Институт информационных технологий

Кафедра: Математическое и программное обеспечение ЭВМ

Дисциплина: Основы Data Science

ЛАБОРАТОРНАЯ РАБОТА № 2

Тема: Структура и описание данных. Описательные характеристики

Выполнил:

студент гр. 1ПИб-02-2оп-23

Кринкин Олег Алексеевич

Проверил:

Юдина Ольга Вадимовна

# 1. Задание

**Цель**. На основе результатов предыдущей работы составить представление о структуре и распределении данных.

Задача этапа – понять, что именно содержится в этих данных, какие факты они описывают, понять, что они могут дать.

**Задания**

Вспомните формулировки задач (исследований), которые вы сделали на прошлом занятии. Запишите предположения, которые нужно проверить. Подумайте, что может их подтвердить или опровергнуть.

Например, мы предполагаем, что в игры «с головой» чаще уходят дети, у которых нет интереса к учебе, а родители тоже в учебе не были заинтересованы.

Что может подтверждать эти предположения?

*Например, определим чрезмерную увлеченность. Как это можно сделать? Посмотрим* *на распределение – можно ли выделить группу «игроков» по времени? А другие группы?* *Можем в этих группах посчитать среднее время в игре, среднюю частоту…Посчитать* *доверительные интервалы – сопоставить*. Нужна проверка гипотез о среднем?

Используя результаты прошлой работы по структурированию данных. а также с учетомпостроенных описаний:

Для выделенных для сравнения групп, вычислите (количественных) подходящие средние значения, доверительные интервалы меры разброса

Для интересующих категорий – доли и их доверительные интервалы.

*Например, совпадают или нет доли самых активных игроков среди лучших и не* *лучших? С начала посчитаем доли и доверительные интервалы, а затем примем решение,* *нужна ли проверка статистических гипотез*

**Помните, что в исследовании должны быть задействованы максимум переменных**

Заключение: какие выводы можно сделать по вашей работе

# 2. Ход работы

На основе полученных ранее данных можно выдвинуть и проверить следующие гипотезы:

1. В игры «с головой» чаще уходят дети, у которых нет интереса к учебе, а родители тоже в учебе не были заинтересованы (из примера);
2. Один из полов имеет больший игровой стаж, чем другой;
3. Игроки со стажем в день проводят больше времени за компьютерными играми;
4. Дети из бедных и обеспеченных семей учатся по-разному.

Проверка гипотез

1. В игры «с головой» чаще уходят дети, у которых нет интереса к учебе, а родители тоже в учебе не были заинтересованы (из примера);

*H0* – Дети, которые часто играют в игры не заинтересованы в учёбе, как и их родители

*H1* – Дети, которые часто играют в игры заинтересованы в учёбе, в отличие от их родителей

Уровень значимости – 5%.



|-22.511| > |1.649|, следовательно, гипотеза *H0* для матерей отклоняется



|-19.648| > |1.649|, следовательно, гипотеза *H0* для отцов отклоняется

1. Один из полов имеет больший игровой стаж, чем другой

*H0* – Разные пола имеют одинаковый игровой стаж

*H1* – Разные пола имеют различный игровой стаж

Уровень значимости – 5%.



|-11.699| > |1.647|, следовательно, гипотеза *H0* отклоняется

1. Игроки со стажем в день проводят больше времени за компьютерными играми;

*H0* – Игроки с малым стажем и большим играют одинаково долго

*H1* – Игроки с малым стажем и большим играют в течении разного времени

Уровень значимости – 5%.



|-22.996| > |1.647|, следовательно, гипотеза *H0* отклоняется

1. Дети из бедных и обеспеченных семей учатся по-разному

*H0* – Дети из бедных и обеспеченных семей учатся одинаково

*H1* – Дети из бедных и обеспеченных семей учатся по-разному

Уровень значимости – 5%.



|-1.953| > |1.648|, следовательно, гипотеза *H0* отклоняется

Выводы:

* Ученики разных полов имеют различный в играх;
* Игроки, что начали играть недавно играют не одинаковое время в день с теми, кто начал играть давно;
* Успеваемость детей из обеспеченных и бедных семей различна.