Министерство образования и науки Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего профессионального образования

«Владимирский государственный университет

имени Александра Григорьевича и Николая Григорьевича Столетовых»

(ВлГУ)

Кафедра информационных систем и программной инженерии

Лабораторная работа №6

по дисциплине "Обработка экспериментальных данных"

ТЕМА РАБОТЫ:

Множественный регрессионный анализ

Выполнил:

студент гр. ПРИ-120

Парахин К.В.

Приняла:

преподаватель кафедры ИСПИ

Андрианова В.И.

Владимир 2025 г

Цель работы:

Освоение методов построения регрессионных моделей на основе обработкой статистических данных.

Выполнение работы

*Вариант* 2*.* Представлены данные о рынке строящегося жилья в Санкт-Петербурге (по состоянию на год.). Исследовать зависимость переменной *у* от различных факторов и отобразить их взаимосвязь в форме регрессионной модели с двумя значимыми факторами.

№

п/п *X*1 *X*2 *X*3 *X*4 *X*5 *X*6 *X*7 *X*8 *Y*

1 1 1 39 20 8,2 0 1 0 15,9

2 3 1 68,4 40,5 10,7 0 1 0 27

3 1 1 34,8 16 10,7 0 1 12 13,5

4 1 1 39 20 8,5 0 1 12 15,1

5 2 1 54,7 28 10,7 0 1 12 21,1

6 3 1 74,7 46,3 10,7 0 1 12 28,7

7 3 1 71,7 45,9 10,7 0 0 0 27,2

8 3 1 74,5 47,5 10,4 0 0 0 28,3

9 4 1 137,7 87,2 14,6 0 1 0 52,3

19 1 1 37 17,8 8,3 0 1 3 15,4

20 3 1 69 42,4 8,3 0 1 3 28,6

21 1 1 40 20 8,3 0 0 0 15,6

22 3 1 69,1 41,3 8,3 0 1 0 27,7

23 2 1 68,1 35,4 13 1 1 20 34,1

24 2 1 75,3 41,4 12,1 1 1 20 37,7

25 3 1 83,7 48,5 12,1 1 1 20 41,9

26 1 1 48,7 22,3 12,4 1 1 20 24,4

27 1 1 39,9 18 8,1 1 0 0 21,3

28 2 1 68,6 35,5 17 1 1 12 36,7

29 1 1 39 20 9,2 1 0 0 21,5

30 2 1 48,6 31 8 1 0 0 26,4

Принятые в таблице обозначения: *Y* - цена квартиры, тыс. долл.; *X*1 - число комнат в квартире; *X*2 - район города (1 - Приморский, Шувалове - Озерки, 2 - Гражданка, 3 – Юго -запад, 4 - Красносельский); *X*3 - общая площадь квартиры (м2); *X*4- жилая площадь квартиры (м2); *X*5 - площадь кухни (м2); *Х*6 - тип дома (1 - кирпичный, 0 - другой); *X*7 - наличие балкона (1 - есть, 0 - нет); *X*8 - число месяцев до окончания срока строительства.

Сверху приведены статистические данные по выборке зависимости – в данном случае n = 21, k = 8

1. Для начала осуществим выбор факторных признаков для построения двухфакторной регрессионной модели. Для этого проведем корреляционный анализ данных

Используем формулу корреляции Пирсона:

r_{xy}=\frac{\Sigma(x_i-\bar{x})\times(y_i-\bar{y})}{\sqrt{\Sigma(x_i-\bar{x})^2\times\Sigma(y_i-\bar{y})^2}}

Для начала находим выражение в числителе – для этого для каждого столбца переменных находим среднее, далее составляем промежуточную матрицу со значениями отклонений каждой из строк переменных от среднего – находим суммы полученных квадратов разностей отклонения (для вычисления знаменателя)

Отдельно составляется еще одна матрица, в которой идет использование функции СУММПРОИЗВ для вычисления суммы произведений соответствующих значений строк двух столбцов – и далее идет подсчет итогового значения для каждой из 2 пар переменных и запись их в корреляционную матрицу – так как от перемены мест двух факторов в паре не меняется значение коэффициента корреляции R – то достаточно вычислить только одну половину матрицы

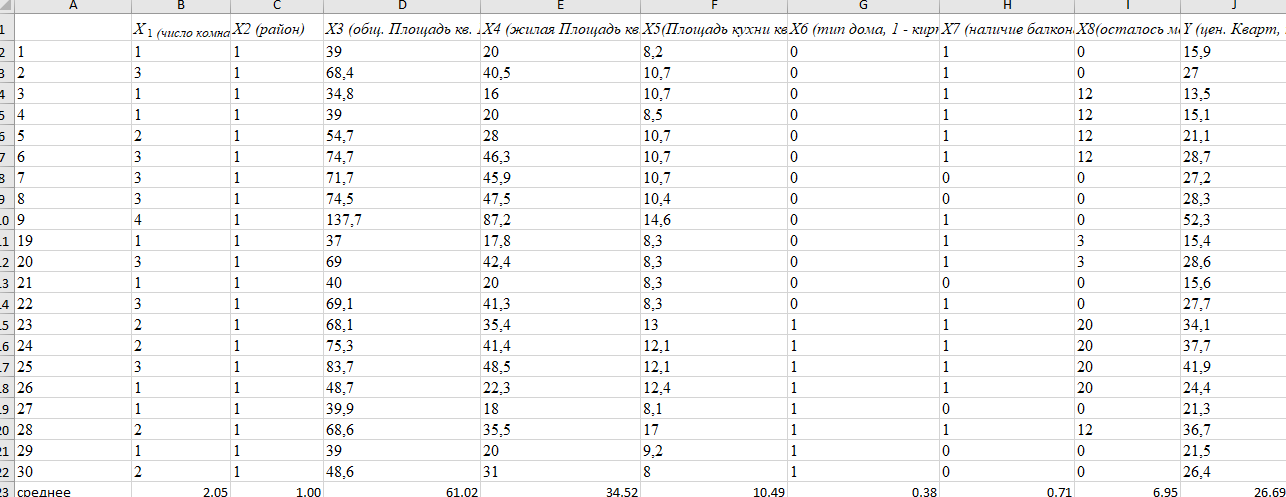


Рис. 1 – Исходная матрица

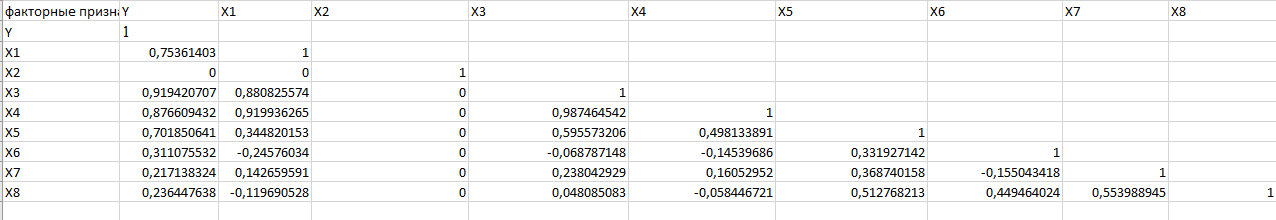


Рисунок 2. Результат корреляционных преобразований

По результатам корреляционного анализа видно, что зависимая переменная Y (цена квартиры) имеет тесную связь с многими факторными переменными. Найдем наибольшие связи с 2 факторами – это связи с факторами X3 (Rx3y = 0.91) и X4 (Rx4y = 0.87) – то есть связи с факторами общей площади квартиры и жилой площади квартиры – что достаточно логично было бы предположить еще до выполнения анализа

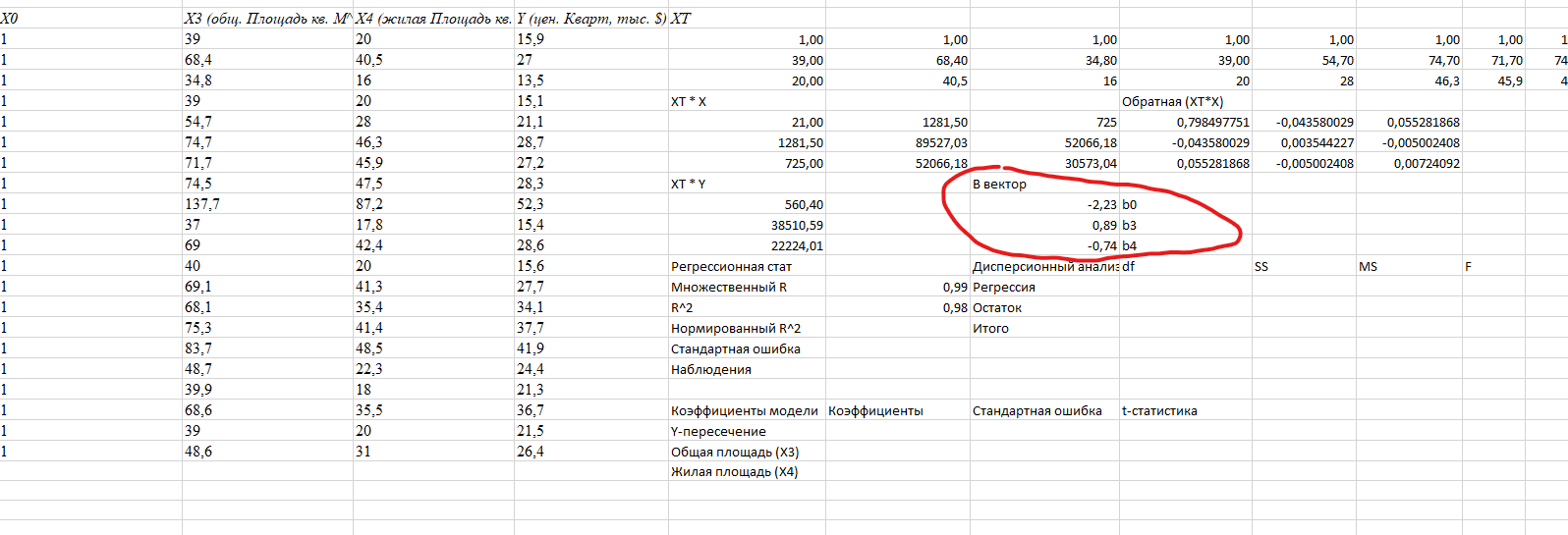
При этом если взять в сравнение еще параметр X1 (число комнат) – то его взаимная связь с фактором X4 (жилая площадь квартиры) приближается к единице (Rx1x4 = 0.987) – это свидетельствует о мультиколлинеарности модели – поэтому в целом можно не включать в исследование фактор X1 – и оставить только факторы X3 (общая площадь квартиры) и X4 (жилая площадь квартиры)

После иcключения незначимых факторов – получаем n = 21, k = 2 (X3 и X4)

Оценим параметры регрессии по методу наименьших квадратов.

То есть нужно получить уравнение для Y = b0 + b3\*X3 + b4\*X4

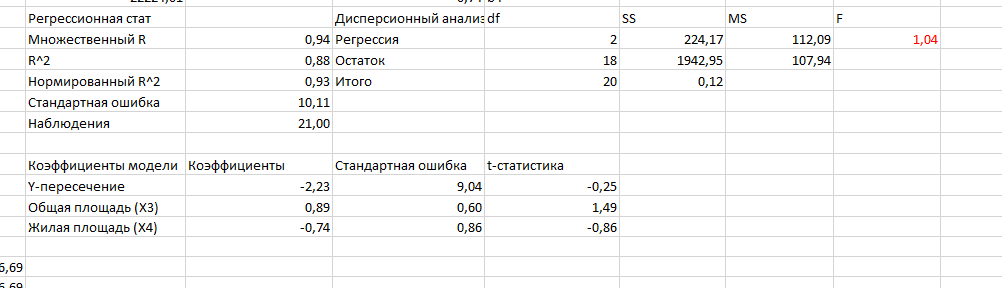
Найдем вектор коэффициентов b из выполнения следующих преобразований: b = ((XT\*X)^(-1))\*XT\*Y – при этом необходимо добавить доп. вектор X0 – для получения значения нулевого коэффициента свободного члена b0



Получаются коэффициенты: b0 = -2.33, b3 = 0.89, b4 = -0,74

Уравнение полученной регрессионной модели принимает вид: Y = -2.33 + 0,89\*X3 – 0,74\*X4

Далее выполняем расчет коэффициентов регрессии, дисперсионного анализа и стандартной ошибки полученных коэффициентов вектора b регрессионной модели



Стандартная ошибка 10.11 указывает на среднюю погрешность предсказаний

Коэффициент множественной корреляции R = 0.94 – он показывает хорошую тесноту связи зависимой переменной Y от факторов X3 и X4

Находим значение критерия Фишера для этой модели = 1.04

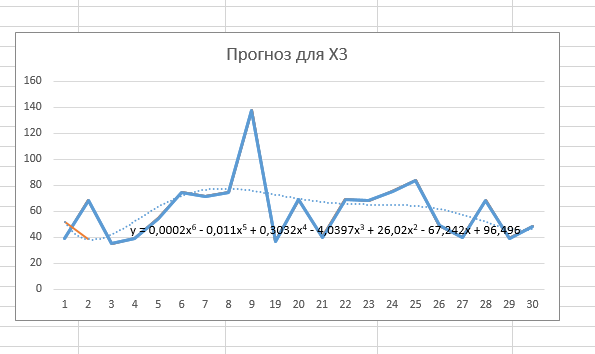
При этом критическим значением для степеней свободы регрессии k = 2, степеней свободы остатка (n – k – 1 = 18) и лямбда = 0,05 = 3.55 – то есть полученное значение критерия Фишера 1.04 < 3.55 - то есть данная регрессионная модель адекватна

Далее проведем анализ полученных значений t-статистики с помощью критерия Стьюдента – для степеней свободы n = 21 и k =2, лямбда = 0,05 – равен 2.11 – то есть так как каждое значение t-статистики меньше этого значения, то каждый коэффициент b по критерию Стьюдента считается незначимым

Далее высчитываю коэффициенты эластичности регрессии – получаем значения -0.08, 2.05, -0.96

Определим точечный и интервальный прогнозные оценки объема реализации на два квартала вперед (*t*0,7 = 1,12)

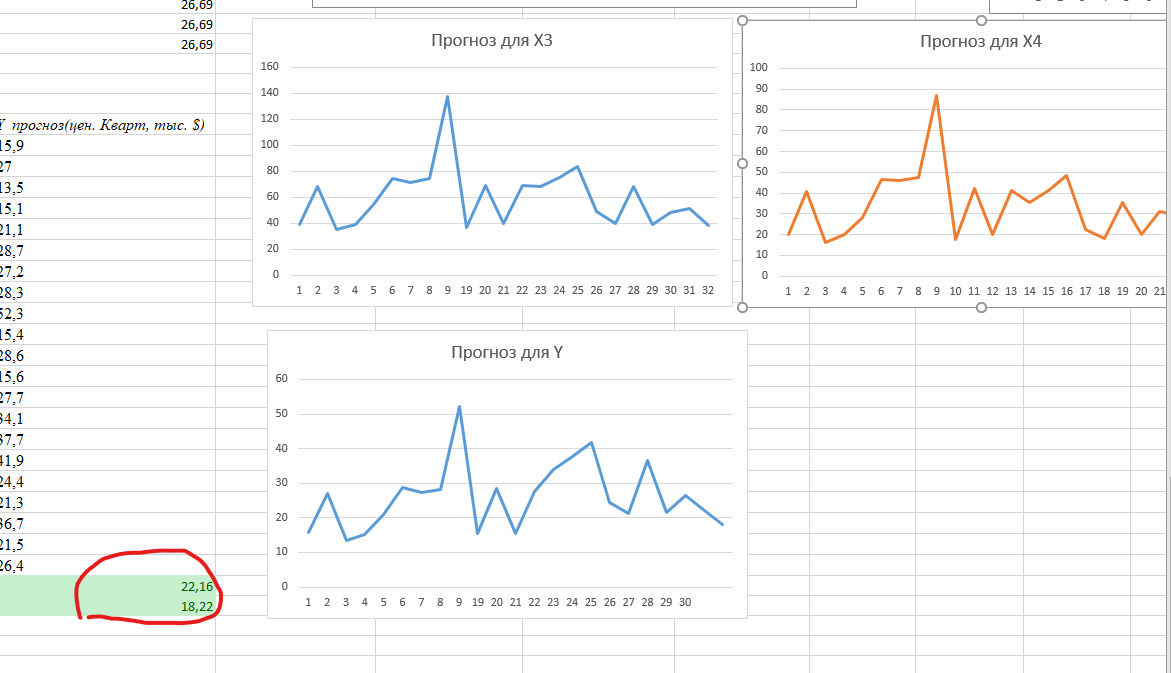
Находим предсказания по функциям факторов X3 и X4 с использованием полиномимального тренда:





Далее остается использовать предсказанные значения функций факторов X3 и X4 для зависимой переменной функции Y = -2.33 + 0,89\*X3 – 0,74\*X4

Подставим в нее найденные прогнозные значения факторов *X*3 и *X*4, находим значения для n = 31 и n = 32 – визуализируем графики прогнозов факторов X3 и X4 – а также главной переменной Y (прогноз на 2 периода)



Вывод

В результате выполнения работы я научился проводить множественный регрессионный анализ переменной в зависимости от различных факторов.