Министерство образования Республики Беларусь

Учреждение образования

Брестский государственный технический университет

Кафедра ИИТ

Лабораторная работа №3

За 7 семестр

По дисциплине «Методы обработки данных»

Тема: «Машинное обучение. Регрессионный анализ»

Выполнил:

Студент 4 курса, группы ПО-7(1)

Вощук Александр Янович

Проверил:

Савицкий Юрий Викторович

2023

**ЛАБОРАТОРНАЯ РАБОТА №3**

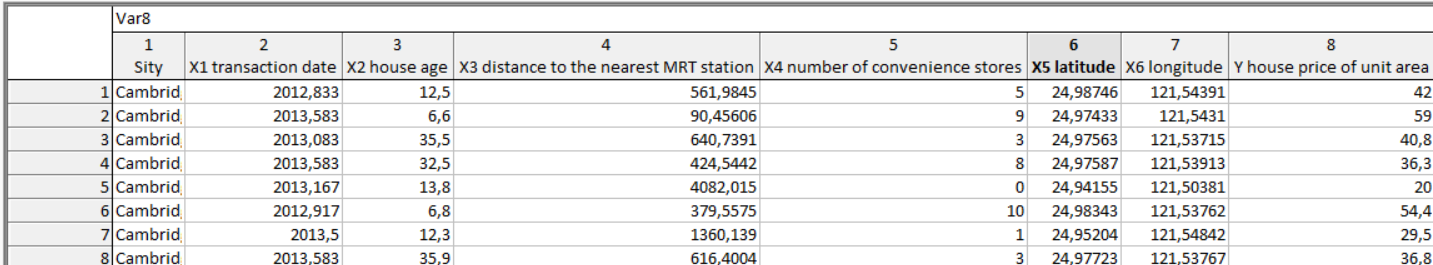
**Машинное обучение. Регрессионный анализ**

**Цель:** изучить основы методов Machine Learning в контексте задачи множественного регрессионного анализа, приобрести навыки работы с методами Machine Learning в системе STATISTICA StatSoft, осуществить обработку методами Machine Learning индивидуального набора данных и интерпретацию результатов.

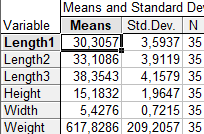
**Вариант 1**

**Ход работы:**

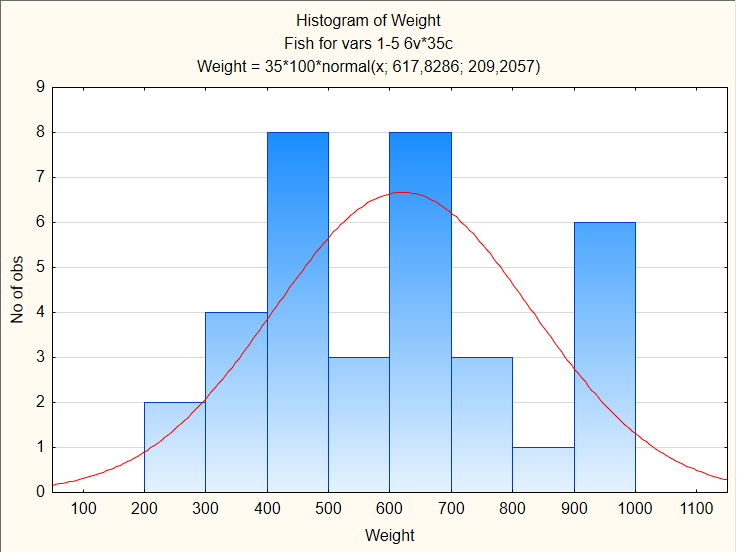
**Задача исследований** - анализ коррелятов цены домов, то есть переменных, которые лучше всего предсказывают цену дома в указанной области. Таким образом, вы будете рассматривать переменную 8 (Y house price of unit area) как зависимую или критериальную переменную, а все остальные переменные как независимые или предикторы.

****

Проведём множественную регрессию, с использованием “Y house price of unit area” в качестве зависимой переменной, а остальные как независимые. В результате выведем “Средние значения и стандартные отклонения”:

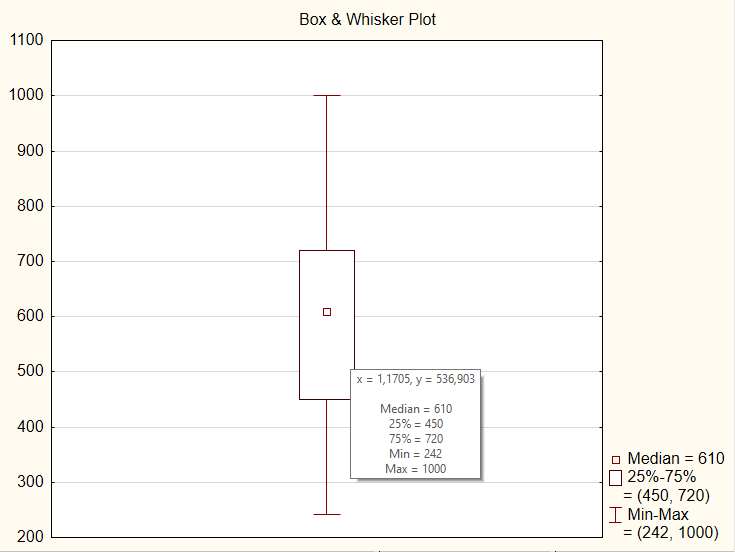


Далее построим гистограмму по полю “Y house price of unit area”, с количество категорий 16. В результате мы должны сравнить полученный результат с нормальным распределением:

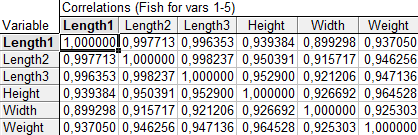


Как можем наблюдать, распределение данной переменной отличается от нормального распределения. Однако, даже несмотря на то, что в 3 округах (два крайних левых и крайний с права столбца) процент семей, живущих за чертой бедности, выше, чем можно было бы ожидать в соответствии с нормальным распределением, они все же кажутся достаточно «в пределах диапазона».

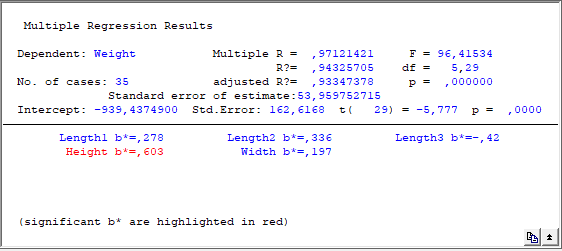
Для того чтобы просмотреть распределение этой переменной, посмотрим её “График c усами”.



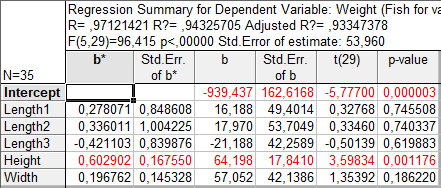
Далее будем получать диаграмму рассеивания, по нему мы можем определить о степени взаимосвязи различных переменных.



Получимрезультат множественной регрессии:

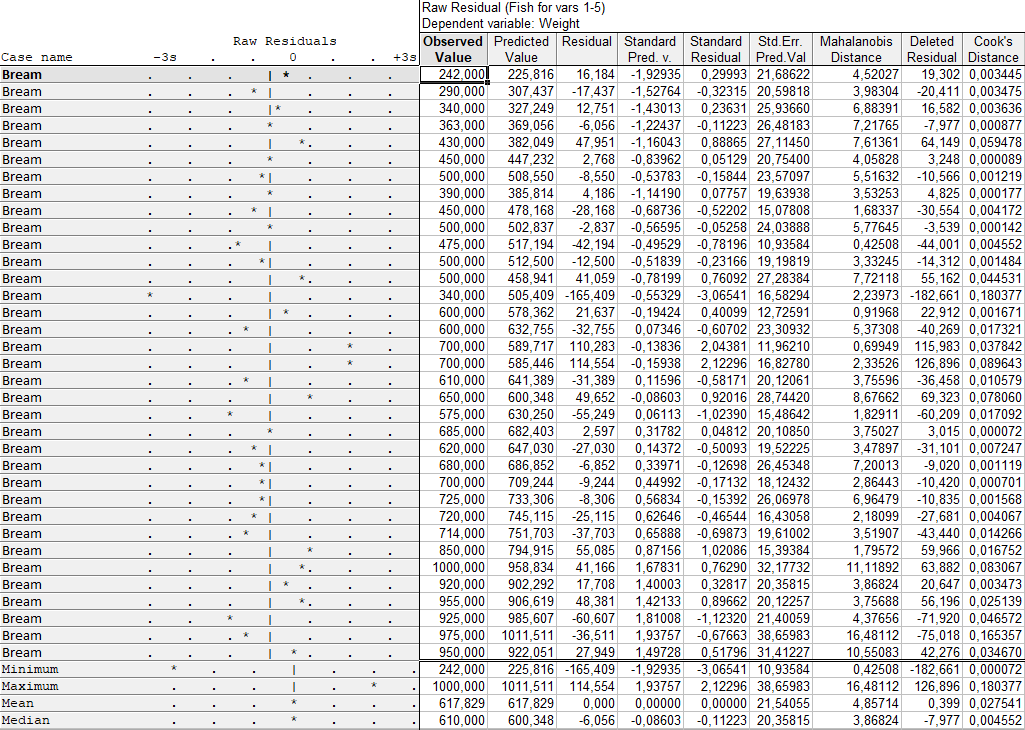


Определим коэффициенты регрессии, чтобы понять какая из независимых переменных, оказывает наибольшее влияние на цену домов.

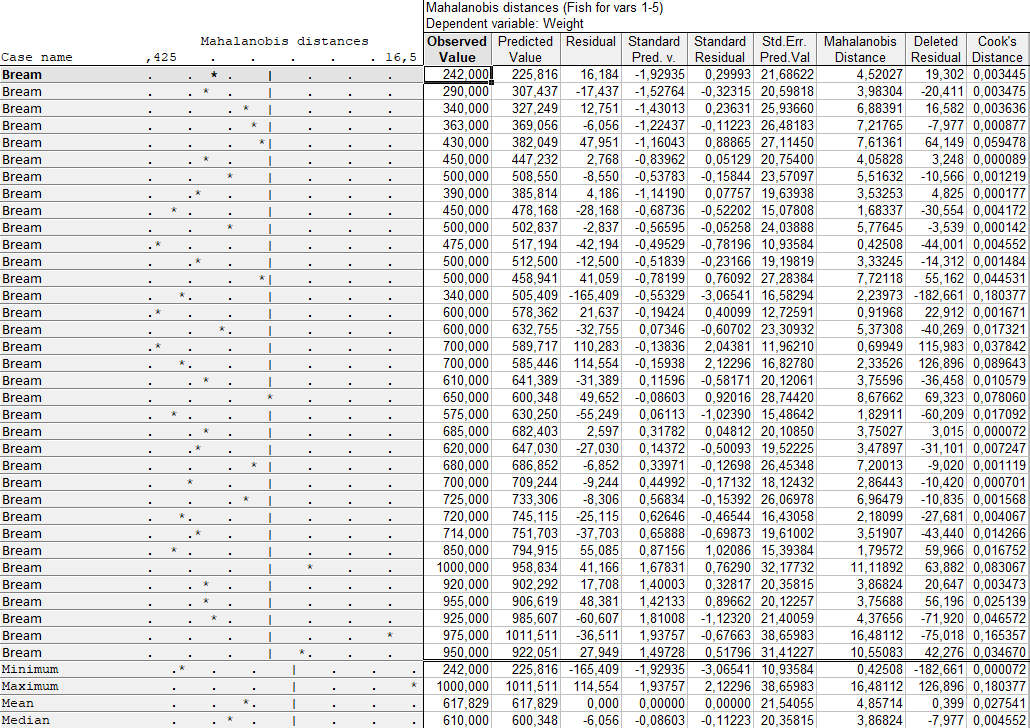


Из результата можем наблюдать, что наиболее значимыми являются переменные X2, X3, X4, которые также являются статически значимыми. Также наблюдается зависимость от X1.

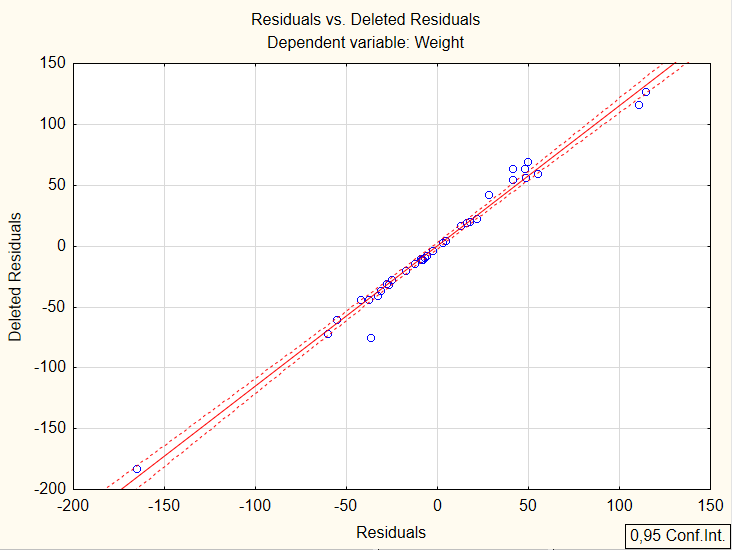
Получив необходимую зависимость, проведём анализ остатков наших данных:



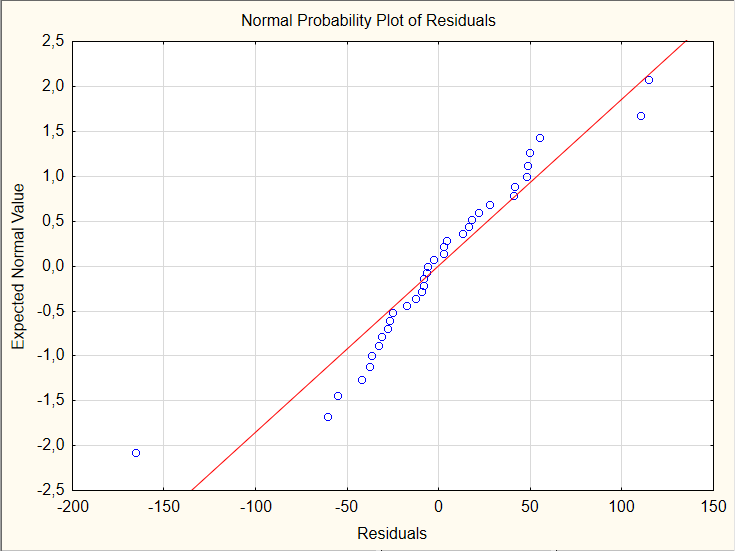
Определим также расстояние Малаханобиса:



Построим график “Остатков по сравнению с удаленными остатками”:



Построим “График нормальной вероятности остатков”:



**Вывод**: изучил основы методов Machine Learning в контексте задачи множественного регрессионного анализа, приобрел навыки работы с методами Machine Learning в системе STATISTICA StatSoft, осуществил обработку методами Machine Learning индивидуального набора данных и интерпретацию результатов.