

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Магнитогорский государственный технический университет им. Г.И. Носова»

Многопрофильный колледж

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ
ДЛЯ ПРАКТИЧЕСКИХ ЗАНЯТИЙ
УЧЕБНОЙ ДИСЦИПЛИНЫ**

ОП.03. ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

**для обучающихся специальности
09.02.13 Интеграция решений с применением технологий искусственного интеллекта**

СОДЕРЖАНИЕ

1 ВВЕДЕНИЕ	3
2 МЕТОДИЧЕСКИЕ УКАЗАНИЯ	4
Практическое занятие №1	4
Практическое занятие № 2	6
Практическое занятие № 3	8
Практическое занятие № 4	11
Практическое занятие № 5	14
Практическое занятие № 6	16
Практическое занятие № 7	18
Практическое занятие № 8	20
Практическое занятие № 9	23
Практическое занятие № 10	26
Практическое занятие № 11	29
Практическое занятие № 12	33
Практическое занятие № 13	35
Практическое занятие № 14	37
Практическое занятие № 15	39
Практическое занятие № 16	40

1 ВВЕДЕНИЕ

Важную часть теоретической и профессиональной практической подготовки обучающихся составляют практические занятия.

Состав и содержание практических занятий направлены на реализацию Федерального государственного образовательного стандарта среднего профессионального образования.

Ведущей дидактической целью практических занятий является формирование профессиональных практических умений (умений выполнять определенные действия, операции, необходимые в последующем в профессиональной деятельности) или учебных практических умений (умений решать задачи по математике), необходимых в последующей учебной деятельности.

В соответствии с рабочей программой учебной дисциплины «Теория вероятностей и математическая статистика» предусмотрено проведение практических занятий.

В результате их выполнения, обучающийся должен уметь:

Уд 1. Применять стандартные методы и модели к решению вероятностных и статистических задач.

Уд 2. Использовать расчетные формулы, таблицы, графики при решении статистических задач.

Уд 3. Применять современные пакеты прикладных программ многомерного статистического анализа.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Содержание практических занятий ориентировано на подготовку обучающихся к освоению профессионального модуля программы подготовки специалистов среднего звена по специальности и овладению **профессиональными компетенциями:**

ПК 3.2. Формировать сценарии обучения готовых моделей искусственного интеллекта.

А также формированию общих компетенций:

ОК 01. Выбирать способы решения задач профессиональной деятельности применительно к различным контекстам.

ОК 02. Использовать современные средства поиска, анализа и интерпретации информации, и информационные технологии для выполнения задач профессиональной деятельности.

Практические занятия проводятся в рамках соответствующей темы, после освоения дидактических единиц, которые обеспечивают наличие знаний, необходимых для ее выполнения.

2 МЕТОДИЧЕСКИЕ УКАЗАНИЯ

Тема 1.1. Основные понятия теории вероятностей

Практическое занятие №1

Построение пространства элементарных исходов для заданных экспериментов

Цель: сформировать умение решать задачи с использованием формул комбинаторики: перестановок, размещений и сочетаний.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:

Решить задачи.

1. Сколько перестановок можно сделать из букв слова «Миссисипи». Ответ: 2520
2. Имеется пять различных стульев и семь рулонов обивочной ткани различных цветов. Сколькими способами можно осуществить обивку стульев. Ответ: 16807
3. На памятные сувениры в «Поле Чудес» спонсоры предлагают кофеварки, утюги, телефонные аппараты, духи. Сколькими способами 9 участников игры могут получить эти сувениры? Сколькими способами могут быть выбраны 9 предметов для участников игры? Ответ: 4^9 , 220
4. Сколькими способами можно расставить на шахматной доске 8 ладей так, чтобы ни одна из них не могла бить другую? Ответ: 40320
5. Сколько может быть случаев выбора 2 карандашей и 3 ручек из пяти различных карандашей и шести различных ручек? Ответ: 200
6. В течение 30 дней сентября было 12 дождливых дней, 8 ветреных, 4 холодных, 5 дождливых и ветреных, 3 дождливых и холодных, а один день был и дождливым, и ветреным, и холодным. В течение скольких дней в сентябре стояла хорошая погода. Ответ: 15
7. На ферме есть 20 овец и 24 свиньи. Сколькими способами можно выбрать одну овцу и одну свинью? Если такой выбор уже сделан, сколькими способами можно сделать его еще раз? Ответ: 480, 437
8. Сколькими способами можно выбрать гласную и согласную буквы из слова «здание»? Ответ: 9
9. Сколько существует четных пятизначных чисел, начинающихся нечетной цифрой? Ответ: 25000
10. В книжный магазин поступили романы Ф. Купера «Прерия», «Зверобой», «Шпион», «Пионеры», «Следопыт» по одинаковой цене. Сколькими способами библиотека может закупить 17 книг на выбранный чек? Ответ: 2985

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Размещения без повторений

Задача 1. Сколько можно составить телефонных номеров из 6 цифр каждый, так чтобы все цифры были различны?

Решение. Это пример задачи на размещение без повторений. Размещаются здесь 10 цифр по 6. А варианты, при которых одинаковые цифры стоят в разном порядке считаются разными. Значит, ответ на задачу будет:

$$A_{10}^6 = \frac{10!}{(10-6)!} = 151200$$

Задача 2. Сколькими способами 4 юноши могут пригласить четырех из шести девушек на танец?

Решение. Два юноши не могут одновременно пригласить одну и ту же девушку. И варианты, при которых одни и те же девушки танцуют с разными юношами, считаются разными, поэтому:

$$A_6^4 = \frac{6!}{(6-4)!} = \frac{720}{2} = 360$$

Возможно 360 вариантов.

Перестановки без повторений

Задача 3. Сколько различных шестизначных чисел можно составить из цифр 0, 1, 2, 3, 4, 5, если цифры в числе не повторяются?

Решение.

1) Найдем количество всех перестановок из этих цифр: $P_6=6!=720$.

2) 0 не может стоять впереди числа, поэтому от этого числа необходимо отнять количество перестановок, при котором 0 стоит впереди. А это $P_5=5!=120$.

$$P_6 - P_5 = 720 - 120 = 600.$$

Сочетания без повторений

Задача 4. Сколько трехкнопочных комбинаций существует на кодовом замке (все три кнопки нажимаются одновременно), если на нем всего 10 цифр.

Решение. Так как кнопки нажимаются одновременно, то выбор этих трех кнопок – сочетание. Отсюда возможно $C_{10}^3 = \frac{10!}{(10-3)! \cdot 3!} = \frac{8 \cdot 9 \cdot 10}{6} = 120$ вариантов.

Задача 5. У одного человека 7 книг по математике, а у второго – 9. Сколькими способами они могут обменять друг у друга две книги на две книги.

Решение. Так как порядок следования книг не имеет значения, то выбор двух книг – сочетание. Первый человек может выбрать 2 книги $C_7^2 = \frac{7!}{(7-2)! \cdot 2!} = \frac{6 \cdot 7}{2} = 21$ способами.

Второй человек может выбрать 2 книги $C_9^2 = \frac{9!}{(9-2)! \cdot 2!} = \frac{8 \cdot 9}{2} = 36$ способами. Значит всего по правилу произведения возможно $21 \cdot 36 = 756$ вариантов.

Задача 6. При игре в домино 4 игрока делят поровну 28 костей. Сколькими способами они могут это сделать?

Решение. Первый игрок делает выбор из 28 костей. Вторым из $28-7=21$ кости, третий из 14, а четвертый игрок забирает оставшиеся кости. Следовательно, возможно $C_{28}^7 \cdot C_{21}^7 \cdot C_{14}^7$.

Размещения и сочетания с повторениями

Задача 7. Сколько трехзначных чисел можно составить из цифр 1, 2, 3, 4, 5?

Решение. Так как порядок цифр в числе существенен, цифры могут повторяться, то это будут размещения с повторениями из пяти элементов по три, а их число равно $\overline{A}_5^3 = 5^3 = 125$.

Задача 8. В кондитерском магазине продавались 4 сорта пирожных: эклеры, песочные, наполеоны и слоеные. Сколькими способами можно купить 7 пирожных.

Решение. Покупка не зависит от того, в каком порядке укладывают купленные пирожные в коробку. Покупки будут различными, если они отличаются количеством купленных

пирожных хотя бы одного сорта. Следовательно, количество различных покупок равно числу сочетаний четырех видов пирожных по семь: $\overline{C}_4^7 = \frac{(7+4-1)!}{7!(4-1)!} = \frac{10!}{7! \cdot 3!} = 120$.

Задача 9. Обезьяну посадили за пишущую машинку с 45 клавишами, определить число попыток, необходимых для того, чтобы она наверняка напечатала первую строку романа Л.Н. Толстого «Анна Каренина», если строка содержит 52 знака и повторений не будет?

Решение. Порядок букв имеет значение. Буквы могут повторяться. Значит, всего есть $\overline{A}_{45}^{52} = 45^{52}$ вариантов.

Перестановки с повторениями

Задача 10. Сколькими способами можно переставить буквы слова «ананас»?

Решение. Всего букв 6. Из них одинаковы: $n_1(a)=3$, $n_2(n)=2$, $n_3(s)=1$. Следовательно, число различных перестановок равно $P_6(3,2,1) = \frac{6!}{3! \cdot 2! \cdot 1!} = 60$.

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 1.1. Основные понятия теории вероятностей

Практическое занятие № 2

Вычисление вероятностей событий на основе классического определения вероятности

Цель: сформировать умение решать задачи на вычисление вероятности с использованием формул комбинаторики.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:

Решить задачи.

1. Какова вероятность того, что наудачу выбранное двузначное число не содержит ни одной двойки?

2. Отряд учащихся из 25 человек участвует в военизированной игре. В отряде 5 следопытов и 4 связиста. В разведку надо направить четырех человек. Какова вероятность того, что в разведгруппу будут включены 2 связиста и 2 следопыта, если включение в разведгруппу равновероятно для любого ученика?

3. На карточках написаны целые числа от 1 до 15 включительно. Наудачу извлекаются две карточки. Какова вероятность того, что сумма чисел, написанных на этих карточках, равна десяти?

4. Для дежурства на вечере путем жеребьевки выделяются 5 человек. Вечер проводит комиссия, в составе которой 10 юношей и 2 девушки. Найдите вероятность того, что в число дежурных войдут обе девушки.

5. Имеется 6 билетов в театр, из которых 4 билета на места первого ряда. Какова вероятность того, что из трех наудачу выбранных билетов два окажутся на места первого ряда?

6. Билет в партер стоит 50 коп., на бельэтаж — 40 коп. и на ярус — 30 коп. Найдите вероятность того, что взятые наудачу два билета стоят вместе не дороже 80 коп.

7. На один ряд из семи мест случайным образом рассаживаются 7 учеников. Найдите вероятность того, что 3 определенных ученика окажутся рядом.

8. Из букв слова событие, составленного с помощью разрезной азбуки, извлекаются наудачу и складываются друг за другом в порядке их извлечения 3 карточки (буквы). Какова вероятность получить при этом слово быт?

9. Из пяти видов открыток, имеющихся в автомате, наудачу выбираются 3 открытки. Какова вероятность того, что все отобранные открытки будут разные?

10. Во время спортивной игры по команде ведущего «становись!» 10 учеников в случайном порядке образовали строй в одну шеренгу. Какова вероятность того, что ученики А и В окажутся отделенными друг от друга тремя учениками?

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Задача 1. Найти вероятность того, что при бросании игральной кости выпадет число, кратное трём (событие А).

Решение: Число всех элементарных событий $n = 6$, число элементарных событий, благоприятствующих данному событию А, $m = 2$ (это выпадение 3 или 6), значит,

$$P(A) = \frac{m}{n} = \frac{1}{3}.$$

В следующих двух примерах мы используем комбинаторные понятия (сочетания и перестановки), разобранные выше (в практическом занятии № 1).

Задача 2. В урне находится 7 шаров, из которых 4 белых и 3 черных. Из урны наугад вынимается три шара. Найти вероятность того, что два шара будут белыми, а один – чёрным.

Решение: Число всех элементарных событий подсчитаем по формуле сочетаний $n = C_7^3 = 35$, число элементарных событий, благоприятствующих данному событию А, по формуле сочетаний, применив правило произведения, $m = C_4^2 C_3^1 = 18$, значит,

$$P(A) = \frac{m}{n} = \frac{C_4^2 C_3^1}{C_7^3} = \frac{18}{35}.$$

Задача 3. На книжной полке случайно размещаются 8 книг, из которых 5 – по математике и 3 – по физике. Найти вероятность того, что слева будут все книги по математике, а справа – все книги по физике.

Решение: Число всех элементарных событий $n = P_8 = 8!$, число элементарных событий, благоприятствующих данному событию А, $m = P_5 P_3 = 5!3!$, значит, $P(A) = \frac{m}{n} = \frac{5!3!}{8!} = \frac{1}{56}.$

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 1.1. Основные понятия теории вероятностей**Практическое занятие № 3****Вычисление условной вероятности и проверка независимости событий**

Цель: сформировать умение решать задачи на вычисление вероятностей сложных событий.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:**Решить задачи.**

1. Вероятность того, что стрелок при одном выстреле выбивает 10 очков, равна 0,4; 9 очков – 0,3 и 8 или меньше очков – 0,3. Найти вероятность того, что при одном выстреле стрелок выбьет не менее 9 очков.

2. В клуб принесли в корзине 9 рыжих и 11 серых котят. Сначала вынимают одного котенка, затем другого, не возвращая первого обратно. Какова вероятность того, что котята разного цвета?

3. В одном ящике 6 синих и 11 зеленых шаров, а в другом – 7 синих и 9 зеленых шаров. Из каждого ящика взяли по одному шару. Какова вероятность того, что оба шара синие?

4. В коробке 4 белых и 5 черных футболок. Наугад вытаскивают две футболки. Найти вероятность того, что одна футболка белая, другая – черная?

5. В урне находятся 10 шаров: 3 белых и 7 черных. Извлекают один шар, а затем другой. Найти вероятность того, что первый шар черный, а второй – синий.

6. В специализированную больницу поступает в среднем 50% больных с заболеванием А, 30% - с заболеванием В, 20% - с заболеванием С. Вероятность полного излечения болезни А равна 0,7, для болезней В и С эти вероятности равны соответственно 0,8 и 0,9. Какой процент пациентов, полностью излечившихся по окончании курса лечения?

7. На трех дочерей Нину, Еву и Айгуль в семье возложена обязанность мыть посуду. Нина выполняет 40% всей работы, а остальные 60% работы Ева и Айгуль делят поровну. С

вероятностью 0.02 Нина может разбить по крайней мере одну тарелку; для Евы и Айгуль эта вероятность равна соответственно 0.03 и 0.04. Родители не знают, кто мыл посуду вечером, но они слышали звон разбитой посуды. Чья очередь мыть посуду в этот вечер наиболее вероятна?

8. Один властелин, которому его звездочет наскучил своими ложными предсказаниями, решил казнить его. Однако, будучи очень добрым повелителем, он решил дать звездочету шанс и предложить ему распределить по двум урнам четыре шара: два белых и два черных. Палач выбирает одну из урн и извлекает из нее один шар. Если шар будет черным, то звездочета казнят, в противном случае его жизнь будет спасена. Каким образом звездочет должен разместить шары в урнах, чтобы обеспечить себе максимальную вероятность спасения?

9. Устройство состоит из трех элементов, работающих независимо. Вероятности безотказной работы для них равны соответственно 0.6, 0.7, 0.8. Найти вероятность того, что отказал один элемент, а два другие - исправны.

10. В электрическую цепь последовательно включены три элемента, работающие независимо. Вероятности отказа каждого из элементов равны соответственно 0.1, 0.15, 0.2. Найти вероятность того, что тока в цепи не будет.

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Задача 1. В урне находятся 6 белых и 5 чёрных шаров. Найти вероятность того, что из пяти случайно взятых шаров

- а) не больше двух белых;
- б) хотя бы один белый.

Решение. а). Событие A – не больше двух белых шаров – равно сумме трёх попарно несовместимых событий A_0, A_1, A_2 – белых шаров соответственно ни одного, один, два: $A = A_0 + A_1 + A_2$. Значит, по теореме сложения для несовместных событий

$$P(A) = P(A_0) + P(A_1) + P(A_2) = \frac{C_6^0 C_5^5}{C_{11}^5} + \frac{C_6^1 C_5^4}{C_{11}^5} + \frac{C_6^2 C_5^3}{C_{11}^5} =$$

$$\frac{1}{462} + \frac{30}{462} + \frac{150}{462} = \frac{181}{462} \quad (\text{по поводу вычисления данных вероятностей см. аналогичную задачу в примере 12}).$$

б). Для нахождения вероятности события B – хотя бы один белый шар – неудобно использовать теорему сложения, т. к. пришлось бы подсчитывать вероятности пяти событий. Поэтому воспользуемся теоремой 5, которая сводит вычисление $P(B)$ к вычислению вероятности противоположного события $\bar{B}$ – ни одного белого шара, т. е. все шары чёрные:

$$P(B) = 1 - P(\bar{B}) = 1 - \frac{C_5^5}{C_{11}^5} = 1 - \frac{1}{462} = \frac{461}{462}.$$

Задача 2. Два стрелка стреляют по мишени. Вероятность попадания первого стрелка равна 0,6 второго – 0,8. Найти вероятности попадания

- а) обоих стрелков;
- б) ровно одного стрелка;
- в) хотя бы одного стрелка.

Решение. Обозначим через A и B события – попал в мишень первый и, соответственно, второй стрелок. Заметим, что эти события *независимы*. Заметим также, что и противоположные события (промахи стрелков) – независимы. С помощью известных нам операций над событиями выразим через A и B события, вероятности которых являются искомыми.

а). Очевидно, интересующее нас событие есть AB . Тогда по теореме умножения для независимых событий $P(AB) = P(A)P(B) = 0,6 \cdot 0,8 = 0,48$.

б). Интересующее нас событие есть $A\bar{B} + \bar{A}B$. Слагаемые в этой сумме – несовместные события, а сомножители в каждом произведении – независимые. Тогда по теоремам сложения для несовместных событий и умножения для независимых событий $P(A\bar{B} + \bar{A}B) = P(A)P(\bar{B}) + P(\bar{A})P(B) = 0,6 \cdot 0,2 + 0,4 \cdot 0,8 = 0,44$.

в). Интересующее нас событие есть $A + B$. По теореме сложения (для двух произвольных событий)

$$P(A + B) = P(A) + P(B) - P(AB) = P(A) + P(B) - P(A)P(B) = 0,6 + 0,8 - 0,6 \cdot 0,8 = 0,92.$$

Заметим, что противоположным событием к $A + B$ является $\bar{A}\bar{B}$ (ни один не попал). Тогда искомую вероятность можно найти и через противоположное событие:

$$P(A + B) = 1 - P(\bar{A}\bar{B}) = 1 - P(\bar{A})P(\bar{B}) = 1 - 0,4 \cdot 0,2 = 0,92.$$

Задача 3. Телеграфное сообщение состоит из сигналов "точка" и "тире". Статистические свойства помех таковы, что искажаются в среднем 2/5 сигналов "точка" и 1/3 сигналов "тире". Известно, что среди передаваемых сигналов "точка" и "тире" встречаются в соотношении 5:3. Определить вероятность того, что переданный сигнал искажён.

Решение: Рассмотрим следующие события: A – сигнал искажён (вероятность которого является искомой), H_1 – передан сигнал "точка", H_2 – передан сигнал "тире" (две гипотезы, образующие полную группу, т. к. они несовместны и одна из них обязательно имеет место). По условию $P(H_1):P(H_2) = 5:3$ и $P(H_1) + P(H_2) = 1$, поэтому $P(H_1) = \frac{5}{8}$ и $P(H_2) = \frac{3}{8}$. Известные вероятности того, что сигналы "точка" и "тире" искажаются, выступают как условные вероятности интересующего нас события при справедливости гипотез: $P_{H_1}(A) = 2/5$, $P_{H_2}(A) = 1/3$. Вероятность события A находим по формуле полной вероятности:

$$P(A) = \frac{5}{8} \cdot \frac{2}{5} + \frac{3}{8} \cdot \frac{1}{3} = \frac{3}{8}.$$

В условиях применения формулы полной вероятности рассмотрим следующую задачу: чему равна условная вероятность гипотезы H_i , если событие A произошло? Ответ даёт следующая *формулой Байеса*:

$$P_A(H_i) = \frac{P(H_i)P_{H_i}(A)}{P(A)}.$$

Продолжение задачи 3. Пусть переданный сигнал искажён. Чему равна вероятность того, что была передана "точка"?

Решение: Искомая вероятность вычисляется по формуле Байеса:

$$P_A(H_1) = \frac{P(H_1)P_{H_1}(A)}{P(A)} = \frac{\frac{5}{8} \cdot \frac{2}{5}}{\frac{3}{8}} = \frac{2}{3}.$$

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 1.2. Случайные величины и распределения

Практическое занятие № 4

Вычисление математического ожидания и дисперсии дискретных случайных величин. Построение и анализ биномиального и нормального распределений

Цель: изучение законов распределения дискретной случайной величины; научиться решать задачи на нахождение числовых характеристик случайной величины.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:

Решить задачи.

1. Найти математическое ожидание, дисперсию, среднее квадратическое отклонение, моду и медиану дискретной случайной величины, представленной законом распределения:

X	10	13	17	20	25
P	0,4	0,3	0,1	0,15	0,05

Составить функцию распределения $F(x)$.

2. Найти математическое ожидание, дисперсию, среднее квадратическое отклонение, моду и медиану дискретной случайной величины, представленной законом распределения:

X	8	14	17	20	23
P	0,2	0,1	0,2	0,4	0,1

Составить функцию распределения $F(x)$.

3. Найти математическое ожидание, дисперсию, среднее квадратическое отклонение, моду и медиану дискретной случайной величины, представленной законом распределения:

X	14	15	17	25	26
P	0,1	0,35	0,3	0,2	0,05

Составить функцию распределения $F(x)$.

4. Найти математическое ожидание, дисперсию, среднее квадратическое отклонение, моду и медиану дискретной случайной величины, представленной законом распределения:

X	16	20	25	30	35
P	0,2	0,15	0,15	0,3	0,2

Составить функцию распределения $F(x)$.

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Задача 1. Задают ли закон распределения ДСВ каждая из следующих таблиц?

а)

X	0	1	2	3	4
P	0,05	0,15	0,20	0,25	0,35

б)

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

Решение. а) Данная таблица задает закон распределения ДСВ, поскольку выполняется равенство (1): $0,05 + 0,15 + 0,20 + 0,25 + 0,35 = 1$.

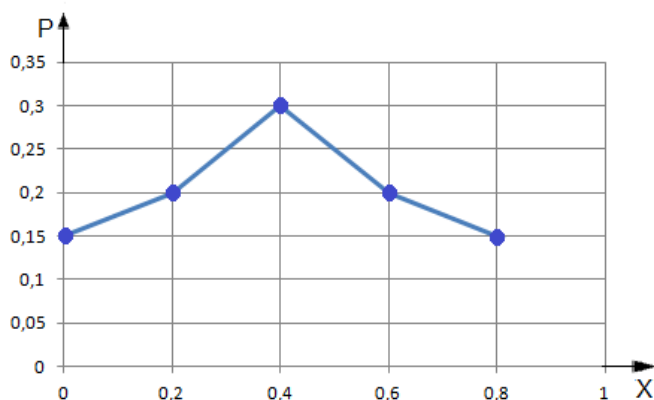
б) Данная таблица не задает закон распределения ДСВ, поскольку не выполняется равенство (1): $0,1 + 0,2 + 0,3 + 0,4 + 0,15 = 1,15 \neq 1$.

Задача 2. ДСВ X имеет закон распределения, заданный таблицей. Чему равна вероятность p_4 ? Постройте многоугольник распределения.

X	0	0,2	0,4	0,6	0,8
P	0,15	0,2	0,3	p_4	0,15

Решение. Поскольку должно выполняться равенство (1), то

$$p_4 = 1 - (0,15 + 0,20 + 0,3 + 0,15) = 0,2$$



Задача 3. В партии из 10 деталей имеется 8 стандартных. Из этой партии наудачу взято 2 детали. Найдите закон распределения ДСВ, равной числу стандартных деталей в выборке.

Решение. Пусть случайная величина X — число стандартных деталей в выборке. Случайная величина X может принимать значения $k=\{0,1,2\}$. Найдём вероятности этих значений.

X	0	1	2
P	p_1	p_2	p_3

$$p_1 = P(X = 0) = \frac{m}{n}$$

Вычислим число всех исходов (n): $C_{10}^2 = \frac{10!}{2!8!} = \frac{9 \cdot 10}{2} = 45$
(из 10 деталей берем 2)

Вычислим число благоприятствующих исходов (m): $C_8^0 \cdot C_2^2 = 1 \cdot 1 = 1$
(из 2 выбранных деталей 0 стандартных)

$$p_1 = P(X = 0) = \frac{1}{45}$$

$$p_2 = P(X = 1) = \frac{m}{n}$$

Вычислим число всех исходов (n): $C_{10}^2 = 45$
(из 10 деталей берем 2)

Вычислим число благоприятствующих исходов (m): $C_8^1 \cdot C_2^1 = 8 \cdot 2 = 16$
(из 2 выбранных деталей 1 стандартная)

$$p_2 = P(X = 1) = \frac{16}{45}$$

$$p_3 = P(X = 2) = \frac{m}{n}$$

Вычислим число всех исходов (n): $C_{10}^2 = 45$
(из 10 деталей берем 2)

Вычислим число благоприятствующих исходов (m): $C_8^2 \cdot C_2^0 = \frac{8!}{2! \cdot 6!} \cdot 1 = \frac{7 \cdot 8}{2} = 28$
(из 2 выбранных деталей 2 стандартные)

$$p_3 = P(X = 2) = \frac{28}{45}$$

X	0	1	2
P	1/45	16/45	28/45

Данная таблица задает закон распределения ДСВ, поскольку выполняется равенство (1):

$$\sum_{i=1}^3 p_i = p_1 + p_2 + p_3 = \frac{1}{45} + \frac{16}{45} + \frac{28}{45} = 1$$

Задача 4. ДСВ X имеет закон распределения, заданный таблицей. Найти функцию распределения этой ДСВ и построить ее график.

X	2	4	5	7	10
P	0,1	0,3	0,25	0,05	0,3

Решение. Для построения функции распределения $F(x)$ дискретной случайной величины X воспользуемся формулой (2):

при $x < 2$ $F(x) = 0$;

при $2 < x < 4$ $F(x) = p_1 = 0,1$;

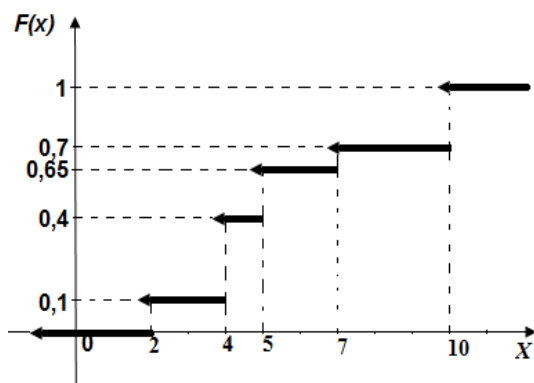
при $4 < x < 5$ $F(x) = p_1 + p_2 = 0,1 + 0,3 = 0,4$;

при $5 < x < 7$ $F(x) = p_1 + p_2 + p_3 = 0,1 + 0,3 + 0,25 = 0,65$;

при $7 < x < 10$ $F(x) = p_1 + p_2 + p_3 + p_4 = 0,1 + 0,3 + 0,25 + 0,05 = 0,7$;

при $x > 10$ $F(x) = p_1 + p_2 + p_3 + p_4 + p_5 = 0,1 + 0,3 + 0,25 + 0,05 + 0,3 = 1$

$$F(x) = \begin{cases} 0 & \text{при } x < 2 \\ 0,1 & \text{при } 2 < x < 4 \\ 0,4 & \text{при } 4 < x < 5 \\ 0,65 & \text{при } 5 < x < 7 \\ 0,7 & \text{при } 7 < x < 10 \\ 1 & \text{при } x > 10 \end{cases}$$



Задача 5. ДСВ X задана законом распределения. Вычислить значения моды, медианы, математического ожидания, дисперсии и среднего квадратического отклонения.

X	2	4	5	7	10
P	0,1	0,3	0,25	0,05	0,3

Решение.

Моды будет две, т.к. наибольшее значение вероятности соответствует двум значениям случайной величины $M_o(X)=4$ и $M_o(X)=10$.

Медиана, то значение случайной величины, которое находится посередине $M_e(X)=5$.

Математическое ожидание вычислим по формуле $M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + \dots + x_n \cdot p_n$

$$M(X) = 2 \cdot 0,1 + 4 \cdot 0,3 + 5 \cdot 0,25 + 7 \cdot 0,05 + 10 \cdot 0,3 = 6$$

Дисперсию вычислим по формуле $D(X) = M(X^2) - (M(X))^2$

$$M(X^2) = 2^2 \cdot 0,1 + 4^2 \cdot 0,3 + 5^2 \cdot 0,25 + 7^2 \cdot 0,05 + 10^2 \cdot 0,3 = 43,90$$

$$D(X) = 43,90 - (6)^2 = 7,90$$

Среднее квадратическое вычислим по формуле $\sigma(X) = \sqrt{D(X)}$

$$\sigma(X) = \sqrt{7,90} \approx 2,81$$

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 1.2. Случайные величины и распределения

Практическое занятие № 5

Применение распределения Пуассона для моделирования редких событий

Цель: изучение законов распределения дискретной случайной величины; научиться решать задачи на нахождение числовых характеристик случайной величины.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:

Решить задачи.

1. На цель сбрасывается 6 бомб, вероятность попадания каждой в цель составляет 0,3. Найти вероятность поражения цели 4 бомбами.

2. Фирма снабжает своей продукцией пять магазинов. От каждого магазина может поступить заявка на очередной день с вероятностью 0,4 независимо от заявок других магазинов. Какова вероятность того, что поступит не менее четырех заявок?

3. Вероятность попадания стрелком в мишень при каждом выстреле не зависит от результатов предыдущих выстрелов и равна 0,8. Стрелок сделал 5 выстрелов. Найти вероятность следующего события: мишень поражена одной пулей.

4. Вероятность попадания бомбы в цель составляет 0,25. Сбрасывается 8 бомб. Найти вероятность того, что будет не менее 7 попаданий.

5. Магазин получил 1000 стеклянных бутылок минеральной воды. Вероятность того, что при перевозке бутылка будет разбита, равна 0,003. Найти вероятность того, что при перевозке будут разбиты не более двух бутылок.

6. На АТС поступают в среднем 12 заказов в минуту. Найти вероятность того, что за 20 мин поступят ровно 10 заказов.

7. Обувной магазин продал 200 пар обуви. Вероятность того, что в магазин будет возвращена бракованная пара равна 0,01. Найти вероятность того, что из проданных пар обуви будет возвращено не более двух пар.

8. В страховую компанию в среднем поступает 2 иска в час. Определите вероятность того, что в течение 1,5 часов не поступит ни одного иска.

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Задача 1. Вероятность попадания в мишень при одном выстреле для данного стрелка 0,7 и не зависит от номера выстрела. Найти вероятность того, что при 5 выстрелах произойдет: а) ровно два попадания в мишень; б) не менее четырех попаданий в мишень.

Решение. Пусть случайная величина X – число попаданий в мишень.

а) По условию задачи $n=5$, $k=2$, $p=0,7$, следовательно, $q=1-p=1-0,7=0,3$. Тогда по формуле Бернулли

$$P_5(X=2) = C_5^2 p^2 q^{5-2} = \frac{5!}{2!3!} \cdot (0,7)^2 \cdot (0,3)^3 = 10 \cdot 0,49 \cdot 0,027 = 0,1323$$

б) По условию задачи $n=5$, $p=0,7$, $q=0,3$, не менее четырех попаданий означает либо четыре, либо пять попаданий, т.е. $k \geq 4$. Тогда по формуле Бернулли

$$P_5(X \geq 4) = P_5(X=4) + P_5(X=5) = C_5^4 p^4 q^{5-4} + C_5^5 p^5 q^{5-5} = 5 \cdot (0,7)^4 \cdot (0,3)^1 + 1 \cdot (0,7)^5 \cdot (0,3)^0 = 0,36015 + 0,16807 = 0,52822 \approx 0,528$$

Задача 2. Вероятность изготовления нестандартной детали 0,004. Найти вероятность того, что среди 1000 деталей окажется 5 нестандартных.

Решение. Пусть случайная величина X – число нестандартных деталей.

По условию задачи $n=1000$, $k=5$, $p=0,004$. Тогда $\lambda=np=1000 \cdot 0,004=4$ и по первой формуле пуассоновского распределения

$$P_{1000}(X=5) = \frac{4^5}{5!} e^{-4} = \frac{4 \cdot 4 \cdot 4 \cdot 4 \cdot 4}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} \cdot \frac{1}{e^4} = \frac{128}{15} \cdot \frac{1}{e^4} \approx 0,1562$$

Задача 3. Среднее число вызовов поступающих на АТС в одну минуту, равно двум. Найти вероятность того, что за 5 минут поступит: а) два вызова; б) менее двух вызовов.

Решение. Пусть случайная величина X – число вызовов за 5 минут.

а) По условию задачи $\lambda=2$, $t=5$, $k=2$. Тогда по второй формуле пуассоновского распределения

$$P_5(X=2) = \frac{(2 \cdot 5)^2}{2!} e^{-2 \cdot 5} = \frac{10^2}{2} \cdot \frac{1}{e^{10}} = \frac{50}{e^{10}}$$

б) По условию задачи $\lambda=2$, $t=5$, $k<2$. Тогда по второй формуле пуассоновского распределения

$$P_5(X < 2) = P_5(X = 0) + P_5(X = 1) = \frac{(2 \cdot 5)^0}{0!} e^{-2.5} + \frac{(2 \cdot 5)^1}{1!} e^{-2.5} = \frac{10^0}{1} \cdot \frac{1}{e^{10}} + \frac{10^1}{1} \cdot \frac{1}{e^{10}} = \frac{1}{e^{10}} + \frac{10}{e^{10}} = \frac{11}{e^{10}}$$

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 1.3. Центральная предельная теорема

Практическое занятие № 6

Демонстрация центральной предельной теоремы на основе генерации выборок и построения гистограмм

Цель: изучение центральной предельной теоремы на основе генерации выборок и построения гистограмм.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; использовать расчетные формулы, таблицы, графики при решении статистических задач.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:

Ответьте на вопросы после выполнения практической работы.

1. Как изменилась форма распределения при переходе от исходных данных к распределению выборочных средних? Почему это произошло?

2. Почему при увеличении объема выборки ($n = 30$) распределение средних стало еще более похожим на нормальное, чем при $n = 5$?

3. Какое практическое значение имеет ЦПТ для специалиста по ИИ? (Подсказка: оценка доверительных интервалов, проверка статистических гипотез, понимание работы многих алгоритмов машинного обучения, которые предполагают нормальность данных).

Порядок выполнения работы:

1. Повторите теоретические сведения.

2. Выполните практические задания, приведенные ниже, и оформите результаты в тетради.

3. Ответьте на вопросы в тетради.

Практические задания

Мы проведем эксперимент на примере экспоненциального распределения. Оно сильно отличается от нормального (асимметричное, с длинным "хвостом"), что делает результат наглядным.

Этап 1: Генерация исходной совокупности

1. Задача: Сгенерировать большую совокупность данных (например, 1000 значений), распределенных по экспоненциальному закону.

2. В электронных таблицах:
 - В ячейку A1 введите формулу для генерации экспоненциального случайного числа. В Excel/Google Sheets это делается через равномерное распределение: `=LN(СЛЧИС())` или `=LN(RAND())`.
 - Протяните эту формулу вниз до ячейки A1000. Вы получили 1000 случайных чисел, представляющих нашу генеральную совокупность.
3. Построение гистограммы:
 - Выделите полученные 1000 значений.
 - Постройте гистограмму (столбчатую диаграмму).
 - Вывод: Вы увидите характерный "хвост" справа. Это явно не нормальное распределение.

Этап 2: Формирование выборок и вычисление их средних

Теперь мы будем многократно извлекать из этой совокупности малые выборки и считать их средние значения.

1. Настройка:
 - Выберем объем выборки $n = 5$.
 - Нам нужно получить, например, 500 таких выборочных средних.
2. Процесс:
 - В ячейку B1 введите формулу для вычисления среднего значения из 5 случайных элементов из нашей совокупности. Формула будет выглядеть так: `=СРЗНАЧ(ИНДЕКС($A$1:$A$1000; СЛУЧМЕЖДУ(1;1000)); ... )` (В Google Sheets аналогично, используя AVERAGE и INDEX). Эту формулу нужно повторить 5 раз внутри СРЗНАЧ, чтобы она брала 5 случайных значений из столбца A и считала их среднее.
 - Протяните эту формулу из ячейки B1 вниз до ячейки B500. Теперь у вас есть 500 значений — это средние значения 500 случайных выборок по 5 элементов.
 - Построение гистограммы:
 - Выделите значения в столбце B (от B1 до B500).
 - Постройте гистограмму.
 - Вывод: Сравните эту гистограмму с предыдущей. Она уже не так сильно скошена. Появляется намек на симметрию и колоколообразную форму.

Этап 3: Увеличение объема выборок

Теперь мы повторим процедуру, но с большим объемом выборки, чтобы увидеть эффект ЦПТ в полной мере.

1. Настройка:
 - Выберем объем выборки $n = 30$.
2. Процесс:
 - В ячейку C1 введите аналогичную формулу СРЗНАЧ, но теперь со ссылками на 30 случайных ячеек из столбца A.
 - Протяните формулу вниз до C500. Вы получили еще один столбец — 500 средних значений, но уже по 30 элементов в каждой выборке.
 - Построение гистограммы:
 - Выделите значения в столбце C (от C1 до C500).
 - Постройте гистограмму.
 - Вывод: Гистограмма стала очень похожа на классическое нормальное распределение ("колокол"). Она симметрична, и большинство значений сосредоточено вокруг центра.

Анализ и выводы. После выполнения всех этапов заполнить сравнительную таблицу.

Параметр	Исходная совокупность	Выборка (n=5)	Выборка (n=30)
Форма распределения	Экспоненциальная	Умеренно скошенная	Близкая к нормальной

Параметр	Исходная совокупность	Выборка (n=5)	Выборка (n=30)
Симметрия	Отсутствует	Частичная	Высокая
Пик распределения	В начале	Смещен к центру	Четко по центру

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 1.3. Центральная предельная теорема

Практическое занятие № 7

Применение центральной предельной теоремы для оценки распределения сумм случайных величин

Цель: сформировать умение применять центральную предельную теорему для оценки распределения сумм случайных величин.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; использовать расчетные формулы, таблицы, графики при решении статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:

Решить задачи.

1. Оценка времени ответа API-сервиса. Вы управляете облачным API-сервисом для распознавания речи. Сервис получает тысячи запросов в час. Время обработки одного запроса — это случайная величина XX . Из исторических данных известно, что среднее время обработки $\mu=200$, а стандартное отклонение $\sigma=40$ мс. Распределение времени не является нормальным и имеет "тяжелые хвосты" из-за редких, но долгих операций (например, обработка очень длинных аудиофайлов). Какова вероятность того, что суммарное время обработки пакета из 500 последовательных запросов превысит 105 секунд?

2. Расчет емкости базы данных. В базе данных хранится информация о миллионах пользователей. Размер записи для каждого пользователя — случайная величина Y , так как пользователи заполняют разное количество полей. Средний размер одной записи $\mu=2$ КБ, стандартное отклонение $\sigma=0.5$ КБ. Инженеру нужно выделить место на диске для хранения

данных 10,000 новых пользователей. Какой объем дискового пространства нужно зарезервировать, чтобы с вероятностью 99% его хватило для всех записей?

3. Анализ энергопотребления дата-центра. Дата-центр состоит из 400 идентичных серверов, выполняющих обучение нейронных сетей. Потребляемая мощность одного сервера – случайная величина WW из-за разной нагрузки на GPU/CPU. Средняя потребляемая мощность одного сервера $\mu=350$ Вт, стандартное отклонение $\sigma=70$ Вт. Какова вероятность того, что пиковое общее энергопотребление этих 400 серверов превысит 145,000 Вт (145 кВт)?

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберитесь решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Центральная предельная теорема утверждает, что сумма $S_n=X_1+X_2+\dots+X_n$ достаточно большого числа независимых одинаково распределённых случайных величин с конечной дисперсией имеет распределение, близкое к нормальному.

Математические параметры этого нормального распределения:

- Математическое ожидание (среднее): $\mu_{\text{sum}}=n \cdot \mu$, где μ — среднее одной величины.
- Дисперсия: $\sigma_{\text{sum}}^2=n \cdot \sigma^2$, где σ^2 — дисперсия одной величины.
- Стандартное отклонение: $\sigma_{\text{sum}}=\sqrt{n} \cdot \sigma$.

Ключевой вывод: разброс суммы растет медленнее, чем сама сумма (пропорционально $\sqrt{n}$, а не n). Это явление называется "усреднением шума".

Сценарий практической задачи: Оценка времени выполнения конвейера задач.

Представьте, что вы — инженер по MLOps. Вы развернули конвейер машинного обучения, который состоит из последовательной обработки множества небольших пакетов данных (например, 10 000 изображений). Время обработки одного изображения — случайная величина. Из-за флуктуаций сети, загрузки CPU и других факторов оно не является постоянным.

Исходные данные модели:

Пусть время обработки одного изображения (X) — это случайная величина, которая с равной вероятностью принимает значения от 50 мс до 70 мс. Это равномерное распределение.

- Минимальное значение (a) = 50 мс
- Максимальное значение (b) = 70 мс

Рассчитаем характеристики для одного изображения:

1. Среднее время (μ): $(a+b)/2=(50+70)/2=60$ мс.
2. Дисперсия (σ^2) для равномерного распределения:
 $((b-a)^2)/12=((70-50)^2)/12=400/12 \approx 33.33$ мс².
3. Стандартное отклонение (σ): $\sqrt{33.33} \approx 5.77$ мс.

Наша задача — оценить общее время обработки всей партии из $N=10,000$ изображений.

Часть 1: Расчет характеристик суммарного времени

Используя формулы ЦПТ, рассчитаем ожидаемое распределение общего времени обработки T_{total} .

1. Ожидаемое общее время (μ_{total}):
 $\mu_{\text{total}}=N \cdot \mu=10,000 \cdot 60 \text{ мс}=600,000 \text{ мс}=600 \text{ секунд (10 минут)}$

2. Дисперсия общего времени (σ_{total}^2):
 $\sigma_{\text{total}}^2=N \cdot \sigma^2=10,000 \cdot 33.33=333,300 \text{ мс}^2$

3. Стандартное отклонение общего времени (σ_{total}):

$$\sigma_{\text{total}} = \sqrt{\sigma_{\text{total}}^2} = \sqrt{333,300} \approx 577.35 \text{ мс} \approx 0.58 \text{ секунды}$$

Вывод на этом этапе: Хотя мы обрабатываем 10 000 отдельных случайных величин, итоговое время работы всего конвейера распределено нормально со средним 10 минут и очень узким стандартным отклонением всего ~0.6 секунды. Индивидуальный шум усреднился.

Часть 2: Применение модели для принятия решений

Теперь используем полученную модель нормального распределения для ответа на практические вопросы.

Вопрос 1: Какова вероятность того, что весь пакет будет обработан менее чем за 10 минут (600 секунд)? Поскольку наше нормальное распределение симметрично относительно среднего (600 сек), ровно половина значений лежит слева от среднего. Ответ: Вероятность составляет 50%.

Вопрос 2: Мы хотим гарантировать клиенту, что 95% всех партий будут обработаны в заявленный срок. Какой срок выполнения (Target Time) мы должны указать? Для нормального распределения 95% значений лежат в интервале $\text{среднее} \pm 1.96 \cdot \text{стандартное отклонение}$.

$$\text{TargetTime} = \mu_{\text{total}} + 1.96 \cdot \sigma_{\text{total}}$$

$$\text{TargetTime} = 600 + 1.96 \cdot 0.58 \approx 600 + 1.14 = 601.14 \text{ секунды}$$

Чтобы выполнить обещание с гарантией 95%, мы должны установить целевой срок равным 601.14 секундам (или округлить вверх для безопасности).

Вопрос 3: Как изменится неопределенность (стандартное отклонение), если мы увеличим размер пакета в 4 раза, до 40 000 изображений? Размер пакета N станет 40 000. Новое стандартное отклонение:

$$\sigma_{\text{new}} = \sqrt{N_{\text{new}}} \cdot \sigma = \sqrt{40,000} \cdot 0.58 = 200 \cdot 0.58 = 116 \text{ секунд}$$

Неопределенность выросла, но только в 2 раза, хотя объем данных вырос в 4 раза. Это прямое следствие роста стандартного отклонения пропорционально N.

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 1.4. Закон больших чисел

Практическое занятие № 8

Моделирование закона больших чисел на основе последовательных испытаний.

Оценка среднего значения выборки и математического ожидания с помощью закона больших чисел

Цель: сформировать умение оценивать среднее значение выборки и математического ожидания с помощью закона больших чисел.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; использовать расчетные формулы, таблицы, графики при решении статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:

Решить задачи.

1. Оценка вероятности "успеха" (Монета). Вы подбрасываете несимметричную монету. Вероятность выпадения "орла" (событие "успех") неизвестна и равна p . Вы решили оценить эту вероятность экспериментально, подбросив монету n раз и вычислив долю выпавших "орлов". Сколько раз (n) нужно подбросить монету, чтобы с вероятностью не менее 95% оценка вероятности отличалась от истинного значения p не более чем на 0.01 (т.е. на 1%)?

2. Оценка среднего времени ответа сервера. Инженеры измеряют время ответа (в миллисекундах) API-сервиса. Истинное среднее время ответа неизвестно и равно $\mu=200$ мс. Стандартное отклонение времени ответа составляет $\sigma=50$ мс. Команда хочет получить оценку среднего времени ответа с точностью до ± 10 мс. Какое количество запросов (n) им нужно отправить, чтобы быть уверенными в своей оценке на 99%?

3. Оценка доли кликов по баннеру (CTR). Маркетолог запускает новую рекламную кампанию и хочет оценить её кликабельность (Click-Through Rate, CTR). Он предполагает, что истинный CTR (вероятность клика по баннеру) равен $p=0.1$ (т.е. 1 из 10 показов приводит к клику). Сколько раз нужно показать баннер (n), чтобы оценить CTR с точностью до $\pm 2\%$ (т.е. $\epsilon=0.02$) с вероятностью не менее 95%?

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Выполните практические задания, приведенные ниже, и оформите результаты в тетради.

3. Решите задачи.

Практические задания

Закон больших чисел (ЗБЧ) утверждает, что при неограниченном увеличении числа независимых испытаний n среднее арифметическое наблюдаемых значений сходится по вероятности к их математическому ожиданию.

Проще говоря: чем больше данных мы собираем, тем ближе их среднее значение к истинному, теоретическому среднему.

• Выборочное среднее ($\bar{X}_n$): Среднее значение, рассчитанное по нашим экспериментальным данным.

• Математическое ожидание ($E[X]$): Теоретическое, "истинное" среднее значение случайной величины.

Формула ЗБЧ:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - E[X]| < \epsilon) = 1$$

Это означает, что с вероятностью, стремящейся к 1, разница между выборочным средним и математическим ожиданием будет меньше любого произвольно малого числа ϵ .

Практическая задача: Моделирование бросков игральной кости

Мы смоделируем классический эксперимент — многократное бросание правильной шестигранной кости.

1. Определение теоретического ожидания Математическое ожидание $E[X]$ для правильной кости вычисляется как среднее всех её граней:

$$E[X] = 1 + 2 + 3 + 4 + 5 + 6 = 21/6 = 3.5$$

Это наше "истинное" значение, к которому мы будем стремиться.

2. Проведение эксперимента (с помощью готовых инструментов)

Для этого задания не нужно писать код. Мы воспользуемся онлайн-калькуляторами для моделирования бросков кости. Такие инструменты легко найти по запросу "dice roll simulator" или "генератор случайных бросков кости".

Ход эксперимента: Мы будем проводить серии бросков и после каждого броска фиксировать текущее среднее значение всех выпавших очков.

- Серия А (Малые выборки):

- 1) Сделайте 10 бросков. Запишите результат каждого броска и вычислите среднее арифметическое этих 10 значений.
- 2) Сделайте еще 10 бросков (всего 20). Вычислите среднее для 20 значений.
- 3) Продолжайте до 50 и 100 бросков.

- Серия Б (Большие выборки):

- 1) Сделайте сразу 1000 бросков. Запишите итоговое среднее значение.
- 2) Сделайте 5000 бросков. Запишите итоговое среднее значение.
- 3) Сделайте 10000 бросков. Запишите итоговое среднее значение.

3. Анализ и визуализация результатов

Заполните таблицу полученными данными:

Количество бросков (n)	Итоговая сумма очков	Выборочное среднее ($\bar{X}_n$)	Отклонение от $E[X]=3.5$
10			
50			
100			
1000			
5000			
10000			

4. Построение графика. На основе таблицы постройте график, где по оси X отложено количество бросков (n), а по оси Y — выборочное среднее ($\bar{X}_n$).

- Вы увидите, что на малых n (10, 50) линия графика сильно колеблется.
- По мере роста n (100, 1000...) колебания затухают, и линия графика стабилизируется, прижимаясь к значению 3.5.

Обсуждение и выводы

После выполнения эксперимента обсудите следующие вопросы:

- 1) Наблюдение: Как менялось выборочное среднее с ростом числа испытаний? Подтвердились ли теоретические предсказания ЗБЧ?
- 2) Стабильность: Почему при малом числе испытаний (например, первые 10 бросков) результат может быть очень далек от 3.5? Как это связано с понятием "шум" в данных?
- 3) Практическое значение для ИИ:
 - Представьте, что вы обучаете модель распознавания изображений на датасете из 10 картинок. Оценка её точности будет очень нестабильной ("шумной").
 - Что произойдет с оценкой точности, если вы увеличите датасет до 10 000 изображений?
 - Вывод: Закон больших чисел — это математическое обоснование того, почему для обучения качественных моделей искусственного интеллекта требуются большие объемы данных. Он гарантирует, что статистика, собранная на обучающей выборке, будет адекватно отражать свойства реального мира.

Таким образом, вы на практике убедились, что Закон больших чисел — это не просто абстрактная теорема, а фундаментальный принцип, обеспечивающий надежность статистических выводов и моделей машинного обучения.

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 2.1. Оценка параметров**Практическое занятие № 9****Построение точечных оценок параметров для различных распределений. Оценка доверительных интервалов для среднего значения и дисперсии**

Цель: сформировать умение оценивать доверительные интервалы для среднего значения и дисперсии.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; использовать расчетные формулы, таблицы, графики при решении статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:**Решить задачи.**

- По данным выборки объема $n=50$ построен вариационный ряд:

X	2	4	6	8	10
n_i	4	12	18	10	6

Проверить при уровне значимости $\alpha=0.05$ гипотезу о том, что генеральная совокупность имеет равномерное распределение на отрезке $[a,b]$.

- По данным выборки объема $n=80$ построен вариационный ряд:

X	0	1	2	3	4
n_i	10	25	20	15	10

При уровне значимости $\alpha=0.01$ проверить гипотезу о том, что генеральная совокупность имеет биномиальное распределение $B(n,p)$.

Порядок выполнения работы:

- Повторите теоретические сведения.
- Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
- Решите задачи.

Примеры решения задач

Задача. Пусть для изучения количественного признака X из генеральной совокупности извлечена выборка объема $n = 100$, имеющая следующее статистическое распределение:

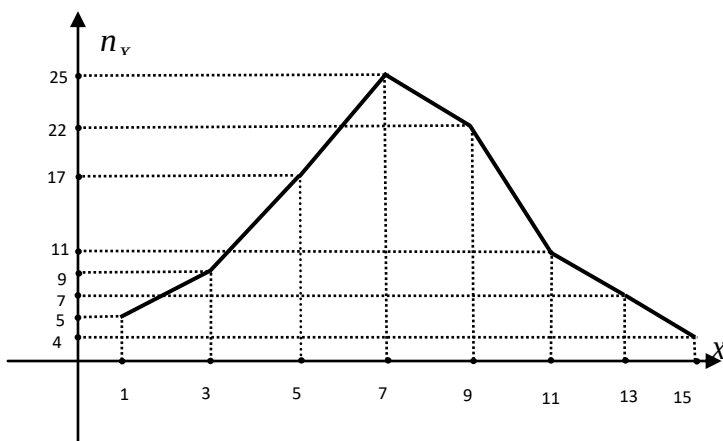
X	1	3	5	7	9	11	13	15
n_i	5	9	17	25	22	11	7	4

Требуется:

- построить полигон частот по данному распределению выборки;
- найти выборочное среднее $\bar{X}$, выборочное среднее квадратичное отклонение $\bar{\sigma}$, «исправленное» выборочное среднее квадратичное отклонение S ;
- при данном уровне значимости α проверить по критерию Пирсона гипотезу о нормальном распределении генеральной совокупности;
- в случае принятия гипотезы о нормальном распределении генеральной совокупности найти доверительные интервалы для математического ожидания m и среднего квадратичного отклонения σ , при уровне надежности $\gamma = 1 - \alpha$.

Решение

а). Отложим на оси абсцисс варианты $x_1, \dots, x_8$, а на оси ординат – соответствующие частоты $n_1, \dots, n_8$. Соединив полученные точки ломаной, получим полигон частот.



б). Выборочное среднее $\bar{X}$, выборочное среднее квадратичное отклонение $\bar{\sigma}$, «исправленное» выборочное среднее квадратичное отклонение S находятся соответственно по формулам:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^8 x_i n_i, \quad \bar{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^8 (x_i - \bar{X})^2 n_i}, \quad S = \sqrt{\frac{1}{n-1} \sum_{i=1}^8 (x_i - \bar{X})^2 n_i}.$$

Результаты вычислений удобно оформлять в виде таблицы.

i	x_i	n_i	$x_i n_i$	$\bar{X}$	$x_i - \bar{X}$	$(x_i - \bar{X})^2$	$(x_i - \bar{X})^2 n_i$	$\bar{\sigma}^2$	$\bar{\sigma}$	S^2
1	1	5	5		-6,62	43,82	219,10			
2	3	9	27		-4,62	21,34	192,06			
3	5	17	85		-2,62	6,86	116,62			
4	7	25	175		-0,62	0,38	9,50			
5	9	22	198		1,38	1,90	41,80			
6	11	11	121		3,38	11,42	125,62			
7	13	7	91		5,38	28,94	202,58			
8	15	4	60		7,38	54,46	217,84			

Σ		100	762	7,62			1125,12	11,25	3,36	11,36
----------	--	-----	-----	------	--	--	---------	-------	------	-------

Итак, $\bar{X} = 7,62$, $\bar{\sigma} = 3,36$, $S = 3,37$. ($S = \sqrt{S^2} = \sqrt{11,36}$)

в). Для проверки гипотезы о нормальном распределении генеральной совокупности применим критерий Пирсона. Учитывая, что варианты равноотстоят друг от друга (с шагом $h = x_i - x_{i-1} = 2$), это можно сделать следующим образом.

Концы интервалов, серединами которых являются x_i , вычисляются по формулам:

$\alpha_{i-1} = x_i - \frac{h}{2}$, $\alpha_i = x_i + \frac{h}{2}$, теоретические вероятности попадания X в интервал (α_{i-1}, α_i) ,

согласно формуле п.14, равны $p_i = \Phi\left(\frac{\alpha_i - \bar{X}}{S}\right) - \Phi\left(\frac{\alpha_{i-1} - \bar{X}}{S}\right)$, теоретические частоты равны $n'_i = np_i$

и, согласно п.14, $\chi^2 = \sum_{i=1}^8 \frac{(n_i - n'_i)^2}{n'_i}$.

Оформим вычисления в виде таблицы (см. ниже).

Из приложения 3 находим при данном уровне значимости $\alpha = 0,05$ и при числе степеней свободы $k = 8 - 3 = 5$ критическое значение $\chi^2_{крит} = 11,1$. Поскольку $\chi^2 = 2,856 < \chi^2_{крит} = 11,1$, гипотеза о нормальном распределении генеральной совокупности принимается.

i	x_i	α_i	$\frac{\alpha_i - \bar{X}}{S}$	$\Phi\left(\frac{\alpha_i - \bar{X}}{S}\right)$	p_i	n'_i	n_i	$\frac{(n_i - n'_i)^2}{n'_i}$
0		0	-2,26	-0,4881				
1	1	2	-1,67	-0,4525	0,0356	3,56	5	0,582
2	3	4	-1,07	-0,3577	0,0948	9,48	9	0,024
3	5	6	-0,48	-0,1844	0,1733	17,33	17	0,006
4	7	8	0,11	0,0438	0,2282	22,82	25	0,208
5	9	10	0,71	0,2611	0,2173	21,73	22	0,003
6	11	12	1,30	0,4032	0,1421	14,21	11	0,725
7	13	14	1,89	0,4706	0,0674	6,74	7	0,010
8	15	16	2,48	0,4934	0,0228	2,28	4	1,298
Σ								$\chi^2 = 2,856$

г). Теперь в предположении, что случайная величина X распределена нормально, найдем доверительные интервалы для математического ожидания a и среднего квадратичного отклонения σ по формулам:

$$a \in \left(\bar{X} - t_\gamma \frac{S}{\sqrt{n}}, \bar{X} + t_\gamma \frac{S}{\sqrt{n}} \right), \quad \sigma \in (S(1 - q_\gamma), S(1 + q_\gamma)) ,$$

где $\bar{X} = 7,62$, $S = 3,37$ – выборочное среднее и исправленное выборочное среднее квадратичное отклонение, найденные в п.б), t_γ, q_γ – коэффициенты, зависящие от уровня надежности $\gamma = 1 - \alpha = 0,95$ и объема выборки $n = 100$, которые находятся из приложений 5, 6:

$$t_\gamma = t(0,95;100) = 1,984, \quad q_\gamma = q(0,95;100) = 0,143 .$$

Итак, с надежностью $\gamma = 0,95$ a и σ принадлежат следующим интервалам: $a \in (6,95; 8,29)$, $\sigma \in (2,89; 3,85)$.

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 2.2. Тестирование гипотез

Практическое занятие № 10

Проверка гипотез с использованием критерия Стьюдента для двух выборок. Применение критерия χ^2 для проверки гипотез о независимости признаков. Оценка ошибок первого и второго рода при тестировании гипотез

Цель: ознакомление студентов с методами оценки различий между двумя независимыми выборками с использованием t-критерия Стьюдента и проверки гипотезы о независимости признаков с помощью критерия χ^2 .

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; использовать расчетные формулы, таблицы, графики при решении статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:

Решить задачи.

1. Команда ИИ-специалистов разработала новый алгоритм оптимизации нейронной сети. Чтобы оценить его эффективность, они решили измерить время обучения модели на одном и том же наборе данных (1000 изображений), но с использованием старого и нового алгоритмов. Для надежности эксперимента было взято 12 разных моделей. В результате для каждой модели было получено два времени обучения.

Данные:

- Объем выборки (количество пар): $n=12$.
- Среднее время обучения со **старым** алгоритмом: $\bar{x}_{old}=120$ секунд.
- Среднее время обучения с **новым** алгоритмом: $\bar{x}_{new}=105$ секунд.
- Стандартное отклонение **разностей** (времени нового и старого): $s_d=15$ секунд.

Вопрос: Можно ли утверждать, что новый алгоритм статистически значимо сокращает время обучения? Проверьте гипотезу при уровне значимости $\alpha=0.01$.

2. Инженер по качеству данных проверяет гипотезу о том, что ошибки в предсказаниях модели (остатки) распределены нормально. Это важное допущение для многих статистических методов. Он собрал выборку из 200 ошибок и сгруппировал их в интервалы.

Данные (Эмпирические частоты):

Интервал ошибок	Частота (n_i)
Менее -20	8

Интервал ошибок	Частота (ni)
от -20 до -10	22
от -10 до 0	60
от 0 до 10	80
от 10 до 20	25
Более 20	5
Итого	200

На основе предварительных расчетов известно, что если бы ошибки были распределены нормально, то теоретические вероятности попадания в эти интервалы были бы следующими: $P_1=0.04$, $P_2=0.12$, $P_3=0.30$, $P_4=0.32$, $P_5=0.16$, $P_6=0.04$.

Вопрос: При уровне значимости $\alpha=0.05$ согласуются ли эмпирические данные с гипотезой о нормальном распределении ошибок?

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Задача 1. Компания-разработчик ИИ выпустила обновление своего алгоритма сжатия изображений. Маркетинговый отдел утверждает, что новая версия сжимает изображения в среднем на 15% эффективнее, чем старая. Команда инженеров хочет проверить это утверждение на основе реальных данных.

Данные: Было проведено тестирование на 20 случайных изображениях для каждой версии алгоритма.

Новая версия (Выборка 1): Средний процент сжатия $\bar{x}_1=45\%$, стандартное отклонение $s_1=8\%$, объем выборки $n_1=20$.

Старая версия (Выборка 2): Средний процент сжатия $\bar{x}_2=38\%$, стандартное отклонение $s_2=7\%$, объем выборки $n_2=20$.

Задание: Проверить маркетинговое утверждение при уровне значимости $\alpha=0.05$.

Решение:

1. **Формулировка гипотез:**

○ H_0 (Нулевая гипотеза): Новая версия НЕ эффективнее старой на 15% или более. Математически: $\mu_1 - \mu_2 \leq 15$.

○ H_1 (Альтернативная гипотеза): Новая версия эффективнее старой на 15% или более. Математически: $\mu_1 - \mu_2 > 15$.

Это односторонний тест.

2. **Выбор критерия и расчет статистики:** Так как дисперсии неизвестны и, вероятно, не равны, используем t-критерий Уэлча (t-test for unequal variances). Статистика критерия:

$$t_{\text{набл}} = \frac{(\bar{x}_1 - \bar{x}_2) - C}{SE}$$

где $C=15$ (заявленное улучшение), а SE — стандартная ошибка разности средних:

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

3. **Вычисления:**

○ Разность выборочных средних: $\bar{x}_1 - \bar{x}_2 = 45 - 38 = 7\%$

○ Стандартная ошибка:

$$SE = \sqrt{\frac{8^2}{20} + \frac{7^2}{20}} = \sqrt{\frac{64}{20} + \frac{49}{20}} = \sqrt{\frac{113}{20}} = \sqrt{5.65} \approx 2.377$$

○ Наблюдаемое значение t-статистики:

$$t_{\text{набл}} = \frac{7 - 15}{2.377} = \frac{-8}{2.377} \approx -3.365$$

4. Определение критического значения и вывод:

- Находим число степеней свободы (приближенная формула Уэлча):

$$df \approx \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}} \approx 37$$

- По таблице t-распределения Стьюдента для одностороннего теста при $\alpha=0.05$ и $df=37$ критическое значение $t_{\text{крит}} \approx -1.687$.

- **Сравнение:** Наше наблюдаемое значение $t_{\text{набл}} \approx -3.365$ находится левее критического ($t_{\text{набл}} < t_{\text{крит}}$).

Вывод: Мы отвергаем нулевую гипотезу H_0 . Статистически значимые данные **не** подтверждают маркетинговое утверждение о том, что новая версия эффективнее на 15% или более.

Часть 2. Критерий χ^2 для проверки независимости признаков

Задача 2. Аналитики исследуют, есть ли связь между использованием новой библиотеки для обработки естественного языка (NLP) и скоростью разработки чат-ботов в компании. Были опрошены 100 разработчиков.

Данные (Таблица сопряженности):

	Используют новую библиотеку	Не используют	Итого
Проект сдан в срок	40	30	70
Проект сдан с задержкой	10	20	30
Итого	50	50	100

Задание: При уровне значимости $\alpha=0.05$ проверить гипотезу о том, что использование библиотеки и своевременность сдачи проекта являются независимыми признаками.

Решение:

1. Формулировка гипотез:

- H_0 : Признаки независимы (использование библиотеки не влияет на сроки).
- H_1 : Признаки зависимы.

2. **Расчет ожидаемых частот (E_{ij}):** Если признаки независимы, то ожидаемая частота для каждой ячейки рассчитывается как:

$$E_{ij} = \frac{(\text{Итого по строке}) \cdot (\text{Итого по столбцу})}{\text{Общий итог}}$$

- Для ячейки "Используют / В срок": $E_{11} = \frac{70 \cdot 50}{100} = 35$
- Для ячейки "Не используют / В срок": $E_{12} = \frac{70 \cdot 50}{100} = 35$
- Для ячейки "Используют / С задержкой": $E_{21} = \frac{30 \cdot 50}{100} = 15$
- Для ячейки "Не используют / С задержкой": $E_{22} = \frac{30 \cdot 50}{100} = 15$

3. Расчет статистики $\chi^2_{\text{набл}}$:

$$\chi^2_{\text{набл}} = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- Для (Используют, В срок): $\frac{(40-35)^2}{35} = \frac{25}{35} \approx 0.714$
- Для (Не используют, В срок): $\frac{(30-35)^2}{35} = \frac{25}{35} \approx 0.714$
- Для (Используют, С задержкой): $\frac{(10-15)^2}{15} = \frac{25}{15} \approx 1.667$
- Для (Не используют, С задержкой): $\frac{(20-15)^2}{15} = \frac{25}{15} \approx 1.667$

$$\chi^2_{\text{набл}} = 0.714 + 0.714 + 1.667 + 1.667 = 4.762$$

4. Определение критического значения и вывод:

- Число степеней свободы: $df=(R-1)(C-1)=(2-1)(2-1)=1$
- По таблице χ^2 -распределения при $\alpha=0.05$ и $df=1$, критическое значение $\chi^2_{\text{крит}}=3.841$.
- **Сравнение:** Наше наблюдаемое значение $\chi^2_{\text{набл}}=4.762$ больше критического ($4.762>3.841$).

Вывод: Мы отвергаем нулевую гипотезу H_0 . Есть статистически значимая связь между использованием новой библиотеки и своевременностью сдачи проектов.

Часть 3. Оценка ошибок I и II рода

Обсудим ошибки на примере первой задачи (тест алгоритма сжатия).

Ошибка I рода (α): Это вероятность отвергнуть верную нулевую гипотезу. В нашем контексте это означает, что мы бы решили, что новая версия алгоритма НЕ достигает заявленного улучшения в 15%, хотя на самом деле она его достигает.

Последствия: Компания может отказаться от внедрения действительно эффективного обновления, посчитав его бесполезным.

Ошибка II рода (β): Это вероятность принять ложную нулевую гипотезу. В нашем контексте это означает, что мы бы решили, что новая версия достигает заявленного улучшения, хотя на самом деле она его НЕ достигает (например, улучшение всего на 7%).

Последствия: Компания потратит ресурсы на внедрение обновления, которое не принесет ожидаемой выгоды, или будет давать клиентам ложные обещания.

Мощность теста ($1-\beta$): Это вероятность правильно отвергнуть ложную нулевую гипотезу. В контексте ИИ высокая мощность теста означает, что наш эксперимент с большой вероятностью обнаружит реальное улучшение алгоритма, если оно существует. Для повышения мощности обычно увеличивают размер выборки (n).

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 2.3. Корреляция и ковариация Практическое занятие № 11

Вычисление коэффициента корреляции Пирсона для анализа зависимостей между признаками. Построение корреляционной матрицы для многомерных данных и её интерпретация

Цель: освоение методов расчета коэффициентов корреляции Пирсона для определения линейных связей между численными признаками, а также изучение способов визуализации и интерпретации корреляций в многомерных наборах данных.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; использовать расчетные формулы, таблицы, графики при решении статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Задание:

Решить задачи.

1. Финансовый аналитик изучает поведение двух активов (например, акции компании "Техно" и акции компании "Нефть-Ресурс") на фондовом рынке, чтобы понять, как они движутся относительно друг друга. Это важно для формирования инвестиционного портфеля.

Данные: За 5 торговых дней зафиксирована доходность (в процентах) каждого актива.

День	Доходность Актива А (XX), %	Доходность Актива Б (YY), %
1	+2	+1
2	-1	-2
3	+3	+2
4	0	-1
5	-4	-4

Задание:

- 1) Вычислить ковариацию между доходностями активов.
- 2) Вычислить коэффициент корреляции Пирсона.
- 3) Интерпретировать результат с точки зрения управления рисками.

2. Маркетинговый отдел интернет-магазина проанализировал данные о поведении пользователей и построил корреляционную матрицу для следующих показателей:

- X1: Время, проведенное на сайте (в минутах).
- X2: Количество просмотренных страниц.
- X3: Сумма покупки (в рублях).

Полученная корреляционная матрица выглядит так:

$$\begin{pmatrix} X_1(\text{Время}) & X_1(\text{Время}) & X_2(\text{Страницы}) & X_3(\text{Сумма}) \\ X_1(\text{Время}) & 1 & 0.85 & 0.60 \\ X_2(\text{Страницы}) & 0.85 & 1 & 0.75 \\ X_3(\text{Сумма}) & 0.60 & 0.75 & 1 \end{pmatrix}$$

Задание: Проанализировать матрицу и ответить на вопросы:

- 1) Какой из поведенческих факторов (Время или Страницы) сильнее связан с итоговой суммой покупки?
- 2) Есть ли признаки избыточности (мультиколлинеарности) в данных?

3) Какой вывод может сделать маркетолог для оптимизации сайта?

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Задача 1. Вычисление коэффициента корреляции Пирсона. Вы — аналитик данных в компании, которая разрабатывает систему умного дома. У вас есть данные о 5 домах, и вы хотите проверить, есть ли связь между потребляемой мощностью (в кВт·ч) и ежемесячной стоимостью электроэнергии.

Данные:

Дом	Потребление (XX), кВт·ч	Стоимость (YY), руб.
1	150	750
2	180	900
3	200	1000
4	220	1100
5	250	1250

Задание:

- 1) Вычислить коэффициент корреляции Пирсона r_{XY} .
- 2) Интерпретировать полученный результат.

Решение:

1. Подготовка и расчеты (в табличном редакторе) Для расчета коэффициента корреляции Пирсона используется формула:

$$r_{XY} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Для удобства расчетов составим таблицу в Excel или Google Sheets.

Дом	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
1	150	750	-70	-250	17500	4900	62500
2	180	900	-40	-100	4000	1600	10000
3	200	1000	-20	0	0	400	0
4	220	1100	0	100	0	0	10000
5	250	1250	30	250	7500	900	62500
Сумма ($\sum$)	1000	5000	-160	-160	29000	7800	145000

- Среднее значение $\bar{x} = 1000/5 = 200$
- Среднее значение $\bar{y} = 5000/5 = 1000$

Теперь подставим суммы из таблицы в формулу:

$$r_{XY} = \frac{29\,000}{\sqrt{7\,800 \cdot 145\,000}}$$

$$r_{XY} = \frac{29\,000}{\sqrt{1\,131\,000\,000}}$$

$$r_{XY} = \frac{29\,000}{33\,630.3} \approx +1.0$$

Примечание: В реальных данных из-за округлений результат может быть, например, +1. или +1. Но в данном идеализированном примере он равен ровно +1.

2. Интерпретация результата Коэффициент корреляции Пирсона $r_{XY}=+1$.

- Знак "+" означает, что связь между признаками положительная: с ростом потребления электроэнергии растет и ее стоимость.

- Значение, равное 1, означает, что связь является полной (функциональной) и строго линейной.

- Это говорит о том, что стоимость электроэнергии напрямую и без отклонений зависит от объема потребления. В контексте ИИ это означает, что признак "Стоимость" является прямой линейной функцией от признака "Потребление" и не несет новой информации. При построении модели для прогноза стоимости достаточно знать только потребление.

Задача 2. Построение и интерпретация корреляционной матрицы. Вы работаете над проектом по предсказанию цены подержанного автомобиля. У вас есть данные по трем ключевым признакам для нескольких автомобилей.

Данные:

Автомобиль	Возраст (X1), лет	Пробег (X2), тыс. км	Цена (Y), тыс. руб.
А	1 год	15 тыс. км	2500
Б	3 года	45 тыс. км	2155
В	5 лет	75 тыс. км	1815

Задание: Построить корреляционную матрицу для этих трех признаков и сделать выводы о том, какие факторы сильнее всего влияют на цену автомобиля.

Решение:

Корреляционная матрица — это таблица, где на пересечении строки i и столбца j стоит коэффициент корреляции между признаком i и признаком j . Матрица всегда симметрична, а на главной диагонали стоят единицы (так как корреляция признака с самим собой равна 1).

Рассчитаем коэффициенты корреляции для всех пар признаков:

1. Корреляция между Возрастом (X1) и Пробегом (X2):
 - Логика подсказывает, что чем старше машина, тем больше у нее пробег.
 - Расчет показывает $r_{X1X2}=+1$. Это сильная положительная линейная связь.
2. Корреляция между Возрастом (X1) и Ценой (Y):
 - Логика подсказывает, что чем старше машина, тем она дешевле.
 - Расчет показывает $r_{X1Y}=-1$. Это сильная отрицательная линейная связь.
3. Корреляция между Пробегом (X2) и Ценой (Y):
 - Логика подсказывает, что чем больше пробег, тем дешевле машина.
 - Расчет показывает $r_{X2Y}=-1$. Это сильная отрицательная линейная связь.

Соберем все коэффициенты в матрицу:

$$\begin{pmatrix} & X_1(\text{Возраст}) & X_2(\text{Пробег}) & Y(\text{Цена}) \\ X_1(\text{Возраст}) & +1 & +1 & -1 \\ X_2(\text{Пробег}) & +1 & +1 & -1 \\ Y(\text{Цена}) & -1 & -1 & +1 \end{pmatrix}$$

Интерпретация матрицы

Анализ матрицы дает нам важную информацию для построения модели машинного обучения:

- Сильная мультиколлинеарность: Мы видим, что признаки "Возраст" и "Пробег" имеют коэффициент корреляции +1. Это означает, что они дублируют информацию друг друга. Зная возраст автомобиля, мы с абсолютной точностью можем предсказать его пробег по этим данным.

- Вывод для ИИ-специалиста: В модель машинного обучения не имеет смысла включать оба этих признака одновременно. Это может привести к нестабильности модели и неверной

оценке важности признаков. Следует оставить только один из них (например, "Возраст"), так как он также сильно коррелирует с целевой переменной "Цена" ($r=-1$).

- Оба признака являются отличными предикторами (предсказателями) для цены автомобиля.

Таким образом, корреляционный анализ помог нам очистить данные от избыточности еще до начала обучения модели.

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 2.3. Корреляция и ковариация

Практическое занятие № 12

Вычисление ковариации и её применение для оценки совместной изменчивости признаков

Цель: научиться рассчитывать ковариацию между двумя признаками, понимать смысл и назначение данной меры, уметь интерпретировать результаты для принятия обоснованных выводов о совместном поведении величин.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; использовать расчетные формулы, таблицы, графики при решении статистических задач.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Задача 1. Вычислить ковариацию вручную

Используя следующий набор данных, рассчитайте ковариацию доходов сотрудников и стажа работы:

Сотрудник	Доход (тыс. руб.)	Стаж работы (лет)
А	30	2

B	40	5
C	50	8
D	60	10
E	70	12

Шаг 1. Запишите формулу для расчета ковариации.

Шаг 2. Найдите среднее значение доходов и средний стаж работы.

Шаг 3. Используя формулу, вычислите ковариацию и сделайте вывод о взаимосвязи дохода и стажа.

Решение задачи №1:

Средние значения:

Средний доход: $\bar{X} = (30+40+50+60+70)/5 = 50$

Средний стаж: $\bar{Y} = (2+5+8+10+12)/5 = 7.4$

Подставляем в формулу:

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1},$$

$\text{Cov}(X, Y) = ((30-50) \cdot (2-7.4) + (40-50) \cdot (5-7.4) + \dots + (70-50) \cdot (12-7.4)) / (5-1) = 120$

Таким образом, положительная ковариация (+120) свидетельствует о прямой связи между стажем и доходом.

Задача №2: Автоматизированный расчет ковариации средствами Python

Даны следующие показатели потребления электроэнергии и температуры воздуха:

Потребление (кВт/ч)	Температура (°C)
150	10
180	15
200	20
220	25
250	30

Напишите программу на Python для автоматического расчета ковариации.

Алгоритм:

Загрузите данные в списки или массивы.

Используйте библиотеку numpy для быстрого расчета ковариации.

```
import numpy as np
```

```
consumption = np.array([150, 180, 200, 220, 250])
```

```
temperature = np.array([10, 15, 20, 25, 30])
```

```
cov_xy = np.cov(consumption, temperature)[0][1]
```

```
print(f"Ковариация: {cov_xy:.2f}")
```

Результат:

Ковариация равна примерно 450, что подтверждает прямую связь между потреблением энергии и температурой.

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 2.4. Регрессионный анализ

Практическое занятие № 13

Построение линейной регрессионной модели на основе экспериментальных данных. Интерпретация коэффициентов линейной регрессии и оценка её качества

Цель: научиться строить линейную регрессию, используя реальные данные, интерпретировать коэффициенты уравнения регрессии и оценивать качество полученной модели.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; применять современные пакеты прикладных программ многомерного статистического анализа.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Предположим, у вас имеются экспериментальные данные о расходе топлива автомобилем в зависимости от пройденного расстояния.

Необходимо построить простую линейную регрессионную модель, позволяющую предсказывать расход топлива автомобиля в зависимости от километража.

Исходные данные:

Пробег (км)	Расход топлива (литры)
100	8
200	15
300	22
400	28
500	35

Последовательность выполнения задания

Шаг 1: Импорт необходимых модулей

Для реализации будем использовать библиотеки Python:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score
```

Шаг 2: Подготовка данных

Загружаем данные в виде списков или таблиц:

```
data = {
```

```

        'пробег': [100, 200, 300, 400, 500],
        'расход': [8, 15, 22, 28, 35]
    }

df = pd.DataFrame(data)
X = df['пробег'].values.reshape(-1, 1) # Независимая переменная
Y = df['расход']                       # Зависимая переменная

```

Шаг 3: Создание и обучение модели

Используем класс `LinearRegression` из библиотеки `Scikit-Learn`:

```

model = LinearRegression() # Создаем объект класса
model.fit(X, Y)            # Тренируем модель

```

Шаг 4: Получение коэффициентов уравнения регрессии

Выведем свободные члены и коэффициенты наклона:

```

intercept = model.intercept_
coefficient = model.coef_[0]

```

```
print(f"Уравнение регрессии: Y = {intercept:.2f} + {coefficient:.2f}X")
```

Полученное уравнение может выглядеть так:

$Y = 3.5 + 0.06X$

Шаг 5: Прогнозирование и проверка точности модели

Сделаем прогнозы и оценим точность с помощью метрик MSE и R^2 :

```

predictions = model.predict(X)
mse = mean_squared_error(Y, predictions)
r2 = r2_score(Y, predictions)

```

```
print(f"Средняя квадратичная ошибка (MSE): {mse:.2f}")
```

```
print(f"Коэффициент детерминации (R²): {r2:.2f}")
```

Если MSE близка к нулю, а R^2 близко к единице, значит модель хорошо описывает данные.

Шаг 6: Графическое представление

Отобразим исходные точки и линию регрессии:

```

plt.scatter(df['пробег'], df['расход'], color="blue") # Данные
plt.plot(df['пробег'], predictions, color="green")   # Линия регрессии
plt.xlabel('Пробег (км)')
plt.ylabel('Расход топлива (литры)')
plt.title('Линия регрессии расхода топлива')
plt.grid(True)
plt.show()

```

Интерпретация результатов:

Коэффициент b (наклон) показывает, насколько изменяется расход топлива при увеличении пробега на один километр.

Свободный член a отражает базовую величину расхода при нулевом километраже.

Метрика R^2 оценивает долю вариабельности зависимой переменной, объясняемой моделью. Чем ближе R^2 к 1, тем лучше модель объясняет поведение данных.

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 2.4. Регрессионный анализ

Практическое занятие № 14

Применение нелинейной регрессии для аппроксимации данных

Цель: овладеть техникой подбора наилучшей нелинейной модели для описания данных, научиться выбирать оптимальную форму нелинейной зависимости и интерпретировать результаты.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; применять современные пакеты прикладных программ многомерного статистического анализа.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Задача:

Пусть нам даны экспериментальные данные о скорости роста бактерий в культуре среды в зависимости от времени культивирования.

Время (часы)	Число клеток (млрд)
0	0.1
1	0.3
2	0.7
3	1.5
4	3.2
5	6.5
6	12.5

Требуется подобрать подходящую нелинейную модель, объяснить ход выбора формы модели и провести оценку качества подобранной модели.

Этапы выполнения задания:

Шаг 1: Предварительное рассмотрение данных

Анализ графика зависимости числа клеток от времени покажет рост численности бактерий в геометрической прогрессии. Поэтому имеет смысл попробовать следующие типы нелинейных функций:

Экспоненциальная модель: $N(t) = ae^{bt}$

Полиномиальная модель: $N(t) = a + bt + ct^2 + \dots$

Шаг 2: Выбор оптимальной модели

Для начала попробуем применить экспоненциальную модель, так как скорость размножения бактерий часто соответствует экспоненциальному росту.

Шаг 3: Реализация нелинейной регрессии на Python

Импортируем необходимые модули:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit
```

Подготовка данных:

```
data = {'Время': [0, 1, 2, 3, 4, 5, 6],
        'Количество клеток': [0.1, 0.3, 0.7, 1.5, 3.2, 6.5, 12.5]}
```

```
df = pd.DataFrame(data)
```

```
X = df['Время']
```

```
Y = df['Количество клеток']
```

Подбор экспоненциальной модели:

```
def exponential(x, a, b):
    return a * np.exp(b*x)
```

```
popt, pcov = curve_fit(exponential, X, Y)
```

```
predicted_values = exponential(X, *popt)
```

Шаг 4: Оценка качества модели

Можно воспользоваться средней квадратичной ошибкой (MSE) и визуализировать график:

```
from sklearn.metrics import mean_squared_error
```

```
mse = mean_squared_error(Y, predicted_values)
```

```
print(f"Модель экспоненциального роста:\nПараметры: a={popt[0]:.2f},\nb={popt[1]:.2f}\nОшибка (MSE): {mse:.2f}")
```

```
plt.figure(figsize=(8, 6))
```

```
plt.scatter(X, Y, label='Экспериментальные данные')
```

```
plt.plot(X, predicted_values, color='red', label=f'Аппроксимирующая кривая\n$a*e^{(bx)}$')
```

```
plt.legend()
```

```
plt.xlabel('Время (часы)')
```

```
plt.ylabel('Количество клеток (млрд)')
```

```
plt.title('Нелинейная регрессия экспоненциальной моделью')
```

```
plt.grid(True)
```

```
plt.show()
```

Шаг 5: Интерпретация полученных результатов

Оценив графики и ошибку MSE, делаем вывод о пригодности выбранной модели. Обычно хорошую модель отличает гладкость кривой и малая средняя квадратичная ошибка.

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 2.5. Анализ дисперсии**Практическое занятие № 15****Проведение однофакторного дисперсионного анализа (ANOVA) для проверки различий между группами**

Цель: научиться проводить однофакторный ANOVA-анализ для выявления существенных различий между средними значениями трех или более групповых выборок.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; применять современные пакеты прикладных программ многомерного статистического анализа.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач**Задача:**

Исследователи изучают влияние различных удобрений на урожайность пшеницы. Имеются три типа удобрений, каждая группа состоит из пяти участков земли. Необходимо проверить, различаются ли удобрения по эффективности.

Удобрение А	Удобрение Б	Удобрение В
100 кг/га	110 кг/га	120 кг/га
105 кг/га	115 кг/га	125 кг/га
110 кг/га	120 кг/га	130 кг/га
115 кг/га	125 кг/га	135 кг/га
120 кг/га	130 кг/га	140 кг/га

Ход выполнения задания:**Шаг 1: Формулирование гипотез**

Нулевая гипотеза (H₀): Все группы имеют одинаковые средние значения урожая.

Альтернативная гипотеза (H_A): Хотя бы одна группа отличается средним значением урожая от остальных.

Шаг 2: Подготовка данных

Реализуем пример на Python:

```
import numpy as np
import pandas as pd
from scipy import stats

# Данные
fertilizerA = np.array([100, 105, 110, 115, 120])
fertilizerB = np.array([110, 115, 120, 125, 130])
fertilizerC = np.array([120, 125, 130, 135, 140])

groups = [fertilizerA, fertilizerB, fertilizerC]
```

Шаг 3: Проведение ANOVA-тестирования

Применяем встроенную функцию ANOVA из модуля SciPy:

```
statistic, p_value = stats.f_oneway(*groups)
```

```
print(f"F-статистика: {statistic:.2f}")
```

```
print(f"P-значение: {p_value:.2f}")
```

Шаг 4: Интерпретация результатов

При уровне значимости $\alpha=0.05$ сравниваем полученное p-значение с пороговым уровнем:

Если $p < \alpha$, отвергаем нулевую гипотезу и считаем, что существуют значимые различия между группами.

Если $p > \alpha$, принимаем нулевую гипотезу и утверждаем, что разница между группами незначительна.

Например, если получено p-значение около 0.001, то мы можем уверенно утверждать, что удобрения оказывают значительное влияние на урожайность.

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.

Тема 2.5. Анализ дисперсии

Практическое занятие № 16

Применение дисперсионного анализа для оценки влияния различных факторов на результаты экспериментов

Цель: научиться применять однофакторный и многофакторный дисперсионный анализ (ANOVA) для выявления значимых различий между группами и оценки вклада отдельных факторов в общую изменчивость результатов эксперимента.

Выполнив работу, вы будете уметь: применять стандартные методы и модели к решению вероятностных и статистических задач; применять современные пакеты прикладных программ многомерного статистического анализа.

Уо 01.01 распознавать и анализировать задачу и/или проблему в профессиональном и/или социальном контексте; выделять её составные части; определять этапы решения задачи; составлять план действий; определять необходимые ресурсы; реализовывать составленный план.

Уо 02.01 определять необходимые источники информации; эффективно искать информацию, необходимую для решения задачи и/или проблемы, оформлять результаты поиска.

Материальное обеспечение: Технические средства обучения: компьютер с лицензионным программным обеспечением и мультимедиа проектор.

Порядок выполнения работы:

1. Повторите теоретические сведения.
2. Разберите решение задач, приведенных ниже, и запишите их в тетрадь.
3. Решите задачи.

Примеры решения задач

Допустим, компания проводит маркетинговый эксперимент, проверяя эффективность трёх рекламных каналов ("Реклама на телевидении", "Контекстная реклама", "Социальные сети"). Для каждого канала фиксируют количество привлечённых клиентов за месяц.

Также дополнительно проводится анализ с учётом региона распространения рекламы ("Москва", "Санкт-Петербург", "Другие регионы").

Исходные данные представлены в следующей таблице:

Канал рекламы	Москва	СПб	Другие регионы
Телевизионная	100	120	80
Контекстная	150	160	130
Социальные сети	180	190	170

Требуется выяснить:

Есть ли существенные различия в привлечении клиентов между разными каналами рекламы?

Оказывают ли различные регионы воздействие на итоговый результат?

Последовательность выполнения задания:

Шаг 1: Формулировка гипотез

Общая гипотеза (H₀): Нет значительных различий между эффектами разных каналов рекламы и регионов.

Альтернативная гипотеза (H₁): Существуют значительные различия между результатами каналов и регионов.

Шаг 2: Подготовка данных

Запишем данные в удобной форме для обработки на Python:

```
import pandas as pd
import numpy as np
from scipy import stats

# Данные
channels = ["Телепередача", "Контекстная", "Социальные сети"]
regions = ["Москва", "СПб", "Другие регионы"]

# Матрица данных
data = [
    [100, 120, 80], # Телевизионная реклама
    [150, 160, 130], # Контекстная реклама
```

```

    [180, 190, 170]    # Реклама в социальных сетях
]

```

Преобразуем данные в удобный формат

```
df = pd.DataFrame(data, index=channels, columns=regions)
```

Шаг 3: Одномерный ANOVA (однофакторный анализ)

Сначала рассмотрим каналы отдельно от регионов.

Приводим данные в вид, пригодный для анализа

```
values_tv = data[0]
```

```
values_context = data[1]
```

```
values_social = data[2]
```

ANOVA-тест

```
f_value, p_value = stats.f_oneway(values_tv, values_context,
values_social)
```

```
print(f"Одномерный ANOVA:")
```

```
print(f"F-статистика: {f_value:.2f}")
```

```
print(f"P-значение: {p_value:.2f}")
```

Если р-значение меньше 0.05, то можно считать, что различия между рекламными каналами значимы.

Шаг 4: Двухмерный ANOVA (многофакторный анализ)

Теперь проведём двухфакторный анализ с учетом обоих факторов одновременно.

```
import statsmodels.api as sm
```

```
from statsmodels.formula.api import ols
```

Готовим фрейм данных

```
data_flat = []
```

```
for ch in channels:
```

```
    for reg in regions:
```

```
        data_flat.append((ch, reg, df.loc[ch, reg]))
```

```
df_flat = pd.DataFrame(data_flat, columns=["Channel", "Region", "Clients"])
```

Проводим ANOVA-анализ

```
model = ols('Clients ~ Channel + Region', data=df_flat).fit()
```

```
anova_table = sm.stats.anova_lm(model, typ=2)
```

```
print("\nДвухмерный ANOVA:")
```

```
print(anova_table)
```

Посмотрим, какое влияние оказывает каждый фактор и взаимодействие между ними.

Шаг 5: Интерпретация результатов

Изучим таблицу ANOVA:

F-statistics и соответствующие p-values покажут, являются ли различия значимыми.

Сумма квадратов SS скажет нам, какой вклад вносят отдельные факторы и их взаимодействие.

Если р-значения будут малыми (<0.05), это будет свидетельствовать о существовании значительной зависимости количества привлечённых клиентов от фактора (канала рекламы или региона).

Форма представления результата:

Представить выполненные задания в тетради для практических работ преподавателю.

Критерии оценки:

«Отлично» – умения сформированы, все задания выполнены правильно, без арифметических ошибок, решение оформлено аккуратно, с необходимыми обоснованиями.

«Хорошо» – некоторые умения сформированы недостаточно, все задания выполнены, некоторые виды заданий выполнены с ошибками. Безошибочно выполнено 80-89% всех заданий.

«Удовлетворительно» – необходимые умения в основном сформированы, большинство заданий выполнено, некоторые из выполненных заданий содержат ошибки. Безошибочно выполнено 70-79% всех заданий.

«Неудовлетворительно» – теоретическое содержание курса не освоено, необходимые умения не сформированы, выполненные учебные задания содержат грубые ошибки. Безошибочно выполнено менее 70% всех заданий.