

Лабораторная работа №5

Изучение связей между переменными

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 выявления нелинейных зависимостей, зависимостей между качественными переменными, качественными и количественными переменными. Если такие есть в наборе данных – вычислите коэффициенты, оцените результаты применения
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.

Контрольные вопросы

1. Когда может быть использован коэффициент корреляции Пирсона
2. Что такое ранговая корреляция. Когда она применяется
3. Когда применяется коэффициент Фехнера? Как им воспользоваться?
4. На основании чего может проводиться анализ регрессии?
5. Какое значение имеют остатки