Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский университет ИТМО»
Факультет Программной Инженерии и Компьютерной Техники

Вариант №76
Лабораторная работа №7
по теме
Построение оценки линейной регрессии
по дисциплине
Математическая статистика

Выполнил Студент группы P3212
Кобелев Р.П.
Балин А.А.
Пархоменко К.А.
к. т. н. Преподаватель:
Милованович Е.А.

г. Санкт-Петербург
2024г.

Цель работы

Цель работы:

На основании анализа двумерной выборки

1. Построить точечную оценку линейной функции регрессии по методу средних и методу наименьших квадратов.

2. Проверить статистическую гипотезу об адекватности выбранной модели эксперементальным данным.

3. Построить доверительные интервалы для коэффиценков функции регрессии и для всей функции.

Данные

Закон: Закон распределения прямоугольного треугольника

Выборка X: 9.0 15.0 22.0 27.0 34.0 39.0 48.0 57.0 66.0

Выборка Y: $-6.9\ -7.4\ -11.7\ -14.3\ -19.3\ -24.4\ -36.6\ -40.1\ -41.1$
Объём выборки n $= 9$
Доверительная вероятность $\beta = 0.95$

Линейная модель

Формула:

$$y = a_0 + a_1 x$$

Метод средних:

Исходя из таблицы, составили уравнения и сложили первые 4 и последнии 5
1)

$$-6.9 = \tilde{a_0} + 9\tilde{a_1}$$
$$+$$
$$-7.4 = \tilde{a_0} + 15\tilde{a_1}$$
$$+$$
$$-11.7 = \tilde{a_0} + 22\tilde{a_1}$$
$$+$$
$$-14.3 = \tilde{a_0} + 27\tilde{a_1}$$
$$=$$
$$-40.3 = 4\tilde{a_0} + 73\tilde{a_1}$$

2)

$$-19.3 = \tilde{a_0} + 34\tilde{a_1}$$
$$+$$
$$-24.4 = \tilde{a_0} + 39\tilde{a_1}$$
$$+$$
$$-36.6 = \tilde{a_0} + 48\tilde{a_1}$$
$$+$$
$$-40.1 = \tilde{a_0} + 57\tilde{a_1}$$
$$+$$
$$-41.1 = \tilde{a_0} + 66\tilde{a_1}$$

$$=$$

$$-161.5 = 5\tilde{a_0} + 244\tilde{a_1}$$

3)

$$\begin{cases} 4\tilde{a_0} + 73\tilde{a_1} = -40.3 \\ 5\tilde{a_0} + 244\tilde{a_1} = -161.5 \end{cases}$$

$$\begin{cases} \tilde{a_0} \approx 3.2 \\ \tilde{a_1} \approx -0.727 \end{cases}$$

Получена точечная оценка:

$$\tilde{y} = 3.2 - 0.727x$$

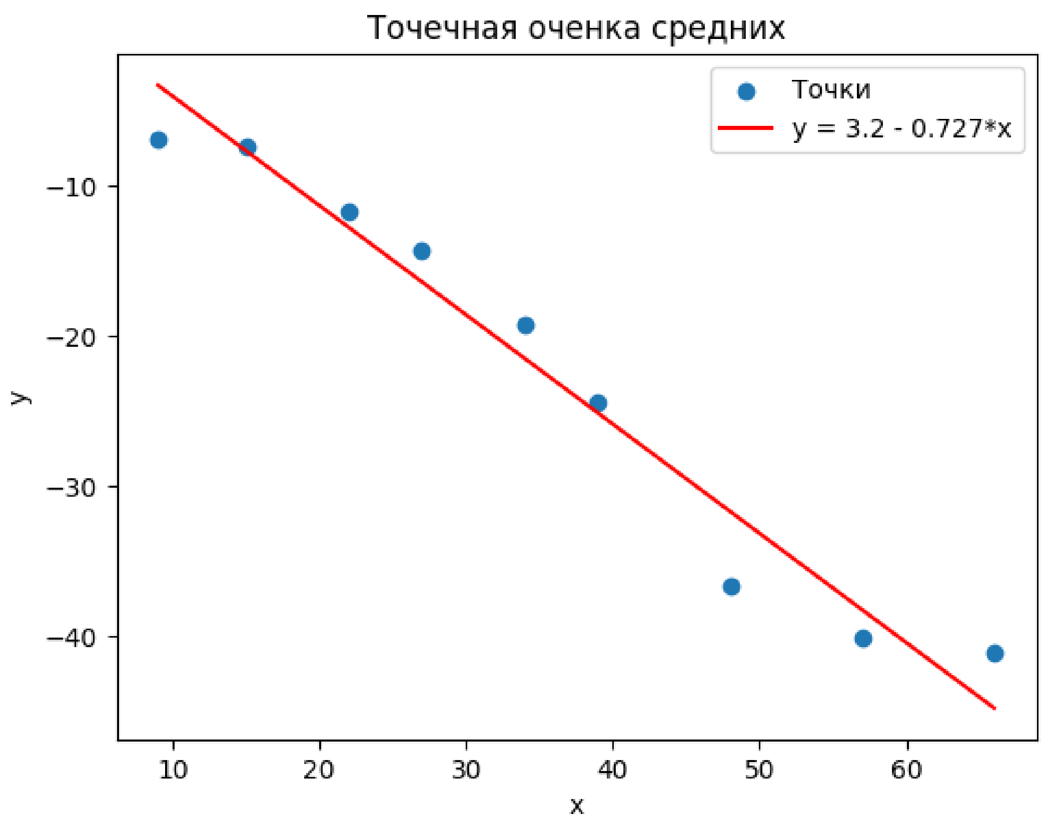

График 1. Точечная оценка метод средних

Метод наименьших квадратов для линейной модели:

$$S(a_0,\ a_1) = \sum_{i=1}^{n}(y_i - \tilde{y}(x_i))^2 = \sum_{i=1}^{9}(y_i - \tilde{a_0} - \tilde{a_1}x_i)^2 -> \ min$$

Найдём экстремум:

$$\begin{cases} \frac{\partial S}{\partial a_0} = -2\left(\sum_{i=1}^{9} y_i - 9\tilde{a_0} - \tilde{a_1}\sum_{i=1}^{9} x_i\right) = 0 \\ \frac{\partial S}{\partial a_1} = -2\left(\sum_{i=1}^{9} x_i y_i - \tilde{a_0}\sum_{i=1}^{9} x_i - \tilde{a_1}\sum_{i=1}^{9} x_i^2\right) = 0 \end{cases}$$

После подсчёта сумм получили систему:

$$\begin{cases} 9\tilde{a_0} + 317\tilde{a_1} = -201.8 \\ 317\tilde{a_0} + 14105\tilde{a_1} = -9179.5 \end{cases}$$

Подсчитали неизвестные:

$$\begin{cases} \tilde{a_0} = 2.4 \\ \tilde{a_1} = -0.705 \end{cases}$$

Подставили коэффиценты и получили точечную оценку:

$$\tilde{y} = 2.4 - 0.705x$$

$$S_{min}^{(1)} = (-6.9 - 2.4 + 0.705 \cdot 9)^2 + (-7.4 - 2.4 + 0.705 \cdot 15)^2 + (-11.7 - 2.4 + 0.705 \cdot 22)^2 +$$
$$(-14.3 - 2.4 + 0.705 \cdot 27)^2 + (-19.3 - 2.4 + 0.705 \cdot 34)^2 + (-24.4 - 2.4 + 0.705 \cdot 39)^2 +$$
$$(-36.6 - 2.4 + 0.705 \cdot 48)^2 + (-40.1 - 2.4 + 0.705 \cdot 57)^2 + (-41.1 - 2.4 + 0.705 \cdot 66)^2$$
$$\approx 63.575$$

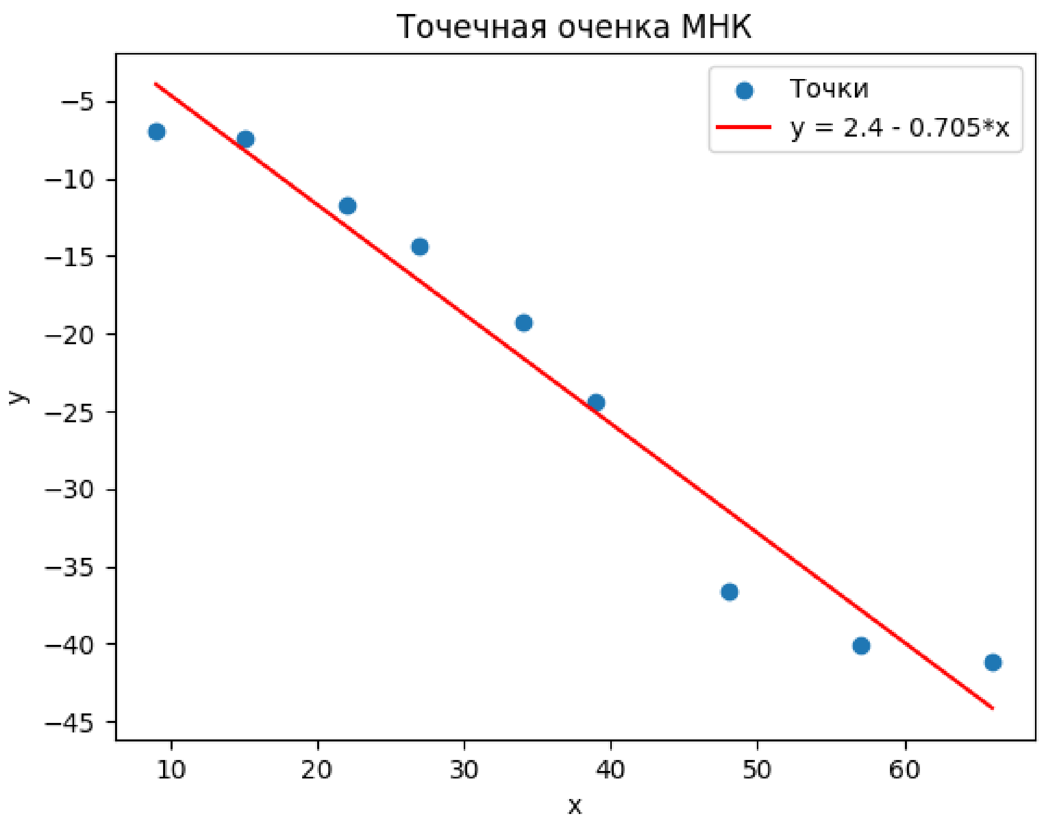

График 2. Точечная оценка МНК

Квадратичная модель

Формула:

$$y = a_0 + a_1 x + a_2 x^2$$

$$S(a_0, a_1, a_2) = \sum_{i=1}^{n} (y_i - \tilde{y}(x_i))^2 = \sum_{i=1}^{9} (y_i - \tilde{a_0} - \tilde{a_1} x_i - \tilde{a_2} x_i^2)^2 -> \ min$$

$$\begin{cases} \frac{\partial S}{\partial a_0} = -2 \left(\sum_{i=1}^{9} y_i - 9\tilde{a_0} - \tilde{a_1} \sum_{i=1}^{9} x_i - \tilde{a_2} \sum_{i=1}^{9} x_i^2 \right) = 0 \\ \frac{\partial S}{\partial a_1} = -2 \left(\sum_{i=1}^{9} x_i y_i - \tilde{a_0} \sum_{i=1}^{9} x_i - \tilde{a_1} \sum_{i=1}^{9} x_i^2 - \tilde{a_2} \sum_{i=1}^{9} x_i^3 \right) = 0 \\ \frac{\partial S}{\partial a_2} = -2 \left(\sum_{i=1}^{9} x_i^2 y_i - \tilde{a_0} \sum_{i=1}^{9} x_i^2 - \tilde{a_1} \sum_{i=1}^{9} x_i^3 - \tilde{a_2} \sum_{i=1}^{9} x_i^4 \right) = 0 \end{cases}$$

$$\begin{cases} 5\tilde{a_0} + 317\tilde{a_1} + 14105\tilde{a_2} = -201.8 \\ 317\tilde{a_0} + 14105\tilde{a_1} + 716339\tilde{a_2} = -9179.5 \\ 14105\tilde{a_0} + 716339\tilde{a_1} + 39311813\tilde{a_2} = -471377.5 \end{cases}$$

$$\begin{cases} \tilde{a_0} = -0.28 \\ \tilde{a_1} = -0.545 \\ \tilde{a_2} = -0.002 \end{cases}$$

Получена точечная оценка:

$$\tilde{y} = -0.28 - 0.545x - 0.002x^2$$

$$S_{min}^{(2)} = 24074.209$$

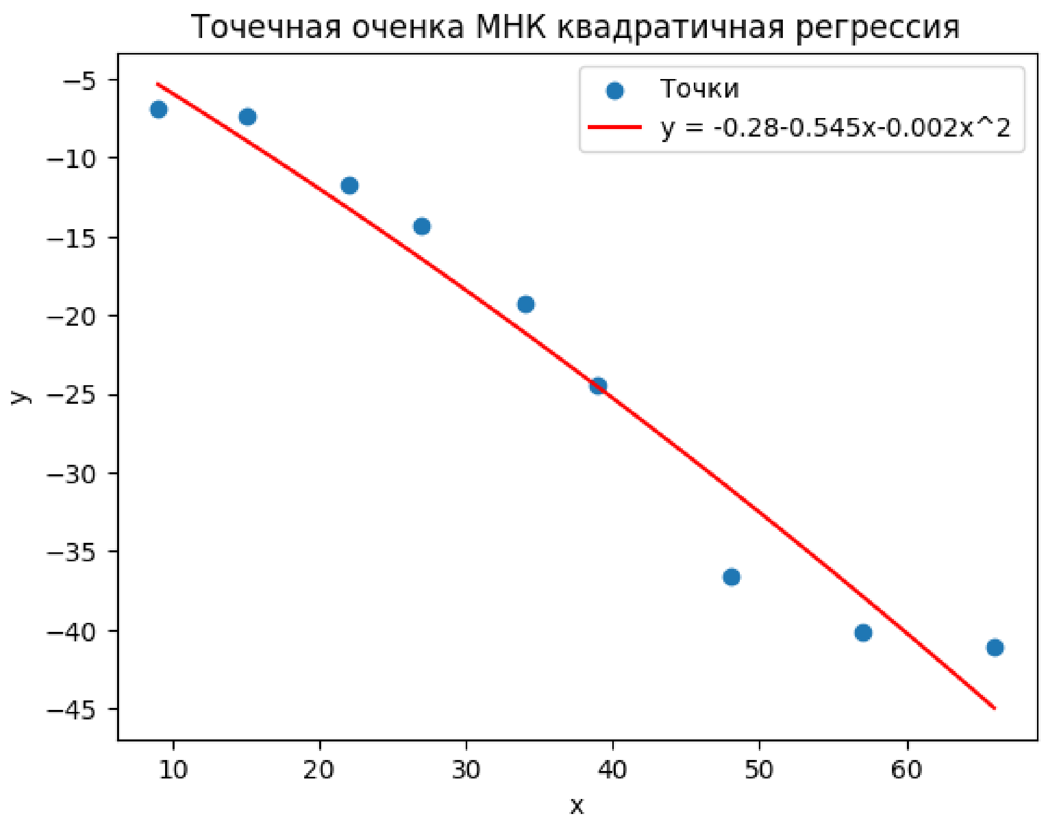

График 3. Точечная оценка МНК квадратичная регрессия

Сравнение графиков:

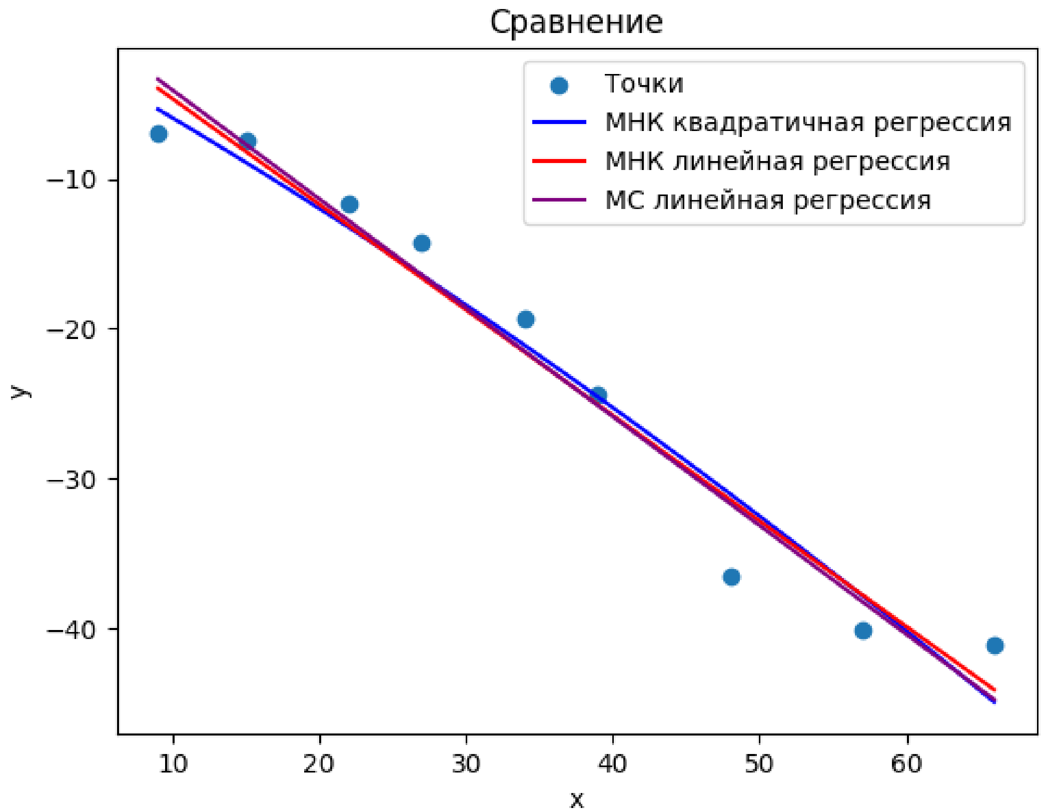

График 4. Сравнение МНК квадратичная регрессия(синий), МНК линейная регрессия(красный), МС линейная регрессия(фиолетовый)

Гипотеза

Проверка гипотезы об адекватности модели в задаче регрессии:

H_0 : Линейная модель хорошо согласуется с данными эксперемента и можно для дальнейшего исследования оставить её. Переход к квадратичной не требуется.

H_1 : Линейная модель плохо согласуется с данными эксперемента и можно для дальнейшего исследования оставить её. Переход к квадратичной требуется.

Введём статистический критерий Фишера:

$$F = \frac{\frac{1}{k-m}(S^{(1)}_{min} - S^{(2)}_{min})}{\frac{1}{n-k-1}S^{(2)}_{min}}$$

$$k = 2;\ m = 1;\ n = 9$$

(n - количество эксперементальных данных)
По теореме Фишера с уровнем значимости $\alpha = 0.05$ и степенями свободы $r_1 = k - m = 1$ и $r_2 = n - k - 1 = 6$.
По таблице найдём:

$$F_{kr} = 5.99$$

$$F = \frac{\frac{1}{1}(63.575 - 24074.209)}{\frac{1}{6} \cdot 24074.209} \approx -5.98$$

Получили что F входит в допустимую область:

$$O = (0, F_{kr}) = (0, 5.99)$$

Тогда H_0 принимается и мы оставляем линейную модель.

Интервальные оценки параметров и функции регрессии

$$y_i = a_0 + a_1 x_i + \varepsilon_i$$

ε_i– ошибка измерения. Будем считать измерения равноточными.

$$\varepsilon_i \in N(0, D(\varepsilon_i) = \sigma^2)$$

$$\tilde{\sigma}^2 = \frac{S_{min}}{n-2} = \frac{63.575}{9-2} = 9.082$$

Определим оценку метрицы корреляционных моментов:

$$\tilde{K} = \begin{pmatrix} \tilde{\sigma}^2[\tilde{a_0}] & \tilde{K}[\tilde{a}_0, \tilde{a}_1] \\ \tilde{K}[\tilde{a}_0, \tilde{a}_1] & \tilde{\sigma}^2[\tilde{a_1}] \end{pmatrix} = \tilde{\sigma}^2 P^{-1}$$

$$P = \begin{pmatrix} 9 & \sum_{i=1}^{9} x_i \\ \sum_{i=1}^{9} x_i & \sum_{i=1}^{9} x_i^2 \end{pmatrix} = \begin{pmatrix} 9 & 317 \\ 317 & 14105 \end{pmatrix}$$

$$P^{-1} = \frac{1}{detP} \cdot \begin{pmatrix} 14105 & -317 \\ -317 & 9 \end{pmatrix} = \frac{1}{9 \cdot 14105 - 317^2} \begin{pmatrix} 14105 & -317 \\ -317 & 9 \end{pmatrix} = \frac{1}{26456} \begin{pmatrix} 14105 & -317 \\ -317 & 9 \end{pmatrix}$$

Получили:

$$\tilde{\sigma}^2[\tilde{a_0}] = \frac{9.082}{26456} \cdot 14105 \approx 4.842$$

$$\tilde{\sigma}^2[\tilde{a_1}] = \frac{9.082}{26456} \cdot 9 \approx 0.003$$

$$\tilde{K}^2[\tilde{a_0}, \tilde{a_1}] = \frac{9.082}{26456} \cdot (-317) \approx -0.109$$

Оценка параметров:

По теореме Стьюдента с доверительной вероятностью $\beta = 0.9$ и степенью свободы $r = 3$ нашли по таблице:

$$t_{0.9;3} = 1.638$$

Получили следующие оценки: Для параметра a_0:

$$2.4 - 1.638\sqrt{4.842} < a_0 < 2.4 + 1.638\sqrt{4.842}$$

$$-1.204 < a_0 < 6.004$$

Для параметра a_1:

$$-0.705 - 1.638\sqrt{0.003} < a_1 < -0.705 + 1.638\sqrt{0.003}$$

$$-0.795 < a_1 < -0.615$$

Оценим функцию

$$\tilde{\sigma}^2[\tilde{y}(x)] = \tilde{\sigma}^2[\tilde{a_0}] + 2\tilde{K}[\tilde{a_0}, \tilde{a_1}]x + \tilde{\sigma}^2[\tilde{a_1}]x^2 = 4.842 - 0.218x + 0.003x^2$$

Доверительный интервал на функцию регрессии:

$$2.4 - 1.204x - 1.638\sqrt{4.842 - 0.218x + 0.003x^2} < M[y/x] < 35.054 + 0.124x + 1.638\sqrt{4.842 - 0.218x + 0.003x^2}$$

Для $x_1 = 9$:

$$-11.3 < M[y/x] < 37.481$$

Для $x_2 = 15$:

$$-18.12 < M[y/x_2] < 39.37$$

Для $x_3 = 22$:
$$-26.09 < M[y/x_3] < 39.79$$
Для $x_4 = 27$:
$$-31.86 < M[y/x_4] < 40.15$$
Для $x_5 = 34$:
$$-40.09 < M[y/x_5] < 40.82$$
Для $x_6 = 39$:
$$-46.11 < M[y/x_6] < 41.45$$
Для $x_7 = 48$:
$$-57.25 < M[y/x_7] < 42.87$$
Для $x_8 = 57$:
$$-68.64 < M[y/x_8] < 44.53$$
Для $x_9 = 66$:
$$-80.14 < M[y/x_9] < 46.31$$

Итоговый график

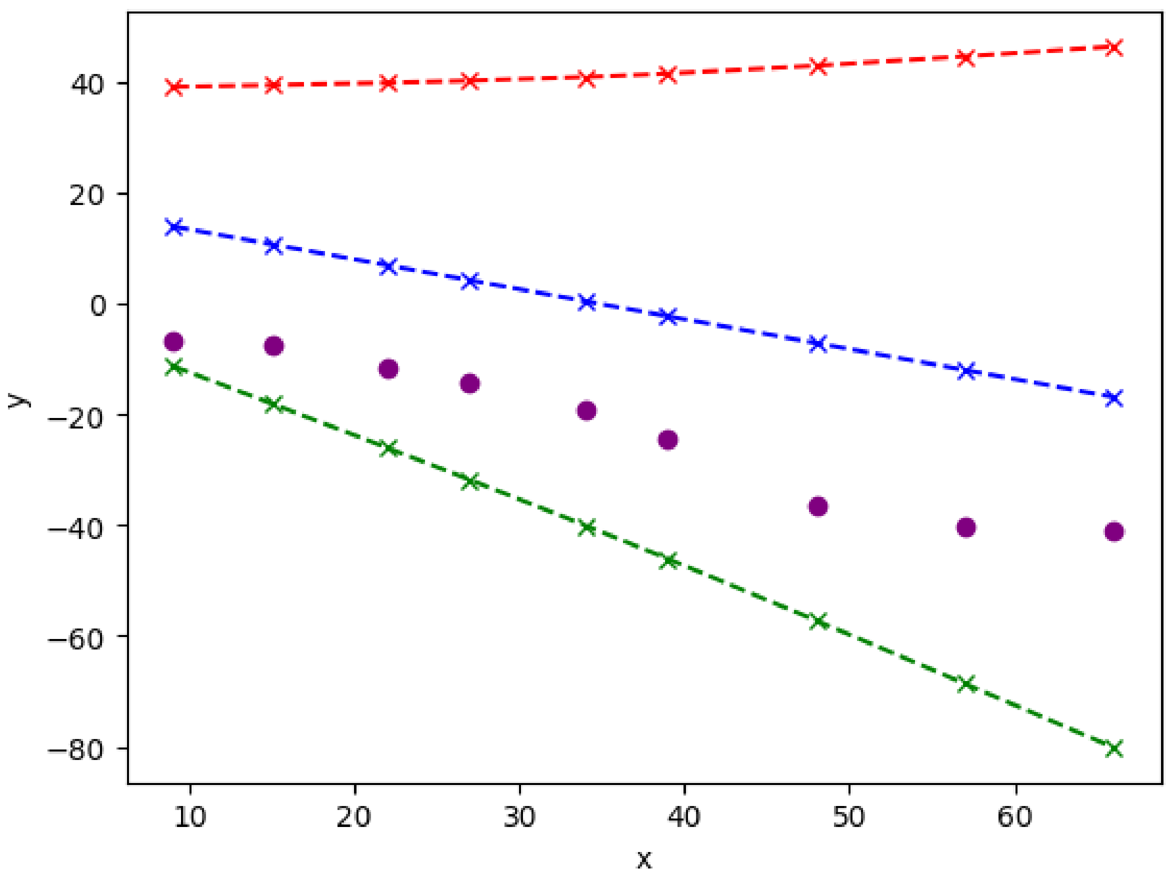

График 5. Доверительная область

Вывод

Используя метод наименьших квадратов, сгладили предложенную табличную зависимость при помощи формул. Помимо этого, вычислили невязки с точностью до сотых и отобразить на графике табличные данные и сглаживающую кривую.