МИНОБРАНУКИ РОССИИ

федеральное государственное бюджетное

образовательное учреждение высшего образования

«ЧЕРЕПОВЕЦКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Институт информационных технологий

*наименование института (факультета)*

Математическое и программное обеспечение ЭВМ

*наименование кафедры*

Основы Data Science

*наименование дисциплины в соответствии с учебным планом*

ЛАБОРАТОРНАЯ РАБОТА №4

Изучение связей между переменными

|  |  |
| --- | --- |
| Руководитель | Юдина О. В. |
|  | ФИО преподавателя |
| Исполнитель |  |
| студент | 1ПИб-02-1оп-22 |
|  | группа |
|  | Харламов Денис |
|  | Алексеевич |
|  | Фамилия, имя, отчество |
| Оценка |  |
| Подпись |  |

Череповец, 2024 г.

**Цель:**

Знакомство с этапом понимания данных стандарта CRISP-DM.Задача этапа – исследовать связи между переменными между переменными.

**Задание:**

Для обнаружения зависимостей воспользуйтесь графическими иллюстрациями, а также результатами предыдущих работ

1. Выделите переменные, для которых может быть посчитана

параметрическая корреляция ( Пирсона). Посчитайте. Оцените

значимость коэффициента Пирсона.

2. Выберите зависимости, для которых можно воспользоваться ранговым

коэффициентом Спирмена. Вычислите те, которые могут быть

использованы при описании данных. Вычислите и оцените

коэффициент.

3. Подберите подходящие инструменты для нелинейных зависимостей.

Если такие есть в наборе данных – вычислите коэффициенты, оцените

результаты.

4. Оцените статистическую значимость найденных зависимостей.

Воспользовавшись программой Excel, постройте регрессионную

зависимость. Для построения зависимости может быть использована часть

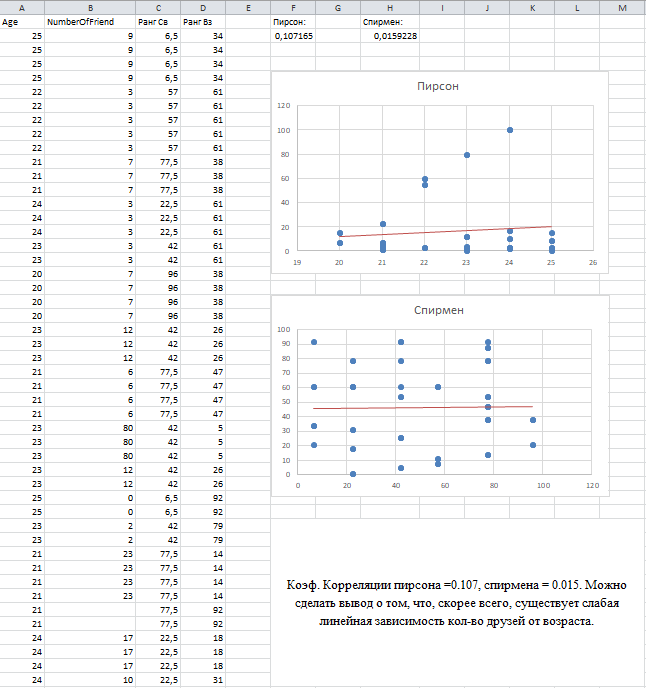
выборки, группы записей, для которых вы установили закономерности. Попробуйте использовать качественную переменную.

Для анализа результатов воспользуйтесь оценками Excel. Поясните оценки и

результаты. Сделайте выводы о пригодности регрессии, обоснуйте их.

Результаты оформите в виде отчета.

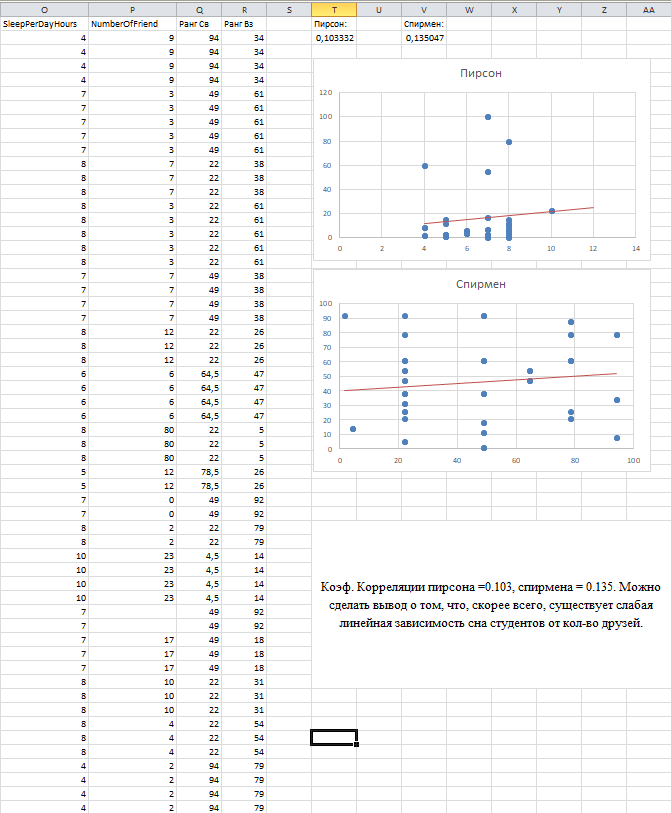
Ход работы



Коэф. Корреляции пирсона =0.107, спирмена = 0.015.

Значения коэффициентов корреляции Пирсона близки к нулю. Можно сделать вывод о том, что, скорее всего, существует слабая линейная зависимость кол-во друзей от возраста.

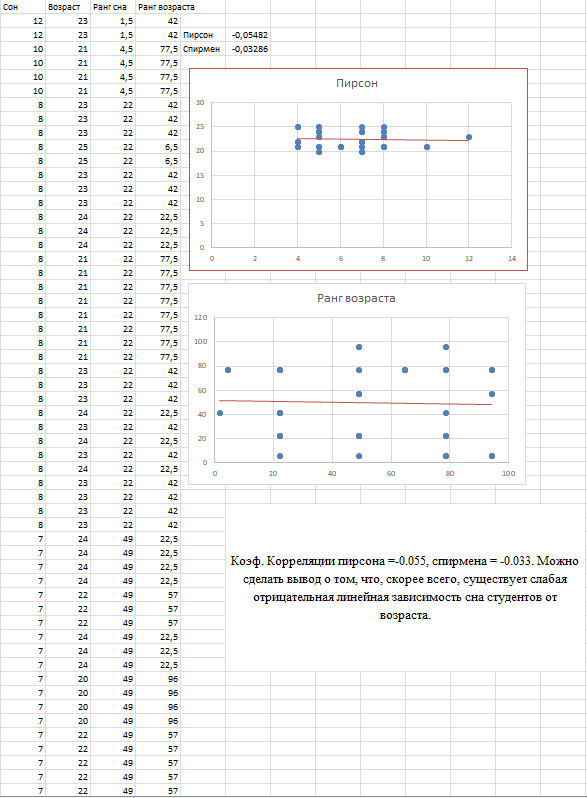
Вывод – существует слабая линейная зависимость кол-во друзей от возраста



Коэф. Корреляции пирсона =0.103, спирмена = 0.135.

Значения коэффициентов корреляции Пирсона близки к нулю. Можно сделать вывод о том, что, скорее всего, существует слабая линейная зависимость сна студентов от кол-во друзей.

Вывод – существует слабая линейная зависимость сна студентов от кол-во друзей.



Коэф. Корреляции пирсона =-0.055, спирмена = -0.033.

Значения коэффициентов корреляции Пирсона близки к нулю. Можно сделать вывод о том, что, скорее всего, существует незначительная отрицательная линейная зависимость сна студентов от возраста.

Вывод – существует незначительная отрицательная линейная зависимость сна студентов от возраста.

Контрольные вопросы

1. Когда может быть использован коэффициент корреляции Пирсона?

Коэффициент корреляции Пирсона используется для измерения линейной зависимости между двумя непрерывными переменными.

1. Что такое ранговая корреляция. Когда она применяется?

Ранговая корреляция — это мера связи между двумя порядковыми переменными или двумя непрерывными переменными с ненормальным распределением. В случае, когда силу зависимости между случайными величинами невозможно определить численно при помощи обычного коэффициента корреляции, используют ранговую корреляцию (Спирмена), которая имеет дело не с непосредственными значениями, а с рангами (порядковыми номерами).

1. Когда применяется коэффициент Фехнера? Как им воспользоваться?

Коэффициент корреляции (знаков) Фехнера — это некоторое число от -1 до 1, характеризующее степень согласованности направлений отклонений значений зависимой и независимой случайных величин.

1. На основании чего может проводится анализ регрессии?

Регрессионный анализ используется для решения следующих типов проблем:

Выявить, какая независимая переменная связана с зависимой.

Понять отношения между зависимой и независимыми переменными.

Определение вклада отдельных независимых переменных в вариацию зависимой.

1. Какое значение имеют остатки?

Разница между наблюдаемым и предсказанным значением. Остатки применяются для получения информации об адекватности построенной модели многомерной линейной регрессии. Чем лучше регрессионная модель согласуется с данными, тем меньше величина остатков.