МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

УЧРЕЖДЕНИЕ ОБРАЗОВАНИЯ

ГОМЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ П. О. СУХОГО

Факультет автоматизированных и информационных систем

Кафедра «Информационные технологии»

ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ №3

по дисциплине «Введение в математическое моделирование»

на тему: «Регрессионные модели с одной входной переменной»

Выполнил: студент гр. ИТП-31

Пронуза М.Ю.

Принял: ассистент

Карась О. В.

Гомель 2024

**Цель работы:** приобрести навыки построения парной линейной регрессии и корреляционного анализа, проверки качества уравнения линейной регрессии и прогнозирования индивидуальных значений зависимой переменной на основании линейной регрессии.

**Задание:**

Для проведения регрессионного анализа и прогнозирования необходимо:

1) построить график исходных данных и попытаться зрительно, приближенно определить характер зависимости;

2) выбрать вид функции регрессии, которая может описывать связь исходных данных;

3) определить численные коэффициенты функции регрессии методом наименьших квадратов;

4) оценить силу найденной регрессионной зависимости на основе

коэффициента детерминации R2;

5) сделать прогноз (при R2>75%) или сделать вывод о невозможности прогнозирования с помощью найденной регрессионной зависимости. При этом не рекомендуется использовать модель регрессии для тех значений независимого параметра X, которые не принадлежат интервалу, заданному в исходных данных.

На рисунке 1 представлено задание варианта 21.



Рисунок 1 – Задание варианта 21

**Ход работы**

Уравнение линейной парной регрессии выглядит следующим образом: Y=a0+а1X. При помощи этого уравнения переменная Y выражается через константу a0 и угол наклона прямой (или угловой коэффициент) а1, умноженный на значение переменной X. Константу a0 также называют свободным членом, а угловой коэффициент – коэффициентом регрессии. Параметры уравнения могут быть определены с помощью метода наименьших квадратов (МНК) Метод наименьших квадратов (в справочных системах англоязычных программ – Least Squares Мethod, LS) является одним из основных методов определения параметров регрессионных уравнений, дающий наилучшие линейные несмещенные оценки. Линейные – относится к характеру взаимосвязи переменных. Несмещенные значит, что ожидаемые значения коэффициентов регрессии должны быть истинными коэффициентами. То есть точки, построенные по исходным данным (xi ,yi) , должны лежать как можно ближе к точкам линии регрессии. Сущность данного метода заключается в нахождении параметров модели, при которых сумма квадратов отклонений эмпирических (фактических) значений результирующего признака от теоретических, полученных по выбранному уравнению регрессии.

Направление связи между переменными определяется на основании знаков коэффициента регрессии. Если знак при коэффициенте регрессии – положительный, связь зависимой переменной с независимой будет положительной. В нашем случае знак коэффициента регрессии положительный, следовательно, связь также является положительной. Если знак при коэффициенте регрессии – отрицательный, связь зависимой переменной с независимой является отрицательной. Для анализа общего качества уравнения уравнения регрессии используют обычно множественный коэффициент детерминации R2, называемый также квадратом коэффициента множественной корреляции R2. (мера определенности) всегда находится в пределах интервала [0;1]. Если значение R 2 близко к единице, это означает, что построенная модель объясняет почти всю изменчивость соответствующих переменных. И наоборот, значение R-квадрата, близкое к нулю, означает плохое качество построенной модели. Коэффициент детерминации R 2 показывает, на сколько процентов ( R2 ⋅100% ) найденная функция регрессии описывает связь между исходными значениями факторов X и Y

Результаты выполнения задания представлен на рисунке 2.

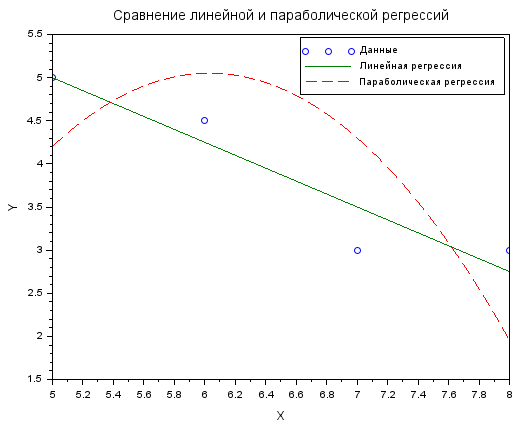


Рисунок 2 – Результат выполнения задания

На рисунке 3 представлены вычисленные значения R2 и прогноз.

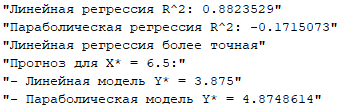


Рисунок 3 – Значения R2 и прогноз

Листинг программы представлен в приложении А.

**Вывод:** в результате выполнения лабораторной работы были освоены основные приемы построения парной линейной регрессии и корреляционного анализа, проверки качества уравнения линейной регрессии и прогнозирования индивидуальных значений зависимой переменной на основании линейной регрессии., используя системы *MathCAD*, *Scilab*, *MS* *Excel*.

**ПРИЛОЖЕНИЕ А**

**Текст программы**

X = [5, 6, 7, 8];

Y = [5, 4.5, 3, 3];

X\_star = 6.5;

n = length(X);

sumX = sum(X);

sumY = sum(Y);

sumXY = sum(X .\* Y);

sumX2 = sum(X .^ 2);

a1\_linear = (n \* sumXY - sumX \* sumY) / (n \* sumX2 - sumX^2);

a0\_linear = (sumY - a1\_linear \* sumX) / n;

// Коэффициент детерминации R^2 для линейной регрессии

Y\_reg\_linear = a0\_linear + a1\_linear \* X;

R2\_linear = 1 - sum((Y - Y\_reg\_linear).^2) / sum((Y - mean(Y)).^2);

Y\_star\_linear = a0\_linear + a1\_linear \* X\_star;

X2 = X.^2;

A = [sum(X2), sum(X), n;

sum(X .\* X2), sum(X2), sum(X);

sum(X.^3), sum(X2), n];

B = [sum(Y); sum(X .\* Y); sum(X2 .\* Y)];

coeffs = A \ B;

a2\_parabola = coeffs(1);

a1\_parabola = coeffs(2);

a0\_parabola = coeffs(3);

// Коэффициент детерминации R^2 для параболической регрессии

Y\_reg\_parabola = a2\_parabola \* X.^2 + a1\_parabola \* X + a0\_parabola;

R2\_parabola = 1 - sum((Y - Y\_reg\_parabola).^2) / sum((Y - mean(Y)).^2);

Y\_star\_parabola = a2\_parabola \* X\_star^2 + a1\_parabola \* X\_star + a0\_parabola;

clf;

X\_fit = linspace(min(X), max(X), 100);

Y\_fit\_linear = a0\_linear + a1\_linear \* X\_fit;

Y\_fit\_parabola = a2\_parabola \* X\_fit.^2 + a1\_parabola \* X\_fit + a0\_parabola;

plot(X, Y, 'o', X\_fit, Y\_fit\_linear, '-', X\_fit, Y\_fit\_parabola, '--');

xlabel('X'); ylabel('Y');

title('Сравнение линейной и параболической регрессий');

legend(['Данные', 'Линейная регрессия', 'Параболическая регрессия'], 'location', 'lower right');

disp("Линейная регрессия R^2: " + string(R2\_linear));

disp("Параболическая регрессия R^2: " + string(R2\_parabola));

if R2\_parabola > R2\_linear then

disp("Параболическая регрессия более точная");

else

disp("Линейная регрессия более точная");

end

disp("Прогноз для X\* = " + string(X\_star) + ":");

disp("- Линейная модель Y\* = " + string(Y\_star\_linear));

disp("- Параболическая модель Y\* = " + string(Y\_star\_parabola));