Институт информационных технологий

Кафедра: Математическое и программное обеспечение ЭВМ

Дисциплина: Основы Data Science

ЛАБОРАТОРНАЯ РАБОТА № 3

Тема: Описание данных. Статистический вывод

Выполнил:

студент гр. 1ПИб-02-2оп-23

Кринкин Олег Алексеевич

Проверил:

Юдина Ольга Вадимовна

# 1. Задание

***Цель***: Знакомство с этапом понимание данных стандарта CRISP-DM.

Задача этапа – найти, описать основные закономерности, которые содержатся в данных и статистически их подтвердить, попытаться выявить связи между этими данными, фактами, знаниями, получить что-то новое, понять, что они могут дать.

***Задание***: Получив первоначальное представление о данных, рассмотрите закономерности, присущие данным.

Для задачи, которую вы сформулировали в предыдущих работах:

Подумайте, какие статистические гипотезы могут подтвердить ил опровергнуть ваши предположения.

Сформулируйте гипотезы о равенстве выборочных средних и долей категориальных, разделяя переменные. Используйте закономерности, особенности, которые вы смогли обнаружить в предыдущих работах. Проследите правильность формулировок гипотез *H0* и *H1*. Проведите испытание, используя соответствующие критерии.

Для того, чтобы сформулировать полезные гипотезы: Подумайте, что вы хотите выяснить, проверить. Можно использовать фильтры разного уровня. Например, чтобы сравнить учебные успехи студентов, проводящих разное время за играми– отфильтруйте по этому признаку показатели успеваемости, посчитайте среднее и доверетельные интервалы. Если доверительные интервалы пересекаются – сформулируйте гипотезы. Результат позволит сделать заключение – отличается ли успеваемость игроков. Аналогично используйте возможность разделить на доли.

В отчете должны быть приведены: вид используемого теста (одновыборочный, двухвыборочный, ранговый и т.д.), формулировки нулевой и альтернативной гипотез, вычисленные и взятые из таблиц статистики.

Должны быть приведены 2-3 варианта гипотез о среднем и о доле. Не забудьте о контрольных вопросах.

# 2. Ход работы

Для работы были сформулированы следующие гипотезы:

1. Имеется взаимосвязь между кодом школы и частотой игры;
2. Имеется взаимосвязь между обеспеченностью семьи и частотой игры;
3. Имеется взаимосвязь между полом и увлечённостью играми.

Проверка гипотез

1. Имеется взаимосвязь между кодом школы и частотой игры

*H0* – Не существует связи между кодом школы и частотой игры

*H1* – Существует связь между кодом школы и частотой игры

Уровень значимости – 5%.

1. Имеется взаимосвязь между обеспеченностью семьи и частотой игры

*H0* – Не существует взаимосвязи между обеспеченностью семьи и частотой игры

*H1* – Существует взаимосвязь между обеспеченностью семьи и частотой игры

Уровень значимости – 5%.

1. Имеется взаимосвязь между полом и увлечённостью играми

*H0* – Не существует взаимосвязи между полом и увлечённостью играми

*H1* – Существует взаимосвязь между полом и увлечённостью играми

Уровень значимости – 5%.