



**АКАДЕМИЯ**  
искусственного интеллекта  
для школьников

#AIAcademy

**План и методические рекомендации по проведению урока  
по теме «Статистический анализ данных. Корреляционный анализ».**

**Основная идея:** привлечение внимания обучающихся к одной из перспективных областей профессиональной деятельности в ИТ-индустрии — **науке о данных**.

**Цель урока:**

- сформировать у учащихся представление о статистическом анализе данных и его роль в интерпретации полученных результатов;
- научить использовать программу Microsoft Excel для работы с табличными данными.

**Задачи урока:**

- расширить представление у учащихся о табличных данных и методах работы с ними;
- сформировать навыки статистического анализа данных;

- сформировать навыки использования базовых средств табличного процессора Microsoft Excel для статистического анализа данных (визуализация данных с помощью диаграммы разброса и расчета коэффициента корреляции (коэффициент корреляции Пирсона));
- сформировать метапредметные результаты, связанные с:
  - умением работать с информацией, анализировать и структурировать полученные знания и синтезировать новые, устанавливать причинно-следственные связи;
  - ставить цель и находить оптимальные способы ее достижения, проводить ситуационную и ретроспективную рефлексия, участвуя в подведении итогов отдельных этапов и урока в целом;
  - умением вступать в диалог и вести его, взаимодействовать в команде.

### **Планируемые результаты:**

личностные:

- готовность и способность к осознанному выбору и построению дальнейшей индивидуальной траектории образования благодаря ориентации в мире профессий и профессиональных предпочтений с учётом устойчивых познавательных интересов;

метапредметные умения и опыт:

- выдвигать версии решения проблемы, формулировать гипотезы, предвосхищать конечный результат;
- определять цель деятельности на основе определенной проблемы и существующих возможностей;
- строить рассуждение на основе сравнения данных, выделяя общие признаки;
- излагать полученную информацию, интерпретируя ее в контексте решаемой задачи;
- выделять информационный аспект задачи, оперировать данными, использовать модель решения задачи и др.

Тема урока **«Статистический анализ данных. Корреляционный анализ»** раскрывается в два академических часа и состоит из вводной теоретической части и практической. Сопровождается

презентацией в рамках содержательной части урока, а также подкрепляется проектами в Microsoft Excel для учителя (с заданиями, комментариями и ответами) и ученика (с заданиями, комментариями, но без ответов) в практической части урока.

**Практическая работа** включает алгоритм, справочный материал и пошаговые инструкции по выполнению заданий. Содержательное наполнение заданий основано на реальной практической задаче (сюжет и данные, используемые в задаче, вымышленные; все совпадения случайны). Кроме того, учитель имеет доступ к файлам с исходными данными задач, что позволит значительно оптимизировать процесс, не тратя времени на создание массивов данных (ввод информации). Подробное описание действий в инструкциях («выделите, откройте» и т.п.) позволит включить в работу даже школьников, не имеющих опыта работы с табличным процессором.

Практическую работу можно выполнять как во фронтальном (синхронном) режиме, объясняя рекомендуемые приемы работы, так и организуя индивидуальную работу школьников в удобном для них темпе.

**Рекомендации по распределению времени:** учитель может распределить учебное время следующим образом: теория – 0,5 часа, практика – 1,5 часа или теория и практика по 1 часу, ориентируясь на уровень подготовки учебной группы. Если есть дополнительное время, то на практическую часть можно заложить и больше времени (на усмотрение учителя).

**Важно:** на уроке можно выполнять не все предложенные из практикума. Все зависит от подготовленности группы. Учитель может самостоятельно формировать набор заданий для урока. Часть разобрать и выполнить на уроке, а другую часть дать учащимся для домашнего выполнения.

Также можно выбрать другую стратегию. Увеличить количество часов для практической части до 2 часов.

**Для проведения урока учителю понадобится:**

- компьютер с проекционным оборудованием;
- компьютерный класс с выходом в интернет и с установленным пакетом Microsoft Excel;
- опорная презентация;
- загруженный файл с заданиями для учеников на каждый учебный компьютер.

Этап урока	Номер слайда	Комментарии для учителя	Дополнительные материалы
Мотивация к учебной деятельности	---	<p>Учителю следует создать условия, чтобы ученики внутренне собрались, подготовились и нацелились на «покорение новых вершин».</p> <p><b>Вступительное слово.</b></p> <p><i>Вы научились работать с табличными данными: группировать, сортировать, рассчитывать показатели, визуализировать и даже проводить первичный анализ. Для работы использовали приложение Microsoft Excel. Но чтобы мы не делали в процессе работы с данными, требуется подтверждать математически все полученные результаты. Как же это делать? Давайте разбираться.</i></p> <p><b>Или просмотр видео ролика.</b></p>	<p><b>Подробнее:</b></p> <p>Коэффициент корреляции - Борис Миркин. ПостНаука. [Электронный ресурс] – Режим доступа: <a href="https://www.youtube.com/watch?v=cd-7k4LNKTO">https://www.youtube.com/watch?v=cd-7k4LNKTO</a></p> <p>(Математик Борис Миркин о регрессе к среднему значению, ложной корреляции и адаптивной системе).</p>

Актуализация знаний	1	<p>После просмотра ролика или вводной лекции учителя, следует обратить внимание на то, что корреляционный анализ позволяет выявить линейную связь между признаками и тесно связан с регрессионным анализом.</p> <p><b>Учитель может задать вопросы для обсуждения:</b></p> <p><i>В школьных предметах можно увидеть корреляцию, хотя может она для вас и не очевидна (связь между величинами, объектами и т.п.). Подумайте и приведите пример.</i></p> <p><i>математика:</i></p> <ul style="list-style-type: none"> <li>● <b>путь — время</b> (график зависимости этих двух величин есть <b>скорость</b>);</li> <li>● <b>ширина — длина</b> (график зависимости этих двух величин есть <b>площадь</b>);</li> </ul> <p><i>физика:</i></p> <ul style="list-style-type: none"> <li>● <b>сила тока</b> в участке цепи — <b>сопротивление</b> участка (график зависимости двух величин есть <b>напряжение</b> на этом участке);</li> </ul> <p><i>и т.п.</i></p>	<p><b>Рекомендация для учителя:</b></p> <ul style="list-style-type: none"> <li>● рекомендуемая презентация: <i>6_Статистический_анализ_данных_Корреляционный.pptx</i>;</li> <li>● организовать групповую работу в классе и предоставить самостоятельно изучить поставленные вопросы для обсуждения (у учащихся должен быть компьютер с выходом в Интернет).</li> </ul>
---------------------	---	--	--

		<p><i>Как можно увидеть данную связь? Надо отметить, что не любая диаграмма может нам показать наличие связи между величинами. Для этого используют диаграмму разброса (рассеяния), которая помогает определить примерный характер связи между двумя величинами при ее наличии. Сегодня на уроке мы будем строить такой тип диаграмм и анализировать связи.</i></p>	
<p>Целеполагание, постановка проблемы</p>	<p><b>2-11</b></p>	<p>После обсуждения учителю следует обратить внимание на то, что ошибки визуализации данных возникают при неправильном масштабировании. Из-за этого несущественные различия кажутся существенными. Это и вводит в заблуждение. Учитель формирует вместе с учащимися основную цель урока:</p> <p><i>«Сегодня нам предстоит разобраться, как при анализе данных помогает правильно интерпретировать результаты корреляционный анализ. Также понять, в каких случаях корреляционный анализ не помогает, а даже мешает делать правильные выводы».</i></p>	<p><b>Рекомендация для учителя:</b></p> <ul style="list-style-type: none"> <li>● рекомендуемая презентация: <i>6_Статистический_анализ_данных_Корреляционный.pptx.</i></li> </ul> <p><b>Подробнее:</b></p>

	<p><b>Учитель раскрывает тему урока. Транслирует новый материал.</b></p> <p><i>В научных исследованиях часто возникает необходимость в нахождении связи между результативными и факторными переменными (урожайностью какой-либо культуры и количеством осадков, ростом и весом человека в однородных группах по полу и возрасту, частотой пульса и температурой тела и т.д.). Вторые представляют собой признаки, способствующие изменению таковых, связанных с ними (первыми).</i></p> <p><i>Существует множество определений термина. Можно сказать, что корреляционный анализ — это метод, применяющийся с целью проверки гипотезы о статистической значимости двух и более переменных, если исследователь их может измерять, но не изменять.</i></p> <p><i>Есть и другие определения рассматриваемого понятия. Корреляционный анализ — это метод обработки статистических данных, заключающийся в изучении коэффициентов корреляции между переменными. При этом сравниваются коэффициенты корреляции между одной парой или</i></p>	<p>Коэффициент корреляции. StatSoft. Портал Знаний: глобальный интеллектуальный ресурс.[Электронный ресурс] – Режим доступа: <a href="http://statistica.ru/theory/koeffitsient-korrelyatsii/">http://statistica.ru/theory/koeffitsient-korrelyatsii/</a></p>
--	--	--

		<p>множеством пар признаков, для установления между ними статистических взаимосвязей. Корреляционный анализ — это метод по изучению статистической зависимости между случайными величинами с не обязательным наличием строгого функционального характера, при которой динамика одной случайной величины приводит к динамике математического ожидания другой.</p>	
	7	<p><b>Разбор кейс-задач на слайдах 7-9.</b></p> <p>Обратите внимание, что при построении графика мы можем увидеть изменение пути, а время сливается в прямую. Этот эффект из-за разного масштаба исходных данных.</p>	
	8	<p>При анализе второй кейс-задачи, также можно увидеть эффект из-за разности масштаба исходных данных, но четко отслеживается тенденция падения энергии.</p> <p>При обсуждении, что показывают параметры прямой, надо подвести учащихся к следующим идеям:</p>	

	9	<ul style="list-style-type: none"> <li>● когда собираемся в поход, есть первоначальная жизненная энергия - 2540,3;</li> <li>● с каждым часом энергия падает, то есть энергия уменьшается на 163,45 с каждым часом.</li> </ul> <p>При обсуждении третьей задачи, надо учащихся подвести к идее, что данные показатели никак не связаны. Даже если папа и мама работают в одной фирме, они выполняют разную работу. У них разная квалификация, образование и т.д.</p> <p>Одним словом, эти данные не описывают никакого общего события.</p> <p>Учитель может продемонстрировать все области диаграммы в MS Excel.</p>	
Поиск путей решения проблемы	---	<p>Ребята, вы уже знаете, что для аналитика важно уметь работать с данными (собирать, сохранять, извлекать, обрабатывать их), также получать новые расчетные данные для дальнейших выводов и прогнозирования.</p>	<p><b>Рекомендация для учителя:</b></p> <ul style="list-style-type: none"> <li>● рекомендуемый раздаточный материал для учителя: 6_Статистический_анализ_данных_Корреляционный (для учителя).xlsx;</li> <li>● рекомендуемый раздаточный материал для ученика:</li> </ul>

		<p><i>Но очень важно, чтобы данные были корректны и анализ подтверждался статистическими методами. Это позволит избегать ошибок в исследовании и подтасовки данных.</i></p> <p><i>Мы с вами будем работать в приложении Microsoft Excel. Будем данные визуализировать с помощью диаграммы разброса и определять наличие связи между признаками. А также подтверждать наличие связи с помощью коэффициента Пирсона. Для этого мы будем использовать встроенные функции. Переходим к практической части.</i></p> <p><b>Консультирует. Демонстрирует работу в приложении Microsoft Excel.</b></p> <p>Учитель повторяет с учащимся интерфейс приложения Microsoft Excel (книга, лист, столбцы, строки, ячейка, адресация (имя ячейки), поле имени, строка формул). Для получения новой информации (обогащения данных) требуется не только использовать формулы и относительную адресацию, но и встроенные функции Microsoft Excel.</p> <p><b>Учитель должен сделать акцент</b> на том, что в рамках урока потребуются команды, находящиеся</p>	<p><i>6_Статистический_анализ_данных_Корреляционный.xlsx</i></p> <p><i>Учащиеся могут работать одновременно с учителем. Знакомиться с интерфейсом. Если учащиеся знакомы с электронной таблицей, то данный этап на уроке можно сократить как по времени, так и по содержанию.</i></p> <p>Планирование путей достижения намеченной цели. Осуществление учебных действий по плану. Решение практических задач.</p>
--	--	--	--

		на ленте «Главная» (это упростит поиск необходимых команд для решения поставленной задачи). При работе с диаграммами будет доступна вспомогательная панель « <b>Конструктор</b> ».	
Решение проблемы	---	<p><i>Вам предстоит самостоятельно познакомиться с коэффициентом корреляции Пирсона и разобраться в его информативности. Решить несколько кейс-задач. Давайте познакомимся с содержанием практической работы.</i></p> <p><i>У каждого этапа есть инструкция и справочный материал, которые помогут решить поставленную задачу. Также вы можете использовать и внешние источники информации (Интернет).</i></p> <p><i>Напоминаю, что все команды, которые потребуются для решения задачи, находится на ленте «Главная». Можете пользоваться справочным материалом Microsoft Excel.</i></p> <p><i>При работе с диаграммами появляется дополнительная панель «<b>Конструктор</b>».</i></p> <p><b>Задача 1. Исследование рисков от курения.</b></p> <p><b>Задача 2. Исследование профилактическим мер вакцинации.</b></p>	<p><b>Рекомендации для учителя:</b> в зависимости от уровня базовой подготовки учащихся, можно организовать фронтальную работу (осуществлять контроль и объяснять методы решения поставленной задачи) или индивидуальную, в удобном для каждого темпе с внешним контролем.</p> <p>Чтобы учащиеся могли самостоятельно осваивать материал, необходимо им показать, как использовать встроенный справочный материал Microsoft Excel (при вызове встроенной функции можно раскрыть справку, также при наведение курсора на команду появляются подсказки).</p>

		<p><b>Задача 3. Исследования здоровья и благополучия подростков.</b></p> <p><b>Задача 4. Исследование качества образования.</b></p> <p>Не спешите. Прочитайте внимательно описание данных. В задачах предложены реальные данные, которые ранее были использованы в исследованиях. В каждой задаче есть содержание исследования и описание ожидаемых результатов. Необходимо подтвердить или опровергнуть наличие тех или иных связей. Сделать итоговый вывод.</p> <p>Выполняя задание шаг за шагом, вы увидите, как анализ становится более глубоким и информативным. Подумайте, почему так получается?</p> <p>Вам потребуется переходить на другую вкладку «<b>Данные</b>». Будьте внимательны.</p> <p>Желаю вам удачи. Если у вас что-то не получается, я вам обязательно помогу.</p>	
--	--	---	--

		<p>Учитель должен сделать акцент на том, что задания нужно выполнять последовательно, это позволит им разобраться с встроенными функциями, которые агрегируют данные по одному или нескольким параметрам и после рассчитывают статистические показатели.</p> <p>Надо обратить внимание учащихся, что данные по исследованию постепенно дополняются. Это дает возможность получать новые показатели и делать дополнительный анализ.</p> <p><b>Консультирует</b></p> <p>В предложенном практическом задании исходные данные произвольные.</p>	
<p>Самостоятельная работа с использованием полученных знаний</p>	---	<p><b>Направляет, советует, консультирует.</b></p>	<p><b>Рекомендации для учителя:</b></p> <p><i>организовать самостоятельную работу таким образом, что если ребенок не знает как выполнить задание, чтобы он смог найти информацию или пример решения подобной задачи (в книге, справке, Интернете и т.п.) Это позволит ему получить навык поиска</i></p>

			необходимой информации в различных источниках.
Систематизация знаний	---	<p><b>Консультирует, направляет.</b></p> <p><i>При проведении корреляционного анализа необходимо учитывать, что его можно провести по отношению к любой совокупности признаков, зачастую абсурдных по отношению друг к другу. Порой они не имеют никакой причинной связи друг с другом.</i></p> <p>Просмотр видео, в котором показывают абсурдные отношения между несвязанными признаками, и как это в дальнейшем влияет на конечный результат исследования.</p>	<p><b>Рекомендации для учителя:</b></p> <p><i>создать условия учащимся, для работы по выявлению связи изученной на уроке темы с изученным ранее материалом, связи с жизнью.</i></p> <p>Корреляция <math>\neq</math> Каузальность.  AsapSCIENCE [Электронный ресурс]  – Режим доступа:  <a href="https://www.youtube.com/watch?v=un00kHP8IxE">https://www.youtube.com/watch?v=un00kHP8IxE</a></p>
Объяснение домашнего задания	---	<p><b>Домашнее задание после первого урока:</b></p> <p><i>доклад на тему: «Статистические методы анализа данных;</i></p> <p><i>Ошибки в статистическом анализе данных».</i></p> <p><b>Домашнее задание после второго урока:</b></p>	<p>У учеников должна быть возможность выбора домашнего задания в соответствии со своими предпочтениями.</p>

		<p>найти примеры зависимостей из школьных предметов; собрать данные, рассчитать коэффициент корреляции (не более 10 записей) и построить диаграмму разброса.</p> <p>Также можно дать задание на самостоятельный поиск исходных данных (или получение их), с последующим анализом.</p>	Необходимо наличие заданий разного уровня сложности.
Оценивание	---	<b>Консультирует, обосновывает оценки.</b>	Учащиеся самостоятельно оценивают работу на (самооценка, взаимооценивание результатов работы одноклассников).
Рефлексия учебной деятельности	---	Благодарит учеников за урок.	Учащиеся называют тему урока, его этапы, перечисляют виды деятельности на каждом этапе, определяют предметное содержание. Делятся мнением о своей работе на уроке.