

Министерство образования и науки Российской Федерации

Северный (Арктический) федеральный университет

Кафедра математики

Теория вероятностей и математическая статистика

Методические указания и рекомендации по выполнению контрольной работы
для студентов заочной формы обучения



Архангельск

2011

Рассмотрены и рекомендованы к изданию методической комиссией
института строительства и архитектуры
Северного (Арктического) федерального университета

3 декабря 2010 г.

Составители: Баданина Л.А., ст. преподаватель,
Серова Г.В., ст. преподаватель.

Рецензент: В.Н. Попов, доктор физ.-мат. наук, доцент.

Баданина Л.А., Серова Г.В. **Теория вероятностей и математическая статистика**: Методические указания и рекомендации по выполнению контрольной работы по математике для студентов заочной формы обучения.- Архангельск: Изд-во С(А)ФУ, 2011, 75с.

Методические указания и рекомендации по выполнению контрольной работы по теме «Теория вероятностей и математическая статистика» включают основные теоретические положения, разбор примеров.

Методические указания предназначены для студентов второго курса заочной формы обучения специальностей 270102.65 «Промышленное и гражданское строительство», 140106.65 «Энергообеспечение предприятий», 140211.65 «Электроснабжение», а также могут быть использованы студентами других технических специальностей.

©Северный (Арктический)

федеральный университет, 2011

© Л.А. Баданина, Г.В. Серова, 2011

Оглавление

Теория вероятностей.....	4
1. Случайные события.....	4
1.1. Алгебра событий.....	4
1.2. Элементы комбинаторики.....	6
1.3. Классическое определение вероятности.....	7
1.4. Теоремы сложения и умножения вероятностей.....	9
1.5. Формула полной вероятности. Формула Байеса.....	13
1.6. Повторение испытаний.....	15
2. Случайные величины.....	21
2.1. Дискретная случайная величина.....	21
2.1.1. Закон распределения дискретной случайной величины.....	21
2.1.2. Функция распределения.....	23
2.1.3. Действия над случайными величинами.....	24
2.1.4. Числовые характеристики дискретной случайной величины.....	26
2.1.5. Основные законы распределения дискретных случайных величин.....	29
2.2. Непрерывная случайная величина.....	30
2.2.1. Дифференциальная и интегральная функции распределения.....	30
2.2.2. Числовые характеристики непрерывной случайной величины.....	31
2.2.3. Основные законы распределения непрерывных случайных величин.....	32
Математическая статистика.....	36
3. Обработка результатов статистических наблюдений.....	37
3.1. Вариационный ряд	37
3.2. Эмпирическая функция распределения.....	38
3.3. Полигон и гистограмма.....	39
3.4. Числовые характеристики статистического	

распределения.....	40
3.5. Точечные оценки параметров распределения. Метод моментов.....	41
3.6. Проверка гипотезы о нормальном распределении генеральной совокупности. Критерий согласия Пирсона.....	42
3.7. Критерий согласия Колмогорова.....	45
3.8. Интервальные оценки параметров нормально распределенной случайной величины	46
4. Элементы корреляционного анализа.....	54
4.1. Корреляционная таблица.....	54
4.2. Числовые характеристики.....	56
4.3. Выборочный коэффициент корреляции и проверка гипотезы о его значимости	57
4.4. Уравнение прямой регрессии.....	59
Простейшие задачи по теории вероятностей	68
Простейшие задачи по математической статистике.....	70
Приложения.....	72
Список литературы.....	78

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

1. Случайные события

1.1. Алгебра событий

Теория вероятностей – математическая наука, изучающая закономерности, присущие массовым случайным явлениям.

В теории вероятностей первичными понятиями являются испытания и события. Пусть проводится некоторое **испытание** со случайным исходом. Множество всех возможных взаимоисключающих исходов данного испытания называется **множеством элементарных событий**. Совокупность всех элементарных событий называется **достоверным событием**. Всякое подмножество множества элементарных событий называется **случайным событием**.

Различают три типа событий: достоверное, невозможное и случайное.

Событие называется **достоверным**, если в результате испытания оно обязательно произойдет. Обозначение достоверного события: Ω .

Событие называется **невозможным**, если в результате испытания оно никогда не произойдет. Обозначение: I (impossible) или $\emptyset$.

Событие называется **случайным**, если в результате испытания оно может произойти, а может и не произойти. Обычно случайные события обозначают латинскими буквами: A, B, C и т.д.

Например, проводится **испытание**: подбрасываем одну монету, в результате этого испытания может быть всего два **исхода**: выпал герб, выпала решка (выпадение на ребро достаточно редкий исход, мы такие исходы рассматривать не будем), т. е. в результате испытания может произойти одно из **элементарных событий**: $A = \{\text{выпал герб}\}$, $B = \{\text{выпала решка}\}$.

Например, проводим испытание: случайным образом вынимаем два шара из ящика, в котором 1 белый и 3 черных шара, тогда событие $\Omega = \{\text{вынули хотя бы один черный шар}\}$ является достоверным, потому что двух белых в ящике нет, событие $I = \{\text{вынули зеленый шар}\}$ - невозможным, а событие $A = \{\text{вынули два разных по цвету шара}\}$ - случайным.

Суммой двух событий A и B называется такое событие $C = A + B$ (или $A \cup B$), которое состоит в наступлении хотя бы одного из событий A или B .

Произведением двух событий A и B называется такое событие $C = AB$ (или $A \cap B$), которое состоит в наступлении событий A и B вместе.

Например, проводится **испытание**: подбрасываем игральную кость, в результате этого испытания может произойти одно из шести **элементарных событий**: $A_1 = \{\text{выпало 1 очко}\}$, $A_2 = \{\text{выпало 2 очка}\}$, $A_3 = \{\text{выпало 3 очка}\}$, $A_4 = \{\text{выпало 4 очка}\}$, $A_5 = \{\text{выпало 5 очков}\}$, $A_6 = \{\text{выпало 6 очков}\}$, выпадение на ребро достаточно редкий исход, мы такие исходы рассматривать не будем. Событие $\Omega = \{1, 2, 3, 4, 5, 6\}$ является достоверным. На множестве Ω можно ввести и другие случайные события: $A = \{\text{выпало нечетное число очков}\}$, $B = \{\text{выпало четное число очков}\}$, $C = \{\text{выпало простое число очков}\}$, и т.п., но они уже не являются элементарными событиями, т.к., их можно представить в виде объединения элементарных событий

$$A = A_1 \cup A_3 \cup A_5; \quad B = A_2 \cup A_4 \cup A_6; \quad C = A_1 \cup A_2 \cup A_3 \cup A_5.$$

Два события называются **несовместными**, если в результате одного испытания они не могут произойти одновременно, т.е. $A \cap B = \emptyset$.

Несколько событий $A_1, A_2, \dots, A_n$ называются **попарно несовместными**, если каждая пара не может произойти одновременно, т.е. каждое не может произойти с остальными.

События $A_1, A_2, \dots, A_n$ образуют **полную группу**, если в результате испытания происходит хотя бы одно из них, и ничего кроме этих событий не может произойти, т.е. $\bigcup_{i=1}^n A_i = \Omega$.

События A и $\bar{A}$ называются **противоположными**, если $A + \bar{A} = \Omega$, $A \cdot \bar{A} = I$. Очевидно, что события A и $\bar{A}$ несовместны.

Пример 1. Проведем испытание: один раз бросаем игральную кость. Рассмотрим события: $A_1 = \{\text{выпало четное число очков}\}$, $A_2 = \{\text{выпало нечетное число очков}\}$, $A_3 = \{\text{выпало три очка}\}$ и $A_4 = \{\text{выпало пять очков}\}$. Указать события, которые являются совместными, несовместными, попарно несовместными и т.п.

Решение.

События A_2 и A_3 являются совместными, т.к. они могут произойти одновременно, в случае, если выпадет три очка. События A_3 и A_4 - несовместные. События A_1 , A_3 и A_4 - попарно несовместные. События A_2 , A_3 и A_4 - не являются попарно несовместными, поскольку, например, A_2 и A_4 могут произойти одновременно. События A_1 , A_2 , A_3 и A_4 образуют полную группу, поскольку $A_1 \cup A_2 \cup A_3 \cup A_4 = \Omega = \{1, 2, 3, 4, 5, 6\}$. События A_1 и A_2 образуют полную группу несовместных событий, поскольку $A_1 \cup A_2 = \Omega$, а $A_1 \cap A_2 = \emptyset$. События A_1 , A_3 и A_4 не образуют полную группу, так как $A_1 \cup A_3 \cup A_4 = \{2, 3, 4, 5, 6\} \neq \Omega$, поскольку может выпасть одно очко.

1.2. Элементы комбинаторики

Комбинаторика — раздел математики, посвященный решению задач выбора и расположения k элементов из некоторого множества n элементов в соответствии с заданными правилами. Ниже будем рассматривать множества, состоящие из n различных элементов.

Перестановками называются комбинации, состоящие из n элементов и отличающиеся только порядком их расположения. Число перестановок равно

$$P_n = n!, \quad (1.1)$$

где $n! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (n-1) \cdot n$, причем принято считать, что $0! = 1$ (читается «ноль факториал»).

Сочетаниями называются комбинации, состоящие в выборе k элементов из n элементов ($k \leq n$), отличающиеся хотя бы одним элементом. Число сочетаний равно

$$C_n^k = \frac{n!}{k! \cdot (n-k)!}. \quad (1.2)$$

Размещениями называются комбинации, состоящие в выборе k элементов из n элементов ($k \leq n$), отличающиеся либо составом элементов, либо их порядком. Число размещений равно

$$A_n^k = P_k \cdot C_n^k = \frac{n!}{(n-k)!}. \quad (1.3)$$

Пример 2. Имеем множество, состоящее из трех элементов $\{1, 2, 3\}$. Сколько из трех предложенных цифр можно составить различных чисел с неповторяющимися цифрами: а) трехзначных; б) двузначных?

Решение. а) Количество различных комбинаций из трех цифр вычисляем по формуле (1.1), т.е. $P_3 = 3! = 6$. И действительно ровно шесть трехзначных чисел можно составить: 123, 132, 213, 231, 312, 321.

б) Порядок в выборе двух элементов из трех важен, т.к. 23 и 32 различные числа, поэтому количество способов выбрать 2 из трех вычисляем по формуле (1.3). Получаем $A_3^2 = \frac{3!}{1!} = 6$. И действительно ровно шесть двузначных чисел можно составить из трех: 12, 13, 21, 23, 31, 32.

Пример 3. Сколько существует способов выбора трех студентов из 10 на конференцию?

Решение. Порядок в выборе трех элементов из десяти не важен, т.к. делегация из Иванова, Петрова и Сидорова от делегации Петрова, Иванова и Сидорова не отличается, поэтому число способов выбора равно числу сочетаний из 10 студентов по 3

$$C_{10}^3 = \frac{10!}{3! 7!} = 120.$$

1.3. Классическое определение вероятности

Вероятность события – это число, характеризующее степень возможности появления события. Вероятность появления события A вычисляется по формуле

$$P(A) = \frac{m}{n}. \quad (1.4)$$

где m - число исходов, благоприятствующих наступлению события A , n - число всех возможных исходов.

Свойства вероятности

1. Вероятность случайного события A : $0 < P(A) < 1$.
2. Вероятность достоверного события $P(\Omega) = 1$.
3. Вероятность невозможного события $P(I) = 0$.

Пример 4. Монету бросают один раз. Вычислить вероятность того, что выпадет герб.

Решение. Рассмотрим событие $A = \{\text{выпал герб}\}$. При бросании монеты возможны два исхода ($n = 2$): выпал герб и выпала решка, и только первый исход благоприятствует наступлению события A ($m = 1$). Тогда по классическому определению вероятности $P(A) = \frac{1}{2}$.

Пример 5. Игральную кость подбрасывают один раз. Вычислить вероятность того, что выпадет простое число очков.

Решение. Рассмотрим событие $A = \{\text{выпало простое число очков}\}$. При бросании игральной кости возможны шесть исходов – выпадение 1, 2, 3, 4, 5, 6 очков, т.е. $n = 6$. Из шести всевозможных исходов только четыре исхода {выпало 1 очко}, {выпало 2 очка}, {выпало 3 очка} и {выпало 5 очков} являются благоприятствующими наступлению события A ($m = 4$). Вероятность появления простого числа очков равна $P(A) = \frac{4}{6} = \frac{2}{3}$.

Пример 6. В урне 4 белых и 6 черных шара. Извлекли одновременно 3 шара. Найти вероятность того, что: а) все шары белые; б) все шары черные; в) один белый и два черных.

Решение. Будем считать, что все шары разные и пронумерованы 1, 2, ..., 10. Пусть первые четыре номера - белые, последние - черные. Тогда количество всевозможных способов извлечь три шара из десяти равно $C_{10}^3 = 240$.

а) Рассмотрим событие $A = \{\text{из урны извлекли три белых шара}\}$. Тогда количество способов, благоприятствующих появлению события A равно $C_4^3 = 4$. Получаем $P(A) = \frac{4}{240} = \frac{1}{60}$.

б) Рассмотрим событие $B = \{\text{из урны извлекли три черных шара}\}$. Тогда количество способов, благоприятствующих появлению этого события равно $C_6^3 = 20$. Получаем $P(B) = \frac{20}{240} = \frac{1}{12}$.

в) Рассмотрим событие $C = \{\text{из урны извлекли один белый и два черных шара}\}$. Тогда количество способов, благоприятствующих появлению этого события равно произведению $C_4^1 C_6^2 = 60$. Получаем $P(C) = \frac{60}{240} = \frac{1}{4}$.

Статистическое определение вероятности

Пусть проводятся n испытаний, в результате которых событие A наступает m раз. Тогда отношение $\frac{m}{n}$, при $n \rightarrow \infty$ называется статистической вероятностью.

Пример 7. Изучали вероятность рождения мальчика. Среди 1000 новорожденных мальчик появился в 515 случаях. Тогда статистическая вероятность (относительная частота) рождения мальчика равна $\frac{515}{1000} = 0,515$.

Геометрическое определение вероятности

Геометрической вероятностью события A называется отношение меры области s , благоприятствующей появлению события A , к мере всей области S_D

$$P(A) = \frac{s}{S_D}. \quad (1.5)$$

Области могут быть одномерными, двумерными, трехмерными.

Пример 8. Пусть в квадрат, со стороной 3 см вписан круг. Найти вероятность того, что точка, случайным образом брошенная в квадрат, попадет в круг.

Решение. Обозначим событие $A = \{\text{точка попала в круг}\}$. Площадь квадрата равна 9 см^2 , а площадь круга радиусом 1,5 см равна $2,25\pi \text{ см}^2$. Тогда $P(A) = \frac{2,25\pi}{9} = 0,25\pi$.

1.4. Теоремы сложения и умножения вероятностей

Теоремы сложения.

Теорема 1. Если события A и B несовместные, то

$$P(A + B) = P(A) + P(B). \quad (1.6)$$

Теорема 2. Если события $A_1, A_2, \dots, A_n$ попарно-несовместны, то

$$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i). \quad (1.7)$$

Теорема 3. Если события $A_1, A_2, \dots, A_n$ образуют полную группу попарно-несовместных событий, то

$$\sum_{i=1}^n P(A_i) = 1. \quad (1.8)$$

Следствие. $P(\bar{A}) + P(A) = 1$.

Теорема 4. Если события A и B совместные, то

$$P(A + B) = P(A) + P(B) - P(AB). \quad (1.9)$$

Теоремы умножения

События A и B называются **независимыми**, если появление одного из них не влияет на вероятность появления другого (в противном случае события **зависимы**).

События $A_1, A_2, \dots, A_n$ называются **попарно независимыми**, если каждые два из них независимы.

Теорема 1. Если события A и B независимые, то

$$P(AB) = P(A) \cdot P(B) \quad (1.10)$$

Теорема 2. Если события $A_1, A_2, \dots, A_n$ попарно-независимы, то

$$P\left(\prod_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i). \quad (1.11)$$

Теорема 3. Если события A и B зависимые, то

$$P(AB) = P(A) \cdot P_A(B) = P(B)P_B(A), \quad (1.12)$$

где $P_B(A)$ - условная вероятность события A при условии, что событие B уже произошло. Для условной вероятности $P_B(A)$ есть еще обозначение $P(A|B)$.

Теорема 4. Вероятность появления хотя бы одного из событий $A_1, A_2, \dots, A_n$, независимых в совокупности, равна

$$P = 1 - P(\bar{A}_1) \cdot P(\bar{A}_2) \cdot \dots \cdot P(\bar{A}_n). \quad (1.13)$$

Пример 9. Вероятность попадания в мишень при каждом выстреле для первого стрелка равна 0,7, а для второго – 0,8. Оба они делают по одному выстрелу по мишени. Найти вероятности событий: A – в мишени две пробоины; B – в мишени только одна пробоина; C – в мишени хотя бы одна пробоина.

Решение. Введем два события: $A_1 = \{\text{в мишень попал первый стрелок}\}$, $A_2 = \{\text{в мишень попал второй стрелок}\}$. Тогда $A = A_1 \cdot A_2$, события A_1 и A_2 – независимые, применяя формулу (1.10), находим вероятность события A

$$P(A) = P(A_1 \cdot A_2) = P(A_1) \cdot P(A_2) = 0,7 \cdot 0,8 = 0,56.$$

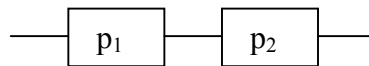
Теперь найдем вероятность события B . В мишени одна пробоина означает, что попал либо первый, либо второй, т.е. $B = A_1 \overline{A_2} + \overline{A_1} A_2$. Выстрелы производятся независимо, поэтому события A_1 и $\overline{A_2}$ – независимые, события $\overline{A_1}$ и A_2 – так же независимые, а два события $A_1 \overline{A_2}$ и $\overline{A_1} A_2$ – несовместные. Используя теоремы сложения и умножения, имеем

$$P(B) = P(A_1 \overline{A_2} + \overline{A_1} A_2) = P(A_1 \overline{A_2}) + P(\overline{A_1} A_2) = 0,7 \cdot 0,2 + 0,3 \cdot 0,8 = 0,38.$$

Вероятность события C находим по формуле (1.13)

$$P(C) = 1 - P(\overline{A_1})P(\overline{A_2}) = 1 - 0,3 \cdot 0,2 = 0,94.$$

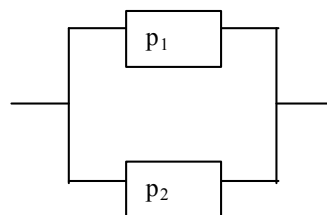
Пример 10. Найти вероятность прохождения тока через цепь при последовательном соединении, если вероятности исправной работы элементов равны p_1 и p_2 . Элементы работают независимо друг от друга.



Решение. При последовательном соединении ток пройдет по цепи только тогда, когда оба элемента работают. Следовательно,

$$P = p_1 \cdot p_2$$

Пример 11. Найти вероятность прохождения тока через цепь при параллельном соединении, если вероятности исправной работы элементов равны p_1 и p_2 . Элементы работают независимо друг от друга.



Решение. При параллельном соединении ток пройдет по цепи только тогда, когда хотя бы один из двух элементов работает. Введем два события: $A_1 = \{\text{первый элемент работает}\}$, $A_2 = \{\text{второй элемент работает}\}$. Тогда событие $C = \{\text{ток идет по всей цепи}\}$ можно записать в виде

$$C = A_1 \cdot A_2 + A_1 \cdot \overline{A_2} + \overline{A_1} \cdot A_2.$$

Однако противоположное событие $\overline{C}$ можно представить в более компактном виде

$$\overline{C} = \overline{A_1} \cdot \overline{A_2}.$$

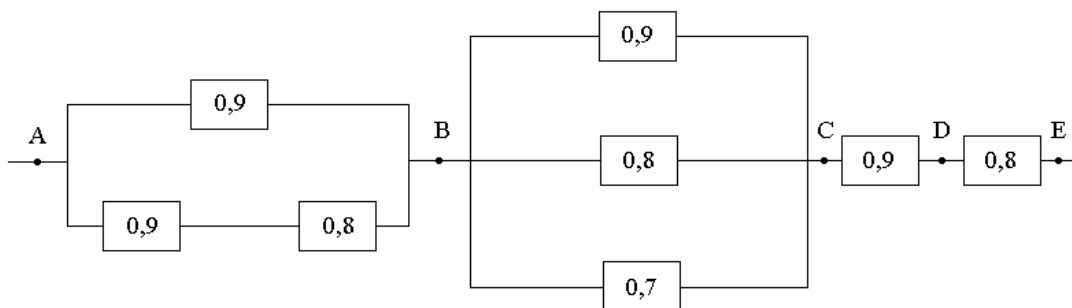
События $\overline{A_1}$ и $\overline{A_2}$ - независимые. Обозначим $P(\overline{A_1}) = q_1$ $P(\overline{A_2}) = q_2$, тогда

$$P(\overline{C}) = P(\overline{A_1} \cdot \overline{A_2}) = P(\overline{A_1}) \cdot P(\overline{A_2}) = q_1 \cdot q_2.$$

По следствию теоремы сложения 3 имеем

$$P(C) = 1 - P(\overline{C}) = 1 - q_1 \cdot q_2 = 1 - (1 - p_1)(1 - p_2).$$

Пример 12. Найти вероятность прохождения тока через цепь, если вероятности исправной работы элементов указаны на схеме. Элементы работают независимо друг от друга.



Решение. Введем события: A_1 - ток идет по участку АВ, A_2 - ток идет по участку ВС, A_3 - ток идет по участку СЕ. Согласно примеру 10, вероятность того, что ток идет по участку СЕ равна

$$P(A_3) = 0,9 \cdot 0,8 = 0,72.$$

Следуя примеру 11, вероятность того, что ток по участку ВС не идет равна

$$P(\overline{A_2}) = (1 - 0,9) \cdot (1 - 0,8)(1 - 0,7) = 0,006.$$

Тогда вероятность того, что ток по участку ВС идет равна

$$P(A_2) = 1 - P(\overline{A_2}) = 0,994.$$

Участок АВ состоит из параллельного и последовательного соединения. Вероятность того, что ток идет по нижнему участку АВ равна

$$P = 0,9 \cdot 0,8 = 0,72.$$

Тогда вероятность того, что по всему участку АВ ток не идет, вычисляем как для параллельного соединения (см. пример 11)

$$P(\overline{A_1}) = (1 - 0,9)(1 - 0,72) = 0,028.$$

А вероятность того, что по участку АВ ток идет равна

$$P(A_1) = 1 - 0,028 = 0,972.$$

Соединение АВ, ВС, СЕ – последовательное, поэтому вероятность того, что ток идет по всему участку АЕ равна

$$P = P(A_1) \cdot P(A_2) \cdot P(A_3) = 0,972 \cdot 0,994 \cdot 0,072 = 0,696.$$

1.5. Формула полной вероятности. Формула Байеса

Теорема. Если событие A может произойти только при условии появления одного из попарно несовместных событий (гипотез) $B_1, B_2, \dots, B_n$, образующих полную группу, то вероятность события A равна сумме произведений вероятностей каждого из этих событий (гипотез) на соответствующие условные вероятности события A

$$P(A) = \sum_{i=1}^n P(B_i)P(A/B_i) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + \dots + P(B_n)P(A/B_n) \quad (1.14)$$

Из теоремы умножения (1.12) и формулы полной вероятности (1.14) легко получить формулу Байеса, которая позволяет «переставить причину и следствие»: по известному факту события позволяет вычислить вероятность того, что оно было вызвано данной причиной.

Теорема. Условная вероятность гипотезы B_i , при условии того, что событие A произошло, вычисляется по формуле Байеса

$$P(B_i/A) = \frac{P(B_i) \cdot P(A/B_i)}{\sum_{i=1}^n P(B_i) \cdot P(A/B_i)}, \quad (1.15)$$

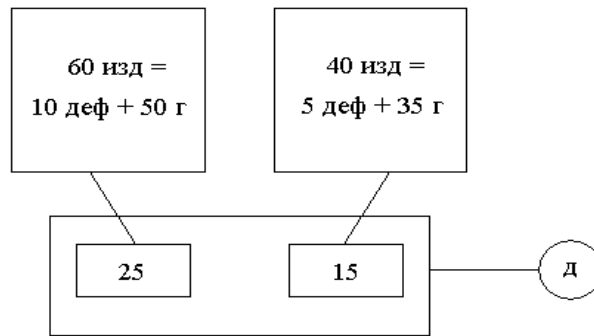
где $B_1, B_2, \dots, B_n$ - попарно несовместные события, образующие полную группу.

Вероятности $P(B_i)$ называют априорными вероятностями, т.е. вероятностями событий до выполнения опыта, а условные вероятности этих событий $P(B_i/A)$ апостериорными, т.е. уточненными в результате опыта, исходом которого послужило появление события A .

Пример 13. Имеются две партии однотипных изделий. Первая партия состоит из 60 изделий, среди которых 10 бракованных, вторая из 40 изделий, среди которых 5 бракованных. Из первой партии берется случайным образом 25 изделий, а из второй – 15. Эти изделия смешиваются и образуется новая смешанная партия,

из которой берется наугад одно изделие. Найти вероятность того, что оно будет бракованным.

Решение. Введем события $B_1 = \{\text{извлеченное из смешанной партии изделие оказалось из первой партии}\}$, $B_2 = \{\text{извлеченное из смешанной партии изделие оказалось из второй партии}\}$, $A = \{\text{извлеченное из смешанной партии изделие бракованное}\}$.



Тогда по формуле (1.4) имеем $P(B_1) = \frac{25}{40}$, $P(B_2) = \frac{15}{40}$. Вероятность выбрать дефектное изделие из смешанной партии, при условии что осуществилось событие B_1 , равна $P(A/B_1) = \frac{10}{60}$; а при условии что осуществилось событие B_2 , равна $P(A/B_2) = \frac{5}{40}$. По формуле полной вероятности (1.14) получаем

$$P(A) = \frac{25}{40} \cdot \frac{10}{60} + \frac{15}{40} \cdot \frac{5}{40} \approx 0,1042.$$

Пример 14. Среди поступающих на сборку деталей с I станка 0,1% бракованных, со II-0,2%; с III- 0,25%, с IV – 0,5%. Производительности их относятся соответственно как 4:3:2:1. Взятая наудачу деталь оказалась стандартной. На каком станке вероятнее всего она изготовлена?

Решение. Введем события: $A = \{\text{взятая деталь стандартная}\}$, $B_i = \{\text{взятая деталь изготовлена на } i\text{-ом станке}\}$, $i = 1, 2, 3, 4$. Вероятности гипотез B_i будут равны:

$$P(B_1) = \frac{4}{4+3+2+1} = \frac{4}{10}, \quad P(B_2) = \frac{3}{10}, \quad P(B_3) = \frac{2}{10}, \quad P(B_4) = \frac{1}{10}.$$

Условные вероятности:

$$P(A/B_1) = 1 - \frac{0,1}{100} = 0,999, \quad P(A/B_2) = 0,998, \quad P(A/B_3) = 0,9975, \quad P(A/B_4) = 0,995.$$

Тогда по формуле (1.14) найдем вероятность того, что наудачу взятая деталь оказалась стандартной.

$$P(A) = \frac{4}{10} \cdot 0,999 + \frac{3}{10} \cdot 0,998 + \frac{2}{10} \cdot 0,9975 + \frac{1}{10} \cdot 0,995 = 0,998.$$

Применим формулу Байеса (1.15), чтобы найти вероятность того, что извлеченная деталь, оказавшаяся стандартной, изготовлена на I станке.

$$P(B_1 / A) = \frac{P(B_1) \cdot P(A / B_1)}{P(A)} = \frac{\frac{4}{10} \cdot 0,999}{0,998} \approx 0,4004.$$

Аналогично, $P(B_2 / A) \approx 0,3000$, $P(B_3 / A) \approx 0,1999$, $P(B_4 / A) \approx 0,0997$.

Таким образом, вероятнее всего, что деталь изготовлена на первом станке.

1.6. Повторение испытаний

Пусть проводится конечное число n последовательных независимых испытаний, в каждом из которых событие A может появиться, а может и не появиться. Испытания называются **независимыми**, если вероятность появления события A в каждом из испытаний не зависит от того, появилось или не появилось это событие в других испытаниях. Вероятность появления события A в каждом испытании одинакова и равна p , а вероятность противоположного события обозначим через $q = P(\bar{A}) = 1 - p$.

Теорема. Вероятность того, что в n испытаниях событие A появится ровно k раз вычисляется по формуле Бернулли

$$P_n(k) = C_n^k p^k q^{n-k}, \quad (1.16)$$

где C_n^k вычисляется по формуле (1.2). Легко доказать, что $\sum_{k=0}^n P_n(k) = 1$.

Замечание. Конечную серию n повторных независимых испытаний с двумя исходами, вероятности которых равны p и q , называют **схемой Бернулли**.

Наивероятнейшее число k_0 появления события A в n независимых испытаниях находится по формуле

$$np - q \leq k_0 \leq np + p. \quad (1.17)$$

Если $np - q$ - целое число, то наивероятнейших чисел ровно два: $np - q$ и $np + p$.

Пример 15. Игральную кость подбрасывают 10 раз. Найти вероятность того, что: а) шесть очков выпадет ровно 3 раза; б) шесть очков выпадет хотя бы один раз.

Решение. Введем событие $A = \{\text{в одном испытании выпадет ровно шесть очков}\}$. Тогда $P(A) = p = \frac{1}{6}$, $P(\bar{A}) = q = 1 - \frac{1}{6} = \frac{5}{6}$.

а) По формуле (1.16) вычисляем вероятность того, что шесть очков выпадет ровно 3 раза из десяти $P_{10}(3) = C_{10}^3 \cdot \frac{1}{6^3} \cdot \left(\frac{5}{6}\right)^7 = \frac{1}{18 \cdot 30^6} \approx 0$.

б) По формуле (1.13) вычисляем вероятность того, что шесть очков выпадет хотя бы один раз из десяти $P = 1 - q^{10} = 1 - \left(\frac{5}{6}\right)^{10} = 0,8385$.

Пример 16. В семье 5 детей. Вероятность рождения мальчика считаем равным $1/2$. Найти вероятность того, что в семье ровно три мальчика. Какое наиболее вероятное количество мальчиков в семье?

Решение. Имеем $n = 5$, $p = 1/2$, $q = 1/2$. Тогда

$$P_5(3) = C_5^3 p^3 q^2 = \frac{5!}{3!2!} \cdot \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^2 = \frac{10}{32}.$$

По формуле (1.17) находим наивероятнейшее число мальчиков в семье из 5 детей

$$\frac{5}{2} - \frac{1}{2} \leq k_0 \leq \frac{5}{2} + \frac{1}{2},$$

т.е. наивероятнейшее число равно 2 или 3.

Пример 17. Для нормальной работы автобазы на линии должно быть не менее 8 автомашин, а их имеется 10. Вероятность того, что автомашина на линию не выйдет равна 0,1. Найти вероятность нормальной работы автобазы в ближайший день.

Решение. Событие $A = \{\text{автобаза работает нормально}\}$. Для нормальной работы автобазы на линии должно быть или 8 автомашин, или 9, или 10. Введем три события $B_1 = \{\text{на линии работает 8 машин}\}$, $B_2 = \{\text{на линии работает 9 машин}\}$, $B_3 = \{\text{на линии работает 10 машин}\}$. Тогда событие A можно представить в виде объединения трех несовместных событий

$$A = B_1 + B_2 + B_3.$$

Используя формулу (1.16), и учитывая, что в нашем случае $n = 10$, $p = 0,9$, $q = 0,1$, вычисляем

$$P(B_1) = P_{10}(8) = C_{10}^8 0,9^8 0,1^2 = 0,1937,$$

$$P(B_2) = P_{10}(9) = C_{10}^9 0,9^9 0,1^1 = 0,3874,$$

$$P(B_3) = P_{10}(10) = C_{10}^{10} 0,9^{10} 0,1^0 = 0,3487.$$

Таким образом,

$$P(A) = P(B_1) + P(B_2) + P(B_3) = 0,9298.$$

Локальная формула Лапласа

Пользоваться формулой Бернулли (1.16) при больших значениях n достаточно трудно. Например, для того, чтобы найти вероятность того, что при 30 испытаниях герб выпадет 10 раз, необходимо вычислить $P_{30}(10) = C_{30}^{10} \left(\frac{1}{2}\right)^{10} \left(\frac{1}{2}\right)^{20}$. Первый сомножитель требует громоздких вычислений.

При большом числе опытов по схеме Бернулли ($n \geq 25$) удобнее пользоваться приближенными формулами.

Локальная теорема Лапласа. Если $npq \geq 10$, то вероятность того, что в n независимых испытаниях событие A наступит k раз, находится по формуле

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \varphi(x), \quad (1.18)$$

где $x = \frac{k - np}{\sqrt{npq}}$, а функция $\varphi(x)$ определяется равенством $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ (рис.2.1).

Для облегчения вычислений функция $\varphi(x)$ представлена в виде таблицы (см. приложение 1).

Свойства функции Гаусса $\varphi(x)$.

1). Функция $\varphi(x)$ - четная, т.е. $\varphi(-x) = \varphi(x)$.

2). $\lim_{x \rightarrow +\infty} \varphi(x) = 0$, поэтому для всех значений $x \geq 4$ можно считать $\varphi(x) \approx 0$.

3). $\int_{-\infty}^{+\infty} \varphi(x) dx = 1$.

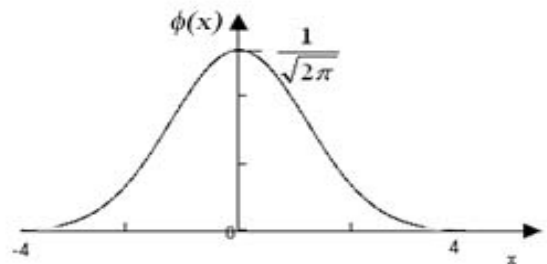


Рис.2.1. Функция Гаусса.

Пример 18. Найти вероятность того, что при 600 выстрелах мишень будет поражена 250 раз, если вероятность поражения мишени при одном выстреле равна 0,4.

Решение. Поскольку $n = 600$ - велико, и $npq = 144 \geq 10$, то вероятность будем вычислять по формуле (1.18).

Находим $x = \frac{250 - 600 \cdot 0,4}{\sqrt{600 \cdot 0,4 \cdot 0,6}} = \frac{10}{12} \approx 0,833$. По таблицам определяем $\varphi(0,833) \approx 0,2820$. Тогда по локальной теореме Лапласа $P_{600}(250) \approx \frac{1}{12} \cdot 0,2820 \approx 0,0235$.

Формула Пуассона (закон редких событий)

Теорема. Если $npq < 10$ и $p < 0,1$ ($p \rightarrow 0$), то вероятность того, что в n независимых испытаниях событие A наступит k раз, находится по формуле Пуассона

$$P_n(k) \approx \frac{\lambda^k \cdot e^{-\lambda}}{k!}, \quad (1.19)$$

где $\lambda = np$ - среднее число появлений события A в n испытаниях.

Пример 19. Завод отправил на базу 5000 доброкачественных изделий. Вероятность того, что в пути изделие повредится, равна 0,0002. Найти вероятность того, что в пути будет повреждено: а) 3 изделия; б) не более 3-х изделий; в) более 3-х изделий.

Решение. По условию задачи имеем $n = 5000, p = 0,0002$. Тогда $q = 0,9998$ и $npq = 0,9998 < 10, p < 0,1$.

а) $k = 3, \lambda = np = 1$, тогда по формуле (1.19) вычисляем $P_{5000}(3) \approx \frac{e^{-1}}{3!} \approx 0,0613$.

б) Введем событие $B = \{\text{в пути будет повреждено не более 3-х изделий}\}$. Это событие равно объединению четырех несовместных событий $\{\text{в пути повреждено ровно } i \text{ изделий}\}$, где $i = 0, 1, 2, 3$. Тогда по формулам (1.7) и (1.19) имеем

$$\begin{aligned} P(B) &= P(k \leq 3) = P_{5000}(0) + P_{5000}(1) + P_{5000}(2) + P_{5000}(3) = \\ &= \frac{e^{-1} \cdot 1^0}{0!} + \frac{e^{-1} \cdot 1^1}{1!} + \frac{e^{-1} \cdot 1^2}{2!} + \frac{e^{-1} \cdot 1^3}{3!} = e^{-1} \left(1 + 1 + \frac{1}{2} + \frac{1}{6} \right) = e^{-1} \cdot 2 \frac{2}{3} = e^{-1} \cdot \frac{8}{3} \approx 0,9810. \end{aligned}$$

в) Противоположным событием к событию B является следующее: $\overline{B} = \{\text{в пути будет повреждено более 3-х изделий}\}$. Тогда

$$P(\overline{B}) = P(k > 3) = 1 - P(B) = 1 - 0,9810 = 0,0190.$$

Интегральная теорема Муавра-Лапласа

Если n – большое и $npq \geq 10$, то вероятность того, что событие A произойдет от k_1 раз до k_2 раз вычисляют по интегральной формуле Муавра-Лапласа

$$P_n(k_1 \leq \kappa \leq k_2) \approx \Phi(x_2) - \Phi(x_1), \quad (1.20)$$

где $x_2 = \frac{k_2 - np}{\sqrt{npq}}$, $x_1 = \frac{k_1 - np}{\sqrt{npq}}$ и $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ – функция Лапласа (рис.2.2).

Таблица значений функции $\Phi(x)$ приведена в приложении 2.

Свойства функции Лапласа

1) $\Phi(x)$ – нечетная функция, т.е.

$$\Phi(-x) = -\Phi(x).$$

2) $\Phi(x)$ – возрастающая функция на всей числовой оси.

3) $\lim_{x \rightarrow \infty} \Phi(x) = 0,5$, поэтому для всех

значений $x \geq 5$ можно считать $\Phi(x) \approx 0,5$.

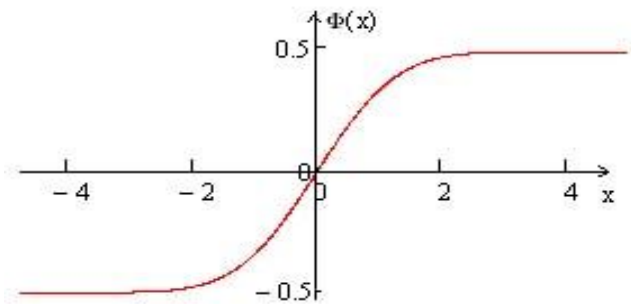


Рис.2.2. Интегральная функция Лапласа.

Пример 21. Вероятность того, что деталь не прошла проверку ОТК, равна $p = 0,2$. Найти вероятность того, что среди 400 случайно отобранных деталей окажется непроверенных от 70 до 100 деталей.

Решение. При вычислении будем использовать формулу (1.20), предварительно вычислив значения

$$x_1 = \frac{70 - 400 \cdot 0,2}{\sqrt{400 \cdot 0,2 \cdot 0,8}} = -1,25, \quad x_2 = \frac{100 - 400 \cdot 0,2}{\sqrt{400 \cdot 0,2 \cdot 0,8}} = 2,5. \quad \text{По приложению 2 находим}$$

$$\Phi(-1,25) = -\Phi(1,25) = -0,3944, \quad \Phi(2,5) = 0,4938. \quad \text{Тогда}$$

$$P_{400}(70,100) = \Phi(x_2) - \Phi(x_1) = \Phi(2,5) + \Phi(1,25) = 0,8882.$$

Следствия интегральной формулы Муавра-Лапласа

Следствие 1. Вероятность того, что в n независимых испытаниях абсолютная величина отклонения числа наступления события A (m) от среднего числа np

появления события A не превзойдёт положительного числа ε , находится по формуле

$$P(|m - np| < \varepsilon) \approx 2\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right). \quad (1.21)$$

Следствие 2. Вероятность того, что в n независимых испытаниях абсолютная величина отклонения относительной частоты $\frac{m}{n}$ наступления события A от его вероятности p не более чем на некоторое число ε , находится по формуле

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) \approx 2\Phi\left(\varepsilon \sqrt{\frac{n}{pq}}\right). \quad (1.22)$$

Пример 22. Мастерская по гарантийному ремонту телевизоров обслуживает 2000 абонентов. Вероятность того, что купленный телевизор потребует гарантийного ремонта, равна 0,3. Предполагая, что событие, вероятность которого 0,9973, достоверно, найти границы числа телевизоров, которые потребуют гарантийного ремонта.

Решение. $n = 2000$, $p = 0,3$, $q = 0,7$, $np = 600$. По формуле (1.21) находим

$$P_{2000}(|m - 600| < \varepsilon) \approx 2\Phi\left(\frac{\varepsilon}{\sqrt{2000 \cdot 0,3 \cdot 0,7}}\right) = 0,9973.$$

$$\Phi\left(\frac{\varepsilon}{20,49}\right) = 0,49865, \text{ по таблицам (приложение 2) находим аргумент } x, \text{ при}$$

$$\text{котором значение функции } \Phi(x) = 0,49865. \text{ Получаем } x = 3 \Rightarrow \frac{\varepsilon}{20,49} = 3 \Rightarrow$$

$$\varepsilon = 61,47 \approx 62. \text{ Следовательно, } |m - np| \leq 62 \text{ и } 538 \leq m \leq 662.$$

Пример 23. Вероятность поражения мишени при одном выстреле равна 0,6. Найти такое число выстрелов по мишени, при котором с вероятностью 0,993 можно ожидать, что отклонения относительно частоты попаданий от вероятности 0,6 по абсолютной величине не превзойдет 0,03.

Решение. Имеем $p = 0,6$, $\varepsilon = 0,03$, тогда

$$P\left(\left|\frac{m}{n} - 0,6\right| \leq 0,03\right) = 2\Phi\left(0,03 \sqrt{\frac{n}{0,6 \cdot 0,4}}\right) = 0,993,$$

откуда получаем

$$\Phi(0,0612 \sqrt{n}) = 0,4965.$$

По приложению 2 находим аргумент функции Лапласа, при котором ее значение равно 0,4965. $x = 2,7. \Rightarrow 0,0612\sqrt{n}=2,7 \Rightarrow \sqrt{n} \approx 44,09 \Rightarrow n \geq 1944$.

2. Случайные величины

Наряду с понятием случайного события и вероятности понятие случайной величины является важнейшим в теории вероятностей. Случайная величина – это числовая функция, определенная на множестве элементарных событий Ω , значения которой есть вещественные числа. Обозначают случайные величины большими последними буквами латинского алфавита X, Y, Z .

Случайными величинами, например, являются: X – число выпавших очков при бросании игральной кости, Y – число вызовов на телефонной станции за некоторый промежуток времени, Z – прибыль или убытки фирмы и т. д.

2.1. Дискретная случайная величина

Дискретной называется случайная величина, значения которой образуют конечную или бесконечную последовательность чисел $x_1, x_2, \dots, x_n, \dots$. Например, случайная величина X – число дней в году – дискретная (она принимает два значения 365;366); случайная величина Y – число испытаний, которое надо сделать до первого попадания в цель (она принимает бесконечное число значений 1, 2, 3 и т.д.).

2.1.1. Закон распределения дискретной случайной величины

Рассмотрим дискретную случайную величину X с возможными значениями $x_1, x_2, x_3, \dots, x_n$. В результате опыта случайная величина примет только одно из этих значений. Введем события $A_i = \{X = x_i\}$, образующие полную группу попарно несовместных событий. Вероятность того, что событие A_i наступит, обозначим через $p_i = P(X = x_i)$.

Законом распределения дискретной случайной величины X называется соотношение между ее возможными значениями x_i и их вероятностями p_i . Закон распределения дискретной случайной величины X можно задать в виде формулы,

например, формулы Бернулли, Пуассона и др. Удобнее задавать закон дискретной случайной величины в виде таблицы

Таблица 2.1. Закон распределения дискретной случайной величины

x_i	x_1	x_2	$\dots$	x_n
p_i	p_1	p_2	$\dots$	p_n

Так как события A_i образуют полную группу попарно несовместных событий, то по формуле (1.8) имеем

$$p_1 + p_2 + \dots + p_n = \sum_{i=1}^n p_i = 1. \quad (2.1)$$

Если случайная величина принимает бесконечное число значений

$x_1, x_2, x_3, \dots, x_n, \dots$, то

$$\sum_{i=1}^{\infty} p_i = 1. \quad (2.2)$$

Графическое представление таблицы 2.1 называется многоугольником распределения вероятностей: фигуры, состоящей из точек (x_i, p_i) , соединенных отрезками (рис.2.3).

Пример 2.1. Стрелок производит три выстрела в мишень. Вероятность попадания в цель при каждом выстреле одинакова и равна 0,8. Составить закон распределения случайной величины X – число попаданий в цель при 3-х выстрелах. построить многоугольник распределения вероятностей.

Решение. Случайная величина может принимать четыре значения: 0, 1, 2 и 3. Вероятности $P(X = x_i)$ будем искать по формуле Бернулли (1.16). Имеем $n = 3$, $p = 0,8$, вероятность того, что стрелок промахнулся, равна $q = 1 - p = 0,2$. Тогда

$$P(X = 0) = P_3(0) = C_3^0 p^0 q^3 = 0,8^0 \cdot 0,2^3 = 0,008.$$

$$P(X = 1) = P_3(1) = C_3^1 p^1 q^2 = 3 \cdot 0,8^1 \cdot 0,2^2 = 0,096.$$

$$P(X = 2) = P_3(2) = C_3^2 p^2 q^1 = 3 \cdot 0,8^2 \cdot 0,2^1 = 0,384.$$

$$P(X = 3) = P_3(3) = C_3^3 p^3 q^0 = 0,8^3 \cdot 0,2^0 = 0,512.$$

x_i	0	1	2	3
p_i	0,008	0,096	0,384	0,512

Контроль: равенство (2.1)
выполняется.

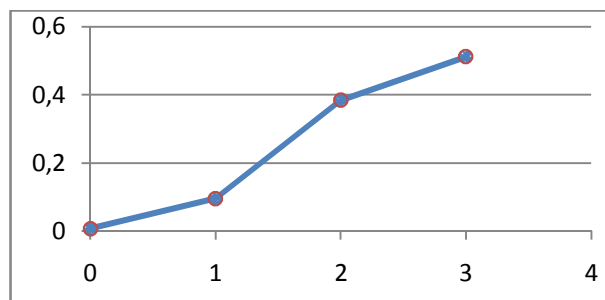


Рис.2.3. Многоугольник распределения вероятностей.

2.1.2. Функции распределения

Закон распределения дискретной случайной величины можно задать функцией распределения $F(x)$.

Функцией распределения случайной величины X называется функция $F(x)$, которая для любого вещественного числа x равна вероятности того, что случайная величина X примет значение, меньшее x

$$F(x) = P(X < x). \quad (2.3)$$

Функция распределения дискретной случайной величины – ступенчатая функция, определенная на всей числовой оси.

Свойства функции распределения

1. Функция распределения принимает значения только из интервала $[0; 1]$:
 $0 \leq F(x) \leq 1$.
2. $F(x)$ - неубывающая функция, т.е. если $x_1 < x_2$, то $F(x_1) \leq F(x_2)$.
3. Вероятность того, что случайная величина примет значение из интервала $[a; b)$ равна $P(a \leq X < b) = F(b) - F(a)$.
4. $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$.
5. $F(x)$ - непрерывна слева, т.е. $\lim_{x \rightarrow x_0 - 0} F(x) = F(x_0)$.

Пример 2.2. В условиях задачи 2.1 составить функцию распределения $F(x)$ и построить её график.

Решение. Очевидно, что вероятность того, что случайная величина примет значение меньшее, чем 0 равна нулю, $F(0) = P(X < 0) = 0$. Найдем значения функции $F(x)$ в точках $x = 1$, $x = 2$ и $x = 3$.

$$F(1) = P(X < 1) = 0,008;$$

$$F(2) = P(X < 2) = P(X = 0) + P(X = 1) = 0,008 + 0,096 = 0,104;$$

$$F(3) = P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = 0,008 + 0,096 + 0,384 = 0,488.$$

Тогда функция распределения $F(x)$ имеет вид

$$F(x) = \begin{cases} 0, & \text{при } x \leq 0, \\ 0,008, & \text{при } 0 < x \leq 1, \\ 0,104, & \text{при } 1 < x \leq 2, \\ 0,488, & \text{при } 2 < x \leq 3, \\ 1, & \text{при } x > 3. \end{cases}$$

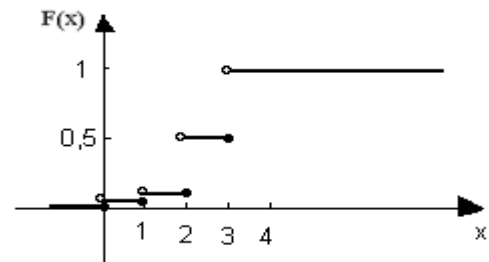


Рис.2.4. Функция распределения дискретной случайной величины.

2.1.3. Действия над случайными величинами

Пусть законы распределения двух *независимых* случайных величин X и Y известны. Пусть случайная величина X принимает значения x_i с вероятностями $p_i = P(X = x_i)$, а случайная величина Y принимает значения y_j с вероятностями $p_j = P(Y = y_j)$.

Произведением случайной величины X на **постоянное число** C называется такая случайная величина Z , значения которой получаются в результате умножения всех значений случайной величины X на число C , соответствующие вероятности при этом не меняются.

Суммой двух случайных величин X и Y , называется такая случайная величина Z , значения которой получаются в результате сложения всех возможных значений случайной величины X и всех возможных значений случайной величины Y , т.е. $z_{ij} = x_i + y_j$, соответствующие вероятности при этом перемножаются, т.е.

$$P(Z = z_{ij}) = P(X = x_i)P(Y = y_j).$$

Произведением двух случайных величин X и Y называется такая случайная величина Z , значения которой получаются в результате произведения всех возможных значений случайной величины X и всех возможных значений случайной величины Y , т.е. $z_{ij} = x_i y_j$, соответствующие вероятности при этом перемножаются, т.е. $P(Z = z_{ij}) = P(X = x_i)P(Y = y_j)$.

Замечание. Если при вычислении получаются одинаковые значения z_{ij} , то в таблице записывают только одно, вероятности при этом складываются.

Пример 2.3. Команда состоит из двух стрелков. Первый стрелок выбивает 8, 9, 10 очков с вероятностями 0,1; 0,4; 0,5. Второй стрелок выбивает 9, 10 очков с вероятностями 0,4; 0,6. Результаты стрельбы одного стрелка не влияют на результаты стрельбы второго. Составить закон распределения числа очков, выбиваемых этой командой, если стрелки сделают по одному выстрелу.

Решение. Из условия задачи результаты стрельбы одного стрелка не влияют на результаты стрельбы второго, следовательно, X и Y – независимые случайные величины. Запишем законы распределения случайных величин X и Y .

x_i	8	9	10
p_i	0,1	0,4	0,5

y_i	9	10
g_i	0,4	0,6

Пусть $Z = X + Y$. Тогда по определению суммы двух независимых случайных величин имеем

z_i	17	18	18	19	19	20
P_i	0,04	0,06	0,16	0,24	0,20	0,30

Значение $z=18$ встречается два раза с вероятностями 0,06 или 0,16, поэтому в таблицу записываем одно значение $z=18$, вероятности при этом складываются. Аналогично запишем $z=19$ с вероятностью 0,44. Получаем окончательную таблицу

z_i	17	18	19	20
P_i	0,04	0,22	0,44	0,30

2.1.4. Числовые характеристики дискретной случайной величины

Математическое ожидание случайной величины X - это число, равное сумме произведений всех значений случайной величины x_i на соответствующие вероятности p_i

$$M(X) = \sum_{i=1}^n x_i \cdot p_i . \quad (2.4)$$

Математическое ожидание определяет среднее значение случайной величины. Размерность математического ожидания совпадает с размерностью случайной величины X .

Свойства математического ожидания

1. Математическое ожидание постоянной случайной величины C равно самой этой постоянной: $M(C) = C$.
2. Постоянный множитель можно вынести за знак математического ожидания: $M(CX) = CM(X)$.
3. Математическое ожидание суммы случайных величин равно сумме их математических ожиданий: $M(X + Y) = M(X) + M(Y)$.
4. Математическое ожидание произведения *независимых* случайных величин равно произведению их математических ожиданий: $M(XY) = M(X)M(Y)$.

Дисперсия случайной величины X - это число, равное математическому ожиданию квадрата отклонения случайной величины от ее математического ожидания

$$D(X) = M\left((X - M(X))^2\right) = \sum_{i=1}^n (x_i - M(X))^2 \cdot p_i . \quad (2.5)$$

Дисперсия определяет средний квадрат отклонений случайной величины от их среднего значения. Размерность дисперсии равна квадрату размерности случайной величины X .

Из формулы (2.5) можно вывести более удобную формулу для вычисления дисперсии

$$D(X) = M(X^2) - M^2(X) = \sum_{i=1}^n x_i^2 \cdot p_i - M^2(X) . \quad (2.6)$$

Свойства дисперсии

- 1 Дисперсия постоянной случайной величины равна нулю: $D(C) = 0$.
- 2 Постоянный множитель можно выносить за знак дисперсии, предварительно возведя его в квадрат: $D(CX) = C^2 D(X)$.
- 3 Дисперсия суммы *независимых* случайных величин равна сумме их дисперсий: $D(X + Y) = D(X) + D(Y)$.

Для того, чтобы размерность степени рассеяния случайной величины относительно среднего значения совпадала с размерностью самой случайной величины, вводят **среднее квадратическое отклонение**

$$\sigma_x = \sqrt{D(X)}. \quad (2.7)$$

Пример 2.4. В условиях примера 2.3. вычислить числовые характеристики случайных величин X , Y и $Z = X + Y$.

Решение. Согласно формуле (2.4) имеем

$$M(X) = 8 \cdot 0,1 + 9 \cdot 0,4 + 10 \cdot 0,5 = 9,4; \quad M(Y) = 9 \cdot 0,4 + 10 \cdot 0,6 = 9,6;$$

$$M(Z) = 17 \cdot 0,04 + 18 \cdot 0,22 + 19 \cdot 0,44 + 20 \cdot 0,30 = 19.$$

Дисперсию вычислим по формуле (2.6)

$$D(X) = 8^2 \cdot 0,1 + 9^2 \cdot 0,4 + 10^2 \cdot 0,5 - 9,4^2 = 0,44; \quad D(Y) = 9^2 \cdot 0,4 + 10^2 \cdot 0,6 - 9,6^2 = 0,24;$$

$$D(Z) = 17^2 \cdot 0,04 + 18^2 \cdot 0,22 + 19^2 \cdot 0,44 + 20^2 \cdot 0,30 - 19^2 = 0,68.$$

Из формулы (2.7) находим средние квадратические отклонения

$$\sigma_x = \sqrt{0,44} = 0,663; \quad \sigma_y = \sqrt{0,24} = 0,49; \quad \sigma_z = \sqrt{0,68} = 0,825.$$

Т.е. в среднем первый стрелок выбивает 9,4 очка, второй – 9,6 очков, а вместе - 19 очков. По средним значениям можно судить, что второй стрелок более меткий. По средним квадратическим отклонениям также можно судить, что второй стрелок лучше стреляет, так как $\sigma_y < \sigma_x$ и $\sigma_y < \sigma_z$.

Видно, что свойство математического ожидания $M(X + Y) = M(X) + M(Y)$ выполнено: $M(X + Y) = 19$, $M(X) = 9,4$, $M(Y) = 9,6$. Кроме того, поскольку случайные величины X и Y - независимы, то свойство дисперсии $D(X + Y) = D(X) + D(Y)$ - так же имеет место. Этими свойствами можно пользоваться для контроля вычислений или для нахождения $M(X + Y)$ и $D(X + Y)$ без составления закона распределения случайной величины $X + Y$.

Пример 2.5. В ящике находятся 5 белых и 15 черных шаров одинаковых на ощупь. Случайным образом вынули сразу три шара. Найти математическое ожидание случайной величины X - число белых шаров среди трех отобранных.

Решение. Случайная величина X может принимать четыре значения 0, 1, 2 и 3. Вероятности вычислим по классическому определению вероятности (1.4). Число всевозможных исходов равно C_{20}^3 . Число благоприятствующих появлению события $A = \{\text{среди трех отобранных шаров ровно } k \text{ белых шаров}\}$ равно произведению $C_5^k \cdot C_{15}^{3-k}$. Тогда получим

$$P_3(0) = \frac{C_5^0 C_{15}^3}{C_{20}^3} = \frac{91}{228}; \quad P_3(1) = \frac{C_5^1 C_{15}^2}{C_{20}^3} = \frac{35}{76};$$

$$P_3(2) = \frac{C_5^2 C_{15}^1}{C_{20}^3} = \frac{5}{38}; \quad P_3(3) = \frac{C_5^3 C_{15}^0}{C_{20}^3} = \frac{1}{114}.$$

Таким образом, закон распределения случайной величины X имеет вид

x_i	0	1	2	3
p_i	$\frac{91}{228}$	$\frac{35}{76}$	$\frac{5}{38}$	$\frac{1}{114}$

Для контроля используем формулу (2.1).

В заключение, по формуле (2.4) находим математическое ожидание

$$M(X) = 0 \cdot \frac{91}{228} + 1 \cdot \frac{35}{76} + 2 \cdot \frac{5}{38} + 3 \cdot \frac{1}{114} = 0,75.$$

Кроме перечисленных числовых характеристик $M(X)$, $D(X)$ и $\sigma(X)$ существуют и другие. К таким относятся начальные моменты и центральные моменты.

Начальным моментом k - того порядка случайной величины X называется величина равная математическому ожиданию случайной величины X^k , т.е.

$$\nu_k = M(X^k) = \sum_{i=1}^n x_i^k \cdot p_i. \quad (2.8)$$

Из формулы (2.8) следует, что математическое ожидание - это начальный момент первого порядка.

Центральным моментом k - того порядка случайной величины X называется величина равная математическому ожиданию центрированной случайной величины в степени $(X - M(X))^k$, т.е.

$$\mu_k = M(X - M(X))^k = \sum_{i=1}^n (x_i - M(X))^k \cdot p_i. \quad (2.9)$$

Из формулы (2.9) следует, что дисперсия - это центральный момент второго порядка.

2.1.5. Основные законы распределения дискретных случайных величин

Биномиальный закон распределения

Пусть проводится n независимых испытаний, в каждом из которых событие A может произойти с одной и той же вероятностью p . Случайная величина X - это число испытаний, в которых событие A наступило. Вероятность того, что событие A наступит ровно k раз в n испытаниях вычисляется по формуле Бернулли (1.16). Биномиальный закон распределения вероятностей случайной величины X имеет вид

x_i	0	1	...	k	...	n
p_i	$C_n^0 p^0 q^{n-0}$	$C_n^1 p^1 q^{n-1}$	...	$C_n^k p^k q^{n-k}$	...	$C_n^n p^n q^0$

Здесь $q = 1 - p$.

Математическое ожидание: $M(X) = np$. Дисперсия: $D(X) = npq$.

Среднее квадратическое отклонение: $\sigma(X) = \sqrt{npq}$.

Биномиальный закон распределения – однопараметрическое распределение, задаётся параметром p .

Закон распределения Пуассона

Пусть проводится большое количество n независимых испытаний, в каждом из которых появление события A одинаково и маловероятно ($p \rightarrow 0$). Случайная величина X - это число испытаний, в которых событие A наступило. Вероятность того, что событие A наступит ровно k раз в n испытаниях вычисляется по формуле Пуассона (1.19). Закон распределения Пуассона имеет вид

x_i	0	1	...	k	...
p_i	$\frac{e^{-\lambda} \cdot \lambda^0}{0!}$	$\frac{e^{-\lambda} \cdot \lambda^1}{1!}$	...	$\frac{e^{-\lambda} \cdot \lambda^k}{k!}$	...

Здесь $\lambda = np$.

Математическое ожидание: $M(X) = \lambda$. Дисперсия: $D(X) = \lambda$.

Среднее квадратическое отклонение: $\sigma(X) = \sqrt{\lambda}$.

Закон распределения Пуассона задаётся одним параметром λ .

Геометрический закон распределения

Пусть вероятность появления события A в каждом испытании одинакова и равна p . Вероятность того, что событие A не наступит в каждом испытании так же одинакова и равна $q = 1 - p$. Испытания проводятся до первого появления события A . Случайная величина X - это количество проведенных испытаний. Вероятность того, что событие A наступит с первого раза равна p , со второго - qp и т.д. Геометрический закон распределения имеет вид

x_i	1	2	3	...	k	...
p_i	p	qp	$q^2 p$	...	$q^{k-1} p$	...

Математическое ожидание: $M(X) = \frac{1}{p}$. Дисперсия: $D(X) = \frac{q}{p^2}$.

Среднее квадратическое отклонение: $\sigma(X) = \frac{\sqrt{q}}{p}$.

Геометрический закон распределения задаётся одним параметром p .

2.2. Непрерывная случайная величина

Непрерывная случайная величина принимает множество значений вещественной прямой.

Закон распределения непрерывной случайной величины задается либо интегральной функцией распределения, либо дифференциальной функцией распределения.

2.2.1. Дифференциальная и интегральная функции распределения

Функция распределения непрерывной случайной величины определяется также как и для дискретной - формулой (2.3). Функция распределения непрерывной случайной величины – непрерывная.

Дифференциальная функции распределения связана с интегральной равенством

$$f(x) = F'(x) . \quad (2.10)$$

И наоборот, если известна дифференциальная функции распределения, то можно найти интегральную функцию распределения по формуле

$$F(x) = \int_{-\infty}^x f(t)dt . \quad (2.11)$$

Дифференциальную функцию распределения $f(x)$ еще называют функцией плотности вероятности.

Свойства интегральной функции распределения такие же, как и свойства 1-4 для функции распределения дискретной случайной величины.

Свойства дифференциальной функции распределения

1. $f(x) \geq 0$, так как $F(x)$ - неубывающая функция.

$$2. \int_{-\infty}^{+\infty} f(x)dx = 1 .$$

$$3. P(a \leq X \leq b) = \int_a^b f(x)dx .$$

2.2.2. Числовые характеристики непрерывной случайной величины

Математическое ожидание

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx ; \quad (2.12)$$

Дисперсия

$$D(X) = M(X - M(X))^2 = \int_{-\infty}^{\infty} (x - M(X))^2 f(x)dx . \quad (2.13)$$

или после преобразования

$$D(X) = M(X^2) - (M(X))^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - (M(X))^2 . \quad (2.14)$$

Среднее квадратическое отклонение

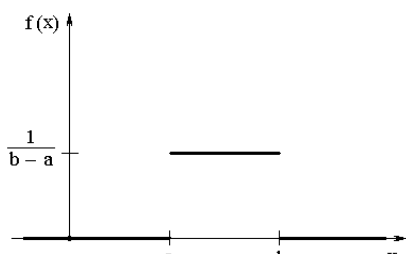
$$\sigma(X) = \sqrt{D(X)} . \quad (2.15)$$

Свойства математического ожидания, дисперсии и среднего квадратического отклонения непрерывных случайных величин аналогичны свойствам математического ожидания и дисперсии дискретных случайных величин.

2.2.3 Основные законы распределения непрерывных случайных величин

Равномерный закон распределения

Непрерывная случайная величина X имеет равномерный закон распределения, на отрезке $[a, b]$, если её плотность вероятности $f(x)$ постоянна на этом отрезке и равна нулю вне его.

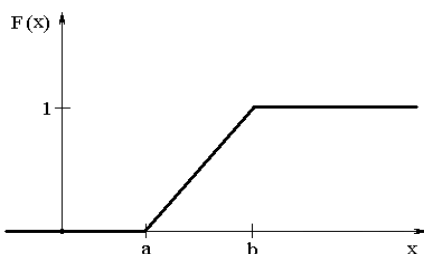


$$f(x) = \begin{cases} 0, & \text{если } x \leq a, \\ \frac{1}{b-a}, & \text{если } a < x \leq b, \\ 0, & \text{если } x > b. \end{cases} \quad (2.16)$$

Рис.2.4. Дифференциальная функция распределения случайной величины, распределенной по равномерному закону.

График дифференциальной функции (2.16) равномерного распределения изображен на рис. 2.4.

Интегральная функция равномерного распределения задается в виде (2.17), ее график изображен на рис.2.5.



$$F(x) = \begin{cases} 0, & \text{если } x \leq a, \\ \frac{x-a}{b-a}, & \text{если } a < x \leq b, \\ 1, & \text{если } x > b. \end{cases} \quad (2.17)$$

Рис.2.5. Интегральная функция распределения случайной величины, распределенной по равномерному закону.

Числовые характеристики:

Математическое ожидание: $M(X) = \frac{a+b}{2}$. Дисперсия: $D(x) = \frac{(b-a)^2}{12}$

Среднее квадратическое отклонение: $\sigma(x) = \frac{b-a}{2\sqrt{3}}$.

Вероятность попадания случайной величины в интервал $[a, b]$ равна

$$P(a \leq X \leq b) = F(b) - F(a).$$

Равномерное распределение имеет 2 параметра - a и b .

Пример 2.5. Цена деления шкалы амперметра равна 0,1 ампера. Показания амперметра округляют до ближайшего целого деления. Найти вероятность того, что при вычислениях будет сделана ошибка, не превышающая по абсолютной величине 0,02 ампера.

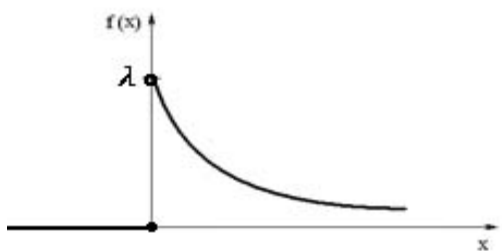
Решение. Ошибку округления отсчета можно рассматривать, как случайную величину X , которая распределена равномерно между двумя соседними целыми делениями. Все значения случайной величины заключены в интервале $[a; b]$, где $a = -0,05$, $b = 0,05$. Используя формулу (2.16), получаем функцию плотности

$$f(x) = \begin{cases} 0, & x \leq -0,05, \\ 10, & -0,05 < x \leq 0,05, \\ 0, & x > 0,05. \end{cases}$$

$$\text{Тогда } P(|x| \leq 0,02) = \int_{-0,02}^{0,02} 10 dx = 10x \Big|_{-0,02}^{0,02} = 0,4.$$

Показательный (экспоненциальный) закон распределения

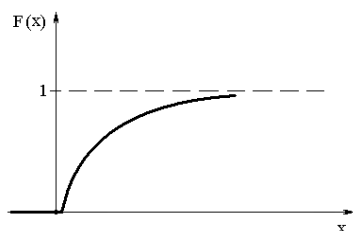
Непрерывная случайная величина X имеет показательный (экспоненциальный) закон распределения с параметром λ , если плотность вероятности $f(x)$ задана функцией (2.18).



$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & \text{если } x \geq 0, \\ 0, & \text{если } x < 0. \end{cases} \quad (2.18)$$

Рис.2.6. Дифференциальная функция распределения случайной величины, распределенной по показательному закону.

Интегральная функция распределения имеет вид



$$F(x) = \begin{cases} 0, & \text{если } x < 0, \\ 1 - e^{-\lambda x}, & \text{если } x \geq 0. \end{cases} \quad (2.19)$$

Рис.2.7. Интегральная функция распределения случайной величины, распределенной по показательному закону.

Математическое ожидание: $M(X) = \frac{1}{\lambda}$. Дисперсия: $D(X) = \frac{1}{\lambda^2}$

Среднее квадратическое отклонение: $\sigma(X) = \frac{1}{\lambda}$.

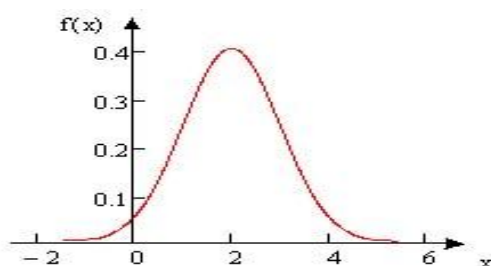
Вероятность попадания случайной величины в интервал $[a, b]$ равна

$$P(a \leq X \leq b) = F(b) - F(a) = e^{-\lambda a} - e^{-\lambda b}. \quad (2.20)$$

Показательный закон распределения имеет 1 параметр - λ .

Нормальный закон распределения

Нормальное распределение вероятностей задается функцией плотности распределения



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-a)^2}{2\sigma^2}}. \quad (2.21)$$

Рис.2.8. Дифференциальная функция распределения случайной величины, распределенной по нормальному закону при $a = 2$, $\sigma = 1$.

Математическое ожидание: $M(X) = a$. Дисперсия: $D(X) = \sigma^2$.

Среднее квадратическое отклонение: σ .

Нормальный закон распределения имеет два параметра – a и σ .

Для удобства вычисления значений $f(x)$ ее можно записать через функцию $\varphi(x)$ следующим образом

$$f(x) = \frac{1}{\sigma} \varphi\left(\frac{x-a}{\sigma}\right), \quad (2.22)$$

и использовать таблицы приложения 1.

Интегральная функция распределения имеет вид

$$F(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt. \quad (2.23)$$

Для вычисления значений $F(x)$ удобнее ее записать через функцию Лапласа $\Phi(x)$

$$F(x) = \frac{1}{2} + \Phi\left(\frac{x-a}{\sigma}\right), \quad (2.24)$$

и использовать таблицы приложения 2.

Вероятность попадания случайной величины X , распределенной нормально, в интервал $[\alpha, \beta]$, находится по формуле

$$P(\alpha \leq X \leq \beta) = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right). \quad (2.25)$$

Правило «трёх сигм»: практически все значения нормально распределённой случайной величины лежат в интервале $[a - 3\sigma; a + 3\sigma]$. Так как по формуле (2.25) имеем $P(a - 3\sigma \leq X \leq a + 3\sigma) = 0,9973$, поэтому не менее чем с достоверностью 99,7%, все значения нормально распределенной случайной величины лежат в указанном интервале.

Нормально распределенные случайные величины играют важнейшую роль во многих областях знаний. Физическая величина часто подчиняется нормальному распределению, когда она подвержена влиянию значительного числа случайных помех. Ясно, что такая ситуация крайне распространена, поэтому можно сказать, что из всех распределений в природе чаще всего встречается именно нормальное распределение — отсюда и произошло его название.

Пример 2.6. Закон распределения случайной величины X задается дифференциальной функцией $f(x) = \frac{1}{3\sqrt{2\pi}} e^{-\frac{(x-2)^2}{18}}$. Найти: а) числовые характеристики случайной величины; б) интегральную функцию распределения; в) интервал, в который с вероятностью 0,96 попадет случайная величина в результате испытаний.

Решение. а) Сравнивая данную дифференциальную функцию с формулой (2.21), можно заключить, что случайная величина X распределена по нормальному закону с математическим ожиданием $M(X) = a = 2$, и средним квадратическим отклонением $\sigma = 3$. Тогда дисперсия равна $D(X) = \sigma^2 = 9$.

б) Из (2.23) находим интегральную функцию распределения

$$F(x) = \frac{1}{3\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-2)^2}{18}} dt.$$

в) По формуле (2.25) находим интервал, в который с вероятностью 0,96 попадет случайная величина. Возьмем в качестве $\alpha = a - \Delta$, а $\beta = a + \Delta$.

$$\begin{aligned} \text{Тогда } P(2 - \Delta < X < 2 + \Delta) &= \Phi\left(\frac{2+\Delta-2}{3}\right) - \Phi\left(\frac{2-\Delta-2}{3}\right) = \Phi\left(\frac{\Delta}{3}\right) - \Phi\left(-\frac{\Delta}{3}\right) = \\ &= 2\Phi\left(\frac{\Delta}{3}\right) = 0,96. \end{aligned}$$

Откуда $\Phi\left(\frac{\Delta}{3}\right) = 0,48$. По таблицам приложения 2 по значению функции Лапласа $\Phi(x) = 0,48$ находим $x = 1,75$. $\Rightarrow \frac{\Delta}{3} = x$. $\Rightarrow \Delta = 3 \cdot 1,75$. $\Rightarrow \Delta = 5,25$. Окончательно получаем интервал $[-3,25; 7,25]$, в который с вероятностью 0,96 случайная величина X попадет в результате испытания.

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Математическая статистика - наука, изучающая методы обработки результатов наблюдений массовых случайных явлений. Основной целью математической статистики при решении задач планирования, прогнозирования и т.д. является сбор статистических данных, их обработка, исследование и принятие решения.

3. Обработка результатов статистических наблюдений

3.1. Вариационный ряд

Пусть требуется изучить некоторую случайную величину, закон распределения которой неизвестен. Требуется приближенно определить этот закон по опытным данным и проверить гипотезу о том, что случайная величина X подчинена этому закону.

В результате *случайного* отбора из всей совокупности объектов (**генеральной совокупности**) можно получить **выборку** (числовые данные). Количество объектов выборки, называемое **объемом выборки**, будем обозначать через n . Для того, чтобы по выборке можно было уверенно судить о генеральной совокупности, выборка должна быть **репрезентативной** (хорошо отражать свойства генеральной совокупности). Репрезентативность улучшается при увеличении объема выборки.

Перед обработкой статистических данных, если число значений признака велико ($n \geq 25$), их упорядочивают, группируют и строят интервальный вариационный ряд. Существует много способов группировки данных.

Предложим самый простой. Просматривая данные наблюдений, находят наименьшее значение выборки X_{min} и наибольшее - X_{max} . Затем находят размах варьирования по формуле

$$R = X_{max} - X_{min}. \quad (3.1)$$

Для построения интервального вариационного ряда необходимо весь интервал $[X_{min}, X_{max}]$ разбить на m частичных интервалов. Для определения числа m по заданному значению объема выборки n можно воспользоваться **формулой Стерджеса**

$$m \geq 1 + [3,322 \cdot \lg n], \quad (3.2)$$

где квадратные скобки - целая часть числа. Для нахождения ориентировочного значения количества интервалов m можно использовать следующую таблицу

n	25-30	30-60	60-100	100-125	125-200
m	5-6	6-7	7-8	8-9	9-10

Считая, что все частичные интервалы имеют одинаковую длину, для каждого интервала следует установить его верхнюю и нижнюю границы, а затем в

соответствии с полученной упорядоченной совокупностью частичных интервалов сгруппировать результаты наблюдений.

Длину частичного интервала (шаг) определяем по формуле

$$h = \frac{X_{max} - X_{min}}{m}. \quad (3.3)$$

Промежуточные интервалы получают, прибавляя к концу предыдущего интервала длину частичного интервала h , обозначим границы интервалов через X_i , тогда $X_0 = X_{min}$, $X_1 = X_{min} + h$, $X_2 = X_{min} + 2h, \dots, X_m = X_{min} + mh$. Затем подсчитывают количество наблюдаемых значений (частоту n_i), которые попали в i -тый интервал, причем $\sum_{i=1}^m n_i = n$. Все вычисления записывают в виде таблицы.

Таблица 3.1. Интервальный вариационный ряд

$[X_i, X_{i+1}]$	$[X_0, X_1]$	$[X_1, X_2]$		$[X_{m-1}, X_m]$	Итого
n_i	n_1	n_2	$\dots$	n_m	n

Замечание. Если значения случайной величины X (варианты) попали на границу интервала $[X_i, X_{i+1}]$, то их количество делят пополам и распределяют половину значений в левый интервал, половину - в правый.

Из таблицы 3.1. можно получить вариационный ряд, если заменить первую строку на значения *середин* каждого интервала (обозначим середины через x_i).

Таблица 3.2. Вариационный ряд

x_i	x_1	x_2	x_3	$\dots$	x_m	Итого
n_i	n_1	n_2	n_3	$\dots$	n_m	n

3.2. Эмпирическая функция распределения

Эмпирической функцией распределения (функцией распределения выборки) называют функцию $F^*(x)$, которая равна относительной частоте появления события $\{X < x\}$, найденную по данной выборке

$$F(x) = \frac{n(x)}{n}, \quad (3.4)$$

где $n(x)$ - количество вариантов, меньших x ; n - объем выборки.

Для нахождения эмпирической функции распределения выборки удобнее сначала вычислить относительные частоты попадания случайной величины в i -тый интервал

$$w_i = \frac{n_i}{n}. \quad (3.5)$$

Затем вычислить накопленные относительные частоты $n(x_i)$ - сумму относительных частот $w_1 + w_2 + \dots + w_{i-1}$. Результаты вычислений записывают в виде таблицы

Таблица 3.3. Распределение относительных частот

x_i	x_1	x_2	x_3	x_4	$\dots$	x_m	
w_i	w_1	w_2	w_3	w_4	$\dots$	w_m	$\sum_{i=1}^m w_i = 1$
Накопленные относительные частоты $n(x_i)$	0	w_1	$w_1 + w_2$	$w_1 + w_2 + w_3$		$w_1 + w_2 + \dots + w_{m-1}$	1, при $X > x_m$

Из таблицы 3.3 находим эмпирическую функцию распределения

$$F^*(x) = \begin{cases} 0, & \text{если } x \leq x_1, \\ w_1, & \text{если } x_1 < x \leq x_2, \\ w_1 + w_2, & \text{если } x_2 < x \leq x_3, \\ w_1 + w_2 + w_3, & \text{если } x_3 < x \leq x_4, \\ w_1 + w_2 + w_3 + w_4, & \text{если } x_4 < x \leq x_5, \\ \dots & \dots \\ 1, & \text{если } x > x_m. \end{cases} \quad (3.6)$$

Эмпирическая функция распределения обладает теми же свойствами, как и функция распределения дискретной случайной величины.

3.3. Полигон и гистограмма

Используя таблицы (3.2) и (3.3) можно построить полигон частот и полигон относительных частот. **Полигоном частот** называют ломаную, соединяющую точки с координатами $(x_1, n_1), (x_2, n_2), \dots, (x_m, n_m)$, **полигоном относительных частот** – с координатами $(x_1, w_1), (x_2, w_2), \dots, (x_m, w_m)$.

Гистограммой частот (относительных частот) называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длины h , а высоты равны отношению $\frac{n_i}{h} \left(\frac{w_i}{h} \right)$. Очевидно, что вся площадь гистограммы частот равна n , а площадь гистограммы относительных частот равна единице.

Гистограмма относительных частот, построенная на основе выборки, может служить оценкой неизвестной плотности вероятностей. При увеличении объема выборки и уменьшении длин интервалов гистограмма относительных частот приближается к графику неизвестной функции плотности вероятности генеральной совокупности $f(x)$.

По виду гистограммы или полигона частот можно выдвинуть гипотезу о виде распределения генеральной совокупности.

3.4. Числовые характеристики статистического распределения

Выборочной средней $\bar{X}_B$ называют среднее арифметическое значений признака выборочной совокупности. Пусть получено статистическое распределение (табл. 3.2). Тогда выборочное среднее равно

$$\bar{X}_B = \frac{\sum_{i=1}^m x_i \cdot n_i}{n} . \quad (3.7)$$

Выборочная средняя имеет те же единицы измерения, что и варианты x_i .

Выборочной дисперсией D_B называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\bar{X}_B$. Выборочную дисперсию вычисляется по формуле

$$D_B = \frac{\sum_{i=1}^m (x_i - \bar{X}_B)^2 \cdot n_i}{n} . \quad (3.8)$$

Более удобная формула для вычисления выборочной дисперсии

$$D_B = \frac{\sum_{i=1}^m x_i^2 \cdot n_i}{n} - (\bar{X}_B)^2 . \quad (3.9)$$

Кроме дисперсии для характеристики рассеяния значений признака выборочной совокупности вокруг своего среднего значения пользуются средним квадратическим отклонением.

Выборочное среднее квадратическое отклонение равно

$$\sigma_B = \sqrt{D_B} . \quad (3.10)$$

Выборочное среднее квадратическое отклонение имеет те же единицы измерения, что и варианты x_i .

3.5. Точечные оценки параметров распределения. Метод моментов

Метод моментов точечной оценки неизвестных параметров заданного распределения состоит в приравнивании теоретических моментов соответствующим эмпирическим моментам того же порядка. Согласно закону больших чисел, при увеличении объема выборки среднее арифметическое выборки стремится к математическому ожиданию генеральной совокупности. Поэтому выборочная средняя может служить оценкой математического ожидания генеральной совокупности.

Если распределение определяется двумя параметрами, то приравнивают два теоретических момента двум соответствующим эмпирическим моментам того же порядка. Например, можно приравнять начальный теоретический момент первого порядка (математическое ожидание) начальному эмпирическому моменту первого порядка (выборочному среднему) и центральный теоретический момент второго порядка (дисперсию) центральному эмпирическому моменту второго порядка (выборочной дисперсии)

$$\begin{cases} M(X) = \bar{X}_B \\ D(X) = D_B \end{cases} . \quad (3.11)$$

Левые части этих равенств являются функциями от неизвестных параметров, а правые части – числовые характеристики, которые вычислены в п.3.4, поэтому, решив систему (3.11) относительно неизвестных параметров, получают их точечные оценки.

Таким образом, зная числовые характеристики нормального распределения $\alpha = M(X)$, $\sigma = \sqrt{D(X)}$, получаем искомые **точечные оценки параметров нормального распределения** (если $n \geq 25$)

$$\alpha = \bar{X}_B, \sigma = \sqrt{D_B}. \quad (3.12)$$

Если объем выборки мал ($n < 25$), то выборочная дисперсия значительно отличается от дисперсии генеральной совокупности нормально распределенной случайной величины, тогда в качестве точечной оценки среднего квадратического отклонения σ генеральной совокупности используют **исправленное выборочное среднее квадратическое отклонение**

$$S = \sqrt{\frac{n}{n-1} \cdot D_B}. \quad (3.13)$$

Точечной оценкой параметра **экспоненциального распределения** является следующая

$$\lambda = \frac{1}{\bar{X}_B}. \quad (3.14)$$

И наконец, точечные оценки параметров **равномерного распределения**

$$a = \bar{X}_B - \sqrt{3}\sigma_B, b = \bar{X}_B + \sqrt{3}\sigma_B. \quad (3.15)$$

3.6. Проверка гипотезы о нормальном распределении генеральной совокупности. Критерий согласия Пирсона

Критерием согласия называют критерий проверки гипотезы о предполагаемом законе неизвестного распределения. Есть основные гипотезы и альтернативные. Альтернативная гипотеза – это гипотеза, которая противоречит основной. Наприсмер, основная гипотеза H_0 - генеральная совокупность распределена по нормальному закону, тогда альтернативная гипотеза H_1 - генеральная совокупность не распределена по нормальному закону. Основную гипотезу можно выдвинуть по виду гистограммы частот (или относительных частот). Если гистограмма куполообразная (рис.3.1), то выдвигают гипотезу H_0 - генеральная совокупность распределена по нормальному закону с плотностью распределения (2.21). Если высоты прямоугольников гистограммы приблизительно на одном уровне (рис.3.2), то проверяют гипотезу H_0 - генеральная совокупность распределена по равномерному закону с плотностью распределения (2.16). Если высоты прямоугольников гистограммы резко (экспоненциально) падают с ростом X (рис.3.3), то выдвигают гипотезу H_0 - генеральная совокупность распределена по экспоненциальному закону с плотностью распределения (2.18).

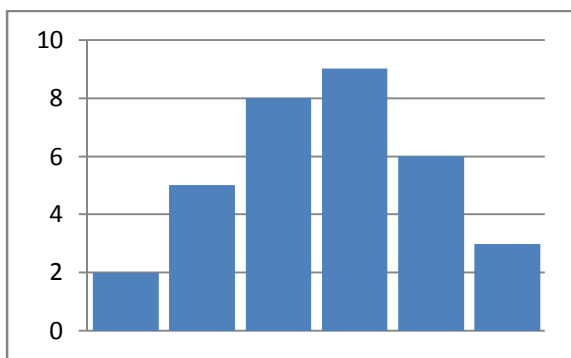


Рис.3.1. Вид гистограммы частот для нормального распределения

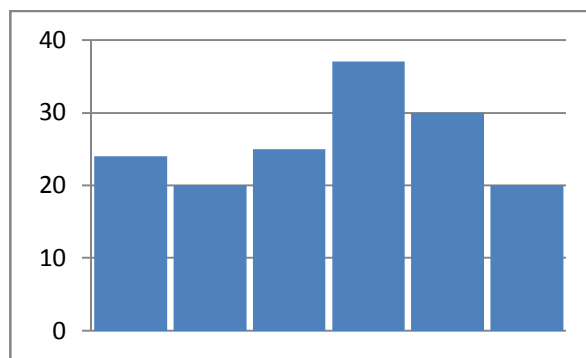


Рис.3.2. Вид гистограммы частот для равномерного распределения

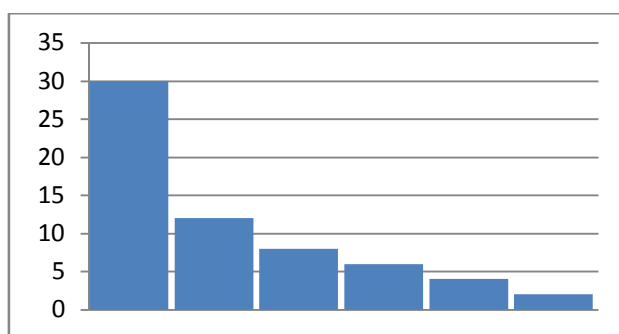


Рис.3.3. Вид гистограммы частот для экспоненциального распределения

Критерий согласия Пирсона: по выборке составляется специальная характеристика $\chi^2_{набл}$ (читается: «хи-квадрат наблюдаемая») и сравнивается с табличным критическим значением $\chi^2_{крит}(\alpha, k)$ (приложение 3). Здесь α - заданный уровень значимости, k - число степеней свободы, который можно вычислить по формуле

$$k = m - r - 1, \quad (3.16)$$

где m - число интервалов, r - число параметров распределения (для нормального распределения два параметра: математическое ожидание и среднее квадратическое отклонение, поэтому $r = 2$). Затем сравниваются значения $\chi^2_{набл}$ и $\chi^2_{крит}(\alpha, k)$.

Если $\chi^2_{набл} < \chi^2_{крит}(\alpha, k)$, то нет оснований отвергнуть нулевую гипотезу.

Если $\chi^2_{набл} > \chi^2_{крит}(\alpha, k)$, то нулевую гипотезу отвергают и принимают альтернативную.

Величина $\chi^2_{\text{набл}}$ зависит от выборки и поэтому является случайной, причем распределена она по закону Хи-квадрат. Вычисляют ее по формуле

$$\chi^2_{\text{набл}} = \sum_{i=1}^m \frac{(n_i - n'_i)^2}{n'_i} \quad (3.17)$$

или по более удобной

$$\chi^2_{\text{набл}} = \sum_{i=1}^m \frac{n_i^2}{n'_i} - n, \quad (3.18)$$

где n_i - эмпирические частоты из таблицы (3.2), n'_i - теоретические частоты, n - объем выборки, m - число частичных интервалов. Теоретические частоты вычисляются по формуле $n'_i = nP'_i$, где P'_i - теоретические вероятности попадания случайной величины X в интервал $[X_i, X_{i+1}]$, $i = 0, 1, 2, \dots, m-1$.

Теоретические вероятности нормального распределения можно вычислить двумя способами.

1. С помощью интегральной функции Лапласа $\Phi(x)$ (по формуле (2.25))

$$P'_i = \Phi(U_{i+1}) - \Phi(U_i), \quad (3.19)$$

здесь $U_i = \frac{X_i - \bar{X}_B}{\sigma_B}$, $U_{i+1} = \frac{X_{i+1} - \bar{X}_B}{\sigma_B}$, $i = 0, 1, 2, \dots, m-1$. σ_B - выборочное среднее квадратическое отклонение, $\bar{X}_B$ - выборочное среднее, X_i, X_{i+1} - концы интервалов $[X_i, X_{i+1}]$ (из таблицы 3.1).

2. С помощью дифференциальной функции Лапласа $\varphi(x)$ (2.22)

$$P'_i = \frac{h}{\sigma_B} \cdot \varphi(u_i), \quad (3.20)$$

здесь $u_i = \frac{x_i - \bar{X}_B}{\sigma_B}$, где x_i - середины интервалов (из таблицы 3.2).

Чем меньше различаются эмпирические и теоретические частоты, тем меньше величина критерия, и, следовательно, он в известной степени характеризует близость эмпирического и теоретического распределения.

Замечания.

1. Необходимым условием применения критерия Пирсона является наличие в каждом из интервалов не менее 5 наблюдений. Интервалы с частотами меньше 5 в таблице (3.1) следует объединять, частоты при этом складываются. При

вычислении числа степеней свободы считать количество интервалов m уже после объединения.

2. Критерий Пирсона является наиболее состоятельным при *большом* числе наблюдений ($n \geq 50$). Он почти всегда опровергает неверную гипотезу, обеспечивая минимальную ошибку в принятии неверной гипотезы по сравнению с другими критериями.

3. Если выдвигается основная гипотеза H_0 - случайная величина распределена по экспоненциальному закону, то теоретические вероятности вычисляют, используя формулу (2.20), или в новых обозначениях

$$P'_i = e^{-\lambda X_i} - e^{-\lambda X_{i+1}}, i = 0, 1, \dots, m - 1.$$

Здесь $\lambda = \frac{1}{\bar{X}_B}$; X_i, X_{i+1} - границы i - того интервала.

3.7. Критерий согласия Колмогорова

Критерий согласия Колмогорова является наиболее простым критерием проверки гипотезы о виде закона распределения. Он связывает эмпирическую функцию распределения $F^*(x)$ с теоретической функцией распределения $F^T(x)$ некоторой случайной величиной (статистика Колмогорова)

$$D_n = \max_i |F^*(x_i) - F^T(x_i)|, \quad (3.21)$$

где D_n - максимальная разность значений эмпирической и теоретической функций распределения, n - объем выборки, x_i - середины интервалов (из табл. 3.2).

Затем находят критическое значение $D_{\text{крит}} = \frac{\lambda_0}{\sqrt{n}}$, где λ_0 находят по таблицам Колмогорова. λ_0 - это такое значение, при котором функция Колмогорова $Q(\lambda)$ равна надежности γ , т.е. $Q(\lambda) = \gamma = 1 - \alpha$ (α - уровень значимости).

Если $D_n < D_{\text{крит}}$, то нет оснований отвергнуть нулевую гипотезу, в противном случае нулевую гипотезу отвергают и принимают альтернативную.

Таблица 3.4. Значения λ_0 при заданном уровне значимости α .

α	0,20	0,10	0,05	0,02	0,01	0,001
λ_0	1,073	1,224	1,358	1,510	1,627	1,950

3.8. Интервальные оценки параметров нормально распределенной случайной величины

Интервальное оценивание параметров распределения генеральной совокупности состоит в построении доверительных интервалов. Доверительным интервалом для параметра θ называется интервал (θ_1, θ_2) , покрывающий истинное значение параметра с заданной надежностью $\gamma = 1 - \alpha$, здесь α - уровень значимости. Таким образом,

$$P(\theta_1 < \theta < \theta_2) = \gamma = 1 - \alpha.$$

Оценка математического ожидания генеральной совокупности, распределенной по нормальному закону

Пусть случайная величина X имеет нормальное распределение с параметрами a и σ . Тогда интервальной оценкой при заданном уровне значимости α (с надежностью γ) математического ожидания при *неизвестном* параметре σ служит доверительный интервал

$$\overline{X}_B - t \frac{S}{\sqrt{n}} < a_{ген} < \overline{X}_B + t \frac{S}{\sqrt{n}}, \quad (3.22)$$

где $\overline{X}_B$ - выборочная средняя, S - исправленное выборочное среднее квадратическое отклонение (при $n \geq 25$ вычисляется по формуле (3.10), а при $n < 25$ - по формуле (3.13)), n - объем выборки, $t = t(k, \alpha)$ находят по таблицам приложения 4 по заданным значениям α (или в некоторой литературе по надежности γ) и числу степеней свободы k , которое на единицу меньше, чем объем выборки n , т.е. равно $k = n - 1$.

Оценка среднеквадратического отклонения генеральной совокупности нормально распределенной случайной величины.

Интервальной оценкой (с надежностью γ) среднего квадратического отклонения генеральной совокупности служит доверительный интервал

$$S \cdot (1 - q) < \sigma_{ген} < S \cdot (1 + q), \text{ если } q < 1, \quad (3.23)$$

$$0 < \sigma_{ген} < S \cdot (1 + q), \text{ если } q > 1. \quad (3.24)$$

где S – исправленное выборочное среднее квадратическое отклонение (см. предыдущий пункт), значение q находится по заданным n и γ по таблицам приложения 5.

Пример 3.1. В результате измерений некоторой случайной величины X получена выборка: 165; 167; 163; 158; 170; 169; 174; 185; 176; 177; 180; 176; 175; 163; 170; 165; 175; 169; 173; 180; 172; 156; 168; 171; 160; 165; 170; 178; 182; 150; 155; 171; 166; 162; 160; 175; 172; 170; 165; 167; 184; 169; 177; 161; 174; 175; 170; 172; 171; 154.

- а) Составить интервальный ряд распределения частот.
- б) Найти эмпирическую функцию распределения выборки и построить ее график.
- в) Построить полигон и гистограмму относительных частот.
- г) Вычислить числовые характеристики выборки:
 - выборочную среднюю;
 - выборочную дисперсию;
 - выборочное среднее квадратическое отклонение.
- д) Найти точечные оценки параметров распределения выборки.
- е) Выдвинув гипотезу о виде распределении выборки, проверить ее критерием согласия Пирсона и критерием согласия Колмогорова при уровне значимости $\alpha = 0,05$.
- ж) Построить на одном чертеже с гистограммой относительных частот график теоретической плотности вероятностей. Сделать выводы.
- з) Найти интервальные оценки параметров распределения выборки при уровне значимости $\alpha = 0,05$.

Решение.

а) Поскольку объем выборки $n \geq 25$, то данные сгруппируем. Число интервалов m примем равным 6.

Находим $X_{\min}=150$, $X_{\max}=185$. По формуле (3.3) находим $h = \frac{185-150}{6} = 5,83$.

Далее, в качестве X_1 возьмем 150, тогда $X_2 = 150 + 5,83 = 155,83$.

Таблица 3.5.

$[X_i, X_{i+1}]$	[150 – 155,83]	[155,83 – 161,67]	[161,67 – 167,5]	[167,5 – 173,33]	[173,33 – 179,17]	[179,17 – 185]
n_i	3	5	10	16	11	5

б) Найдем середины частичных интервалов из таблицы 3.5 и получим вариационный ряд распределения частот

Таблица 3.6.

x_i	152,92	158,75	164,58	170,41	176,24	182,07
n_i	3	5	10	16	11	5

Все дальнейшие вычисления проводятся с использованием данных таблицы 3.6, которые можно провести с использованием пакета прикладных программ EXCEL (результаты см. приложение 6).

Используя формулу (3.5), из таблицы 3.6 имеем

Таблица 3.7.

x_i	152,92	158,75	164,58	170,41	176,24	182,07	Контроль
w_i	0,06	0,1	0,2	0,32	0,22	0,1	$\sum_{i=1}^6 w_i = 1$
Накопленные относительные частоты $\frac{n(x_i)}{n}$	0	0,06	0,16	0,36	0,68	0,9	1
$\frac{w_i}{h}$	0,01	0,02	0,04	0,055	0,038	0,02	-

По таблице 3.7, используя накопленные частоты, находим эмпирическую функцию распределения, график которой показан на рис.3.4.

$$F^*(x) = \begin{cases} 0, & \text{если } x \leq 152,92; \\ 0,06, & \text{если } 152,92 < x \leq 158,75; \\ 0,16, & \text{если } 158,75 < x \leq 164,58; \\ 0,36, & \text{если } 164,58 < x \leq 170,41; \\ 0,68, & \text{если } 170,42 < x \leq 176,24; \\ 0,9, & \text{если } 176,25 < x \leq 182,07; \\ 1, & \text{если } x > 182,07. \end{cases}$$

в) Используя таблицу 3.7 строим полигон относительных частот (рис.3.5).

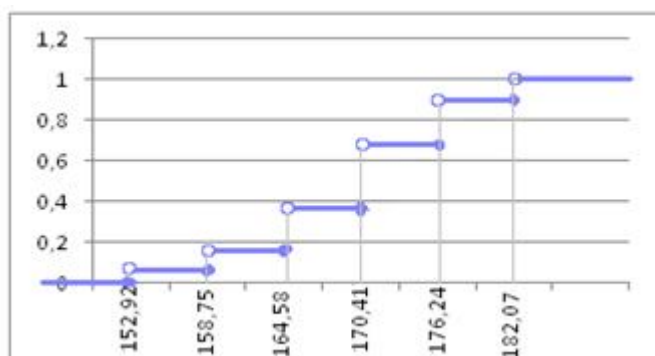


Рис.3.4. График эмпирической функции распределения

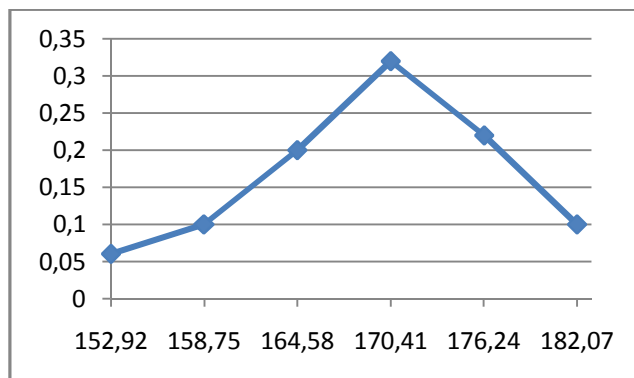


Рис.3.5. Полигон относительных частот

Построим гистограмму относительных частот с помощью результатов из таблицы 3.7. По оси X напротив интервала $[150 - 155,83]$ с серединой 152,92 строим прямоугольник высотой 0,01, затем аналогично еще пять прямоугольников с высотами 0,02; 0,04; 0,055; 0,038, 0,02. Полученная гистограмма относительных частот изображена на рис. 3.6.

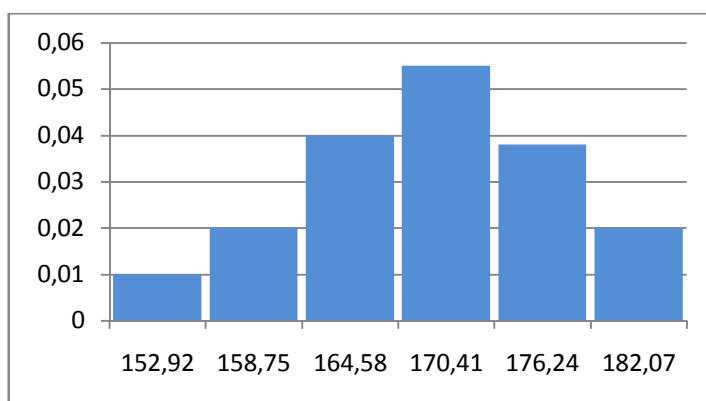


Рис.3.6. Гистограмма относительных частот

г) Вычислим числовые характеристики. По формуле (3.7) находим выборочную среднюю

$$\bar{X}_B = \frac{152,92 \cdot 3 + 158,75 \cdot 5 + 164,58 \cdot 10 + 170,41 \cdot 16 + 176,24 \cdot 11 + 182,07 \cdot 5}{50} = 169,48.$$

Выборочную дисперсию находим по формуле (3.9)

$$D_B = (152,92^2 \cdot 3 + 158,75^2 \cdot 5 + 164,58^2 \cdot 10 + 170,41^2 \cdot 16 + 176,24^2 \cdot 11 + 182,07^2 \cdot 5) / 50 - 169,48^2 = 58,9504.$$

И выборочное среднее квадратическое отклонение (по формуле (3.10))

$$\sigma_B = \sqrt{58,9504} = 7,68.$$

д) Найдем точечные оценки параметров распределения. Для этого надо определиться, параметры какого распределения нас интересуют. Гистограмма и полигон относительных частот (рис.3.5, 3.6), являющиеся статистическими оценками плотности вероятностей генеральной совокупности, схожи с кривой плотности вероятностей нормального закона (рис.2.6), поэтому точечные оценки будем находить для параметров нормального распределения – математического ожидания и среднего квадратического

$$a=169,48, \sigma=7,68.$$

е) По виду гистограммы выдвигаем основную гипотезу H_0 - генеральная совокупность, из которой взята выборка, распределена по нормальному закону с параметрами $a=169,48, \sigma=7,68$. Проверим ее **критерием Пирсона**. Имеем: объем выборки $n = 50$, точечные оценки $a=169,48, \sigma=7,68$, длина частичных интервалов $h = 5,83, \frac{h}{\sigma} = 0,759$.

Используя формулы (3.18, 3.20), найдем величину $\chi^2_{\text{набл}}$. Результаты вычислений запишем в таблицу 3.8.

Таблица 3.8.

i	1	2	3	4	5	6	сумма
середины x_i	152,92	158,75	164,58	170,41	176,24	182,07	
n_i	3	5	10	16	11	5	50
$u_i = \frac{x_i - a}{\sigma}$	-2,16	-1,40	-0,64	0,12	0,88	1,64	-
$\varphi(u_i)$	0,0387	0,1497	0,3251	0,3961	0,2709	0,1040	-
$P'_i = \frac{h}{\sigma} \varphi(u_i)$	0,0296	0,1136	0,2468	0,3007	0,2055	0,0789	0,9760
теор. частоты $n'_i = nP'_i$	1,4808+5,7073 7,188		12,3585	15,0349	10,2763	3,9461	48,8038
$\frac{(n_i - n'_i)^2}{n'_i}$	0,0917		0,4501	0,0620	0,0510	0,2815	0,9362

Замечание. Значения функции $\varphi(u_i)$ в точках u_i можно найти с помощью пакета прикладных программ EXCEL, используя статистическую функцию НОРМРАСП с логическим значением ЛОЖЬ: $\varphi(u_i) = \text{НОРМРАСП}(\text{номер ячейки } u_i; 0; 1; \text{ЛОЖЬ})$.

Так как в первом интервале частота $n_1 = 3 < 5$, то объединим его со вторым интервалом. При объединении интервалов значения n'_1 и n'_2 суммируются. Из таблицы 3.8 находим $\chi^2_{\text{набл}} = 0,9362$.

Число степеней свободы после объединения интервалов $k = m - r - 1 = 5 - 2 - 1 = 2$ (здесь $m = 5$ - количество интервалов после объединения, $r = 2$ - число параметров нормального распределения) По таблице критических точек χ^2 - распределения (приложение 3) при уровне значимости $\alpha = 0,05$ и числу степеней свободы $k = 2$ находим критическое значение

$$\chi^2_{\text{крит}}(0,05; 2) = 6,0.$$

Вывод: поскольку $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то нет оснований отвергнуть гипотезу о распределении случайной величины X по нормальному закону распределения с параметрами $\mu = 169,48$, $\sigma = 7,68$ при уровне значимости $\alpha = 0,05$. Вероятность того, что принятая гипотеза верна, равна $\gamma = 0,95$.

Проверим гипотезу о виде распределения **критерием согласия Колмогорова**. Критерий Колмогорова можно применять, не объединяя интервалы с малочисленными частотами. Из таблицы (3.7) запишем в третью строку таблицы (3.9) значения эмпирической функции распределения. Теоретические вероятности из таблицы 3.8 запишем в четвертую строку. В следующей строке запишем накопленные теоретические вероятности, получим теоретическую функцию распределения $F^T(x)$. В последней строке таблицы 3.9 запишем модуль разности между значениями теоретической и эмпирической функциями распределения.

Таблица 3.9.

i	1	2	3	4	5	6	сумма
середины x_i	152,92	158,75	164,58	170,41	176,24	182,07	
$F^*(x)$	0	0,06	0,16	0,36	0,68	0,9	1
$P'_i = \frac{h}{\sigma} \varphi(u_i)$	0,0296	0,1136	0,2468	0,3007	0,2055	0,0789	0,976
$F^T(x_i)$	0	0,0296	0,1438	0,3909	0,6916	0,8972	0,976
D_i $= F^*(x_i) - F^T(x_i) $	0	0,0304	0,0162	0,0309	0,0116	0,0029	0,024

Получаем, что максимальное отклонение эмпирической функции распределения от теоретической по модулю равно $D_n = 0,0304$.

При уровне значимости $\alpha = 0,05$ находим по таблице 3.4 значение $\lambda_0 = 1,358$. Тогда критическое значение $D_{\text{крит}} = \frac{1,358}{\sqrt{50}} = 0,192$.

Вывод. Получили, что $D_n < D_{\text{крит}}$, поэтому нет оснований отвергать нулевую гипотезу и гипотезу о нормальном законе распределения генеральной совокупности X принимаем. Вероятность того, что принятая гипотеза верна, равна $\gamma = 1 - 0,05 = 0,95$.

ж) Построим гистограмму относительных частот и график теоретической функции плотности вероятностей. Для этого из середин частичных интервалов восстановим перпендикуляры высотой $f(x_i)$ (табл. 3.10). Для построения теоретической функции плотности вероятностей вычислим дополнительно значение $f(\bar{X}_B)$ - наибольшее значение функции плотности.

Таблица 3.10.

i	1	2	3	4	5	6	7
середины x_i	152,92	158,75	164,58	170,41	176,24	182,07	$\bar{X}_B = 169,48$
$u_i = \frac{x_i - a}{\sigma}$	-2,16	-1,40	-0,64	0,12	0,88	1,64	
$\varphi(u_i)$	0,0387	0,1497	0,3251	0,3961	0,2709	0,1040	0,3989
$f(x_i) = \frac{\varphi(u_i)}{\sigma_B}$	0,0050	0,0195	0,0423	0,0516	0,0353	0,0135	0,0519

На рисунке 3.7 отметим точки $(x_i, f(x_i))$, $i = 1, 2, \dots, 7$, соединим их плавной линией.

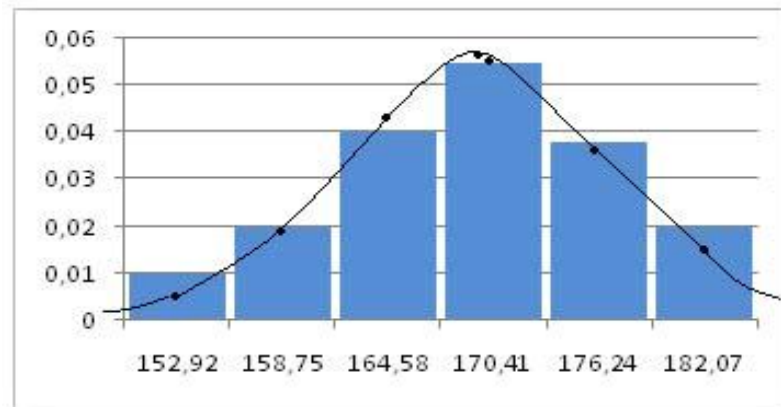


Рис.3.7. Гистограмма эмпирического распределения и теоретическая функция плотности.

Из рисунка 3.7 видно, что график теоретической функции плотности вероятностей и гистограмма достаточно хорошо совпадают.

з) После принятия гипотезы о том, что генеральная совокупность X распределена по нормальному закону, вычислим интервальные оценки параметров распределения выборки (если основную гипотезу отвергают, то интервальные оценки не находят). По формуле (3.22) вычисляем доверительный интервал для математического ожидания генеральной совокупности. В качестве исправленного среднего квадратического отклонения возьмем $S = 7,68$, так как у нас $n \geq 25$. По таблицам приложения 4, находим $t = 2,00$ при уровне значимости $\alpha = 0,05$ (надежность равна $\gamma = 0,95$) и числу степеней свободы $k = 50 - 1 = 49$ (объем выборки $n = 50$). Получаем интервальную оценку для математического ожидания

$$169,49 - 2,00 \cdot \frac{7,68}{\sqrt{50}} < a_{\text{ген}} < 169,49 + 2,00 \cdot \frac{7,68}{\sqrt{50}}$$

или после вычислений

$$167,31 < a_{\text{ген}} < 171,67.$$

Оценим среднеквадратическое отклонение генеральной совокупности. Взяв надежность $\gamma = 0,95$, по таблицам приложения 5 находим $q = 0,21 < 1$, и, используя для оценки $\sigma_{\text{ген}}$ формулу (3.23), получаем

$$7,68 \cdot (1 - 0,21) < \sigma_{\text{ген}} < 7,68 \cdot (1 + 0,21)$$

или

$$6,06 < \sigma_{\text{ген}} < 9,30.$$

4. Элементы корреляционного анализа

Многие процессы зависят от нескольких параметров их характеризующих, например, температура окружающей среды от влажности, коэффициент прочности древесины от ее плотности, урожайность от количества внесенных удобрений. Во многих задачах требуется установить влияние одного фактора на другой, есть ли корреляционная связь между факторами, сильная ли эта связь, какой формы.

4.1. Корреляционная таблица

Пусть из двумерной генеральной совокупности (X, Y) получена выборка $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$, объем которой равен n . Перед обработкой статистических данных, если $n \geq 25$, их группируют и строят корреляционную таблицу. Для этого, просматривая данные наблюдений, находят наименьшее значение x_{min} , наибольшее - x_{max} ; наименьшее - y_{min} , наибольшее - y_{max} . Затем по формуле (3.2) находят число интервалов m , на которые разбиваются два интервала $[x_{min}, x_{max}]$ и $[y_{min}, y_{max}]$.

Шаг по x обозначим через h_x , шаг по y - через h_y , их вычисляют по формуле (3.3). Далее вычисляем все концы m частичных интервалов по x (записываем их в верхнюю строку таблицы 4.1) и по y (записываем их в самый правый столбец). Затем заполняем основную часть таблицы (изображено жирной линией). Подсчитываем частоту n_{ij} - количество пар, которые попали в данный сектор. Для этого берут первую пару и в ту клетку, которую она попадет, ставят точку (черточку и т.п.), затем отмечают сектор, куда попадет вторая пара и так для всех n пар. Затем суммируем по столбцам $\sum_{j=1}^m n_{ij} = n_i^x$ и записываем результаты n_i^x в следующую строку после основной таблицы, у нас это значения $n_1^x, n_2^x, \dots, n_m^x$, причем $\sum_{i=1}^m n_i^x = n$. Аналогично, находим суммы по строкам $\sum_{i=1}^m n_{ij} = n_j^y$ записываем $n_1^y, n_2^y, \dots, n_m^y$ в столбец после основной таблицы, для контроля $\sum_{j=1}^m n_j^y = n$.

Таблица 4.1 Вспомогательная корреляционная таблица

Y \ X	$[x_{min}, x_{min} + h_x]$	$[x_{min} + h_x, x_{min} + 2h_x]$	...	$[x_{min} + (m-1)h_x, x_{max}]$	n_j^y
$[y_{min}, y_{min} + h_y]$			..		n_1^y
$[y_{min} + h_y, y_{min} + 2h_y]$			..		n_2^y
.....		.		...	...
$[y_{min} + (m-1)h_y, y_{max}]$			...		n_m^y
n_i^x	n_1^x	n_2^x	...	n_m^x	n

Из таблицы 4.1 можно получить корреляционную таблицу 4.2, для этого подсчитаем в основной части частоты в каждой клетке, и заменим на их численные значения n_{ij} , верхнюю строку заменим на значения *середины* каждого интервала по x (обозначим x_i), а правый столбец - на значения *середины* каждого интервала по y (обозначим y_j).

Таблица 4.2 Корреляционная таблица

Y \ X	x_1	x_2	...	x_m	n_j^y
y_1	n_{11}	n_{21}	..	n_{m1}	n_1^y
y_2	n_{12}	n_{22}	..		n_2^y
		.		...	...
y_m	n_{1m}	n_{2m}	...	n_{mm}	n_m^y
n_i^x	n_1^x	n_2^x	...	n_m^x	n

4.2. Числовые характеристики

Из таблицы 4.2 можно построить вариационный ряд для случайной величины X , выписав первую и последнюю строчку.

Таблица 4.3. Безусловный закон распределения частот выборки X

x_i	x_1	x_2	$\dots$	x_m
n_i^x	n_1^x	n_2^x	$\dots$	n_m^x

По формулам (3.7-3.10) по данным таблицы 4.3 можно найти числовые характеристики: безусловную выборочную среднюю $\bar{X}_B$, выборочную дисперсию D_x , выборочное среднее квадратическое отклонение σ_x .

Аналогично, из безусловного закона распределения случайной величины Y (таб.4.4) можно найти $\bar{Y}_B$, D_y , σ_y .

Таблица 4.4. Безусловный закон распределения частот выборки Y

y_j	y_1	y_2	$\dots$	y_m
n_j^y	n_1^y	n_2^y	$\dots$	n_m^y

В качестве оценок условных математических ожиданий принимают условные средние. Условным средним $\bar{Y}_i$ называют среднее арифметическое наблюдавшихся значений Y , при фиксированном значении $X = x_i$. Условные средние вычисляются по формуле

$$\bar{Y}_i = \frac{\sum_{j=1}^m y_j n_{ij}}{n_i^x} . \quad (4.1)$$

в частности, например,

$$\bar{Y}_1 = \frac{\sum_{j=1}^m y_j n_{1j}}{n_1^x}, \quad \bar{Y}_2 = \frac{\sum_{j=1}^m y_j n_{2j}}{n_2^x}, \quad \dots, \quad \bar{Y}_m = \frac{\sum_{j=1}^m y_j n_{mj}}{n_m^x}. \quad (4.2)$$

Ломаную, отрезки которой соединяют точки $(x_1, \bar{Y}_1)$, $(x_2, \bar{Y}_2)$, ..., $(x_m, \bar{Y}_m)$, называют **эмпирической кривой регрессии Y на X** .

Аналогично находят условные средние $\bar{X}_j$ для каждого фиксированного значения $Y = y_j$. Ломаную, отрезки которой соединяют точки $(y_1, \bar{X}_1)$, $(y_2, \bar{X}_2)$, ..., $(y_m, \bar{X}_m)$, называют **эмпирической кривой регрессии X на Y** .

4.3. Выборочный коэффициент корреляции и проверка гипотезы о его значимости

О тесноте связи между случайными величинами X и Y можно судить по **коэффициенту корреляции**, который вычисляется по формуле

$$r_{xy} = \frac{M(XY) - M(X)M(Y)}{\sigma_x \sigma_y}. \quad (4.3)$$

Коэффициент корреляции – величина безразмерная. Коэффициент корреляции определяет степень линейной связи случайных величин. Коэффициент корреляции обладает следующими свойствами:

1. Если коэффициент корреляции равен нулю, то между случайными величинами X и Y отсутствует линейная корреляционная зависимость.
2. Если $r_{xy} = \pm 1$, то между величинами X и Y существует функциональная линейная зависимость.
3. Абсолютная величина коэффициента корреляции двух величин, связанных линейной корреляционной зависимостью, удовлетворяет неравенству $0 < |r_{xy}| < 1$. При этом если $r_{xy} > 0$, то случайные величины X и Y одновременно возрастают или убывают, в этом случае говорят о положительной корреляции, если $r_{xy} < 0$, то с ростом одной величины - другая убывает, тогда корреляция отрицательная. Чем больше по абсолютной величине коэффициент корреляции (ближе к единице), тем сильнее **линейная корреляционная связь** между величинами X и Y , а если ближе к нулю - связь отсутствует или является существенно нелинейной.
4. Если случайные величины независимы, то коэффициент корреляции равен нулю. Если случайные величины независимы, то они и некоррелированы, но из некоррелированности нельзя сделать вывод о их независимости.

Теорема. Если двумерная случайная величина (X, Y) распределена нормально, то X и Y связаны линейной корреляционной зависимостью.

Оценкой коэффициента корреляции генеральной совокупности является выборочный коэффициент корреляции, который вычисляется по выборочным данным по формуле

$$r_B = \frac{\overline{XY} - \overline{X}_B \overline{Y}_B}{\sigma_x \sigma_y}, \quad (4.4)$$

где $\overline{X}_B$ - выборочная средняя количественного признака X , σ_x - выборочное среднее квадратическое отклонение количественного признака X , $\overline{Y}_B$ - выборочная средняя количественного признака Y , σ_y - выборочное среднее квадратическое отклонение количественного признака Y , $\overline{XY}$ - среднее произведения XY , которое вычисляется по формуле

$$\overline{XY} = \frac{\sum_{j=1}^m \sum_{i=1}^m (n_{ij} \cdot x_i \cdot y_j)}{n}. \quad (4.5)$$

Вспомогательное средство анализа выборочных данных – корреляционное поле. При составлении корреляционного поля по данной выборке объемом, равным n , наносят все n точек (x_1, y_1) , (x_2, y_2) , $(x_3, y_3), \dots, (x_n, y_n)$ на плоскость. Расположение точек позволяет сделать предварительное заключение о характере и форме зависимости.

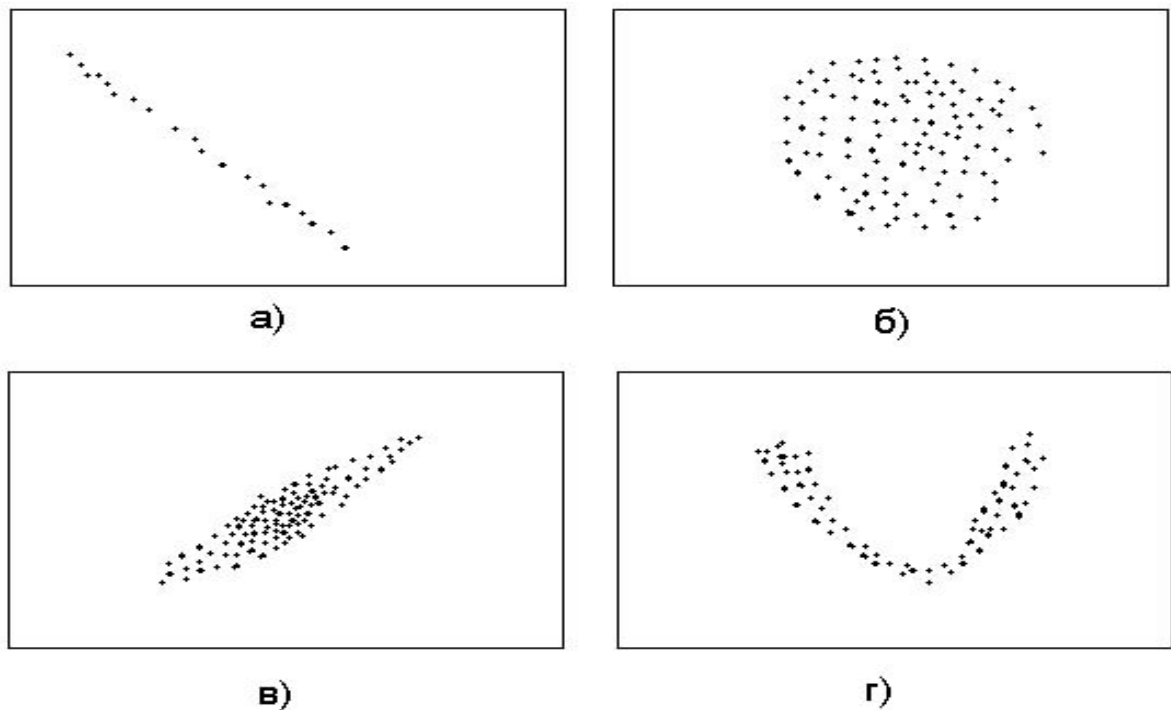


Рис. 4.1. Корреляционные поля с различными выборочными коэффициентами корреляции:

а) $r_B = -1$; б) $r_B = 0$; в) $r_B = 0,75$; г) $r_B = 0$.

На рис. 4.1 приведены четыре диаграммы рассеяния, построенные по парным выборкам. На рис. 4.1 (а) приведен пример, когда между величинами X и Y существует функциональная линейная связь, $r_B = -1$. В случае рис. 4.1 (б) следует предполагать отсутствие какой-либо корреляции между X и Y , $r_B = 0$. На рис. 4.1 (в) видно, что X и Y линейно коррелированы, $r_B = 0,75$. На рис. 4.1 (г) приведен пример, когда между величинами X и Y существует корреляция в виде параболической кривой, но так как $r_B = 0$, то предположение о линейной корреляции было бы неверным.

Проверим **значимость выборочного коэффициента корреляции**. Пусть из двумерной генеральной совокупности (X, Y) , распределенной нормально, извлечена выборка объема n , по которой вычислен выборочный коэффициент корреляции $r_{\text{ген}} \neq 0$. Требуется проверить нулевую гипотезу $H_0: r_{\text{ген}} = 0$, т.е. о равенстве нулю коэффициента корреляции генеральной совокупности.

Для проверки гипотезы о значимости выборочного коэффициента корреляции по выборке (X, Y) составляется специальная характеристика

$$T_{\text{набл}} = \frac{r_B \cdot \sqrt{n-2}}{\sqrt{1-r_B^2}}. \quad (4.6)$$

и сравнивается с табличным критическим значением $t_{\text{крит}}(\alpha, k)$, которое находится по таблице критических точек распределения Стьюдента (приложение 4) по заданному уровню значимости α и числу степеней свободы $k = n - 2$, где n – объем выборки.

Критерий.

Если $|T_{\text{набл}}| < t_{\text{крит}}$ - нет оснований отвергнуть нулевую гипотезу, следовательно, случайные величины X и Y некоррелированы.

Если $|T_{\text{набл}}| > t_{\text{крит}}$ - нулевую гипотезу отвергают, следовательно, выборочный коэффициент корреляции значимо отличается от нуля, т.е. X и Y коррелированы.

4.4. Уравнение прямой регрессии

Уравнение линейной регрессии представляет собой уравнение прямой, аппроксимирующей (приблизительно описывающей) зависимость между

случайными величинами X и Y . Если коэффициент корреляции оказывается значимым, то находят **выборочное уравнение прямой линии регрессии Y на X** по формуле

$$y = \bar{Y}_B + r_B \cdot \frac{\sigma_y}{\sigma_x} (x - \bar{X}_B). \quad (4.7)$$

Если считать, что величина Y свободная, а X зависимая от Y , то **выборочное уравнение прямой линии регрессии X на Y** находят по формуле

$$x = \bar{X}_B + r_B \cdot \frac{\sigma_x}{\sigma_y} (y - \bar{Y}_B). \quad (4.8)$$

В уравнении (4.7) коэффициент $r_B \frac{\sigma_y}{\sigma_x}$ называется коэффициентом регрессии Y на X , в уравнении (4.8) коэффициент $r_B \frac{\sigma_x}{\sigma_y}$ называется коэффициентом регрессии X на Y . Прямые (4.7) и (4.8) пересекаются в точке $(\bar{X}_B, \bar{Y}_B)$. Чем меньше угол между прямыми, тем сильнее линейная связь между X и Y . При линейно коррелированных величинах X и Y при помощи прямой регрессии (4.7) можно сделать наилучшее предсказание для Y при заданном значении X или при помощи прямой регрессии (4.8) предсказать X при заданном значении Y .

Для сравнения теоретических расчетов и данных наблюдений можно построить эмпирические и теоретические линии регрессии, корреляционное поле (диаграмму рассеяния).

Пример 4.1. Имеются 100 наблюдений двумерной случайной величины (X, Y) .

- а) Составить корреляционную таблицу.
- б) Вычислить для каждой выборки числовые характеристики: выборочные средние $\bar{X}_B$ и $\bar{Y}_B$, выборочные дисперсии D_x и D_y , выборочные средние квадратические отклонения σ_x и σ_y .
- в) Вычислить выборочный коэффициент корреляции и проверить его значимость (уровень значимости $\alpha = 0,05$).
- г) Найти выборочные уравнения прямых линий регрессии Y на X и X на Y , изобразить их на корреляционном поле.
- д) Построить эмпирические и теоретические линии регрессии.

X	Y	X	Y	X	Y	X	Y	X	Y
168	64	169	57	176	71	166	62	177	72
167	71	165	55	172	73	178	80	167	63
171	62	165	72	181	71	165	68	164	66
172	64	171	67	167	61	162	55	167	61
169	64	173	71	168	63	164	60	180	69
168	68	171	65	169	66	165	68	170	61
163	66	158	57	173	77	176	70	169	61
166	63	162	58	168	71	165	61	175	80
168	50	170	62	163	63	159	60	169	57
162	58	170	65	175	67	178	73	165	55
173	67	157	56	174	74	168	62	156	54
182	76	162	52	174	75	165	69	166	61
179	71	170	73	171	66	163	63	168	65
165	69	173	71	172	66	166	57	173	75
164	59	169	61	163	59	167	67	172	68
175	80	163	67	171	60	173	60	175	80
169	64	169	62	169	75	175	75	164	60
160	61	166	57	161	52	165	63	167	64
167	58	178	73	170	61	167	76	167	62
171	76	165	66	163	58	164	57	165	69

Решение.

а) Составим корреляционную таблицу. Исходя из того что количество испытаний $n > 20$, то данные группируем. Находим $X_{\min}=156$, $X_{\max}=181$, $Y_{\min}=50$, $Y_{\max}=80$. Пусть число частичных интервалов $m = 8$. Тогда по формуле (3.3) шаг по оси x равен $h_x = \frac{181-156}{8} = 3,125$, по оси y - $h_y = \frac{80-50}{8} = 3,75$. Просматривая каждую пару, определяем, в какой сектор она попала. Например, пару (168, 64) отметим чертой в сектор $[165,38-168,5] \times [61,25-65]$ и т.д. Если пара попала на границу интервала, то ставим вместо черты в секторе, точку на границе. Например, две пары (170, 65) и (171,65) попали на границу интервала $[168,5-171,63]$ при $Y=65$, поэтому при составлении корреляционной таблицы мы их разделим на соседние секторы поровну. А пару (168, 65), попавшую на границу интервала $[165,38-168,50]$, в таблице (4.5) отметим точкой, а в следующей таблице запишем ее в сектор, ближе к центру рассеяния (верхний).

Таблица 4.5. Предварительная корреляционная таблица

X Y	[156- 159,13]	[159,13- 162,25]	[162,25 165,38]	[165,38 -168,5]	[168,5- 171,63]	[171,63 174,75]	[174,75 177,88]	[177,88 -181]
[50- 53,75]								
[53,75- 57,5]								
[57,5- 61,25]								
[61,25- 65]								
[65- 68,75]								
[68,75- 72,5]								
[72,5- 76,25]								
[76,25- 80]								

Вычислим середины частичных интервалов по X и по Y

$$x_1 = X_{\min} + \frac{h}{2} = 157,56, x_2 = x_1 + h = 160,69 \text{ и т.д.}$$

Подсчитав все частоты, и распределив граничные точки должным образом, получаем следующую таблицу

Таблица 4.6. Корреляционная таблица

$x_i \backslash y_j$	157,56	160,69	163,81	166,94	170,06	173,19	176,31	179,44	n_j^y
51,88		3		1					4
55,63	3	1	3	2	2				11
59,38	1	2	6	4	5	1			19
63,13			3	9	6	1			19
66,88			6	2	4	3			15
70,63			4	2		2	4	3	15
74,38				1	3	4	1	3	12
78,13						1	3	1	5
n_i^x	4	6	22	21	20	12	8	7	100

б) Составим статистические законы распределения частот выборки X и Y отдельно.

Таблица 4.7.

x_i	157,56	160,69	163,81	166,94	170,06	173,19	176,31	179,44	Итого
n_i^x	4	6	22	21	20	12	8	7	100

Таблица 4.8.

y_j	51,88	55,63	59,38	63,13	66,88	70,63	74,38	78,13	Итого
n_j^y	4	11	19	19	15	15	12	5	100

Используя формулы (3.7-3.10), вычислим числовые характеристики для X и Y.

Выборочные средние

$$\bar{X}_B = (157,56 \cdot 4 + 160,69 \cdot 6 + 163,81 \cdot 22 + 166,94 \cdot 21 + 170,06 \cdot 20 + 173,19 \cdot 12 + 176,31 \cdot 8 + 179,44 \cdot 7) / 100 = 168,5;$$

$$\bar{Y}_B = (51,88 \cdot 4 + 55,63 \cdot 11 + 59,38 \cdot 19 + 63,13 \cdot 19 + 66,88 \cdot 15 + 70,63 \cdot 15 + 74,38 \cdot 12 + 78,13 \cdot 5) / 100 = 64,93;$$

Выборочные дисперсии

$$D_x = (157,56^2 \cdot 4 + 160,69^2 \cdot 6 + 163,81^2 \cdot 22 + 166,94^2 \cdot 21 + 170,06^2 \cdot 20 + 173,19^2 \cdot 12 + 176,31^2 \cdot 8 + 179,44^2 \cdot 7) / 100 - 168,5^2 = 30,1758.$$

и $D_y = 47,6663$;

Выборочные средние квадратические отклонения

$$\sigma_x = \sqrt{30,1758} = 5,49; \text{ и } \sigma_y = \sqrt{47,6663} = 6,90.$$

Найдем условные средние $\bar{Y}_i$, при фиксированных значениях $X = x_i$.

$$\bar{Y}_1 = \frac{55,63 \cdot 3 + 59,38 \cdot 1}{4} = 56,56, \quad \bar{Y}_2 = \frac{51,875 \cdot 3 + 55,625 \cdot 1 + 59,375 \cdot 2}{6} = 55;$$

$$\bar{Y}_3 = \frac{55,63 \cdot 3 + 59,38 \cdot 6 + 63,13 \cdot 3 + 66,88 \cdot 6 + 70,63 \cdot 4}{22} = 63,47;$$

$$\bar{Y}_4 = \frac{51,88 \cdot 1 + 55,63 \cdot 2 + 59,38 \cdot 4 + 63,13 \cdot 9 + 66,88 \cdot 2 + 70,63 \cdot 2 + 74,38 \cdot 1}{21} = 62,77,$$

$$\bar{Y}_5 = \frac{55,63 \cdot 2 + 59,38 \cdot 5 + 63,13 \cdot 6 + 66,88 \cdot 4 + 74,38 \cdot 3}{20} = 63,88;$$

Аналогичным образом находим $\bar{Y}_6 = 70,00$; $\bar{Y}_7 = 73,91$; $\bar{Y}_8 = 73,30$.

Вычисления занесем в таблицу

Таблица 4.9.

x_i	157,56	160,69	163,81	166,94	170,06	173,19	176,31	179,44
$\bar{Y}_i$	56,56	55	63,47	62,77	63,88	70	73,91	73,30

Если точки с координатами $(x_i, \bar{Y}_i)$ из таблицы 4.9 отметить на плоскости и соединить ломаной линией, то получим эмпирическую кривую регрессии Y на X .

Аналогично найдем условные средние $\bar{X}_j$, при фиксированных значениях $Y = y_j$.

$$\bar{X}_1 = \frac{160,69 \cdot 3 + 166,94 \cdot 1}{4} = 162,25, \quad \bar{X}_2 = \frac{157,56 \cdot 3 + 160,69 \cdot 1 + 163,81 \cdot 166,94 \cdot 2 + 170,06 \cdot 2}{11} = 163,53;$$

$$\bar{X}_3 = 165,95; \bar{X}_4 = 167,76; \bar{X}_5 = 167,77; \bar{X}_6 = 171,94; \bar{X}_7 = 176,31; \bar{X}_8 = 176,31.$$

Вычисления занесем в таблицу

Таблица 4.10.

y_j	51,88	55,63	59,38	63,13	66,88	70,63	74,38	78,13
$\bar{X}_j$	162,25	163,53	165,95	167,76	167,77	171,94	173,71	176,31

Точки $(y_j, \bar{X}_j)$ из таблицы 4.10, отмеченные на плоскости и соединенные ломаной линией, изображают эмпирическую кривую регрессии X на Y .

в) Вычислим выборочный коэффициент корреляции. Для этого найдем среднее произведения $\overline{XY}$ по формуле (4.5).

$$\begin{aligned}\overline{XY} = & (51,88(3 \cdot 160,69 + 1 \cdot 166,94) + 55,63(3 \cdot 157,56 + 1 \cdot 160,69 + 3 \cdot 163,81 \\ & + 2 \cdot 166,94 + 2 \cdot 170,06) + 59,38(1 \cdot 157,56 + 2 \cdot 160,69 + 6 \cdot 163,81 + 4 \cdot 166,94 + \\ & + 5 \cdot 170,06 + 1 \cdot 173,193) + 63,13(3 \cdot 163,81 + 9 \cdot 166,94 + 6 \cdot 170,06 + 1 \cdot 173,19) + \\ & + 66,88(6 \cdot 163,81 + 2 \cdot 166,94 + 4 \cdot 170,06 + 3 \cdot 173,19) + \\ & + 70,63(4 \cdot 163,81 + 2 \cdot 166,94 + 2 \cdot 173,19 + 4 \cdot 176,31 + 3 \cdot 179,44) + \\ & + 74,38(1 \cdot 166,94 + 74,38 \cdot 3 \cdot 170,06 + 4 \cdot 173,19 + 1 \cdot 176,31 + 3 \cdot 179,44) + \\ & + 78,13(1 \cdot 173,19 + 3 \cdot 176,31 + 1 \cdot 179,44))/100 = 10964,94\end{aligned}$$

Тогда по формуле (4.4) выборочный коэффициент корреляции равен

$$r_B = \frac{10964,94 - 168,5 \cdot 64,93}{5,49 \cdot 6,90} = 0,66.$$

Так как $r_B > 0$, то с ростом X растет и Y .

Проверим гипотезу $H_0: r_r = 0$, т.е. выясним, значим ли выборочный коэффициент корреляции. По таблицам приложения 4 по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = 100 - 2 = 98$ находим $t_{\text{крит}}(0,05, 98) = 1,98$. По формуле (4.6) находим

$$T_{\text{набл}} = \frac{0,66 \cdot \sqrt{98}}{\sqrt{1 - 0,66^2}} = 8,72.$$

Поскольку $|T_{\text{набл}}| > t_{\text{крит}}$, то нулевую гипотезу отвергаем, и считаем, что выборочный коэффициент корреляции значимо отличается от нуля, т.е. X и Y коррелированы.

г) Найдем выборочные уравнения прямых линий регрессии Y на X и X на Y . Подставив наши результаты в (4.7), получаем выборочное уравнение регрессии Y на X

$$y = 64,93 + 0,66 \cdot \frac{6,90}{5,49}(x - 168,5)$$

или после преобразования

$$y = 0,83x - 74,93.$$

Используя формулу (4.8), находим выборочное уравнение регрессии X на Y

$$x = 168,5 + 0,66 \cdot \frac{5,49}{6,90}(y - 64,93)$$

или после преобразования

$$x = 0,53y + 134,09.$$

Изобразим на одном рисунке (рис.4.1) графики полученных прямых линий регрессии Y на X и X на Y , а точками – статистические данные (корреляционное поле).

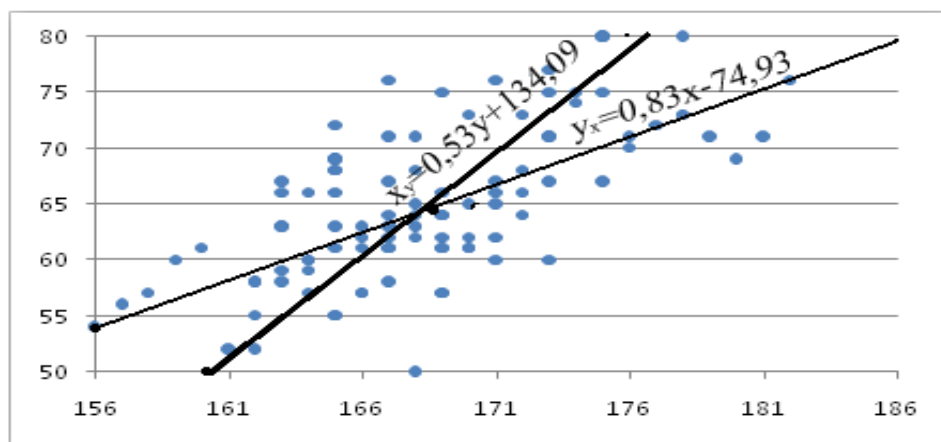


Рис. 4.2. Корреляционное поле и графики теоретических линий регрессии Y на X и X на Y .

д) Построим эмпирическую и теоретическую линии регрессии Y на X на одном чертеже (рис. 4.3). Эмпирическую кривую регрессии Y на X построим, соединяя ломаной восемь точек с координатами $(x_i, \bar{Y}_i)$, взятыми из таблицы 4.9. Теоретическую линию регрессии Y на X построим по двум точкам с координатами (156; 154,72) и (186; 79,65).

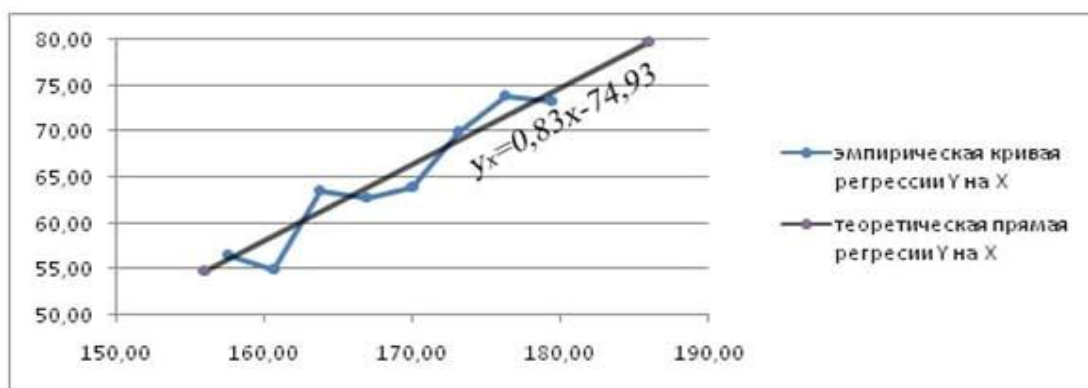


Рис. 4.3. Эмпирическая и теоретическая линии регрессии Y на X .

Аналогично, используя таблицу 4.10, построим эмпирическую линию регрессии X на Y , соединяя прямой две точки с координатами (160,4; 50) и (176,18; 80), теоретическую линию регрессии X на Y (рис. 4.4).

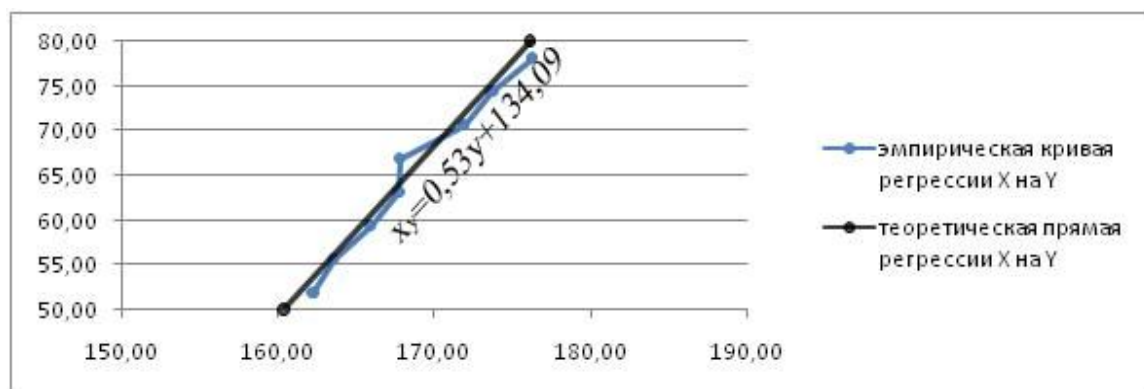


Рис. 4.4. Эмпирическая и теоретическая линии регрессии X на Y.

Простейшие задачи по теории вероятностей

Задача 1. Бросают две игральные кости. Какова вероятность того, что сумма выпавших очков: а) кратна 3; б) равна 7, а разность равна 3; в) равна 7, если известно, что разность их равна 3; г) не менее 7, если известно, что разность их равна 3.

Ответы: $1/3$; $1/18$; $1/3$; $2/3$.

Задача 2. В коллекции из 20 грампластинок имеется 5 пластинок с произведениями Моцарта. Наугад выбирают 4 пластинки. Какова вероятность того, что 2 из них с произведениями Моцарта?

Ответ: 0,2167.

Задача 3. Из букв разрезной азбуки составлено слово «ремонт». Перемешаем карточки, затем, вытаскивая их наудачу, разложим в порядке вытаскивания. Какова вероятность того, что при этом получится слово «море»?

Ответ: 0,0028.

Задача 4. Охотники Александр, Виктор, Павел попадают в летящую утку с вероятностями, соответственно равными $2/3$, $3/4$ и $1/4$. Все одновременно стреляют по пролетающей утке. Какова вероятность, что утка будет подбита?

Ответ: $15/16$.

Задача 5. Вероятность того, что студент сдаст первый экзамен, равна 0,9, второй – 0,9, третий – 0,8. Вычислить вероятность того, что хотя бы два экзамена будут сданы.

Ответ: 0,954.

Задача 6. На сборку поступают детали с трех автоматов. Первый автомат делает 0,3% брака, второй – 0,2%, третий – 0,4%. Найти вероятность попадания на сборку бракованной детали, если с первого автомата поступает 1000 деталей, со второго – 2000 деталей, а с третьего – 2500 деталей.

Ответ: 0,0031.

Задача 7. Монету бросают 5 раз. Найти вероятность того, что герб выпадает хотя бы один раз; хотя бы три раза.

Ответы: $31/32$; $1/2$.

Задача 8. Вероятность появления события в каждом из 100 независимых испытаний равна 0,8. Найти вероятность того, что событие появится в этих испытаниях: а) 90 раз; б) не менее 80 раз и не более 90 раз.

Ответы: 0,0044; 0,4938.

Задача № 9. Завод отправил на базу 5000 доброкачественных изделий. Вероятность того, что в пути изделие повредится, равна 0,0002. Найти вероятность того, что на базу поступит ровно три негодных изделия.

Ответ: 0,0613.

Задача 10. Из винтовки производят 19 выстрелов. Вероятность попадания в цель при каждом выстреле равно 0,8. Найти наивероятнейшее число попаданий в цель.

Ответы: 15; 16.

Задача 11. Написать закон распределения вероятностей и функцию распределения числа попаданий мячом в корзину при двух бросках, если вероятность попаданий $p = 0,4$.

Ответ:

X	0	1	2
P	0,36	0,48	0,16

Задача 12. Найти математическое ожидание $M(X)$, дисперсию $D(X)$, среднее квадратическое отклонение $\sigma(X)$, функцию распределения дискретной случайной величины X:

X	1,4	1,8	2,3	3,2
P	0,3	0,4	0,2	0,1

Ответы: $M(X) = 1,92$; $D(X) = 0,2796$; $\sigma(X) = 0,53$.

Задача 13. Непрерывная случайная величина X задана интегральной функцией распределения

$$F(x) = \begin{cases} 0 & \text{при } x \leq 0, \\ \frac{x}{4} & \text{при } 0 < x \leq 4, \\ 1 & \text{при } x > 4. \end{cases}$$

Найти функцию плотности распределения $f(x)$, $M(X)$, $D(X)$, $\sigma(X)$, $P(1/2 < X < 1)$.
Построить графики функций $F(x)$ и $f(x)$.

Ответы: $M(X) = 2$; $D(X) = 4/3$; $\sigma(X) = 1,1547$; $1/8$.

Задача 14. У нормально распределенной случайной величины X известны $M(X) = 10$ и $D(X) = 4$. Найти вероятность $P(12 < X < 14)$.

Ответ: 0,1359.

Простейшие задачи по математической статистике

Задача 1. Пользуясь данными выборки, определить числовые характеристики: выборочную среднюю; выборочную дисперсию; выборочное среднее квадратическое отклонение; коэффициенты эксцесса и асимметрии.

x_i	0-2	2-4	4-6	6-8	8-10
n_i	3	4	10	5	3

Ответ: 5,08; 5,1736; 2,275; -0,08; -0,49.

Задача 2. На контрольных испытаниях $n=15$ ламп была определена средняя продолжительность горения ламп $\bar{X} = 3000$ ч. Считая, что срок службы лампы распределен нормально с $\sigma=16$ (известно), определить доверительную вероятность того, что точность средней равна 10 ч.

Ответ: 0,9844.

Задача 3. Найти минимальный объем выборки из нормальной генеральной совокупности, при котором с надежностью, не меньшей $\gamma=0,9$, погрешность средней, найденной по этой выборке, будет меньше 0,3, если среднее квадратическое отклонение $\sigma=2$.

Ответ: $n \geq 121$.

Задача 4. По результатам измерений диаметра 51 корпуса электродвигателей было получено $S^2 = 2,56$. Предполагая, что распределение результата измерения нормальное, найти вероятность того, что интервал $(S - 0,16; S + 0,16)$ накроет среднее квадратическое отклонение σ .

Ответ: 0,7.

Задача 5. Проверить с уровнем значимости $\alpha=0,1$ гипотезу о том, что случайная величина X имеет распределение Пуассона. Здесь n'_i - выравнивающие частоты.

x_i	0	1	2	3	4	5	6	Сумма частот
n_i	4	12	9	11	13	9	1	59
n'_i	3,19	9,26	13,57	13,22	9,62	5,61	2,71	57,18

Задача 6. Пользуясь данными выборки, найти выборочное уравнение линейной регрессии.

x_i	1	2	3	4	5
y_i	2	3,2	2,8	4	5

Ответ: $y=0,68x+1,36$. Проверить значимость коэффициента корреляции.

Задача 7. Пользуясь данными выборки, найти выборочное уравнение линейной регрессии.

x_i	10	20	25	28	30
y_i	4	8	7	12	14

Ответ: $y=0,447x-1,1$. Проверить значимость коэффициента корреляции.

Задача 8. Вычислить коэффициент корреляции

X	1	4	6	9
Y	1	3	5	7

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

	С о т ы е д о л и а р г у м е н т а x									
x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3726	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

Приложение 2

Значения функции Лапласа $\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-\frac{t^2}{2}} dt$.

x	С о т ы е д о л и x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2703	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
3,1	0,4990	0,4991	0,4991	0,4991	0,4992	0,4992	0,4992	0,4992	0,4993	0,4993
3,2	0,4993	0,4993	0,4994	0,4994	0,4994	0,4994	0,4994	0,4995	0,4995	0,4995
3,3	0,4995	0,4995	0,4995	0,4996	0,4996	0,4996	0,4996	0,4996	0,4996	0,4997
3,4	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4997	0,4998
3,5	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998	0,4998

Значения $\chi^2(\alpha, k)$ критерия Пирсона

k число степеней свободы	Уровень значимости α			
	0,01	0,025	0,05	0,1
1	6,635	5,024	3,841	2,706
2	9,210	7,378	5,991	4,605
3	11,345	9,348	7,815	6,251
4	13,277	11,143	9,488	7,779
5	15,086	12,832	11,070	9,236
6	16,812	14,449	12,592	10,645
7	18,475	16,013	14,067	12,017
8	20,090	17,535	15,507	13,362
9	21,666	19,023	16,919	14,684
10	23,209	20,483	18,307	15,987
11	24,725	21,920	19,675	17,275
12	26,217	23,337	21,026	18,549
13	27,688	24,736	22,362	19,812
14	29,141	26,119	23,685	21,064
15	30,578	27,488	24,996	22,307
16	32,000	28,845	26,296	23,542
17	33,409	30,191	27,587	24,769
18	34,805	31,526	28,869	25,989
19	36,191	32,852	30,144	27,204
20	37,566	34,170	31,410	28,412
21	38,932	35,479	32,676	29,615
22	40,289	36,781	33,924	30,813
23	41,638	38,076	35,172	32,007
24	42,980	39,364	36,415	33,196
25	44,314	40,646	37,652	34,382
26	45,642	41,923	38,885	35,563
27	46,963	43,194	40,113	36,741
28	48,278	44,461	41,337	37,916
29	49,588	45,722	42,557	39,087
30	50,892	46,979	43,773	40,256

Квантили t – распределения Стьюдента

k число степеней свободы	Уровень значимости α					
	0,01	0,02	0,05	0,1	0,2	0,25
1	63,657	31,821	12,706	6,314	3,078	2,414
2	9,925	6,965	4,303	2,920	1,886	1,604
3	5,841	4,541	3,182	2,353	1,638	1,423
4	4,604	3,747	2,776	2,132	1,533	1,344
5	4,032	3,365	2,571	2,015	1,476	1,301
6	3,707	3,143	2,447	1,943	1,440	1,273
7	3,499	2,998	2,365	1,895	1,415	1,254
8	3,355	2,896	2,306	1,860	1,397	1,240
9	3,250	2,821	2,262	1,833	1,383	1,230
10	3,169	2,764	2,228	1,812	1,372	1,221
11	3,106	2,72	2,201	1,796	1,363	1,214
12	3,055	2,681	2,179	1,782	1,356	1,209
13	3,012	2,65	2,160	1,771	1,350	1,204
14	2,977	2,624	2,145	1,761	1,345	1,200
15	2,947	2,60	2,131	1,753	1,341	1,197
16	2,921	2,583	2,120	1,746	1,337	1,194
17	2,898	2,57	2,110	1,740	1,333	1,191
18	2,878	2,552	2,101	1,734	1,330	1,189
19	2,861	2,54	2,093	1,729	1,328	1,187
20	2,845	2,528	2,086	1,725	1,325	1,185
22	2,819	2,508	2,074	1,717	1,321	1,182
24	2,797	2,492	2,064	1,711	1,318	1,179
26	2,779	2,479	2,056	1,706	1,315	1,177
28	2,763	2,467	2,048	1,701	1,313	1,175
30	2,750	2,457	2,042	1,697	1,310	1,173
35	2,724	2,44	2,032	1,690	1,306	1,170
40	2,708	2,42	2,021	1,684	1,303	1,167
45	2,692	2,41	2,016	1,679	1,300	1,165
50	2,679	2,40	2,009	1,676	1,299	1,164
60	2,669	2,39	2,001	1,671	1,296	1,162
70	2,649	2,39	1,996	1,667	1,294	1,160
80	2,639	2,38	1,991	1,664	1,292	1,159
90	2,633	2,38	1,987	1,662	1,291	1,158
100	2,627	2,37	1,984	1,660	1,290	1,157
120	2,617	2,36	1,980	1,658	1,289	1,156
150	2,609	2,35	1,976	1,655	1,287	1,155
∞	2,576	2,326	1,960	1,645	1,282	1,151

**Таблица значений $q = q(\gamma, n)$, для нахождения доверительного интервала
для оценки среднего квадратического отклонения σ нормального
распределения**

n	Надежность γ		
	0,95	0,99	0,999
10	0,65	1,08	1,80
11	0,59	0,98	1,60
12	0,55	0,90	1,45
13	0,52	0,83	1,33
14	0,48	0,78	1,23
15	0,46	0,73	1,15
16	0,44	0,70	1,07
17	0,42	0,66	1,01
18	0,40	0,63	0,96
19	0,39	0,60	0,92
20	0,37	0,58	0,88
25	0,32	0,49	0,73
30	0,28	0,43	0,63
35	0,26	0,38	0,56
40	0,24	0,35	0,50
45	0,22	0,32	0,46
50	0,21	0,30	0,43
60	0,188	0,269	0,38
70	0,174	0,245	0,34
80	0,161	0,226	0,31
90	0,151	0,211	0,29
100	0,143	0,198	0,27
150	0,115	0,160	0,211
200	0,099	0,136	0,185

Приложение 6

Реализация обработки наблюдаемых с помощью пакета прикладных программ EXCEL

	В	С	Д	Е	Ф	Г	Н	И	Ж	
		шаг	5,83							
									сумма	
5	середины x_i	152,92	158,75	164,58	170,41	176,24	182,07	-		
6	частоты n_i	3	5	10	16	11	5	50		
7	относ част w_i	0,06	0,1	0,2	0,32	0,22	0,1	1		
8	накоп отн част $n(w_i)$	0	0,06	0,16	0,36	0,68	0,9	1		
9	w_i/h	0,0103	0,0172	0,0343	0,0549	0,0377	0,0172			
10	$x_i \cdot n_i$	458,76	793,75	1645,80	2726,56	1938,64	910,35	8473,86	169,4772	выб ср
11	$x_i \cdot x_i \cdot n_i$	70153,58	126007,81	270865,76	464633,09	341665,91	165747,42	1439073,58	58,9503	дисперсия
12	$u_i = (x_i - \bar{X})/s$	-2,1565	-1,3972	-0,6378	0,1215	0,8808	1,6401		7,6779	ср кв откл
13	$\Phi(u_i)$	0,0390	0,1503	0,3255	0,3960	0,2707	0,1039			
14	$f(x_i) = \Phi(u_i)/s$	0,0051	0,0196	0,0424	0,0516	0,0353	0,0135			
15	теор вер $P'_i = h \cdot f(x_i)$	0,0296	0,1141	0,2472	0,3007	0,2055	0,0789	0,9761		
16	накоп теор вер-ти	0,0000	0,0296	0,1438	0,3909	0,6916	0,8972	0,9761		
17	$n' \cdot i$	1,4808	5,7073	12,3585	15,0349	10,2763	3,9461	48,8038		
18	$n' \cdot i - n_i$		-0,8120	2,3585	-0,9651	-0,7237	-1,0539			
19	$(n' \cdot i - n_i)^2$		0,6593	5,5623	0,9314	0,5238	1,1106			
20	$(n' \cdot i - n_i)^2 / n_i'$		0,0917	0,4501	0,0620	0,0510	0,2815	0,9362		хи квадрат набл
21	D_i	0,0000	-0,0304	-0,0162	0,0309	0,0116	-0,0028			

I:6=СУММ(C6:H6) C:5=C6/I6 G:8=F8+F7	I:10=СУММ(C11:H11) J:10=I10/I6 I:11=СУММ(C11:H11)	J:11=(I11/I6)-(J10*J10) J:12=КОРЕНЬ(J11) C:13=НОРМРАСП(C12;0;1;ЛОЖЬ)
---	---	--

Список литературы

1. Данко П. Е., Попов А. П., Кожевникова Т. Я. Высшая математика в упражнениях и задачах. - М.: Высшая школа, 1986/ 2 часть. – с. 415.
2. Гмурман В.Е. Руководство к решению задач по теории вероятности и математической статистике. - М.: Высшая школа, 2002. – с. 373.
3. Гмурман В.Е. Теория вероятностей и математическая статистика. - М.: Высшая школа, 1977. – с. 479.
4. Кремер Н.Ш. Теория вероятностей и математическая статистика.- М.: ЮНИТИ, 2002. – с. 543.
5. Письменный Д.Т. Конспект лекций по теории вероятностей и математической статистике. – М.:АЙРИС ПРЕСС, 2005. – с. 245.

Авторы выражают благодарность учебному мастеру кафедры математики Смирнову Михаилу Анатольевичу за помощь в построении рисунков к методическому пособию.