**РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ**

**Факультет физико-математических и естественных наук**

**Кафедра прикладной информатики и теории вероятностей**

Практикум по математической статистике

Лабораторная работа №1

**Тема: «Множественный регрессионный анализ»**

Вариант 10

Выполнил

Студент: Феоктистов Владислав

Группа: НПМбд-01-19б

№ c/б: 1032192939

Преподаватель: Матюшенко Сергей Иванович

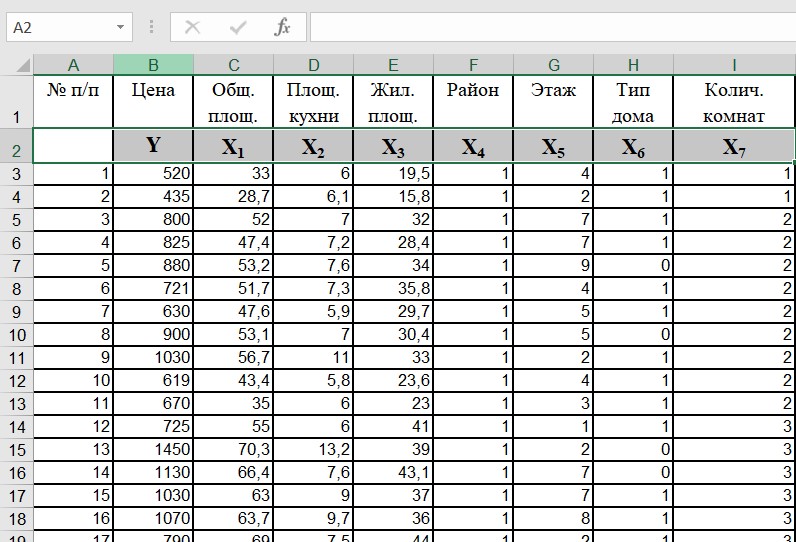
**МОСКВА**

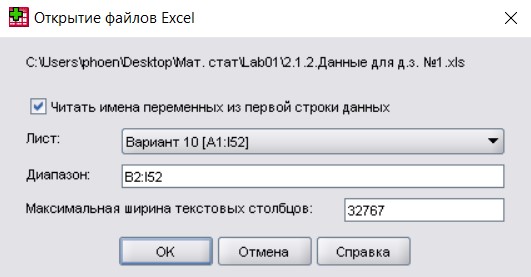
2022 г.

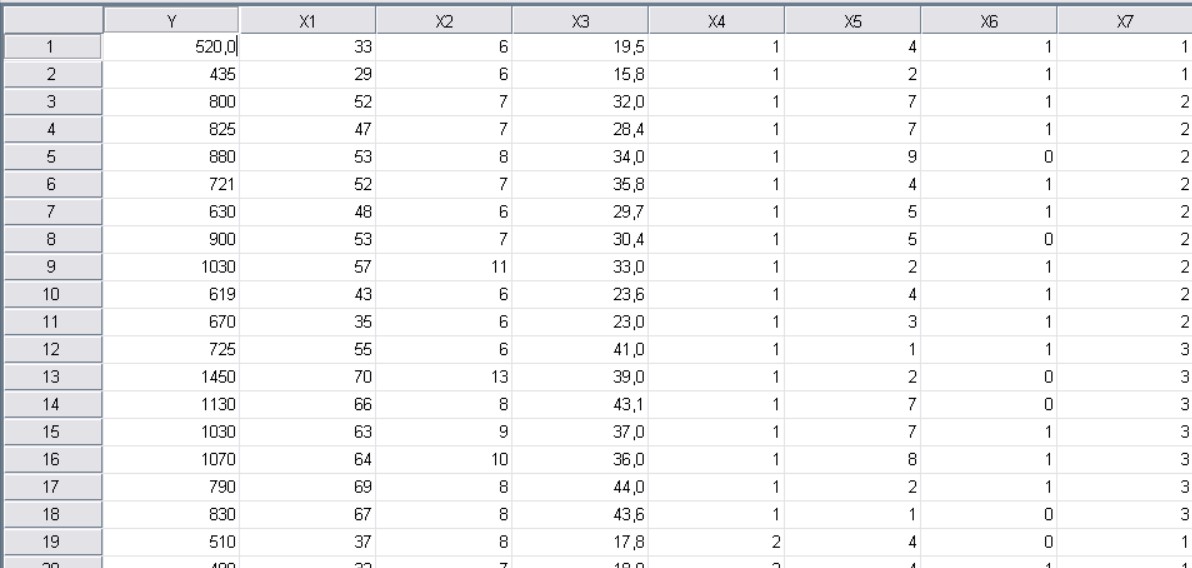
**Цель работы:** приобрести практические навыки применения множественного регрессионного анализа для решения конкретных задач с использованием статистического пакета SPSS.

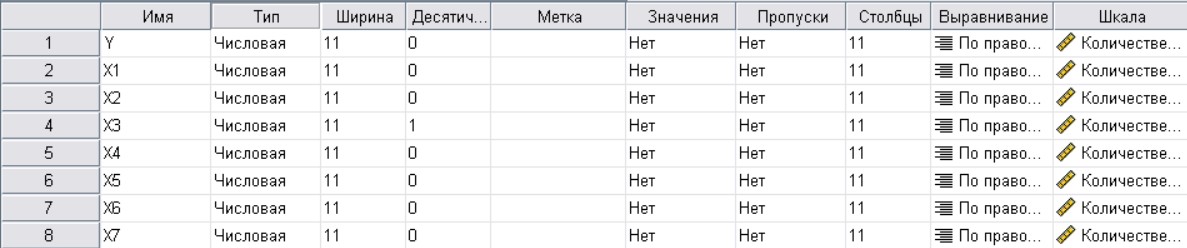
**Ход работы:**

1. Повторил теоретические основы множественного регрессионного анализа.
2. Разобрал пример использования SPSS для построения модели множественной линейной регрессии.
3. Импортировал данные из файла в формате Excel в файл SPSS.



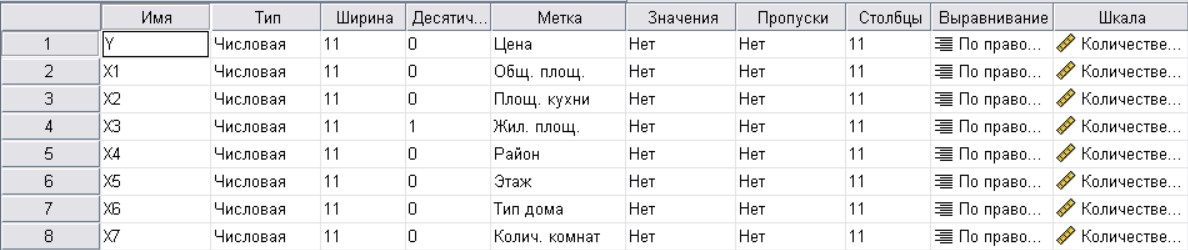




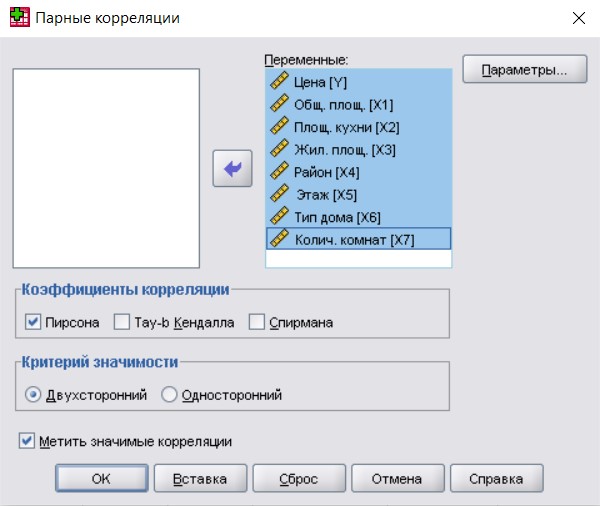


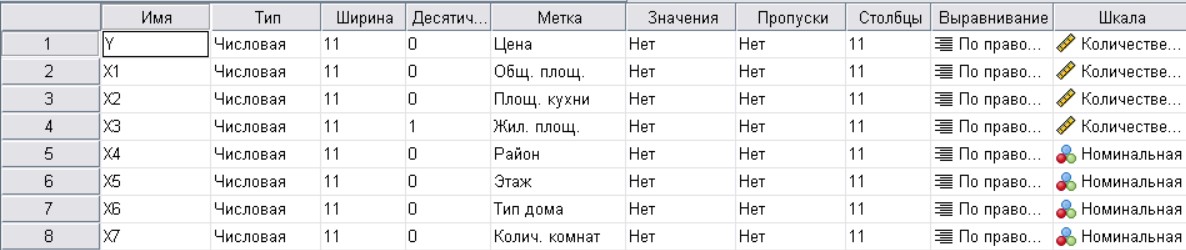
Важно отметить, что перед импортом Excel файла необходимо в начале просмотреть содержимое файла, определить лист и диапазон выборки. Поскольку у меня 10 вариант и в исходных данных первый столбец отвечает за индексы, а первая строка за метки (более подробное описание параметров Y, X1, X2, …, X7), то я выбрал лист с названием «Вариант 10» и установил диапазон выборки B2:I52.

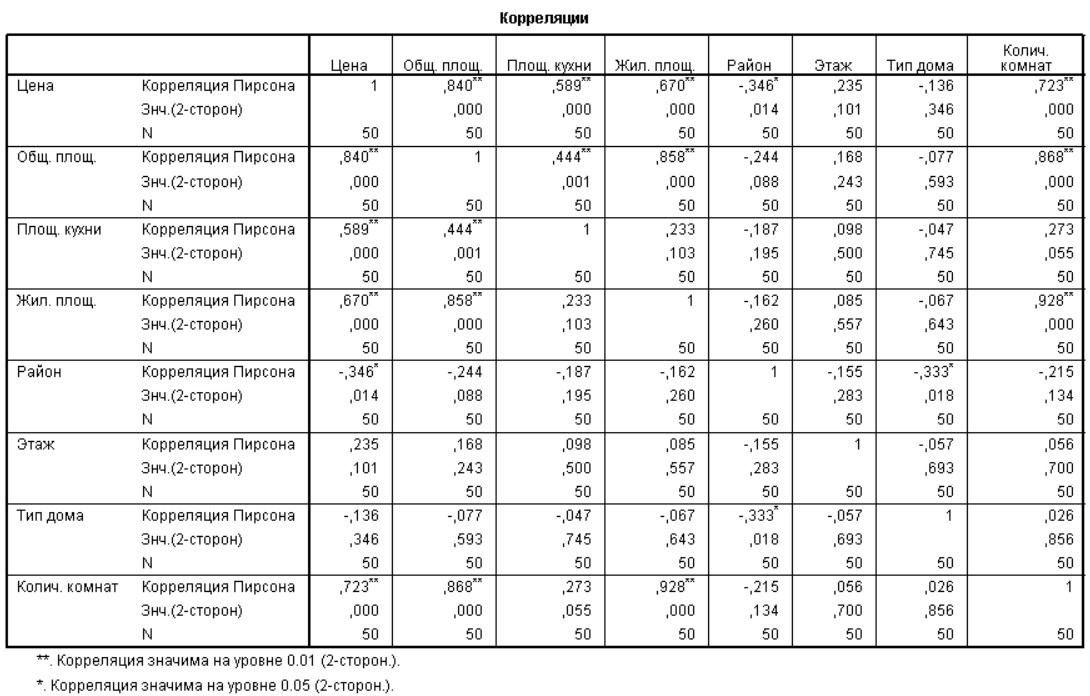
Для удобства дальнейшего работы, параметрам Y, X1, X2, …, X7 можно задать метки.



Далее я построил матрицу парной корреляции всех переменных.







Как можно заметить, SPSS сам установил для параметров X4, X5, X6, X7 шкалу «Номинальная», поскольку эти параметры носят категориальный характер.

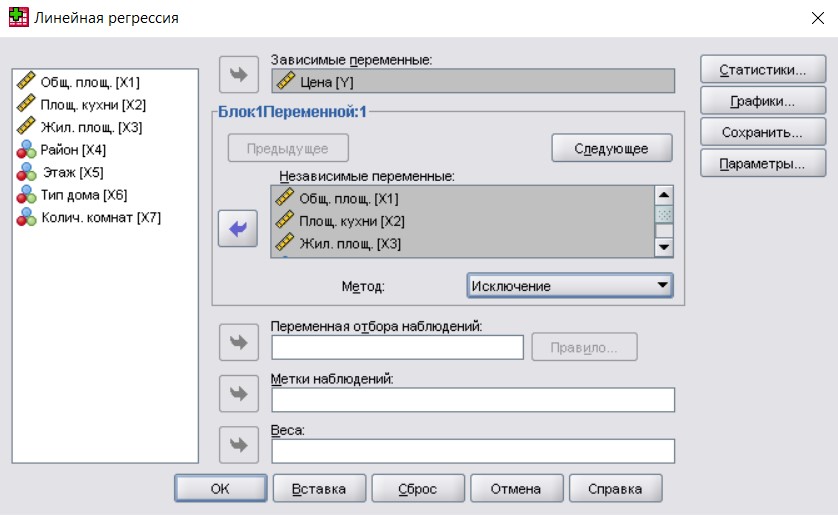
Из таблицы корреляции видно, что зависимая переменная «Цена» сильнее всего связана с независимыми переменными «Общ. площ.» (), «Колич. комнат» (), «Жил. площ» () и «Площ. кухни» (). Однако, в то же время переменная «Общ. площ.» сильно коррелирует с переменными «Жил. площ» (), «Колич. комнат» (), что свидетельствует о наличии мультиколлинеарности (это очевидно, поскольку общая площадь напрямую зависит от жилой площади и количества комнат и их площадь, как правило больше, чем площадь кухни, а значит и их корреляция выше). Поэтому, т.к. зависимая переменная «Цена» сильнее зависит от независимой переменной «Общ. площ.», то переменные «Жил. площ» и «Колич. комнат» нужно будет исключить.

Стоит заметить, что независимая переменная «Этаж» не имеет значимой корреляции на уровне 0,01 и 0,05 ни для одной из других переменных, в том числе и зависимой. Следовательно, эту переменную тоже можно исключить.

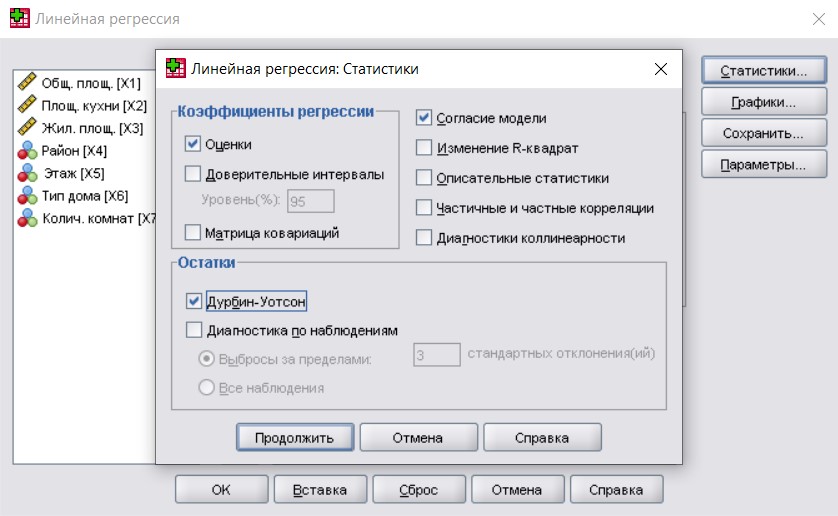
В итоге, воспользовавшись методом исключения, приходим к выводу, что целесообразно построить многофакторное регрессионное уравнение от переменных:

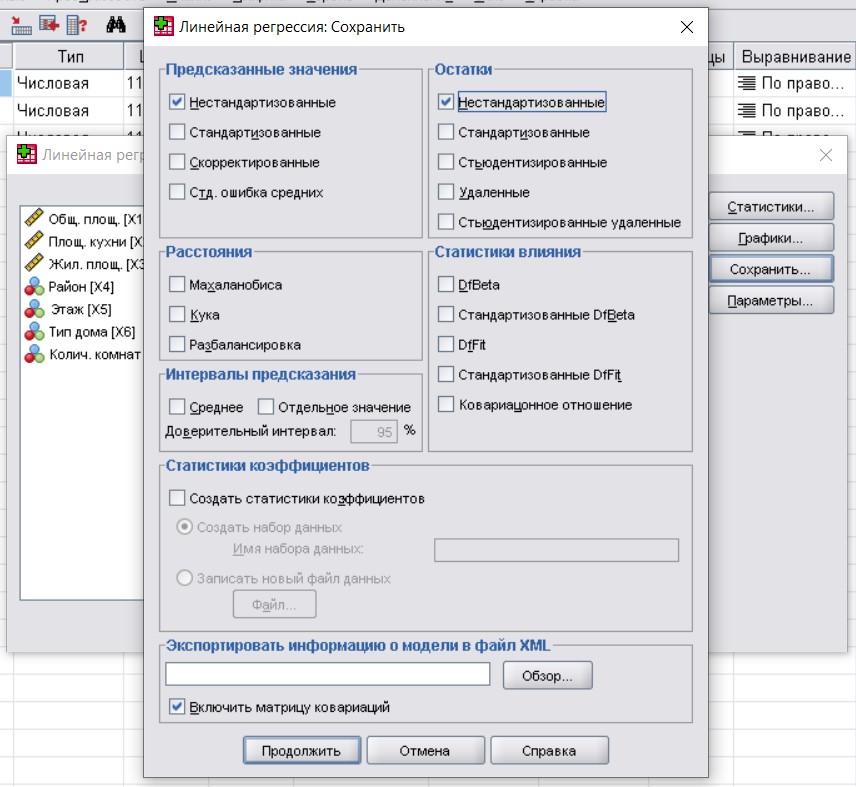
«Общ. площ», «Площ. кухни», «Район» и «Тип дома».

Построим линейное уравнение регрессии, используя метод исключения. В качестве зависимой переменной выбираем «Цена», а в качестве независимых – все остальное.

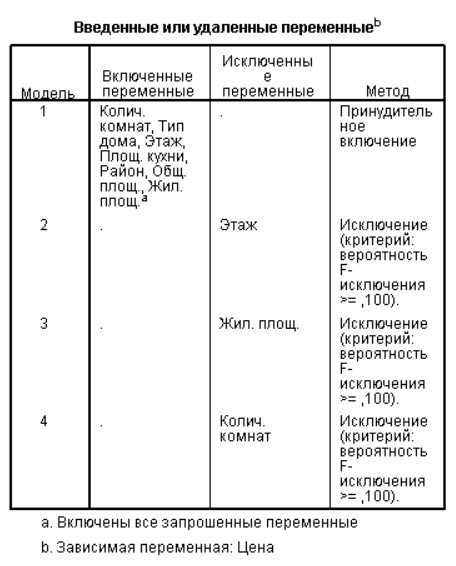


В поле панели «Статистики…» добавляем галочку перед «Дурбин-Уотсон», а в поле панель «Сохранить…» - перед «Нестандартизованные» в разделах «Предсказательные значения» и «Остатки».

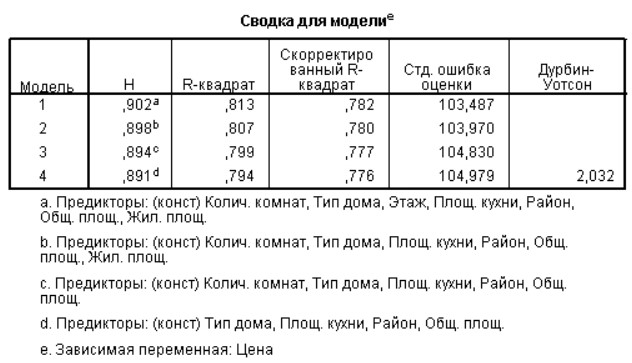




В итоге получаем следующие результаты:



Из таблицы «Введенные и удаленные переменные» видно, что были исключены переменные «Этаж», «Жил. Площ.» и «Колич. комнат». Это сходится с моими исключениями при анализе таблицы корреляции.



С помощью коэффициентов детерминации и множественной корреляции в таблице «Свода для модели» оценим качество модели.

Коэффициент детерминации показывает, что около 79,4% вариаций зависимых переменных учтено в модели и обусловлено влиянием включенных факторов, что достаточно хорошо.

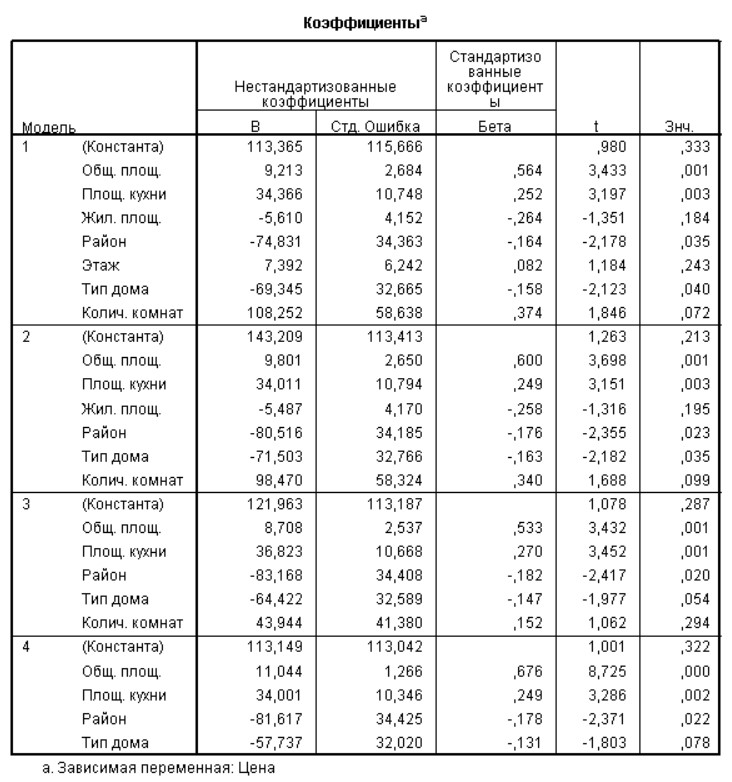
Коэффициент множественной корреляции показывает, что зависимая переменная «Цена» достаточно тесно связана со включенными в модель факторами.

Несмотря на то, что коэффициент детерминации в первой модели (до исключения коррелирующих и незначимых переменных) незначительно больше, чем в четвертой, принимаем за верную мы последнюю, поскольку эта разница незначительна, а мультиколлинеарность нежелательна и от нее нужно избавляться.

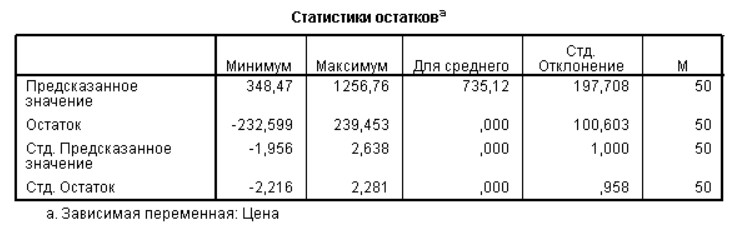


На основании F-критерия Фишера в дисперсионном анализе сделаем проверку значимости уравнения регрессии.

Значение критерия Фишера . Расчетное значение F-критерия меньше (Округленное до тысячных значение равно 0,000. Это не говорит о том, что значение равно 0, а только о том, что оно очень маленькое), следовательно, модель значима.







На основе данных в таблице «Коэффициенты» составил уравнение регрессии:

При увеличении общей площади на 1 квадратный метр, цена увеличится на 11,044 у.е.; при увеличении площади кухни на 1 квадратный метр, цена увеличится на 34,001 у.е.; при увеличении номера района на 1, цена уменьшится на 81,617 у.е.; при увеличении номера типа дома на 1, цена уменьшается на 57,737 у.е.

**Вывод:** при выполнении лабораторной работы были приобретены практические навыки применения множественного регрессионного анализа для решения конкретных задач с использованием статистического пакета SPSS.