

Государственной автономное образовательное учреждение «Центр
поддержки одаренных детей»
Структурное подразделение «Детский технопарк «Кванториум»
IT-квантум

Кейс «Развитие биопопуляции крыс»

Авторы:

Попова Светлана Сергеевна,
педагог дополнительного
образования

Марасанова Елена Вячеславовна,
педагог дополнительного
образования

Липецк, 2020

Кейс «Развитие биопопуляции крыс»

1. О кейсе

Решая задачу о развитии биопопуляции крыс, обучающийся познакомится с основами математической статистики: ее основными понятиями, научиться применять методы математической статистики на практике.

2. Текст кейса

Вы владеете небольшим зоомагазином, и вам подарили пару белых мышей разного пола для их разведения и дальнейшей продажи. Вы дождались пополнения их численности. И вот перед вами стоит выбор: продать пару крыс сейчас, или дождаться следующего прироста численности... В каком случае можно получить большую прибыль? На сколько больше будет эта прибыль?

3. Категория кейса: углубленный. Рассчитан на обучающихся 14-17 лет.

4. Место в структуре программы:

Данный кейс используется при изучении темы «Статистический анализ» образовательной программы по математике (углубленный модуль).

Количество учебных часов/занятий, на которые рассчитан кейс: 5.

Продолжительность занятия – 40 минут.

Учебно-тематическое планирование

Занятие 1	
Цель: освоить основные понятия и характеристики математической статистики	
Что делаем: Деление учащихся на группы. Знакомство с задачей кейса, планирование работы. Осуществление поиска необходимой информации с помощью информационных источников.	Компетенции: Soft: 4К-компетенции, умение генерировать идеи указанными методами, слушать и слышать собеседника; искать информацию в свободных источниках и структурировать ее Hard: освоение основных понятий математической статистики (мода, медиана, генеральная и выборочная средняя, показатели вариации, генеральная и выборочная дисперсия)
Занятие 2	
Цель: научиться выполнять расчеты величин, в соответствии с полученным вариантом набора данных; анализировать числовую информацию, представленную графически	

Что делаем: Расчет величин в соответствие с полученным вариантом данных. Построение диаграммы (графики) в программе MS Excel и анализ полученной информации	Компетенции: Soft: 4К-компетенции, умение слушать и слышать собеседника; умение грамотно письменно формулировать свои мысли Hard: умение строить графики, диаграммы и боксплоты для имеющихся числовых данных; умение анализировать информацию о числовом ряде при чтении графиков, диаграмм
Занятие 3	
Цель: научиться вычислять показатели вариации;	
Что делаем: Изучение основных показателей вариации с помощью источников информации, решение задач на вычисление. Анализ временных рядов, в том числе определение Фиббоначиевой последовательности	Компетенции: Soft: 4К-компетенции, умение грамотно письменно формулировать свои мысли Hard: освоение основных понятий математической статистики (показатели вариации, генеральная и выборочная дисперсия); умение анализировать информацию о числовом ряде при чтении графиков, диаграмм
Занятие 4	
Цель: научиться строить прогнозы на основании данных.	
Что делаем: решение задачи на построение прогнозов на основании данных. Проверка данных на наличие выбросов и на наличие тренда, Построение и анализ регрессионной модели.	Компетенции: Soft: 4К-компетенции, умение грамотно письменно формулировать свои мысли, опыт публичных выступлений Hard: построение прогнозов с помощью регрессионных моделей.
Занятие 5	
Цель: представить результаты решения кейса	
Что делаем: Создание презентаций и представление решений кейса.	Компетенции: Soft: командная работа, коммуникативность, основы ораторского искусства, опыт публичных выступлений, умение отвечать на вопросы, умение грамотно отстаивать свою точку зрения, умение оценивать себя. Hard: связывание математических понятий и окружающего мира, анализ полученных данных, формулирование статистических гипотез и их проверка

5. Минимально необходимый уровень входных компетенций: иметь представление о числовых и арифметических последовательностях; знание основ работы в программах Microsoft Office Excel и Wolframe Alpha.

6. Предполагаемые результаты обучающихся, формируемые навыки: Soft Skills: 4К критическое мышление, креативное мышление, коммуникация, кооперация; умение работать с информацией в свободном доступе; умение анализировать информацию, формулировать проблему и

строить гипотезы; навык анализа промежуточных результатов разработки; умение структурировано преподнести результаты собственной разработки; умение анализировать результаты других разработчиков. **Hard Skills:** представление о методах математической статистики, их использовании при решении задач обработки числовой информации; навыки математического моделирования реальных технических объектов и процессов в проектной деятельности; умение использовать приобретенные знания и умения в практической деятельности и повседневной жизни для анализа реальных числовых данных, представленных в виде диаграмм, графиков; для анализа информации статистического характера с использованием методов математической статистики.

7. Способ выявления образовательного результата

Представление результатов образовательной деятельности пройдет в форме публичной презентации решений кейса командами и последующих ответов выступающих на вопросы наставника и других команд.

8. Необходимые материалы и оборудование

Компьютер с установленным программным обеспечением и доступом к сети интернет.

Оборудование для презентации: проектор, экран.

9. Источники информации:

1. Балдин Е.М., Шипунов А.Б., Наглядная статистика. Используем R!, [Электронный ресурс]. URL: <https://cran.r-project.org/doc/contrib/Shipunov-rbook.pdf>.
2. Чубукова И.А. Data Mining: учебное пособие – 2-е изд., испр. – М.: Интернет-Университет Информационных Технологий; БИНОМ. Лаборатория знаний, 2008.

Дополнительные материалы:

1. Презентация «Основные понятия математической статистики»
2. Файл в формате .xlsx с исходными данными

Руководство наставника

Наставники знакомятся с заданием кейса. Чтобы ответить на его вопросы, необходимо решить поставленные задачи:

1. Построить математическую модель, отражающую увеличение численности крыс.
2. Провести анализ построенной модели.

Кейс состоит из 6 блоков.

На первом занятии из обучающихся необходимо сформировать группы по 3-5 человек.

В 1 блоке обучающиеся изучают основные характеристики математической статистики, используя рекомендуемые источники информации.

Во 2 блоке выполняются расчеты величин, в соответствии с полученным вариантом набора данных. Для этого преподаватель генерирует наборы данных из 5/10 элементов для каждого варианта (должны быть наборы как с четным, так и с нечетным количеством элементов), допускается повторение нескольких элементов.

В 3 блоке для построения диаграмм (графиков) преподаватель генерирует наборы данных для каждого варианта в виде таблиц в программе MS Excel, получая файл с данными (примеры таких таблиц в презентации «Основные понятия математической статистики»). Линейный график – данные требуется сгенерировать таким образом, чтобы на графике наблюдалось увеличение/уменьшение количества крыс. Точечная диаграмма – данные требуется сгенерировать таким образом, чтобы наблюдалось по 2-3 (или 4) кластера, для усложненных вариантов можно добавить спорные объекты, которые можно отнести сразу к 2-м.

В 4 блоке обучающиеся с помощью источников информации самостоятельно знакомятся с терминами.

В 5 блоке преподаватель рассчитывает для групп последовательные n элементов Фибоначиевой последовательности (начиная, например, с 5-10 элемента), и подготавливает последовательности, которые не являются Фибоначиевой последовательностью (например, сгенерировать

последовательность и заменить несколько элементов случайными числами, близкими к начальным). Таким образом, для каждого варианта будет 2 последовательности 1 - Ф, 2 - не Ф.

Например: Дан временной ряд a_1, a_2, \dots, a_n , где a_i - это численность крыс в i -й период, например, даны числа: 3,5,8,13,21 и 3,5,8,11,21).

Для проверки гипотезы потребуется самостоятельное изучение материалов о Фиббоначиевой последовательности, изучение формул, описывающих данную последовательность, понимать взаимосвязь между набором чисел (математической моделью) и реальным объектом изучения: как найти среднее значение прироста популяции для данного ряда? В какой месяц родилось максимальное/минимальное значение крыс?

В 6 блоке необходимо подготовить временные ряды (сгенерировать (минимум по 50 значений) выбрав математические модели или подготовить собранную Биоквантумом информацию. В конце занятия провести сравнение точности построенных прогнозов учащимися с сгенерированными.

После формирования представления о способах нахождения выбросов и эффективности использовании каждого метода, для построения математической модели, отражающей увеличение численности крыс и построения прогноза учащиеся приступают к последовательному выполнению действий: обработка данных, построение регрессионной модели, построение прогноза, анализ точности прогноза.

При ответе на вопрос: «Каких теоретических и практических знаний вам не хватает, чтобы воспользоваться выбранными вами способами?» обучающийся должен воспользоваться источниками информации, чтобы сформировать терминологический словарь тех определений, которые им не известны, но необходимы для решения задач.

Рекомендации по возможным проблемам в рамках реализации кейса

В ходе выполнения реализации кейса могут встретиться новые термины, что потребует самостоятельного изучения учащимися.

В случае невозможности выбора правильной модели или возникновении сложности при подготовке данных к работе допускается построение нескольких вариантов прогноза с последующим выбором наилучшего решения

Руководство для учащегося

Старт

Вы владеете небольшим зоомагазином, и вам подарили пару белых мышей разного пола для их разведения и дальнейшей продажи. Вы дождались пополнения их численности. И вот перед вами стоит выбор: продать пару крыс сейчас, или дождаться следующего прироста численности... В каком случае можно получить большую прибыль? На сколько больше будет эта прибыль?

Задачи проекта:

1. Построить математическую модель, отражающую увеличение численности крыс.
2. Провести анализ построенной модели.

Планирование

Чтобы спланировать работу, ответьте на вопросы:

1. Какая наука занимается прогнозированием?
2. Как выполняется расчет величин, в соответствии с полученным вариантом набора данных?
3. В каких программах можно производить построение математических моделей с использованием численных методов? И каким образом?
4. Как найти среднее, максимальное (минимальное) значение для ряда (последовательности) данных?
5. Как строятся прогнозы на основании математической модели?

Блок 1. Введение

Изучите основные характеристики математической статистики, используя рекомендуемые источники информации.

Математическая статистика изучает _____

Генеральная средняя – _____

Выборочное среднее – _____

Мода – _____

Медиана – _____

Характеристика	Формула
Генеральная средняя	
Выборочное среднее	
Мода	
Медиана	

Блок 2. Подготовка

Выполните расчеты величин, в соответствии с полученным вариантом набора данных.

Вариант № « ____ »

Дано:

Генеральная совокупность: _____

Выборка: _____

Решение

1) Генеральная средняя = _____

2) Выборочное среднее = _____

3) Мода:

Запишем в порядке возрастания:

Мода = _____

4) Медиана:

Запишем в порядке возрастания:

Количество элементов _____ (четное/нечетное?)

Медиана =

Ответ: 1) Генеральная средняя = _____;

2) Выборочное среднее = _____;

3) Мода = _____;

4) Медиана = _____.

Блок 3. Анализ

Согласно данным варианта постройте диаграммы (графики) в программе MS Excel и ответьте на поставленные задачи, заполнив пропуски:

Вариант № « ____ »	
Линейный график	Численность популяции: увеличивается/уменьшается/остаётся неизменной. Наименьший прирост численности был в _____ месяце. Наибольший прирост численности был в _____ месяце.
Точечная диаграмма	Крыс можно разделить на _____ групп, так как _____ _____ _____ В первую группу можно отнести крыс под номерами _____ Во вторую группу можно отнести крыс под номерами _____ _____ _____
Круговая диаграмма	По диаграмме видно, что: чаще всего встречаются крысы _____ цвета, а реже _____ цвета. Крыс с окрасом _____ и _____ приблизительно поровну/ таких не наблюдается
Столбчатая диаграмма	По диаграмме видно, что: чаще всего встречаются крысы _____ цвета, а реже _____ цвета. Крыс с окрасом _____ и _____ приблизительно поровну/ таких не наблюдается

Блок 4. Показатели вариации

Выпишите основные понятия и определения

Основными показателями, характеризующими вариацию, являются:

- _____
- _____
- _____
- _____

	Определение	Формула, метод нахождения
<i>Размах вариации</i>		
<i>Дисперсия</i>		

Блок 5. Последовательности

Дан временной ряд a_1, a_2, \dots, a_n , где a_i - это численность крыс в i -й период (например, период – месяц, даны числа: $3, 5, 8, 13, 21$ и $3, 5, 8, 11, 21$).

- Как найти среднее значение прироста популяции для данного ряда?
- В какой месяц родилось максимальное/минимальное значение крыс?

Вариант № « ____ »

Дано: _____

Последовательность № 1 – последовательность Фиббоначи

Последовательность № 2

Доказать, что *Последовательность №2* является Фиббоначиевой последовательностью.

Решение

Гипотеза: данная последовательность является Фиббоначиевой последовательностью.

Числа Фиббоначи это _____

Ими можно описать _____

Формула для вычисления i -го элемента последовательности:

Где: _____

Проверим *Последовательность №1*:

Проверим *Последовательность №2*:

Итак, на основании расчетов, можно сделать вывод, что _____

Блок 6. Решение задачи

Подумайте над предложенной задачей

На основании генерированных данных: Построить прогнозы на основании данных: зная численность крыс за последние n месяцев (т.е. имеется временной ряд, согласно варианту группы) сказать какова будет численность крыс на $n+k$ месяц.

На основании реальных данных: В ходе наблюдений за численностью была собрана информация об уходе за питомцами в каждом периоде времени, требуется выявить самые благоприятные/неблагоприятные условия для популяции, а также выдвинуть гипотезы о влиянии внешних условий на популяцию крыс.

Уточняющие вопросы для построения прогнозов с помощью математических моделей:

1. Для чего проводится анализ временного ряда? Что такое выброс, удаление выбросов, сглаживание временного ряда? Зачем удаляют выбросы/сглаживают временные ряды? Нужно ли проводить подготовку данных? Будут ли отличаться прогнозы, построенные по исходным данным, сглаженным, данным после удаления выбросов?

2. Тренд, сезонная и циклическая компонента: что это, как определить по графику, возможно ли одновременное наличие тренда и сезонности? Можно и нужно ли избавляться от тренда, сезонной и циклической компоненты ряда? На что это может повлиять?

3. Что такое регрессионная модель, какие бывают регрессионные модели, как выбрать подходящую модель? Как построить регрессионную модель?

4. Как определить точность прогноза, что влияет на точность? Что такое средняя ошибка, средняя абсолютная ошибка, сумма квадратов ошибок/среднеквадратическая, относительная ошибка? Что показывают эти ошибки. Как сравниваются эти величины, какие выводы можно сделать после их сравнения?

5. На что влияет качество построенной модели и как его улучшить?

Проверьте данные на наличие выбросов, выберите способ по удалению или сглаживанию выбросов.

Какие есть способы нахождения выбросов, и какой способ лучше подойдет:

1. Удаление выбросов вручную.
2. С помощью Excel.
3. Использование математических пакетов.
4. Использование нейросетей.

Найденные выбросы стоит удалить или провести сглаживание ряда? Какой способ даст лучший результат?

Проверьте данные на наличие тренда, сезонной и циклической компоненты.

В данном временном ряду (наблюдается/не наблюдается)

Какая регрессионная модель подходит данному набору данных?

Постройте математическую модель, отражающую увеличение численности крыс. Постройте прогнозы на основании исходных данных: зная численность крыс за последние n месяцев (т.е. имеется временной ряд) дайте прогноз численности крыс на $n+k$ месяц.

Постройте регрессионные модели. Установите период, горизонту, интервал прогнозирования. Постройте прогнозы по этим моделям и проведите сравнения полученных результатов. Выберите наиболее удачный прогноз, оценив точность прогноза, для этого сравните:

По данным наблюдений: численность и информация об уходе за питомцами в каждом периоде времени, используя графики, выявить самые благоприятные/неблагоприятные условия для популяции, выдвинуть гипотезы о влиянии внешних условий на популяцию крыс.

Рефлексия

- Как можно доработать проект?
- Какие программы существуют для автоматизации расчетов?

Что если?

- В качестве прогнозируемых данных взять что-то другое? Что можно попытаться прогнозировать?
- Выполнить прогноз данных без удаления выбросов, без сглаживания и без удаления тренда, сезонности?
- Регрессионная модель выбрана неверно?