Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Институт информационных технологий и анализа данных
наименование института

Отделение прикладной математики и информатики
наименование отделения

Отчет по дисциплине
«Вычислительная математика»
по теме:
«Статистическая обработка опытных данных»

Выполнил студент группы	АСУб-20-2		Арбакова А.В.
	Шифр группы	Подпись	И.О. Фамилия
Проверил преподаватель			И.А. Огнёв
		Подпись	И.О. Фамилия

Отчет по НИР защищен с оценкой ________________________

Иркутск 2021 г.

ЗАДАНИЕ

Вариант: 6

Условия задания:

По заданной таблице результирующего признака y и одного из факториальных признаков x_1, x_2, x_3 построить методом наименьших квадратов две различные эмпирические формулы и сравнить качество полученных приближений.

y	55,65	67,68	105,20	85,02	52,76	58,86	72,19	61,09	70,44	51,67
x_1	9,11	9,35	8,90	9,22	8,74	8,98	8,77	9,31	8,81	9,14
x_2	1,52	3,24	6,63	7,15	2,96	1,73	7,44	3,70	2,00	2,63
x_3	2,51	3,74	8,70	5,36	1,89	3,01	3,59	2,64	4,77	1,60

Рисунок 1 – Задание.

Корреляция

Выбираем для признака Y признак X_1 потому что корреляционная оценка показала, что по теоретическим подсчётам выборка Y зависит от выборки X_3 на 88%, что больше, чем от X_2 и от X_1.

Корреляция

	y	*x1*	*x2*	*x3*
y	1			
x1	-0,09454	1		
x2	0,737003	-0,09811	1	
x3	0,978845	-0,1402	0,588451	1

Рисунок 2 – Корреляция зависимости Y от X_1, X_2, X_3.

Анализ данных

Используем метод наименьших квадратов. Мы выполняем регрессионный анализ, используя выборку наблюдений, где a и b – выборочные оценки истинных параметров.

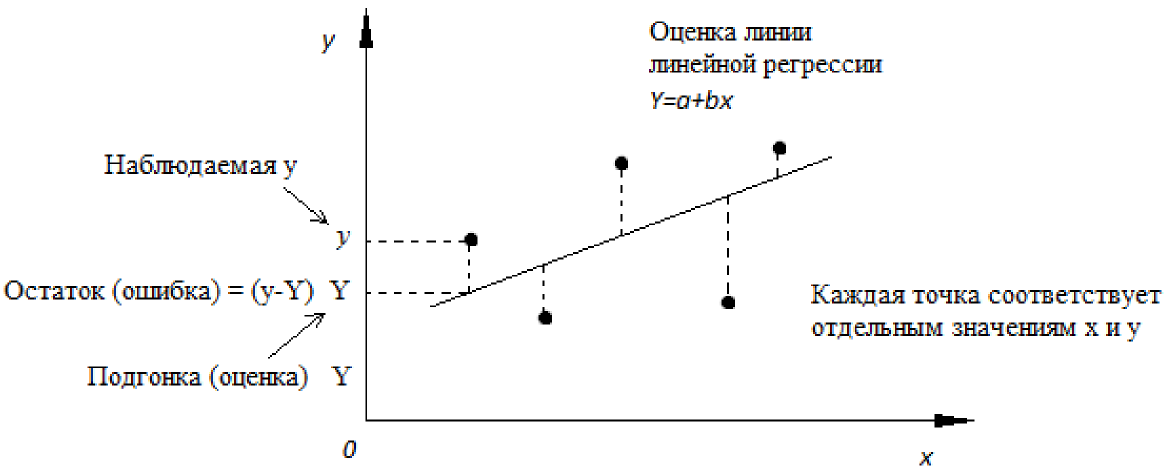

Рисунок 3 – Линия линейной регрессии с изображенными остатками для каждой точки.

Можно применять регрессионную линию для прогнозирования Y значения по значению X в пределе наблюдаемого диапазона.

Итак, если $X = x_{0+}$ прогнозируем Y, как $Y_0 = bx_0 + a$ или $\sum(-y_i + ax_i + + b) = 0$. Используем эту предсказанную величину и ее стандартную ошибку.

Повторение этой процедуры для различных величин x позволяет построить доверительные границы для этой линии. Это полоса или область, которая содержит истинную линию, например, с 95% доверительной вероятностью.

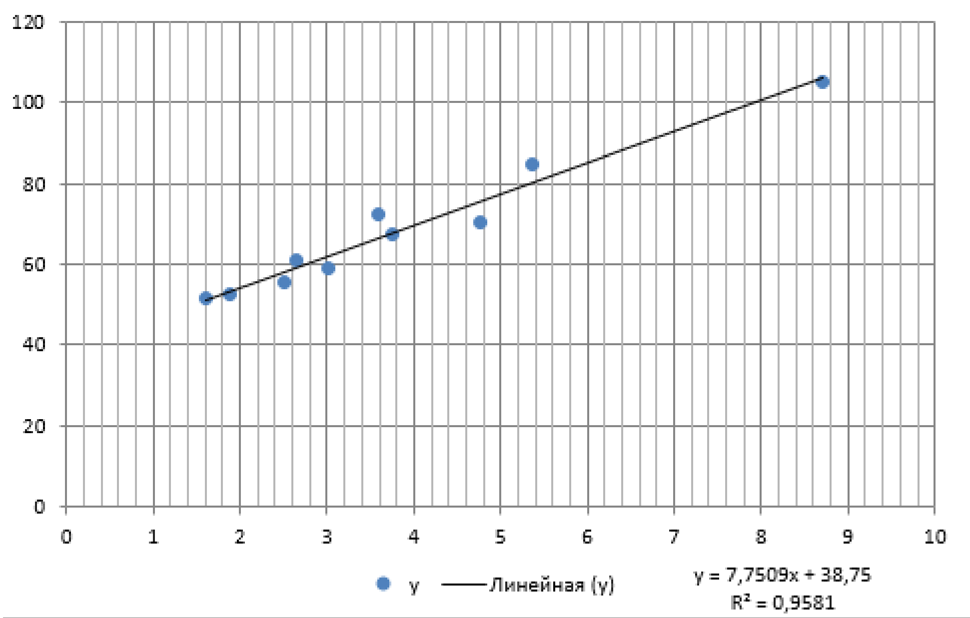

Рисунок 4 – Модель линейной регрессии.

Найдём процент отклонения построенной модели с помощью линейной регрессии. Для этого найдём абсолютное значение разности $Y - Y_0$, и посчитаем частное остатка $\Delta Y = Y - Y_0$ и признака Y, получая $\delta(Y)$.

Δ	δ
103,8135	3,72%

Рисунок 5 – Процент отклонения построенной модели линейной регрессии.

Аналогично находим следующие регрессии, а также проценты отклонения их регрессионных моделей.

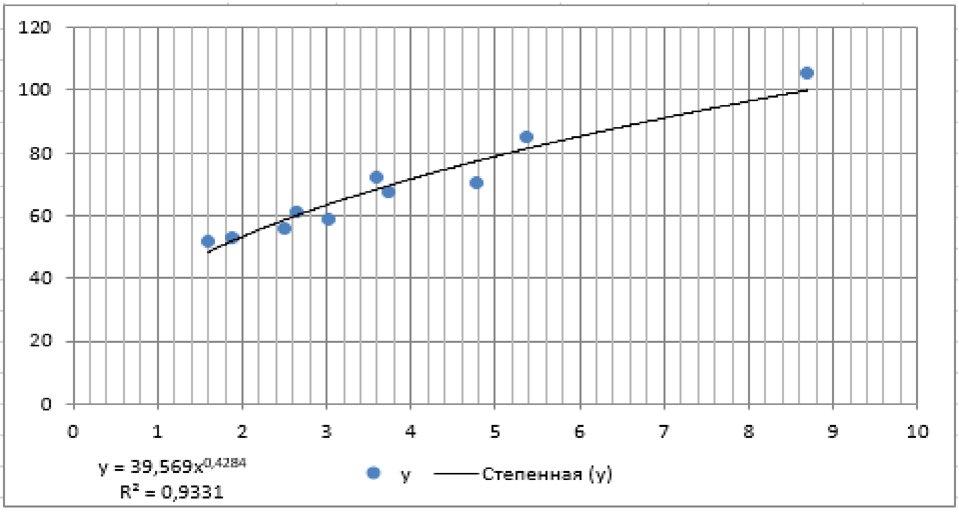

Рисунок 6 – Модель степенной регрессии.

Δ	δ
149,3166	5,01%

Рисунок 7 – Процент отклонения построенной модели степенной регрессии.

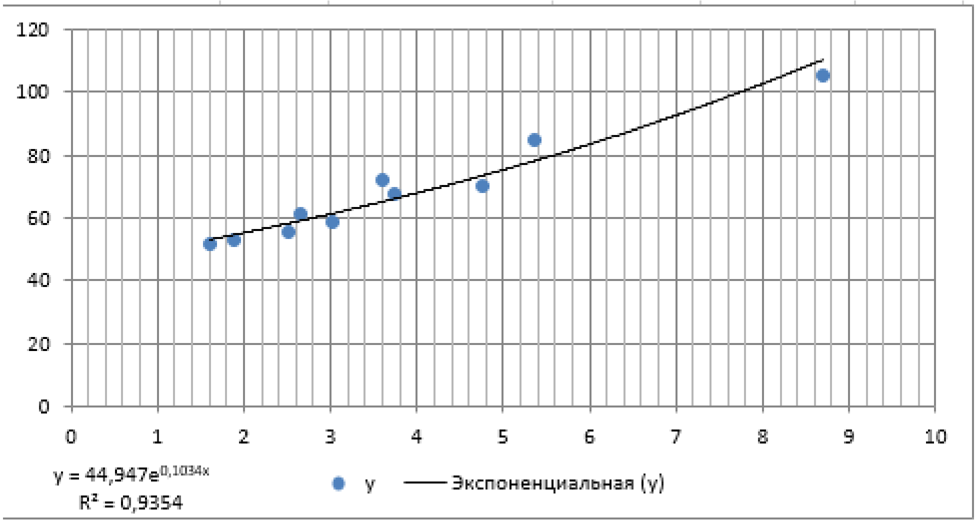

Рисунок 8 – Модель экспоненциальной регрессии.

Δ	δ
158,7016	4,80%

Рисунок 9 – Процент отклонения построенной модели экспоненциальной регрессии.

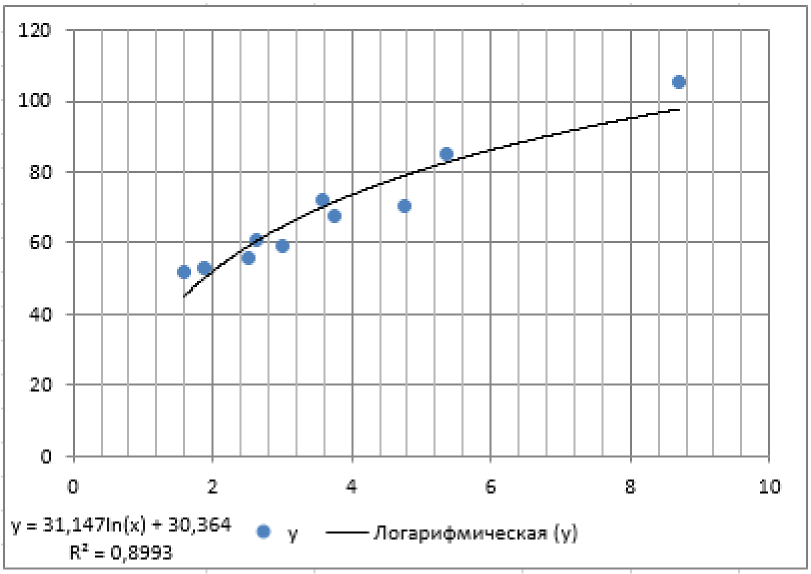

Рисунок 10 – Модель логарифмической регрессии.

Δ	δ
249,788	6,50%

Рисунок 11 – Процент отклонения построенной модели логарифмической регрессии.

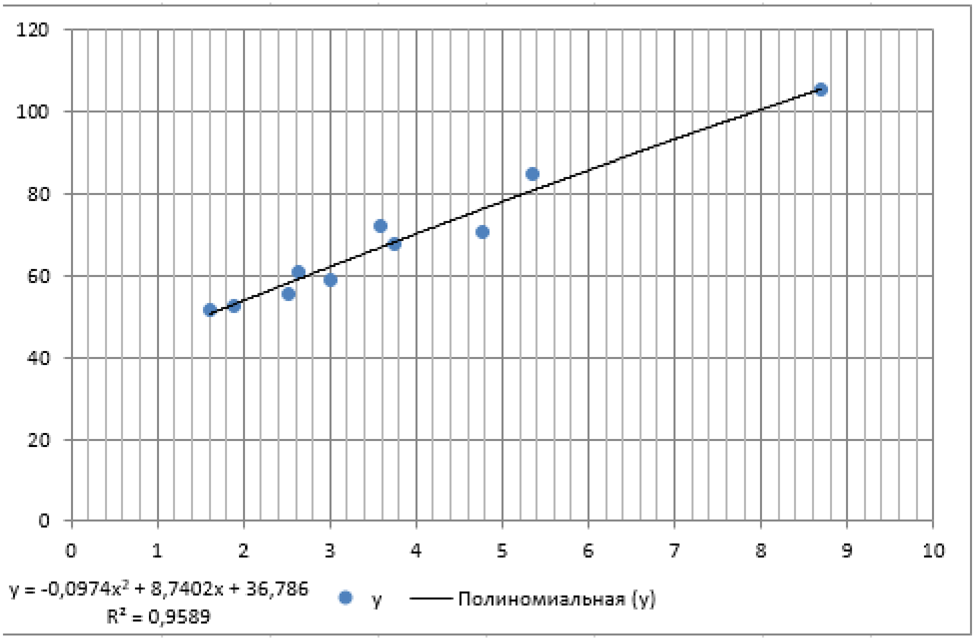

Рисунок 12 – Модель квадратичной регрессии.

Δ	δ
101,8304	3,72%

Рисунок 13 – Процент отклонения построенной модели квадратичной регрессии.

Критерий Фишера

Находим квадрат детерминации, то есть

$$R^2 = \frac{\sum(y_{\text{эксп}} - y_{\text{ср}})^2}{\sum(y_x - y_{\text{ср}})^2}$$

Где в числителе сумма квадратов отклонений прогнозируемых значений от среднего значения переменной, а в знаменателе общая дисперсия прогнозируемых значений.

После находим экспериментальный критерий Фишера $F_{\text{факт}} = \frac{R^2(n-m-1)}{1-R^2}$ и эмпирический критерий Фишера с помощью F-распределения.

Критерий Фишера			
	R^2		
Линейная	0,9581	$F_{\text{факт}}$ =	182,9308
		$F_{\text{табл}}$ =	5,317655
Степенная	0,9331	$F_{\text{факт}}$ =	111,5815
		$F_{\text{табл}}$ =	5,317655
экспоненциальная	0,9354	$F_{\text{факт}}$ =	115,839
		$F_{\text{табл}}$ =	5,317655
логарифмическая	0,8993	$F_{\text{факт}}$ =	71,44389
		$F_{\text{табл}}$ =	5,317655
квадратичная	0,9589	$F_{\text{факт}}$ =	186,6472
		$F_{\text{табл}}$ =	5,317655

Рисунок 16 – Критерий Фишера.

Вывод

Решая поставленную задачу, мы удостоверились с коэффициентом корреляции 0,97, что выборка x_3 подходит лучше, чем выборки x_1 и x_2. Мы построили модели регрессии, а именно: линейной, степенной, экспоненциальной, логарифмической, полиномиальной (квадратичной).

Оказалось, что, полиномиальная регрессия подходит больше всего, как регрессионная модель с процентом отклонения 3,72%. Также мы сделали критерий Фишера, который оценил проверку равенства дисперсий двух выборок и показал, что $F_{\text{факт}} > F_{\text{табл}}$ – это означает, построенная нами модель является статистически значимой регрессией.