МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ

ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ им. А. И. ГЕРЦЕНА»

**Институт информационных технологий и технологического образования**

**кафедра информационных технологий и электронного обучения**

Основная профессиональная образовательная программа

Направление подготовки 09.03.01 Информатика и вычислительная техника

Направленность (профиль) «Технологии разработки программного обеспечения»

форма обучения – очная

**Курсовая работа**

по дисциплине «Пакетам прикладных программ»

Анализ опросных данных:

Воздействие контента на психическое здоровье

Обучающегося 3 курса

Зухир Амиры Саидовны

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Руководитель:

д.п.н, профессор

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Власова Е. З.

«\_\_\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2024 г.

Санкт-Петербург

2024

ОГЛАВЛЕНИЕ

стр.

Введение 3

Теоретическая часть: 4

1.1 Первый закон Ньютона 4

1.2 Равномерное движение 4

1.3 Равноускоренное движение 5

1.3 Свободное падение 8

Практическая часть 10

Заключение 14

Литература 15

ВВЕДЕНИЕ

Современное общество переживает период значительных изменений, вызванных активным развитием информационных технологий и интернета. С появлением новых социальных платформ и возросшим объемом потребляемого контента становится всё более актуальным вопрос о воздействии этого контента на психологическое здоровье человека. Несмотря на множество положительных аспектов цифровой эры, существует опасность, что избыточное воздействие разнообразных медийных форматов может оказывать негативное воздействие на эмоциональное, психическое и физическое состояние человека.

Данная курсовая работа посвящена анализу опросных данных, направленных на выявление воздействия контента на психическое здоровье. Задачей исследования является выявление тенденций в восприятии и воздействии различных видов контента, а также оценка степени и характера возможных негативных последствий на психическое состояние человека.

Анализ опросных данных представляет собой важный метод исследования, который позволяет получить количественные и качественные данные о восприятии контента, а также выявить возможные связи между типами контента, различными социальными сетями и психическим здоровьем человека.

1 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ  
 1.1 Корреляционный анализ:

Корреляционный анализ — это метод по изучению статистической зависимости между случайными величинами с необязательным наличием строгого функционального характера, при которой динамика одной случайной величины приводит к динамике математического ожидания другой.

Ложная корреляция — это явление, при котором может наблюдаться статистическая связь между двумя переменными, не имеющими объективной причинной связи друг с другом. При проведении корреляционного анализа необходимо осознавать, что статистическая взаимосвязь может возникнуть случайно или из-за наличия третьего фактора, не имеющего непосредственного отношения к рассматриваемым переменным. Такие корреляции могут ввести исследователя в заблуждение, если не учитывать контекст и особенности данных. Однако, важно помнить, что статистическая корреляция не всегда указывает на наличие причинно-следственной связи между переменными.

Основные понятия в корреляционном анализе:

1. Корреляция: это степень взаимосвязи между двумя переменными. Корреляция может быть положительной, отрицательной или равной нулю. Положительная корреляция указывает на то, что увеличение одной переменной сопровождается увеличением другой. Отрицательная корреляция означает, что увеличение одной переменной связано с уменьшением другой. Корреляция равная нулю говорит о том, что между переменными нет статистической связи.
2. Коэффициент корреляции: это числовая мера, которая выражает степень корреляции между переменными. Коэффициент корреляции принимает значения от -1 до 1. Значение -1 указывает на полную отрицательную корреляцию, 1 - на положительную, а 0 - на отсутствие корреляции.
3. Корреляционная матрица: это таблица, в которой показаны коэффициенты корреляции между всеми парами переменных в наборе данных.

Корреляционный анализ широко применяется в различных областях, таких как экономика, психология, медицина и другие, для выявления связей между переменными и более глубокого понимания структуры данных.

1.1.1 Коэффициент корреляции Спирмена:

Коэффициент корреляции Спирмена - это статистическая мера, используемая для оценки силы и направления монотонной (не обязательно линейной) связи между двумя ранжированными переменными. Этот метод основан на рангах значений переменных вместо их фактических числовых значений.

Процесс вычисления коэффициента корреляции Спирмена включает в себя следующие шаги:

1. Ранжирование значений каждой переменной по возрастанию. Присвоение каждому значению его ранга. Ранг представляет собой порядковый номер. Далее, вычисление разностей рангов для соответствующих пар значений переменных.
2. Квадратирование этих разностей и суммирование квадратов разностей рангов.
3. Применение формулы (1) для вычисления коэффициента корреляции Спирмена.

, (1)

где: d – разность рангов, n – количество признаков, участвовавших в ранжировании.

Коэффициент корреляции Спирмена принимает значения от -1 до 1. Значение -1 указывает на полную обратную монотонную связь, 1 - на положительную монотонную связь, а 0 - на отсутствие монотонной связи. Данный коэффициент часто используется в случаях, когда данные не соответствуют предположению о нормальном распределении или когда связь между переменными не является линейной.

1.1.2 Коэффициент V Крамера:

Коэффициент V Крамера является мерой силы и направления ассоциации между двумя категориальными переменными, измеренной в кросс-таблице. Кросс-таблица представляет собой двумерную таблицу, где каждая ячейка показывает количество наблюдений, попавших в конкретную комбинацию категорий обеих переменных.

Процесс вычисления коэффициента V Крамера начинается с применения критерия хи-квадрат для оценки статистической значимости связи между двумя переменными. Критерий хи-квадрат проверяет гипотезу о том, что две категориальные переменные независимы.

Формула для вычисления коэффициента V Крамера (2).

(2)

где: – значение статистики хи-квадрат, полученное при анализе кросс-таблицы,

– общее количество наблюдений,

– количество категорий (уровней) в одной переменной,

– количество категорий (уровней) в другой переменной.

Коэффициент V Крамера лежит в диапазоне от 0 до 1. Значение 0 означает отсутствие ассоциации между переменными, а значение 1 указывает на полную ассоциацию. Чем ближе значение V Крамера к 1, тем сильнее связь между переменными.

Этот коэффициент часто используется в исследованиях социальных наук, маркетинге, медицинских исследованиях и других областях, где необходимо оценить степень ассоциации между двумя категориальными переменными.

1.2 Тест Хи-квадрат:

Тест хи-квадрат — статистический метод, используемый для оценки статистической значимости ассоциаций между категориальными переменными в виде таблиц сопряженности (кросс-таблиц). Он позволяет проверить гипотезу о том, что наблюдаемое распределение частот в кросс-таблице не отличается от ожидаемого распределения при предположении независимости переменных.

Процесс теста хи-квадрат включает в себя следующие шаги:

1. Постановка гипотез:

Нулевая гипотеза (H0): нет статистически значимой связи между переменными (независимость).

Альтернативная гипотеза (H1): существует статистически значимая связь между переменными (зависимость).

2. Построение кросс-таблицы: данные разбиваются на категории, создавая кросс-таблицу, где значения ячеек представляют собой частоты наблюдений для каждой комбинации категорий.

3. Расчет ожидаемых частот: ожидаемые частоты вычисляются на основе предположения независимости переменных. Это делается путем умножения сумм частот по строкам и столбцам на соответствующие вероятности.

4. Вычисление статистики хи-квадрат: вычисляется статистика хи-квадрат, которая представляет собой сумму квадратов отклонений между наблюдаемыми и ожидаемыми частотами, нормированных на ожидаемые частоты.

5. Определение степеней свободы: определяются степени свободы для теста, которые зависят от размеров кросс-таблицы.

6. Определение критического значения и принятие решения: сравнивается значение статистики хи-квадрат с критическим значением из таблицы распределения хи-квадрат. Если статистика хи-квадрат превышает критическое значение, то нулевая гипотеза отклоняется в пользу альтернативной.

Тест хи-квадрат часто используется в медицинских исследованиях, социологии, маркетинге и других областях, где требуется оценка влияния одной переменной на другую в виде категорий.

2 ПРАКТИЧЕСКАЯ ЧАСТЬ

Для проведения исследования, был создан специальный опрос и распространён среди 50 человек. Были заданы следующие вопросы:

1. Какое среднее количество часов в день вы обычно проводите в социальных сетях?
2. В каких социальных сетях Вы сидите?
3. Делаете ли Вы "детокс" от социальных сетей? Хотя бы на несколько часов
4. Какой контент Вы чаще всего потребляете?
5. Каково ваше самочувствие после проведенного времени в социальных сетях?
6. Хотели бы Вы проводить меньше времени в соц.сетях ?
7. Бывали ли случаи, когда вы испытывали стресс из-за событий в социальных сетях?

С помощью языка программирования Python и IDE VScode были построены гистограммы и круговые диаграммы результатов ответов представленные на рисунках 1-7.

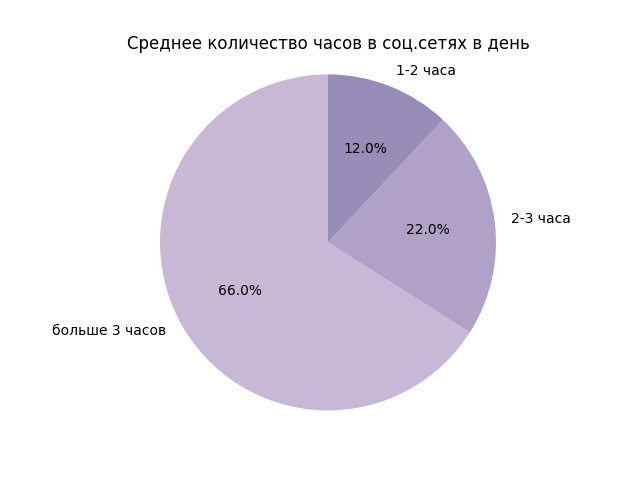


Рис.1:Круговая диаграмма среднего кол-ва часов проведенных в соц.сетях

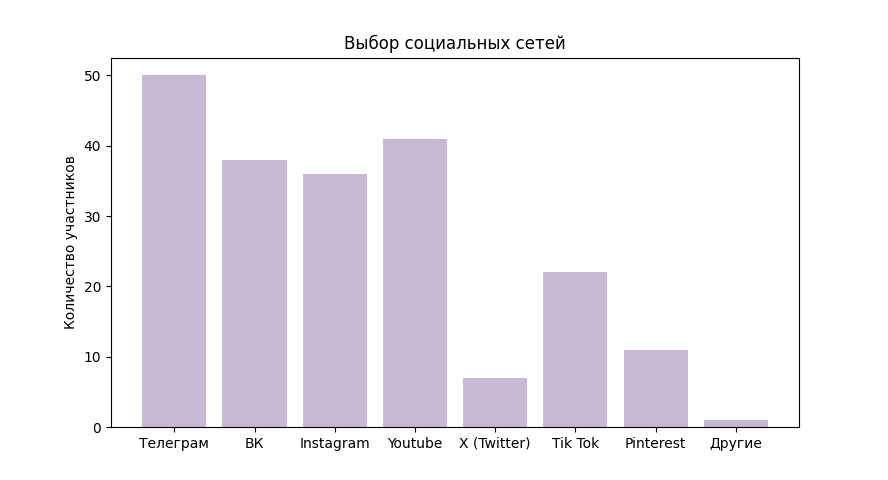


Рис.2: Выбор соц.сетей участниками опроса.

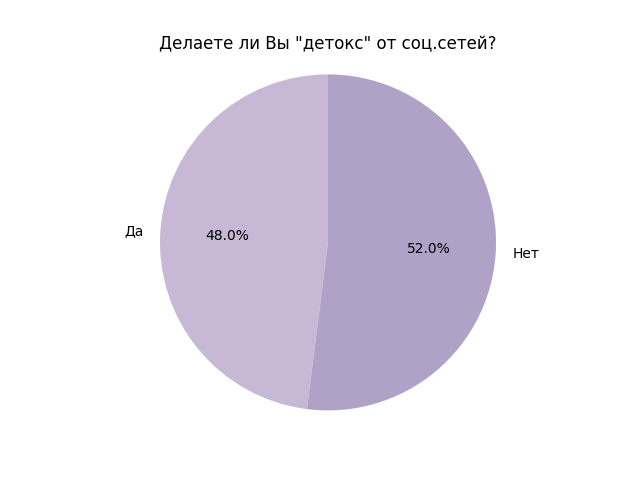


Рис.3: Процент опрошенных, выполняющих диджитал детокс (да/нет)



Рис.4: Выбор потребляемого контента.

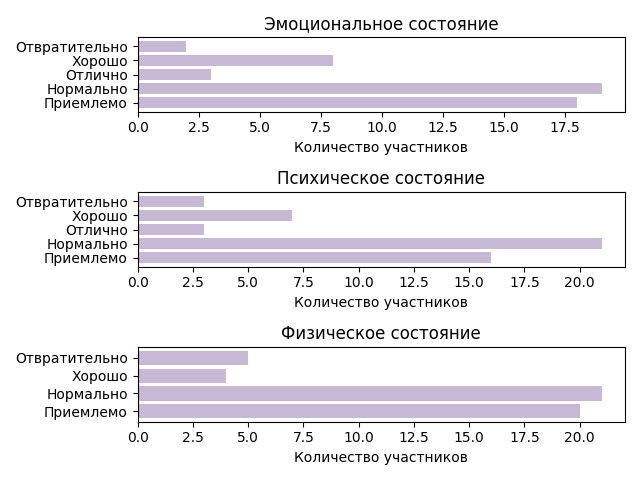


Рис.5: Эмоциональное, психическое и физическое состояния после времени, проведенного в социальных сетях.



Рис.6: Желание сократить время в социальных сетях.

Рис.7: Стресс из-за событий в социальных сетях.

Исследуем, есть ли статистически значимая связь между желанием уменьшить время в социальных сетях и опытом стресса из-за событий в социальных сетях.

Для анализа связи между желанием уменьшить время в социальных сетях и опытом стресса из-за событий в социальных сетях, воспользуемся тестом Хи-квадрат.

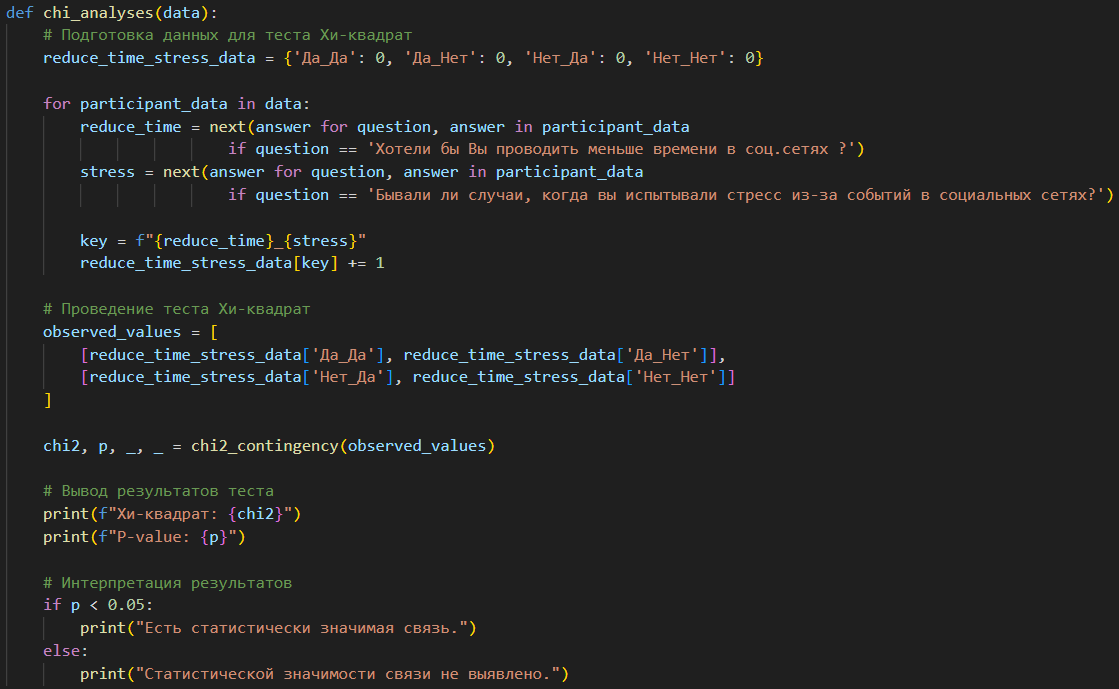


Рис.8: Код выполнения теста Хи-квадрат.

На рисунке 8 представлен код, в нем используем библиотеку scipy, которая предназначена для выполнения научных и инженерных вычислений.

В этом коде chi2\_contingency используется для выполнения теста Хи-квадрат на категориальных данных. Результаты теста включают значение Хи-квадрат и p-значение. Если p-значение меньше выбранного уровня значимости (обычно 0.05), то можно отклонить нулевую гипотезу и считать, что существует статистически значимая связь.



Рис.10: результат выполнения функции

На рис.10 виден результат выполнения кода для вычисления теста хи-квадрата.

Значение Хи-квадрата равно 0.0, что указывает на то, что нет статистических различий между наблюдаемыми и ожидаемыми частотами. Это может произойти, когда все наблюдаемые значения точно соответствуют ожидаемым.

Значение P-value равно 1.0, говорит о том, что нет статистической значимости и нельзя отклонить нулевую гипотезу.

Исходя из этого, можно сказать, что данные не представляют статистически значимых доказательств для отклонения нулевой гипотезы в гипотезы в контексте теста хи-квадрат.