МЕЖГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Высшая математика»

ВЫСШАЯ МАТЕМАТИКА. МАТЕМАТИКА. ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические рекомендации к самостоятельной работе для студентов всех специальностей заочной формы обучения

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Могилев 2020

УДК 519.21
ББК 22.171
В93

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Высшая математика» «28» мая 2020 г., протокол № 9

Составитель Д. В. Роголев

Рецензент канд. техн. наук И. Д. Камчицкая

Методические рекомендации содержат необходимые теоретические сведения по теме «Теория вероятностей и математическая статистика», примеры с решениями и примеры для самостоятельной работы.

Учебно-методическое издание

ВЫСШАЯ МАТЕМАТИКА. МАТЕМАТИКА.
ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Ответственный за выпуск	В. Г. Замураев
Корректор	И. В. Голубцова
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, Могилев.

Содержание

1 Элементы комбинаторики. Алгебра событий

Результат выбора k элементов из группы, содержащей n элементов, называют ***выборкой*** k элементов из n.

Перестановками называют комбинации, состоящие из одних и тех же n различных элементов и отличающиеся только порядком их расположения.

Число всех возможных перестановок вычисляется по формуле $P_n = n!$

Размещениями называют комбинации, составленные из n различных элементов по k элементов, которые отличаются либо составом элементов, либо их порядком.

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

Выражение $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$ называется ***факториалом*** числа n, $0! = 1$.

Сочетаниями называют комбинации, составленные из n различных элементов по k элементов, которые отличаются хотя бы одним элементом.

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

Также справедлива формула $C_n^k = C_n^{n-k}$.

Ранее предполагалось, что все элементы различны. Если же некоторые элементы повторяются, то такие комбинации называются ***комбинациями с повторениями*** и их число находят уже по другим формулам.

Если при выборе k элементов из n элементы возвращаются обратно и упорядочиваются, то говорят, что это ***размещения с повторениями***: $\overline{A}_n^k = n^k$.

Если при выборе k элементов из n элементы возвращаются обратно без последующего упорядочивания, то говорят, что это ***сочетания с повторениями***: $\overline{C}_n^k = C_{n+k-1}^k$.

Если в множестве из n элементов есть k различных элементов, при этом первый элемент повторяется n_1 раз, второй – n_2 раз,… k-й – n_k раз, причём $n_1 + n_2 + ... + n_k = n$, то перестановки элементов данного множества называют ***перестановками с повторениями***: $P_n(n_1, n_2, ..., n_k) = \dfrac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!}$.

Правило суммы. Если некоторый объект A может быть выбран из совокупности объектов a способами, а другой объект B может быть выбран b способами, то выбрать либо A, либо B можно $a + b$ способами.

Правило произведения. Если объект A можно выбрать из совокупности объектов a способами и после каждого такого выбора объект B можно выбрать b способами, то пара объектов (A, B) в данном порядке может быть выбрана $a \cdot b$ способами.

Случайным событием называют некоторый результат опыта. Событие, которое обязательно происходит в данном опыте, называют ***достоверным*** и обозначают Ω; событие, которое никогда не происходит в данном опыте, называют

невозможным и обозначают символом $\varnothing$.

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

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

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

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

Событием, противоположным событию A, называют событие $\overline{A}$, состоящее в том, что событие A не произошло, т. е. $\overline{A} = \Omega \setminus A$.

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

Пример 1 – Сколько четырёхзначных чисел можно составить из цифр 1, 2, 3, 4, если каждая цифра входит в запись числа только один раз?

Решение

Так как используются все цифры, найдём число перестановок из четырёх элементов: $P_4 = 4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$.

Пример 2 – В высшей лиге по футболу 16 команд. Сколькими способами могут быть распределены медали чемпионата?

Решение

Поскольку медали различаются, то порядок распределения важен. Найдём число размещений из 16 по 3: $A_{16}^{3} = \dfrac{16!}{(16-3)!} = \dfrac{16!}{13!} = 16 \cdot 15 \cdot 14 = 3360$.

Пример 3 – Сколькими способами можно выбрать три детали из ящика, содержащего 10 деталей?

Решение

Порядок выбора деталей не важен, используем формулу сочетаний: $C_{10}^{3} = \dfrac{10!}{3!(10-3)!} = \dfrac{10!}{3! \cdot 7!} = \dfrac{8 \cdot 9 \cdot 10}{2 \cdot 3} = 120$.

Задачи для самостоятельного решения

1 В розыгрыше первенства страны по футболу принимают участие 15 команд. Сколькими способами могут быть распределены золотая, серебряная и бронзовая медали? *Ответ*: 2730.

2 В турнире принимали участие 12 шахматистов и каждые два шахматиста встретились один раз. Сколько партий было сыграно в турнире? *Ответ*: 66.

3 В партии из 12 деталей девять стандартных. Сколькими способами можно выбрать шесть деталей, из которых четыре были бы стандартными? *Ответ*: 378.

4 Сколько существует различных пятизначных чисел с неповторяющимися цифрами? *Ответ*: 27216.

5 На базе 20 наименований товаров. Сколькими способами их можно распределить по трём магазинам, если известно, что в первый магазин должно быть доставлено восемь, во второй – семь и в третий – пять наименований товаров? *Ответ*: 99768240.

6 Из спортивного клуба, насчитывающего 30 человек, надо составить команду из четырёх человек для участия в беге на 1000 м. Сколькими способами это можно сделать? Сколькими способами можно составить команду из четырёх человек для участия в эстафете $100+200+400+800$? *Ответ*: 27405; 657720.

7 Сколько различных диагоналей можно провести в восьмиугольнике; двенадцатиугольнике; n-угольнике? *Ответ*: 20; 54; $n\cdot(n-3)/2$.

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 влечёт за собой появление события A.

Свойства вероятности.

1 Вероятность случайного события A есть число от 0 до 1: $0\le P(A)\le 1$.

2 Вероятность невозможного события равна нулю: $P(\varnothing)=0$.

3 Вероятность достоверного события равна 1: $P(\Omega)=1$.

4 Если события A и B несовместны $(AB=\varnothing)$, то $P(A+B)=P(A)+P(B)$.

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

Пример 1 – Производится розыгрыш лотереи «5 из 35». Найти вероятность того, что игрок правильно угадает все пять чисел.

Решение

Обозначим через A событие, состоящее в том, что игрок правильно угадал все пять чисел. Искомая вероятность будет равна отношению числа исходов, благоприятствующих событию, к числу всех элементарных исходов:

$$P(A)=\frac{m}{n};\ \ m=1;\ \ n=C_{35}^{5}=\frac{35!}{5!(35-5)!}=324\,632;\ \ P(A)=\frac{1}{324\,632}\approx 0{,}000003.$$

Пример 2 – В партии из 10 деталей семь стандартных. Найти вероятность того, что среди шести взятых наудачу деталей четыре стандартные.

Решение

Общее число возможных элементарных исходов испытания равно числу способов, которыми можно извлечь шесть деталей из 10, то есть C_{10}^{6}.

Определим число исходов, благоприятствующих интересующему нас событию A (среди шести взятых деталей четыре стандартные). Четыре стандартные детали можно взять из семи стандартных деталей C_7^4 способами; при этом остальные $6-4=2$ детали должны быть нестандартными; взять же две нестандартные детали из $10-7=3$ нестандартных деталей можно C_3^2 способами. Следовательно, число благоприятствующих исходов равно $C_7^4\cdot C_3^2$.

Искомая вероятность равна отношению числа исходов, благоприятствующих событию, к числу всех элементарных исходов:

$$P(A)=\frac{C_7^4\cdot C_3^2}{C_{10}^{6}}=\frac{1}{2}.$$

Задачи для самостоятельного решения

1 Четырёхтомное собрание сочинений расположено на полке в случайном порядке. Чему равна вероятность того, что тома стоят в должном порядке справа налево или слева направо? *Ответ*: 0,083.

2 Студент пришёл на экзамен, зная лишь 20 из 25 вопросов программы. Экзаменатор задал студенту три вопроса. Найти вероятность того, что студент знает все три вопроса. *Ответ*: 0,496.

3 Имеется шесть отрезков, длины которых равны соответственно 2, 4, 6, 8, 10, 12 единицам. Определить вероятность того, что с помощью взятых наугад трёх отрезков из данных шести можно построить треугольник. *Ответ*: 0,35.

4 В конверте среди 100 фотокарточек находится одна разыскиваемая. Из конверта наудачу извлечены 10 карточек. Найти вероятность того, что среди них окажется нужная. *Ответ*: 0,1.

5 Из десяти изделий два бракованных. Определить вероятность того, что среди пяти взятых для проверки изделий окажется хотя бы одно бракованное. *Ответ*: 0,778.

Геометрическое определение вероятности.

Классическое определение неприменимо 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.

В случае пространства R^1 мерой является длина соответствующих отрезков, для пространства R^2 – площадь областей, для пространства R^3 – объём соответствующих тел.

Пример 3 – Два студента условились встретиться в определённом месте между 12 и 13 часами дня. Пришедший первым ждёт второго в течение 1/4 часа, после чего уходит. Найти вероятность того, что встреча состоится, если каждый студент наудачу выбирает момент своего прихода в течение указанного часа.

Решение

Пусть *x* – момент прихода первого студента, *y* – момент прихода второго студента. Тогда рассмотрим области на плоскости (рисунок 1):

$$\Omega=\left\{(x,y)\in R^2: 0\le x\le 1,\quad 0\le y\le 1\right\},$$

$$A=\left\{(x,y)\in R^2: 0\le x\le 1,\quad 0\le y\le 1,\quad |x-y|\le\frac{1}{4}\right\}.$$

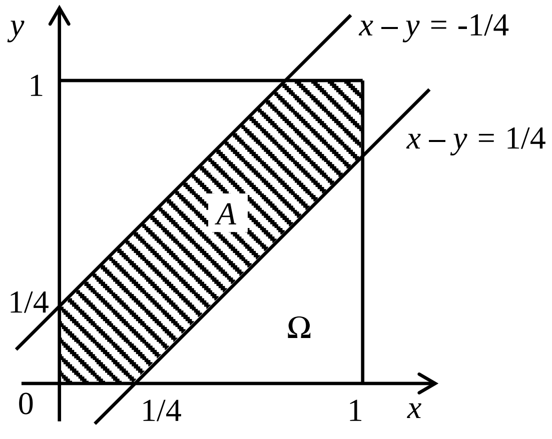

Рисунок 1

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

Задачи для самостоятельного решения

1 Найти вероятность того, что сумма двух наугад взятых положительных правильных дробей не больше единицы, а их произведение не больше 3/16. *Ответ*: 0,456.

2 Задача Бюффона.

На плоскость, разграфленную параллельными прямыми линиями, отстоящими друг от друга на расстояние $2a$, наудачу бросается игла длиной $2l$. Какова вероятность того, что игла пересечёт одну из параллельных прямых, если $l \le a$? *Ответ*: $2l/\pi a$.

3 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. *Ответ*: 0,121.

3 Сложение и умножение вероятностей. Полная (средняя) вероятность. Формула Байеса

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

Теорема 2. *Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления:* $P(A+B)=P(A)+P(B)-P(AB)$.

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, вычисленная при условии, что имело место событие A, называется ***условной вероятностью*** события B: $P_A(B)=P(B\,|\,A)=\dfrac{P(AB)}{P(A)}$.

Теорема 3 (умножения вероятностей). *Вероятность произведения двух событий (совместного появления этих событий) равна произведению вероятности одного из них на условную вероятность другого, вычисленную при условии, что первое событие уже наступило:* $P(AB)=P(A)\cdot P(B\,|\,A)$.

Теорема 4. *Полная вероятность события А, которое может произойти вместе с одним из событий* $H_1, H_2, ..., H_n$, *равна сумме парных произведений вероятностей каждого из этих событий на соответствующие им условные вероятности наступления события А:* $P(A)=\sum_{i=1}^{n} P(H_i)\cdot P(A\,|\,H_i)$.

Теорема 5. *Вероятность гипотезы после испытания равна произведению вероятности гипотезы до испытания на соответствующую ей условную вероятность события, которое произошло при испытании, делённому на полную вероятность этого события:*

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

Пример 1 – Два стрелка стреляют по мишени. Вероятность попадания в мишень при одном выстреле для первого стрелка равна 0,7, а для второго – 0,8. Найти вероятность того, что при одном залпе в мишень попадает только один из стрелков.

Решение

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

$$P(A) = 0,7; \quad P(\overline{A}) = 0,3; \quad P(B) = 0,8; \quad P(\overline{B}) = 0,2.$$

Вероятность того, что первый стрелок попадёт в мишень, а второй – нет, $P(A)P(\overline{B}) = 0,7\cdot 0,2 = 0,14$.

Вероятность того, что второй стрелок попадёт в цель, а первый – нет, $P(\overline{A})P(B) = 0,3\cdot 0,8 = 0,24$.

Тогда вероятность попадания в цель только одним стрелком $P = 0,14 + 0,24 = 0,38$.

Пример 2 – Вероятности того, что нужная деталь находится в первом, втором, третьем или четвёртом ящике, соответственно $p_1 = 0,6$, $p_2 = 0,7$, $p_3 = 0,8$, $p_4 = 0,9$. Найти вероятность того, что эта деталь находится только в одном ящике.

Решение

Применим теоремы 1 и 4, приняв $q_i = 1 - p_i$, $i = \overline{1,4}$:

$$P = p_1q_2q_3q_4 + q_1p_2q_3q_4 + q_1q_2p_3q_4 + q_1q_2q_3p_4,$$

$$P = 0,6\cdot 0,3\cdot 0,2\cdot 0,1 + 0,4\cdot 0,7\cdot 0,2\cdot 0,1 + 0,4\cdot 0,3\cdot 0,8\cdot 0,1 + 0,4\cdot 0,3\cdot 0,2\cdot 0,9 =$$
$$= 0,0036 + 0,0056 + 0,0096 + 0,0216 = 0,0404.$$

Пример 3 – Известно, что 5 % всех мужчин и 0,25 % всех женщин – дальтоники. На обследование прибыло одинаковое число мужчин и женщин. Наудачу выбранное лицо оказалось дальтоником. Какова вероятность, что это мужчина?

Решение

Рассмотрим событие A – наудачу выбранное лицо оказалось дальтоником.

По условию задачи возможны следующие две гипотезы:

1) H_1 – выбранное для обследования лицо оказалось мужчиной;

2) H_2 – выбранное для обследования лицо оказалось женщиной.

Так как на обследование прибыло одинаковое число мужчин и женщин, то вероятности наступления гипотез будут равны между собой, т. е. $P(H_1) = P(H_2) = 0{,}5$.

Вероятности быть дальтоником для мужчины и женщины известны из условия задачи. Они соответственно $P(A|H_1) = 0{,}05$, $P(A|H_2) = 0{,}0025$.

Для отыскания вероятности события A воспользуемся формулой полной вероятности. Имеем

$$P(A) = 0{,}5 \cdot 0{,}05 + 0{,}5 \cdot 0{,}0025 = 0{,}025 + 0{,}00125 = 0{,}02625.$$

По условию задачи необходимо найти вероятность события, состоящего в том, что дальтоником окажется мужчина, т. е. $P(H_1 \mid A)$. Эту вероятность вычислим по формулам Байеса. Таким образом, получаем

$$P(H_1 \mid A) = \frac{0{,}5 \cdot 0{,}05}{0{,}02625} \approx 0{,}95.$$

Задачи для самостоятельного решения

1 Баскетболист выполняет два штрафных броска. Вероятность попадания при каждом броске равна 0,8. Найти вероятность:

а) двух попаданий; б) одного попадания; в) хотя бы одного попадания.

Ответ: а) 0,64; б) 0,32; в) 0,96.

2 Имеется коробка с девятью новыми теннисными мячами. Для игры берут три мяча; после игры их кладут обратно. При выборе мячей игранные от неигранных не отличают. Какова вероятность того, что после трёх игр в коробке не останется неигранных мячей? *Ответ*: 0,003.

3 В ящике 10 деталей, среди которых две нестандартные. Найти вероятность того, что в наудачу отобранных шести деталях окажется не более одной нестандартной. *Ответ*: 0,667.

4 Отдел технического контроля проверяет изделия на стандартность. Вероятность того, что изделие нестандартно, равна 0,1. Найти вероятность того, что:

а) из трёх проверенных изделий только два окажутся нестандартными;

б) нестандартным окажется только четвёртое по порядку проверенное изделие.

Ответ: а) 0,027; б) 0,073.

5 33 буквы русского алфавита написаны на карточках разрезной азбуки:

а) три карточки вынимают наугад одну за другой и укладывают на стол в порядке появления. Найти вероятность того, что получится слово «мир»;

б) вытянутые три карточки можно поменять местами произвольным образом. Какова вероятность того, что из вытянутых трёх карточек можно сложить слово «мир»?

Ответ: а) 0,00003; б) 0,00018.

6 В первой коробке содержится 20 микросхем, из них 18 исправных; во второй коробке – 10 микросхем, из них девять исправных. Из второй коробки наудачу взята микросхема и переложена в первую. Найти вероятность того, что:

а) наугад извлечённая из первой коробки микросхема исправна;

б) из второй коробки переложили исправную микросхему, если из первой извлекли исправную.

Ответ: а) 0,9; б) 0,905.

4 Повторение испытаний

Пусть в результате n независимых испытаний, проведённых в одинаковых условиях, событие A наступает с вероятностью $P(A)=p$, а противоположное ему событие $\overline{A}$ – с вероятностью $P(\overline{A})=1-p=q$.

Вероятность $P_n(k)$ того, что в результате n испытаний событие A наступило ровно k раз, определяется по ***формуле Бернулли*** $P_n(k)=\dfrac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$.

Формула Пуассона. Если число испытаний n велико, а вероятность p появления события A в каждом испытании очень мала, то используют приближённую формулу $P_n(k)\approx\dfrac{\lambda^k\cdot e^{-\lambda}}{k!}$, где $\lambda=np$ – среднее число появления события A.

Локальная теорема Муавра – Лапласа. Вероятность того, что в n независимых испытаниях, в каждом из которых вероятность появления события A равна p ($0<p<1$ – вероятность далека от 0 или 1), событие A наступит k раз, равна (тем точнее, чем больше n) $P_n(k)\approx\dfrac{1}{\sqrt{npq}}\cdot\varphi(x)$.

Здесь $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}\cdot e^{-\frac{x^2}{2}}$ – функция Гаусса; $x=\dfrac{k-np}{\sqrt{npq}}$.

Существует таблица значений функции $\varphi(x)$ для $0\le x<4$; $\varphi(x\ge)\approx 0$; функция $\varphi(x)$ – чётная ($\varphi(-x)=\varphi(x)$).

Интегральная теорема Лапласа. Вероятность того, что в n независимых испытаниях, в каждом из которых вероятность появления события A равна p $(0 < p < 1)$, событие A наступит не менее k_1 и не более k_2 раз, приближённо равна $P_n(k_1, k_2) \approx \Phi(x_2) - \Phi(x_1)$, где $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{x^2}{2}} dx$ – функция Лапласа; $x_1 = \dfrac{k_1 - np}{\sqrt{npq}}$; $x_2 = \dfrac{k_2 - np}{\sqrt{npq}}$. Значения функции Лапласа вычисляют по таблице.

Пример 1 – По цели производится пять выстрелов. Вероятность попадания для каждого выстрела равна 0,4. Найти вероятность того, что в цель попали не менее трёх раз.

Решение

Вероятность не менее трёх попаданий складывается из вероятности пяти, четырёх и трёх попаданий. Так как выстрелы независимы, то можно применить формулу Бернулли:

$$P_5(5) = p^5 = 0,4^5 = 0,01024; \; P_5(4) = \frac{5!}{4! \cdot 1!} p^4 (1-p) = 0,0768;$$

$$P_5(3) = \frac{5!}{3! \cdot 2!} p^3 (1-p)^2 = 0,2304.$$

Окончательно получаем вероятность не менее трёх попаданий из пяти выстрелов $P = 0,01204 + 0,0768 + 0,2304 = 0,31744$.

Пример 2 – При въезде в новый дом в осветительную сеть было включено 400 новых электроламп. Каждая лампа в течение года перегорит с вероятностью 0,1. Найти вероятность того, что в течение года не менее 50 первоначально включённых ламп придётся заменить новыми.

Решение

Воспользуемся интегральной формулой Лапласа.

Событие A – электролампа в течение года перегорит.

$P(A) = p = 0,1$; $n = 400$; $k_1 = 50$; $k_2 = n = 400$; $q = 1 - p = 1 - 0,1 = 0,9$.

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = \frac{50 - 400 \cdot 0,1}{\sqrt{400 \cdot 0,1 \cdot 0,9}} = \frac{10}{20 \cdot 0,3} = \frac{10}{6} \approx 1,67;$$

$$x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{400 - 400 \cdot 0,1}{\sqrt{400 \cdot 0,1 \cdot 0,9}} = \frac{360}{6} = 60;$$

$$P_{400}(50;400) \approx \Phi(x_2) - \Phi(x_1) \approx \Phi(60) - \Phi(1{,}67) \approx 0{,}5 - 0{,}4525 = 0{,}0475.$$

Задачи для самостоятельной работы

1 Магазин получил 1000 бутылок минеральной воды. Вероятность того, что при перевозке бутылка окажется разбитой, равна 0,003. Найти вероятность того, что магазин получит разбитых бутылок:

а) ровно две; б) менее двух; в) более двух; г) хотя бы одну.

Ответ: а) 0,224; б) 0,199; в) 0,577; г) 0,95.

2 В мастерской имеется 10 моторов. При существующем режиме работы вероятность того, что мотор в данный момент работает с полной нагрузкой, равна 0,8. Найти вероятность того, что в данный момент не менее восьми моторов работают с полной нагрузкой. *Ответ*: 0,636.

3 Всхожесть семян пшеницы составляет 80 %. Найти вероятность того, что из 100 посеянных семян число взошедших окажется:

а) равным 80; б) заключённым между 80 и 90.

Ответ: а) 0,1; б) 0,4938.

4 Коммутатор учреждения обслуживает 600 абонентов. Вероятность того, что в течение одной минуты абонент позвонит на коммутатор, равна 0,005. Найти вероятность того, что в течение одной минуты позвонят:

а) три абонента; б) хотя бы один абонент; в) никто не позвонит.

Ответ: а) 0,224; б) 0,95; в) 0,05.

5 Вероятность выигрыша по лотерейному билету равна 0,2. Куплено 250 билетов. Найти вероятность того, что выигрышными окажутся:

а) ровно 60 билетов; б) не более 60 билетов.

Ответ: а) 0,018; б) 0,9431.

5 Дискретные случайные величины

Дискретной случайной величиной называется такая величина, которая в результате опыта может принимать определённые значения с определённой вероятностью, образующие счётное множество.

Соотношение между возможными значениями случайной величины и их вероятностями называется ***законом распределения*** дискретной случайной величины. Закон распределения может быть задан аналитически, в виде таблицы или графически.

Таблица соответствия значений случайной величины и их вероятностей называется ***рядом распределения***.

Графическое представление этой таблицы называется ***многоугольником распределения.***

Функцией распределения называют функцию $F(x)$, определяющую вероятность того, что случайная величина X в результате испытания примет значение, меньшее x: $F(x) = P(X < x)$.

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

Дисперсией (рассеиванием) дискретной случайной величины называется математическое ожидание квадрата отклонения случайной величины от её математического ожидания: $D(X) = M\left[X - M(X)\right]^2$ или $D(X) = M\left(X^2\right) - \left[M(X)\right]^2$.

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

$\alpha_k = M\left(X^k\right)$ – ***начальный момент порядка k***. В частности, $M(X) = \alpha_1$.

$\mu_k = M\left(\left[X - M(X)\right]^k\right)$ – ***центральный момент порядка k***, $D(X) = \mu_2$.

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

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

Пример 1 – По цели производится пять выстрелов. Вероятность попадания для каждого выстрела равна 0,4. Найти вероятности числа попаданий и построить многоугольник распределения.

Решение

Вероятности пяти попаданий из пяти возможных, четырёх из пяти и трёх из пяти были найдены в примере 1 пункта 4 и равны соответственно

$$P_5(5) = 0{,}01024, \quad P_5(4) = 0{,}0768, \quad P_5(3) = 0{,}2304.$$

Аналогично найдём следующее:

$$P_5(2) = \frac{5!}{2!\cdot 3!} 0{,}4^2 \cdot 0{,}6^3 = 0{,}3456; \; P_5(1) = \frac{5!}{1!\cdot 4!} 0{,}4^1 \cdot 0{,}6^4 = 0{,}2592;$$

$$P_5(0) = \frac{5!}{0!\cdot 5!} 0{,}4^0 \cdot 0{,}6^5 = 0{,}6^5 = 0{,}0778.$$

Представим графически зависимость числа попаданий от их вероятностей (рисунок 2).

Пример 2 – Испытывается устройство, состоящее из четырёх независимо работающих приборов. Вероятности отказа каждого из приборов равны соответственно $p_1 = 0{,}3$, $p_2 = 0{,}4$, $p_3 = 0{,}5$, $p_4 = 0{,}6$. Найти математическое ожидание и дисперсию числа отказавших приборов.

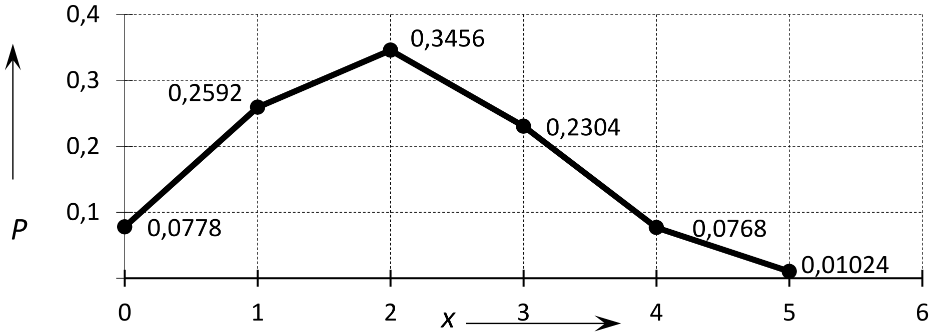

Рисунок 2

Решение

Принимая за случайную величину число отказавших приборов, видим, что эта случайная величина может принимать значения 0, 1, 2, 3 или 4.

$$p(0) = q_1 q_2 q_3 q_4 = 0,7 \cdot 0,4 \cdot 0,5 \cdot 0,4 = 0,084.$$

$$p(1) = p_1 q_2 q_3 q_4 + q_1 p_2 q_3 q_4 + q_1 q_2 p_3 q_4 + q_1 q_2 q_3 p_4 = 0,302\,.$$

$$p(2) = p_1 p_2 q_3 q_4 + p_1 q_2 p_3 q_4 + p_1 q_2 q_3 p_4 + q_1 p_2 p_3 q_4 + q_1 p_2 q_3 p_4 + q_1 q_2 p_3 p_4 = 0,38.$$

$$p(3) = p_1 p_2 p_3 q_4 + p_1 p_2 q_3 p_4 + p_1 q_2 p_3 p_4 + q_1 p_2 p_3 p_4 = 0,198.$$

$$p(4) = p_1 p_2 p_3 p_4 = 0,036.$$

Получаем следующий закон распределения (таблица 1).

Таблица 1

x	0	1	2	3	4
p	0,084	0,302	0,38	0,198	0,036

$$M(X) = 0,302 + 2 \cdot 0,38 + 3 \cdot 0,198 + 4 \cdot 0,036 = 1,8.$$

$$M(X^2) = 0,302 + 4 \cdot 0,38 + 9 \cdot 0,198 + 16 \cdot 0,036 = 4,18.$$

$$D(X) = M(X^2) - \left[M(X)\right]^2 = 4,18 - 3,24 = 0,94.$$

Задачи для самостоятельного решения

1 Охотник, имеющий пять патронов, стреляет в цель до первого попадания (или пока не израсходует все патроны). Найти математическое ожидание и дисперсию СВ X – число израсходованных патронов. Построить график функции распределения этой СВ, если вероятность попадания при каждом выстреле равна 0,4. *Ответ*: $M(X)=2,31$, $D(X)=1,96$.

2 Два стрелка стреляют по одной мишени, делая независимо друг от друга по два выстрела. Вероятность попадания в мишень для первого стрелка – 0,5, для второго – 0,6. Построить ряд распределения случайной величины X – общего числа попадания и найти её числовые характеристики $M(X)$ и $D(X)$. *Ответ*: $M(X)=2,2$, $D(X)=0,98$.

3 В партии 10 % нестандартных деталей. Наудачу отобраны четыре детали. Составить закон распределения ДСВ X – числа нестандартных деталей среди четырёх отобранных (табличный, графический).

4 Рабочий обслуживает четыре станка. Вероятность того, что в течение часа станок не потребует внимания рабочего, для первого станка равна 0,7, для второго – 0,75, для третьего – 0,8, для четвёртого – 0,9. Найти математическое ожидание и дисперсию СВ X – числа станков, которые не потребуют внимания рабочего в течение часа. *Ответ*: $M(X)=3,15$, $D(X)=0,65$.

5 Из урны, содержащей три белых и пять черных шаров, наугад извлекают три шара. Пусть СВ X – число вынутых черных шаров. Построить ряд распределения СВ X и найти её математическое ожидание. *Ответ*: $M(X)=1,875$.

6 Имеется шесть ключей, из которых только один подходит к замку. Составить ряд распределения СВ X – числа попыток при открывании замка, если ключ, не подошедший к замку, в последующих опробованиях не участвует. Чему равны математическое ожидание и дисперсия этой случайной величины? *Ответ*: $M(X)=3,5$, $D(X)=2,92$.

7 Найти закон распределения ДСВ X, которая может принимать только два значения: x_1 – с вероятностью $p_1=0,4$ и x_2 (причём $x_2<x_1$), если известны математическое ожидание $M(X)=3,2$ и дисперсия $D(X)=0,96$.

6 Непрерывные случайные величины

Непрерывной случайной величиной называется величина, которая может принимать любые значения из некоторого конечного или бесконечного промежутка.

Плотностью распределения вероятностей непрерывной случайной величины X называется функция $f(x)$ – первая производная от функции распределения $F(x)$: $f(x)=F'(x)$.

Вероятность того, что непрерывная случайная величина X примет значение, принадлежащее интервалу (a, b), равна определённому интегралу от плотности распределения, взятому в пределах от a до b: $P(a<X<b)=\int_a^b f(x)dx$.

Функция распределения может быть найдена по формуле $F(x)=\int\limits_{-\infty}^{x} f(x)dx$.

Несобственный интеграл от плотности распределения в пределах от $-\infty$ до $+\infty$ равен единице: $\int\limits_{-\infty}^{\infty} f(x)dx=1.$

Математическим ожиданием непрерывной случайной величины X называется интеграл $M(X)=\int\limits_{-\infty}^{\infty} xf(x)dx$.

Дисперсией непрерывной случайной величины называется математическое ожидание квадрата её отклонения $D(X)=\int\limits_{-\infty}^{\infty}[x-M(X)]^2 f(x)dx$ или $D(X)=\int\limits_{-\infty}^{\infty} x^2 f(x)dx-[M(X)]^2$.

Остальные характеристики находятся аналогично дискретным СВ.

Пример 1 **–** Непрерывная случайная величина задана плотностью распределения

$$f(x)=\begin{cases} A\cos 2x & \text{при} \quad -\dfrac{\pi}{4}\le x\le\dfrac{\pi}{4}; \\ 0 & \text{при} \quad |x|>\dfrac{\pi}{4}. \end{cases}$$

Требуется найти коэффициент A, функцию распределения, определить вероятность того, что случайная величина x попадёт в интервал $\left(\dfrac{\pi}{6};2\right)$.

Решение

Найдём коэффициент A:

$$\int\limits_{-\infty}^{\infty} f(x)dx=\int\limits_{-\infty}^{-\pi/4} 0dx+\int\limits_{-\pi/4}^{\pi/4} A\cos 2xdx+\int\limits_{\pi/4}^{\infty} 0dx=\left.\frac{A\sin 2x}{2}\right|_{-\pi/4}^{\pi/4}=A=1.$$

Тогда $f(x)=\begin{cases} \cos 2x & \text{при} \quad -\dfrac{\pi}{4}\le x\le\dfrac{\pi}{4}; \\ 0 & \text{при} \quad |x|>\dfrac{\pi}{4}. \end{cases}$

Найдём функцию распределения:

1) на участке $x<-\dfrac{\pi}{4}$: $\quad F(x)=\int\limits_{-\infty}^{x} f(x)dx=\int\limits_{-\infty}^{x}0dx=0$;

2) на участке $-\dfrac{\pi}{4}\le x\le\dfrac{\pi}{4}$: $F(x)=\int\limits_{-\infty}^{-\pi/4}0dx+\int\limits_{-\pi/4}^{x}\cos 2xdx=\left.\dfrac{\sin 2x}{2}\right|_{-\pi/4}^{x}=\dfrac{\sin 2x}{2}+\dfrac{1}{2}$;

3) на участке $x>\dfrac{\pi}{4}$: $F(x)=\int\limits_{-\infty}^{-\pi/4}0dx+\int\limits_{-\pi/4}^{\pi/4}\cos 2xdx+\int\limits_{\pi/4}^{x}0dx=\left.\dfrac{\sin 2x}{2}\right|_{-\pi/4}^{\pi/4}=1.$

Итого

$$F(x)=\begin{cases}0 & \text{при} \quad x<-\dfrac{\pi}{4};\\ \dfrac{\sin 2x+1}{2} & \text{при} \quad -\dfrac{\pi}{4}\le x\le\dfrac{\pi}{4};\\ 1 & \text{при} \quad x>\dfrac{\pi}{4}.\end{cases}$$

Найдём вероятность попадания случайной величины в интервал $\left(\dfrac{\pi}{6};2\right)$:

$$P\left(\frac{\pi}{6}<x<2\right)=\int\limits_{\pi/6}^{2}f(x)dx=\int\limits_{\pi/6}^{\pi/4}\cos 2xdx+\int\limits_{\pi/4}^{2}0dx=\left.\frac{\sin 2x}{2}\right|_{\pi/6}^{\pi/4}=\frac{1}{2}-\frac{\sqrt{3}}{4}=0{,}067\,.$$

Ту же самую вероятность можно искать и другим способом:

$$P\left(\frac{\pi}{6}<x<2\right)=F(2)-F(\pi/6)=1-\frac{\sin(\pi/3)+1}{2}=\frac{1}{2}-\frac{\sqrt{3}}{4}=0{,}067.$$

Задачи для самостоятельного решения

1 По известной модели распределения для СВ *X*:

– найти значение константы «*C*»;

– найти $F(x)$ и $f(x)$;

– построить графики $F(x)$ и $f(x)$;

– вычислить $M(X)$, $D(X)$, асимметрию *A*, эксцесс *E;*

– найти $P(a\le X\le b)$,

если СВ задана следующим образом:

а) $f(x)=\begin{cases}C, & \text{если } x\in[1,7];\\ 0, & \text{если } x\notin[1,7];\end{cases}$ $\qquad a=2,\ b=5;$

б) $f(x)=\begin{cases} Cx^2, & \text{если } 0\le x\le 10, \\ 0, & \text{если } x\le 0, x>10. \end{cases}$ $a = 1, b = 4;$

в) $F(x)=\begin{cases} 1-\dfrac{C}{x^2}, & \text{если } x\ge 1; \\ 0, & \text{если } x<1; \end{cases}$ $a = 1, b = 3;$

г) $F(x)=\begin{cases} 0, & \text{если } x<0; \\ Cx^2, & \text{если } 0\le x<2; \\ x-\dfrac{7}{4}, & \text{если } 2\le x<\dfrac{11}{4}; \\ 1, & \text{если } x\ge\dfrac{11}{4}. \end{cases}$ $a = 0, b = 3.$

7 Основные законы распределения случайных величин

Биноминальное распределение ДСВ X – числа успехов в схеме Бернулли при проведении n независимых испытаний с вероятностью успеха p:

$$P_n(k)=C_n^k p^k q^{n-k}, \qquad k=0,1,2,...,n;$$

$$MX=np;\ DX=npq.$$

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

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

$$MX=DX=\lambda.$$

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:

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

Тогда $F(x)=\begin{cases}0 & \text{при} \quad x<a; \\ \dfrac{x-a}{b-a} & \text{при} \quad a\le x\le b; \\ 1 & \text{при} \quad x>b;\end{cases}$ $MX=\dfrac{a+b}{2}$; $DX=\dfrac{(b-a)^2}{12}$.

Показательным (экспоненциальным) называется распределение вероятностей непрерывной СВ *X*, которое описывается плотностью

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

где λ – положительное число.

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

Нормальным называется распределение вероятностей непрерывной СВ *X*, которое описывается плотностью вероятности

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

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

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

Вероятность попадания нормальной СВ *X* в интервал $(\alpha;\beta)$ $P(\alpha\le X\le\beta)=\Phi\left(\dfrac{\beta-m_x}{\sigma}\right)-\Phi\left(\dfrac{\alpha-m_x}{\sigma}\right)$.

Вероятность отклонения СВ от среднего $P(|X-m_x|<\varepsilon)=2\Phi\left(\dfrac{\varepsilon}{\sigma}\right)$ $(\varepsilon>0)$.

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

Пример 1 **–** Цена деления шкалы измерительного прибора равна 0,2. Показания прибора округляются до ближайшего деления. Найти вероятность того, что при отсчёте будет сделана ошибка менее 0,04.

Решение

Ошибку округления можно рассматривать как случайную величину *X*, которая имеет равномерное распределение в промежутке между двумя соседними делениями, следовательно, её плотность вероятности имеет вид:

$$f(x)=\begin{cases}5 \quad \text{при } 0<x\le 0,2;\\ 0 \quad \text{при } x<0 \text{ или } x>0,2.\end{cases}$$

Ошибка отсчёта будет менее 0,04, если случайная величина заключена в интервале (0; 0,4) или (0,16; 0,2). Следовательно, получаем, что вероятность того, что при отсчёте будет сделана ошибка меньше 0,04, $p=0,4$.

Пример 2 **–** Время безотказной работы прибора распределено по показательному закону $f(t)=\lambda\cdot e^{-\lambda\cdot t}$ при $t\ge 0$ (*t* – время). Найти вероятность того, что прибор проработает безотказно 100 ч, если в среднем данные приборы работают безотказно 50 ч.

Решение

Из условия следует, что $\frac{1}{\lambda}=M(X)=50$. Тогда $\lambda=0,02$. В результате получаем, что $P(X\ge 100)=1-F(100)=e^{-2}=0,13534$.

Итак, вероятность безотказной работы прибора в течение 100 ч равна 0,14.

Пример 3 – Рост мужчин некоторой возрастной группы распределён по нормальному закону с математическим ожиданием *a* = 165 см и средним квадратичным отклонением $\sigma=5$ см. Какую долю костюмов 3-го роста следует предусмотреть в общем объёме производства для данной возрастной группы (3-й рост – 170–176 см)?

Решение

X – рост в сантиметрах представителя данной возрастной группы. Тогда доля *P* мужчин ростом 170–176 см

$$P=\Phi\left(\frac{176-165}{5}\right)-\Phi\left(\frac{170-165}{5}\right)=\Phi(2,2)-\Phi(1)=0,1448.$$

Значит, доля костюмов 3-го роста должна составлять приблизительно 14,5 % общего числа.

Пример 4 – Определить среднеквадратичную ошибку прибора, если систематических ошибок он не имеет, а случайные ошибки X распределены по нормальному закону и с вероятностью 0,95 не выходят за пределы ± 20 мк.

Решение

Из условия задачи следует, что $P(|x| \le 20) = 0{,}95$. С другой стороны, т. к. $a = M(X) = 0$, то $P(|x| \le 20) = P(|x - a| \le 20) = 2\Phi\left(\frac{20}{\sigma}\right)$. Таким образом, $\Phi\left(\frac{20}{\sigma}\right) = \frac{0{,}95}{2} = 0{,}475$. Воспользовавшись таблицей функции $\Phi(x)$, находим, что $\Phi(x) = 0{,}475$ при $x = 1{,}96$, следовательно, $\frac{20}{\sigma} = 1{,}96$, откуда $\sigma = \frac{20}{1{,}96} \approx 10{,}2$ мк.

Задачи для самостоятельного решения

1 Троллейбусы некоторого маршрута идут строго по расписанию. Интервал движения составляет 9 мин. Найти вероятность того, что пассажир, подошедший к остановке, будет ожидать очередного троллейбуса менее 3 мин. *Ответ*: 1/3.

2 Случайная величина X равномерно распределена на отрезке [–3; 2]. Найти функцию распределения $F(x)$ этой случайной величины.

3 Среднее время работы каждого из трёх элементов технического устройства равно 750 ч. Для безотказной работы устройства необходима безотказная работа хотя бы одного из трёх этих элементов. Определить вероятность того, что устройство будет работать от 450 до 600 ч, если время T работы каждого из трёх элементов независимо и распределено по показательному закону. *Ответ*: 0,27.

4 Случайная величина X – ошибка измерительного прибора – распределена по нормальному закону со средним квадратичным отклонением 3 мм. Систематическая ошибка прибора отсутствует. Найти вероятность того, что в двух независимых измерениях ошибка хотя бы один раз окажется в интервале (0; 2,4). *Ответ*: 0,493.

5 Случайная величина X подчинена нормальному закону распределения с математическим ожиданием $M(X) = 2$ и дисперсией $D(X) = 9$. Написать выражения для плотности вероятности и функции распределения этой случайной величины.

6 Диаметр валиков есть нормальная случайная величина X с параметрами $m_x = 10$ мм и $\sigma = 0{,}1$ мм. Найти интервал, в котором с вероятностью 0,9973 будут заключены диаметры изготовленных валиков. *Ответ*: $(9{,}7; 10{,}3)$.

8 Векторные случайные величины. Числовые характеристики векторных случайных величин

Случайную величину (X,Y), возможные значения которой есть пары чисел (x,y), называют ***системой двух случайных величин*** или ***двумерной случайной величиной***. Одномерные величины X и Y – составляющие или компоненты.

Законом распределения дискретной двумерной СВ называют перечень возможных значений этой величины, т. е. пар чисел $\left(x_i, y_j\right)$ и их вероятностей $p\left(x_i, y_j\right)$ $\left(i=1,2,...,n;\ j=1,2,...,m\right)$. Обычно закон распределения задают в виде таблицы с двумя входами, называемой ***матрицей распределения***. Суммируя вероятности таблицы по столбцам или по строкам, получаем законы распределения составляющих.

Двумерную СВ (дискретную или непрерывную) можно задать с помощью ***функции распределения*** $F(x,y)$: $F\left(x,y\right)=P\left(X<x, Y<y\right)$.

Свойства функции распределения:

1) $0\le F\left(x,y\right)\le 1$;

2) $F\left(-\infty;y\right)=F\left(-\infty;\infty\right)=F\left(x;-\infty\right)=0$; $F\left(+\infty;+\infty\right)=1$;

3) $F\left(x,+\infty\right)=F_1\left(x\right)$, где $F_1\left(x\right)$ – функция распределения составляющей X;
$F\left(+\infty;y\right)=F_2\left(y\right)$, где $F_2\left(y\right)$ – функция распределения составляющей Y.

Непрерывную двумерную СВ можно также задать ***двумерной плотностью вероятности*** $f\left(x,y\right)=\dfrac{\partial^2 F\left(x,y\right)}{\partial x\partial y}$.

Свойства плотности вероятности:

1) $f\left(x,y\right)\ge 0$;

2) $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f\left(x,y\right)dxdy=1$;

3) плотности распределения составляющих $f_1\left(x\right)=\int\limits_{-\infty}^{+\infty} f\left(x,y\right)dy$; $f_2\left(y\right)=\int\limits_{-\infty}^{+\infty} f\left(x,y\right)dx$;

4) вероятность попадания случайной точки (X,Y) в область D определяется по формуле $P\left(\left(X,Y\right)\in D\right)=\iint\limits_D f\left(x,y\right)dxdy$.

Математическое ожидание и ***дисперсия*** системы есть совокупность составляющих. Для дискретной двумерной СВ

$$M\left(X\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i p_{ij}=\sum_{i=1}^{n}\left(x_i\sum_{j=1}^{m}p_{ij}\right);\quad M\left(Y\right)=\sum_{i=1}^{n}\sum_{j=1}^{m}y_j p_{ij}=\sum_{j=1}^{m}\left(y_j\sum_{i=1}^{n}p_{ij}\right);$$

$$D(X)=\sum_{i=1}^{n}\sum_{j=1}^{m}\left(x_i-M(X)\right)^2 p_{ij}=\sum_{i=1}^{n}\left(x_i^2\sum_{j=1}^{m}p_{ij}\right)-\left(M(X)\right)^2;$$

$$D(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_j-M(Y)\right)^2 p_{ij}=\sum_{j=1}^{n}\left(y_j^2\sum_{i=1}^{n}p_{ij}\right)-\left(M(Y)\right)^2.$$

Если система двух непрерывных СВ задана плотностью $f(x,y)$, то

$$M(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy;\quad M(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy;$$

$$D(X)=\int_{-\infty}^{+\infty}\left(x-M(X)\right)^2 f(x,y)dxdy=\int_{-\infty}^{+\infty}x^2 f(x,y)dxdy-\left(M(X)\right)^2;$$

$$D(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(y-M(Y)\right)^2 f(x,y)dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y^2 f(x,y)dxdy-\left(M(Y)\right)^2.$$

Корреляционным моментом μ_{XY} случайных величин X и Y называют математическое ожидание произведения отклонений этих случайных величин:

$$\mu_{XY}=M\left(\left(X-M(X)\right)\left(Y-M(Y)\right)\right)=M(XY)-M(X)\cdot M(Y).$$

Для вычисления корреляционного момента дискретных величин используют формулу

$$\mu_{XY}=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i y_j p_{ij}-M(X)\cdot M(Y),$$

а для непрерывных величин – формулу

$$\mu_{XY}=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(x-M(X)\right)\left(y-M(Y)\right)f(x,y)dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy-M(X)M(Y).$$

Коэффициентом корреляции r_{XY} величин X и Y называют отношение корреляционного момента к произведению средних квадратических отклонений этих величин:

$$r_{XY}=\frac{\mu_{XY}}{\sigma_X\cdot\sigma_Y};\qquad \left|r_{XY}\right|\le 1.$$

Пример 1 – Матрица распределения системы двух дискретных случайных величин (X,Y) задана таблицей 2.

Таблица 2

X	Y		
	0	2	5
1	0,1	0,1	0,2
2	0,2	0,3	0,1

Найти математическое ожидание, дисперсию, корреляционный момент и коэффициент корреляции системы (X,Y).

Решение

Находим математические ожидания составляющих X и Y:

$$M(X)=\sum_{i=1}^{2}\sum_{j=1}^{3}x_i p_{ij}=\sum_{i=1}^{2}\left(x_i\sum_{j=1}^{3}p_{ij}\right)=x_1\left(p_{11}+p_{12}+p_{13}\right)+$$

$$+x_2\left(p_{21}+p_{22}+p_{23}\right)=1\cdot\left(0,1+0,1+0,2\right)+2\cdot\left(0,2+0,3+0,1\right)=1,6,$$

$$M(Y)=\sum_{i=1}^{2}\sum_{j=1}^{3}y_j p_{ij}=\sum_{j=1}^{3}y_j\sum_{i=1}^{2}p_{ij}=y_1\left(p_{11}+p_{21}\right)+y_2\left(p_{12}+p_{22}\right)+y_3\left(p_{13}+p_{23}\right)=$$

$$=0\cdot\left(0,1+0,2\right)+2\cdot\left(0,1+0,3\right)+5\cdot\left(0,2+0,1\right)=2,3.$$

Вычисляем дисперсии:

$$D(X)=\sum_{i=1}^{2}\sum_{j=1}^{3}x_i^2 p_{ij}-\left(M(X)\right)^2=x_1^2\left(p_{11}+p_{12}+p_{13}\right)+$$

$$+x_2^2\left(p_{21}+p_{22}+p_{23}\right)-\left(M(X)\right)^2=$$

$$=1^2\cdot\left(0,1+0,1+0,2\right)+2^2\cdot\left(0,2+0,3+0,1\right)-\left(1,6\right)^2=0,24,$$

$$D(Y)=\sum_{i=1}^{2}\sum_{j=1}^{3}y_j^2 p_{ij}-\left(M(Y)\right)^2=y_j^2\left(p_{11}+p_{21}\right)+$$

$$+y_2^2\left(p_{12}+p_{22}\right)+y_3^2\left(p_{13}+p_{23}\right)-\left(M(Y)\right)^2=$$

$$=0^2\cdot\left(0,1+0,2\right)+2^2\cdot\left(0,1+0,3\right)+5^2\cdot\left(0,2+0,1\right)-\left(2,3\right)^2=3,81.$$

Корреляционный момент

$$\mu_{XY}=\sum_{i=1}^{2}\sum_{j=1}^{3}x_i y_j p_{ij}-M(X)\cdot M(Y)=x_1y_1p_{11}+x_1y_2p_{12}+x_1y_3p_{13}+$$

$$+x_2y_1p_{21}+x_2y_2p_{22}+x_2y_3p_{23}-M(X)\cdot M(Y)=$$

$$=1\cdot0\cdot0,1+1\cdot2\cdot0,1+1\cdot5\cdot0,2+2\cdot0\cdot0,2+2\cdot2\cdot0,3+2\cdot5\cdot0,1-1,6\cdot2,3=$$

$$=3,4-3,68=-0,28.$$

Коэффициент корреляции

$$r_{XY} = \frac{\mu_{XY}}{\sigma_x \cdot \sigma_y} = \frac{\mu_{XY}}{\sqrt{D(X)} \cdot \sqrt{D(Y)}} = \frac{-0,28}{\sqrt{0,24} \cdot \sqrt{3,81}} = -0,29.$$

Задачи для самостоятельного решения

1 Случайная величина (X,Y) задана таблицей распределения 3.

Написать законы распределения случайных компонент X и Y. Найти коэффициент корреляции между X и Y.

2 Передаются два сообщения, каждое из которых может быть искажено независимо друг от друга с вероятностями $p_1 = 0,6$ и $p_2 = 0,3$ соответственно. Рассматривается система двух случайных величин (X,Y), определяемых так:

$$X = \begin{cases} 1, \text{ если 1-е сообщение искажено;} \\ 0, \text{ если 1-е сообщение не искажено;} \end{cases}$$

$$Y = \begin{cases} 1, \text{ если 2-е сообщение искажено;} \\ 0, \text{ если 2-е сообщение не искажено.} \end{cases}$$

Найти закон распределения двумерной случайной величины (X,Y).

Таблица 3

x_i	y_i				
	–2	–1	0	1	2
–1	0,05	0,15	0,15	0,10	0,05
0	0,05	0,10	0,10	0,05	0
2	0	0,05	0,05	0,05	0,05

3 Число X выбирается случайным образом из множества целых чисел $\{1;2;3\}$. Затем из того же множества наудачу выбирается число Y, большее или равное X. Построить таблицу распределения двумерной случайной величины (X,Y), найти коэффициент корреляции.

4 Закон распределения дискретной двумерной случайной величины (X,Y) задан таблицей 4.

Таблица 4

X	Y		
	–1	0	1
1	0,15	0,3	0,35
2	0,05	0,05	0,1

Найти законы распределения составляющих X и Y. Вычислить вероятности $P(X = 2, Y = 0)$ и $P(X > Y)$. Установить, зависимы или нет составляющие X и Y.

5 Двумерная случайная величина (X,Y) задана таблицей 5 распределения.

Таблица 5

X	Y		
	3	4	5
0	0,02	0,12	0,06
1	0,03	0,18	0,09
2	0,05	0,30	0,15

Найти законы распределения составляющих X и Y, а также условный закон распределения величины Y при $X = 0$.

9 Выборка и её характеристики. Статистические оценки параметров распределения. Точечное и интервальное оценивание

Постановка задачи. Дана выборка: значения случайной величины X – результата n измерений некоторой величины. Требуется:

1) определить размах варьирования R;

2) составить статистический ряд распределения частот СВ X;

3) составить интервальные статистические ряды частот и относительных частот;

4) построить полигон и гистограмму относительных частот;

5) найти эмпирическую функцию распределения и построить её график;

6) вычислить выборочные значения числовых характеристик СВ X: математического ожидания $M(X)$, дисперсии $D(X)$, среднего квадратического отклонения $\sigma(X)$;

7) по виду полигона и гистограммы относительных частот, построенных в занятии 10, сделать выбор закона распределения СВ X;

8) найти точечные оценки параметров предполагаемого распределения, записать функцию распределения и плотность вероятности СВ X.

Решение типового варианта

Случайная величина X – диаметр головки заклёпки. Получены результаты измерения $n=90$ головок (в миллиметрах).

13,39	13,42	13,38	13,48	13,51	13,33	13,44	13,49	13,46
13,40	13,50	13,43	13,37	13,42	13,45	13,50	13,52	13,51
13,42	13,51	13,39	13,40	13,32	13,38	13,41	13,46	13,39
13,45	13,43	13,37	13,49	13,57	13,38	13,41	13,35	13,47
13,38	13,42	13,39	13,41	13,58	13,54	13,39	13,39	13,38
13,43	13,39	13,45	13,41	13,57	13,55	13,40	13,41	13,32
13,40	13,42	13,47	13,62	13,59	13,43	13,48	13,38	13,34
13,39	13,51	13,48	13,33	13,36	13,40	13,52	13,40	13,60
13,37	13,42	13,34	13,37	13,43	13,46	13,36	13,32	13,56
13,49	13,34	13,43	13,42	13,47	13,54	13,39	13,41	13,48

1 Найдём минимальное $x_{\min} = 13{,}32$ и максимальное $x_{\max} = 13{,}62$ значения вариант вариационного ряда и определим размах варьирования как разность между крайними значениями вариант: $R = x_{\max} - x_{\min} = 13{,}62 - 13{,}32 = 0{,}3$.

2 Подсчитаем частоты вариант и составим статистический ряд распределения частот (таблица 6).

Таблица 6

Диаметр головки x_i, мм	Частота m_i	Диаметр головки x_i, мм	Частота m_i
13,32	1	13,48	4
13,33	2	13,49	3
13,34	3	13,50	2
13,35	1	13,51	4
13,36	3	13,52	2
13,37	5	13,53	0
13,38	5	13,54	2
13,39	11	13,55	1
13,40	6	13,56	1
13,41	6	13,57	1
13,42	7	13,58	1
13,43	5	13,59	1
13,44	1	13,60	1
13,45	3	13,61	0
13,46	4	13,62	1
13,47	2		Контроль: $\sum_i m_i = 90$

3 Разобьём отрезок $[x_{\min}; x_{\max}]$ на k частичных интервалов равной длины и подсчитаем частоты попадания наблюдаемых значений СВ X в частичные интервалы. Число интервалов обычно выбирают произвольно $(5 \le k \le 15)$ или рассчитывают по формуле $h = \dfrac{R}{1 + 3{,}2\lg n}$, где n – ***объём выборки***, $n = 90$; h – длина интервала.

Число интервалов равно округлённому до целого значению $\dfrac{R}{h}$. Имеем

$$h = \frac{0{,}3}{1 + 3{,}2\lg 90} \approx \frac{0{,}3}{7{,}24}; \quad k = \frac{R}{h} = 1 + 3{,}2 \cdot \lg 90 \approx 7{,}24 \quad (k \in N).$$

Выберем для удобства число интервалов $k = 6$. Тогда $h = \dfrac{0{,}3}{6} = 0{,}05$.

Составим интервальные статистические ряды распределения частот и относительных частот (таблица 7, строки 2, 4 и 2, 5).

В строках 6–10 дополнительно подсчитаны числа, нужные при выполнении следующих заданий.

Контроль правильности вычислений: $\sum_{i=1}^{6} n_i = 90 = n$, $\sum_{i=1}^{6} w_i = 1$.

4 Построим полигон относительных частот (рисунок 3) по данным строк 5

и 10 (см. таблицу 7), соединяя точки $(y_i; w_i)$, y_i – середины интервалов.

По данным строк 2 и 9 таблицы 7 построим гистограмму относительных частот (рисунок 4) – ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы $(x_i; x_{i+1})$, а высоты равны отношению w_i/h – плотности относительной частоты. Площадь гистограммы относительных частот $\sum_{i=1}^{6} h \cdot \frac{w_i}{h} = \sum_{i=1}^{6} w_i = 1$.

Таблица 7

Номер интервала	1	1	2	3
Интервал $[x_i, x_{i+1})$	2	$[13{,}32; 13{,}37)$	$[13{,}37; 13{,}42)$	$[13{,}42; 13{,}47)$
Подсчёт частот	3	1 + 2 + 3 + 1 + + 3	5 + 5 + 11 + + 6 + 6	7 + 5 + 1 + 3 + + 4
Частота n_i в интервале	4	10	33	20
Относительная частота $w_i = \frac{n_i}{n}$	5	10/90	33/90	20/90
Накопленная частота n_x	6	10	43	63
Относительная накопленная частота $\frac{n_x}{n}$	7	10/90	43/90	63/90
Плотность частоты $\frac{n_i}{h}$	8	200	660	400
Плотность относительной частоты $\frac{w_i}{h}$	9	20/9	22/3	40/9
Середина интервала $\frac{x_i + x_{i+1}}{2}$	10	13,345	13,395	13,445

Продолжение таблицы 7

Номер интервала	1	4	5	6
Интервал $[x_i, x_{i+1})$	2	$[13{,}47; 13{,}52)$	$[13{,}52; 13{,}57)$	$[13{,}57; 13{,}62)$
Подсчёт частот	3	2 + 4 + 3 + 2 + + 4	2 + 2 + 0 + 1 + + 1	2 + 1 + 1 + 1 + + 0 + 1
Частота n_i в интервале	4	15	6	6
Относительная частота $w_i = \frac{n_i}{n}$	5	15/90	6/90	6/90
Накопленная частота n_x	6	78	84	90
Относительная накопленная частота $\frac{n_x}{n}$	7	78/90	84/90	1
Плотность частоты $\frac{n_i}{h}$	8	300	120	120
Плотность относительной частоты $\frac{w_i}{h}$	9	10/3	4/3	4/3
Середина интервала $\frac{x_i + x_{i+1}}{2}$	10	13,495	13,545	13,595

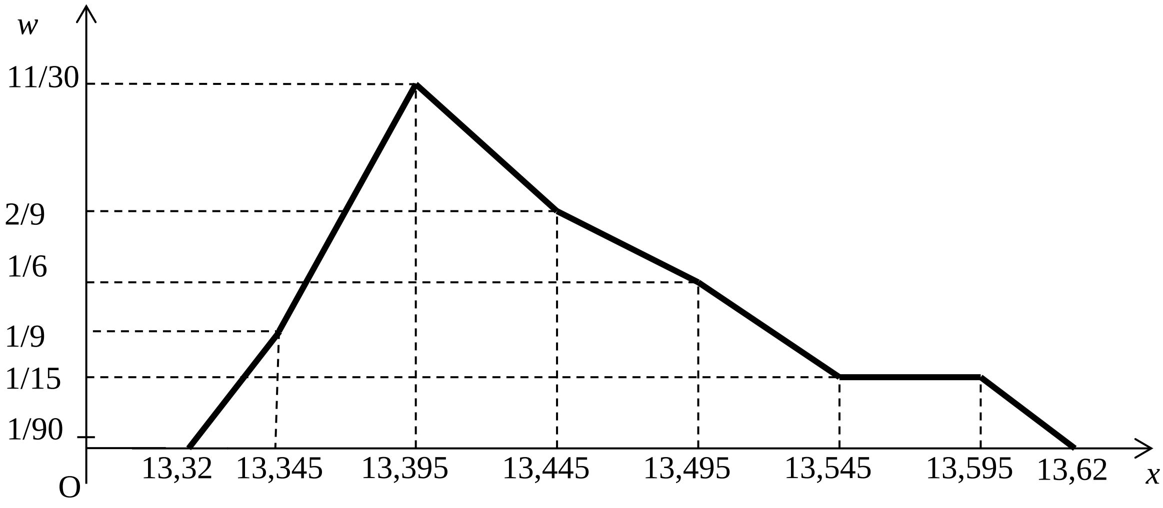

Рисунок 3

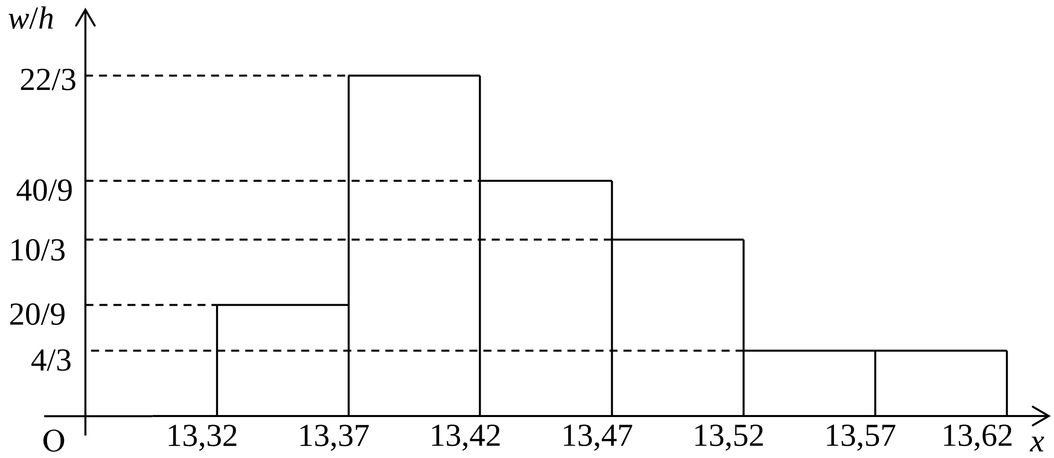

Рисунок 4

По гистограмме и полигону относительных частот можно судить о форме эмпирической кривой распределения – графике функции $f^*(x)$ (эмпирической плотности вероятности).

5 По данным таблицы 7 (строки 2 и 7) найдём эмпирическую функцию распределения (таблица 8) по формуле $F^*(x) = W(X < x) = \dfrac{n_x}{n}$.

Таблица 8

X	$\le 13{,}32$	13,37	13,42	13,47	13,52	13,57	>13,62
$F^*(x)$	0	1/9	43/90	7/10	39/45	42/45	1

Построим график $F^*(x)$: сначала на интервалах $(-\infty;13{,}32)$ и $(13{,}62;\infty)$, а затем в указанных в таблице 8 точках. Учитывая непрерывность функции $F^*(x)$, полученные точки соединим (рисунок 5).

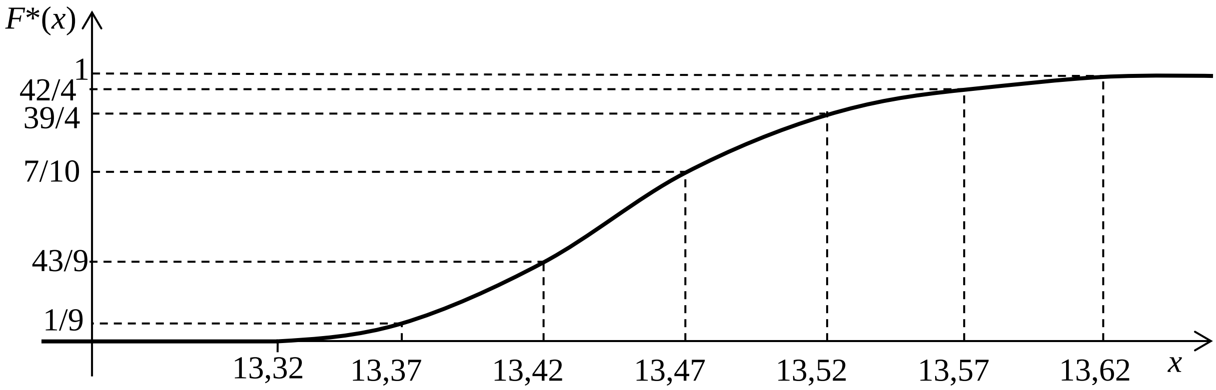

Рисунок 5

6 Найдём выборочную среднюю $M_в(X)=\overline{X}$ по формуле $\overline{X}=\frac{1}{n}\sum_{i=1}^{6} y_i\cdot n_i$,

где y_i – середина i-го интервала;

n_i – его частота;

n – объём выборки (таблица 7, строки 4, 10).

$$\overline{X}=\frac{13,345\cdot 10+13,395\cdot 33+13,445\cdot 20+13,495\cdot 15+13,545\cdot 6+13,595\cdot 6}{90}=$$

$$=\frac{133,45+442,035+268,9+202,425+81,27+81,57}{90}=\frac{1209,65}{90}\approx 13,44.$$

Вычислим выборочную дисперсию $D_в(X)$ по формуле

$$D_в(X)=\frac{1}{n}\sum_{i=1}^{6}\left(y_i-\overline{X}\right)^2\cdot n_i.$$

$$D_в(X)=\frac{1}{90}\Big((13,345-13,44)^2\cdot 10+(13,395-13,44)^2\cdot 33+(13,445-13,44)^2\cdot 20+$$

$$+(13,495-13,44)^2\cdot 15+(13,545-13,44)^2\cdot 6+(13,595-13,44)^2\cdot 6\Big)=$$

$$=\frac{1}{90}\left(0,09025+0,066825+0,0005+0,045375+0,06615+0,14415\right)=$$

$$=\frac{1}{90}\cdot 0,4794\approx 0,0053.$$

$$\sigma_в(X)=\sqrt{D_в(X)}=\sqrt{0,0053}\approx 0,073.$$

7 Полигон и гистограмма относительных частот (см. рисунки 3 и 4) напоминают нормальную кривую (рисунок 6, кривая Гаусса). Поэтому предположим, что распределение СВ X (диаметра головки заклёпки) является *нормальным*.

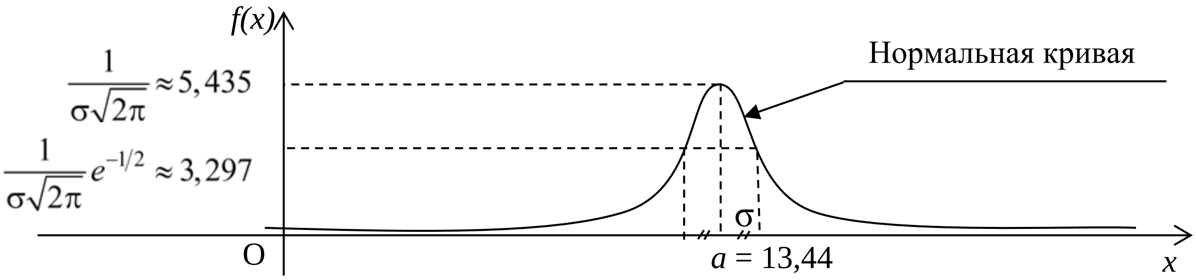

Рисунок 6

8 Плотность вероятности и функция распределения СВ X, распределённой по нормальному закону, имеют вид:

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

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-a)^2}{2\sigma^2}} dt. \tag{2}$$

Найдём точечные оценки параметров $a = M(X)$ и $\sigma = \sigma(X)$ нормального распределения:

$$\begin{cases} \tilde{a} = \tilde{M}(X) = M_в(X) = \overline{X} \approx 13{,}44; \\ \sigma_в = \sqrt{\tilde{D}(X)} = \sqrt{\frac{n}{n-1} D_в(X)} = \sqrt{\frac{90}{89} \cdot 0{,}0053} \approx 0{,}0734, \end{cases} \tag{3}$$

где $\tilde{M}(X)$ и $\tilde{D}(X)$ – точечные оценки соответственно для $M(X)$ и $D(X)$.

Следовательно, плотность вероятности предполагаемого распределения $N(a,\sigma)$ имеет вид:

$$f(x) = \frac{1}{0{,}0734 \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-13{,}44)^2}{0{,}0108}},$$

её график изображён на рисунке 6, и *функция распределения*

$$F(x) = \frac{1}{0{,}0734\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-13{,}44)^2}{0{,}0108}} dt$$

(учли формулы (1)–(3)).

Варианты заданий для самостоятельной работы

1 При обработке наружного диаметра 100 карданных валов получены следующие размеры (в миллиметрах).

42,2	42,0	43,3	43,7	42,5	42,9	43,0	42,7	42,4	42,6
41,7	41,9	41,8	43,2	42,4	42,5	42,5	42,5	43,0	42,5
41,2	41,6	42,6	40,3	41,6	43,4	43,5	42,7	43,1	42,6
42,5	40,6	44,1	42,0	40,4	42,0	43,1	42,5	43,4	42,5
43,2	42,8	40,7	43,6	42,5	41,1	41,3	43,6	43,8	42,8
42,5	43,3	42,9	40,8	41,1	42,3	40,2	43,1	43,7	42,6
42,8	43,6	42,5	41,5	40,0	41,4	42,1	44,0	42,4	43,9
42,4	42,5	42,4	42,9	42,2	43,5	43,0	42,8	43,2	43,4
42,5	42,6	42,6	42,4	43,5	43,8	43,4	42,5	42,8	41,3
42,1	42,5	42,4	43,4	44,0	42,6	42,5	44,0	41,7	43,2

2 Для анализа распределения диаметров отобраны 150 валиков, изготовленных на токарном станке. Получены следующие данные (в сантиметрах).

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

3 Случайная величина X – время работы элемента (в часах) для 100 элементов приняла следующие значения.

2,4	7,3	8,5	27,4	26,1	18,1	19,4	17,3	18,6	19,1
24,3	24,0	9,6	13,2	17,1	16,8	2,9	11,3	2,8	3,1
14,5	5,2	5,7	6,9	4,7	8,1	5,3	10,2	11,4	17,1
2,3	21,3	12,4	16,1	3,1	4,7	5,8	6,2	11,7	10,3
2,7	2,9	4,2	5,6	11,0	17,2	3,8	4,1	3,1	7,3
8,1	13,1	12,7	21,0	3,0	4,2	10,0	9,2	2,5	4,7
4,3	5,9	11,7	10,2	2,9	2,8	6,1	3,2	4,8	12,0
13,2	2,7	19,1	17,3	2,4	4,3	13,2	11,2	7,3	3,2
2,6	6,3	7,1	7,6	8	8,9	12,4	19,2	5,2	5,9
4,1	7,8	12,0	3,3	2,3	7,7	18,1	20,2	2,8	6,1

4 Подсчитаны затраты времени 120 рабочих на обработку одной детали (в минутах).

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

Таблица 9 – Дополнительные сведения к вариантам заданий

Номер варианта	1	2	3	4
Объём выборки	50	50	55	55
Уровень значимости	0,2	0,05	0,02	0,2
Критерий согласия	χ^2	λ	χ^2	χ^2
a_0	a_1	a_1	a_2	a_2
σ_0^2	σ_2^2	σ_1^2	σ_1^2	σ_1^2
Альтернативные гипотезы H_a	$a \neq a_0$, $\sigma^2 > \sigma_0^2$	$a < a_0$, $\sigma^2 \neq \sigma_0^2$	$a > a_0$, $\sigma^2 < \sigma_0^2$	$a \neq a_0$, $\sigma^2 > \sigma_0^2$

Замечание – a_1 *и* σ_1 *соответственно значения* a *и* σ *в правом конце доверительных интервалов,* a_2 *и* σ_2 *– в левом*

10 Статистическая проверка гипотез. Критерии согласия

Постановка задачи. На основе данных из п. 9 требуется:

1) проверить согласие эмпирической функции распределения $F^*(x)$ с теоретической $F(x)$ при помощи критерия согласия χ^2-Пирсона (χ^2 – «хи квадрат») или λ-Колмогорова;

2) в случае нормального распределения СВ X по заданному уровню значимости α:

а) найти интервальные оценки параметров распределения;

б) проверить нулевую гипотезу $H_0 : M(X) = a = a_0$ о математическом ожидании при альтернативной гипотезе $H_a : a \neq a_0$ $(a > a_0;\ a < a_0)$;

в) проверить нулевую гипотезу $H_0 : D(X) = \sigma^2 = \sigma_0^2$ о дисперсии против альтернативной гипотезы $H_a : \sigma^2 \neq \sigma_0^2$ $(\sigma^2 < \sigma_0^2,\ \sigma^2 > \sigma_0^2)$.

Примечание – Использовать гипотезы из таблицы 9 п. 9.

Решение типового варианта

1 Проведём проверку гипотезы о нормальном распределении СВ X – диаметра головки заклёпки.

Вероятность попадания СВ X, распределённой по нормальному закону $N(a,\sigma)$, в интервал $(\alpha;\beta)$ найдём по формуле

$$p = P(\alpha < X < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) - \Phi\left(\frac{\alpha - a}{\sigma}\right), \qquad (4)$$

где $\Phi(x)$ – функция Лапласа.

Вероятность попадания СВ X в первый частичный интервал $(-\infty;13,37)$ (учтём, что $\Phi(-x) = -\Phi(x)$)

$$p_1 = P(-\infty < X < 13,37) = \Phi\left(\frac{13,37 - 13,44}{0,0734}\right) - \Phi\left(\frac{-\infty - 13,44}{0,0734}\right) =$$

$$= \Phi(\infty) - \Phi(0,9537) = 0,5 - 0,3289 = 0,1711.$$

Аналогично вероятности попадания СВ X в остальные интервалы

$$p_2 = P(13,37 \le X < 13,42) = \Phi\left(\frac{13,42 - 13,44}{0,0734}\right) - \Phi\left(\frac{13,37 - 13,44}{0,0734}\right) =$$

$$= \Phi(0,9537) - \Phi(0,2725) \approx 0,3289 - 0,1064 = 0,2225;$$

$$p_3 = P(13,42 \le X < 13,47) = \Phi\left(\frac{13,47 - 13,44}{0,0734}\right) - \Phi\left(\frac{13,42 - 13,44}{0,0734}\right) =$$

$$= \Phi(0,4086) + \Phi(0,2725) \approx 0,1591 + 0,1064 = 0,2655;$$

$$p_4 = P(13,47 \le X < 13,52) = \Phi\left(\frac{13,52 - 13,44}{0,0734}\right) - \Phi\left(\frac{13,47 - 13,44}{0,0734}\right) =$$

$$= \Phi(1,0899) - \Phi(0,4087) \approx 0,3621 - 0,1591 = 0,2030;$$

$$p_5 = P(13,52 \le X < 13,57) = \Phi\left(\frac{13,57 - 13,44}{0,0734}\right) - \Phi\left(\frac{13,52 - 13,44}{0,0734}\right) =$$

$$= \Phi(1,7711) - \Phi(1,0899) \approx 0,4616 - 0,3621 = 0,0995;$$

$$p_6 = P(13,57 \le X < +\infty) = \Phi\left(\frac{\infty - 13,44}{0,0734}\right) - \Phi\left(\frac{13,57 - 13,44}{0,0734}\right) =$$

$$= \Phi(\infty) - \Phi(1,7711) = 0,5 - 0,4616 = 0,0384.$$

Проверим гипотезу о нормальном распределении с помощью критерия согласия χ^2-Пирсона. Вычисления, необходимые для определения наблюдаемого значения выборочной статистики χ^2, проведём в таблице 10.

Таблица 10 – Определение выборочной статистики χ^2

Интервал наблюдаемых значений СВ X	Частота n_i	p_i	np_i	$(n_i - np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
$(-\infty;13,37)$	10	0,17105	15,3945	29,1006	1,8903
$[13,37;13,42)$	33	0,22255	20,0295	168,2339	8,3993
$[13,42;13,47)$	20	0,2655	23,895	15,1710	0,6349
$[13,47;13,52)$	15	0,20305	18,2745	10,7223	0,5867
$[13,52;13,57)$	6	0,0995	8,955	8,7320	0,9751
$[13,57;\ \infty)$	6	0,03835	3,4515	6,4948	1,8817
$\sum$	90	1	90		$\chi^2_{набл.} = 14,368$

Из таблицы распределения χ^2 по уровню значимости $\alpha = 0,05$ (см. таблицу 9) и числу степеней свободы $\nu = k - r - 1 = 6 - 2 - 1 = 3$ (здесь k – число интервалов; r – число параметров распределения) находим $\chi^2_{0,05;\,3} = 7,815$. Так как $\chi^2_{набл.} > \chi^2_{0,05;\,3}$, то отклоняем гипотезу о нормальном распределении диаметров головок заклёпок.

Проверим гипотезу о нормальном распределении с помощью критерия согласия λ – Колмогорова.

Все вспомогательные расчёты, необходимые для нахождения выборочной характеристики λ, сведём в таблицу 11.

Таблица 11 – Нахождение выборочного значения λ

Интервалы наблюдаемых значений СВ X	Частота n_i	n_x	p_i	$F^*(x) = \frac{n_x}{n}$	$F(x) = P(X < x)$	$\lvert F^*(x) - F(x) \rvert$
$(-\infty;13,37)$	10	10	0,17105	$1/9$	0,17105	0,0599
$[13,37;13,42)$	33	43	0,22255	$43/90$	0,3936	0,0842
$[13,42;13,47)$	20	63	0,2655	$7/10$	0,6591	0,0409
$[13,47;13,52)$	15	78	0,20305	$39/45$	0,86215	0,0045
$[13,52;13,57)$	6	84	0,0995	$42/45$	0,96165	0,0283
$[13,57;\ \infty)$	6	90	0,03835	1	1	0
$\sum$	90	–	–	–	–	–

$$D = \max\lvert F^*(x) - F(x)\rvert = 0,0842; \qquad \lambda_{набл.} = D \cdot \sqrt{n} = 0,0842 \cdot \sqrt{90} \approx 0,7988.$$

По таблице квантилей распределения Колмогорова и уровню значимости $\alpha = 0,05$ (надёжность $P = 1-\alpha = 0,95$) находим критическое значение $\lambda_{0,05} = 1,358$.

Так как $\lambda_{набл.} < \lambda_{0,05}$, то нет оснований для отклонения гипотезы о нормальном распределении диаметров головок заклёпок.

2 Найдём интервальные оценки параметров нормального распределения.

Доверительный интервал, накрывающий математическое ожидание СВ X с надёжностью $P = 1-\alpha$, имеет вид:

$$\tilde{a} - t\frac{\sigma_в}{\sqrt{n}} < a < \tilde{a} + t\frac{\sigma_в}{\sqrt{n}}. \tag{5}$$

По таблице квантилей распределения Стьюдента по заданному уровню значимости $\alpha = 0,05$ и числу степеней свободы $\nu = n-1 = 89$ найдём квантиль:

$$t = t_{\frac{\alpha}{2},\nu} = t_{0,025;89} = 1,987.$$

Вычислим ***точность оценки***: $\delta = t\cdot\frac{\sigma_в}{\sqrt{n}} = 1,987\cdot\frac{0,073}{\sqrt{90}} \approx 0,015$.

Искомый доверительный интервал для $M(X) = a$

$$13,44 - 0,015 < a \le 13,44 + 0,015;\ \ 13,425 < a \le 13,455.$$

Полуинтервал $(13,425;\,13,455]$ накрывает неизвестное $M(X)$ с вероятностью $p = 0,95$.

Доверительный интервал, накрывающий среднее квадратическое отклонение СВ X с надёжностью $P = 1-\alpha$,

$$j_1\cdot\sigma_в < \sigma < j_2\cdot\sigma_в\ , \tag{6}$$

где $j_1 = \sqrt{\dfrac{n-1}{\chi^2_{\frac{\alpha}{2},\nu}}}$;

$j_2 = \sqrt{\dfrac{n-1}{\chi^2_{1-\frac{\alpha}{2},\nu}}}$.

По таблице распределения χ^2 по заданной доверительной вероятности $p = 0,95$ и числу степеней свободы $\nu = 89$ найдём числа $j_1 = 0,873$, $j_2 = 1,171$.

Замечание – Если число степеней свободы $\nu > 30$*, то* $\chi^2_{\alpha,\nu}$ *можно найти из равенства Уилсона–Гильферти:*

$$\chi^2_{\alpha,\nu} = \nu\left(1 - \frac{2}{9\nu} + z_\alpha\sqrt{\frac{2}{9\nu}}\right)^3,$$

где z_α *находят из равенства* $\Phi\left(z_\alpha\right) = \dfrac{1-2\alpha}{2}$.

Искомый доверительный интервал для параметра σ

$$0{,}873\cdot 0{,}073 < \sigma < 1{,}171\cdot 0{,}073;\ \ 0{,}0637 < \sigma < 0{,}0855.$$

Интервал $\left(0{,}0637;\, 0{,}0855\right)$ накрывает неизвестное $\sigma(X)$ с вероятностью $p = 0{,}95$.

Проведём проверку статистических гипотез о математическом ожидании и дисперсии нормального распределения.

Для математического ожидания имеем $a_0 = a_1 = 13{,}455$. Надо проверить нулевую гипотезу H_0: $a = a_0 = 13{,}455$ против альтернативной $H_a : a \neq a_0$ $(a > a_0$ или $a < a_0)$.

Правило 1. *Чтобы при заданном уровне значимости* α *проверить нулевую гипотезу* $H_0 : a = a_0$ *(о равенстве* $a = M(X)$ *СВ X, распределённой по закону* $N(a, \sigma)$, *предполагаемому значению* a_0*) при альтернативной гипотезе* $H_a : a \neq a_0$, *надо вычислить наблюдаемое значение U-критерия*

$$\mathrm{U}_{набл.} = \frac{\left(\overline{X} - a_0\right)\sqrt{n}}{\sigma_в} \tag{7}$$

и по таблице значений функции Лапласа найти критическое значение $U_{кр.}$ *двусторонней критической области из равенства*

$$\Phi\left(U_{кр}\right) = \frac{1-\alpha}{2}.$$

Если $\left|U_{набл.}\right| < U_{кр.}$, *то* H_0 *принимают; если* $\left|U_{набл.}\right| > U_{кр.}$, *то нулевую гипотезу отклоняют и принимают альтернативную гипотезу* $H_a : a \neq a_0$.

Правило 2. *Пусть* $H_0 : a = a_0$ *и альтернативная гипотеза* $H_a : a > a_0$. *Критическое значение* $U_{кр.}$ *правосторонней критической области находят из равенства*

$$\Phi\left(U_{кр.}\right) = \frac{1-2\alpha}{2}.$$

Если $U_{набл.} < U_{кр.}$, *то нулевую гипотезу принимают; если* $U_{набл.} > U_{кр.}$, *то* *нулевую гипотезу отвергают, принимают альтернативную.*

Правило 3. *Пусть* $H_0: a = a_0$ *и альтернативная гипотеза* $H_a: a < a_0$. *По правилу 2 находят вспомогательное критическое значение* $U_{кр}$, *полагают* $U^*_{кр.} = -U_{кр.}$ *(граница левосторонней критической области). Если* $U_{набл.} > -U_{кр.}$, H_0 *принимают; если* $U_{набл.} < -U_{кр.}$, H_0 *отвергают (наблюдаемое значение* U *-критерия попадает в критическую область) и принимают альтернативную гипотезу* H_a.

Воспользуемся правилом 1. По таблице значений функции Лапласа $\Phi(u)$ и по уровню значимости $\alpha = 0,05$ находим

$$\Phi\left(U_{кр.}\right) = \frac{1-0,05}{2} = 0,475;\ U_{кр.} = 1,96.$$

Вычисляем по формуле (7)

$$U_{набл.} = \frac{(13,44 - 13,455)\sqrt{90}}{0,073} \approx -1,939.$$

Так как $\left|U_{набл.}\right| < U_{кр.}$, то нулевую гипотезу принимаем и отвергаем альтернативную гипотезу $H_a: a \neq 13,455$.

Для среднего квадратического отклонения имеем $\sigma_0 = \sigma_2 = 0,0641$. Проверим гипотезу $H_0: \sigma^2 = \sigma_0^2 = 0,0041$ против альтернативной $H_a: \sigma^2 > \sigma_0^2$ $\left(\sigma^2 = \sigma_0^2 \text{ или } \sigma^2 < \sigma_0^2\right)$.

Правило 4. *Для того чтобы при заданном уровне значимости* α *проверить нулевую гипотезу* $H_0: \sigma^2 = \sigma_0^2$ *(о равенстве неизвестной дисперсии* σ^2 *предполагаемому значению* σ_0^2*) при альтернативной гипотезе* $H_a: \sigma^2 > \sigma_0^2$ *, вычисляют наблюдаемое значение статистического критерия*

$$\chi^2_{набл.} = \frac{(n-1)\sigma_в^2}{\sigma_0^2} \tag{8}$$

и по таблице распределения χ^2- *Пирсона находят критическое значение (по уровню значимости* α *и числу степеней свободы* $\nu = n-1$) $\chi^2_{кр.} = \chi^2_{\alpha,\nu}$. *Если* $\chi^2_{набл.} < \chi^2_{кр.}$, *то нулевую гипотезу принимают; если* $\chi^2_{набл.} > \chi^2_{кр.}$, *то отвергают нулевую гипотезу в пользу альтернативной.*

Правило 5. *Пусть* $H_0: \sigma^2 = \sigma_0^2$ *и альтернативная гипотеза* $H_a: \sigma^2 \neq \sigma_0^2$. *Находят критические левую и правую точки* $\chi^2_{лев.кр.} = \chi^2_{1-\frac{\alpha}{2},\nu}$ *и* $\chi^2_{прав.\,кр.} = \chi^2_{\frac{\alpha}{2},\nu}$ *по* *таблице распределения* χ^2 *-Пирсона.*

Если $\chi^2_{лев.кр.} < \chi^2_{набл.} < \chi^2_{прав.кр.}$ *(наблюдаемое значение критерия попало в область допустимых значений), то гипотезу* H_0 *принимают. Если* $\chi^2_{набл.} < \chi^2_{лев.кр.}$ *или* $\chi^2_{набл.} > \chi^2_{прав.кр.}$*, то отклоняют гипотезу* H_0 *в пользу гипотезы* H_a*.*

Правило 6*. Пусть* $H_0 : \sigma^2 = \sigma_0^2$ *и альтернативная гипотеза* $H_a : \sigma^2 < \sigma_0^2$*. Находят критическое значение* $\chi^2_{кр.} = \chi^2_{1-\alpha,\nu}$*. Если* $\chi^2_{набл.} > \chi^2_{кр.}$*, то* H_0 *принимают; если* $\chi^2_{набл.} < \chi^2_{кр.}$*, то* H_0 *отклоняют, принимают альтернативную гипотезу* H_a*.*

Воспользуемся правилом 4 и равенством Уилсона–Гильферти:

$$\chi^2_{набл.} = \frac{89 \cdot 0,0734^2}{0,0641^2} \approx 116,6987.$$

По уровню значимости $\alpha = 0,05$ и числу степеней свободы $\nu = 89$ находим $\chi^2_{кр.} = \chi^2_{\alpha,\nu}$.

Так как $\nu > 30$, находим $\Phi(z_\alpha) = \dfrac{1 - 2 \cdot 0,05}{2} = 0,45,$ $z_\alpha = 1,645$ и получаем

$$\chi^2_{\alpha,\nu} = \chi^2_{0,05;\,89} = 89\left(1 - \frac{2}{9 \cdot 89} + 1,645\sqrt{\frac{2}{9 \cdot 89}}\right)^3 \approx 112,022.$$

$\chi^2_{набл.} > \chi^2_{кр.}$ и потому гипотезу H_0 отвергаем, принимаем гипотезу $H_a : \sigma^2 > 0,0041$.

11 Линейная регрессия и корреляция

Выборочное уравнение прямой линии регрессии *Y* на *X* имеет следующий вид:

$$\overline{y}_x - \overline{y} = r_в \frac{\sigma_y}{\sigma_x}\left(x - \overline{x}\right), \qquad (9)$$

где $\overline{y}_x$ – условная средняя;

$\overline{x}$ и $\overline{y}$ – выборочные средние признаков *X* и *Y*;

σ_x и σ_y – выборочные средние квадратические отклонения;

$r_в$ – выборочный коэффициент корреляции, $r_в = \dfrac{\sum n_{xy} xy - n\overline{x}\,\overline{y}}{n\sigma_x\sigma_y}$;

n_{xy} – частота пары значений (x, y).

Аналогично выборочное уравнение прямой линии регрессии *X* на *Y*

$$\overline{x}_y - \overline{x} = r_{\text{в}} \frac{\sigma_x}{\sigma_y}\left(y - \overline{y}\right). \qquad (10)$$

Если данные наблюдений над признаками *X* и *Y* заданы в виде корреляционной таблицы с равноотстоящими вариантами, то целесообразно перейти к условным вариантам:

$$u_i = \frac{x_i - C_1}{h_1};\ v_i = \frac{y_i - C_2}{h_2},$$

где C_1 – «ложный нуль» вариант *X* (новое начало отсчёта); в качестве «ложного нуля» выгодно принять варианту, которая расположена примерно в середине вариационного ряда (условимся принимать в качестве «ложного нуля» варианту, имеющую наибольшую частоту);

h_1 – шаг, т. е. разность между двумя соседними вариантами *X*;

C_2 – «ложный нуль» вариант *Y*;

h_2 – шаг вариант *Y*.

В этом случае выборочный коэффициент корреляции $r_в = \dfrac{\sum n_{uv}uv - n\overline{u}\,\overline{v}}{n\sigma_u\sigma_v}$,

причём слагаемое $\sum n_{uv}uv$ удобно вычислять, используя расчётную таблицу 14 (см. пример 1).

Величины $\overline{u}, \overline{v}, \sigma_u, \sigma_v$ могут быть найдены либо методом произведений (при большом числе данных), либо непосредственно по формулам

$$\overline{u} = \left(\sum n_u u\right)/n;\ \overline{v} = \left(\sum n_v v\right)/n;\ \sigma_u = \sqrt{\overline{u^2} - \left(\overline{u}\right)^2};\ \sigma_v = \sqrt{\overline{v^2} - \left(\overline{v}\right)^2}.$$

Зная эти величины, можно определить входящие в уравнения регрессии (9) и (10) величины по формулам

$$\overline{x} = \overline{u}h_1 + C_1;\ \overline{y} = \overline{v}h_2 + C_2;\ \sigma_x = \sigma_u h_1;\ \sigma_y = \sigma_v h_2.$$

Пример 1 – По результатам наблюдений случайной величины (*X*, *Y*) из корреляционной таблицы 12 необходимо:

– найти выборочные уравнения регрессии *Y* на *X* и *X* на *Y;*

– построить графики прямых линий регрессии *Y* на *X* и *X* на *Y*.

Решение

Составим корреляционную таблицу 13 в условных вариантах, выбрав в качестве «ложных нулей» $C_1 = 30$ и $C_2 = 36$ (каждая из этих вариант расположена в середине соответствующего вариационного ряда).

Таблица 12

Y	X					n_y
	20	25	30	35	40	
16	4	6	—	—	—	10
26	—	8	10	—	—	18
36	—	—	32	3	9	44
46	—	—	4	12	6	22
56	—	—	—	1	5	6
n_x	4	14	46	16	20	n = 100

Таблица 13

v	u					
	–2	–1	0	1	2	n_v
–2	4	6	–	–	–	10
–1	–	8	10	–	–	18
0	–	–	32	3	9	44
1	–	–	4	12	6	22
2	–	–	–	1	5	6
n_u	4	14	46	16	20	n= 100

Найдём средние $\overline{u}$ и $\overline{v}$:

$$\overline{u} = \left(\sum n_u u\right)\Big/ n = \left(4\cdot(-2)+14\cdot(-1)+46\cdot 0+16\cdot 1+20\cdot 2\right)\Big/100 = 0,34;$$
$$\overline{v} = \left(\sum n_v v\right)\Big/ n = \left(10\cdot(-2)+18\cdot(-1)+44\cdot 0+22\cdot 1+6\cdot 2\right)\Big/100 = -0,04.$$

Найдём вспомогательные величины $\overline{u^2}$ и $\overline{v^2}$:

$$\overline{u^2} = \left(\sum n_u u^2\right)\Big/ n = \left(4\cdot 4+14\cdot 1+16\cdot 1+20\cdot 4\right)\Big/100 = 1,26;$$
$$\overline{v^2} = \left(\sum n_v v^2\right)\Big/ n = \left(10\cdot 4+18\cdot 1+22\cdot 1+6\cdot 4\right)\Big/100 = 1,04.$$

Найдём σ_u и σ_v:

$$\sigma_u = \sqrt{\overline{u^2}-\left(\overline{u}\right)^2} = \sqrt{1,26-0,34^2} = 1,07;\ \sigma_v = \sqrt{\overline{v^2}-\left(\overline{v}\right)^2} = \sqrt{1,04-0,04^2} = 1,02.$$

Для вычисления $\sum n_{uv}uv$ составим расчётную таблицу 14. Суммируя числа последнего столбца таблицы 14, находим $\sum\limits_v v\cdot U = \sum n_{uv}uv = 82$.

Для контроля вычислений находим сумму чисел последней строки:

$$\sum_u u\cdot V = \sum n_{uv}uv = 82.$$

Совпадение сумм свидетельствует о правильности вычислений.

Таблица 14

v	u					$U=\sum n_{uv}\cdot u$	$v\cdot U$
	–2	– 1	0	1	2		
–2	–8 4 –8	–6 6 –12				–14	28
–1	–	–8 8 –8	0 10 –10			–8	8
0	–	–	0 32 0	3 3 0	18 9 0	21	0
1	–	–	0 4 4	12 12 12	12 6 6	24	24
2	–	–	–	1 1 2	10 5 10	11	22
$V=\sum n_{uv}\cdot v$	–8	–20	–6	14	16		$\sum_v v\cdot U=82$
$u\cdot V$	16	20	0	14	32	$\sum_u u\cdot V=82$	Контроль

Пояснения к составлению таблицы 14.

1 Произведение частоты n_{uv} на варианту u, т. е. $n_{uv}\cdot u$, записывают в правом верхнем углу клетки, содержащей значение частоты. Например, в правых верхних углах клеток первой строки записаны произведения: $4\cdot(-2)=-8$; $6\cdot(-1)=-6$.

2 Складывают все числа, помещённые в правых верхних углах клеток одной строки, и их сумму помещают в клетку этой же строки «столбца *U*». Например, для первой строки $u=-8+(-6)=-14$.

3 Умножают варианту v на U и полученное произведение записывают в соответствующую клетку «столбца *vU*». Например, в первой строке таблицы $v=-2$, $U=-14$, следовательно, $vU=(-2)\cdot(-14)=28$.

4 Сложив все числа «столбца *vU*», получают сумму $\sum uV$, которая равна искомой сумме $\sum n_{uv}uv$. Например, для таблицы 14 $\sum_v uV=82$, следовательно, искомая сумма $\sum n_{uv}uv=82$.

5 Для контроля аналогичные вычисления производят по столбцам: произведения $n_{uv}v$ записывают в левый нижний угол клетки, содержащей значение частоты; все числа из левых нижних углов клеток одного столбца складывают и их сумму помещают в «строку *V*»; умножают каждую варианту *u* на *V* и результат записывают в клетках последней строки.

6 Сложив все числа последней строки, получают сумму $\sum_{u} vU$, которая также равна искомой сумме $\sum n_{uv}uv$. Например, для таблицы 14 $\sum_{v} vU = 82$, следовательно, $\sum n_{uv}uv = 82$.

Найдём искомый выборочный коэффициент корреляции:

$$r_в = \frac{\sum n_{uv}uv - n\overline{u}\,\overline{v}}{n\sigma_u\sigma_v} = \frac{82 - 100 \cdot 0,34 \cdot (-0,04)}{100 \cdot 1,07 \cdot 1,02} = 0,76.$$

Найдём шаги h_1 и h_2 (разности между любыми двумя соседними вариантами):

$$h_1 = 25 - 20 = 5;\ h_2 = 26 - 16 = 10.$$

Найдём $\overline{x}$ и $\overline{y}$, учитывая, что $C_1 = 30$, $C_2 = 36$:

$$\overline{x} = \overline{u} \cdot h_1 + C_1 = 0,34 \cdot 5 + 30 = 31,70;\ \overline{y} = \overline{v} \cdot h_2 + C_2 = (-0,04) \cdot 10 + 36 = 35,60.$$

Найдём σ_x и σ_y: $\sigma_x = h_1 \cdot \sigma_u = 5 \cdot 1,07 = 5,35$; $\sigma_y = h_2 \cdot \sigma_v = 10 \cdot 1,02 = 10,2$.

Подставив найденные величины в соотношение (9), получим искомое уравнение прямой линии регрессии *Y* на *X*

$$\overline{y}_x - 35,60 = 0,76 \cdot \frac{10,2}{5,35}(x - 31,70) \text{ или } \overline{y}_x = 1,45x - 10,36.$$

Аналогично из (10) получим уравнение прямой линии регрессии *X* на *Y*

$$\overline{x}_y - 31,70 = 0,76 \cdot \frac{5,35}{10,2}(y - 35,60) \text{ или } \overline{x}_y = 0,40y + 17,51.$$

Для построения из последнего уравнения можно выразить y: $y = 2,51\overline{x}_y - 43,92$.

Строим графики прямых линий регрессии (рисунок 7).

Задачи для самостоятельного решения

По результатам наблюдений случайной величины (*X*, *Y*) из корреляционных таблиц 15 (вариант 1), 16 (вариант 2) и 17 (вариант 3) необходимо:

1) найти выборочные уравнения регрессии *Y* на *X* и *X* на *Y;*

2) построить графики прямых линий регрессии *Y* на *X* и *X* на *Y*.

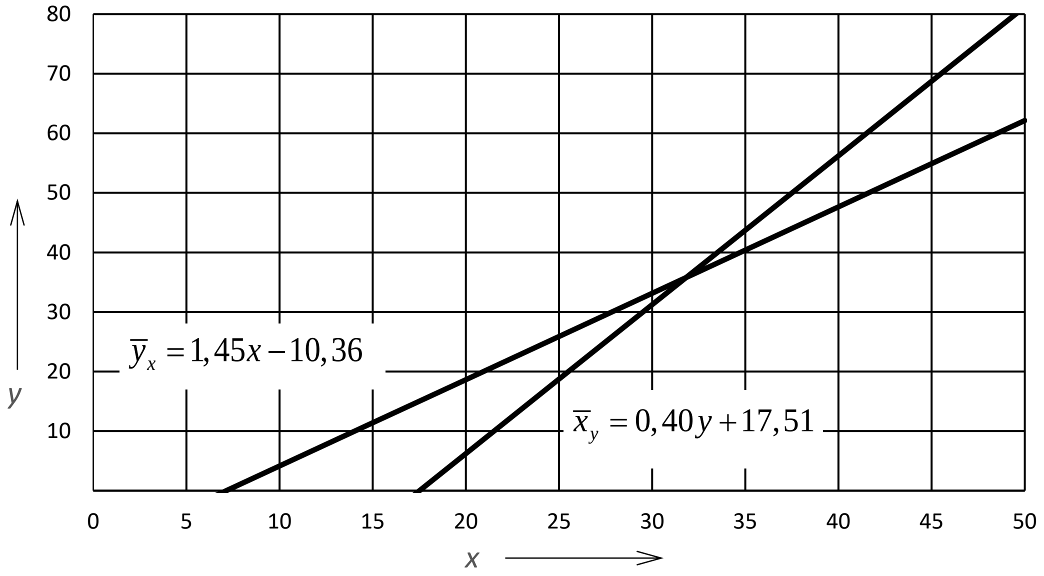

Рисунок 7

Таблица 15

Y	X					
	5	8	11	14	17	20
10	2	2	–	–	–	–
20	–	3	7	–	–	–
30	–	–	5	30	10	–
40	–	–	7	10	8	–
50	–	–	–	5	6	5

Таблица 16

Y	X					
	10	15	20	25	30	35
7	3	3	–	–	–	–
10	–	4	6	–	–	–
13	–	–	8	28	9	–
16	–	–	7	10	8	–
19	–	–	–	5	6	3

Таблица 17

Y	X					
	5	10	15	20	25	30
20	1	5	–	–	–	–
23	–	2	8	–	–	–
26	–	–	10	25	10	–
29	–	–	5	12	8	–
32	–	–	–	5	6	3

Список литературы

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

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

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

4 **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.

5 **Герасимович, А. И**. Математическая статистика / А. И. Герасимович. – Минск: Вышэйшая школа, 1983. – 276 с.