,31	0,25	0,09	0,04

Здесь $n = \sum_{i=1}^{8} n_i = 105$.

Доопределим таблицу 6 серединами частичных интервалов $\tilde{x}_i = \dfrac{x_{i-1} + x_i}{2}$, $i = 1, 2, \ldots, k$. Запишем таблицу 7.

Таблица 7

Рост	149…155	155…161	161…167	167…173	173…179	179…185	185…191	191…197
n_i	3	1	6	22	33	26	10	4
w_i	0,03	0,01	0,06	0,21	0,31	0,25	0,09	0,04
$\tilde{x}_i$	152	158	164	170	176	182	188	194

Для построения гистограммы относительных частот найдем высоты f_i^* прямоугольников как отношение $\dfrac{w_i}{h}$: $f_1^* = \dfrac{0,03}{6} \approx 0,005$; $f_2^* = \dfrac{0,01}{6} \approx 0,002$; $f_3^* = \dfrac{0,06}{6} \approx 0,01$; $f_4^* = \dfrac{0,21}{6} \approx 0,035$; $f_5^* = \dfrac{0,31}{6} \approx 0,052$; $f_6^* = \dfrac{0,25}{6} \approx 0,042$; $f_7^* = \dfrac{0,09}{6} \approx 0,015$; $f_8^* = \dfrac{0,04}{6} \approx 0,067$.

Гистограмма относительных частот представлена на рисунке 6.

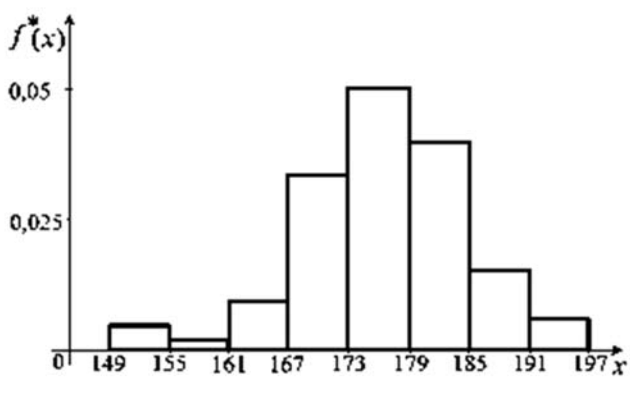

Рисунок 6

Значения эмпирической функции $F^*(x)$:

$$F^*(x)=\begin{cases}0, & x\le 149;\\ 0{,}03, & x=155;\\ 0{,}04, & x=161;\\ 0{,}10, & x=167;\\ 0{,}31, & x=173;\\ 0{,}62, & x=179;\\ 0{,}87, & x=185;\\ 0{,}96, & x=191;\\ 1, & x\ge 197.\end{cases}$$

График эмпирической функции $F^*(x)$ представлен на рисунке 7.

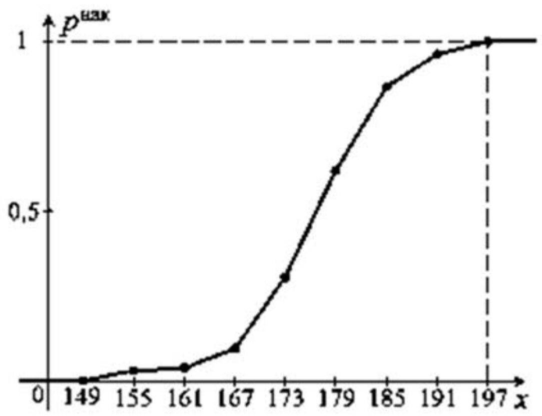

Рисунок 7

Задачи для самостоятельного решения

1 Изучается случайная величина X – число выпавших очков при бросании игральной кости. Кость подбросили 60 раз и получили следующие результаты: 3, 2, 5, 6, 6, 1, 4, 6, 4, 6, 3, 6, 4, 2, 1, 5, 3, 1, 6, 4, 5, 4, 2, 2, 4, 2, 6, 3, 1, 5, 6, 1, 6, 6, 4, 2, 5, 4, 3, 6, 4, 1, 5, 6, 3, 2, 4, 4, 5, 2, 5, 6, 2, 3, 5, 4, 1, 2, 5, 3. Найти эмпирическую функцию распределения выборки. Построить интервальный статистический ряд. Построить полигон частот и гистограмму частостей.

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

ряд распределения частот количественного признака X для данной совокупности.

3 Построить интервальный вариационный ряд распределения процентного содержания жира в молоке от 20 коров, если наблюдения за процентным содержаним жира в молоке дали следующие результаты: 3,86 4,06 3,67 3,97 3,76 3,61 3,86 4,04 3,84 3,94 3,98 3,57 3,87 4,07 3,99 3,69 3,76 3,71 3,94 3,82.

7 Числовые характеристики вариационного ряда. Точечные и интервальные оценки параметров распределения случайной величины по данным выборки

Пусть имеется статистическое распределение выборки объемом n с вариантами x_i и соответствующими им частотами n_i (или частостями w_i).

Характеристики положения: *выборочное среднее значение*, *мода и медиана*.

Выборочным средним $\overline{x}_B$ называется среднее арифметическое всех значений выборки $\overline{x}_B = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$ или $\overline{x}_B = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot w_i$, где $w_i = \frac{n_i}{n}$.

В случае *интервального статистического ряда* в качестве x_i берут середины его интервалов, а n_i – соответствующие им частоты.

Выборочная средняя, вычисленная по данным всех элементов совокупности, равна ***взвешенной средней*** для так называемых частичных средних, т. е. средних, найденных для отдельных частей совокупности, причем частота для каждой частичной средней равна количеству элементов в соответствующей части совокупности.

Если выборка состоит из элементов $x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_l, z_1, z_2, \ldots, z_m$, причем $k + l + m = n$, тогда выборочная средняя этой выборки равна $k\overline{x}_B + l\overline{y}_B + m\overline{z}_B$.

В качестве дополнительной (описательной) числовой характеристики вариационного ряда применяется ***медиана*** (***выборочная медиана***) M_e^*. Это варианта (значение случайной величины X), приходящаяся на середину ряда.

Для ее вычисления все наблюдения располагают в порядке возрастания. При этом если число вариант нечётно, т. е. $n = 2k + 1$, то медианой является $k+1$ варианта $M_e^* = x_{k+1}$. Если же число вариант чётное, то медиана равна среднему арифметическому двух средних значений $M_e^* = \frac{1}{2}\left(x_k + x_{k+1}\right)$.

Для интервального упорядоченного вариационного ряда медиана вычисляется по формуле $M_e^* = x_n + \frac{\frac{n}{2} - S_{M_{e-1}}}{n_{M_e}} \cdot h$, где x_n – начало медианного интервала,

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

Модой (***выборочной модой***) M_0^* называется число, которое наиболее часто встречается в вариационном ряду (варианта, имеющая наибольшую частоту).

Размахом вариации называется число $R = x_n - x_1$, где $x_1 = \min\limits_{1 \le k \le n} x_k$, $x_n = \max\limits_{1 \le k \le n} x_k$ или $R = x_{\max} - x_{\min}$, где $x_{\max}$ – наибольший, $x_{\min}$ – наименьший вариант ряда.

Характеристики разброса или рассеяния.

Выборочной дисперсией D_B (***статистической дисперсией***) называется среднее арифметическое квадратов отклонений значений выборки от выборочной средней $\overline{x}_B$: $D_B = \frac{1}{n}\sum\limits_{i=1}^{k}\left(x_i - \overline{x}_B\right)^2 \cdot n_i$ или $D_B = \sum\limits_{i=1}^{k}\left(x_i - \overline{x}_B\right)^2 \cdot w_i$. Выборочную дисперсию удобно находить по формуле $D_B = \frac{1}{n}\sum\limits_{i=1}^{n} x_i^{\,2} n_i - \overline{x}_B^{\,2}$.

Выборочным средним квадратическим отклонением (стандартной ошибкой) называется квадратный корень из выборочной дисперсии $\sigma_B = \sqrt{D_B}$.

Исправленной выборочной дисперсией называется величина $S^2 = \frac{n}{n-1} D_B$ или $S^2 = \frac{1}{n-1} \cdot \sum\limits_{i=1}^{k}\left(x_i - \overline{x}_B\right)^2 \cdot n_i$.

Исправленным выборочным средним квадратическим отклонением называется величина $S = \sqrt{S^2}$.

Обобщающие характеристики выборочных распределений.

Начальные статистические моменты k ***-го порядка***: $\tilde{\alpha}_k = \frac{1}{n}\sum\limits_{i=1}^{n} x_i^k$ или $\tilde{\alpha}_k = \frac{1}{n}\sum\limits_{i} x_i^k n_i \quad \left(n = \sum\limits_{i} n_i\right)$. Тогда $\tilde{\alpha}_0 = 1$; $\tilde{\alpha}_1 = \overline{x}$; $\tilde{\alpha}_2 = \overline{x}^{\,2}$.

Центральные статистические моменты k ***-го порядка***: при $n = \sum\limits_{i} n_i$

$\tilde{\mu}_k = \frac{1}{n}\sum\limits_{i=1}^{n}\left(x_i - \overline{x}_B\right)^k$ или $\tilde{\mu}_k = \frac{1}{n}\sum\limits_{i} n_i\left(x_i - \overline{x}_B\right)^k$, тогда $\tilde{\mu}_0 = 1$; $\tilde{\mu}_1 = 0$; $\tilde{\mu}_2 = D_B$.

Мера асимметрии распределения выборки определяется $\tilde{\mu}_3$: $A_B = \frac{\tilde{\mu}_3}{\sigma_B^3}$. Если распределение симметрично, то $\tilde{\mu}_3 = 0$.

Эксцесс выборочного распределения определяется $\tilde{\mu}_4$: $E_B = \frac{\tilde{\mu}_4}{\sigma_B^4} - 3$.

Пример 1 – Группа из десяти абитуриентов по результатам тестирования набрала баллы 5, 3, 0, 1, 4, 2, 5, 4, 1, 5. Найти характеристики выборки.

Решение

$$\overline{x}_n = \frac{0\cdot 1+1\cdot 2+2\cdot 1+3\cdot 1+4\cdot 2+5\cdot 3}{10}=3;$$

$$D_B = \frac{(0-3)^2\cdot 1+(1-3)^2\cdot 2+(2-3)^2\cdot 1+(3-3)^2\cdot 1+(4-3)^2\cdot 2+(5-3)^2\cdot 3}{10}==3,2;$$

$$\sigma_B=\sqrt{3,2}\approx 1,79;\quad S^2=\frac{10}{9}\cdot 3,2\approx 3,56;\quad S=\sqrt{3,56}\approx 1,87;$$

$$R=5-0=5;\quad M_0^*=5;\quad M_e^*=\frac{3+4}{2}=3,5.$$

Выборочная числовая характеристика, которая применяется для получения оценки неизвестного параметра θ генеральной совокупности, называется ***точечной оценкой*** (***оценкой одним числом***). Пусть $x_1, x_2, \ldots, x_n$ – выборка, полученная в результате проведения n независимых за случайной величиной X. Чтобы подчеркнуть случайный характер величин $x_1, x_2, \ldots, x_n$, перепишем их в виде $X_1, X_2, \ldots, X_n$, т. е. под X_i будем понимать значение случайной величины X в i-м опыте (это n независимых «экземпляров» величины X).

К оценке любого параметра предъявляют ряд *требований*, чтобы она была приближена к истинному значению параметра и имела практическую ценность: *несмещённость*, *эффективность* и *состоятельность*.

Оценка $\tilde{\theta}$ параметра θ называется ***несмещенной***, если $M(\tilde{\theta})=\theta$. Если $M(\tilde{\theta})\neq\theta$, то оценка $\tilde{\theta}$ называется ***смещенной***.

Среди несмещенных оценок одного и того же параметра выделяют эффективные оценки. Несмещенная оценка $\tilde{\theta}_n$ параметра θ называется ***эффективной***, если она имеет (при заданном объеме выборки) *наименьшую* возможную дисперсию D при $n\to\infty$.

Если существует больше одной несмещённой оценки, то выбирают более эффективную оценку. Оценка $\tilde{\theta}_1$ называется ***более эффективной***, чем оценка $\tilde{\theta}_2$, если $D(\tilde{\theta}_1)<D(\tilde{\theta}_2)$.

Оценка $\tilde{\theta}$ называется ***состоятельной оценкой*** параметра θ, если при $n\to\infty$ она сходится по вероятности к θ, т. е. $\lim\limits_{n\to\infty} P\left(\left|\tilde{\theta}_n-\theta\right|<\varepsilon\right)=1$.

Для *математического ожидания* $M(X)$ несмещённой, эффективной и состоятельной оценкой является ***выборочное среднее*** $\overline{X}_B=\frac{1}{n}\cdot\sum\limits_{i=1}^{n} X_i$.

Оценкой для дисперсии $D(X)$ случайной величины X служит её выборочная дисперсия D_B. Эта оценка состоятельная, смещённая. Несмещённой и состоятельной оценкой *дисперсии* является ***исправленная статистическая (выборочная) дисперсия*** $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{n}{n-1}D_B$.

Оценкой *среднего квадратического отклонения* случайной величины X служат σ_B и исправленное среднеквадратическое отклонение $S_X = \sqrt{\frac{n}{n-1}}\sigma_B$.

Интервальное оценивание. Оценка неизвестного параметра называется интервальной, если она определяется двумя числами – концами интервала. Задача интервального оценивания формулируется следующим образом: по данным выборки построить числовой интервал, относительно которого с заранее выбранной вероятностью можно утверждать, что этот интервал накрывает оцениваемый параметр. Интервальные оценки особенно необходимы при малом числе экспериментов, когда точечные оценки малонадёжны.

Пусть $\theta -$ оцениваемый параметр генеральной совокупности; $\tilde{\theta}$ – статистическая оценка этого параметра по выборке.

Доверительным интервалом называется интервал $\left(\tilde{\theta}-\varepsilon;\tilde{\theta}+\varepsilon\right)$, который покрывает оцениваемый параметр θ с заданной надёжностью $1-\alpha=\gamma$. Число $\gamma = P\left(\left|\tilde{\theta}-\theta\right| < \varepsilon\right)$ называют также доверительной вероятностью; $\varepsilon -$ точность оценки.

Очевидно, что чем меньше для данного γ будет ε (длина интервала), тем точнее оценка $\tilde{\theta}$. В качестве доверительной вероятности обычно берут $\gamma = 0{,}9; 0{,}95; 0{,}99$ или $0{,}999$. Тогда практически достоверно нахождение параметра θ в доверительном интервале $\left(\tilde{\theta}-\varepsilon, \tilde{\theta}+\varepsilon\right)$.

Для точного нахождения доверительных интервалов необходимо знать закон распределения случайной величины X, тогда как для применения приближенных методов это необязательно.

Доверительный интервал для математического ожидания нормально распределённой случайной величины X.

1 Если σ известно, доверительная вероятность γ задана, то доверительным интервалом для $M\left(X\right)$ является интервал $\left(\bar{X} - t\frac{\sigma}{\sqrt{n}}, \bar{X} + t\frac{\sigma}{\sqrt{n}}\right)$.

Значение t находим, используя таблицу функции Лапласа, из равенства $2\Phi\left(t\right)=\gamma$, $\Phi\left(t\right)=\frac{\gamma}{2}$.

2 Если σ неизвестно, γ задана, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{n}{n-1}D_B -$ исправленное среднее квадратическое отклонение случайной величины X, то доверительным интервалом для $M\left(X\right)$ является интервал $\left(\bar{X} - t_\gamma \cdot \frac{s}{\sqrt{n}}, \bar{X} + t_\gamma \cdot \frac{s}{\sqrt{n}}\right)$.

На практике значение t_γ находим по таблице распределения Стьюдента в зависимости от заданной доверительной вероятности γ и числа степеней свободы $n-1$, где t_γ– квантиль уровня $1-\gamma$.

Доверительный интервал для среднего квадратического отклонения нормально распределённой случайной величины X.

1 Пусть среднее квадратическое отклонение σ неизвестно, γ задана, $M(X)=a$ известно, тогда доверительным интервалом для σ является интервал $\left(\frac{\sqrt{n}\cdot S_0}{\chi_2}, \frac{\sqrt{n}\cdot S_0}{\chi_1}\right)$, где n – объем выборки, $S_0^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-a\right)^2$, $\chi_1^2=\chi^2_{\frac{1+\gamma}{2},n}$, $\chi_2^2=\chi^2_{\frac{1-\gamma}{2},n}$ – квантили χ^2-распределения с n степенями свободы (находятся по таблице квантилей $\chi^2_{a,n}$ распределения χ_n^2) [1–6].

2 Если $M(X)=a$ неизвестно, то доверительный интервал для неизвестного σ имеет вид $\left(\frac{\sqrt{n-1}\cdot S}{\chi_2}; \frac{\sqrt{n-1}\cdot S}{\chi_1}\right)$, где n – объем выборки, $S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$, квантили $\chi_1^2=\chi^2_{\frac{1+\gamma}{2},n-1}$, $\chi_2^2=\chi^2_{\frac{1-\gamma}{2},n-1}$ определяются по таблице $\chi^2_{a,k}$ при $k=n-1$, $a=\frac{1+\gamma}{2}$ и $a=\frac{1-\gamma}{2}$ соответственно.

Пример 2 – Произведено пять независимых наблюдений над случайной величиной X. Получены следующие результаты: $x_1=-25$, $x_2=34$, $x_3=-20$, $x_4=10$, $x_5=21$, $\sigma=20$. Найти точечную оценку для $a=M(X)$, а также построить для $M(X)$ доверительный интервал с надежностью 0,95 (т. е. 95 % доверительный интервал). Каким будет доверительный интервал указанной надежности, если значение σ неизвестно?

Решение

$X=\overline{X}_B=\frac{1}{5}\left(-25+34-20+10+21\right)=4.$ По условию $\gamma=0{,}95$, $\Phi_0(t)=\frac{\gamma}{2}.$ Значит, $\Phi_0(t)=0{,}475.$ По таблице значений функций $\Phi_0(t)$ выясняем, что $t=1{,}96$. Так как σ известно по условию задачи, то для нахождения доверительного интервала для $M(X)$ используем формулу $\left(\overline{X}-t\frac{\sigma}{\sqrt{n}}, \overline{X}+t\frac{\sigma}{\sqrt{n}}\right).$

Найдем $t\cdot\frac{\sigma}{\sqrt{n}}=\frac{1{,}96\cdot 20}{\sqrt{5}}\approx 1{,}75,$ тогда доверительным интервалом для $a=M(X)$ является интервал $\left(4-17{,}5;4+17{,}5\right)$, т. е. $\left(-13{,}5;21{,}5\right)$.

Если σ неизвестно, то используем формулу $\left(\overline{X}-t_\gamma\cdot\frac{s}{\sqrt{n}},\overline{X}+t_\gamma\cdot\frac{s}{\sqrt{n}}\right)$.

Найдем $S^2=\frac{1}{4}\left((-25-4)^2\cdot 1+(34-4)^2+(-20-4)^2+(10-4)^2+(21-4)^2\right)=660,5;$ $S\approx 25,7.$ По таблице для $\gamma=0,95$ и $n-1=4$ находим $t_\gamma=2,78.$ Следовательно, $\varepsilon=2,78\cdot\frac{25,7}{2,24}\approx 31,9.$ Доверительный интервал имеет вид $(-27,9;35,9)$.

Пример 3 **–** Для оценки среднего квадратического отклонения нормально распределенной случайной величины X была сделана выборка объема в 30 единиц. Найденное значение $S=1,5$. Найти доверительный интервал, покрывающий σ с вероятностью $\gamma=0,90$.

Решение

По условию $n=30$ и $\gamma=0,90$, тогда по таблице значений для $\chi^2_{a,n}$ находим $\chi_1^2=\chi^2_{\frac{1+0,9}{2};30-1}=\chi^2(0,95;29)=17,7$ и $\chi_2^2=\chi^2_{\frac{1-0,9}{2};30-1}=\chi^2(0,05;29)=42,6$. Тогда доверительный интервал имеет вид $\left(\frac{\sqrt{30-1}\cdot 1,5}{\sqrt{42,6}},\frac{\sqrt{30-1}\cdot 1,5}{\sqrt{17,7}}\right)$ или $1,238<\sigma<1,920.$

Задачи для самостоятельного решения

1 Монету подбрасывают n раз с вероятностью p выпадения герба при каждом подбрасывании. Монета выпала гербом n_A раз. Показать несмещенность оценки $\tilde{\theta}=\frac{n_A}{n}$ вероятности $\theta=p$ выпадения герба в каждом опыте.

2 Для оценки параметра нормально распределенной случайной величины была сделана выборка объема 30 единиц и вычислено $S=1,5.$ Найти доверительный интервал, покрывающий σ с вероятностью $\gamma=0,90.$

3 Глубина моря измеряется прибором, систематическая ошибка которого равна нулю, а случайные ошибки распределены нормально с $\sigma=15$ м. Сколько надо сделать независимых измерений, чтобы определить глубину моря с ошибкой не более 5 м при надежности $\gamma=0,9$.

4 Производятся независимые испытания с одинаковой, но с неизвестной вероятностью p появления события A в каждом испытании. Найти доверительный интервал для оценки p с надежностью 0,95, если в 400 испытаниях событие A появилось 80 раз.

8 Проверка статистических гипотез. Критерий согласия Пирсона

Под ***статистической*** гипотезой (или просто гипотезой) понимается всякое высказывание (предположение) о генеральной совокупности, проверяемое по выборке. Статистические гипотезы делятся на гипотезы о параметрах распределения известного вида (***параметрические гипотезы***) и гипотезы о виде неизвестного распределения (***непараметрические гипотезы***).

Гипотезу, однозначно фиксирующую распределение наблюдений, называют ***простой*** (в ней речь идет об одном значении параметра), в противном случае – ***сложной***.

Одну из гипотез выделяют в качестве ***основной*** (или ***нулевой***) и обозначают H_0, а другую (логическое отрицание H_0, т. е. противоположную H_0) выделяют в качестве ***конкурирующей*** (или ***альтернативной***) гипотезы и обозначают H_1. Имея две гипотезы H_0 и H_1, следует на основе выборки $X_1, X_2, \ldots, X_n$ принять либо основную гипотезу H_0, либо конкурирующую H_1.

Функция выборки $T_n = T\left(X_1, X_2, \ldots, X_n\right)$ называется ***статистикой критерия***. Правило, по которому принимается решение принять или отклонить гипотезу H_0(соответственно H_1), называется ***статистическим критерием*** (критерием) проверки гипотезы H_0.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

Областью принятия гипотезы называют совокупность значений критерия, при которых гипотезу принимают.

Основной принцип проверки статистических гипотез: если наблюдаемое значение критерия принадлежит критической области, то гипотезу отвергают; если наблюдаемое значение критерия принадлежит области принятия гипотезы – гипотезу принимают.

Критическими точками $t_{\text{кр}}$ называют точки, отделяющие критическую область от области принятия гипотезы.

Основной принцип проверки гипотез. Множество возможных значений статистики критерия T_n разбивается на два непересекающихся подмножества: *критическую область* S, т. е. область отклонения гипотезы H_0, и область $\overline{S}$ принятия этой гипотезы. Если фактически наблюдаемое значение статистики критерия, вычисляемое по выборке $T_{\text{набл}} = T\left(x_1, x_2, \ldots, x_n\right)$, попадает в критическую область S, то основная гипотеза H_0 отклоняется, а альтернативная гипотеза H_1 принимается; если же $T_{\text{набл}}$ попадает в $\overline{S}$, то принимается H_0 и отклоняется H_1.

Типы ошибок при проверке гипотез. Ошибка первого рода состоит в том, что отвергается нулевая гипотеза H_0, когда на самом деле она верна. *Ошибка* *второго рода* состоит в том, что отвергается альтернативная гипотеза H_1, когда на самом деле она верна.

Вероятность ошибки первого рода (обозначается через α) называется ***уровнем значимости критерия***, $\alpha = P(H_1 \setminus H_0)$. Чем меньше α, тем меньше вероятность отклонить верную гипотезу. Обычно для α используются стандартные значения $\alpha = 0{,}05;\ 0{,}01;\ 0{,}005;\ 0{,}001$.

Вероятность ошибки второго рода обозначается через β, $\beta = P(H_0 \setminus H_1)$. Величина $1-\beta$ (вероятность недопущения ошибки второго рода) называется ***мощностью критерия***. Чем больше мощность критерия, тем вероятность ошибки второго рода меньше.

Одновременное уменьшение ошибок первого и второго рода возможно лишь при увеличении объема выборок.

Этапы проверки гипотезы.

1 Располагая выборкой $X_1, X_2, \ldots, X_n$, формулируют нулевую гипотезу H_0 и альтернативную гипотезу H_1.

2 В каждом конкретном случае подбирают статистику критерия $T_n = T(X_1, X_2, \ldots, X_n)$. Для этого используют специальным образом подобранную случайную величину (выборочную статистику). Её обозначают U или Z – если случайная величина распределена по нормальному закону; F – если она имеет распределение Фишера – Снедекора; T – если она имеет распределение Стьюдента; χ^2 – если она имеет распределение «хи-квадрат» Пирсона.

3 По статистике критерия T_n и уровню значимости α определяют *критическую область* S (и $\overline{S}$). Для ее отыскания достаточно найти *критическую точку* $t_{\text{кр}}$, т. е. границу (или квантиль), отделяющую S от $\overline{S}$.

Границы областей определяются, соответственно, из соотношений: $P(T_n > t_{\text{кр}}) = \alpha,\ t_{\text{кр}} > 0$ для правосторонней критической области S; $P(T_n < t_{\text{кр}}) = \alpha$, $t_{\text{кр}} < 0$ для левосторонней критической области S; $P(T_n < t_{\text{кр}}^{\text{л}}) = P(T_n > t_{\text{кр}}^{\text{п}}) = \dfrac{\alpha}{2}$ для двусторонней критической области S.

Для каждого критерия есть соответствующие таблицы, по которым и находят критическую точку, удовлетворяющую приведенным выше соотношениям.

4 Для полученной реализации выборки $x = (x_1, x_2, \ldots, x_n)$ подсчитывают наблюдаемое значение критерия, т. е. $T_{\text{набл}} = T(x_1, x_2, \ldots, x_n) = t$.

5 Если $t \in S$ (например, $t > t_{\text{кр}}$ для правосторонней области S), то нулевую гипотезу H_0 отвергают; если же $t \in \overline{S}$ $(t < t_{\text{кр}})$, то нет оснований, чтобы отвергать гипотезу H_0.

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

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

Проверка гипотезы о нормальном распределении для *дискретного вариационного ряда* проводится по следующему алгоритму:

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

2) находим число степеней свободы $k = l - 3$, где l – число различных значений вариант;

3) выбрав уровень значимости α критерия по таблице χ^2 – распределения, находим критическую точку (квантиль) $\chi^2_{\alpha,k}$ [3, приложение 5];

4) если $\chi^2_{\text{набл}} \le \chi^2_{\alpha,k}$, то гипотеза H_0 не противоречит опытным данным; если $\chi^2_{\text{набл}} > \chi^2_{\alpha,k}$, то гипотеза H_0 отклоняется.

Проверка гипотезы о нормальном распределении для *непрерывного вариационного ряда* проводится по следующему алгоритму:

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

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

3) вычисляем теоретические вероятности попадания в интервалы $\left(y_i; y_{i+1}\right)$ $P\left(y_i \le y \le y_{i+1}\right) = \Phi\left(y_{i+1}\right) - \Phi\left(y_i\right)$, где $\Phi\left(y\right)$ – функция Лапласа;

4) по формуле для χ^2 вычисляем $\chi^2_{\text{набл}}$ – выборочное значение статистики критерия;

5) находим число степеней свободы $k = l - 3$, где l – число интервалов выборки;

6) выбрав уровень значимости α критерия по таблице χ^2 – распределения, находим критическую точку (квантиль) $\chi^2_{\alpha,k}$ [3, приложение 5];

7) если $\chi^2_{\text{набл}} \le \chi^2_{\alpha,k}$, то гипотеза H_0 не противоречит опытным данным; если $\chi^2_{\text{набл}} > \chi^2_{\alpha,k}$, то гипотеза H_0 отклоняется.

Пример – Измерены сто обработанных деталей. Отклонения от заданного размера представлены в таблице 8.

При заданном уровне значимости $\alpha = 0{,}01$ проверить гипотезу о том, что отклонение от заданного размера подчиняется нормальному закону распределения. Так как интервалы должны содержать не менее пяти-восьми вариантов, то интервалы с меньшим количеством вариант следует объединить. Объединим крайние интервалы, число наблюдений в которых меньше 5, с соседними и получим ряд распределения, представленный в таблице 9.

Таблица 8

$[x_i; x_{i+1})$	$[-3;-2)$	$[-2;-1)$	$[-1;0)$	$[0;1)$	$[1;2)$	$[2;3)$	$[3;4)$	$[4;5)$
n_i	3	10	15	24	25	13	7	3

Таблица 9

$[x_i; x_{i+1})$	$[-3;-1)$	$[-1;0)$	$[0;1)$	$[1;2)$	$[2;3)$	$[3;5)$
n_i	13	15	24	25	13	10

Пусть случайная величина X – отклонение от заданного размера. Найдем параметры, определяющие нормальный закон распределения (a и σ). Их точечные оценки: $x_B = \dfrac{1}{100}\left(-2\cdot 13 + (-0{,}5)\cdot 15 + 0{,}5\cdot 24 + \ldots + 4\cdot 10\right) \approx 0{,}885 \approx 0{,}9$,

$$D_B = \frac{\left((-2)^2\cdot 13 + (-0{,}5)^2\cdot 15 + \ldots + 4^2\cdot 10\right)}{100} - (0{,}885)^2 \approx 2{,}809,\ \sigma = \sqrt{2{,}809} \approx 1{,}7.$$

Так как случайная величина X определена на $(-\infty;+\infty)$, то крайние интервалы заменяем интервалами $(-\infty;-1)$ и $(3;+\infty)$. Тогда $p_1 = p(-\infty < X < -1) = = \Phi_0\left(\dfrac{-1-0{,}9}{1{,}7}\right) - \Phi_0(-\infty) = \dfrac{1}{2} - \Phi_0(1{,}12) = 0{,}1314.$ Аналогично получаем $p_2 = 0{,}1667;$ $p_3 = 0{,}2258;$ $p_4 = 0{,}2183;$ $p_5 = 0{,}1503;$ $p_6 = p(3 \le X < \infty) = = \Phi_0(\infty) - \Phi_0\left(\dfrac{3-0{,}9}{1{,}7}\right) = 0{,}5 - \Phi_0(1{,}24) = 0{,}1075.$

Полученные результаты представим в таблице 10.

Таблица 10

$[x_i; x_{i+1})$	$[-\infty;-1)$	$[-1;0)$	$[0;1)$	$[1;2)$	$[2;3)$	$[3;+\infty)$
n_i	13	15	24	25	13	10
p_i	0,1314	0,1667	0,2258	0,2183	0,1503	0,1075
$n\cdot p_i$	13,14	16,67	22,58	21,83	15,03	10,75

По формуле $\chi^2_{\text{набл}} = \sum_{i=1}^{k} \frac{\left(n_i - np_i\right)^2}{np_i}$ вычисляем значение $\chi^2_{\text{набл}}$. Получаем $\chi^2_{\text{набл}} = \frac{\left(13-13{,}14\right)^2}{13{,}14} + \frac{\left(15-16{,}67\right)^2}{16{,}67} + \ldots + \frac{\left(10-10{,}75\right)^2}{10{,}75} \approx 1{,}045$.

Находим число степеней свободы: $k = l - 3 = 6 - 3 = 3$. Для значений $\alpha = 0{,}01$ и $k = 3$ по таблице χ^2 -распределения находим $\chi^2_{\alpha,k} = 11{,}3$. Получили, что $\chi^2_{\text{набл}} < \chi^2_{\alpha,k}$. Следовательно, нет оснований отвергнуть проверяемую гипотезу и данные не противоречат нормальному распределению.

Список литературы

1 **Белько, И. В.** Теория вероятностей, математическая статистика, математическое программирование: учебное пособие / И. В. Белько, И. М. Морозова, Е. А. Криштапович. – Минск: Новое знание; Москва: ИНФРА-М, 2016. – 299 с.

2 **Герасимович, А. И.** Математическая статистика / А. И. Герасимович, Я. И. Матвеева. – Минск: Вышэйшая школа, 1978. – 200 с.

3 **Гмурман, В. Е.** Теория вероятностей и математическая статистика / В. Е. Гмурман. – Москва: Высшая школа, 2004. – 479 с.

4 **Гмурман, В. Е.** Руководство к решению задач по теории вероятностей и математической статистике / В. Е. Гмурман. – Москва: Высшая школа, 2004. – 404 с.

5 **Письменный, Д. Т.** Конспект лекций по теории вероятностей и математической статистике / Д. Т. Письменный. – Москва: Айрис-пресс, 2004. – 256 с.

6 **Письменный, Д. Т.** Конспект лекций по теории вероятностей, математической статистике и случайным процессам / Д. Т. Письменный. – Москва: Айрис-пресс, 2008. – 288 с.