МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

ФАКУЛЬТЕТ ПРОГРАММНОЙ ИНЖЕНЕРИИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ

ЛАБОРАТОРНАЯ РАБОТА №7

по дисциплине
«Математическая статистика»
Вариант № 92

Выполнили:
Векшин А. И. P3216
Дашкевич Е.В. P3208
Кононова В.В. P3211
Преподаватель:
Танченко Ю.В.

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 экспериментальным данным
3) Построить доверительные интервалы для коэффициентов и для всей функции (β=0,95)

Исходные данные

$x(i)$	4,0	11,0	16	25	32	41	49
$y(i)$	24,6	22,1	21,4	17,7	14,7	11,8	10,1

n = 7

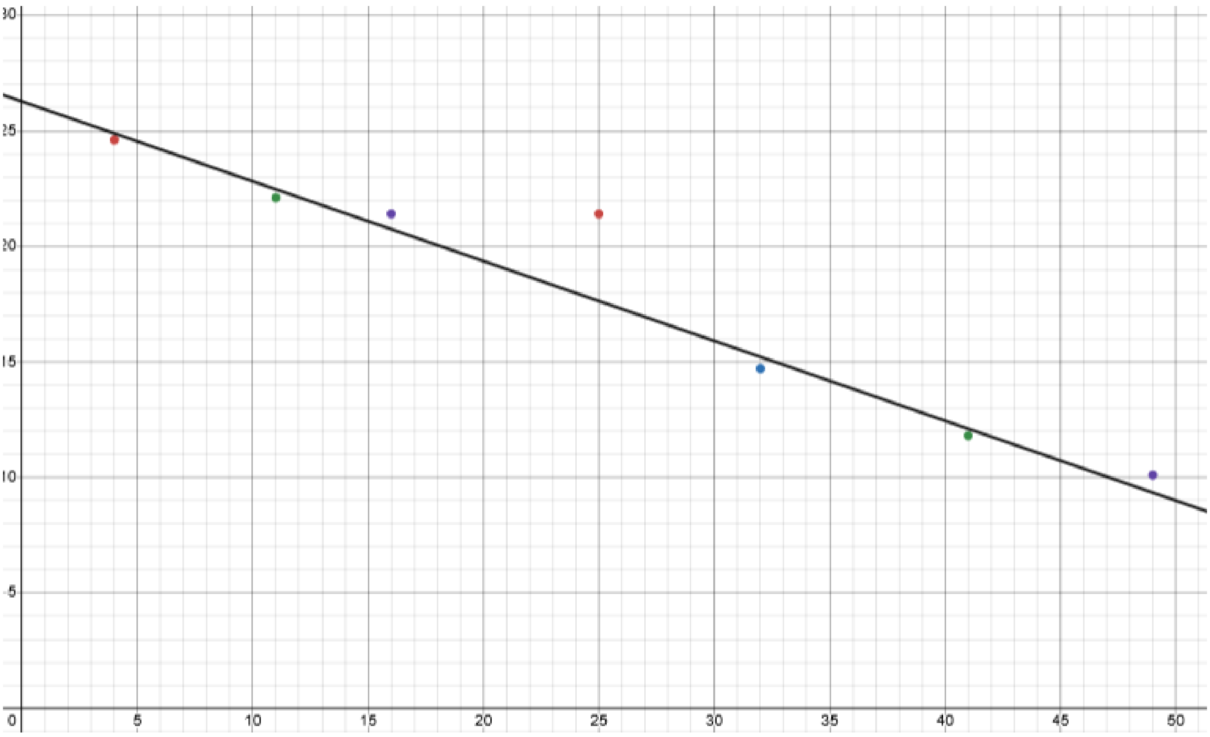
Линейная модель

Примем за модель линейную, с формулой $y = a + bx$

Метод средних

Из таблицы получим систему:

$$\begin{cases} 68.1 = 3a + 31b \\ 54.3 = 4a + 147b \end{cases} \quad \text{решим систему, получив точечную оценку :} \begin{cases} a = 26.2694 \\ b = -0.3454 \end{cases}$$



Метод наименьших квадратов

$$S(a,\ b) = \sum_{i=1}^{n} \left(y_i - \tilde{y}(x_i)\right)^2 = \sum_{i=1}^{7} \left(y_i - \tilde{a} - \tilde{b}x_i\right)^2 \ \to min$$

Найдем экстремум:

$$\frac{\partial S}{\partial a} = -2\left(\sum_{i=1}^{7} y_i - 7\tilde{a} - \tilde{b}\sum_{i=1}^{7} x_i\right) = 0$$

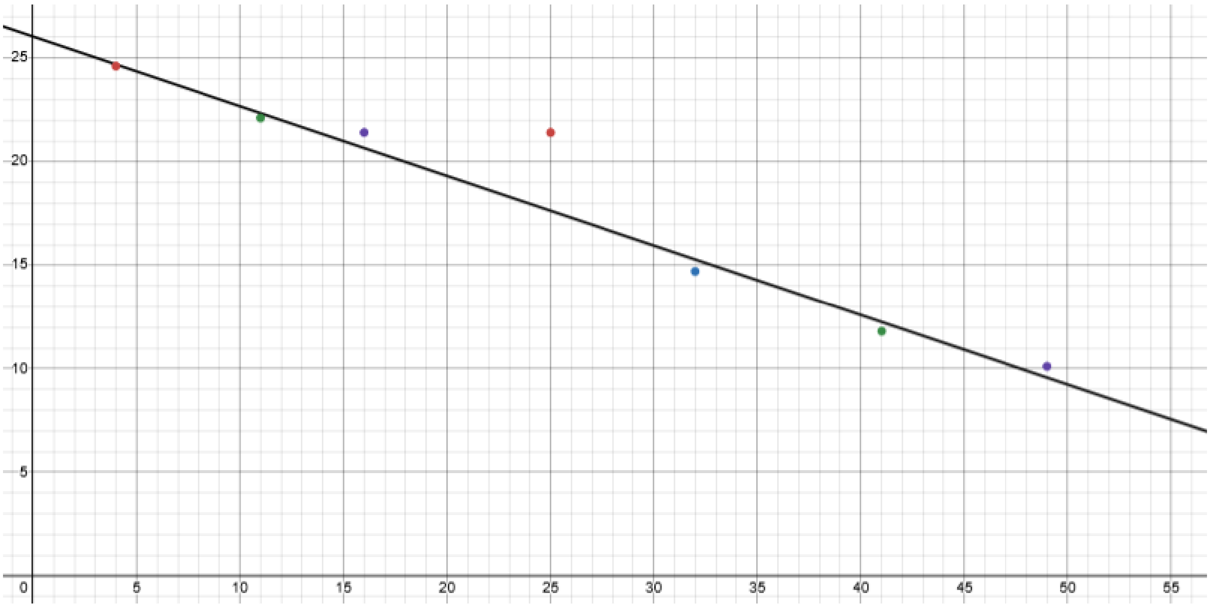
$$\frac{\partial S}{\partial a} = -2\left(\sum_{i=1}^{7} x_i\, y_i - \tilde{a}\sum_{i=1}^{7} x_i - \tilde{b}\sum_{i=1}^{7} x_i^2\right) = 0$$

Получим систему

$$\begin{cases} 7a + 178b = 122.4 \\ 178a +\ 6124b = 2575.5 \end{cases} \quad \to \quad \begin{cases} a = -0.3360 \\ b = 26.0317 \end{cases}$$

Подсчитаем отклонение:

$$S(a,\ b) = \sum_{i=1}^{n} \left(y_i - \tilde{y}(x_i)\right)^2 = \sum_{i=1}^{7} \left(y_i - \tilde{a} - \tilde{b}x_i\right)^2 \approx 1.449$$



Квадратичная модель

Формула: $y = ax^2 + bx + c$

Метод наименьших квадратов

$$S(a,b,c) = \sum_{i=1}^{n}\left(y_i - \tilde{y}(x_i)\right)^2 = \sum_{i=1}^{7}\left(y_i - \tilde{c} - \tilde{b}x_i - \tilde{a}x_i^2\right)^2 \ \rightarrow \ min$$

Найдем экстремум:

$$\frac{\partial S}{\partial a} = -2\left(\sum_{i=1}^{7} x_i^2\, y_i - \tilde{c}\sum_{i=1}^{7} x_i^2 - \tilde{b}\sum_{i=1}^{7} x_i^3 - \tilde{a}\sum_{i=1}^{7} x_i^4\right) = 0$$

$$\frac{\partial S}{\partial b} = -2\left(\sum_{i=1}^{7} x_i\, y_i - \tilde{c}\sum_{i=1}^{7} x_i - \tilde{b}\sum_{i=1}^{7} x_i^2 - \tilde{a}\sum_{i=1}^{7} x_i^3\right) = 0$$

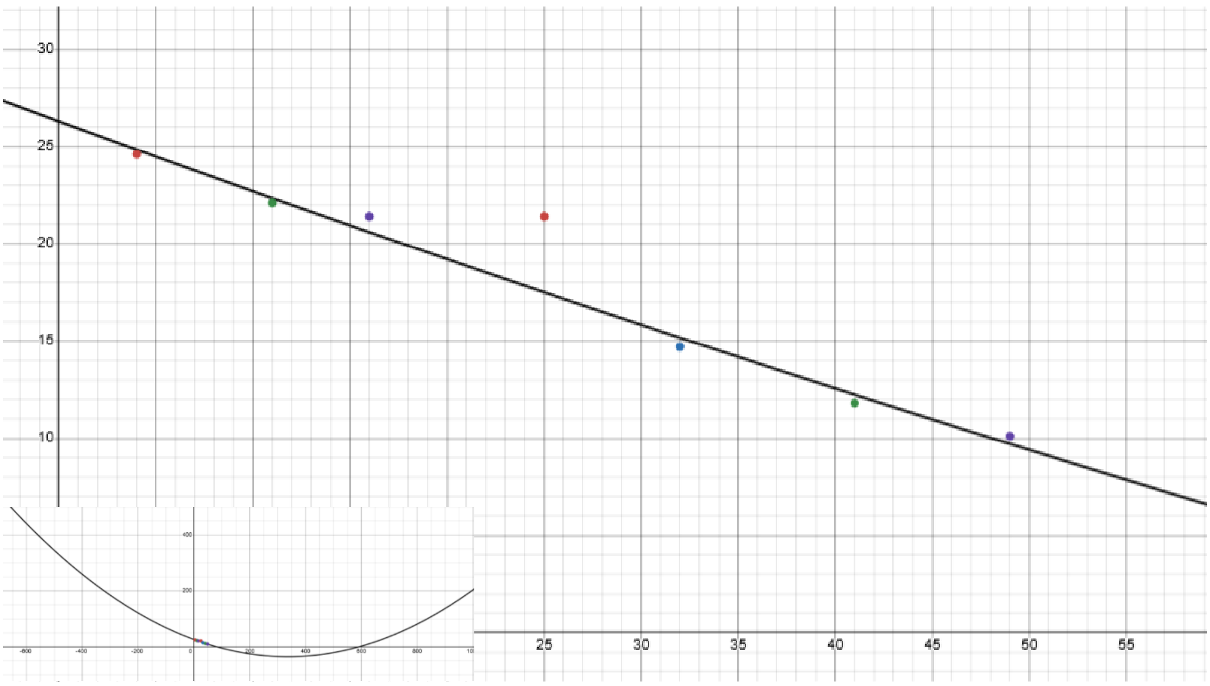
$$\frac{\partial S}{\partial c} = -2\left(\sum_{i=1}^{7} y_i - 7\tilde{c} - \tilde{b}\sum_{i=1}^{7} x_i - \tilde{a}\sum_{i=1}^{7} x_i^2\right) = 0$$

Получим систему

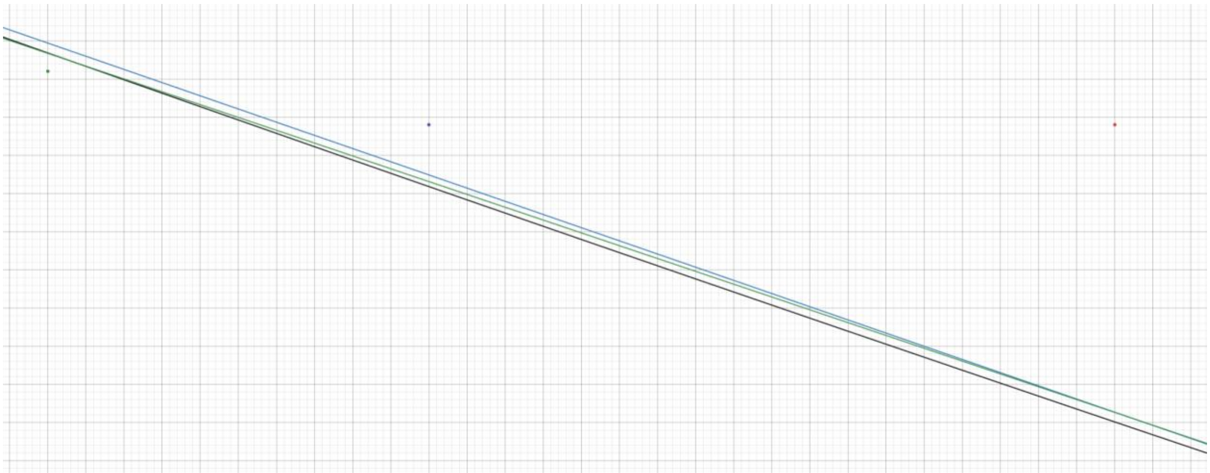
$$\begin{cases} 10110196\,a + 240454\,b + 6124\,c = 78747{,}3 \\ 240454\,a + 6124\,b + 178\,c = 2575.5 \\ 6124\,a + 178\,b + 7\,c = 122.4 \end{cases} \rightarrow \begin{cases} a = 0{,}0005420308213 \\ b = -0{,}3648231163 \\ c = 26{,}28844542 \end{cases}$$

Подсчитаем отклонение:

$$S(a,b,c) = \sum_{i=1}^{n}\left(y_i - \tilde{y}(x_i)\right)^2 = \sum_{i=1}^{7}\left(y_i - \tilde{c} - \tilde{b}x_i - \tilde{a}x_i^2\right)^2 \ \approx 1{,}343133$$



Сравнение графиков



Черный – квадратичная линейная регрессия
Синий – МС линейная регрессия
Зеленый – МНК линейная регрессия

https://www.desmos.com/calculator/fmlesbswia?lang=ru

Проверка гипотезы

Проверка гипотезы об адекватности модели в задаче регрессии:

- H0 : Линейная модель хорошо согласуется с данными эксперемента и можно для дальнейшего исследования оставить её. Переход к квадратичной не требуется.
- H1 : Линейная модель плохо согласуется с данными эксперемента и можно для дальнейшего исследования оставить её. Переход к квадратичной требуется.

Введём статистический критерий Фишера:

$$\mathrm{F} = \frac{\frac{1}{k-m}\left(S_{min}^{(1)} - S_{min}^{(2)}\right)}{\frac{1}{n-k-1}S_{min}^{(2)}} \quad \text{где } \mathrm{k} = 2;\ \mathrm{m} = 1;\ \mathrm{n} = 7$$

По т. Фишера с уровнем значимости $\alpha = 0.05$ и степенями свободы $\mathrm{r}_1 = k - m = 1$ и $r_2 = n - k - 1 = 4$, по таблице найдем

$$\mathrm{F}_{kr} = 7.71$$

$$F = \frac{\frac{1}{1}(1.449 - 1{,}343133)}{\frac{1}{4}\cdot 1{,}343133} \approx 0.315237$$

Так как $0 < F < F_{kr}$, то гипотеза H0 принята, используем линейную модель.

Интервальные оценки параметров регрессии

$$y_i = a_0 + a_1 x_i + \varepsilon_i \quad \text{где } \varepsilon_i - \text{ошибка измерения.}$$

$$\varepsilon_i \in N(0, D(\varepsilon_i) = \sigma^2) \qquad \widetilde{\sigma^2} = \frac{S_{min}}{n-2} = \frac{1{,}449}{7-2} = 0{,}2898$$

Определим оценку метрик корреляционных моментов:

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

$$P = \begin{pmatrix} 7 & \sum_{i=1}^{7} x_i \\ \sum_{i=1}^{7} x_i & \sum_{i=1}^{7} x_i^2 \end{pmatrix} = \begin{pmatrix} 7 & 178 \\ 178 & 6124 \end{pmatrix}$$

$$P^{-1} = \frac{1}{detP}\cdot\begin{pmatrix} 6124 & -178 \\ -178 & 7 \end{pmatrix} = \frac{1}{11184}\cdot\begin{pmatrix} 6124 & -178 \\ -178 & 7 \end{pmatrix}$$

Получим:

$$\widetilde{\sigma^2}[\widetilde{a_0}] = \frac{0{,}2898}{11184} * 6124 \approx 0{,}15869 \qquad \widetilde{\sigma^2}[\widetilde{a_1}] = \frac{0{,}2898}{11184} * 7 \approx 0{,}00018$$

$$\widetilde{K^2}[\widetilde{a_0}, \widetilde{a_1}] = \frac{0{,}2898}{11184}\cdot(-178) \approx -0{,}00461$$

Оценим параметры.

По теореме Стьюдента с доверительной вероятностью β = 0.95 и степенью свободы r=5 из таблицы получим $t_{0.95\ \ 5} = 2.015$. Отсюда получим оценки:

$$\widetilde{a_0} - t_{0.95\,;5}\sqrt{\widetilde{\sigma^2}[\widetilde{a_0}]} < a_0 < \widetilde{a_0} + t_{0.95\,;5}\sqrt{\widetilde{\sigma^2}[\widetilde{a_0}]} \qquad 25.2289 < a_0 < 26.8344$$

$$\widetilde{a_1} - t_{0.95\,;5}\sqrt{\widetilde{\sigma^2}[\widetilde{a_1}]} < a_1 < \widetilde{a_1} + t_{0.95\,;5}\sqrt{\widetilde{\sigma^2}[\widetilde{a_1}]} \qquad -0{,}36311 < a_1 < -0{,}3090$$

Оценим функцию:

$$\widetilde{\sigma^2}[\tilde{y}(x)] = \widetilde{\sigma^2}[\widetilde{a_0}] + 2\widetilde{K}[\widetilde{a_0}, \widetilde{a_1}]x + \widetilde{\sigma^2}[\widetilde{a_1}]x^2 = 0{,}15869 - 0{,}00922\,x + 0{,}00018x^2$$

x	y-ε	y + ε
4	23,9758	25,3989
11	21,7682	22,9013
16	20,1723	21,1365
25	17,2232	18,0362
32	14,8358	15,7185
41	11,6710	12,8338
49	8,81236	10,3153