МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное

образовательное учреждение высшего образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт информационных технологий

*наименование института(факультета)*

Кафедра математического и программного обеспечения ЭВМ

*наименование кафедры*

Основы Data Science

*наименование дисциплины в соответствии с учебным планом*

ЛАБОРАТОРНАЯ РАБОТА №2

“Описание данных. Статистический вывод”

|  |  |
| --- | --- |
| Исполнитель |  |
| Студент | 1ПИб-02-1оп-22 |
|  | *группа* |
|  | Рябкова А.Л. |
|  | *ФИО* |
| Руководитель | Юдина О.В. |
|  | *ФИО преподавателя* |
| Оценка |  |
| Подпись |  |

2024 год

**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 предыдущей работе: 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 предыдущей работе.

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, имеющих двоих и более друзей и тех, у кого друзей меньше – отфильтруйте по этому признаку показатели успеваемости, посчитайте среднее и доверительные интервалы. Если доверительные интервалы пересекаются – сформулируйте гипотезы. Результат позволит сделать заключение – отличается ли успеваемость студентов, которые общаются мало и их более общительных товарищей.

Ход работы:

Задача: Исследовать влияние депрессии на студентов

Гипотеза 1: Доля людей

H0: Девочки чаще страдают депрессией

H1: Мальчики чаще страдают депрессией

**Гипотеза по выборочным долям – две выборки**

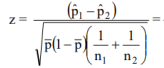
Кол-во девочек, страдающих депрессией – 25

Кол-во мальчиков, страдающих депрессией – 9

Р1 = 25/43 = 0,58

Р2 = 9/56 = 0,16

Р = 34/99 = 0,34



1, 02

Критическое значение = 1,64

1,02 < 1,64

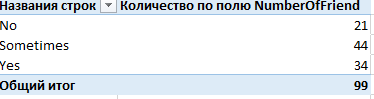
Значит верна гипотеза о том, что девочки чаще страдают от депрессии.

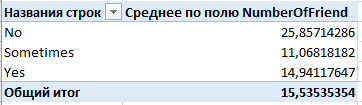
Гипотеза 2:

1. Испытание гипотезы по выборочным средним – независимые выборки. Двух выборочный t-тест с одинаковыми дисперсиями.

Были отобраны две независимые выборки и получены следующие результаты

|  |  |
| --- | --- |
| DepressionStatus | NumberOfFriend |
| No | 7 |
| No | 3 |
| No | 7 |
| No | 3 |
| No | 7 |
| No | 3 |
| No | 3 |
| No | 3 |
| No | 3 |
| No | 100 |
| No | 23 |
| No | 3 |
| No | 3 |
| No | 100 |
| No | 23 |
| No | 3 |
| No | 3 |
| No | 100 |
| No | 23 |
| No | 100 |
| No | 23 |
| Sometimes | 15 |
| Sometimes | 4 |
| Sometimes | 0 |
| Sometimes | 0 |
| Sometimes | 15 |
| Sometimes | 4 |
| Sometimes | 0 |
| Sometimes | 0 |
| Sometimes | 15 |
| Sometimes | 4 |
| Sometimes | 0 |
| Sometimes | 10 |
| Sometimes | 12 |
| Sometimes | 55 |
| Sometimes | 17 |
| Sometimes | 7 |
| Sometimes | 4 |
| Sometimes | 9 |
| Sometimes | 0 |
| Sometimes | 12 |
| Sometimes | 15 |
| Sometimes | 7 |
| Sometimes | 0 |
| Sometimes | 10 |
| Sometimes | 12 |
| Sometimes | 55 |
| Sometimes | 17 |
| Sometimes | 7 |
| Sometimes | 4 |
| Sometimes | 9 |
| Sometimes | 0 |
| Sometimes | 12 |
| Sometimes | 15 |
| Sometimes | 7 |
| Sometimes | 10 |
| Sometimes | 12 |
| Sometimes | 55 |
| Sometimes | 17 |
| Sometimes | 7 |
| Sometimes | 4 |
| Sometimes | 9 |
| Sometimes | 7 |
| Sometimes | 4 |
| Sometimes | 9 |
| Yes | 2 |
| Yes | 3 |
| Yes | 1 |
| Yes | 3 |
| Yes | 2 |
| Yes | 6 |
| Yes | 2 |
| Yes | 2 |
| Yes | 3 |
| Yes | 1 |
| Yes | 3 |
| Yes | 2 |
| Yes | 6 |
| Yes | 2 |
| Yes | 2 |
| Yes | 3 |
| Yes | 1 |
| Yes | 3 |
| Yes | 2 |
| Yes | 6 |
| Yes | 3 |
| Yes | 1 |
| Yes | 3 |
| Yes | 2 |
| Yes | 6 |
| Yes | 80 |
| Yes | 6 |
| Yes | 60 |
| Yes | 80 |
| Yes | 6 |
| Yes | 60 |
| Yes | 80 |
| Yes | 6 |
| Yes | 60 |





Значит n1 = 21, x1 = 25,857, s1 = 37,624

n2 = 44, x2 = 11,068, s2 = 13,142

n3 = 34, x3 = 14,941, s3 = 26,270