ИНСТИТУТ СОЦИАЛЬНЫХ И ГУМАНИТАРНЫХ ЗНАНИЙ
КАФЕДРА МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ЭКОНОМИКЕ

0072.04.01

Кит Ю.В.

МАТЕМАТИКА

Часть II
ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УЧЕБНОЕ ПОСОБИЕ
для студентов всех специальностей и форм обучения

4-е издание, пересмотренное

Казань

2012

УДК 519.2
ББК 22.07
К45

Рецензенты:
Л.Р. Галяутдинова — к.ф.-м.н.
С.П. Курзин — к.ф.-м.н., доцент ИСГЗ

Кит Ю.В.
К45 Математика: часть II. Теория вероятностей и математическая статистика: Учебное пособие для студентов всех специальностей и форм обучения / Кит Ю.В. — 4-е изд., пересмотр. — Казань: Изд-во «Юниверсум», 2012. — 116 с.

Учебное пособие по дисциплине «Математика» составлено в соответствии с требованиями федерального компонента к обязательному минимуму содержания и уровню подготовки дипломированного специалиста по циклу общих математических и естественнонаучных дисциплин государственного образовательного стандарта высшего профессионального образования РФ и является обязательным для изучения.

УДК 519.2
ББК 22.07

СОДЕРЖАНИЕ

Часть 1. «ТЕОРИЯ ВЕРОЯТНОСТЕЙ»

ВВЕДЕНИЕ

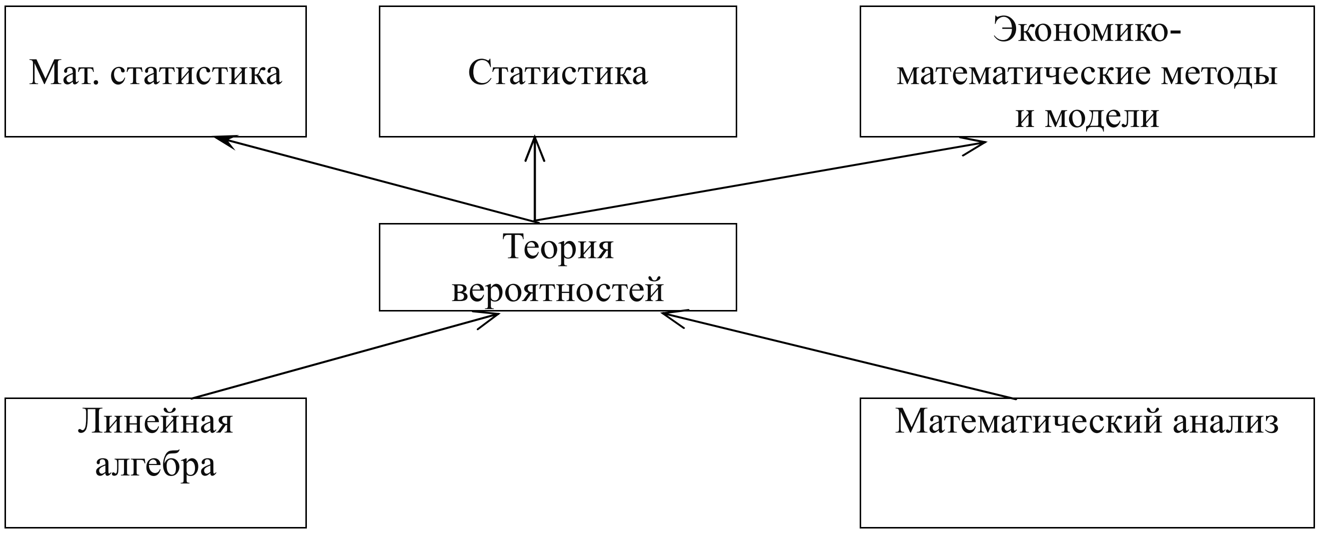
Целью изучения курса «Теория вероятностей» является развитие у студентов навыков:

- математического мышления;
- использования математических методов и основ математического моделирования;
- математической культуры.

Фундаментальность математической подготовки включает в себя достаточную общность математических понятий и конструкций, обеспечивающую широкий спектр их применимости, точность формулировок математических свойств изучаемых объектов, логическую строгость изложения математики, опирающуюся на адекватный современный математический язык.

Студент должен иметь представление о важнейших понятиях теории вероятностей, на основе которых возможны корректное их применение в практической деятельности, а также повышение им своей квалификации.

Карта межпредметных связей



ВЫПИСКА ИЗ ГОСУДАРСТВЕННОГО ОБРАЗОВАТЕЛЬНОГО СТАНДАРТА ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ СПЕЦИАЛЬНОСТЬ 060500 — «БУХГАЛТЕРСКИЙ УЧЕТ, АНАЛИЗ И АУДИТ»

Общие математические и естественнонаучные дисциплины
Федеральный компонент

Математика (ЕН.Ф.01)

Теория вероятностей и математическая статистика: случайные события; частота и вероятность; основные формулы для вычисления вероятностей; случайные величины; числовые характеристики дискретной и непрерывной случайных величин; нормальный закон распределения; генеральная совокупность и выборка; оценки параметров; корреляция и регрессия.

РАБОЧАЯ ПРОГРАММА КУРСА

Тема 1. Множества и действия над ними

Предмет теории вероятностей. Понятие множества. Пересечение, объединение, разность, умножение множеств.

Тема 2. Комбинаторика

Выборки без повторений. Размещения, перестановки и сочетания без повторений.

Тема 3. Выборки с повторениями

Размещения, перестановки и сочетания с повторениями

Тема 4. Алгебра событий

Пространство элементарных событий.

Тема 5. Вероятность события

Понятие случайного события. Классическое и геометрическое определение вероятности. Бином Ньютона. Элементарная теория вероятностей.

Тема 6. Основные теоремы теории вероятностей

Теорема сложения вероятностей. Теорема умножения вероятностей. Формула полной вероятности.

Тема 7. Методы вычисления вероятностей

Схема Бернулли.

Тема 8. Дискретные случайные величины

Функция распределения, ее свойства. Математическое ожидание и дисперсия непрерывной случайной величины.

Тема 9. Непрерывные случайные величины

Функция распределения, плотность распределения, их взаимосвязь и свойства. Математическое ожидание и дисперсия непрерывной случайной величины.

Тема 10. Основные законы распределения

Биномиальный закон распределения. Равномерный закон распределения. Распределение Пуассона. Показательный закон распределения. Нормальное распределение. Стандартное (нормированное) нормальное распределение. Вероятность попадания в заданный интервал нормально распределенной случайной величины. Вероятность заданного отклонения нормально распределенной случайной величины от своего математического ожидания. Правило трех сигм. Локальная и интегральная теоремы Лапласа.

Тема 11. Предельные теоремы теории вероятностей

Понятие о законе больших чисел. Неравенства Маркова, Чебышева. Теоремы Чебышева, Бернулли, Пуассона. Понятие о «центральной предельной теореме» Ляпунова.

Тема 12. Числовые характеристики случайных векторов

Условные математические ожидания Функции регрессии. Ковариационная матрица. Коэффициенты корреляции. Функции случайных величин и случайных векторов, их законы распределения.

Тема 13. Элементы теории случайных процессов и теории массового обслуживания

Понятия случайного процесса Цепи Маркова. Применение в социально-экономических исследованиях.

КРАТКИЙ КУРС ЛЕКЦИЙ

ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ МНОЖЕСТВ

Множество представляет собой соединение, совокупность, собрание некоторых предметов, объединенных по какому-либо признаку. Например, множество студентов одной группы, буквы алфавита и т.д.

Примеры бесконечных числовых множеств:

N — множество натуральных чисел (1, 2, 3, 4,…);

Z — множество целых чисел (-1, -5, 10, …);

Q — множество рациональных чисел (-2, 5.6, 100.56, …);

R — множество действительных чисел (π, $\sqrt{2}$, 2, 100);

C — множество комплексных чисел ($2+i$, $5-3i$).

Предметы, из которых состоит множество, называются его элементами. Например, «K» — элемент множества букв русского алфавита.

Элементы множества обозначаются прописными буквами латинского или греческого алфавита, например: $\{a_1, a_2, a_3\}$. Для обозначения множеств используют заглавные буквы латинского алфавита: A, B… Запись $\alpha \in A$ означает, что элемент α принадлежит множеству A, $\alpha \notin A$ означает, что элемент α не принадлежит множеству A. Так, $2 \in N$ — число 2 принадлежит множеству натуральных чисел.

Множество считается заданным, если перечислены все его элементы или задано свойство (признак) принадлежности элементов данному множеству. Например, если A — множество четных натуральных чисел, его можно задать следующим образом: $A = \{x \notin N | x \,\vdots\, 2\}$ ($\vdots$ — кратно).

Если множество не содержит ни одного элемента, то оно называется пустым и обозначается: $\varnothing$. Множество B называется подмножеством множества A, если все элементы множества B являются элементами множества A: $B \subset A$, это означает, что множество B «включается» в множество A. Пустое множество является подмножеством любого множества. Любое множество является подмножеством самого себя.

Пример 1.

Пусть $A = \{1, 2, 3, 4\}$. Найти все подмножества множества A.

Решение:

Подмножествами множества A являются:

$\varnothing$, $\{1,2,3,4\}$, $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, $\{2,4\}$, $\{1,3\}$, $\{1,2\}$ $\{1,4\}$, $\{4,3\}$, $\{2,3\}$, $\{3,4,1\}$, $\{4,1,2\}$, $\{1,2,3\}$, $\{2,3,4\}$.♦

Равными (одинаковыми) являются множества, состоящие из одних и тех же элементов.

ОПЕРАЦИИ НАД МНОЖЕСТВАМИ

Пересечение множеств

Пересечением двух множеств A и B называется множество C, состоящее из всех элементов, принадлежащих обоим множествам A и B. Обозначается $C = A \cap B$, графически изображено на рис. 1.

Пример 2.

A — множество натуральных чисел, кратных 2: $A = \{x \in N | x \vdots 2\}$, B — множество натуральных чисел, кратных 3: $B = \{x \in N | x \vdots 3\}$. Найти $C = A \cap B$.

Решение:

C — множество чисел, кратное шести. $C = \{x \in N | x \vdots 6\}$.♦

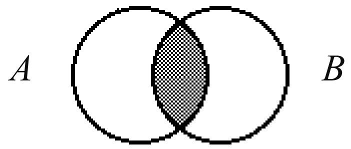


Рис. 1.

Объединение множеств

Объединением множеств A и B называется множество C, которое состоит из всех элементов множеств A и B. Обозначается

$C = A \cup B$, графически представлено на рис. 2.

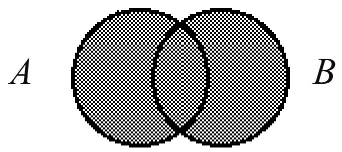


Рис. 2.

Пример 3.

$A = \{1, 2, 3\}$, $B = \{2, 3, 4\}$.

Найти: $C = A \cup B$

Решение:

$C = \{1, 2, 3, 4\}$ ♦

Свойства операций над множествами

Коммутативность: $A \cap B = B \cap A$, $B \cup A = A \cup B$

Ассоциативность: $(A \cup B) \cup C = A \cup (B \cup C)$,

$(A \cap B) \cap C = A \cap (B \cap C)$

Дистрибутивность: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$,

$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$.

Вычитание множеств. Дополнение множеств

Разностью множеств A и B называется множество, состоящее из элементов множества A, не принадлежащее множеству B. Обозначается $A \backslash B$ (рис. 3 см. ниже).

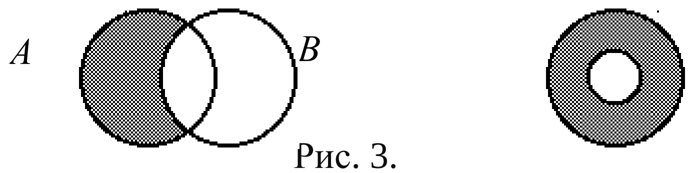


Рис. 3.

Пример 4.

а) $A = \{1, 2, 3, 4\}$, $B = \{1, 4\}$

б) $A = \{1, 2, 3\}$, $B = \{1, 2, 3, 4\}$

Найти $A \backslash B$:

Решение:

а) $A \backslash B = \{2, 3\}$;

б) $A \backslash B = \{ \varnothing \}$ ♦

Если $B \subset A$, то $A \backslash B$ называется дополнением B до множества A.

Умножение множеств

Произведением двух множеств A и B является множество C элементами которого являются пары, составленные из всех элементов множеств A и B.

Пример 5.

$A = \{1, 2\}$, $B = \{3, 4\}$. Найти $A*B$.

Решение.

$A*B = \{(1, 3), (1, 4), (2, 3), (2, 4)\}$ ♦

Задания для практики

1. Найдите множество корней уравнения $(x^2 - 1)(x^2 + 5x + 6) = 0$

2. Найдите множество всех целых чисел, удовлетворяющих неравенству $x^2 \leq 5$.

3. Пусть M — множество всех корней уравнения $x^5 + 3x^4 + x^3 - 1 = 0$. Какие из чисел 1, -1, ½, -½ являются элементами множества M?

4. Найдите все подмножества множества $A = \{3, 4, 5\}$

5. Найдите $A \cap B$, $A \cup B$, $B \backslash A$, если

а) $A = \{3, 4, 5\}$, $B = \{3, 5, 6\}$ $A \backslash B$

б) $A = \{0, 1, 7, 8\}$, $B = \{-7, 0, 6, 9\}$

в) $A = \{1, 3, 5, 7\}$, $B = \{2, 4, 6, 8\}$

г) $A = \{1, 2, 3\}$, $B = \{-1, 0, 1, 2, 3\}$.

6. Пусть M — множество всех корней уравнения $2x^6 + x^3 + x = 0$.

Найдите пересечение этого множества с множествами $A = \{1, 2, 3\}$, $B = \{0, 1, -1\}$, $C = \{-2, -1, 1\}$.

7. Найдите $A \backslash M$, $B \backslash M$, $C \backslash M$ по данным предыдущего номера.

8. Найдите дополнение множества A до множества В, если

а) $A = \{1, 2, 3\}$, $B = \{0, 1, 2, 3, 5\}$

б) $A = \{1, 2, 3\}$, $B = (0, 1, 2, 3, 4\}$

в) $A = \{0, 1\}$, $B = \{-1, 0, 1, 2\}$.

9. Чему равны $A \cap B$, $A \cup B$, $B \backslash A$ если $A \subset B$?

10. Найдите множества $A \cap B$, $A \cup B$, $A \cap C$, $A \cup C$, $C \cap B$, $C \cup B$, $A \cap B \cap C$, $A \cup B \cup C$.

$A = \{-4, -3, -2, -1, 0, 1, 2\}$,

$B = \{-2, -1, 0, 1, 2, 3, 4\}$,

$C = \{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$.

11. Пусть A — множество параллелограммов, B — множество прямоугольников, C — множество ромбов, D — множество квадратов. Найдите $A \cap B$, $C \cap B$, $A \cap B \cap C \cap D$, $A \cup B \cup C \cup D$.

КОМБИНАТОРИКА

Общие правила комбинаторики

Правило суммы: если некоторый объект A можно выбрать k способами, а некоторый объект B — n способами, то объект «либо A, либо B», можно выбрать $(k + n)$ способами.

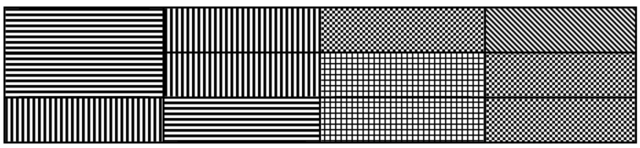
Правило произведения: если объект A можно выбрать m способами, а после такого выбора каждый объект можно выбрать k способами (независимо от выбора объекта A), то пары объектов A и B можно выбрать $m*n$ способами.

Генеральная совокупность без повторений и выборки без повторений

Генеральная совокупность — это набор некоторого конечного числа различных элементов a_1, a_2, a_3, a_4, a_5,… a_n.

Выборкой объема m ($m \le n$) называется произвольная группа из m элементов данной генеральной совокупности.

Сравним узоры нескольких пестрых лент, построенных из одинакового количества прямоугольников.



Эти ленты могут отличаться либо окраской, по крайней мере, одного квадрата, либо порядком расположения прямоугольников в линейном строю, либо и тем, и другим.

Таким образом, минимальным признаком, отличающим одну выборку объема m от другой выборки такого же объема, может быть: их различие, по крайней мере, одним элементом или их различие порядком расположения элементов.

Размещениями без повторений из n элементов по m называются такие выборки, которые, имея по m элементов, выбранных из числа данных n элементов генеральной совокупности без повторений, отличаются одна от другой либо составом элементов, либо порядком их расположения.

Пример 1.

Размещением из 10 элементов по три является совокупность трехзначных номеров машин, без повторяющихся цифр.

Пример 2.

Из трех элементов a, b, c можно составить 3 размещения по одному элементу: a, b, c; 6 размещений по два элемента, ab, ac, bc, cb, ca, ba; 6 размещений по три элемента abc, bca, cab, bac, cba, acb. ♦

Количество размещений из n элементов по m без повторений считается по формуле:

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

Перестановками из n элементов называются размещения из n элементов по n, т.е. размещения, отличающиеся друг от друга только порядком расположения элементов.

Пример 3.

Перестановка из 10 элементов — совокупность десятизначных номеров машин.

Количество перестановок без повторений из n элементов считается по формуле: $P_n=n!$

Пример 4.

Сколькими различными способами можно рассадить 10 человек на одной скамейке?

Решение.

$P_{10}=10!=1*2*3*4*5*6*7*8*9*10=3628800$ ♦

Сочетаниями из n элементов по m без повторений называются такие размещения из n элементов по m без повторений, отличающиеся друг от друга хотя бы одним элементом.

Число таких сочетаний подсчитывается по формуле: $C_n^m = \frac{n!}{m!(n-m)!}$

Пример 5.

На тренировках занимается 12 баскетболистов. Сколько может быть образовано стартовых пятерок.

Решение:

$$C_{12}^5 = \frac{12!}{5'(12-5)'} = \frac{8\cdot 9\cdot 10\cdot 11\cdot 12}{1\cdot 2\cdot 3\cdot 4\cdot 5} = 792$$

Обобщим полученные сведения в таблице.

Выборки без повторений

Название	Характерный признак отличия	Пример	Формула подсчета вариантов
Размещения	Состав Порядок	a, b, c из 3 по 2 ab, bc, ca, ba, cb, ac	$A_n^m = \frac{n!}{(n-m)!}$
Перестановки	Порядок	a, b, c из 3 abc, bca, cba, cab, bac, acb	$P_n=n!$
Сочетания	Состав	a, b, c из 3 по 2 ab, bc, ca	$C_n^m = \frac{n!}{m!(n-m)!}$

Задания для практики

1. Вычислить 4!, 5!, 6!

2. Вычислить A_7^5, P_5, C_7^5.

3. В классе 30 учеников, необходимо избрать старосту, культорга и казначея класса. Сколькими способами можно образовать руководящую тройку, если одно лицо может занимать только один пост? (24360).

4. Сколько разных пятизначных чисел можно составить из цифр 1, 2, 3, 4, 5 при условии, что ни одна цифра не повторяется? (120).

5. Сколько разных стартовых шестерок можно образовать из 10 волейболистов? (210).

6. В кружке математиков 25 человек. Необходимо избрать председателя кружка, его заместителя, редактора стенгазеты и секретаря. Сколькими способами можно образовать руководящую четверку, если одно лицо может занимать только один пост? (303600).

7. Школьная молодежная организация, в которой 50 человек, выбирает 6 делегатов на конференцию. Сколькими способами может быть избрана делегация?

8. В колоде 32 карты, раздается по 3 карты. Сколько может быть способов появления туза среди розданных карт?

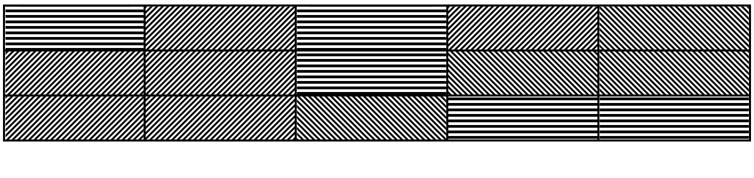
9. Для полета на Марс необходимо укомплектовать следующий экипаж: командир корабля, первый помощник, второй помощник, два бортинженера и один врач. Командная тройка может быть отобрана из 25 летчиков, 2 бортинженера из 20 специалистов, в совершенстве знающих устройство корабля, и врач — из 8 медиков. Сколькими способами можно укомплектовать экипаж?

Генеральная совокупность с повторениями и выборки с повторениями

Генеральная совокупность с повторениями — это набор элементов различных классов, когда элементы, принадлежащие одному классу, считаются одинаковыми. Число элементов в каждом классе неограниченно.

Выборкой с повторениями объема m называется произвольная группа m элементов с повторениями.

Рассмотрим несколько пестрых лент, составленных из одинакового числа прямоугольников с разными узорами. Эти ленты могут отличаться порядком расположения прямоугольников, различным набором прямоугольников, либо и тем, и другим.



Таким образом, две выборки с повторениями могут отличаться друг от друга либо составом, либо порядком, либо и тем , и другим.

Размещениями с повторениями из элементов n классов по m, называются такие выборки, которые, имея по m элементов, выбранных из числа элементов данных n классов генеральной совокупности с повторениями, отличаются друг от друга либо составом элементов, либо порядком их расположения.

Число таких размещений, где n — число классов, m — число элементов выборки подсчитывается по формуле $A_n'^m = n^m$

Пример 6.

Сколько можно составить пятизначных телефонных номеров?

Решение: $A'^{5}_{10} = 10^5 = 100000$. ♦

Перестановками с повторениями называются такие размещения из элементов n классов, которые отличаются друг от друга только порядком расположения элементов.

$$\underbrace{a, a, a, \ldots a}_{k_1} \quad \underbrace{b, b, b \ldots b}_{k_2} \quad \underbrace{l, l, l \ldots l, l}_{k_n}$$

$$k_1 + k_2 + \quad + k_n = k$$

Число таких перестановок обозначается $P'\, k_1, k_2, .. k_n = \dfrac{k!}{k_1! k_2! \ldots k_n!}$

Пример 7.

Сосчитать, сколько можно сделать перестановок в словах: замок, топор, ротор, колокол.

Решение.

замок: $P' = \dfrac{5!}{1!1!1!1!1!} = 120$; ротор: $P' = \dfrac{5!}{1!2!2!} = 30$

топор: $P' = \dfrac{5!}{1!1!2!} = 60$; колокол: $P' = \dfrac{7!}{2!2!2!} = 210$

Пример 8.

Я помню, что нужный мне телефонный номер начинается с цифры 9 и содержит три четверки и две пятерки. Однако расположение этих пяти цифр забыто. Сколько нужно сделать проб?

Решение:

$$P' = \frac{5!}{3!2!} = 10$$

Сочетаниями с повторениями из элементов n классов по m называются такие размещения с повторениями из n классов по m, которые отличаются одно от другого хотя бы одним элементом. Их число подсчитывается по формуле:

$$C_n^m = \frac{(m+n-1)!}{m!(n-1)!}$$

Пример 9.

В продажу поступили открытки 10 разных видов. Сколькими способами можно образовать набор из 12 открыток?

Решение: $C_{10}^{12} = \dfrac{(10+12-1)!}{12!(10-1)!} = 293930$.

Обобщим полученные сведения в таблице

Выборки с повторениями

Название	Характерный признак отличия	Пример	Формула подсчета вариантов
Размеще-ния	состав порядок	*a, b, c* из 3 по 2 *ab, bc, ca, ba, cb, ac, aa, cc, bb*	$A_n^m = n^m$
Переста-новки	порядок	*a, b* из 2 *ab, ba, aa, bb*	$P' = \frac{k!}{k_1!k_2!*...*k_n!}$
Сочетания	состав	*a, b, c* из 3 по 2 *ab, bc, ca, aa, cc, bb*	$C_n^m = \frac{(m+n-1)!}{m!(n-1)!}$

Задания для практики

1. Мать купила фрукты: 2 яблока, 3 груши и 4 апельсина. 9 дней подряд она каждый день предлагает сыну по фрукту. Сколькими способами можно выдать фрукты? (1260).

2. Для несения почетного караула из 10 человек могут быть приглашены офицеры пехотных войск, авиации, погранвойск, артиллерии, офицеры морского флота ракетных войск. Сколькими способами можно избрать состав караула? (3003).

3. В гастрономе имеются конфеты трех наименований. Конфеты упакованы в коробки трех видов, для каждого наименования своя коробка. Сколькими способами можно заказать набор из пяти коробок? (21)

4. Сколько машин можно обеспечить шестизначными номерами? (1000000)

5. Четыре студента сдают экзамен. Сколько может быть вариантов распределения оценок, если известно, что так или иначе они экзамен сдали? (81).

6. Сколько разных четырехзначных чисел можно составить из цифр 0, 1, 2, если цифры могут повторяться? (54)

7. На конференцию собрались школьники 9,10,11 классов. В президиум приглашаются 10 человек. Сколькими способами можно его составить при условии участия в нем хотя бы одного 11-классника? (55)

8. На Всемирный фестиваль молодежи прибыли представители пяти континентов мира. Сколькими способами можно образовать делегацию из 8 человек, при условии участия в ней представителей всех континентов? (35)

9. Имеется неограниченное количество монет по 10, 15 и 20 копеек. Сколькими способами можно образовать набор из 20 монет? (231)

10. Сколько разных трехзначных чисел можно составить из цифр 1, 2, 3, 4, 5 при условии, что цифры не повторяются? (60)

11. Буквы азбуки Морзе образуют последовательность точек и тире. Сколько различных групп можно образовать, если использовать 5 символов? (32)

12. Требуется составить расписание отправления поездов на разные дни недели. При этом необходимо, чтобы 3 дня отправлялось по 2 поезда в день, 2 дня по 1 поезду в день, 2 дня по 3 поезда в день. Сколько можно составить расписаний? (210)

13. Надо рассадить на одной скамейке 5 мальчиков и 5 девочек так, чтобы не было двух рядом сидящих мальчиков и двух рядом сидящих девочек. (660)

14. Сколькими различными способами можно составить разведывательную группу из трех солдат и одного командира, если имеется 12 солдат и 3 командира? (1680).

Сколькими различными способами можно разместить в 9 клетках следующие 9 букв: *а, а, а, в, в, в, с, с, с*?

ТЕОРИЯ ВЕРОЯТНОСТИ

Основные понятия

Человека окружает мир событий. Он часто замечает такой факт: одни события при реализации какого-то комплекса условий обязательно происходят, другие же могут произойти, а могут не произойти. Рассмотрим следующую группу событий:

Событие	Реализация комплекса условий	Исход
A_1	При нагревании проволоки	ее длина увеличивается
A_2	При бросании игральной кости	выпало 4 очка
A_3	При бросании монеты	выпал герб
A_4	При осмотре почтового ящика	найдено 4 письма
A_5	При температуре ниже 0^0 С	вода превращается в лед

Очевидно, что события A_1 и A_5 происходят закономерно, а события A_2, A_3, A_4 могут произойти, а могут и не произойти.

Поэтому наблюдаемые нами события можно подразделить на достоверные, случайные, невозможные.

Достоверным называется событие, которое при определенных условиях обязательно произойдет.

Случайным называется событие, которое при определенных условиях может произойти, а может и не произойти.

Невозможным называется событие, которое при определенных условиях заведомо не произойдет.

Отношения и операции над событиями

Сравним следующие события: A — при бросании игральной кости выпало 4 очка, B — при бросании игральной кости выпало четное число очков. Из того, что произошло событие A следует, что произошло и событие B. Говорят, что A включено в B. Таким образом, $A \subset B$, если из того, что произошло событие A следует, что произошло и B.

Два события называются несовместными, если появление одного из них исключает появление другого в одном и том же опыте, например, выпадение герба и цифры при бросании монеты.

Объединением несовместных событий A и B называется событие C, состоящее в наступлении, по крайней мере, одного из событий A или B:

$C = A \cup B$.

Для большего числа событий $A = A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n$, событие A состоит в том, что произошло или A_1, или A_2 , ... , или A_n.

Пересечением событий A и B называется событие C, означающее, что произошло и A, и B: $C = A \cap B$.

При большем числе событий $A = A_1 \cap A_2 \cap A_3 \cap A_n$, событие A состоит в том, что произошло и событие A_1, и A_2, и A_n.

Пример 1.

Событие A — попадание в мишень первым выстрелом, B — попадание в мишень вторым выстрелом. В чем состоят события

$A \cup B$, $A \cap B$?

Решение:

$A \cup B$ — попадание в мишень хотя бы одним выстрелом,

$A \cap B$ — оба выстрела попали в цель.

События A_1, A_2 , ..., A_n образуют полную группу событий, если хотя бы одно из них обязательно произойдет в данном испытании.

Противоположными называются два единственно возможных события, образующих полную группу.

Задания для практики

1. Какие из событий являются частью другого события:

A — попадание в мишень первым выстрелом;

B — попадание в мишень по меньшей мере одним из четырех выстрелов;

C — попадание в мишень одним из двух выстрелов;

D — попадание в мишень по меньшей мере одним из пяти выстрелов.

2. Событие A — лотерейный выигрыш в 1 рубль, B — лотерейный выигрыш в 2 рубля, C — лотерейный выигрыш в 3 рубля, D — лотерейный выигрыш в 4 рубля. В чем состоит событие $A \cup B \cup C \cup D$?

3. Событие A — появление нечетного числа очков при бросании игральной кости, B — непоявление 3 очков при бросании, C — непоявление 5 очков. В чем состоят события $A \cap B \cap C$, $A \cap B$, $A \cap C$, $B \cap C$?

Вероятность события

Вероятность — это количественная мера возможности появления рассматриваемого события.

Классическое определение вероятности события

Равновозможными называются такие события, любое из которых по отношению к другим событиям не обладает никаким преимуществом появляться чаще другого в многократно проводимых испытаниях в одинаковых условиях.

Исход называется благоприятствующим данному событию, если его появление влечет за собой наступление этого события.

Вероятность случайного события H равна отношению числа m равновозможных, единственно возможных и несовместных исходов, благоприятствующих этому событию H, к общему числу n всех равновозможных исходов, определяемых данным испытанием.

$$P(H)=\frac{m}{n}$$

Пример 2.

Пусть в урне лежат 23 белых и 2 черных шара. Наугад вынимаем один шар. Поскольку из 25 возможных исходов, 23 благоприятствуют появлению белого шара и лишь 2 благоприятствуют появлению черного шара. Вероятность того, что вынем белый шар будет 23/25, а черный шар — 2/25.

Пример 3.

При бросании игральной кости равновозможно выпадение любого из 6 очков. Поэтому вероятность выпадения грани (например, с 5 очками) будет 1/6. Все 6 случаев выпадение любой грани являются единственно возможными, равновозможными и несовместными.

Пример 4.

Бросают две игральные кости. Какова вероятность выпадения суммы очков равной 7?

Решение.

Рассмотрим возможные случаи выпадения очков. На первом месте — количество очков на первой кости, на втором — на второй.

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Все 36 случаев единственно возможны, равновозможны и несовместны. Поэтому вероятность появления, например, случая (2,5) равна 1/36. Определим вероятность выпадения в сумме 7 очков. Таких будет 6 случаев. Поэтому вероятность выпадения такой суммы очков будет 6/36=1/36.

Статистическое определение вероятности

Статистическая частота появления события H вычисляется по формуле

$$P^*(H)=\frac{k}{l},$$

где k — число появления события H в серии из l опытов.

Вероятностью события H называется число, относительно которого стабилизируется (устанавливается) относительная частота $P^*(H)$ при неограниченном увеличении числа опытов.

Пример 5.

В стрелковом кружке занимаются Алеша и Сережа. У кого больше вероятность выиграть соревнования, если данные предварительных туров о попадании в цель следующие:

Стрелки	Число выстрелов				
	10	20	30	40	50
Алеша	8	17	26	33	41
Сережа	3	5	8	12	15

Решение.

Событие H — попадание в цель. Определим статистическую частоту попадания в цель у Алеши и Сережи.

Алеша: $P_1{}^* = 8/10 = 0{,}8$; $P_2{}^* = 17/20 = 0{,}85$; $P_3{}^* = 26/30$; $P_4{}^* = 33/40$;$P_5{}^* = 41/50$.

Сережа: $P_1{}^* = 3/10 = 0{,}3$; $P_2{}^* = 5/20 = 0{,}25$; $P_3{}^* = 8/30$; $P_4{}^* = 12/40$;$P_5{}^* = 15/50$.

Статистическая частота попадания в цель Алеши сосредотачивается около числа 0,8. Статистическая частота попадания в цель Сережи сосредотачивается около числа 0,3. Поэтому вероятность попадания в цель у Алеши больше, чем у Сережи.

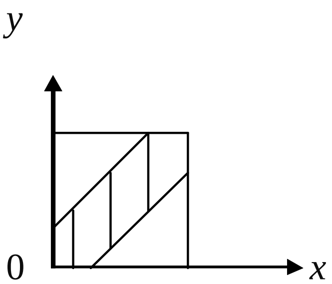
Геометрическое определение вероятности

Геометрической вероятностью события А называется отношение меры области, благоприятствующей появлению события А, к мере всей области.

Пример 6.

Два лица A и B договорились встретиться в определенном месте в промежутке времени от 9.00 до 10.00 часов. Каждый из них приходит наудачу, независимо от другого, и ожидает 15 минут. Какова вероятность, что они встретятся?

Решение.



Рассмотрим прямоугольную систему координат XOY, в качестве единиц масштаба — часы. Обозначим моменты прихода в определенное место лиц, соответственно через x и y. За начало отсчета возьмем 9.00. По условию $0 \le x \le 1$, $0 \le y \le 1$.

Всевозможные исходы будут являться точками квадрата со стороной 1. Встреча двух лиц произойдет, если разность между x и y не превзойдет 0,25 часа. Например, x=9.20, y=9.30 (первый дождался второго), а значит $|x\text{-}y| \le 0{,}25$. Это неравенство можно записать $—0{,}25 \le y\text{-}x \le 0{,}25$, а значит, $x—0{,}25 \le y \le x+0{,}25$. Изобразим данную область на координатной плоскости.

Искомая вероятность равна отношению заштрихованной полосы к площади всего квадрата. Площадь всего квадрата равна 1. Найдем площадь заштрихованной полосы: 1-2*0,5*0,75*0,75=0,4375 (площадь квадрата минус площади двух равных не заштрихованных треугольников).

Значит, $P = \dfrac{0{,}4375}{1}$

Ответ: P = 0,4375.

Задачи:

1. Два лица A и B договорились встретиться в определенном месте в промежутке времени от 11.00 до 12.00 часов. Каждый из них приходит наудачу, независимо от другого, и ожидает 30 минут. Какова вероятность, что они встретятся. (0,75).

2. Два лица A и B договорились встретиться в определенном месте в промежутке времени от 9.00 до 10.00 часов. Каждый из них приходит наудачу, независимо от другого, и ожидает 15 минут. Какова вероятность, что они встретятся?

Два лица A и B договорились встретиться в определенном месте в промежутке времени от 8.00 до 9.00 часов. Каждый из них приходит наудачу, независимо от другого, и ожидает 10 минут. Какова вероятность, что они встретятся? (11/36)

3. В круг радиуса R помещен меньший круг радиуса r. Найти вероятность того, что точка, наудачу брошенная в большой круг, попадет так же и в малый круг. Предполагается, что вероятность попадания точки в круг пропорциональна площади круга и не зависит от его расположения.

ОПЕРАЦИИ НАД ВЕРОЯТНОСТЯМИ

Теорема сложения вероятностей несовместных событий

Теорема. Вероятность появления одного из двух несовместных событий равна сумме вероятностей этих событий: $P(A \cup B) = P(A) + P(B)$.

Пример 7.

Для отправки груза со склада может быть выделена одна из двух машин различного вида. Известны вероятности выделения каждой машины: $P(A) = 0,2$; $P(B) = 0,4$. Тогда вероятность того, что к складу будет подана одна из этих машин $P(A \cup B) = 0,2 + 0,4 = 0,6$.

Теорема верна и для любого конечного числа несовместных событий.

$P\ (A_1 \cup A_2 \cup A_3 \cup _{\dots} \cup A_n) = P\ (A_1\ ) + P(A_2\ ) + P(A_3) +_{\dots} +P(A_n)$.

Пример 8.

В лотерее выпущено 10000 билетов и установлено 10 выигрышей по 200 рублей, 100 по 100 рублей, 500 по 25 рублей, 1000 по 5 рублей. Гражданин купил один билет. Какова вероятность того, что он выиграет не менее 25 рублей?

Решение.

A — выигрыш не менее 25 рублей;

A_1 —выигрыш равен 25 рублей, $P(A_1\ )=500/1000=0,05$

A_2 — выигрыш равен 100 рублей, $P(A_2\ )=100/1000=0,01$

A_3 — выигрыш равен 200 рублей, $P(A_3\ )=10/1000=0,001$.

Поскольку куплен только один билет, то $A= A_1 \cup A_2 \cup A_{3.}$ По теореме сложения вероятностей несовместных событий (события несовместны, так как куплен только один билет)

$P(A_1 \cup A_2 \cup A_3)=P(A_1)+P(A_2)+P(A_3)=0,05+0,01+0,001=0,061$.

Следствие из теоремы сложения: Сумма вероятностей противоположных событий равна 1.

Теорема сложения вероятностей совместных событий

Два события называются совместными, если появление одного из них не исключает появление другого в одном и том же испытании.

Теорема.

Вероятность появления хотя бы одного из двух совместных событий равна сумме вероятностей этих событий без вероятности их совместного появления.

$$P(A \cup B)=P(A)+P(B)-P(A \cap B).$$

Теорема умножения вероятностей независимых событий

Вероятность совместного появления двух независимых событий A и B равна произведению вероятностей этих событий

$P(A \cap B) = P(A)*P(B)$.

Теорема распространяется и для случая событий больше двух.

Пример 9.

Какова вероятность того, что при десятикратном бросании монеты герб выпадет 10 раз?

Решение.

Событие A — выпадение герба или надписи при десятикратном бросании не зависит от результата предыдущих бросаний. Поэтому здесь идет речь о совмещении десяти независимых событий: A_1 выпал герб при первом бросании, A_2 выпал герб при втором бросании и т.д. Вероятность выпадения герба при однократном бросании — ½, поэтому искомая вероятность равна:

$$P(A) = P(A_1)*P(A_2)*...*P(A_{10}) = S = \frac{1}{2}\cdot\frac{1}{2}...-\frac{1}{2} = \left(\frac{1}{2}\right)^{10} = \frac{1}{1024}$$

Теорема умножения вероятностей зависимых событий

Два события называются **зависимыми**, если вероятность появления одного из них зависит от наступления или ненаступления другого.

Пример 10.

Пусть в урне 3 белых и 10 черных шаров. Из урны наудачу извлекают один шар, а затем извлекают другой. Обозначим через событие A — при первом извлечении появился белый шар, через событие B — при втором извлечении появился белый шар. Если событие А произошло, то в урне из 12 оставшихся шаров есть 2 белых шара, поэтому $P(B)$=2/12, если A не произошло, то $P(B)$=3/12. Таким образом, вероятность события B зависит от появления или непоявления события A.

Условной вероятностью $P_A(B)$ называют вероятность события B, вычисленную в предположении, что событие A уже наступило.

Теорема. Вероятность совместного появления двух зависимых событий равна произведению вероятности одного из них на условную вероятность другого, вычисленную в предположении, что первое событие уже наступило

$P(A \cap B)=P(A)* P_A(B)$.

Эта теорема распространяется и на случай числа зависимых событий больше двух, например, для трех зависимых событий $P(A \cap B \cap C) = P(A) * P_A(B) * * P_{AB}(C)$, где $P_{AB}(C)$ — вероятность события C, вычисленная в предположении, что A и B уже произошли.

Пример 11.

В ящике a белых и b черных шаров, последовательно вынимают два шара. Какова вероятность того, что оба они черные?

Решение.

Событие A — «первый шар черный», B — «второй шар черный»,

$$P(A \cap B) = P(A)*P_A(B) \cdot P(A) = \frac{b}{a+b} \; P_A(B) = \frac{b-1}{a+b-1}$$

$$P(A \cap B) = \frac{b}{(a+b)} * \frac{(b-1)}{(a+b=1)}$$

Формула полной вероятности

Пусть требуется найти вероятность события A, которое может происходить вместе с одним из независимых событий $B_1, B_2, \ldots B_n$. Тогда

$$P(A) = P(B_1)*P_A(B_1) + P(B_2) * P_A(B_2) + \ldots + P(B_n) * P_A(B_n).$$

Пример 12.

В магазин поступает одна и та же продукция от первого предприятия в количестве 20 изделий, от второго предприятия — 10 и от третьего предприятия — 70. Вероятности некачественного изготовления изделий на предприятиях, соответственно, равны 0,02, 0,03 и 0,05. Случайным образом отбирается одно изделие. Требуется определить вероятность того, что это изделие некачественное.

Решение.

Событие A — выбранное изделие некачественное, события B_1, B_2, B_3 — выбор изделия из продукции соответствующего предприятия.

$P(B_1) = 0{,}2$, $P(B_2) = 0{,}1$, $P(B_3) = 0{,}7$

$P_A(B_1) = 0{,}02$, $P_A(B_2) = 0{,}03$, $P_A(B_3) = 0{,}05$, тогда

$P(A) = 0{,}2*0{,}02 + 0{,}1*0{,}03 + 0{,}7*0{,}05 = 0{,}042$

Вероятность повторения событий

Пусть в результате некоторого опыта может произойти или не произойти событие A. Опыт должен быть произведен n раз. Известно, что в каждом опыте вероятность появления события A равна p. Вероятность того, что при n испытаниях событие A появится ровно m раз равна:

$$P = C_n^m \, p^m \, (1 - p)^{n-m}$$

Пример 13.

Подбрасывают монету 10 раз. Какова вероятность двукратного появления герба?

Решение:

A — появление герба, n=10, m=2, p=1/2

$$P = C_{10}^2 \left(\frac{1}{2}\right)^2 \left(1 - \frac{1}{2}\right)^{10-2} = \frac{45}{1024}$$

Пример 14.

Вероятность того, что изделие не пройдет контроля, равна 0,1. Какова вероятность того, что среди 5 изделий не будет ни одного забракованного?

Решение.

A — изделие забраковано, $p(A) = 0{,}1$, $n = 5$, $m = 0$,

$P = C_5^0 \cdot 0{,}1^0 \cdot (1 - 0{,}1)^5 = 1 \cdot 1 \cdot 0{,}9^5 \approx 0{,}6$

Задачи для практики

1. В данном пункте имеют остановку трамваи шести маршрутов: №7, 12, 16, 24, 31, 49. Пассажир ждет трамвай либо 12, либо 16. Какова вероятность того, что первым подойдет трамвай нужного маршрута? (1/3)

2. Подбросили 2 игральные кости. Какова вероятность того, что сумма выпавших очков будет больше 5? (13/18)

3. Подбросили три монеты. Какова вероятность того, что хотя бы одна из них упадет гербом вверх? (7/8)

4. В лотерее 4 выигрышных билета и 96 пустых. Какова вероятность того, что на 10 купленных билетов выпадет хотя бы один выигрыш? (0,3439)

5. Работают три станка. Вероятность того, что в течение часа станок не потребует остановки, равна для первого — 0,85, для второго — 0,9 и для третьего — 0,95. Найти вероятность того, что в течение часа ни один станок не потребует остановки? (0,73)

6. Вероятность сбить самолет противника выстрелом из винтовки — 0,004. Найти вероятность уничтожения самолета при одновременной стрельбе из 25 винтовок. ($1-(0{,}996)^{25}$)

7. Вероятность попасть в цель равна 0,8. Какова вероятность того, что из пяти выстрелов будет хотя бы один промах? ($1-0{,}8^5$)

8. Вынимают 100 раз карту из колоды в 52 карты и кладут ее обратно. Какова вероятность того, что червонный валет не появится ни разу? $(51/52)^{100}$

9. Вынимают 100 раз карту из колоды и кладут ее обратно. Какова вероятность того, что червонный валет появится хотя бы один раз?($1-(51/52)^{100}$)

10. На военных учениях летчик получил задание уничтожить три склада. На борту самолета — одна бомба. Вероятность попадания в первый склад — 0,01, во второй — 0,08, в третий — 0,025. Любое попадание в результате детонации вызывает взрыв всех трех складов. Какова вероятность того, что склады противника будут уничтожены? (0,115)

11. Из колоды в 36 карт одну за другой вынимают 2 карты. Найти вероятность, что вынуты: а) 2 валета, б) 2 карты пиковой масти, в) вынуты валет и дама. (1/105, 2/35, 4/315)

12. В лотерее выпущено n билетов, из них m — выигрышных. Гражданин купил k билетов. Какова вероятность того, что по крайней мере один из купленных билетов — выигрышный?

13. Какова вероятность вытащить короля из колоды 2 раза подряд, если после первого извлечения карты ее не возвращают обратно? (1,221)

14. Вероятность того, что взятое наугад изделие является пригодным, равна 92/100. Вероятность того, что взятое наугад годное изделие является

изделием первого сорта, равна 72/100. Какова вероятность того, что взятое наугад изделие является изделием первого сорта? (0,6624)

15. В урне 4 белых и 7 черных шаров. Вынимают 2 шара, не возвращая обратно. Какова вероятность того, что первый шар белый, а другой черный? (14/55)

16. На 5 карточках написаны буквы *а, г, и, к, н*. Вынимают одну за другой карточки и кладут в том порядке, в каком они были вынуты. Какова вероятность того, что получится слово «книга»? (1/120)

17. На 7 карточках написаны буквы *к, к, л, л, о, о, о*. Вынимают одну за другой карточки и кладут в том порядке, в каком они были вынуты. Какова вероятность того, что получится слово «колокол»? (1/1050)

18. В ящике 10 белых и 8 красных шаров. Одновременно наугад вынимают 2 шара. Какова вероятность того, что они разных цветов? (40/153)

19. В ящике 7 белых и 9 черных шаров. Вынимают один шар и кладут его обратно, снова вынимают и снова кладут обратно. Какова вероятность того, что шары белые? (49/256).

20. В телевизоре 10 ламп, для любой лампы вероятность быть исправной в течение года равна р. Какова вероятность: а) что в течение года хотя бы одна лампа выйдет из строя; б) в течение года выйдут из строя ровно 2 лампы? ($48p^8(1-p)^2$)

ПРИБЛИЖЕННЫЕ ФОРМУЛЫ В СХЕМЕ БЕРНУЛЛИ

При большом числе опытов по схеме Бернулли удобнее пользоваться приближенными формулами.

Формула Пуассона

Если вероятность p наступления события A в каждом испытании постоянна и мала, число испытаний n-велико и число $\lambda=np$ — незначительно (будем полагать, что $\lambda=np\leq10$), то вероятность $P_{m,n}$ того, что событие A появится m раз в n независимых испытаниях:

$$P_{m,n} \approx \frac{\lambda^m e^{-\lambda}}{m!}$$

Пример 1.

Завод отправил 5000 доброкачественных изделий. Вероятность того, что в пути разбили одно изделие — 0,0002. Найти вероятность того, что в пути будет повреждено 3 изделия.

Решение:

$\lambda=np=5000*0{,}0002=1$. Применим формулу Пуассона:

$$P_{3,5000} \approx e^{-1}\frac{1^3}{3!}$$

$$P_{3,5000}\approx 1/6e=0{,}061.$$

Локальная и интегральная формулы Лапласа

Локальная формула Лапласа.

Если $npq \geq 10$, то вероятность $P_{m,n}$ того, что событие A появится m раз в n независимых испытаниях при достаточно большом числе n:

$$P_{m,n} \approx f(x)\frac{1}{\sqrt{npq}}; \quad f(x) = \frac{1}{\sqrt{2\pi}} * e^{-\frac{x^2}{2}}; \quad x = \frac{m-np}{\sqrt{npq}}$$

Для облегчения вычислений значения данной функции представляются в таблице.

Пример 2.

Стрелок выполнил 400 выстрелов. Найти вероятность 325 попаданий, если вероятность попадания при каждом выстреле равна 0,8.

Решение.

npq=400*0,8*0,2=64≥10, следовательно, применяем формулу Лапласа.

Найдем значение $x = \frac{325-400*0{,}8}{\sqrt{64}} = \frac{5}{8} = 0{,}63$

$P_{325,400} \approx f(0{,}63)*(1/8) = 0{,}3271*(1/8) \approx 0{,}041$.

Интегральная формула Лапласа.

Если вероятность наступления события A в каждом испытании постоянна и отлична от 0 и 1, то вероятность того, что число m наступления события в n независимых испытаниях заключено в пределах от a до b (включительно), при достаточно большом числе n равна

$$P_n = (a \leq m \leq b) \approx (0{,}5)*(\Phi(x_2) - \Phi(x_1)),$$

где $\Phi(x)$ — функция Лапласа вычисляемая по формуле

$$\hat{O}(x) = \frac{2}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$$

Значения данной функции обычно берут из таблицы.

$$x_1 = \frac{a-np}{\sqrt{npq}}; \quad x_2 = \frac{b-np}{\sqrt{npq}}$$

Функция $\Phi(x)$ — нечетная, $\Phi(-x) = -\Phi(x)$.

Следствие из интегральной формулы Лапласа:

Вероятность отклонения относительной частоты (частости) от постоянной вероятности в независимых испытаниях не более, чем на некоторое число $\varepsilon > 0$

Пример 3.

Стрелок выполнил 400 выстрелов, вероятность одного попадания равна 0,8. Найти вероятность того, что он попадет от 310 до 325 раз.

Решение.

$P_{400} = (310 \leq m \leq 325) \approx (0{,}5)*(\Phi(x_2) - \Phi(x_1))$,

$$x_2 = \frac{325-320}{\sqrt{64}} = \frac{5}{8} = 0{,}63$$

$$x_1 = \frac{310-320}{\sqrt{64}} = \frac{-10}{8} = -1{,}25$$

$P_{400}=(310 \leq m \leq 325) \approx (0,5)*(\Phi(0,63)-\Phi(-1,25))=(0,5)*(\Phi(0,63)+\Phi(1,25))=$

$=0,5(0,4713+0,7887)=0,63.$

Задачи

1. Вероятность того, что при сортировке изделий одно из них будет разбито, равна 0,005. Найти вероятность того, что из 200 изделий окажутся разбитыми три изделия. (0,061)

2. Вероятность поражения мишени при одном выстреле равна 0,8. Найти вероятность того, что при 100 выстрелах мишень будет поражена ровно 75 раз. (0,04565)

3. Известно, что 80% специалистов в районе имеет высшее образование. Найти вероятность того, что из 100 наудачу отобранных человек высшее образование имеет от 65 до 90 человек. (0,9937).

4. Установлено, что виноградник поражен вредителями в среднем на 10%. Определить вероятность того, что из 10 проверенных кустов винограда один будет поражен. Вычислить по формулам Бернулли, Лапласа, Пуассона. Сделать выводы. (0,387, 0,42, 0,368)

5. Всхожесть семян составляет 80%. Какова вероятность того, что из 1000 посеянных семян взойдут от 650 до 760?

6. Вероятность рождения мальчика равна 0,51. Найти вероятность того, что среди 100 новорожденных окажется 50 мальчиков.

7. Всхожесть зерна 90%. Определить вероятность того, что для отобранных, случайным образом, 100 зерен относительная частота всхожести будет отличаться от вероятности p=0,9 по абсолютной величине не более, чем на 0,1.

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

Виды случайных величин

Величина называется случайной, если в результате опыта она может принимать любое заранее неизвестное значение.

Случайные величины делятся на дискретные и непрерывные.

Дискретными называются величины, если их возможные значения представляют отдельные изолированные числа. Например, число очков при бросании игральной кости: 1, 2, 3, 4, 5, 6.

Если же возможные значения случайной величины сплошь заполняют некоторый промежуток, то такую случайную величину называют непрерывной. Например, расстояние, которое пролетит снаряд при выстреле из орудия, — непрерывная случайная величина.

Дискретные случайные величины

Распределение дискретной случайной величины

Пусть дискретная случайная величина X может принимать n значений x_1, $x_2,\ldots, x_n$. Для характеристики этой случайной величины должны быть заданы вероятности появления указанных значений $p_1, p_2,\ldots, p_n$.

Дискретные значения случайной величины и вероятности их появления удобно записывать в следующем виде:

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

Для дискретной случайной величины, так же как и для непрерывной вводится понятие функции распределения, которая представляет собой вероятность события $X < x$, где x — задаваемые непрерывно изменяющиеся значения, т.е.

$$F(x)=P(X<x).$$

Если дискретные значения случайной величины $x_1, x_2, \dots, x_n$ расположены в порядке возрастания, то каждому значению x_i этих величин ставится в соответствие сумма вероятностей всех предыдущих значений и вероятности p_i.

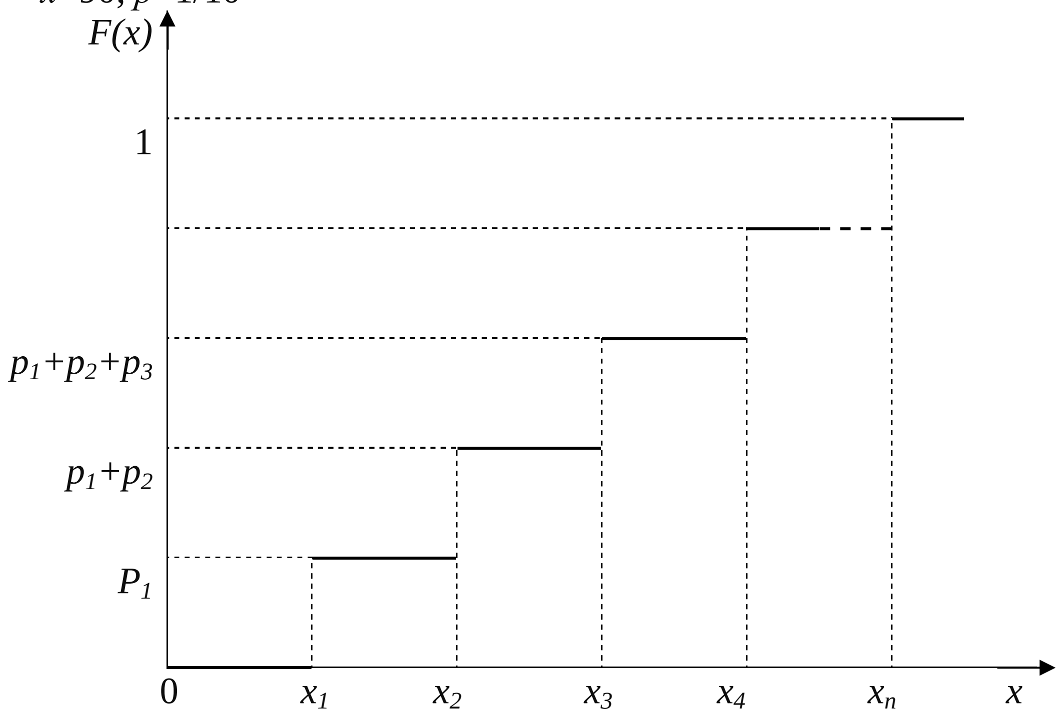
Нанося на график возможные дискретные значения случайной величины x и соответствующие суммы вероятностей, получаем ступенчатую фигуру, оторая и является графиком функции распределения вероятностей.

Пример 1.

Построить статистическую функцию распределения результатов 10 измерений.

I	1	2	3	4	5	6	7	8	9	10
X_i	30	10	10	20	20	60	70	90	60	50

x=10, p=2/10
x=20, p=2/10
x=30, p=1/10
x=50, p=1/10
x=60, p=2/10
x=70, p=1/10
x=90, p=1/10



Математическое ожидание случайной величины

Математическим ожиданием дискретной случайной величины X называется сумма произведений всех ее значений на соответствующие вероятности:

$$M(x) = x_1 p_1 + x_2 p_2 + \dots + x_n p_n.$$

Пример 2.

В магазин поступает ежедневно не более пяти радиоприемников. Известны вероятности их поступления: p_0=0,1, p_1=0,2, p_2=0,1, p_3=0,15, p_4=0,2, p_5=0,25.

Найти математическое ожидание числа поступлений радиоприемников.

Решение:

$M(x)$=0*0,1+1*0,2+2*0,1+3*0,15+4*0,2+5*0,25=2,9

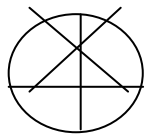
Математическое ожидание случайной величины — это постоянная величина, которая показывает - какое значение случайной величины можно ожидать в среднем при проведении серии опытов.

Пример 3.

Мишень установлена так, что может вращаться вокруг своей оси. При достаточно большой скорости стрелок не может различить сектора мишени, он вынужден стрелять наугад. Мишень поделена на 8 равных секторов. При попадании в первый сектор он выигрывает 1 рубль, во второй сектор — два рубля и т.д. Стоит ли участвовать в игре, если один выстрел стоит 5 рублей?

Случайная величина x принимает значения 1, 2, 3, 4, 5, 6, 7, 8. Так как мишень поделена на 8 равных частей, то вероятность попадания в каждый сектор рана 1/8.

X	1	2	3	4	5	6	7	8
p	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8



Математическое ожидание выигрыша подсчитываем по формуле:

$M(x)$=1*1/8+2*1/8+3*1/8+4*1/8+5*1/8+6*1/8+7*1/8+8*1/8=4,5.

Так как стоимость выстрела превышает математическое ожидание выигрыша, то стрелять невыгодно.

Свойства математического ожидания

$M(C)=C$

$M(Cx)=CM(x)$

$M(x+y)=M(x)+M(y)$

$M(x*y)=M(x)*M(y)$, где x и y независимые случайные величины.

Дисперсия

Рассмотрим две дискретные случайные величины X и Y. Первая принимает значения —1 и 1 с вероятностями 0,5. Вторая принимает значения —5 и 5 с теми же вероятностями 0,5. Математические ожидания этих величин одинаковы и равны 0:

$M(X)$=-1*0,5+(-1)*0,5=0

$M(Y)$=-5*0,5+(-5)*0,5=0.

Очевидно, что вторая величина сильнее отклоняется от своего математического ожидания в конкретных реализациях, чем первая. Чтобы учесть и оценить эти отклонения, используют понятие дисперсии.

Дисперсия случайной величины X вычисляется по формуле

$D(X) = M(X - M(X))^2 = M(X^2) - M^2(X)$

В рассмотренном выше случае $D(X) = (-1)^2 * 0,5 + 1^2*0,5 = 1$

$D(Y) = (-5)^2 * 0,5 + 5^2 * 0,5 = 25$

Свойства дисперсии

$D(C) = 0$

$D(CX) = C^2D(X)$

$D(X+Y) = D(X)+ D(Y)$, где X и Y — независимые случайные величины.

Пример 4.

Дано следующее распределение дискретной случайной величины:

X	1	2	4	5
P	0,2	0,1	0,4	0,3

Найти ее дисперсию.

Решение.

$M(X)$=1*0,2+2*0,1+4*0,4+5*0,3=3,5

$M(X^2)=1^2*0,2+2^2*0,1+4^2*0,4+5^2*0,3=14,5$

$D(X)=14,5-3,5^2=2,25$

Задания для практики

1. Найти математическое ожидание случайной величины попаданий при 5 выстрелах, если она задана рядом распределения:

X	0	1	2	3	4	5
P	0,01024	0,0768	0,2304	0,3456	0,2592	0,7776

2. Законы распределения случайных величин X и Y следующие:

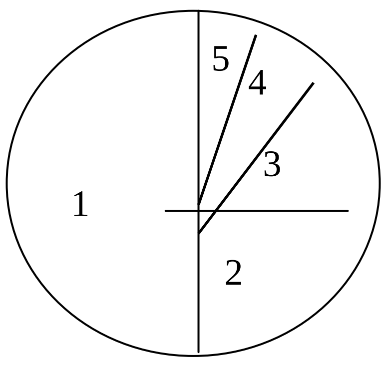
X	0	1	2	3	4	5	6	7
P	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

Y	1	2	3	4	5	6	7	8
P	1/4	1/8	1/10	1/16	1/16	1/16	1/8	1/4

Найти $M(X + Y)$, $M(X - Y)$, $M(X * Y)$. (8, -1, 15,75).

3. Автомобиль встретит 4 светофора, каждый из которых пропустит его с вероятностью 0,5. Найти математическое ожидание числа светофоров до первой остановки. (1,625)

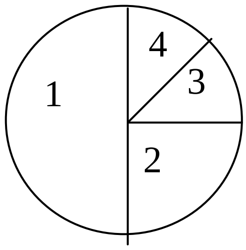
4. У охотника 4 патрона, он стреляет по зайцу, пока не попадет или пока не кончатся патроны. Найти математическое ожидание количества выстрелов, если вероятность попадания 0,25. (1,47)

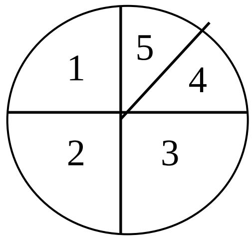


5. Мишень установлена так, что может вращаться вокруг оси. Стрелок стреляет наугад. При попадании в первый сектор (рис. справа) — выигрывает

1 рубль, во второй — проигрывает 2 рубля, в третий — выигрывает 3 рубля, четвертый — проигрывает 4 рубля, в пятый — выигрывает 5 рублей. Стоит ли участвовать в такой игре? (7/10)

6. Мишени установлены так, что могут вращаться вокруг оси. Стрелок стреляет наугад.





Секторы: 1 — выигрыш 1 рубль
2 — проигрыш 2 рубля
3 — выигрыш 3 рубля
4 — проигрыш 4 рубля
5 — ни выигрыш, ни проигрыш

1 — проигрыш 1 рубль
2 — проигрыш 2 рубля
3 — проигрыш 3 рубля
4 — выигрыш 4 рубля

Стоит ли участвовать в такой игре?

7. Пусть X — сумма очков, получаемая при бросании двух игральных костей. Постройте закон распределения и найдите математическое ожидание.

8. Распределение случайных величин имеет вид:

X	1	2	3	4	5	6	7	8
P	0,15	0,2	0,15	0,1	0,15	0,05	0,15	0,05
$У$	9	8	7	6	5	4	3	2
P	0,15	0,1	0,15	0,1	0,15	0,1	0,15	0,1

Найдите $M(x+y)$, $D(x+y)$. (9,5; 10,03).

ОСНОВНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЙ

Дискретные случайные величины

1. Биномиальное распределение

Биномиальное распределение — одно из распространенных дискретных распределений. Оно возникает в тех случаях, когда нас интересует, сколько раз происходит некоторое событие в серии из определенного числа независимых наблюдений (опытов), выполняемых в одинаковых условиях.

Например, рассмотрим массовое производство, при котором производятся как стандартные, так и дефектные детали. Пусть доля дефектных деталей будет в среднем равна p ($0<p<1$). Вероятность того, что среди n деталей окажется k бракованных, будет подсчитываться по формуле Бернулли: $P(X=k)=C_n^k p^k(1-p)^{n-k}$

Таким образом, процесс обнаружения бракованной детали подчиняется биномиальному закону распределения.

Опр. Случайная величина X имеет биномиальное распределение с параметрами n и p, если она принимает значения 0, 1,…, , n с вероятностями:

$P(X=k)=C_n^k p^k(1-p)^{n-k}$

$k=0, 1, …, n$.

Свойства:

Математическое ожидание и дисперсия случайной величины, имеющей биномиальное распределение, равны:

$M(X)=np$, $D(X)=np(1-p)$

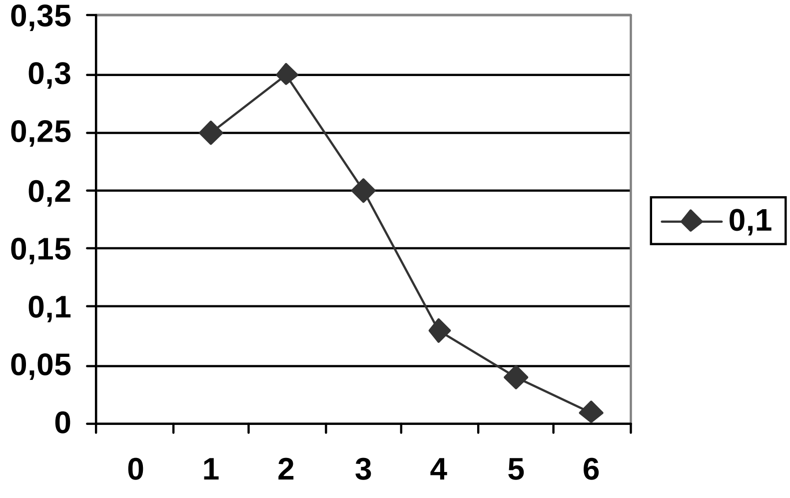


Рис. 1. Вид биномиального распределения для p=0,2 при n=10

2. *Распределение Пуассона*.

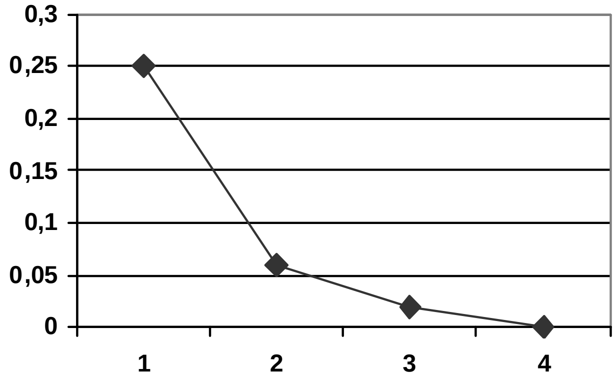
Распределение Пуассона играет важную роль в ряде вопросов физики, теории связи, теории надежности, теории массового обслуживания и т.д. — там, где в течение определенного времени может происходить случайное число каких-то событий (телефонных вызовов, отказов оборудования и т.д.). При данном распределении вероятность появления события за малый интервал времени пропорциональна длине этого интервала. Если за данный интервал времени уже произошло одно событие, то условная вероятность появления в этом же интервале другого события стремиться к нулю.

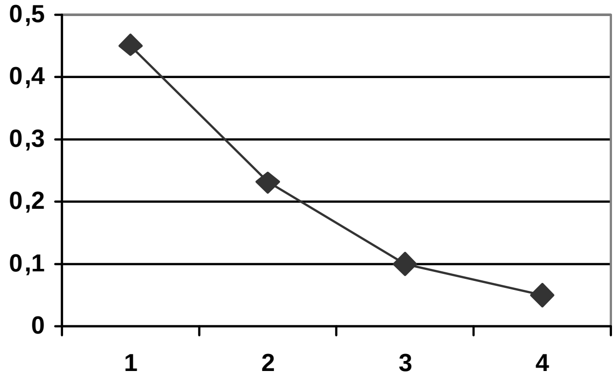
Опр. Дискретная случайная величина X имеет закон распределения Пуассона. Если она принимает значения 0, 1, …, m ,… с вероятностями

$$P(X=m)=\frac{\lambda^m e^{-\lambda}}{m!}$$

Свойства: Математическое ожидание и дисперсия случайной величины, распределенной по закону Пуассона, совпадают и равны параметру λ этого закона, т.е. $M(X)=\lambda$, $D(X)=\lambda$.

λ=0,5 λ=1





При достаточно большом n и близким к нулю p распределение Пуассона хорошо аппроксимирует биномиальное распределение.

Непрерывные случайные величины

Для непрерывной случайной величины X вероятность того, что $P(X=x_i)\to 0$, поэтому удобнее использовать вероятность того, что $X< x_i$, где x_i — текущее значение переменной. Эта вероятность называется функцией распределения:

$P(X< x_i)=F(x)$ (интегральной).

Свойства функции распределения:

1. $F(x)$ не убывает (если $x_2> x_1,$ то $F(x_2)\geq F(x_1)$)
2. $F(-\infty)=0$
3. $F(+\infty)=1$
4. Вероятность попадания случайной величины X в интервал $a<X<b$ определяется по формуле:

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

Случайная величина непрерывна, если ее функция распределения непрерывна на всей числовой оси.

Производная функции распределения случайной величины называется функцией плотности вероятности: $f(x)=F'(x)$ (дифференциальная функция распределения).

Свойства функции плотности:

1. $f(x)\geq 0$
2. $F(x)=\int\limits_{-\infty}^{x} f(x)dx$
3. $F(x)=\int\limits_{-\infty}^{\infty} f(x)dx=1$

Аналогом графика функции плотности является полигон распределения для дискретной случайной величины.

Числовые характеристики непрерывной случайной величины

1. Математическое ожидание непрерывной случайной величины определяется по формуле

$$M(X)=\int\limits_{-\infty}^{\infty} xf(x)dx$$

2. Если непрерывная случайная величина определена на интервале $(a;b)$

$$M(X)=\int\limits_{a}^{b} xf(x)dx$$

3.Дисперсия непрерывной случайной величины определяется по формуле:

$$D(X)=\int\limits_{-\infty}^{\infty}(x-M(x))^2 f(x)dx$$

$$D(X)=\int\limits_{-\infty}^{\infty}x^2 f(x)dx-(M(X))^2$$

Асимметрия и эксцесс

Коэффициент асимметрии: $A=(M(x-M(x)))^3/\sigma^3$;

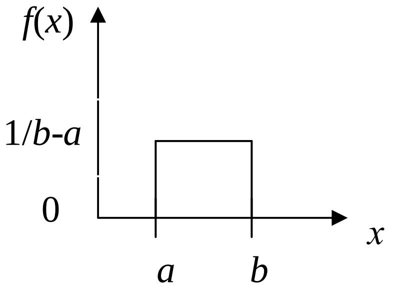
Эксцесс: $Ex=(M(x-M(x)))^4/\sigma^4-3$

где $\sigma=\sqrt{D}$ среднеквадратическое отклонение. В зависимости от значений асимметрии график плотности имеет положительную или отрицательную асимметрию, в зависимости от знака коэффициента эксцесса распределение имеет более заостренную, либо более плоскую вершину.

ОСНОВНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

Равномерный закон распределения

Случайная величина X распределена по равномерному закону, если все ее значения лежат внутри некоторого интервала $(a;b)$ и все они равновероятны. На этом интервале плотность вероятности равна постоянному числу $f(x)=1/b-a$, вне этого интервала $f(x)=0$. График плотности вероятности имеет вид:



$M(X)=(b+a)/2$; $D(X)=(b-a)^2$

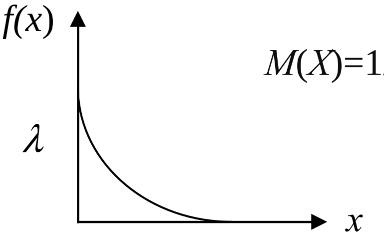
1. Показательное распределение

Показательным называют распределение вероятностей непрерывной случайной величины X, которое описывается плотностью

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

где λ — постоянная положительная величина. Показательное распределение играет большую роль при статистических исследованиях медико-биологических процессов, связанных с данными типа «времени жизни»

График плотности вероятности имеет вид:



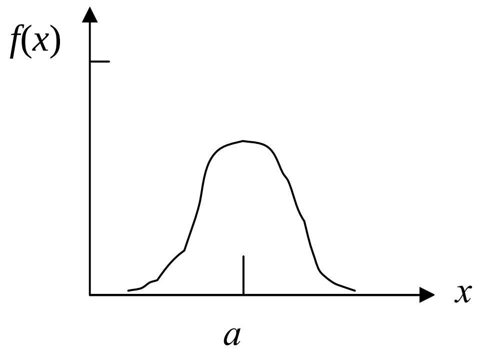
$M(X)=1/\lambda$; $D(X)=1/(\lambda)^2$

2. Нормальный закон распределения

Нормальный закон распределения играет исключительную роль в теории вероятностей. Это наиболее часто встречающийся закон распределения, главной особенностью которого является то, что он является предельным законом, к которому, при определенных условиях, приближаются другие законы распределения.

Нормальным называют распределение вероятностей непрерывной случайной величины X, плотность которого имеет вид, где a — математическое ожидание, σ — среднеквадратическое отклонение X.

График функции плотности имеет вид:



Свойства плотности нормального распределения:

1. Область определения плотности R.
2. Ось Ox — горизонтальная асимптота.
3. $x=a\pm\sigma$ — две точки перегиба.
4. Максимум в точке с координатами а; $1/(\sigma\sqrt{2\pi})$
5. Вероятность попадания нормально распределенной случайной величины в заданный интервал определяется по свойству функции распределения
6. Вероятность попадания нормально распределенной случайной величины в заданный интервал определяется по формуле:

$$P(\left|X-a\right|)<\varepsilon = 2\hat{O}\left(\frac{\varepsilon}{\sigma}\right)$$

где $\Phi(x)$ — функция Лапласа.

Правило трех сигм

Найдем вероятность того, что нормально распределенная случайная величина X отклонится от $M(X)$ на σ, 2σ, 3σ.

$P(\left|x-a\right|<\sigma)=2\Phi(1)=0{,}6826$,

$P(\left|x-a\right|<2\sigma)=2\Phi(2)=0{,}9544$,

$P(\left|x-a\right|<3\sigma)=2\Phi(3)=0{,}9973$.

Если случайная величина X имеет нормальное распределение, то отклонение этой случайной величины от ее математического ожидания по абсолютной величине не превышает утроенное среднеквадратическое отклонение.

Асимметрия и эксцесс нормального распределения.

Коэффициент асимметрии: $A=(M(x-M(x)))3/\sigma^3=0$

Эксцесс: $Ex=(M(x-M(x)))^4/\sigma^4-3=0$

Примеры и задачи

1. Случайная величина X задана плотностью распределения $f(x)=2x$ в интервале (0,1); вне этого интервала $f(x)=0$. Найти математическое ожидание величины X.

Решение: Используем формулу: $M(X)=\int_a^b xf(x)dx$

$$M(X)=2\int_0^1 xxdx=2/3$$

2. Математическое ожидание и среднее квадратическое отклонение нормально распределенной случайной величины X, соответственно равны 10 и 2. Найти вероятность того, что в результате испытания X примет значение, заключенное в интервале (12;14).

Решение: Воспользуемся формулой:

Подставив $\alpha=12$, $\beta=14$, $a=10$, $\sigma=2$, получим $P\,(12 < X < 14) = \Phi(2) - \Phi(1) = 0{,}1359$

Задачи

1. Вероятность рождения мальчика равна 0,515. Составить закон распределения случайной величины X — числа мальчиков в семьях, имеющих четырех детей. Построить график распределения вероятностей. Найти математическое ожидание, дисперсию и среднеквадратическое отклонение.
2. Автобусы некоторого маршрута идут строго по расписанию. Интервал движения — 5 минут. Найти вероятность того, что пассажир, подошедший к остановке, будет ожидать очередной автобус менее 3 минут (0,6).
3. Автомат штампует детали. Контролируется длина детали X, которая распределена нормально с математическим ожиданием (проектная длина), равным 50 мм. Фактическая длина изготовленных деталей не менее 32 и не более 68 мм. Найти вероятность того, что длина наудачу взятой детали: а) больше 55 мм; б) меньше 40 мм.
4. Написать плотность и функцию распределения показательного закона, если параметр $\lambda=5$.
5. Непрерывная случайная величина X распределена по показательному закону, заданному плотностью вероятности $f(x)=3e^{-3x}$ при $x>0$, при $x<0$ $f(x)=0$. Найти вероятность того, что в результате испытания X попадет в интервал (0,13;0,7).

Закон больших чисел по Колмогорову

Совокупное действие большого числа случайных факторов приводит к результату, почти не зависящему от случая, т.е. при большом числе случайной величины их средней результат перестает быть случайным и может быть предсказан с определенной долей вероятности.

Теорема Бернулли

Частость события в n повторных испытаниях, в каждом из которых оно может произойти с одной и той же вероятностью *p* при неограниченном увеличении числа опытов (*n*) сходится по вероятности к вероятности p этого события в отдельном испытании.

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}\right| \le \varepsilon\right) = 1$$

Неравенство Чебышева

Для любой случайной величины.

$$P\left(\left|x-a\right| \succ \varepsilon\right) \le \frac{D(x)}{\varepsilon^2}$$

ПЛАНЫ ПРОВЕДЕНИЯ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

Практическое занятие 1. «Множества. Комбинаторика».
Решение задач № 1-11 с.6, №1-4 с.9. Кит Ю.В. Теория вероятностей в примерах и задачах.
На с/р: №5-9 с.9. Кит Ю.В. Теория вероятностей в примерах и задачах.

Практическое занятие 2. «Выборки с повторениями. Алгебра событий».
Решение задач №1-8 с.12, №1, 2 с.15. Кит Ю.В. Теория вероятностей в примерах и задачах.
На с/р: №9-15 с.13, №3 с.15. Кит Ю.В. Теория вероятностей в примерах и задача.

Практическое занятие 3. «Вероятность события. Основные теоремы теории вероятностей».
Решение задач №1-15 с.22. Кит Ю.В. Теория вероятностей в примерах и задачах.
№12-19. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.:»Высшая школа». — 2002.
На с/р: №16-20 с.22. Кит Ю.В. Теория вероятностей в примерах и задачах.
№1.39-1.46. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.

Практическое занятие 4. «Методы вычисления вероятностей».
Решение задач: №50-60. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.:»Высшая школа». — 2002.
На с/р: №1.1.47-1.54, 1.75, 1.76. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.

Практическое занятие 5. Контрольная работа №1.

Практическое занятие 6. «Дискретные и непрерывные случайные величины».
Решение задач: №154-172. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.: «Высшая школа». — 2002.
Решение задач: №1-8 с.27. Кит Ю.В. Теория вероятностей для гуманитариев в примерах и задачах.
На с/р: 2.13-2.15, 3.25-3.27. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.

Практическое занятие 7. «Основные законы распределения. Предельные теоремы теории вероятностей».
Решение задач: №309, 314, 330, 334. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.: «Высшая школа». — 2002.

На с/р: № 4.11-4.17, 6.10-6.12. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.

Практическое занятие 8. «Системы случайных величин».
Решение задач: №374, 393, 408, 434. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.: «Высшая школа». — 2002.
На с/р: №5.10, 5.11. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.

Практическое занятие 9. «Элементы теории массового обслуживания».
Решение задач: №7.10, №7.11. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.

САМОСТОЯТЕЛЬНАЯ РАБОТА

Методические указания

Организация самостоятельной работы студентов имеет цель:

- систематизировать и расширить их теоретические знания;
- закрепить практические и организаторские способности;
- научить работать с учебной и научной литературой;
- стимулировать профессиональный рост студентов, воспитывать творческую активность и инициативу.

Самостоятельная работа студентов организуется преподавателями в соответствии с календарным планом изучения дисциплины и предполагает изучение лекционного материала, чтение рекомендуемых литературных источников, решение задач, ответы на контрольные вопросы или тесты и т.д.

Самостоятельная работа представляет собой дополнительное изучение дисциплины для полного и глубокого усвоения материала на основе анализа учебной, методической и дополнительной литературы.

Самостоятельная работа студентов включает повторение пройденного материала и подготовку к контрольной работе. Вся самостоятельная работа студентов оценивается в течение семестра на фактических занятиях и учитывается при сдаче экзамена.

Повторение пройденного материала осуществляется в процессе выполнения домашнего задания и самостоятельной проработки теоретического материала. Домашняя работа выполняется в соответствии с номерами заданий, представленных в плане практических заданий, методические рекомендации к которым изложены в учебном пособии.

В этом же пособии после каждого теоретического пункта определены те литературные источники (с указанием номеров страниц), по которым можно углубить свои знания по данным вопросам самостоятельно.

Контрольная работа
Методические рекомендации

При выполнении контрольной работы номер варианта совпадает с последней цифрой номера вашей зачетной книжки.

1. Работа выполняется в тетради в клетку. Титульный лист оформляется в соответствии с требованиями деканата по оформлению контрольных работ. В графе «работу проверил» следует писать: к.п.н. Кит Ю.В. На титульном листе указывается номер варианта.

2. Условия, все пояснения и формулы следует писать полностью. В выводах по заданию необходимо указать интерпретацию полученных числовых значений.

I. «Основы теории множеств»

1. Найти $A \cup B$, $A \cap B$, A/B, B/A, если A и B следующие:

Вариант	*A*	*B*
1	2, 3, 4, 5	5, 6, 7, 8
2	3, 4, 5, 6	6, 7, 8, 9
3	4, 5, 6, 7	7, 8, 9, 10
4	5, 6, 7, 8	8, 9, 10, 11
5	6, 7, 8, 9	9, 10, 11, 12
6	7, 8, 9, 10	10, 11, 12, 13
7	8, 9, 10, 11	11, 13, 14, 15
8	9, 10, 11, 12	10, 14, 15, 16
9	10, 11, 12, 13	11, 15, 16, 17
10	11, 13, 14, 15	15, 16, 17, 18
11	13, 14, 15, 16	14, 17, 18, 19
12	14, 15, 16, 17	17, 18, 19, 20
13	15, 16, 17, 18	18, 19, 20, 21
14	16, 17, 18, 19	19, 20, 21, 22
15	17, 18, 19, 20	20, 21, 22, 23
16	18, 19, 20, 21	21, 22, , 23, 24
17	19, 20, 21, 22	22, 23, 24, 25
18	20, 21, 22, 23	21, 5, 6, 45
19	21, 22, , 23, 24	23, 25, 65, 30
20	22, 23, 24, 25	23, 26, 28, 29

2. Найти $A \cup B$, $A \cap B$, A/B, B/A, если A и B следующие:

Вариант	*A*	*B*
1	2, 3, 4, 5	2, 3, 4, 5, 6, 7, 8
2	3, , 4, 5, 6	3, , 4, 5, 6, 7, 8, 9
3	4, 5, 6, 7	4, 5, 6, 7, 8, 9, 10
4	5, 6, 7, 8	5, 6, 7, 8, 9, 10, 11
5	6, 7, 8, 9	6, 7, 8, 9, 10, 11, 12
6	7, 8, 9, 10	7, 8, 9, 10, 11, 12, 13
7	8, 9, 10, 11	8, 9, 10, 11, 12, 13, 14
8	9, 10, 11, 12	9, 10, 11, 12, 13, 14, 15
9	10, 11, 12, 13	10, 11, 12, 13, 14, 15, 17
10	11, 13, 14, 15	11, 13, 14, 15, 17, 16, 19
11	13, 14, 15, 16	13, 14, 15, 16, 17, 18, 19
12	14, 15, 16, 17	14, 15, 16, 17, 15, 16, 17, 18
13	15, 16, 17, 18	15, 16, 17, 18, 19, 20, 21
14	16, 17, 18, 19	16, 17, 18, 19, 20, 21, 22
15	17, 18, 19, 20	17, 18, 19, 20, 21, 22, 23
16	18, 19, 20, 21	18, 19, 20, 21, 22, 23, 24
17	19, 20, 21, 22	19, 20, 21, 22, 23, 32, 45
18	20, 21, 22, 23	20, 21, 22, 23, 22, 45, 67
19	21, 22, 23, 24	21, 22, 23, 24, 25, 26, 27
20	22, 23, 24, 25	22, 23, 24, 25, 26,27,28

II. «Комбинаторика»

Варианты 1, 3, 5, 7, 9, 11, 13, 15, 17, 19

1. В правлении акционерного общества **m** человек. Необходимо избрать председателя правления, его заместителя и казначея. Сколькими способами можно образовать эту руководящую тройку, если одно лицо может занимать только один пост?
2. Сколько разных буквосочетаний можно образовать при перестановке букв следующего слова?
3. В магазине имеются фрукты n видов. Сколькими способами можно сформировать подарочный набор к празднику пожилого человека из m фруктов?
4. Сколько разных n-значных чисел можно составить из следующих цифр при условии, что ни одна цифра не повторяется?
5. Для работы с инвалидами группе социальных работников необходимо командировать m сотрудников. Сколькими способами это можно сделать, если весь штат составляет n человек?
6. Сколько телефонных номеров можно составить при условии, что первые две цифры 4 и 2, всего цифр в номере n?

Варианты 2, 4, 6, 8, 10, 12, 14, 16, 18, 20

1. В группе m человек, необходимо избрать старосту, культурно-массовый сектор и профсоюзного лидера. Сколькими способами можно образовать эту руководящую тройку, если одно лицо может занимать только один пост?
2. Сколько разных буквосочетаний можно образовать при перестановке букв следующего слова?
3. В магазине имеются подарочные товары n разных наименований. Сколькими способами можно сформировать подарочный набор к празднику пожилого человека из m предметов?
4. Сколько разных n-значных чисел можно составить из следующих цифр при условии, что ни одна цифра не повторяется?
5. Для участия в теледебатах депутатам нижней палаты Госдумы необходимо избрать делегацию в составе m человек, а всего на заседании присутствует n депутатов. Сколько способов избрания делегации существует?
6. Сколько машин можно обеспечить n-значными номерами при условии, что последние две цифры 1 и 6?

Задача	**1**	**2**	**3**		**4**		**5**		**6-**
Вариант	***m***	**слово**	***m***	***n***	***n***	**цифры**	***m***	***n***	***n***
1, 2	10	Уравнение	3	5	3	1, 2, 3	3	10	6
3, 4	12	Приближение	4	6	4	2, 3, 4, 5	4	12	7
5, 6	14	Измерение	5	7	5	3, 4, 5, 6, 7	5	14	5
7, 8	15	Неравенство	6	8	3	3, 4, 5	6	13	8
9, 10	16	Извлечение	7	9	4	3, 4, 5, 6	7	15	5
11, 12	18	Теорема	8	10	5	5, 6, 7, 8	8	15	6
13, 14	20	Вычисление	4	7	3	6, 7, 8	9	16	7
15, 16	24	Окружность	3	6	4	6, 7, 8, 9	3	18	8

Задача	1	2	3		4		5		6-
Вариант	***m***	**слово**	***m***	***n***	***n***	**цифры**	***m***	***n***	***n***
17, 18	25	Развитие	5	8	5	4, 5, 6, 7, 8, 9	4	20	5
19, 20	28	Степень	6	9	3	2, 5, 7	5	20	6

III. «Основы теории вероятностей»

Варианты 1, 5, 9, 13, 17

1. Бросают две игральные кости. Найти вероятность того, что сумма выпавших цифр будет

а) = m, б) < m.

2. Из колоды в 32 карты наугад вынимают одну за другой 2 карты. Найти вероятность того, что:

а) вынуты 2 короля,

б) вынуты 2 карты пиковой масти,

в) вынуты валет и дама.

3. В цехе n станков. Для любого станка вероятность того, что он останется исправным в течение месяца, равна р. Какова вероятность того, что:

а) в течение месяца хотя бы один станок выйдет из строя,

б) в течение месяца выйдет из строя ровно m станков.

Варианты 2, 6, 10, 14, 18

1. Бросают кубик два раза. Найти вероятность того, что сумма выпавших цифр будет

а) = m; б) > m**.**

2. Из колоды в 36 карт наугад одну за другой вынимают 2 карты. Найти вероятность того, что:

а) вынуты 2 дамы,

б) вынуты 2 карты червовой масти,

в) вынуты король и дама.

3. На предприятии работает n человек. Любой человек не заболеет в течение месяца с вероятностью p. Какова вероятность того, что:

а) в течение месяца хотя бы один человек заболеет,

б) в течение месяца заболеет ровно m человек.

Варианты 3, 7, 11, 15, 19

1. Бросают две игральные кости. Найти вероятность того, что сумма выпавших цифр будет

а) = m; б) < m

2. Из колоды в 48 карт наугад одну за другой вынимают 2 карты

Найти вероятность того, что:

а) вынуты 2 валета,

б) вынуты 2 карты пиковой масти,

в) вынуты валет и дама.

3. В цехе n станков. Для любого станка вероятность того, что он останется исправным в течение месяца, равна p. Какова вероятность того, что:

а) в течение месяца хотя бы один станок выйдет из строя,

б) в течение месяца выйдет из строя ровно m станков.

Вариант 4, 8, 12, 16, 20

1. Бросают кубик два раза. Найти вероятность того, что сумма выпавших цифр будет

а) = m; б) > m.

2. Из колоды в 52 карты наугад одну за другой вынимают 2 карты. Найти вероятность того, что:

а) вынуты 2 валета,

б) вынуты 2 карты бубновой масти,

в) вынуты король и валет.

3. На предприятии работает n человек. Любой человек не заболеет в течение месяца с вероятностью p. Какова вероятность того, что:

а) в течение месяца хотя бы один человек заболеет,

б) в течение месяца заболеет ровно m человек.

Задача	**1**	**3**		
Вариант	***m***	***m***	***n***	***p***
1, 2, 3, 4	5	3	12	0,1
5, 6, 7, 8	6	4	15	0,2
9, 10, 11, 12	7	6	20	0,25
13, 14, 15, 16	8	5	25	0,3
17, 18, 19, 20	9	2	10	0,35

IV. «Элементы статистики»

1. Найдите математическое ожидание и дисперсию случайной величины x, распределенной по следующему закону:

Ва-риант	***x***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	*p*	0, 1	0, 1	0, 1	0, 1	0, 1	0.2	0.1	0, 2
2	*p*	0, 2	0, 2	0, 1	0, 1	0, 1	0.1	0.1	0, 1
3	*p*	0, 3	0	0, 1	0, 1	0, 1	0.2	0.1	0, 1
4	*p*	0, 2	0, 1	0, 3	0, 1	0, 1	0	0.1	0, 1
5	*p*	0	0, 1	0, 1	0, 3	0, 1	0.2	0.1	0, 1
6	*p*	0, 4	0, 1	0, 1	0	0	0.2	0.1	0, 1
7	*p*	0, 1	0, 2	0, 1	0, 1	0, 1	0.2	0.1	0, 1
8	*p*	0, 15	0, 15	0, 1	0, 1	0, 1	0.2	0.1	0, 1
9	*p*	0, 2	0, 1	0, 1	0, 1	0	0.2	0	0, 3
10	*p*	0	0, 2	0, 2	0, 1	0, 1	0.2	0.1	0, 1
11	*p*	0, 2	0, 15	0, 05	0, 1	0, 1	0.2	0.1	0, 1
12	*p*	0	0, 1	0, 2	0, 1	0, 1	0.2	0.2	0, 1
13	*p*	0, 2	0, 15	0, 1	0, 1	0, 1	0.2	0.05	0, 1
14	*p*	0, 2	0, 1	0, 1	0, 1	0, 1	0	0.1	0, 3

Вариант	x	1	2	3	4	5	6	7	8
15	p	0, 4	0	0, 1	0	0, 1	0.2	0.1	0, 1
16	p	0, 1	0, 1	0, 1	0, 2	0, 1	0.2	0.1	0, 1
17	p	0, 2	0, 1	0, 1	0, 1	0, 1	0.2	0.1	0, 1
18	p	0, 25	0, 1	0, 1	0, 05	0, 1	0.2	0.1	0, 1
19	p	0, 15	0, 15	0, 1	0, 1	0, 1	0.2	0.1	0, 1
20	p	0, 3	0, 1	0	0, 1	0, 1	0.2	0.1	0, 1

2. Построить ряд распределения и вычислить математическое ожидание и дисперсию для числа попаданий при стрельбе по мишени до первого попадания, если вероятность попадания при одном выстреле равна p. Количество патронов равно 3.

Вар.	1	2	3	4	5	6	7	8	9	10
p	0,1	0,2	0,3	0,4	0,1	0,2	0,3	0,4	0,1	0,2
Вар.	11	12	13	14	15	16	17	18	19	20
p	0,3	0.4	0,5	0,1	0.2	0,3	0,4	0,5	0,4	0,6

3. Построить статистическую функцию распределения 10 измерений.

Вариант	i	1	2	3	4	5	6	7	8	9	10
1	X_i	20	40	50	10	40	50	70	70	40	30
2	X_i	10	30	20	30	10	80	80	70	50	50
3	X_i	20	20	50	60	20	40	30	70	50	50
4	X_i	30	10	10	20	20	60	70	80	50	50
5	X_i	20	40	50	10	40	50	70	70	40	30
6	X_i	10	30	20	30	10	80	80	70	50	50
7	X_i	20	20	50	60	20	40	30	70	50	50
8	X_i	30	10	10	20	20	60	70	80	50	50
9	X_i	20	40	50	10	40	50	70	70	40	30
10	X_i	10	30	20	30	10	80	80	70	50	50
11	X_i	20	20	50	60	20	40	30	70	50	50
12	X_i	30	10	10	20	20	60	70	80	50	50
13	X_i	20	40	50	10	40	50	70	70	40	30
14	X_i	10	30	20	30	10	80	80	70	50	50
15	X_i	20	20	50	60	20	40	30	70	50	50
16	X_i	30	10	10	20	20	60	70	80	50	50
17	X_i	20	40	50	10	40	50	70	70	40	30
18	X_i	10	30	20	30	10	80	80	70	50	50
19	X_i	20	20	50	60	20	40	30	70	50	50
20	X_i	20	40	50	10	40	50	70	70	40	30

КОНТРОЛЬ ЗНАНИЙ СТУДЕНТОВ
Контрольный тест для промежуточной аттестации студентов
Вариант 1.

1. Выберите формулу вероятности события:
1) $P(H)=m*n$; 2) $P(H)=m+n$; 3) $P(H)=m/n$; 4) $P(H)=m-n$.

2. Вероятность классическая отличается от статистической тем, что:
1) классическая вероятность вычисляется после опыта;
2) статистическая вероятность вычисляется до опыта;
3) статистическая вероятность вычисляется до опыта, а классическая вероятность вычисляется после опыта;
4) классическая вероятность вычисляется до опыта, а статистическая вероятность вычисляется после опыта.

3. Вероятность того, что произошло событие *A* или событие *B* при условии, что события независимы:
1) $\text{Đ}(A \cap B) = P(A) + P(B)$; 2) $\text{Đ}(A \cap B) = P(A)*P(B)$; 3) $\text{Đ}(A \cup B) = P(A)*P(B)$;
4) $\text{Đ}(A \cup B) = P(A) + P(B)$.

4. Вероятность того, что произошло событие *A* или событие *B* при условии, что события несовместны:
1) $\text{Đ}(A \cap B) = P(A) + P(B)$; 2) $\text{Đ}(A \cap B) = P(A)*P(B)$; 3) $\text{Đ}(A \cup B) = P(A)*P(B)$;
4) $\text{Đ}(A \cup B) = P(A) + P(B)$.

5. Математическое ожидание дискретной случайной величины подсчитывается по формуле
а) $M(x)=x_1p_1 + x_2p_2 + \ldots + x_np_n$
б) $M(x)=x_1 + x_2 + \ldots + x_n$
в) $M(x)=p_1 + p_2 + \ldots + p_n$
г) $M(x) = \frac{x_1}{p_1} + \frac{x_2}{p_2} + \ldots + \frac{x_n}{p_n}$
где $x_1 + x_2 + \ldots + x_n$ — случайные величины; $p_1 + p_2 + \ldots + p_n$ — вероятности их проявления, соответственно.

6. Математическое ожидание является аналогом:
1) дисперсии. 2) среднего, 3) разброса, 4)вероятности.

7. Дискретная случайная величина:
1) заполняет промежуток, 2) плавная, 3) отдельные изолированные числа, 4) независимая.

8. Непрерывная случайная величина:
1) заполняет промежуток, 2) плавная, 3) отдельные изолированные числа, 4) независимая.

9. Дисперсия является мерой:
1) математического ожидания; 2) среднего, 3) разброса, 4)вероятности.

10. Дисперсия дискретной случайной величины подсчитывается по формуле:
а) $D(x) = M(x^2) — M^2(x)$
б) $D(x) = M(x^2)$

в) $D(x) = M^2(x)$
г) $D(x) = M(x^2) + M^2(x)$
где $M(x)$ — математическое ожидание случайной величины х.

11. График функции плотности нормального распределения имеет форму: 1) параболы, 2) колокола, 3)вогнутой дуги, 4) гиперболы.

12. Для нормального закона распределения средних значений по отношению к крайним значениям:
1) одинаково, 2) меньше, 3) независимо, 4) больше.

13. Найдите вероятность выпадения четырех очков при бросании двух кубиков:
а) 3/36, б) 4/36, в) 1/2, г) 1,85.

14. Найдите вероятность, что из букв «*з*», «*к*», «*п*», «*а*», «*о*», «*н*» сложится слово «закон»:
а) 120, б) 1/120, в) 120, г) 2/1296.

15. Статистическая вероятность попадания в цель при 50 выстрелах равна 0,5. Какова вероятность попадания при 100 выстрелах:
а) 1, б) 0,5, в) 1/100, г) 190/950?

16. Какова вероятность, что при жеребьевке из номеров от 1 до 60 Вам не достанется номер, содержащий цифру 7:
а) 7/60, б) 1/10, в) 6/61, г) 1/2730?

17. Бросают игральную кость. Найти вероятность того, что выпавших очков будет «6»:
а) 1/3, б) 1/7, в) 1/6, г)1/2.

18. В магазине имеются подарочные товары пяти наименований.

Какой формулой необходимо воспользоваться для подсчета способов формирования подарочного набора к празднику пожилого человека из семи предметов:

а) $C_n^m = \dfrac{(m+n\text{-}1)!}{m!(n\text{-}1)!}$;

б) $C_n^m = \dfrac{n!}{m!(n\text{-}m)!}$;

в) $P_n = n!$;

г) $A_n^m = n^m$?

19. Сумма вероятностей появления различных значений дискретной величины х:
а) 1, б) -1, в) 0, г) 2.

Вариант 2

1. Математическое ожидание дискретной случайной величины подсчитывается по формуле:
а) $M(x)=x_1p_1 + x_2p_2 + \ldots + x_np_n$
б) $M(x)=x_1 + x_2 + \ldots + x_n$
в) $M(x)=p_1 + p_2 + \ldots + p_n$

г) $M(x) = \frac{x_1}{p_1} + \frac{x_2}{p_2} + ... + \frac{x_n}{p_n}$

где $x_1 + x_2 + \ldots + x_n$ — случайные величины; $p_1 + p_2 + \ldots + p_n$ — вероятности их проявления, соответственно.

2. Математическое ожидание является аналогом:
1) дисперсии, 2) среднего, 3) разброса, 4) вероятности.

3. Дискретная случайная величина:
1) заполняет промежуток, 2) плавная, 3) отдельные изолированные числа, 4) независимая.

4. Непрерывная случайная величина:
1) заполняет промежуток, 2) плавная, 3) отдельные изолированные числа, 4) независимая.

5. Дисперсия является мерой
1) математического ожидания, 2) среднего, 3) разброса. 4) вероятности.

6. Дисперсия дискретной случайной величины подсчитывается по формуле:
а) $D(x) = M(x^2) — M^2(x)$
б) $D(x) = M(x^2)$
в) $D(x) = M^2(x)$
г) $D(x) = M(x^2) + M^2(x)$
где $M(x)$ — математическое ожидание случайной величины x.

7. График функции плотности нормального распределения имеет форму:
1) параболы, 2) колокола, 3) вогнутой дуги, 4) гиперболы.

8. Для нормального закона распределения средних значений по отношению к крайним значениям:
1) одинаково, 2) меньше, 3) независимо, 4) больше.

9. Выберите формулу вероятности события:
1) $P(H)=m*n$, 2) $P(H)=m+n$, 3) $P(H)=m/n$, 4) $P(H)=m-n$.

10. Вероятность классическая отличается от статистической тем, что:
1) классическая вероятность вычисляется после опыта;
2) статистическая вероятность вычисляется до опыта;
3) статистическая вероятность вычисляется до опыта, а классическая вероятность вычисляется после опыта;
4) классическая вероятность вычисляется до опыта, а статистическая вероятность вычисляется после опыта.

11. Вероятность того, что произошло событие A и событие B, при условии, что события зависимы:
1) $\text{Đ}(A \cap B) = P(A) + P(B)$; 2) $\text{Đ}(A \cap B) = P(A)*P_A(B)$; 3) $\text{Đ}(A \cap B) = P(A)*P(B)$;
4) $\text{Đ}(A \cup B) = P(A) + P(B)$.

12. Вероятность того, что произошло событие A или событие B, при условии, что события совместны:
1) $\text{Đ}(A \cap B) = P(A) + P(B)$; 2) $\text{Đ}(A \cap B) = P(A)*P(B)$; 3) $\text{Đ}(A \cup B) = P(A)*P(B)$;
4) $\text{Đ}(A \cup B) = P(A) + P(B) - P(A \cap B)$.

13. Вероятность попадания в цель первым стрелком равна 0,9, а второго — 0,95. Какова вероятность, что цель будет поражена хотя бы одним стрелком:
а) 0,855, б) 0,995?

14. Какова вероятность вытащить из 36 карт одновременно двух дам:
а) 1/105, б) 2/36?

15. Из 190 деталей лишь 1 является бракованной. Сколько бракованных изделий можно ожидать в партии товара из 950 изделий:
а) 5/190, б) 5?

16. На группу из пятнадцати человек выделили 3 билета на концерт. Сколько возможно вариантов распределения билетов:
а) 2730, б) 15* 3=45?

17. В группе 32 человека. Необходимо избрать старосту, культурно-массовый сектор и профсоюзного лидера. Для подсчета способов образования руководящей тройки (одно лицо может занимать только один пост) необходимо подсчитать:
а) число размещений без повторений;
б) число перестановок;
в) число сочетаний с повторениями;
г) число размещений с повторениями.

18. Для подсчета того, сколько разных трехзначных чисел можно составить из цифр 3, 4, 5 при условии, что ни одна цифра не повторяются необходимо подсчитать:
а) число перестановок без повторений;
б) число сочетаний;
в) число перестановок с повторениями;
г) число повторений.

19. Вероятность попадания в мишень равна 0,4. Тогда вероятность не попасть в мишень:
а) 0,5; б) 0,6; в) 0,8; г) 0,2.

20.Событие C, означающее что произошло и событие A, и событие B, является:
а) объединением событий A и B;
б) пересечением событий A и B;
в) размещением событий A и B;
г) разность событий A и B.

Вопросы для подготовки к экзамену

1. Множество (Понятие множества. Подмножество. Пустое множество. Равные множества. Объединение двух множеств. Пересечение двух множеств. Разность двух множеств. Дополнение множества).
2. Комбинаторика (Размещения. Перестановки. Сочетания). Выборки без повторений и с повторениями.
3. Теория вероятностей Случайные события. Достоверные события. Невозможные события. Отношения между событиями (Совместные события. Несовместные события. Противоположные события. Независимые события). Операции над событиями (Объединение событий. Пересечение событий).
4. Определение вероятности. Условная вероятность. Классическое определение вероятности события. Статистическое определение вероятности события.
5. Теорема сложения вероятностей несовместных событий. Теорема сложения вероятностей совместных событий.
6. Теорема умножения вероятностей независимых событий. Теорема умножения вероятностей зависимых событий.
7. Формула полной вероятности. Формула Байеса. Формула Бернулли.
8. Понятие случайной величины. Виды случайных величин (Дискретная величина. Непрерывная величина). Распределение дискретной случайной величины. Функция распределения. Понятие математического ожидания (дискретные случайные величины). Свойства математического ожидания. Понятие дисперсии (дискретные случайные величины). Свойства дисперсии.
9. Функция и плотность распределения вероятностей (непрерывные случайные величины). Математическое ожидание. Дисперсия. Мода. Медиана.
10. Нормальное распределение. Биномиальный закон распределения. Равномерный закон распределения. Распределение Пуассона. Показательный закон распределения. Локальная и интегральная теорема Лапласа.
11. Предельные теоремы теории вероятностей.
12. Многомерные случайные величины. Ковариация и коэффициент корреляции.
13. Определение случайного процесса и его характеристики.
14. Основные понятия теории массового обслуживания.
15. Цепи Маркова.

ЛИТЕРАТУРА

Основная:

1. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.
2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.: «Высшая школа». — 2002.
3. Бочаров П.П., Печенкин А.В. Теория вероятностей и математическая статистика. — М.: Гардарика, 1998.
4. Кит Ю.В. Теория вероятностей для гуманитариев в примерах и задачах: Учебное пособие, 2003.
5. Вентцель Е.С. Теория вероятностей: Учебник. — М.: Академия, 2008. — 576 с.
6. Вентцель Е.С., Овчаров Л.А. Задачи и упражнения по теории вероятностей: Учеб. пособие. — М.: Академии, 2008. — 448 с.

Дополнительная:

1. Ковалев В.А., Калинина В.Н. Теория вероятностей и математическая статистика. — М.: ИНФРА-М, 1999.
2. Общий курс высшей математики для экономистов: Учебник под ред. В.И. Ермакова. — М.: ИНФА-М, 2000.
3. Красс М.С. Математика для экономических специальностей: Учебник. — М.: ИНФРА-М, 1999. — 464с.
4. Данко П.Е., Попов А.Г., Кожевникова Т.Я. Высшая математика в упражнениях и задачах. Часть 2. — М.: Мир и образование, 2005. — 416с.
5. Горелова Г.В., Кацко И.А. Теория вероятностей и математическая статистика в примерах и задачах с применением EXCEL / Учебное пособие для вузов. — Ростов н/Д: Феникс, 2002.

Часть 2. «МАТЕМАТИЧЕСКАЯ СТАТИСТИКА»

ВВЕДЕНИЕ

Цель изучения дисциплины — познакомить студентов со статистическими методами обработки результатов измерений, научить их интерпретировать, показать прикладное значение математических знаний

Студент должен иметь представление о важнейших математических понятиях, на основе которых возможны корректное применение математических методов в различных видах исследований.

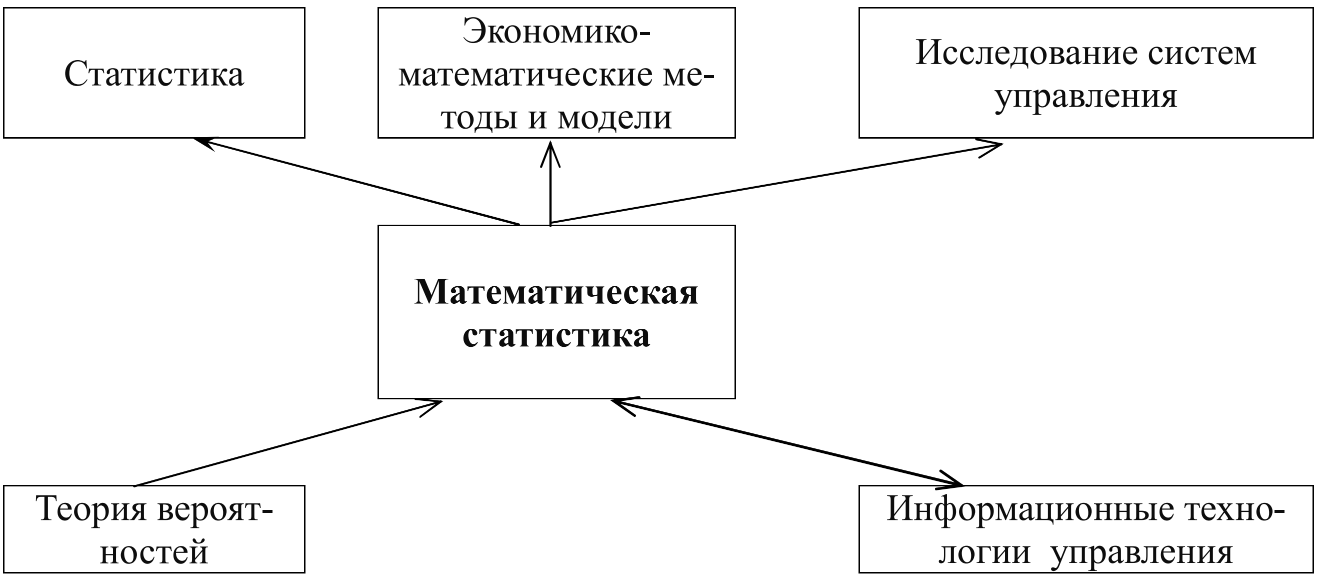
По окончании изучения курса студент должен знать:

- основные законы распределения;
- о нормальном распределении и его свойствах;
- различные формы закона больших чисел;
- основы статистического описания;
- основные понятия дисперсионного анализа.

По окончании изучения курса студент должен уметь:

- вычислять числовые характеристики выборки;
- коэффициенты корреляции;
- составлять уравнение регрессии.
- строить гистограмму и полигон частот;
- вычислять интервальные оценки, доверительные интервалы и области;
- проводить статистическую проверку гипотез.

Карта межпредметных связей



РАБОЧАЯ ПРОГРАММА

Тема 1. Вариационный ряд

Понятие о вариационном ряде. Частоты и частности. Виды вариации. Дискретные и интервальные вариационные ряды. Границы интервалов и величина интервала. Плотность распределения. Накопленные частоты. Графические методы изображения вариационного ряда: полигон, гистограмма, кумулята, огива. Виды шкал.

Тема 2. Числовые характеристики вариационного ряда

Средняя арифметическая и ее свойства. Квантили. Мода. Медиана. Показатели разброса признака: вариационный размах, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение, коэффициент вариации.

Тема 3. Основные законы распределения

Биномиальный закон распределения. Равномерный закон распределения. Распределение Пуассона. Показательный закон распределения. Нормальное распределение. Стандартное (нормированное) нормальное распределение. Вероятность попадания в заданный интервал нормально распределенной случайной величины. Вероятность заданного отклонения нормально распределенной случайной величины от своего математического ожидания. Правило трех сигм. Локальная и интегральная теоремы Лапласа.

Тема 4. Закон больших чисел

Понятие о законе больших чисел. Неравенства Маркова, Чебышева. Теоремы Чебышева, Бернулли, Пуассона. Понятие о «центральной предельной теореме» Ляпунова.

Тема 5. Выборочный метод и его значение

Понятие выборочного метода. Статистическое распределение выборки. Генеральная и выборочная совокупность. Способы отбора: собственно-случайный (повторный и бесповторный), механический, типический, серийный. Ошибки регистрации и репрезентативности (систематические и случайные). Статистические оценки параметров распределения (сущность теории оценивания). Точечные оценки: генеральной средней по выборочной средней, точечная оценка генеральной дисперсии. Предельная и средняя ошибка выборки для средней и доли. Интервальные оценки. Точность оценки. Доверительная вероятность. Доверительный интервал для оценки генеральной средней нормального распределения при известном и неизвестном среднем квадратическом отклонении. Доверительный интервал для оценки генеральной доли. Необходимая численность выборки.

Тема 6. Статистическая проверка гипотез

Законы распределения, применяемые в математической статистике: Стьюдента, Хи-квадрат, Фишера. Статистические гипотезы их виды. Нулевая и конкурирующая гипотезы. Ошибки I и II рода. Уровень значимости. Параметрические и непараметрические гипотезы. Выявление различий в уровне исследуемого признака. U-критерий Манна-Уитни. Оценка достоверности сдвига в значениях исследуемого признака. G-критерий знаков. Критерий χ^2 Фридмана, T-критерий Вилксона. Выявление различий в распределении признака. χ^2-критерий Пирсона. λ-критерий Колмогорова-Смирнова. Многофункциональные статистические критерии. Критерий φ^* — угловое преобразование Фишера.

Тема 7. Корреляционный анализ

Корреляционная связь и ее статистическое изучение. Линейная парная регрессия. Коэффициент корреляции. Линейный и ранговый коэффициенты корреляции. Проверка гипотезы о значимости выборочного коэффициента корреляции. Построение доверительного интервала для коэффициента корреляции. Нелинейная регрессия. Параболическая и гиперболическая зависимости между зависимыми случайными величинами. Множественная корреляция.

Тема 8. Дисперсионный анализ

Понятие дисперсионного анализа. Подготовка данных к дисперсионному анализу. Однофакторный дисперсионный анализ для несвязанных и связанных выборок. Двухфакторный дисперсионный анализ для несвязанных и связанных выборок.

Тема 9. Применение компьютерных программ при статистической обработке данных

Возможности электронных таблиц EXCEL. Возможности прикладного пакета STATISTICA. Компьютерный практикум.

КРАТКИЙ КУРС ЛЕКЦИЙ

Введение

Статистика — наука, изучающая методы сбора, обработки фактов и данных в области человеческой деятельности и природных явлений.

В нашем курсе, который можно считать введением в курс «Статистика», речь будет идти о так называемой прикладной статистике, о сущности специальных методов сбора, обработки и анализа информации и, кроме того, о практических приемах выполнения связанных с этим расчетов с применением информационных технологий.

Основная цель математической статистики — получение и обработка данных для статистически значимой поддержки процесса принятия решений, например, при решении задач планирования, правления, прогнозирования.

Аналитические методы математической статистики позволяют на основании выборочных наблюдений сделать статистически значимые выводы о наличии закономерностей для всей совокупности.

ШКАЛЫ ИЗМЕРЕНИЯ

Измерение — приписывание символов объектам или событиям в соответствии с определенными правилами. Существует четыре типа шкал измерений.

Таблица 1

Типы шкал

Шкала	Характеристика	Пример	Допустимый метод обработки
1	2	3	4
Шкала **наименований**	Объекты классифицированы по названию, а классы обозначены номерами. То, что один номер больше другого не говорит о разнице в свойствах, за исключением того, что они отличаются	Раса, цвет глаз, номера на футболках, номера на машинах	Метод χ^2, биномиальный критерий m, угловое преобразование Фишера φ^*
Шкала **порядка**	Объекты классифицированы по принципу «больше-меньше», т.е. соответствующие значения, присваиваемые предметам, отражают количество, свойства, однако равные разности не означают равные разности в количестве, свойствах	Награды за заслуги, военные ранги, твердость минералов	U-критерий Манна-Уитни, χ^2-критерий Пирсона, λ-критерий Колмогорова-Смирнова, угловое преобразование Фишера φ^* и др.

1	2	3	4
Шкала **интервалов**	Классифицирует объекты по принципу «больше на определенное количество единиц, меньше на определенное количество единиц». Существует единица измерения. Равные разности приписываемых чисел означают равные разности в свойствах. Нуль не означает отсутствие свойства.	Температура по C^0, календарное время	U-критерий Манна-Уитни, χ^2-критерий Пирсона, λ-критерий Колмогорова-Смирнова, угловое преобразование Фишера φ^* и др.
Шкала **отношений**	Классифицирует объекты пропорционально степени выраженности измеряемого свойства. Ноль указывает на отсутствие свойства.	Вес, рост	Любые методы математической статистики

ВАРИАЦИОННЫЕ РЯДЫ

Основные определения

Первый шаг на пути статистической обработки данных заключается в группировке полученных результатов и их представлении в виде сводных аблиц. При распределении членов совокупности в ряд преследуются определенные цели. Одна из них — раскрытие закономерности варьирования изучаемого признака. Поэтому к рядам распределения предъявляются следующие требования:

1) они должны быть легко обозримы;

2) хорошо иллюстрировать закономерности варьирования.

Результаты наблюдений, в общем случае ряд чисел, расположенных в беспорядке, необходимо упорядочить (проранжировать). Ранжировать можно как по возрастанию, так и по убыванию признака. После операции ранжирования опытные данные можно сгруппировать так, чтобы в каждой группе признак принимал одно и то же значение, которое называется варианта (x_i). Число элементов в каждой группе называется частотой варианты (n_i). Размахом выборки называется разность наибольшей и наименьшей вариант.

Такой упорядоченный ряд распределения, в котором указана повторяемость вариант, принадлежащих к данной совокупности, называется вариационным рядом. Таким образом, числа, которые характеризуют встречаемость отдельных вариант в исходной совокупности, называют весами или частотами.

Признаки принято обозначать прописными буквами латинского алфавита — X, Y, Z, ..., а их численные значения, т.е. варианты — соответствующими строчными буквами — x_1, x_2, x_3,...; y_1, y_2, y_3,... и т.д. Частоты обозначаются латинской строчной буквой p. Общее число вариант, входящих в состав данной совокупности, называется объемом выборки и обозначается латинскими буквами n или N. Общая сумма частот равна объему совокупности n: $\Sigma p = p_1 + p_2 +$

$+ p_3 + ... = n$. Частоты — это абсолютные веса отдельных вариант. Они могут быть выражены и в относительных значениях варьирующего признака, т.е. в долях единицы или же в процентах по отношению к общей численности вариант данной совокупности. В таких случаях веса называются относительными частотами или частостями. Сумма частостей, выраженная в долях единицы, равняется единице ($\Sigma p/n = 1$), а сумма частостей, выраженная в процентах, равна 100 ($100 \cdot \Sigma p/n = 100$).

Построение вариационных рядов

Известно два вида вариационных рядов: безинтервальные и интервальные. Но, когда совокупность достаточно большая или состоит из непрерывных значений, безинтервальный вариационный ряд плохо отражает закономерности варьирования признаков. В таких случаях целесообразнее построить интервальный вариационный ряд. Техника построения такого ряда заключается в следующем. Вся вариация признака от минимального до максимального разбивается на равные интервалы или промежутки, называемые также классами. Затем все варианты совокупности распределяются по этим классам. В результате получается интервальный вариационный ряд, в котором частоты (p) относятся уже не к отдельным конкретным вариантам, как в безинтервальном вариационным ряду, а к установленным классовым интервалам, т.е. оказываются частотами не вариант, а классов. Для этого определяем число классов по формуле $K = 1+3{,}32 \cdot lgn$ (по Стерджесу) или по Таблице 2. Затем определяем $R = x_{max} — x_{min}$.

Находим ширину классового интервала h, по формуле h= R/(K-1).

Находим нижнюю границу первого класса по формуле $x_н = x_{min} — 0{,}5 \cdot h$.

Таблица 2

Объем выборки n	Число классов K
6-11	4
12-22	5
23-46	6
47-93	7
94-187	8
188-377	9
378-755	10
756-1515	11
1516-3050	12

Начальные и конечные значения всех последующих классов можно вычислить путем последовательного прибавления величины классового интервала к соответствующим значениям первого класса.

Пример построения интервального вариационного ряда:

Пусть измерен некоторый экономический показатель в 30 регионах:
23 29 35 7 11 18 23 30 36 18 11 8 13 20 25 27 14 30 20 20 24 19 21 26 22 16 26 25 33 27.

Расставим экспериментальные данные в возрастающем порядке:
7 8 11 11 13 14 16 18 18 19 20 20 20 21 22 23 23 24 25 25 26 26 27 27 29 30 30 33 35 36.

По таблице 2 определяем число классов.

Для n=30 число классов K=6. Найдем минимальное и максимальное значения вариант: x_{min}=7, x_{max}=36. Определим вариационный размах

$R= x_{max}-x_{min}$=36-7=29.

Определим величину классового интервала: $\Delta = \frac{R}{K-1} = \frac{29}{51} = 5{,}8$.

$$X_{H1} = x_{\min} - \frac{1}{2}\Delta = 7 - 2{,}9 = 4{,}1;\; X_{B1} = x_{\min} + \frac{1}{2}\Delta = 7 + 2{,}9 = 9{,}9$$

Обобщим полученные данные в таблице:

Таблица 3

Номера классов	Классовые интервалы	Серединные значения классов	Частоты	Накопленные частоты
1	4,1-9,9	7	2	2
2	9,9-15,7	12,8	4	6
3	15,7-21,5	18,6	8	14
4	21,5-27,3	24.4	10	24
5	27,3-33,1	30,2	4	28
6	33,1-38,9	36	2	30

Графическое представление вариационных рядов

Вариационные ряды изображают графически с помощью полигона, гистограммы и кумуляты.

График, называемый гистограммой, получается, если в прямоугольной системе координат отложить по оси абсцисс границы классов, а по оси ординат — их частоты. Гистограмма на рис. 1 построена на основании данных таблицы 3.

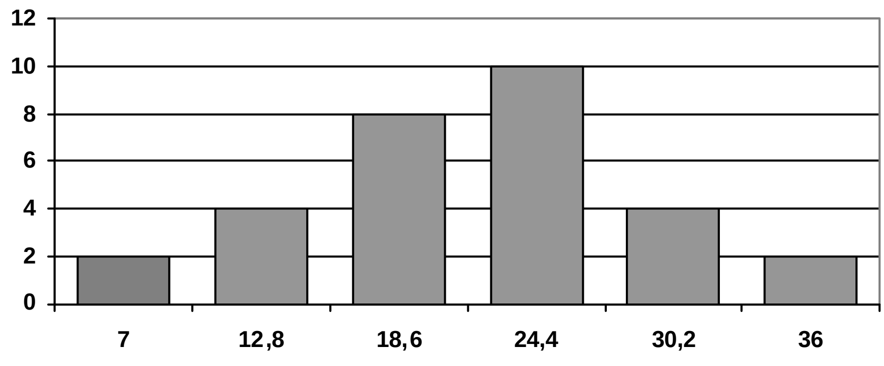


Рис. 1. Гистограмма распределения.

Если серединные точки вершин прямоугольников гистограммы соединить между собой, получится график дискретного варьирования, называемый полигоном распределения. Для дискретного (безинтервального) вариационного ряда полигон частот — это ломаная, отрезки которой соединяют точки (x_1, n_1), (x_2, n_2),…, (x_k, n_k).

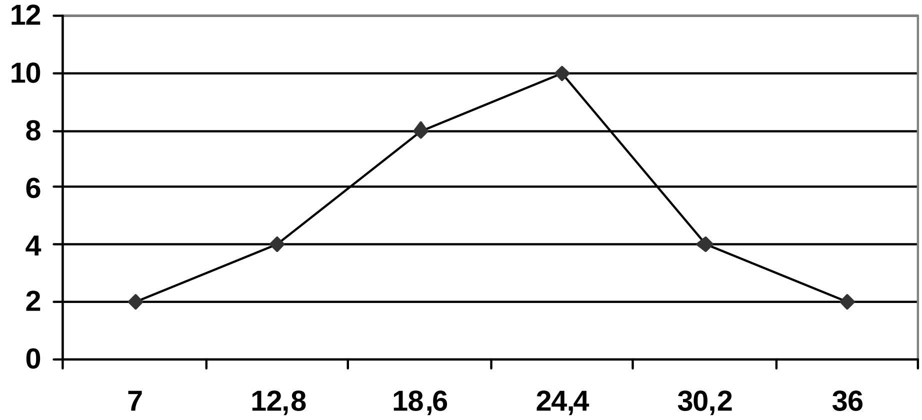


Рис.2. Полигон распределения.

Кумулята — ломаная, соединяющая точки, абсциссы которых — значения вариант (серединные значения ряда), а ординаты — накопленные частоты этих значений.

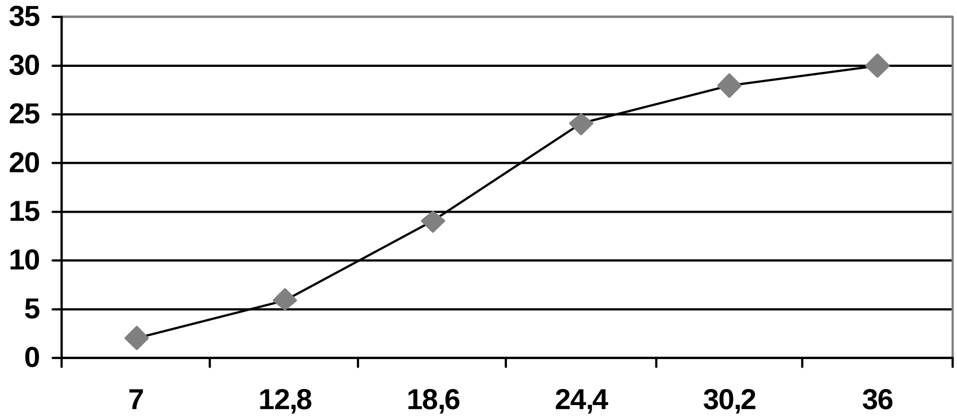


Рис. 3. Кумулята распределения.

Числовые характеристики вариационных рядов

Меры центральной тенденции

а) Среднее арифметическое вариационного ряда находится по формуле

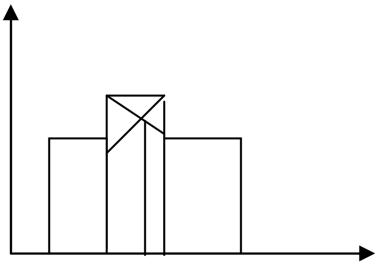
$$x_{ср}= (x_1+x_2+x_3+ \ldots+x_n)/n.$$

б) Мода дискретного вариационного ряда — это наиболее часто встречающееся значение ряда (M_o).

в) Медиана дискретного вариационного ряда — средний член упорядоченного дискретного вариационного ряда (M_e). Если вариантов четное количество, то медиана — среднее из двух центральных значений.

Для интервального вариационного ряда медиана может быть найдена с помощью кумуляты, как абсцисса точки, ордината которой — $n/2$. Мода интервального вариационного ряда может быть приближенно найдена с помощью гистограммы (см. рис. 4 ниже).

При наличии крайних значений использование медианы и моды в качестве меры центральной тенденции предпочтительнее, так как среднее арифметическое наиболее чувствительно к крайним значениям.



M_o

Рис.4. Нахождение медианы по гистограмме.

Меры изменчивости

а) Вариационный размах $R= x_{max}\text{-}x_{min.}$.

б) Дисперсия распределения находится по формуле:

$$D = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_{ñð})^2 .$$

в**)** Стандартное (среднеквадратическое) отклонение: $S = \sqrt{D}$.

г) коэффициент вариации:

$$v = \frac{S}{x_{ср}}100\%$$

Характеристики меры скошенности и островершинности

Коэффициент асимметрии подсчитывается по формуле:

$$A = \frac{1}{nS^3}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^3$$

Коэффициент эксцесса: $E = \frac{1}{nS^4}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^4 - 3$.

Если построен интервальный вариационный ряд, то для подсчета характеристик ряда можно использовать средние значения классов, умножая их на частоту класса.

Задания для самостоятельной работы:

1. Дан ряд распределения хозяйств по количеству рабочих на 100 га с/х угодий (n=60). Построить интервальный вариационный ряд и определить его характеристики (меры центральной тенденции, меры изменчивости, скошенности и островершинности).

12, 6, 8, 6, 10, 11, 7, 10, 12, 8, 7, 7, 6, 7, 8, 6, 11, 9, 11, 9, 10, 11, 8, 10, 7, 8, 8, 8, 11, 9, 8, 7, 5, 9, 7, 7 ,14, 11, 9 ,8 ,7, 5, 5, 10, 7, 7, 5, 8, 10, 10, 15, 10, 10, 13, 12, 11, 15, 6, 13, 15.

2. По списку на предприятии числится 100 рабочих, которые имеют следующие разряды:

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

Составить ряд распределения рабочих по разрядам. Найти накопленные частоты и частости. Вариационный ряд изобразить графически. Определить характеристики ряда.

ВЫБОРОЧНЫЙ МЕТОД

Этот метод состоит в том, что по некоторой части генеральной совокупности (выборки) выводится суждение о свойствах всей генеральной совокупности.

Существуют следующие типы выборок:

1. Собственно случайная: а) повторная, б) бесповторная.

2. Типическая — генеральная совокупность разбивается (предварительно) на группы типических элементов, затем выборка осуществляется в каждой из них.

3. Механическая — отбор через определенный интервал.

4. Серийная — отбор производится не по одному элементу, а сериями для проведения сплошного обследования.

5. Комбинированная — используются перечисленные типы выборки в сочетании.

Точечные оценки

В качестве точечных оценок используют чаще всего математическое ожидание, дисперсию и т.д.

Точечная оценка генеральной совокупности — это число, определяемое по выборке, для ее обозначения используется символ Θ (тетта).

$\widehat{\Theta}_n$ — точечная оценка выборки.

Качество оценки Θ устанавливается по 3 свойствам:

1. *Состоятельность* — предел вероятности того, что при увеличении объема выборки n, выборочная характеристика стремится к соответствующей характеристике генеральной совокупности.

2. *Несмещенность*. Оценка $\widehat{\Theta}_n$ называется несмещенной, если для любого фиксированного числа наблюдений выполняется равенство: $M(\widehat{\Theta})=\Theta$, где $\widehat{\Theta}_n$ — оценка генеральной характеристики Θ.

3. Эффективность. Несмещенная оценка $\widehat{\Theta}_n$ генеральной характеристики Θ будет называться *несмещенной эффективной*, если среди всех несмещенных оценок она будет обладать наименьшей дисперсией.

Ошибки выборки

Чем меньше разность $\left|\widehat{\Theta}_n - \theta\right|$, тем точнее оценка.

Выборка называется репрезентативной, если она максимально точно отображает особенности генеральной совокупности.

Ошибки выборки возникают в результате:

- малого объема выборки;
- большого разброса признака;
- неправильной организации метода обследования (например, субъективность исследователя).

Ошибка среднеарифметической

Пусть из одной и той же генеральной совокупности, распределенной по нормальному закону, отобрано какое-то количество η независимых выборок по n вариант каждой. Среднее арифметическое этих выборок $\overline{X_1}, \overline{X_2}, ...\overline{X_n}$ будет варьироваться рядом с генеральной средней $\overline{X}$.

Основным мерилом вариации является среднеквадратичное отклонение. В математической статистике доказано, что выборочные средние варьируются относительно генерального параметра в n-раз меньше, чем отдельные варианты генеральной совокупности. Тогда ошибка средней арифметической m_x при случайном повторном отборе рассчитывается по формуле:

$$m_x = \sqrt{\frac{s^2}{n}}.$$

При больших выборках (n>300), $\sqrt{n}$, $\sqrt{n-1}$ не различаются.

Интервальной называют оценку, которая определяется двумя числами — границами интервала. Она позволяет ответить на вопрос: внутри какого интервала и с какой вероятностью находится неизвестное значение оцениваемого параметра генеральной совокупности.

Предельную ошибку выборки можно подсчитать по формуле: $\Delta = tm_x$, где t — нормированное отклонение — «коэффициент доверия», зависящий от вероятности, с которой гарантируется предельная ошибка выборки.

На основании теоремы П.Л. Чебышева можно утверждать, что $P\left[\left|X_{ñð.} - \overline{X}\right| \leq \Delta\right] = \Phi(t)$. Наиболее часто встречающиеся значения функции $\Phi(t)$ представлены в таблице 4. Предельная ошибка выборки может быть определена с определенной долей вероятности. Значения функции Лапласа определяют эту вероятность. Ранее рассчитанные значения ошибки соответствуют вероятности при t=1.

Таблица 4.

t	1,000	1,960	2,000	2,580	3,000
Φ	0,683	0,950	0,954	0,990	0,997

Тогда предельная ошибка среднего для повторной выборки равна:

$$\Delta = tm_x = t\sqrt{\frac{s^2}{n}}$$

Предельная ошибка среднего для бесповторной выборки:

$$\Delta\bar{x} = t\sqrt{\frac{s^2}{n}\left(1 - \frac{n}{N}\right)}$$

Тогда $\bar{x} - \Delta\bar{x} \leq \overline{x_{ср.}} \leq \bar{x} + \Delta\bar{x}$

Пример 1.

Методом случайной повторной выборки было взято для проверки на вес 200 штук деталей. В результате был установлен средний вес детали $\bar{x}$=30 гр., при S=4 гр. Рассчитать среднюю ошибку выборки с вероятностью 0,954.

Решение.

С вероятностью Φ=0,954 требуется определить пределы, в которых находится средний вес детали генеральной совокупности.

Предельная ошибка выборки для средней, при повторном отборе, рассчитывается по формуле:

$$\bar{\Delta}x = t\sqrt{\frac{s^2}{n}}\ .$$

Если для нашей задачи, то t=2, $\bar{\Delta}x = 2*0.28 = 0.56$. Тогда доверительный интервал для средней будет рассчитываться по формуле:

$$\bar{x} - \bar{\Delta}x \le \overline{x_{ср.}} \le \bar{x} + \bar{\Delta}x$$

$$30\text{-}0{,}56 \le \overline{x_{ср.}} \le 30 + 0{,}56\,,$$

$$29{,}44 \le \overline{x_{ср.}} \le 30{,}56$$

Расчет численности выборки по формуле бесповторного и повторного случайного отбора

При проведении статистических исследований следует выяснить необходимую численность выборки из генеральной совокупности, чтобы обеспечить ее репрезентативность. Для определения n необходимо задать надежность, доверительную вероятность $\Phi(t)$ оценки и точность (предельную ошибку выборки Δ). Для расчета необходимой численности выборки следует использовать следующие формулы:

при повторном отборе $n = \dfrac{t^2S^2}{\Delta^2}$

для бесповторного отбора: $n = \dfrac{t^2S^2N}{\Delta^2 N + t^2S^2}$

Параметр S не известен до исследования ни для выборки, ни для генеральной совокупности. Поэтому на практике пользуются либо данными предыдущего исследования, либо пилотажного исследования, либо предполагая распределение близким к нормальному:

$$S = \frac{x_{\max} - x_{\min}}{6}\,,$$

либо пробно исследуют некоторое количество данных и подсчитывают дисперсию и среднеквадратическое отклонение этих значений.

Пример 2.

В районе проживает 2000 семей. В порядке случайной бесповторной выборки предполагается определить средний размер семьи при условии, что ошибка выборочной средней не должна превышать 0,8 человека, а среднеквадратическое отклонение равно 2 с вероятностью 0,954.

Решение:

N=2000, Δ=0,8, S=2, p=0,954, а значит, t=2, поэтому $n = \dfrac{2^2 2^2 2000}{0{,}8^2 2000 + 2^2 2^2} = 24$

Задания для самостоятельной работы

1. В порядке случайной повторной выборки было обследовано 900 деревьев, по этим данным установлен средний диаметр одного дерева 235 мм и среднее квадратическое отклонение, равное 27 мм. С вероятностью 0,683 определите границы, в которых будет находиться средний диаметр деревьев в генеральной совокупности.

2. Для определения среднего размера вклада вкладчиков сбербанка, где число вкладчиков равно 5000, необходимо провести выборку лицевых счетов методом механического отбора. Предварительно установлено, что среднее квадратическое отклонение размеров вкладов составляет 120 руб. Определите необходимую численность выборки при условии, что с вероятностью 0,954 ошибка выборки не превысит 10 руб.

3. Для установления среднего возраста 50 тыс. читателей библиотеки необходимо провести выборку из читательских карточек методом механического отбора. Предварительно установлено, что среднее квадратическое отклонение возраста читателей равно 10 годам. Определите необходимую численность выборки при условии, что с вероятностью 0,997 ошибка выборки не превысит двух лет.

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Кривая нормального распределения (рис. 5а) аналитически описывается уравнением:

$$u = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-0,5\cdot\left(\frac{x-M}{\sigma}\right)^2} \qquad (1)$$

где u — высота кривой над любым заданным значением x;

σ — стандартное отклонение;

x — значение варианты;

M — математическое ожидание, или среднее арифметическое значение;

π — число пи, равное 3,142;

e — основание натурального логарифма, равное 2,718.

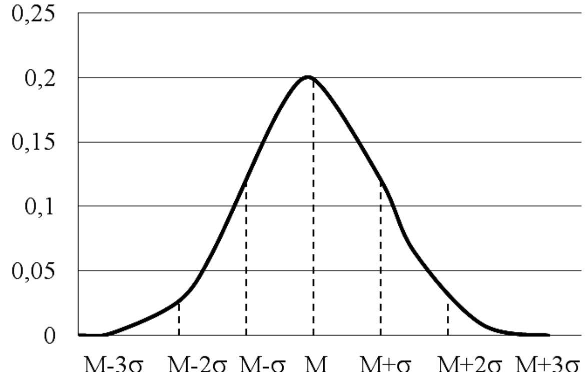


Рис. 5а. Кривая нормального распределения (M=7; σ=2)

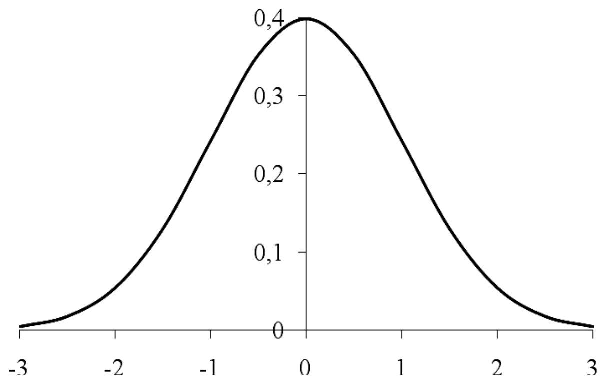


Рис. 5б. Кривая нормального распределения (M=0; σ=1)

Из этой формулы видно, что для различных значений σ и M может существовать бесконечное множество кривых нормального распределения. Выражение $(x\text{-}M)/\sigma$, входящее в состав формулы, называется нормированным отклонением и обозначается буквой t. Нормированное отклонение показывает на сколько сигм (σ) (т.е. единиц меры, которой служит среднее квадратическое отклонение) та или иная варианта совокупности отклоняется от среднего уровня варьирующего признака.

Введем обозначение $t = (x\text{-}M)/\sigma$ в формулу (1). Тогда она примет вид:

$$u = \frac{1}{\sigma\sqrt{2\pi}} \cdot \mathring{a}^{-0,5t^2} \qquad (2)$$

Из всего множества кривых нормального распределения выберем кривую, для которой M=0 и σ=1. Заменив u на $f(t)$ получим формулу:

$$f(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-0,5t^2} \qquad (3)$$

Здесь t — случайная величина, для которой математическое ожидание равно нулю, а среднее квадратическое отклонение — единице. График уравнения (3) для M=0 и σ=1 является также кривой нормального распределения (см. рис. 5б выше) и называется единичной нормальной кривой. Единичная нормальная кривая — особая кривая, так как ее выбрали как стандарт нормального распределения.

Площадь под кривой (рис. 5б), описываемой уравнением (8), равна единице. Если ее принять за 100%, то для значений t в пределах:

от -1 до +1 расположено 68,3% всей площади;

от -2 до +2 расположено 95,4% всей площади;

от -3 до +3 расположено 99,7% всей площади.

Следовательно, вероятность p любой варианты нормального распределения находится в пределах:

от -1 до +1 равна 0,683;

от -2 до +2 равна 0,954;

от -3 до +3 равна 0,997.

Эти выводы можно распространить и на любое нормальное распределение с математическим ожиданием M и средним квадратическим отклонением σ. Для любого нормального распределения (рис. 5):

1) 68,3% площади под кривой лежит в пределах от M-σ до M+σ,

2) 95,4% площади под кривой лежит в пределах от M-2σ до M+2σ,

3) 99,7% площади под кривой лежит в пределах от M-3σ до M+3σ.

Это означает, что случайная величина x находится в диапазоне:

$M-\sigma < x < M+\sigma$ с вероятностью p=0,683;

$M-2\sigma < x < M+2\sigma$ с вероятностью p=0,954;

$M-3\sigma < x < M+3\sigma$ с вероятностью p=0,997.

Проверка нормальности распределения признака по показателям асимметрии и эксцесса

Оценку на нормальность распределения признаков можно произвести и по величине показателей асимметрии и эксцесса. Показатели асимметрии и эксцесса вычисляются по известным формулам:

$$As_{\text{ýìï}} = \frac{\sum\left(x_i - \bar{x}\right)^3}{n \cdot \sigma^3}, \qquad Ex_{\text{ýìï}} = \frac{\sum\left(x_i - \bar{x}\right)^4}{n \cdot \sigma^4} - 3$$

Как и другие выборочные показатели, асимметрия и эксцесс являются величинами случайными. Поэтому необходима статистическая оценка достоверности их выборочных показателей.

Показатели асимметрии и эксцесса свидетельствуют о достоверном отличии эмпирических распределений от нормального в том случае, если они превышают критические значения. Критические значения можно вычислить по формулам, предложенным Е.И. Пустыльником:

$$As_{\text{êð}} = 3 \cdot \sqrt{\frac{6 \cdot (n-1)}{(n+1) \cdot (n+3)}}, \qquad Ex_{\text{êð}} = 5 \cdot \sqrt{\frac{24 \cdot n \cdot (n-2) \cdot (n-3)}{(n+1)^2 \cdot (n+3) \cdot (n+5)}},$$

где n — количество наблюдений.

Если $As_{эмп} > As_{кр}$ и $Ex_{эмп} > Ex_{кр}$, то эмпирическое распределение отличается от нормального.

СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ

Статистической гипотезой называется всякое высказывание о генеральной совокупности, проверяемое по выборке. Статистические гипотезы делятся на:

1) Параметрические — это гипотезы, сформулированные относительно параметров (среднего значения, дисперсии и т.д.) распределения известного вида;

2) Непараметрические — это гипотезы, сформулированные относительно вида распределения (например, определение по выборке степени нормальности генеральной совокупности).

Процесс использования выборки для проверки гипотезы называется статистическим доказательством. Основную выдвигаемую гипотезу называют нулевой H_0. Наряду с нулевой гипотезой рассматривают ей альтернативную H_1.

Параметрические методы — это те методы, использование которых основано на знании определенного закона распределения. Непараметрические могут применяться при любом законе распределения.

Статистические критерии

Статистический критерий — это решающее правило, обеспечивающее принятие истинной и отклонение ложной гипотезы с высокой точностью.

Статистические критерии означают метод расчета определенного числа, по значению которого можно судить о подтверждении или отклонении той или иной гипотезы.

Когда мы говорим, что достоверность различий определялось по t критерию Стьюдента, то это означает, что подсчитано t-фактическое определенное число, рассчитанное по этому методу.

В большинстве случаев критические значения зависят от количества наблюдений и от количества степеней свободы, которые обозначаются как V или η.

$V=n^1 — 1$ (n^1 — количество интервалов).

При проведении исследований уровнем статистической значимости называется число, равное вероятности, с которой мы отклоняем H_0. Например, когда мы указываем, что различие достоверно на 1% уровне значимости, т.е. $p \leq 0{,}01$ то имеем в виду, что вероятность того, что различия недостоверны равна 0,01.

Выявление различий в уровне исследуемого признака

Критерий U Манна-Уитни

Критерий Манна-Уитни применяется для несвязанных выборок и является непараметрическим критерием. Этот критерий предназначен для оценки различий между двумя выборками по уровню какого-либо признака, количественно измеренного. Он позволяет выявлять различия и между малыми выборками. Для оценки по этому критерию все варианты признака сравниваемых совокупностей ранжируют в один общий ряд и находят их ранги. Ранги находят последующим правилам: меньшему значению начисляется меньший ранг. Наименьшему значению начисляется ранг 1. Если несколько значений равны, то им начисляется ранг, представляющий собой среднее из тех рангов, которые они получили бы, если бы они не были равны.

Затем ранги суммируют отдельно по каждой выборке. Если сравниваемые выборки совершенно не отличаются одна от другой, то и суммы их рангов должны быть равны между собой. В противном случае такое равенство наблюдаться не будет. И чем значительнее расхождение между выборками, тем больше разница между суммами их рангов. А так как указанные различия могут быть случайными, они оцениваются с помощью критерия, который вычисляется по следующей формуле:

$$U_ф = (n_1 \cdot n_2) + n_x \cdot (n_x+1)/2 - T_x ,$$

где n_1 и n_2 — количество испытуемых в выборках 1 и 2;

T_x — большая из ранговых сумм;

n_x — объем выборки с большей суммой рангов.

Вычисленное фактическое значение ($U_ф$) сравнивается с табличным стандартным (U_{st}) значением. При этом может быть:

1) $U_ф > U_{st}$ — различия отсутствуют;

2) $U_ф \leq U_{st0,05}$ — различия достоверны на уровне p=0,05;

3) $U_ф \leq U_{st0,01}$ — различия достоверны на уровне p=0,01.

Пример 1. У двух групп студентов Санкт-Петербургского университета (у физиков и психологов) был измерен уровень вербального (ВИ) и невербального (НВИ) интеллекта с помощью методики Д. Векслера, результаты измерения ВИ которых представлены в таблицах 5 и 6 (см. ниже).

Таблица 5

Физики

NN	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Показат. ВИ	132	134	124	132	135	132	131	132	121	127	136	129	136	136

Таблица 6

Психологи

NN	1	2	3	4	5	6	7	8	9	10	11	12
Показат. ВИ	126	127	132	120	119	126	120	123	120	116	123	115

Решение.

Гипотеза H_0 состоит в том, что студенты-физики не превосходят студентов-психологов по уровню вербального интеллекта.

Занесем все показатели в таблицу (табл. 7) в возрастающем порядке, отмечая принадлежность каждого показателя к той или иной группе. Снизу показателей проставим ранги. Если несколько показателей имеют одинаковые значения, то они все должны иметь одинаковые ранги. Для этого складываем все порядковые номера испытуемых с одинаковыми показателями и делим на их число. Мы получим усредненные ранги для одинаковых показателей. После этого определяем сумму рангов для первой и для второй групп.

Таблица 7

NN	1	2	3	4	5	6	7	8	9	10	11	12	13
Пок.	115	116	119	120	120	120	121	123	123	124	126	126	127
Ранг	1	2	3	5	5	5	7	8,5	8,5	10	11,5	11,5	13,5
Код							*ф*			*ф*			

NN	14	15	16	17	18	19	20	21	22	23	24	25	26
Пок.	127	129	131	132	132	132	132	132	134	135	136	136	136
Ранг	13.5	15	16	19	19	19	19	19	22	23	25	25	25
Код	*ф*	*ф*	*ф*	*ф*	*ф*	*ф*	*ф*		*ф*	*ф*	*ф*	*Ф*	*ф*

Для студентов-психологов $S_{r1} = 93,5$.

Для студентов-физиков $S_{r2} = 257,5$.

Определим значение критерия $U_ф$

$$U_ф = 12 \cdot 14 + \frac{14 \cdot (14+1)}{2} - 257,5 = 15,5$$

По таблице 9 (см. Приложение 2) для $n_1 = 14$ и $n_2 = 12$, $U_{st} = 51$ для p=0,05 и $U_{st} = 38$ для p=0,01.

Здесь $U_ф < U_{st}$, то есть выборки отличаются, студенты-физики по вербальному интеллекту превосходят студентов-психологов с достоверностью p=0,01.

Оценка достоверности сдвига в значениях исследуемого признака

Т-критерий Вилкоксона

Данный критерий применяется для сопоставления показателей измеренных в двух разных условиях на одной и той же выборке. Он позволяет установить не только направленность изменений, но и их выраженность, т.е. является ли сдвиг показателей в каком-то одном направлении более интенсивным, чем в другом.

Гипотезы

H_0: Интенсивность сдвигов в типичном направлении не превосходит интенсивности сдвигов в нетипичном направлении.

H_1: Интенсивность сдвигов в типичном направлении превышает интенсивность сдвигов в нетипичном направлении.

Данный метод применяется при объеме выборки $5<n<50$. Если сдвига не произошло, то такие данные исключаются из рассмотрения.

Данный метод состоит в следующем:

- определяется разность между значениями во втором и первом замерах;
- определяется типичность сдвига (в положительную или отрицательную сторону);
- находят ранги абсолютных величин этих сдвигов;
- подсчитывается сумма рангов, соответствующих сдвигам в нетипичном направлении;
- найденная сумма рангов сравнивается с критической, если $T_{эмп} \leq T_{крит}$., то сдвиг в «типичную» сторону по интенсивности достоверно преобладает.

Пример 2.

В выборке государственных служащих г. Казани измерялся уровень эмоциональной напряженности до и после проведения учебных занятий в Институте государственной службы при Президенте РТ. Подтвердилась ли гипотеза экспериментаторов о том, что проведение учебных занятий способствует повышению уровня эмоциональной напряженности? Данные представлены в таблице 8.

Решение:

Как видим из таблицы 8 (см. ниже) типичным является сдвиг в отрицательную сторону, нетипичным — в положительную. Сумма рангов нетипичных сдвигов равна:

T=1+2,5+7=10,5.

Таблица 8

Код имени испытуемого		Уровень эмоциональной напряженности		Разность	Абсолютное значение разности	Ранговый номер разности
		до учебных занятий	после учебных занятий			
1	Г.	64	25	-39	39	11
2	Кос.	77	50	-27	27	8
3	Крив.	74	77	3	3	1

4	Кур.	95	76	-19	19	6
5	Л.	105	67	-38	38	9,5
6	М.	83	75	-8	8	4
7	Р.	73	77	4	4	2,5
8	С.	75	71	-4	4	2,5
9	Т.	101	63	-38	38	9,5
10	Х.	97	122	25	25	7
11	Ю.	78	60	-18	18	5

По таблице критических значений (табл. 4 Приложения 2) для n=11 при уровне значимости p=0,05 $T_{крит.}$=13, а значит, $T_{эмп} \leq T_{крит.}$, и сдвиг в «типичную» сторону по интенсивности достоверно преобладает. Поэтому интенсивность отрицательного сдвига показателя эмоциональной напряженности превышает интенсивность положительного сдвига (p=0,05). Гипотеза экспериментаторов о том, что проведение учебных занятий способствует повышению уровня эмоциональной напряженности не подтвердилась. Напротив, эмоциональная напряженность после проведения занятий снижается.

Выявление различий в распределении признака

λ-критерий Колмогорова-Смирнова

Критерий λ предназначен для сопоставления двух распределений: эмпирического с теоретическим, например, равномерным или нормальным, одного эмпирического распределения с другим эмпирическим распределением.

Критерий позволяет найти точку, в которой сумма накопленных расхождений между двумя распределениями является наибольшей, и оценить достоверность этого расхождения. Данный критерий обычно применяется на достаточно больших выборках.

Гипотезы:

H_0: Различия между двумя распределениями недостоверны.

H_1: Различия между двумя распределениями достоверны.

В данном методе сопоставляются сначала частоты по первому разряду, потом по сумме первого, второго разрядов и т.д., т.е. сопоставляются накопленные частоты: эмпирические и теоретические. Если различия между распределениями существенны, то в какой-то момент разность накопленных частот достигнет критического значения, и тогда различия признаются достоверными. В формулу критерия λ входит эта разность. Чем больше эмпирическое значение λ, тем более существенны различия.

Пример 3.

В выборке студентов технических вузов в возрасте от 19 до 22 лет, проводился тест Люшера в 8-цветном варианте. Установлено, что желтый цвет предпочитается испытуемыми чаще, чем отвергается (табл. 9). Можно ли утверждать, что распределение желтого цвета по восьми позициям отличается от равномерного?

Таблица 9

Разряды	Позиции желтого цвета								Сумма
	1	2	3	4	5	6	7	8	
Эмп. частоты	24	25	13	8	15	10	9	8	102

Решение: Заполним по приведенному выше алгоритму таблицу расчета критерия, определив по последнему столбцу наибольшую абсолютную величину разности между накопленными эмпирическими и теоретическим частостями. Напомним, что при равномерном распределении каждая позиция из восьми будет иметь одинаковую теоретическую частоту, равную 0,125.

d_{max}=0,135. Если $d_{эмп.} \geq d_{крит.}$, то различия между распределениями могут считаться достоверными. Критические значения представлены в Приложении 2 (табл. 6), т.к. n>100, то $d_{крит.}$ рассчитывается по формуле $d_{êðèò.} = d/\sqrt{102}$

При уровне значимости p=0,05 $d_{êðèò.} = 1{,}36/\sqrt{102}$

$d_{крит.}$ =0,135**.**

Так как $d_{эмп.} = d_{крит.}$, то распределение желтого цвета по восьми позициям отличается от равномерного распределения.

Таблица 10

Расчет критерия при сопоставлении распределения выборов желтого цвета с равномерным распределением

Позиция желтого цвета	Эмпирическая частота	Эмпирическая частость	Накопленная эмпирическая частость	Накопленная теоретическая частость	Разность
1	24	0,235	0,235	0,125	0,110
2	15	0,147	0,382	0,250	0,132
3	13	0,128	0,510	0,375	0,135
4	8	0,078	0,588	0,500	0,088
5	15	0,147	0,735	0,625	0,110
6	10	0,098	0,833	0,750	0,083
7	9	0,088	0,921	0,875	0,046
8	8	0,079	1,000	1,000	0,000

φ*-критерий Фишера

Критерий φ* Фишера является многофункциональным статистическим критерием и предназначен для сопоставления уровней исследуемого признака, сдвигов в значениях исследуемого признака и сравнения распределений.

Критерий применим как к независимым, так и к связанным выборкам при использовании любой шкалы измерения, начиная с номинативной.

Для применения критерия необходимо свести любые данные к альтернативной шкале «Есть эффект — нет эффекта».

Критерий φ^* Фишера предназначен для сопоставления двух выборок по частоте встречаемости интересующего исследователя эффекта путем оценки достоверности различий между процентными долями двух выборок, в которых зарегистрирован интересующий нас эффект. При этом проверяются нулевая и альтернативная гипотезы.

Гипотезы

H_0: Доля, у которой проявляется исследуемый эффект, в выборке 1 не больше, чем в выборке 2.

H_1: Доля, у которой проявляется исследуемый эффект, в выборке 1 больше, чем в выборке 2.

Суть углового преобразования Фишера состоит в переводе процентных долей в величины центрального угла, который измеряется в радианах. Большей процентной доле будет соответствовать больший угол φ, а меньшей доле — меньший угол.

Угол φ в радианах вычисляется по следующей формуле:

$$\varphi = 2\arcsin\left(\sqrt{P}\right),$$

где P — процентная доля, выраженная в долях единицы.

При изменении P от 0 до 1 угол φ увеличивается от 0 до π (π = 3,14159...).

Величины угла φ (в радианах) для разных процентных долей приведены в Приложении.

Эмпирическое значение критерия вычисляется по формуле:

$$\varphi^* = \left|(\varphi_1 - \varphi_2)\right| \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

где φ_1 — угол, соответствующий процентной доле первой выборки;

φ_2 — угол, соответствующий процентной доле второй выборки;

n_1 и n_2 — объемы первой и второй выборок, соответственно.

Процедуру определения эмпирического значения φ^* рассмотрим на следующем примере.

Пример 4.

Две разные выборки студентов решали некоторую задачу. В первой выборке из 30 человек с нею справились 10 человек, а во второй выборке из 25 человек 15. Различаются ли две группы студентов по успешности решения задачи?

Решение.

В первом случае процентная доля решивших задачу (есть эффект) составит (10:30) 100 = 33,3%, то есть, P_1 = 33,3% а во втором случае доля решивших задачу составит (15:25) 100 = 60%, то есть, P_2 = 60%. Достоверно ли различаются эти процентные доли при данных n_1 и n_2?

По таблице 1 Приложения 2 определяем величины φ_1 и φ_2, соответствующие процентным долям P_1 = 33,3% и P_2 = 60%.

φ_1 = 1,230;

φ_2 = 1,772.

Таким образом, получим таблицу 11.

Таблица 11

Выборка	Решили задачу	Не решили задачу	Доля решивших задачу	Угол φ, соответствующий доле решивших задачу
Выборка 1	10	20	33,3%	1,772
Выборка 2	15	10	60%	1,230

Определим эмпирическое значение φ^*

$$\varphi^*_{\partial} = \left|(1{,}230 - 1{,}772)\right| \cdot \sqrt{\frac{30 \cdot 25}{30 + 25}} = 0{,}542 \cdot \sqrt{\frac{150}{11}} = 2{,}001$$

Эмпирическое значение критерия $\varphi^*_{эмп}$ сравним с критическими значениями критерия.

Критические значения критерия $\varphi^*_{кр}$ (по Гублеру Е.В., 1978) равны 1,64 (для p=0,05), 2,31 (для p=0,01) и 2,81 (для p=0,001) (таблица 2 Приложения 2).

$\varphi^*_{ф} > \varphi^*_{кр}$, соответствующий уровню значимости p=0,05. Таким образом, H_0 отвергается, принимается H_1. Доля лиц, справившихся с задачей во второй выборке больше, чем в первой.

Однако, не всегда результаты наблюдений представлены в виде «есть эффект — нет эффекта», поэтому возникает задача нахождения той критической, переломной точки, которая бы позволила перевести результаты количественных измерений к данному виду. Такую переломную точку можно найти, если использовать критерий Фишера в сочетании с критерием λ-Колмогорова-Смирнова. Рассмотрим пример иллюстрирующий это.

Пример 5.

Проводилось тестирование знаний по предмету «Математическая статистика» среди студентов, закончивших средние профессиональные учебные заведения и закончивших среднюю школу. Процентная доля выполнения задания представлена в таблице 12.

Таблица 12

Распределение студентов, выполнивших тест

Доля выполнения задания	Число выполнивших данную долю задания в каждой категории	
	Окончившие СПО (n=45)	Окончившие среднюю школу (n=25)
От 0 до 20%	4	5
от 21 до40%	15	11
От41 до 60%	18	5
от 61 до80%	7	4
от 81 до100%	1	0

Определим точку максимального расхождения между двумя распределениями ответов.

Таблица 13

Расчет максимальной разности накопленных частостей в распределениях, выполнивших тест

Доля выполнения задания	Число выполнивших данную долю задания в каждой категории		Частость		Накопленная частость		Разность
	окончили СПО	окончили ср. школу	окончили СПО	окончили ср. школу	окончили СПО	окончили ср. школу	
0-20%	4	5	0,089	0,200	0,089	0,200	0,111
21-40%	15	11	0,333	0,440	0,422	0,640	0,218
41-60%	18	5	0,400	0,200	0,822	0,840	0,018
61-80%	7	4	0,156	0,160	0,978	1,000	0,022
81-100%	1	0	0,022	0	1,000	1,000	0

Максимальная выявленная между двумя накопленными эмпирическими частостями разность составляет 0,218. А, следовательно, граничной точкой «есть эффект — нет эффекта» будем считать 40%, т.е. «эффект есть», если выполнено от 41 до 100% задания и эффекта нет, если выполнено от 0 до 40% задания. Таким образом, распределение по процентному выполнению задания будет следующим:

Таблица 14

Выполнено задания	Частоты	
	Окончивших СПШ	Окончивших ср. школу
От 0 до 40%	19	16
От 41 до 100%	26	9

По рассмотренному выше алгоритму применения критерия Фишера, получим:

Таблица 15

Выборка	От 41 до 100%	От 0 до 40%	Доля выполнивших от 41 до 100%	Угол φ, соответствующий доле решивших задачу
Окончивших СПШ	26	19	57,8%	1,727
окончивших ср. школу	9	16	36,0%	1,287

Определяем эмпирическое значение φ^*:

$$\varphi^*_{\delta} = \left|(1{,}727 - 1{,}287)\right| \cdot \sqrt{\frac{45 \cdot 25}{45 + 25}} = 0{,}440 \cdot 4{,}009 = 1{,}764.$$

Эмпирическое значение критерия $\varphi^*_{эмп}$ сравним с критическими значениями критерия.

Критические значения критерия $\varphi^*_{кр}$ (по Гублеру Е.В., 1978) равны 1,64 (для p=0,05), 2,31 (для p=0,01) .

$\varphi^*_{ф} > \varphi^*_{кр}$, соответствующий уровню значимости p=0,05. Таким образом, H_0 отвергается, принимается H_1. Доля лиц, справившихся с тестом на 41-100% в группе, закончивших СПО больше, чем в выборке, закончивших среднюю школу.

Задание для самостоятельной работы

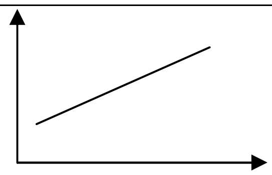
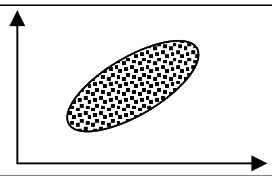
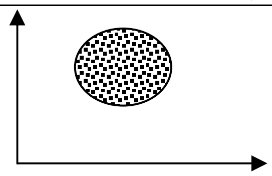
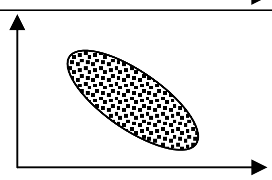
1. По данным таблицы 1 Приложения 1 определить существенность различий между показателями 1985, 1989, 1994, 1998 и 2000 годов.
2. По данным таблицы 4 Приложения 1 определить существенность сдвига продуктивности скота и птицы за периоды 1965-1985, 1985-1990 г.г.
3. По данным таблицы 13 Приложения 1 определить, используя рассмотренные выше критерии, является ли значительным прирост в благоустройстве жилищного фонда (1965 и 1980, 1980 и 1990 годы).
4. По данным таблицы 14 Приложения 1 определить, используя рассмотренные выше критерии, значительность сдвига урожайности по представленным периодам.

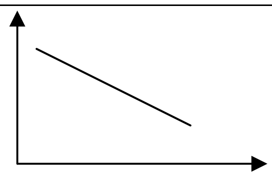
КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Для выявления связи между показателями используют коэффициент корреляции: линейный (по Пирсону) и ранговый (по Спирмену). Линейный коэффициент корреляции используется для выявления связи на тех выборках, где распределение подчинено нормальному закону; ранговый коэффициент применим для любых распределений. Коэффициент корреляции изменяется в пределах от -1 до 1 и обозначается r.

Таблица 16

Интерпретация коэффициента корреляции

Значение r	Сила связи	Графическая интерпретация
1	Строгая прямая связь	
0,5<r<1	Средняя прямая связь	
0	Связь отсутствует	
-1<r<-0,5	Средняя обратная связь	

Значение r	Сила связи	Графическая интерпретация
-1	Строгая обратная связь	

Линейный коэффициент корреляции подсчитывается по формуле:

$$r_{xy} = \frac{\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)}{\sqrt{\sum\left(x-\bar{x}\right)^2 \sum\left(y-\bar{y}\right)^2}}$$

После преобразований получим: $r_{xy} = \dfrac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{x^2 - \overline{(x)}^2}\sqrt{y^2 - \overline{(y)}^2}}$

Однако, для вычисления линейного коэффициента корреляции необходимо, чтобы распределение было нормальным, поэтому на небольших выборках следует применять ранговый коэффициент корреляции, который вычисляется по формуле: $r_s = 1 - \dfrac{6\sum d_i^2}{n(n^2-1)}$, где n — объем выборки, d_i — разность рангов соответствующих значений.

В практических исследованиях о тесноте связи судят по величине выборочного коэффициента корреляции и поскольку вычисленная величина является величиной случайной, то оценивается значимость коэффициента корреляции по t-статистике Стьюдента с (n-2) степенями свободы:

$$t_{ôàêò} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Нулевая гипотеза H_0 — коэффициент корреляции не является статистически значимым, т.е. линейная корреляционная связь между переменными отсутствует, альтернативная гипотеза H_1 — выборочный коэффициент значимо отличается от нуля, т.е. между показателями наблюдается линейная корреляционная зависимость. Нулевая гипотеза отвергается, если $t_{факт.} > t_{кри.}$ на данном уровне значимости. $t_{крит.}$ определяется по таблице 5 Приложения 2.

Пример 1. Расчета рангового коэффициента корреляции.

Пусть при исследовании десяти человек получены следующие показатели X и Y. Выясним, существует ли между ними связь. Для этого подсчитаем ранговый коэффициент корреляции и дадим его графическую интерпретацию.

Таблица 17

X	Y
175	*2*
176	*3*
179	*8*
180	*9*
181	*6*
184	*7*
185	*13*

Найдем ранг (порядковый номер по убыванию) каждого из значений x и y: R_x и R_y. Затем найдем разности соответствующих рангов d, возведем их в квадрат, получим ряд значений d^2. Если значения одинаковые, то приписывается промежуточный средний ранг, например, 6,5.

186	*11*
191	*10*
192	*12*

Просуммируем их и подставим в формулу:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}.$$

Таблица 18

№	X	Y	R_x	R_y	$\lvert d \rvert$	d^2
1	175	2	1	1	0	0
2	176	3	2	2	0	0
3	179	8	3	5	2	4
4	180	9	4	6	2	4
5	181	6	5	3	2	4
6	184	7	6	4	2	4
7	185	13	7	10	3	9
8	186	11	8	8	0	0
9	191	10	9	7	2	4
10	192	12	10	9	1	1
Сумма:						30

В нашем случае: $r_s = 1 - \frac{6*30}{10(10^2-1)} = 0{,}81$.

Оценим значимость коэффициента корреляции:

$$t_{\hat{o}\grave{a}\hat{e}\grave{o}} = \frac{0{,}81\sqrt{10-2}}{\sqrt{1-0{,}81^2}} = 3{,}92.$$

По таблице 5 Приложения 2 определяем, что для уровня значимости p=0,05 $t_{крит.}$=2,31. Следовательно, вычисленный коэффициент корреляции значимо отличается от нуля и между показателями x и y наблюдается линейная связь выше среднего.

Для графической интерпретации по оси x откладываются значения признака x, по оси y — значения признака y.

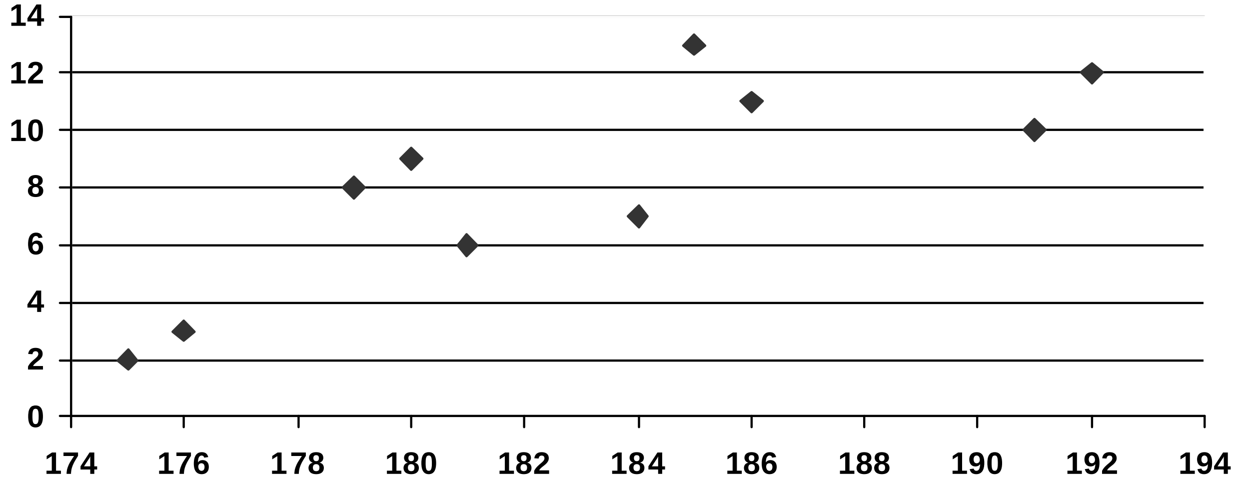


Рис. 6. Графическая интерпретация коэффициента корреляции.

Пример 2. Расчет линейного коэффициента корреляции.

Выясните наличие связи между показателем вложения в рекламу и средним доходом по данным, представленным в таблице 19.

Таблица 19

Х	*У*
5	10
8	1
13	5
3	8
10	4
8	7
4	9
7	0
9	6
2	2
10	3

Таблица 20

Х	*У*	X^2	*ХУ*	$У^2$
5	10	25	50	100
8	1	64	8	1
13	5	169	65	25
3	8	9	24	64
10	4	100	40	16
8	7	64	56	49
4	9	16	36	81
7	0	49	0	0
9	6	81	54	36
2	2	4	4	4
10	3	100	30	9
Сумма 79	**55**	**681**	**367**	**385**

$$r_{xy} = \frac{367\text{-}395}{\sqrt{681\text{-}567}\sqrt{385-275}}$$

r_{xy}=0,25.

Аналогично рассмотренному ранее, анализируется значимость связи и дается графическая интерпретация.

РЕГРЕССИОННЫЙ АНАЛИЗ

Изучение корреляционных зависимостей основывается на исследовании таких связей между переменными, при которых значения одной переменной (ее можно принять за зависимую переменную) «в среднем» изменяются в зависимости от того, какие значения принимает другая переменная.

Теоретической линией регрессии называется та линия, вокруг которой группируются точки корреляционного поля, и которая указывает основное направление, основную тенденцию связи.

Теоретическая линия регрессии должна отображать изменение средних величин результирующего признака y по мере изменения факторного признака x. Следовательно, эта линия должна быть проведена так, чтобы сумма отклонений точек поля корреляции от соответствующих точек теоретической линии регрессии была бы минимальной величиной. Важным этапом регрессионного анализа является определение типа функции, с помощью которой характеризуется зависимость между признаками. Приблизительное представление о линии связи можно получить на основе эмпирической линии регрессии, получаемой графическим методом.

Графический метод построения линии регрессии

Алгоритм построения:

1. Разбить весь диапазон изменения признака X на классы;
2. Определить средние значения x для каждого класса и соответствующее среднее значение y_x для каждого класса;
3. Отметить на координатной плоскости точки с координатами ($x_{ср}$; y_x);
4. Соединить полученные точки линиями.

Однако, чаще всего при построении графическим методом, линия регрессии получается в виде ломаной линии, которую необходимо сгладить, заменив ее более плавной линией. Для этого вычисляют среднее арифметическое двух или трех соседних значений ряда.

Наиболее часто для характеристики связи различных показателей используют следующие типы функций:

линейную $y'=a+bx$;

гиперболическую $y' = a + b\dfrac{1}{x}$

показательную $y'=ab^x$;

параболическую $y'=a+bx+cx^2$;

степенную $y'=ax^b$;

логарифмическую $y'=a+b\lg x$.

Рассмотрим наиболее простой случай — случай линейной функции $y'=a+bx$.

Задача построения линии регрессии сводится к нахождению коэффициентов a и b. Будем использовать для этого метод наименьших квадратов.

Суть метода наименьших квадратов состоит в нахождении такой функции, которая наилучшим образом соответствует эмпирическим данным, считая, что сумма квадратов отклонений эмпирических точек теоретической линии регрессии должна быть величиной минимальной.

Поскольку $y'=a+bx$, то

$$S = \sum_1^n \left(y_i - y'(x_i)\right)^2 \to \min \qquad \sum_1^n \left(y_i - (a + bx_i)\right)^2 \to \min$$

Таким образом, определение параметров a и b прямой, наиболее соответствующей эмпирическим данным, сводится к задаче на экстремум.

$$\frac{dS}{da} = 0 \qquad \frac{dS}{db} = 0$$

Функция двух переменных $S(a;b)$ может достигать экстремума в том случае, когда

$$\frac{dS}{da} = -2\sum_1^n (y_i - a - bx_i) = 0 \qquad \frac{dS}{db} = -2\sum_1^n (y_i - a - bx_i)x_i = 0$$

Следовательно, из первого равенства

$$\sum_1^n y_i = \sum_1^n (a + bx_i) \qquad \sum_1^n y_i = an + b\sum_1^n x_i \qquad a = \frac{\sum_1^n y_i - b\sum_1^n x_i}{n} = \bar{y} - b\bar{x}$$

Подставим найденное значение, а во второе равенство:

$$\sum_1^n y_i x_i = a\sum_1^n x_i + b\sum_1^n x_i^2 \qquad \sum_1^n y_i x_i - \bar{y}\sum_1^n x_i = b\sum_1^n\left(x_i^2 - \bar{x}\sum_1^n x_i\right)$$

$$b = \frac{\sum_1^n y_i x_i - \bar{y}\sum_1^n x_i}{\sum_1^n x_i^2 - \bar{x}\sum_1^n x_i} = \frac{\sum_1^n y_i x_i - \bar{y}\bar{x}}{\sum_1^n x_i^2 - \bar{x}^2 n}$$

Из последнего равенства находим коэффициент b, а затем и $a = \bar{y} - b\bar{x}$.

Пример 1.

Имеются следующие данные (таблица 21) о сменной добыче угля на одного рабочего $y(m)$ и мощности пласта $x(м)$, характеризующие процесс добычи угля в 10 шахтах.

Составить линейное уравнение зависимости сменной добычи от мощности пласта.

Оценить сменную добычу угля на одного рабочего для шахт с мощностью пласта 8 м.

Таблица 21

x	***y***
8	5
11	10
12	10
9	7
x	***y***
8	5
8	6
9	6
9	5
8	6
12	8

Таблица 22

x	***y***	*xy*	x^2
8	5	40	644
11	10	110	121
12	10	120	144
9	7	63	81
x	***y***	*xy*	x^2
8	5	40	64
8	6	48	64
9	6	54	81
9	5	45	81
8	6	48	64
12	8	96	144
Сумма 94	68	664	908
ср.зн. 9,4	6,8	66,4	90,8

Подставим полученные значения в формулу:

$$\sum_1^n y_i = an + b\sum_1^n x_i$$

Получим 68=10a+b94, a=(68-94b)/10,

$$\sum_1^n y_i x_i = a\sum_1^n x_i + b\sum_1^n x_i^2$$

Подставим во второе уравнение

664=a94+b908, тогда 9,4(68-94b)+908b=664

639,2-883,

6b+908b=664,

24,4b=24,8,

b=1,016,
a=(68-94*1,016)/10,
a=-2,754.
Значит, уравнение линии регрессии: y=-2,754+1,016x.
Построим ее, сравнив с эмпирическими значениями.

x	$У$
4	1,31
10	7,406

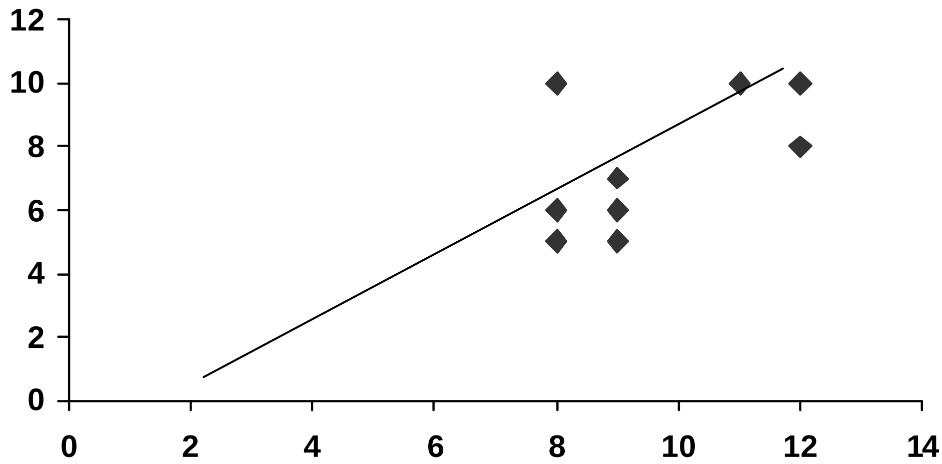


Рис. 7. Линия регрессии.

Таким образом, при увеличении мощности пласта на 1 м добыча угля на одного рабочего в среднем увеличивается на 1,016 т., тогда при x=8
y=-2,75+1,016*8=5,38 т.

Задания для самостоятельной работы

Таблица 23

Возраст работников управления инженерно-технической службы

№	**Возраст (*X*)**	**Число случаев временной нетрудоспособности (*Y*)**
1	41	2
2	43	1
3	31	2
4	33	3
5	33	2
6	50	1
7	28	2
8	38	2
9	35	3
10	30	2
11	33	4
12	64	2
13	32	1

№	Возраст (*X*)	Число случаев временной нетрудоспособности (*Y*)
14	60	2
15	30	3
16	29	2
17	45	2
18	49	1
19	39	2
20	47	1
21	34	3
22	23	1
23	44	1
24	32	1
25	46	2
26	26	1
27	46	1
28	32	1
29	36	1
30	33	1
31	49	1
32	32	1
33	36	1
34	39	2
35	48	1
36	29	1
37	36	1
38	44	1
39	25	1
40	33	1

1. Для выяснения существования связи между показателем возраста работников управления инженерно-технической службы при МЗ РТ и числом случаев временной нетрудоспособности подсчитайте ранговый коэффициент корреляции и дайте его графическую интерпретацию. В случае значимой линейной связи построить линию регрессии.

2. Проанализируйте корреляционную связь между стажем и заработной платой. В случае значимой линейной связи построить линию регрессии.

Таблица 24

Сведения о стаже и заработной плате рабочих
на промышленном предприятии

№ п/п	Стаж, лет	Зарплата, т.р.
1	0,8	575,8
2	2	576,4
3	2	691,8

№ п/п	Стаж, лет	Зарплата, т.р.
4	2	704,5
5	3	619,7
6	4	614,1
7	4	714,5
8	4	764,3
9	4	804
10	4	801,5
11	4	907
12	5	674
13	5	705
14	6	1307,4
15	7	587,3
16	7	714,5
17	7	763,1
18	8	1100,1
19	8	1121,3
20	9	1100,9
21	10	814,4
22	10	860,5
23	10	871,3
24	11	767,5
25	11	904,4
26	12	1409,4
27	15	1499,7
28	16	1607,4
29	17	1500,5
30	19	1598,5

3. Рассмотрим уровень газификации на примере 4 районов: Авиастроительного, Московского, Кировского и Вахитовского районов г. Казани. В указанных районах имеется 27 различных посёлков и населённых пунктов. Выяснить зависит ли уровень газификации от количества домов в поселке. В случае значимой линейной связи построить линию регрессии.

Таблица 25

№№	Наименование	Количество домов	Количество газифицированных домов
	Авиастроительный район		
1.	п. Сухая река	129	17
2.	п. Щербаково	416	146
3.	п. Северный	104	12
4.	п. Грабарский	270	20
5.	п. Крутушка	40	21

№№	Наименование	Количество домов	Количество газифицированных домов
6.	п. Новое-Караваево	26	17
7.	п. Борисоглебская	254	168
8.	п. Кадышева	336	235
9.	п. Авиастроитель	84	73
	Вахитовский район		
10.	п. Ометьево	187	187
11.	п. Калуга	982	982
	Кировский район		
12.	п. Юдино	366	300
13.	п. Н.Юдино	116	150
14.	п. Калининский	275	168
15.	п. Куземетьево	199	187
16.	п. Займище	411	357
17.	п. Красная горка	435	415
18.	п. Новое Аракчино	396	193
19.	п. Старое Аракчино	131	98
20.	п. Игумново-Лагерная	434	400
21.	п. Ново-Савиново	67	43
22.	п. Старое Савиново	64	45
23.	п. Дружба	435	435
	Московский район		
24.	п. им. Урицкого	94	93
25.	п.п. Левченко	74	74
26.	п. Жилплощадка	3	3
27.	п. Краснооктябрьская	187	150

4. 1) Провести статистический анализ по выплате субсидий на основании постановления №50 от 19 февраля 2001.”Возмещение части процентных ставок из бюджета Российской Федерации по кредитам, выданным предприятиям агропромышленного комплекса”(субсидии АПК).

2) Охарактеризовать тесноту связи между налоговыми выплатами и объемом предоставляемых субсидий предприятиям АПК (коэф. корреляции). В случае значимой линейной связи построить линию регрессии.

Для статистического анализа было выбрано случайным методом 30 предприятий АПК из 300, по которым производилось возмещение части процентов по кредитам (субсидирование).

Таблица 26

№	Наименование АПК	Объем субсидий (в тыс. руб.)	Налоговые выплаты (в тыс. руб)
1	Азнакаевский маслодельный завод	1 010	2 500
2	ОАО Кукморский маслодельный завод	900	1 200
3	ОАО “Агрофирма Сосновоборская	35	150
4	ГУП птицефабрика «Юбилейная»	1200	1800
5	ОАО «Атнинский маслодельный завод»	400	850
6	ОАО «Муслюмовский завод СОМ»	45	300
7	АО «СОТ» г. Наб. Челны	1500	5600
8	ООО «Фермерское хозяйство Сафия»	48	160
9	ОАО «ХК Татарстан СЭТЭ»	1000	3200
10	ОАО «Черемшанский сырзавод»	168	200
11	ООО «Большие Ковали»	300	100
12	АКХ Кубня	125	78
13	ОАО «Прогресс»	70	48
14	ФХ Исмагиловых	1400	1800
15	Елабужский консервный завод	560	1000
16	АКХ Марс	90	45
17	СА «Новая жизнь»	1400	800
18	Лаишевский молочный комбинат	1100	1600
19	СХПК «Оч Ойле»	154	320
20	ОАО Дрожжановский сыр завод	650	1200
21	ПК «Камский» Тукаевский район	780	2000
22	ФК «Родник» Балтасинский район	321	450
23	СПК «Игенче» Балтасинский район	200	155
24	ОСХК «Узяк» Актанышский район	560	1200
25	ОАО «Сармановский молочный завод»	1500	1800
26	ООО «Тетра-Инвест»	890	1000
27	ООО «Управляющая компания Золотой колос»	3000	6000
28	АККХ «50 лет Татарии»	770	900
29	Крестьянское хозяйство «Земляки»	122	89
29	АККХ «Юлдуз» Кайбицкий район	500	258

5. В городе Наб. Челны действуют 27 высших и средних специальных учебных учреждений (филиалов), обучающих студентов на платной основе. Проанализировать: существует линейная зависимость между количеством обучающихся и суммой оплаты за обучение.

Таблица 27

№	Учебное заведение	Кол.студ. в тыс. чел.	Оплата в тыс.руб.
1	Камский гос. политехнический институт	12,9	17,85
2	Набережночелнинский гос. педагогический институт	4,160	14,5
3	Камский гос. институт физической культуры	1,041	12,6
4	Набережночелнинский филиал КГУ	2,702	20,2
5	Набережночелнинский филиал КГТУ	0,404	21,5
6	Набережночелнинский филиал Нижегородского университета им. Н.А.Добролюбова	0,260	17,0
7	Институт управления	0,378	12,0
8	Региональный институт передовых технологий и бизнеса	0,220	11,0
9	Волжско-камская академия туризма	0,250	15,0
10	Набережночелнинский филиал СГИ	0,926	19,0
11	Татарско-американский региональный институт	0,283	15,0
12	Филиал ТИСБИ	0,750	12,0
13	Набережночелнинский филиал Казанского ИУЭП	1,169	15,0
14	Набережно-челнинский филиал КГПУ	0,740	15,0
15	Камский институт	0,876	15,0
16	Филиал Московской государственной технологической академии	0,300	11,0
17	Филиал Московского университета культуры и искусства	0,364	8,0
18	Камский институт экономики, статистики и права	0,340	9,0
19	Филиал Московского социально-гуманитарного института	0,250	12,0
20	Училище искусств	0,284	8,0
21	Камский государственный автомеханический техникум	2,142	9,5
22	Набережночелнинский экономико-строительный колледж	0,143	8,0
23	Набережночелнинский медицинский колледж	0,448	9,4
24	Камский юридический колледж	0,172	11,2
25	Колледж «Идель-Урал»	0,260	5,7
26	Камский экономико-правовой колледж	0,170	7,5
27	Колледж при ИУЭП	0,399	7,0

6. По данным таблицы 6 Приложения 1 построить линию регрессии зависимости числа многодетных матерей, а также количества детей у них от года переписи.

ДИСПЕРСИОННЫЙ АНАЛИЗ

При решении многих статистических задач приходится рассматривать зависимость результативного признака от многомерного фактора x=(x_1, x_2, x_p), составляющие $x_к$, $к$=1,…, p которого в дальнейшем будем называть его уровнями.

Примером таких задач могут служить задачи изучения зависимости качества воспроизведения учебного материала от скорости его подачи, урожайности от дозы веденных удобрений, прибыли от вложения в рекламу и т.п.

При решении подобных задач обычно используется дисперсионный метод, заключающийся в сравнении факторной и остаточной дисперсии, на которой разбивается общая дисперсия результативного признака у.

Факторная дисперсия $S^2_{факт.}$ вызвана действием на случайную величину y фактора x, а остаточная ($S^2_{ост}$) целиком обусловлена случайными причинами.

Однофакторный дисперсионный анализ

Пусть на результаты признака y воздействует фактор x, имеющий p постоянных уровней, т.е. x=(x_1, x_2, …x_p). На каждом из p уровней произведем одинаковое число испытаний, равное q, и результаты этих испытаний оформляем в виде таблицы 28 (см. ниже). X разбивает наблюдаемые значения на p групп и значения этих групп образуют столбцы таблицы. В последней строке таблицы приведены найденные значения групповых средних.

Общая выборочная средняя $\bar{y} = \dfrac{\sum_{j=1}^{p} \overline{y_{ir}}}{p}$.

Таблица 28

№ испытания	Уровни фактора			
	X_1	X_2	…	X_p
1	y_{11}	y_{12}	…	y_p
2				
. . . .				
Q	y_{q1}	y_{q2}	…	y_{qp}
Групповая средняя	y_{r1}	y_{r2}	…	y_{r3}

Используя данные таблицы, найдем:

1. Общую сумму квадратов отклонений наблюдаемых значений от общей средней $\bar{ó}$: $R_{îáù} = \sum_{j=1}^{ð} \sum_{i=1}^{q} (y_{ij} - \bar{y})^2$ характеризует рассеяние всех наблюдаемых значений от общей средней.

2. $R_{ôàêò} = q\sum_{j=1}^{ð}(\overline{y_{rj}} - \overline{y})^2$ — факторная сумма квадратов отклонений групповых средних от общей средней и она характеризует рассеяние групп около общей средней.

3. $R_{îñò} = \sum_{j=1}^{p}\sum_{i=1}^{q}(y_{ij} - \overline{y_{ri}})^2$ — остаточная сумма квадратов отношений наблюдаемых значений группы от своей групповой средней, которая уже характеризует рассеяние внутри группы. Общая сумма дисперсии σ^2 складывается из сумм межгрупповой и внутригрупповой дисперсий.

Аналогично:

$$R_{общ}=R_{факт.}+R_{ост.}.$$

При практическом использовании дисперсионного анализа обычно вычисляют $R_{общ}$ и $R_{факт}$, а $R_{ост}$. находят из разности: $R_{общ}$. — $R_{факт}$.

Решение статистических задач методом дисперсионного анализа

Пусть на конечный результативный признак у, имеющий нормальное распределение, воздействует фактор x (x_1, x_2… x_p) с p-уровнями и необходимо установить, оказывает ли этот фактор существенное влияние на признак у.

Если изучаемый фактор x (x_1, x_2… x_p) оказывает существенное влияниена y, то групповые средние $\overline{ó_{rj}}$ $\; j=\overline{1,p}$ существенно различаются между собой.

Поэтому одновременно значимо будут различаться $R_{факт}$ и $R_{ост}$. А значит отношение $\frac{R_{ôàêò}.}{R_{îñò}.}$ будет заметно больше 1.

При решении практических задач обычно рассматривают отношение факторной и остаточной дисперсии, подсчитывается:

$$F_{íàáë} = \frac{S^2_{ôàêò.}}{S^2_{îñò.}}, \text{ где } S_{ôàêò.} = \frac{R_{ôàêò.}}{ð-1}, \quad S^2_{îñò.} = \frac{R_{îñò.}}{p(q-1)}$$

и задача сводится к сравнению факторной и остаточной дисперсий по критерию Фишера-Снедекора. Из таблицы 10 Приложения 2 критических значений находим $F_{кр}$.

Если $F_{набл}.>F_{кр}.$, то нулевая гипотеза о равенстве факторной и остаточной дисперсий отвергаются.

Пример 1.

При уровне значимости α=0,05 исследовать степень влияния стажа рабочих предприятия на мотивацию труда.

Таблица 29

Стаж	1-5 л.	6-10 л.	11 л и более
1 бр.	20	22	24
2 бр.	21	23	25
3 бр.	22	24	26
	21	23	25

Результат признака *у* — мотивация труда зависит от стажа, который имеет 3 уровня

$P=q=3 \quad \bar{б}=23$

$r_{общ}.=\Sigma\Sigma\ 3^2+2^2+1^2+1^2+1^2+1^2+2^2+3^2=4+8+18=12+18=30.$

$r_{факт}.=3(22+22)=24$, $r_{ост}.=30-24=6.$

$S^2_{ôàêò} = \frac{24}{2} = 12$, $S^2_{îñò} = \frac{6}{3*2} = 1$.

$F = \frac{12}{1}$, $F_{кр}(0{,}05;2;6)=5{,}14$ (из таблицы 10 Приложения 2).

$F_{набл}.>F_{кр}.$, поэтому нулевая гипотеза о равенстве факторной и остаточной дисперсий отвергаются. А это означает, что факторная дисперсия значимо отличается от остаточной дисперсии.

Задания для самостоятельной работы

1. В течение шести лет использовались пять различных технологий по выращиванию сельскохозяйственной культуры. Данные по эксперименту (в ц/га) приведены в таблице:

Таблица 30

Номер наблюдения (год)	Технология (фактор А)				
	А1	А2	А3	А4	А5
1	1.2	0.6	0.9	1.7	1.0
2	1.1	1.1	0.6	1.4	1.4
3	1.0	0.8	1.3	1.3	1.1
4	1.3	0.7	1.0	1.5	0.9
5	1.1	0.7	1.0	1.2	1.2
6	0.8	0.9	1.0	1.3	1.5
итого	6.5	4.8	5.4	8.4	7.1

2. Необходимо на уровне значимости *a*=0.05 установить влияние различных технологий на урожайность культуры. На заводе установлено четыре линии по выпуску облицовочной плитки. С каждой линии, случайным образом, в течение смены отобрано по 10 плиток и сделаны замеры их толщины (мм). Отклонения от номинального размера приведены в таблице:

Таблица 31

Линия по выпуску плиток	Номер испытания									
	1	2	3	4	5	6	7	8	9	10
1	0.6	0.2	0.4	0.5	0.8	0.2	0.1	0.6	0.8	0.8
2	0.2	0.2	0.4	0.3	0.3	0.6	0.8	0.2	0.5	0.5
3	0.8	0.6	0.2	0.4	0.9	1.1	0.8	0.2	0.4	0.8
4	0.7	0.7	0.3	0.3	0.2	0.8	0.6	0.4	0.2	0.6

3. По данным Таблицы 6 Приложения 1 определить оказал ли фактор времени существенное влияние на распределение числа детей в многодетных семьях.

СТАТИСТИЧЕСКИЙ АНАЛИЗ В EXCEL

В общем случае запуск приложения производится стандартным образом: Пуск → Программы → Excel.

Форматирование текста в Excel производится таким же образом, как и в текстовом редакторе Word.

Верхняя часть открывшегося листа содержит 5 строк:

Окно Excel: 1 строка — Заголовок

2 строка — Меню (File, Правка,…)

3 строка — Стандартная панель инструментов

4 строка — Форматирование

5 строка — Строка Формул.

Признаком формулы является стоящий впереди знак «=»

Все формулы набираются латинскими буквами.

Рабочая книга состоит из 3-х листов, но можно увеличить до 255. Рабочий лист состоит из 256 столбцов и 65536 строк.

Каждая ячейка однозначно идентифицируется номером столбца: A, B, C, …AA, AB, …. и номером строки (1, 2, 3,..).

Выбрав команду: «Формат ячейки» можно задать формат: %, денежный, цифровой,…

Ячейки могут иметь абсолютную и относительную адресацию.

Абсолютность адресации задает знак «$» (при копировании: — $A1*$A2, т.е. это значение). При абсолютной ссылке копирование сохраняет данный адрес.

Excel имеет большой набор функций, который активизируется следующим образом: на верхней панели с помощью мыши выбирается *fx*, из раскрывшегося окна выбираются Статистические. Для получения справки о формате ввода данных и формуле вычисления функции следует обратиться к информации справочного характера путем нажатия кнопки «?».

Также большими возможностями обладает пакет анализа, который активизируется следующим образом:

Сервис → Настройки → Пакет анализа.

После этого в меню «Сервис» появляется строка «Анализ данных», активируя которую можем выбрать из большого количества предлагаемых статистических процедур необходимую. Разберем пример применения пакета анализа в случае однофакторного дисперсионного анализа.

Для начала введем необходимые данные в электронную таблицу. Задача состоит в следующем:

Проверить статистическую существенность влияния катализатора A на химическую реакцию. Результаты измерений при 5 уровнях фактора A приведены в таблице 32 (см. ниже).

После подключения «пакета анализа», предварительно введя данные в ячейки: A1-E7, запускаем процедуру «однофакторный дисперсионный анализ».

Входной интервал $A1$:E7 указывает ту область по диагонали, в которую мы вводим данные.

Группировка по столбцам означает те уровни фактора, на которые разделены данные.

	A	B	C	D	E
1	A1	A2	A3	A4	A5
2	3,2	2,6	2,9	3,7	3
3	3,1	3,1	2,6	3,4	3,4
4	3,1	2,7	3	3,2	3,2
5	2,8	2,9	3,1	3,3	3,5
6	3,3	2,7	3	3,5	2,9
7	3	2,8	2,8	3,3	3,1
8					

Таблица 32

A1	A2	A3	A4	A5
3,2	2,6	2,9	3,7	3
3,1	3,1	2,6	3,4	3,4
3,1	2,7	3	3,2	3,2
2,8	2,9	3,1	3,3	3,85
3,3	2,7	3	3,5	2,9
3	2,8	2,8	3,3	3,1

Выходной интервал A11 — то место в таблице, в которое выводится результат.

Метки в первой строке — то, что первая строка, указанная во входном интервале, является заголовком. P=0,05 (уровень достоверности) — 5%-ая вероятность ошибки.

В результате получим:

Однофакторный дисперсионный анализ

ИТОГИ

Группы	*Счет*	*Сумма*	*Среднее*	*Дисперсия*
A1	6	18,5	3,08333333	0,02966667
A2	6	16,8	2,8	0,032
A3	6	17,4	2,9	0,032
A4	6	20,4	3,4	0,032
A5	6	19,1	3,18333333	0,05366667

Дисперсионный анализ

Источник вариации	*SS*	*df*	*MS*	*F*	*P-Значение*	*F критическое*
Между группами	1,342	4	0,3355	9,35408922	9,16424E-05	2,758710593
Внутри групп	0,89667	25	0,03586667			
Итого	2,23867	29				

$F_ф$ = 9,35, $F_{кр}$. = 2,76.

Так как $F_ф$. > $F_{кр}$., то фактор катализатора существенно влияет на реакцию.

Регрессия

Допустим, вычисляется прибыль, в зависимости от времени функционирования предприятия, и хотим узнать, какова будет прибыль через 5 лет. Для аналитического выравнивания и прогноза по уравнению прямой $y = ax + b$ можно использовать в категории «Статистические» следующие функции:

А) предсказания;

Б) тенденция.

Введем A1:A5 — время функционирования, B1:B5 — прибыль.

Заполним необходимые строки меню данных функций.

№	A	B
1	1	10
2	10	50
3	8	12
4	3	11
5	6	5

Получить меню регрессии можно также с использованием пакета анализа данных.

Сервис → Настройки → Пакет анализа → Анализ данных → Регрессия.

Корреляция

При помощи функции КОРРЕЛ вычисляется линейный коэффициент корреляции. Можно также использовать в этих целях пакет анализа:

Пакет анализа → Корреляция, либо *fx* → Корреляция → Пирсон.

Для подсчета рангового коэффициента корреляции можно использовать после введения данных *fx*:Ранг, а затем задать формулу.

Описательная статистика

Описательная статистика дает возможность подробного описания данных и осуществляется следующим образом: Пакет анализа → Описательная статистика. Гистограмма распределения строится с помощью мастера диаграмм. Описательная статистика позволяет вычислить размах вариации, максимум, минимум, дисперсию, стандартное отклонение и т.д.

Различие между двумя рядами можно оценивать по критерию Фишера: *fx* → Статистич. → Фишер.

Более качественный автоматизированный статистический анализ проводится с помощью различных статистических пакетов:

STATISTICA, SPSS и др.

СТАТИСТИЧЕСКИЙ АНАЛИЗ В ПРОГРАММЕ STATISTICA

STATISTICA является лидером среди программ статистической обработки данных в среде Windows. В ней реализован так называемый графически-ориентированный подход к анализу данных, который состоит во всестороннем визуальном представлении данных. Разделы, которые будут рассматриваться, соответствуют тем разделам, которые были рассмотрены на теоретическом уровне и требовали долгих утомительных вычислений на калькуляторе:

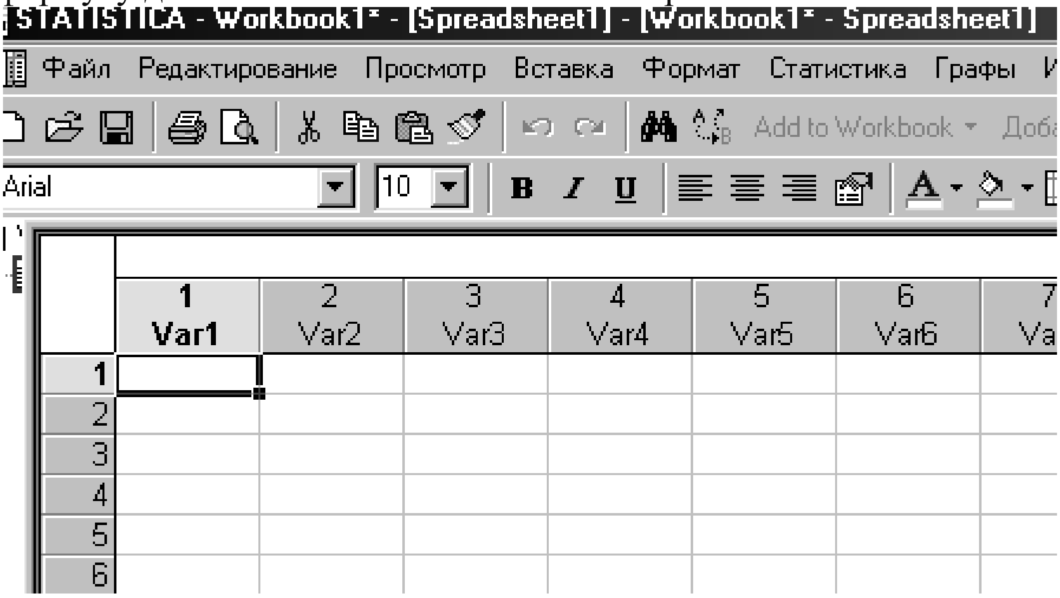
- простейшая статистика — среднее, стандартное отклонение, построение гистограммы;
- оценка значимости сдвига;
- корреляционный анализ;
- дисперсионный (однофакторный анализ);
- регрессионный анализ.

Общие правила работы с программой

После инсталляции программы STATISTICA запускается путем выбора кнопок Пуск → Программы → STATISTICA. На экране появляется рабочий лист, сходный с рабочим листом приложения EXCEL. Создание, открытие и сохранение файлов происходит общепринятым способом для Windows. Файлы имеют расширение sta.

Число необходимых строк и столбцов можно задать сразу, а можно варьировать с помощью мыши.

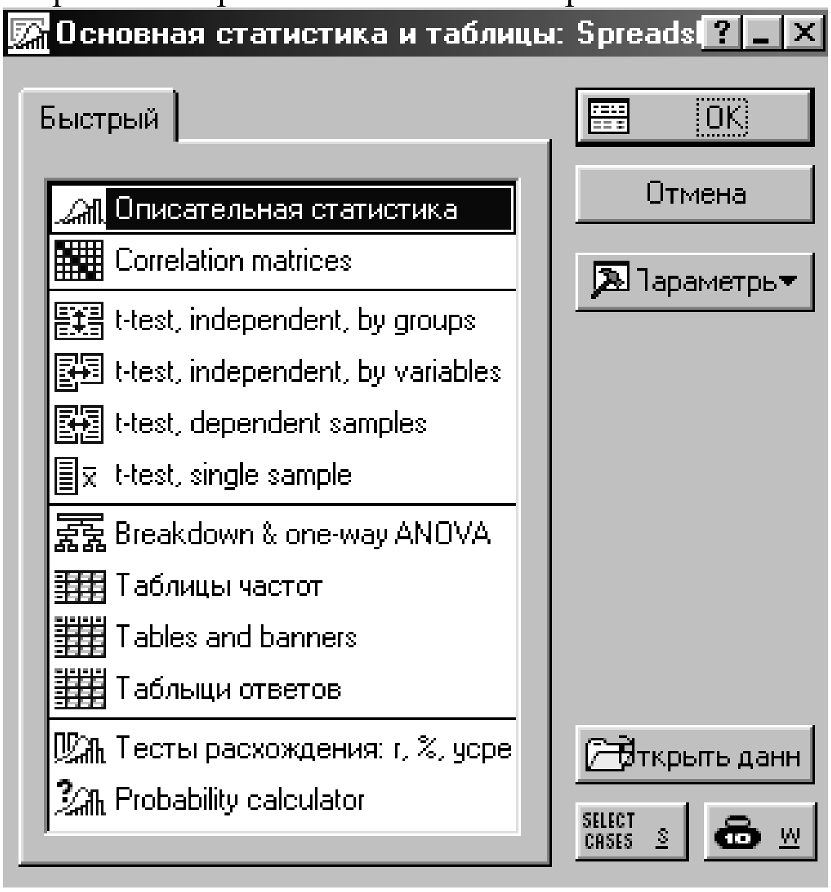
Дважды щелкнув на имени переменной Var1, можно задать ее имя в поле Name. Длина имени не должна превышать 8 символов. В поле Long name можно задать формулу для вычисления той или иной переменной.



Простейшая описательная статистика

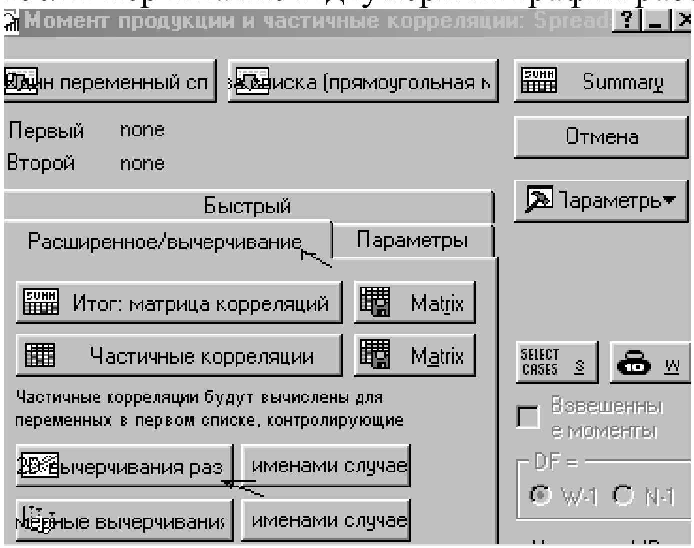
После введения данных в меню **Статистика** выбираем: основная статистика — описательная статистика, затем активизируем окно Variables, выбираем тот столбец, который хотим исследовать, и поставим флажки на тех параметрах, значения которых хотим узнать.

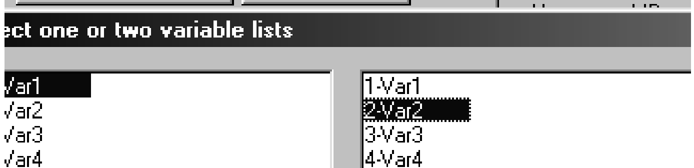
Чтобы получить гистограмму распределения и сравнить его с нормальным, следует выбрать: быстрая статистика-гистограмма.

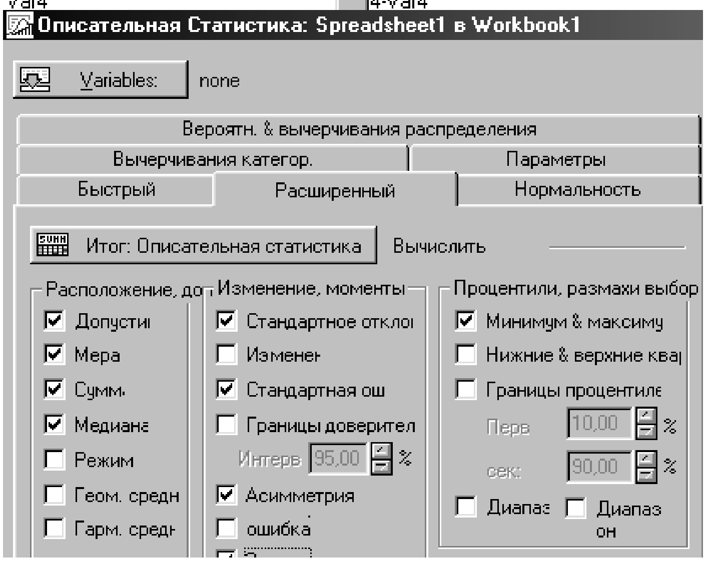


Вычисление корреляции и коэффициентов регрессии

Для вычисления линейного коэффициента корреляции следует выбрать Статистика → Основная, статистика → Correlation matrices. Затем выбрать вкладку расширенное/вычерчивание и двумерный график разброса.







В результате получим значение линейного коэффициента корреляции, линию регрессии и уравнение регрессии.

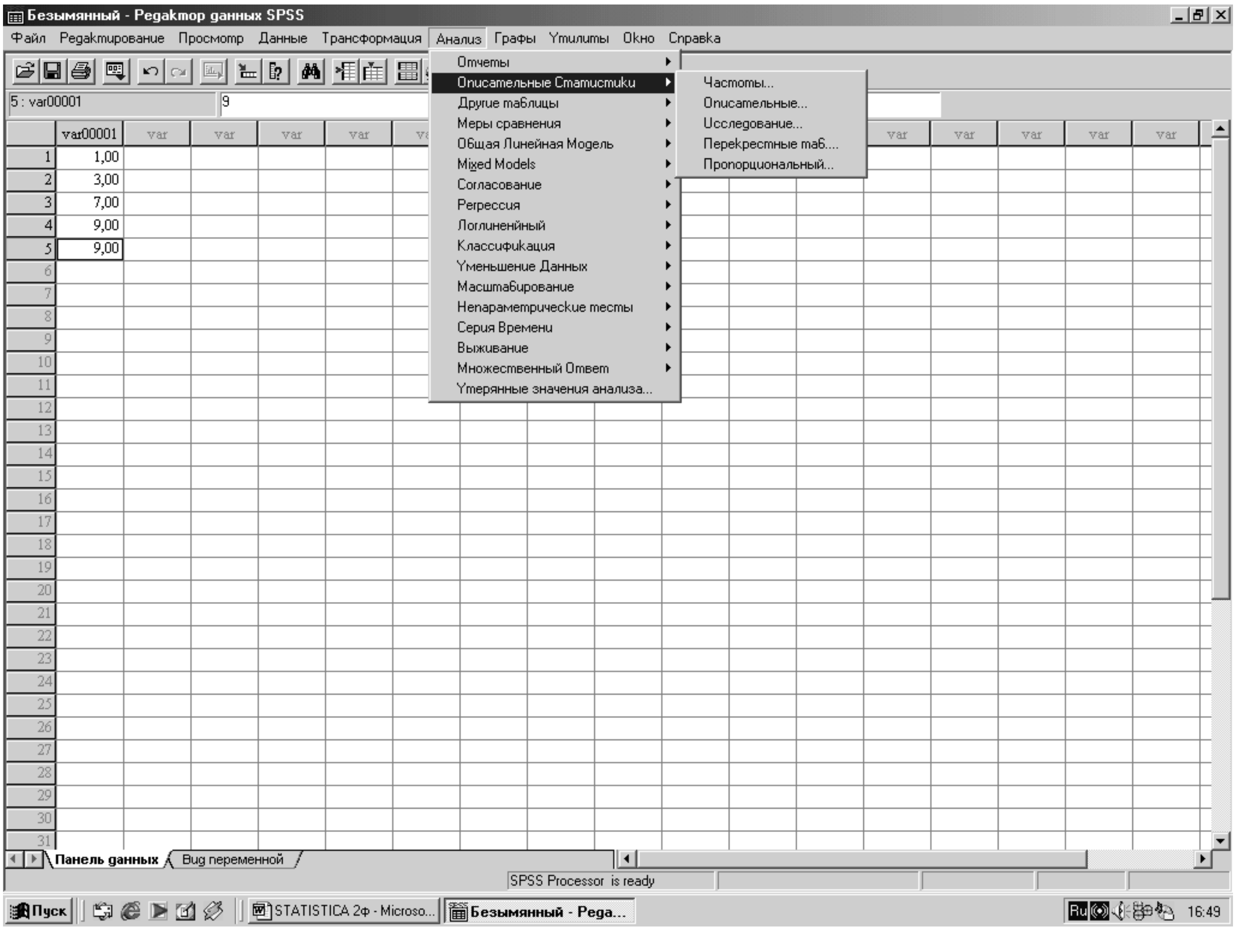
Есть возможность вычислить и ранговый коэффициент корреляции. Для этого: Статистика → непараметрические → correlations → RSpearmen. Затем задать переменные и после вычисления коэффициента корреляции выбрать вкладку — диаграмма рассеяния.

Для того, чтобы запустить модуль «Дисперсионный анализ» нужно выбрать вкладку «Другие виды анализа» из меню «Анализ», процедура «ANOVA» (Analysis of variance).

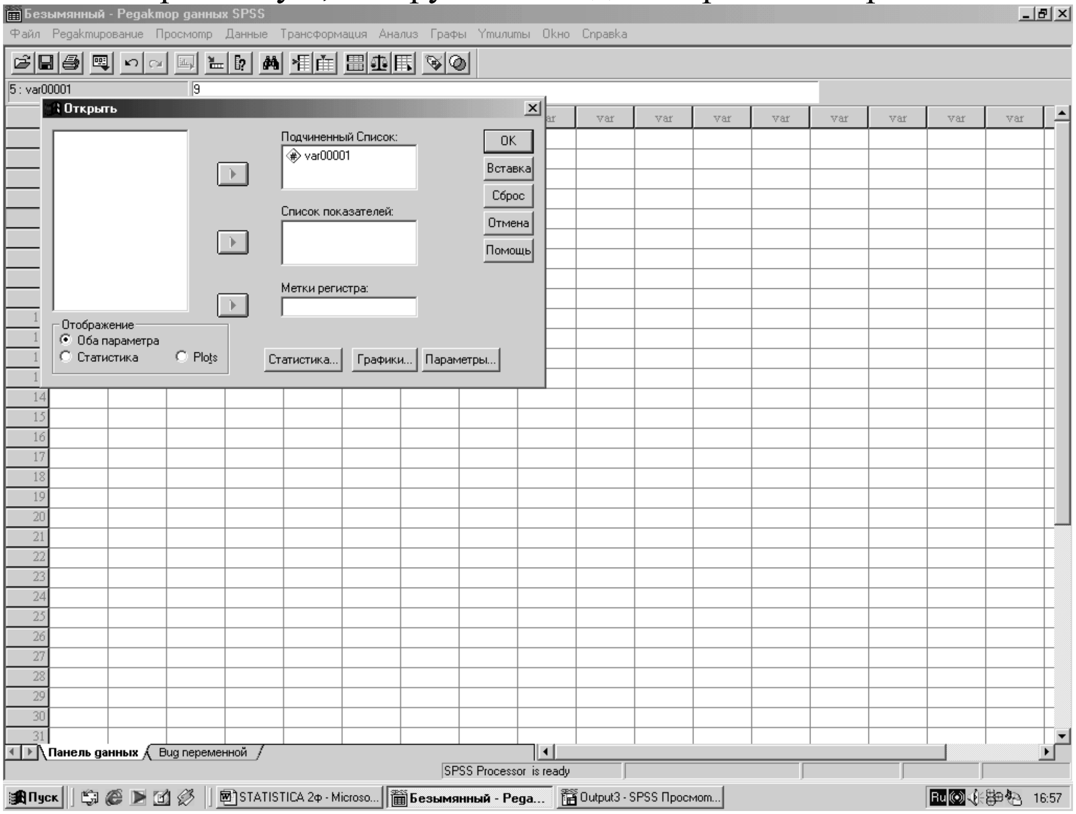
СТАТИСТИЧЕСКИЙ АНАЛИЗ В ПРОГРАММЕ SPSS

Процедуры, рассмотренные ранее, возможны и с помощью статистического пакета **SPSS.**

Так ***описательный анализ*** данных подключается с помощью модуля Анализ → Описательные, статистики → описательные. Предварительно в файл данных должны быть введены данные. Более подробный анализ можно получить выбрав: Анализ → Описательные статистики → Исследование. Рассмотрим пример, представленный на рисунке ниже



Укажем переменную, которую необходимо проанализировать:



В результате получим следующие результаты:

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
VAR00001	5	1,00	9,00	5,8000	3,63318
Valid N (listwise)	5				

Explore

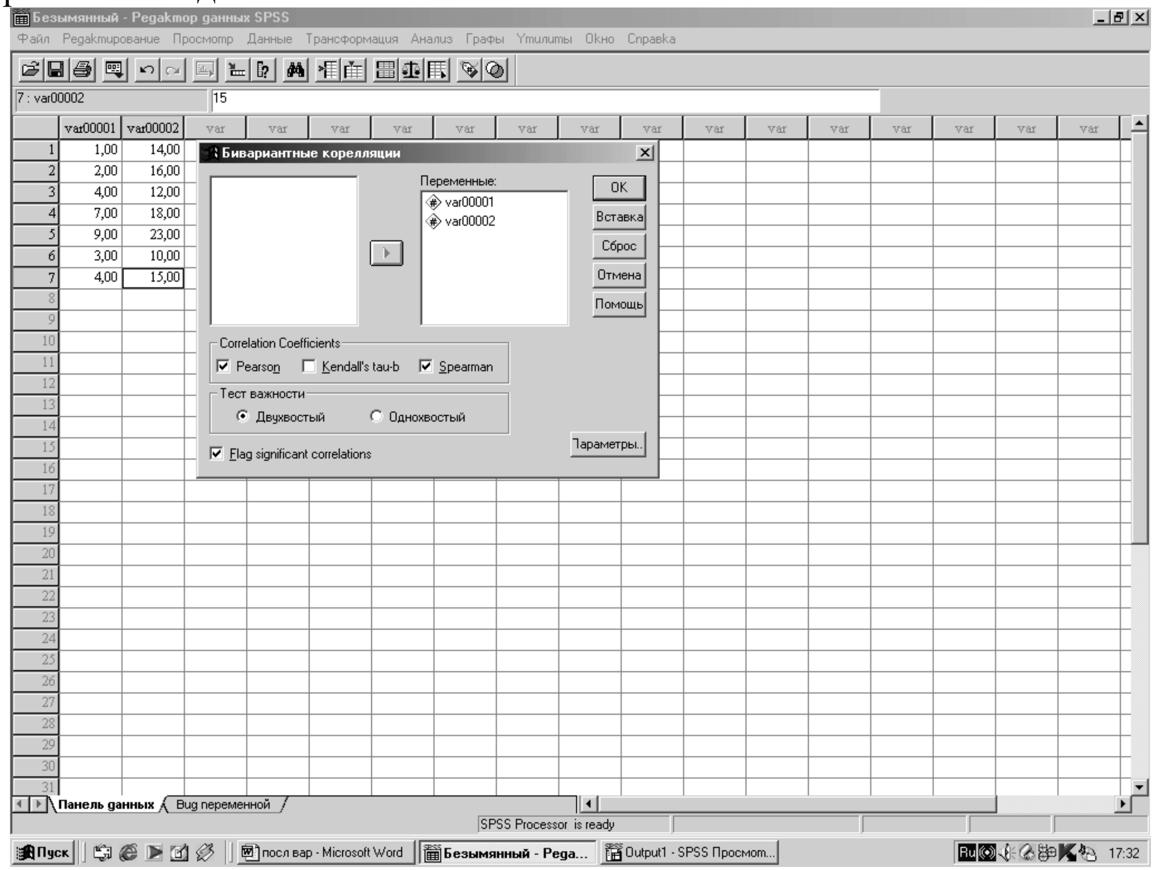
Descriptives

			Statistic	Std. Error Mean
VAR00001	Mean		5,8000	1,62481
	95% Confidence Interval for Mean	Lower Bound	1,2888	
		Upper Bound	10,3112	
	5% Trimmed Mean		5,8889	
	Median		7,0000	
	Variance		13,200	
	Std. Deviation		3,63318	
	Minimum		1,00	
	Maximum		9,00	
	Range		8,00	
	Interquartile Range		7,0000	
	Skewness		-,567	,913
	Kurtosis		-2,231	2,000

Mean — среднее
Median — медиана

Std.deviation — стандартное отклонение
Variance — дисперсия
Range — размах
Std.error mean — стандартная ошибка среднего
Skewness — коэффициент ассиметрии
Kurtosis — коэффициент эксцесса.

Расчет коэффициентов корреляции в пакете SPSS начинается с ввода данных. Допустим, рассчитывается степень связи между признаком 1 (var1) и признаком 2 (var2). Для этого выбирается Анализ → согласование → одновариантность. Затем откроется окно: Бивариантная корреляция. Рассчитаем с помощью данного пакета коэффициенты корреляции: линейный Пирсона и ранговые Спирмена и Кэндэла.



Результаты корреляционного анализа будут представлены в следующем виде:

Correlations

		VAR00001	VAR00002
VAR00001	Pearson Correlation	1	,758
	Sig. (2-tailed)	,	,049
	N	7	7
VAR00002	Pearson Correlation	,758[8]	1
	Sig. (2-tailed)	,049	,
	N	7	7

* Correlation is significant at the 0.05 level (2-tailed).

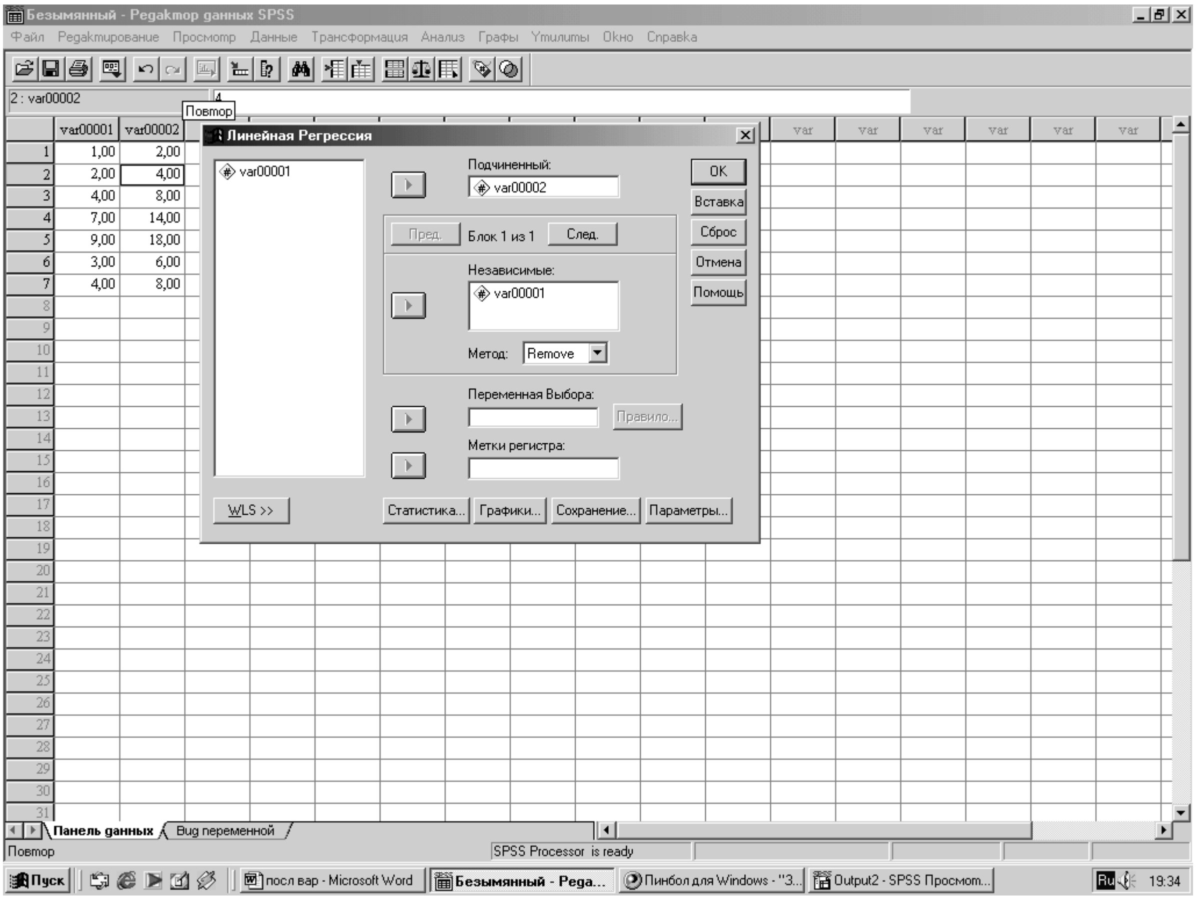
Nonparametric Correlations

Correlations

			VAR00001	VAR00002
Spearman's rho	VAR00001	Correlation Coefficient	1,000	,577
		Sig. (2-tailed)	,	,175
		N	7	7
	VAR00002	Correlation Coefficient	,577	1,000
		Sig. (2-tailed)	,175	,
		N	7	7

В клетках таблицы — три величины. Первая сверху — значение коэффициента корреляции, посередине — его минимальный уровень значимости и снизу — объем выборки. Включенная опция Flag significant correlation (пометка значимых коэффициентов) отмечает звездочкой коэффициенты корреляции, значимо отличные от нуля. На главной диагонали таблиц стоят значения корреляций переменных между самими собой. Интересующие нас значения стоят на побочной диагонали. По значениям коэффициентов корреляции видно, что связь значима. Однако следует заметить, что для достоверности исследований количество значений переменных должно быть больше, чем 7.

Регрессионный анализ осуществляется путем выбора модуля Анализ-регрессия. Рассмотрим наиболее простой случай линейной регрессии. После ввода данных следует определить: какая из переменных будет зависимой, а какая — независимой.



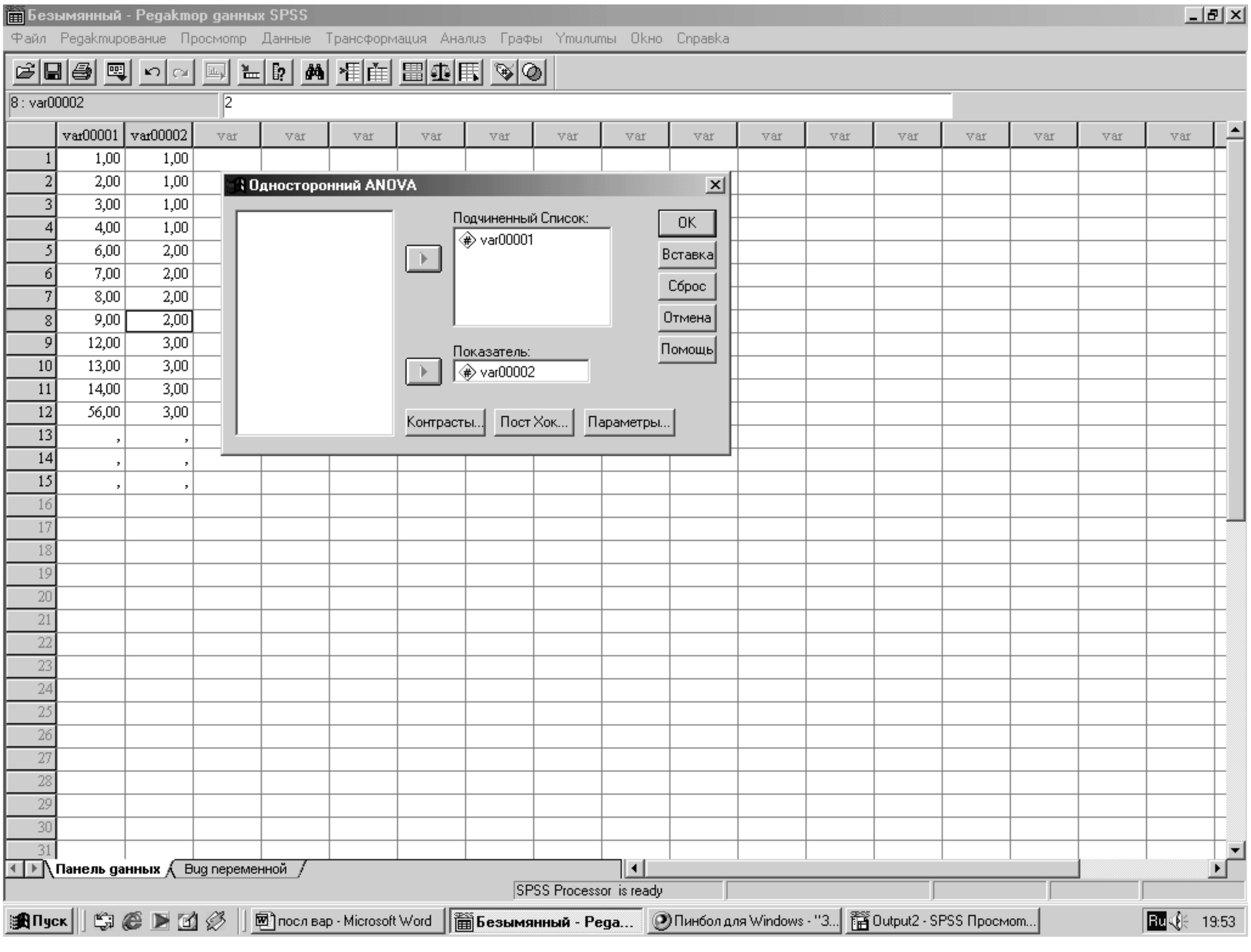
В уравнении линейной регрессии $y=a+bx$, константа — это a, коэффициент b — коэффициент перед переменной var 1.Судя по полученным результатам, уравнение линейной регрессии $y=2x$.

Coefficients

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	,000	,000		,	,
	VAR00001	2,000	,000	1,000	,	,
2	(Constant)	8,571	2,125		4,033	,007

a. Dependent Variable: VAR00002

Для осуществления ***однофакторного дисперсионного анализа*** следует подключить процедуру ANOVA: Анализ → Меры сравнения → Односторонний ANOVA. Перед этим данные следует ввести следующим образом: в первом столбце указываются все имеющиеся данные, а во втором номер группы, к которой они принадлежат (уровни фактора). Тогда подчиненный список — var1, показатель var2. Нулевая гипотеза состоит в том, что влияние фактора отсутствует.



ANOVA
VAR00001

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	987,500	2	493,750	3,177	,090
Within Groups	1398,750	9	155,417		
Total	2386,250	11			

В столбце F выводится значение *F*-критерия, в столбце Sig — его уровень значимости. Если эта величина близка к нулю, то есть основание отвергнуть нулевую гипотезу. В примере, приведенном выше, фактор значительно влияет на распределение признака.

Задания для самостоятельной работы

1. По данным таблицы 2 и таблицы 3 определить, оказал ли фактор погоды (года 1985-1990, 1991-1999) на средний надой молока, яйценоскость кур и годовой настриг шерсти. Определить — есть ли линейная зависимость между этими показателями.
2. Провести простейший статистический анализ данных представленных в таблице 5 Приложения 1.
3. По данным таблицы 7 Приложения 1 определить — оказал ли фактор времени существенное влияние на распределение по типам летних оздоровительных лагерей и на количество отдохнувших в них детей.
4. Провести простейший статистический анализ данных представленных в таблице 8 Приложения 1.
5. Провести, используя данные таблицы 9 Приложения 1, анализ зависимости числа заболевших различными видами болезней между собой. Построить корреляционную матрицу.
6. Провести простейший анализ статистических данных, представленных в таблице 10 Приложения 1.
7. Используя данные, представленные в таблице 11 Приложения 1, определить — присутствует ли связь между численностью мужчин и женщин в районах РТ.
8. Определить, влиял ли фактор времени поступления и выпуска специалистов (1975, 1980, 1985, 1990, 1995, 2000) на количество выбор специальности и количество выпуска по ней.

ПЛАНЫ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

Практическое занятие 1. «Вариационный ряд»
Решение задач: №1, с.10. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.
№446, 447 с.154. Гмурман В.Е. Руководство к решению задач по теории вероятностей и мат. статистике. — М.: «Высшая школа». — 2002.
На с/р: №2, с.10. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.
№448. Гмурман В.Е. Руководство к решению задач по теории вероятностей и мат. статистике. — М.: «Высшая школа». — 2002.

Практическое занятие 2. «Числовые характеристики вариационного ряда»
Решение задач: №1-8 с.12, №1, 2 с.15. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.
№8.10-8.11, С.284. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.
На с/р: №9-15 с.13, №3 с.15. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.
№8.12-8.13 С.284. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.

Практическое занятие 3. «Основные законы распределения»
Решение задач: №4.11-4.17, С.172. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.
На с/р: №4.22-4.23, С.173. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.

Практическое занятие 4. «Выборочный метод и его значение»
Решение задач: №1, 2 с.15. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.
На с/р: №3, с.15. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.
№9.19, с.330. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.

Практическое занятие 5. «Статистическая проверка гипотез»
Решение задач: №1, 2 с.27. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.
На с/р: №3, 4 с.27. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.

Практическое занятие 6. «Корреляционный анализ»
Решение задач: №1, 2, 3, 4 с.36-39. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.

На с/р: № 2.20, 12.21 с.437. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002,
№5, 6 с.40, 41. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.

Практическое занятие 7. «Дисперсионный анализ»
Решение задач: №11.3-11.4, с.390, 391. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.
№1, 2, с.44. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.
На дом: № 11.5, с.391. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.
№3, с.44. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.

Практическое занятие 8. «Применение компьютерных программ при статистической обработке данных (компьютерный практикум)»
Решение задач №1-8, с.55. Кит Ю.В. Прикладная математическая статистика с применением информационных технологий.

Практическое занятие 9. «Линейное и динамическое программирование»
Решение задач со стр.156. Шикин Е.В., Чхартишвили А.Г. Математические методы и модели в управлении: Учебное пособие. — М.: Дело, 2002. — 440с.

САМОСТОЯТЕЛЬНАЯ РАБОТА

Методические указания

Организация самостоятельной работы студентов имеет цель:

- систематизировать и расширить их теоретические знания;
- закрепить практические и организаторские способности;
- научить работать с учебной и научной литературой;
- стимулировать профессиональный рост студентов, воспитывать творческую активность и инициативу.

Самостоятельная работа студентов организуется преподавателями в соответствии с календарным планом изучения дисциплины и предполагает:

- изучение лекционного материала, чтение рекомендуемых литературных источников, решение задач, ответы на контрольные вопросы или тесты и т.д.,
- самостоятельное изучение материала по заданным преподавателем темам;
- написание контрольной работы.

Самостоятельная работа представляет собой дополнительное изучение дисциплины для полного и глубокого усвоения материала на основе анализа учебной, методической и дополнительной литературы.

Самостоятельная работа студентов включает повторение пройденного материала и подготовку к контрольной работе. Вся самостоятельная работа студентов оценивается в течение семестра на фактических занятиях и учитывается при сдаче экзамена.

Повторение пройденного материала осуществляется в процессе выполнения домашнего задания и самостоятельной проработки теоретического материала. Домашняя работа выполняется в соответствии с номерами заданий, представленных в плане практических заданий.

Контрольная работа

Методические рекомендации

При выполнении контрольной работы номер варианта совпадает с номером вашей зачетной книжки.

Работа выполняется в тетради в клетку. Титульный лист оформляется в соответствии с требованиями деканата по оформлению контрольных работ. В графе «работу проверил» следует писать: к.п.н. Кит Ю.В. На титульном листе указывается номер варианта.

Условия, все пояснения и формулы следует писать полностью. В выводах по заданию необходимо указать интерпретацию полученных числовых значений.

Краткое пояснение к первому разделу контрольной работы

Пример построения интервального вариационного ряда:

Пусть измерен некоторый экономический показатель в 30 регионах:
23 29 35 7 11 18 23 30 36 18 11 8 13 20 25 27 14 30 20 20 24 19 21 26 22 16 26 25 33 27

Расставим экспериментальные данные в возрастающем порядке:
7 8 11 11 13 14 16 18 18 19 20 20 20 21 22 23 23 24 25 25 26 26 27 27 29 30 30 33 35 36

По таблице 1 определяем число классов

Таблица 1

Объем выборки *n*	Число классов *K*
6-11	4
12-22	5
23-46	6
47-93	7
94-187	8
188-377	9
378-755	10
756-1515	11

Для n=30 число классов K=6. Найдем минимальное и максимальное значения вариант: x_{min}=7, x_{max}=36.

Определим вариационный размах R=x_{min}-x_{max}=36-7=29.

Определим величину классового интервала: $\Delta = \frac{R}{K-1} = \frac{29}{51} = 5{,}8$.

$\tilde{O}_{\acute{i}1} = x_{\min} - \frac{1}{2}\Delta = 7 - 2{,}9 = 4{,}1$; $X_{\hat{a}1} = x_{\min} + \frac{1}{2}\Delta = 7 + 2{,}9 = 9{,}9$

Обобщим полученные данные в таблице:

Таблица 2

Номера классов	Классовые интервалы	Серединные значения классов	Частоты	Накопленные частоты
1	4,1-9,9	7	2	2
2	9,9-15,7	12,8	4	6
3	15,7-21,5	18,6	8	14
4	21,5-27,3	24.4	10	24
5	27,3-33,1	30,2	4	28
6	33,1-38,9	36	2	30

График 1 (см. рис. 1 ниже), называемый гистограммой, получается, если в прямоугольной системе координат отложить по оси абсцисс границы классов, а по оси ординат их частоты.

Гистограмма на рис.1 построена на основании данных таблицы 2.

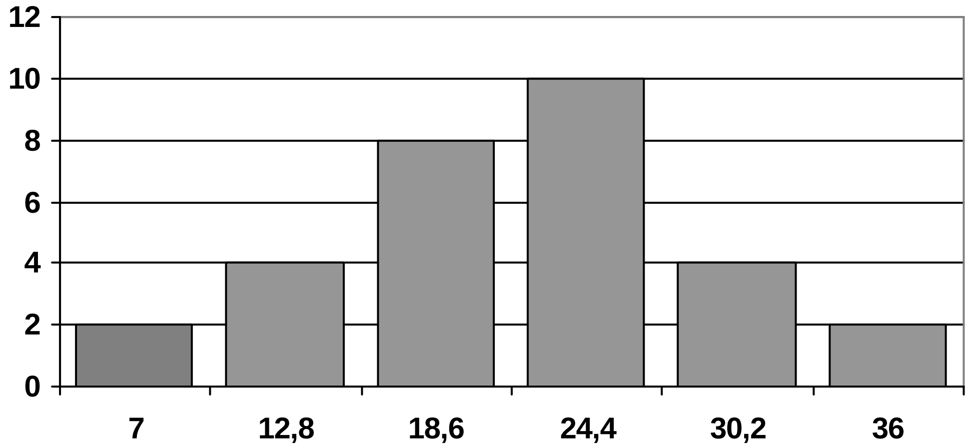


Рис.1. Гистограмма распределения.

Если серединные точки вершин прямоугольников гистограммы соединить между собой, получится график дискретного варьирования, называемый полигоном распределения.

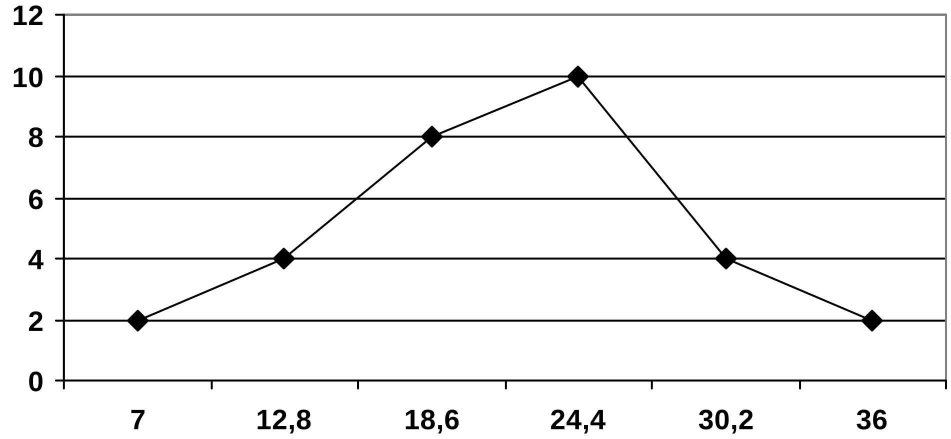


Рис.2. Полигон распределения.

Мода распределения — это наиболее часто встречающееся значение ряда.

Среднее арифметическое распределения находится по формуле

$$x_{cp}= (x_1+x_2+x_3+ \ldots +x_n)/n$$

Дисперсия распределения находится по формуле:

$$D = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_{\text{ñç}})^2$$

Стандартное отклонение $S = \sqrt{D}$

Краткое пояснение ко второму разделу контрольной работы

Пример расчета рангового коэффициента корреляции.

Пусть при исследовании десяти регионов получены следующие показатели X и Y.

Выясним, существует ли между ними связь. Для этого подсчитаем ранговый коэффициент корреляции, дадим его графическую интерпретацию.

Таблица 3

X	Y
175	5
176	7
179	8
180	9
184	3
191	1
181	10
186	6
192	4
185	2

Найдем ранг (порядковый номер по убыванию) каждого из значений x и y: R_x и R_y, затем найдем разности соответствующих рангов d возведем их в квадрат, получим ряд значений d^2. Просуммируем его и подставим в формулу:

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)}.$$

Таблица 4

№	X	Y	R_x	R_y	d	d^2
1	175	5	10	9	1	1
2	176	7	9	7	2	4
3	179	8	8	6	2	4
4	180	9	7	5	2	4
5	184	3	5	10	5	25
6	191	6	2	8	6	36
7	181	10	6	4	2	4
8	186	11	3	3	0	0
9	192	12	1	2	1	1
10	185	13	4	1	3	9
Сумма:						87

В нашем случае: $r_S = 1 - \frac{6*87}{10(10^2-1)} = 0{,}5$.Если значения одинаковые, то приписывается промежуточный средний ранг, например, 6,5.

Для графической интерпретации по оси x откладываются значения признака x, по оси y — значения признака y.

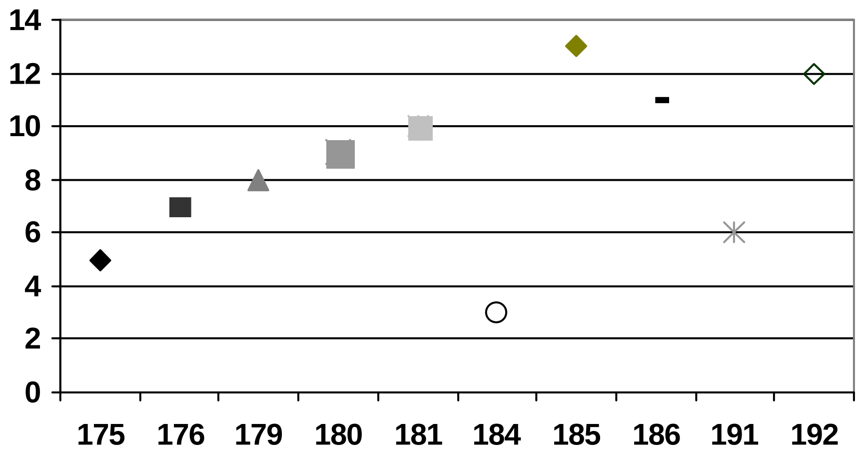


Рис.3. Графическая интерпретация коэффициента корреляции.

По значению коэффициента корреляции и графической интерпретации можем сказать, что между признаками *x* и *y* есть средняя прямая связь.

Контрольная работа

1. При обследовании группы предприятий (10 шт.) были получены следующие данные показателя прибыли (в млн. руб.).

1. Проанализируйте данный ряд (расположите совокупность в порядке возрастания признака).
2. Постройте интервальный вариационный ряд
3. Постройте гистограмму и полигон распределения
4. Найдите моду, медиану и среднее арифметическое данного распределения.
5. Определите дисперсию, среднеквадратическое отклонение и коэффициент вариации данного распределения.

T
13
24
6
4
14
5
5
9
5
7

2. Проводилось испытание 8 сортов озимой пшеницы. Каждый сорт высевался на 6 делянках одинаковой площади. При 5% уровне значимости проверить гипотезу о существенности различий в средней урожайности двух сортов озимой пшеницы (номера сортов даются студенту преподавателем). Урожайность озимой пшеницы, ц/га:

Повторения	Сорт							
	1	2	3	4	5	6	7	8
I	45	51	60	49	63	44	55	60
II	44	50	62	52	61	40	53	55
III	46	56	61	45	62	41	51	53
IV	44	52	56	48	56	43	58	57
V	47	54	61	47	62	45	54	54
VI	45	52	59	46	61	41	53	56

3. Дана таблица результатов наблюдений:

X_i	2	4	6	8	10	12	14
Y_i	3,5	6,0	7,0	6,0	7,5	8,5	10

Найти выборочный коэффициент корреляции и определить его значимость. Рассчитать параметры линейного уравнения регрессии.

4. Доказывает ли опыт влияние различных доз удобрений на урожайность озимой пшеницы?

Урожайность озимой пшеницы с 1 га, ц

Повторения	Дозы удобрений							
	1	2	3	4	5	6	7	8
1	39	41	42	47	55	41	43	39
2	39	40	40	45	50	42	50	40
3	37	39	39	43	50	44	47	37
4	41	40	42	42	48	44	49	43
5	36	38	40	40	44	43	50	39
6	38	38	39	41	43	40	46	41
7	43	41	43	45	51	42	44	40
8	40	42	45	50	53	47	49	41

КОНТРОЛЬ ЗНАНИЙ СТУДЕНТОВ
Контрольный тест для промежуточной аттестации студентов

Вариант 1

1. Поставьте в соответствие каждое из следующих измерений к одному из видов шкал:

а) числа, кодирующие темпераменты	1) шкала наименований
б) академический ранг (ассистент, доцент, профессор)	2) шкала порядка
в) метрическая система расстояний	3) интервальная шкала
г) телефонные номера	4) шкала отношений

а) — ;б) — ; в) — ; г) — .

2. Перечислите меры среднего (центральной тенденции) и способы их вычисления.

3. Перечислите меры разброса (изменчивости) и способы их вычисления.

4. Какой из показателей наиболее чувствителен к наличию крайних значений: 1) мода, 2) медиана, 3) среднее арифметическое.

5. Большее стандартное отклонение показателей в одной совокупности в отличие от другой свидетельствует о:
1) связи значений, 2) меньшем разбросе значений, 3) большем разбросе значений.

- Перечислите типы выборок и укажите их отличия.
- Перечислите критерии качества оценки.
- Если для любого фиксированного числа наблюдений выполняется равенство: М($\widehat{\Theta}_n$)=Θ, где $\widehat{\Theta}_n$ — оценка генеральной характеристики Θ, то оценка $\widehat{\Theta}_n$ называется:

1) несмещенной, 2) эффективной, 3) состоятельной.

6. Какая выборка называется репрезентативной?

7. Оценка тем точнее,
1) чем больше разность $|\widehat{\Theta}_n - \Theta|$
2) чем меньше разность $|\widehat{\Theta}_n - \Theta|$
3) чем выразительнее $|\widehat{\Theta}_n - \Theta|$.

8. Какое из утверждений верное (где t-нормированное отклонение — «коэффициент доверия», зависящий от вероятности, с которой гарантируется предельная ошибка выборки).
1) Предельную ошибку выборки можно подсчитать по формуле: $\Delta=t/m_x$,
2) Чем меньше t, тем больше вероятность, с которой гарантируется предельная ошибка выборки.
3) Чем больше t, тем больше вероятность, с которой гарантируется предельная ошибка выборки.

9. Какие параметры должны быть известны, чтобы определить необходимую численность выборки?

10. Распределение проверяется на близость к нормальному, потому что:
1) может быть аномальным,
2) при большом объеме выборки,
3) т.к. большинство статистических формул применимо только к распределениям близким к нормальным.

11. Чем отличаются параметрические методы от непараметрических?

12. Какой критерий является многофункциональным статистическим критерием и предназначен для сопоставления уровней исследуемого признака, сдвигов в значениях исследуемого признака и сравнения распределений?
1) Критерий φ^* Фишера
2) T-критерий Вилкоксона
3) Критерий U Манна-Уитни

13. Если для некоторой совокупности значений x и y коэффициент корреляции равен 0,9, то какое из утверждений будет ***неверным***:
1) большим значениям x соответствуют большие значения y,
2) большим значениям x соответствуют меньшие значения y,
3) меньшим значениям x соответствуют меньшие значения y.

14. Выберите верное утверждение:
1) Линейный коэффициент корреляции применим для любого распределения, а ранговый только для малых выборок,
2) Ранговый коэффициент корреляции применим только для распределений, подчиняющихся нормальному закону, а линейный — только для малых выборок,
3) Порядковую (ранговую) корреляцию применяют при небольшом объеме выборки и возможности ранжирования, а линейный коэффициент корреляции — для распределений, подчиняющихся нормальному закону.

15. Что показывает теоретическая линия регрессии?

16. Укажите верное утверждение:
1) Факторная дисперсия $S^2_{факт}$. вызвана действием на у случайных причин, а остаточная $S^2_{ост}$ целиком обусловлена влиянием фактора x.
2) И факторная дисперсия $S^2_{факт}$. и $S^2_{ост}$ вызвана действием на y фактора x.
3) Факторная дисперсия $S^2_{факт}$. вызвана действием на y фактора x, а остаточная ($S^2_{ост}$) целиком обусловлена случайными причинами.

17. Возможностями компьютерной статистической обработки данных обладает следующее приложение WINDOWS:
1) WORD,
2) EXCEL,
3) ACCESS.

18. Чтобы автоматизировать дисперсионный анализ следует выполнить следующую последовательность действий, активизируя строки меню:
1) сервис-анализ данных — дисперсионный анализ; сервис-настройки — пакет анализа,
2) сервис-настройки — пакет анализа; сервис-анализ данных — дисперсионный анализ,
3) сервис-настройки — дисперсионный анализ.

Вариант 2

1. Какие параметры должны быть известны, чтобы определить необходимую численность выборки?

2. Распределение проверяется на близость к нормальному, потому что:
1) может быть аномальным,
2) при большом объеме выборки,
3) т.к. большинство статистических формул применимо только к распределениям близким к нормальным.

3. Чем отличаются параметрические методы от непараметрических?

4. Какой критерий является многофункциональным статистическим критерием и предназначен для сопоставления уровней исследуемого признака, сдвигов в значениях исследуемого признака и сравнения распределений.
1) Критерий φ* Фишера.
2) *T*-критерий Вилкоксона.
3) Критерий *U* Манна-Уитни.

5. Если для некоторой совокупности значений *x* и *y* коэффициент корреляции равен 0,9, то какое из утверждений будет ***неверным***:
1) большим значениям *x* соответствуют большие значения *y*,
2) большим значениям *x* соответствуют меньшие значения *y*,
3) меньшим значениям *x* соответствуют меньшие значения *y*.

6. Выберите верное утверждение:
1) Линейный коэффициент корреляции применим для любого распределения, а ранговый только для малых выборок,
2) Ранговый коэффициент корреляции применим только для распределений, подчиняющихся нормальному закону, а линейный — только для малых выборок,
3) Порядковую (ранговую) корреляцию применяют при небольшом объеме выборки и возможности ранжирования, а линейный коэффициент корреляции — для распределений, подчиняющихся нормальному закону.

7. Что показывает теоретическая линия регрессии?

8. Укажите верное утверждение:
1) Факторная дисперсия $S^2_{факт}$. вызвана действием на у случайных причин, а остаточная $S^2_{ост}$ целиком обусловлена влиянием фактора *x*.
2) И факторная дисперсия $S^2_{факт}$. и $S^2_{ост}$ вызвана действием на *y* фактора *x*.
3) Факторная дисперсия $S^2_{факт}$. вызвана действием на *y* фактора *x*, а остаточная ($S^2_{ост}$) целиком обусловлена случайными причинами.

9. Возможностями компьютерной статистической обработки данных обладает следующее приложение WINDOWS:
1) WORD,
2) EXCEL,
3) ACCESS.

10. Чтобы автоматизировать дисперсионный анализ следует выполнить следующую последовательность действий, активизируя строки меню:
1) сервис-анализ данных — дисперсионный анализ; сервис-настройки — пакет анализа,

2) сервис-настройки — пакет анализа; сервис-анализ данных — дисперсионный анализ,
3) сервис-настройки — дисперсионный анализ;

11. Поставьте в соответствие каждое из следующих измерений к одному из видов шкал:

а) числа, кодирующие темпераменты	1) шкала наименований
б) академический ранг (ассистент, доцент, профессор)	2) шкала порядка
в) метрическая система расстояний	3) интервальная шкала
г) телефонные номера	4) шкала отношений

а) — ,б) — , в) — , г) — .

12. Перечислите меры среднего (центральной тенденции) и способы их вычисления.

13. Перечислите меры разброса (изменчивости) и способы их вычисления.

14. Какой из показателей наиболее чувствителен к наличию крайних значений:
1) мода, 2) медиана, 3) среднее арифметическое.

15. Большее стандартное отклонение показателей в одной совокупности в отличие от другой свидетельствует о:
1) связи значений, 2) меньшем разбросе значений, 3) большем разбросе значений.

16. Перечислите типы выборок и укажите их отличия.

17. Перечислите критерии качества оценки.

18. Если для любого фиксированного числа наблюдений выполняется равенство: $M(\hat{\Theta}_n)=\Theta$, где $\hat{\Theta}_n$ — оценка генеральной характеристики Θ, то оценка $\hat{\Theta}_n$ называется
1) несмещенной, 2) эффективной, 3) состоятельной.

19. Какая выборка называется репрезентативной?

20. Оценка тем точнее,
1) чем больше разность $|\hat{\Theta}_n - \Theta|$
2) чем меньше разность $|\hat{\Theta}_n - \Theta|$
3) чем выразительнее $|\hat{\Theta}_n - \Theta|$.

21. Какое из утверждений верное (где *t*-нормированное отклонение, «коэффициент доверия», зависящий от вероятности, с которой гарантируется предельная ошибка выборки).
1) Предельную ошибку выборки можно подсчитать по формуле: $\Delta=t/m_x$,
2) Чем меньше t, тем больше вероятность, с которой гарантируется предельная ошибка выборки.
3) Чем больше t, тем больше вероятность, с которой гарантируется предельная ошибка выборки.

Вариант 3

1. Если для любого фиксированного числа наблюдений выполняется равенство: М($\hat{\Theta}_n$)=Θ, где $\hat{\Theta}_n$ — оценка генеральной характеристики Θ, то оценка $\hat{\Theta}_n$ называется

1) несмещенной, 2) эффективной 3) состоятельной.

2. Какая выборка называется репрезентативной?

3. Оценка тем точнее,

1) чем больше разность $|\hat{\Theta}_n - \Theta|$
2) чем меньше разность $|\hat{\Theta}_n - \Theta|$
3) чем выразительнее $|\hat{\Theta}_n - \Theta|$.

4. Какое из утверждений верное (где t-нормированное отклонение, «коэффициент доверия», зависящий от вероятности, с которой гарантируется предельная ошибка выборки).

1) Предельную ошибку выборки можно подсчитать по формуле: $\Delta=t/m_x$,
2) Чем меньше t, тем больше вероятность, с которой гарантируется предельная ошибка выборки.
3) Чем больше t, тем больше вероятность, с которой гарантируется предельная ошибка выборки.

5. Какие параметры должны быть известны, чтобы определить необходимую численность выборки?

6. Распределение проверяется на близость к нормальному, потому что:

1) может быть аномальным,
2) при большом объеме выборки,
3) т.к. большинство статистических формул применимо только к распределениям близким к нормальным.

7. Чем отличаются параметрические методы от непараметрических?

8. Какой критерий является многофункциональным статистическим критерием и предназначен для сопоставления уровней исследуемого признака, сдвигов в значениях исследуемого признака и сравнения распределений

1) Критерий φ* Фишера,
2) T-критерий Вилкоксона,
3) Критерий U Манна-Уитни.

9. Если для некоторой совокупности значений x и y коэффициент корреляции равен 0,9, то какое из утверждений будет ***неверным***:

1) большим значениям x соответствуют большие значения y,
2) большим значениям x соответствуют меньшие значения y,
3) меньшим значениям x соответствуют меньшие значения y.

10. Выберите верное утверждение:

1) Линейный коэффициент корреляции применим для любого распределения, а ранговый только для малых выборок,
2) Ранговый коэффициент корреляции применим только для распределений, подчиняющихся нормальному закону, а линейный — только для малых выборок,

3) Порядковую (ранговую) корреляцию применяют при небольшом объеме выборки и возможности ранжирования, а линейный коэффициент корреляции — для распределений, подчиняющихся нормальному закону.

11. Что показывает теоретическая линия регрессии?

12. Поставьте в соответствие каждое из следующих измерений к одному из видов шкал:

а) числа, кодирующие темпераменты	1) шкала наименований
б) академический ранг (ассистент, доцент, профессор)	2) шкала порядка
в) метрическая система расстояний	3) интервальная шкала
г) телефонные номера	4) шкала отношений

а) — , б) — , в) — , г) — .

13. Перечислите меры среднего (центральной тенденции) и способы их вычисления.

14. Перечислите меры разброса (изменчивости) и способы их вычисления.

15. Какой из показателей наиболее чувствителен к наличию крайних значений:

1) мода, 2) медиана, 3) среднее арифметическое.

16. Большее стандартное отклонение показателей в одной совокупности в отличие от другой свидетельствует о:

1) связи значений, 2) меньшем разбросе значений, 3) большем разбросе значений.

17. Перечислите типы выборок и укажите их отличия.

18. Перечислите критерии качества оценки.

19. Укажите верное утверждение:

1) Факторная дисперсия $S^2_{факт.}$ вызвана действием на у случайных причин, а остаточная $S^2_{ост}$ целиком обусловлена влиянием фактора x.
2) И факторная дисперсия $S^2_{факт.}$ и $S^2_{ост}$ вызвана действием на y фактора x
3) Факторная дисперсия $S^2_{факт.}$ вызвана действием на y фактора x, а остаточная ($S^2_{ост}$) целиком обусловлена случайными причинами.

20. Возможностями компьютерной статистической обработки данных обладает следующее приложение WINDOWS:

1) WORD,
2) EXCEL,
3) ACCESS.

21. Чтобы автоматизировать дисперсионный анализ следует выполнить следующую последовательность действий, активизируя строки меню:

1) сервис-анализ данных — дисперсионный анализ; сервис-настройки — пакет анализа,
2) сервис-настройки — пакет анализа; сервис-анализ данных — дисперсионный анализ,
3) сервис-настройки — дисперсионный анализ.

Вопросы для подготовки к зачету

1. Понятие о вариационном ряде. Частоты и частости.
2. Виды вариации. Дискретные и интервальные вариационные ряды.
3. Границы интервалов и величина интервала.
4. Плотность распределения.
5. Накопленные частоты.
6. Графические методы изображения вариационного ряда: полигон, гистограмма, кумулята, огива.
7. Виды шкал.
8. Средняя арифметическая и ее свойства. Квантили. Мода. Медиана.
9. Показатели разброса признака: вариационный размах, среднее линейное отклонение, дисперсия, среднее квадратическое отклонение, коэффициент вариации.
10. Биномиальный закон распределения.
11. Равномерный закон распределения.
12. Распределение Пуассона.
13. Показательный закон распределения
14. Нормальное распределение.
15. Стандартное (нормированное) нормальное распределение.
16. Вероятность заданного отклонения нормально распределенной случайной величины от своего математического ожидания. Правило трех сигм.
17. Локальная и интегральная теоремы Лапласа.
18. Понятие о законе больших чисел.
19. Понятие выборочного метода. Статистическое распределение выборки. Генеральная и выборочная совокупность.
20. Способы отбора: собственно-случайный (повторный и бесповторный), механический, типический, серийный.
21. Ошибки регистрации и репрезентативности (систематические и случайные).
22. Статистические оценки параметров распределения (сущность теории оценивания).
23. Интервальные оценки. Точность оценки. Доверительная вероятность.
24. Необходимая численность выборки.
25. Законы распределения, применяемые в математической статистике: Стьюдента, Хи-квадрат, Фишера.
26. Статистические гипотезы, их виды. Нулевая и конкурирующая гипотезы.
27. Ошибки I и II рода. Уровень значимости.
28. Параметрические и непараметрические гипотезы.
29. Выявление различий в уровне исследуемого признака. U-критерий Манна-Уитни.
30. Оценка достоверности сдвига в значениях исследуемого признака. G-критерий знаков. Критерий χ^2 Фридмана, T-критерий Вилксона.
31. Выявление различий в распределении признака. χ^2-критерий Пирсона. λ-критерий Колмогорова-Смирнова.

32. Многофункциональные статистические критерии. Критерий φ^* — угловое преобразование Фишера.
33. Корреляционная связь и ее статистическое изучение.
34. Линейная парная регрессия.
35. Коэффициент корреляции. Линейный и ранговый коэффициенты корреляции.
36. Проверка гипотезы о значимости выборочного коэффициента корреляции.
37. Построение доверительного интервала для коэффициента корреляции.
38. Нелинейная регрессия. Параболическая и гиперболическая зависимости между зависимыми случайными величинами.
39. Множественная корреляция.
40. Понятие дисперсионного анализа.
41. Подготовка данных к дисперсионному анализу.
42. Однофакторный дисперсионный анализ для несвязанных и связанных выборок.
43. Двухфакторный дисперсионный анализ для несвязанных и связанных выборок.
44. Возможности обработки экспериментальных данных в электронных таблиц EXCEL.

ЛИТЕРАТУРА

Основная:

1. Кремер Н.Ш. Теория вероятностей и математическая статистика. — М.: ЮНИТИ, 2002.
2. Горелова Г.В., Кацко И.А. Теория вероятностей и мат. статистика в примерах и задачах с применением EXCEL/ Учебное пособие для вузов. — Ростов н/Д: Феникс, 2006.
3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. — М.: Высшая школа, 2002.

Дополнительная:

1. Бочаров П.П., Печенкин А.В. Теория вероятностей и математическая статистика. — М.: Гардарика, 1998.
2. Калинина В.Н., Панкин В.Н. Математическая статистика. — М.: Высшая школа, 1998.
3. Ковалев В.А., Калинина В.Н. Теория вероятностей и математическая статистика. — М.: ИНФРА-М, 1999
4. Колемаев В.А., Калинина В.Н. Теория вероятностей и математическая статистика. — М.: ИНФРА-М, 1997.
5. Боровиков В.П.: «Популярное введение в программу статистика». М. — Компьютер Пресс. 1998.
6. Общий курс высшей математики для экономистов: Учебник под ред. В.И. Ермакова. — М.:ИНФА-М, 2000.
7. Сидоренко Е.В. Методы математической обработки в психологии. — СПб: СПЦ, 1996. 349 с.
8. Справочник по прикладной статистике. / Под ред. Э. Ллойда и У. Ледермана. Том 2. — М.: Финансы и статистика, 1990. — 526с.
9. Методы математической статистики в психологии / Вагапов Р.Г., Шевцов М.Н. — Казань: КГУ, 2000.

Учебное издание

КИТ Юлия Владимировна

МАТЕМАТИКА

Часть II
ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УЧЕБНОЕ ПОСОБИЕ
для студентов всех специальностей и форм обучения

Корректор *Шамонова А.М.*
Технический редактор, оформление *Александрова М.Н.*

Формат 60*90 1/16. Бумага газетная. Гарнитура New Roman. Печать офсет.
Усл. печ. л. 7,5. Уч.-изд. л. 4,30. Тираж 400 экз. Заказ №

Издательство «Юниверсум».
420012, г. Казань, ул. Достоевского, д. 10.
Отпечатано в полном соответствии с качеством предоставленных материалов
в типографии ОАО «Щербинская типография».
117623, г. Москва, ул. Типографская, д. 10. Тел. 659-2327

ДЛЯ ЗАМЕТОК

ДЛЯ ЗАМЕТОК

ДЛЯ ЗАМЕТОК