Университет ИТМО

Факультет программной инженерии и компьютерной техники

Индивидуальное домашнее задание №7

по «Математической статистике»

Вариант 84

Выполнили:

Студенты группы P3233

Хасаншин Марат

Шикунов Максим

Номер команд: 9

Санкт-Петербург

2024

Цель работы

На основании анализа данных двух выборок X и Y из нормально распределенных генеральных совокупностей проверить статистическую гипотезу о равенстве математических ожиданий этих совокупностей, предполагая, что их дисперсии равны.

Исходные данные

x(i)	7.0	14.0	22.0	30.0	39.0	48.0
y(i)	28.2	21.3	14.2	-1.4	-13.7	-26.5

Ход выполнения

1) Выбираю функцию вида

$$y = a_0 + a_1 x$$

2) Нахождение точечной оценки неизвестных параметров функции регрессии

2.1) Метод средних:

а) у нас 2 неизвестных параметра

б) делим таблицу данных на 2 части

x(i)	7.0	14.0	22.0
y(i)	28.2	21.3	14.2

x(i)	30.0	39.0	48.0
y(i)	-1.4	-13.7	-26.5

в) Составляем системы уравнений для каждой части

$$\begin{cases} 28.2 = a_0 + a_1 * 7 \\ 21.3 = a_0 + a_1 * 14 \\ 14.2 = a_0 + a_1 * 22 \end{cases} \qquad \begin{cases} -1.4 = a_0 + a_1 * 30 \\ -13.7 = a_0 + a_1 * 39 \\ -26.5 = a_0 + a_1 * 48 \end{cases}$$

$$63.7 = 3a_0 + 43a_1 \qquad -41.6 = 3a_0 + 117a_1$$

Составим и решим систему

$$\begin{cases} 3a_0 + 43a_1 = 63.7 \\ 3a_0 + 117a_1 = -41.6 \end{cases}$$

$$a_0 = 41{,}62928;\ a_1 = 1{,}42297$$

2.2) Метод наименьших квадратов:

$S = \sum_{i=1}^{n}(\delta_i)^2$ – критерий близости оценки функции регрессии к экспериментальным данным

Для линейной регрессии:

$$S = \sum_{i=1}^{n}\left(y_i - \tilde{a}_0\varphi_0(x) - \tilde{a}_1\varphi_1(x) - \dots - \tilde{a}_k\varphi_k(x)\right)$$

Рассмотрим

$$y(x) = a_0(x) + a_1(x) \qquad \varphi_0(x) = 1, \varphi_1(x) = x$$

$$S = \sum_{i=1}^{6}(y_i - a_0 - a_1x_i)^2$$

$$\frac{\partial S}{\partial a_0} = 2\sum_{i=1}^{6}(y_i - a_0 - a_1x_i) * (-1) = 0$$

$$\frac{\partial S}{\partial a_1} = 2\sum_{i=1}^{6}(y_i - a_0 - a_1x_i) * (-x_i) = 0$$

$$\begin{cases} \sum_{i=1}^{6} -1 * y_i + \sum_{i=1}^{6} a_0 + \sum_{i=1}^{6} a_1 * x_i = 0 \\ \sum_{i=1}^{6} -x_iy_i + \sum_{i=1}^{6} a_0x_i + \sum_{i=1}^{6} a_1 * x_i^2 = 0 \end{cases} =>$$

$$\begin{cases} 6a_0 + a_1\sum_{i=1}^{6} x_i = \sum_{i=1}^{6} y_i \\ a_0 \sum_{i=1}^{6} x_i + a_1\sum_{i=1}^{6} x_i^2 = \sum_{i=1}^{6} x_iy_i \end{cases} =>$$

$$\begin{cases} 6a_0 + a_1\sum_{i=1}^{6} x_i = \sum_{i=1}^{6} y_i \\ a_0 \sum_{i=1}^{6} x_i + a_1\sum_{i=1}^{6} x_i^2 = \sum_{i=1}^{6} x_iy_i \end{cases} =>$$

$$\begin{cases} 6a_0 + 160a_1 = 22.1 \\ 160a_0 + 5454a_1 = -1040,3 \end{cases}$$

$$a_0 = 40,28375; a_1 = -1,37252$$

Значение критерия близости данной функции и данных эксперимента:

$$S_{min}\sum_{i=1}^{6}(y_i - a_0 - a_1x_i)^2 = 24.37142$$

Рассмотрим

$$y_2(x) = a_0(x) + a_1(x) + a_2(x) \qquad \varphi_0(x) = 1, \varphi_1(x) = x, \varphi_2(x) = x^2$$

$$S = \sum_{i=1}^{6} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)^2$$

$$\frac{\partial S}{\partial a_0} = 2\sum_{i=1}^{6} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) * (-1) = 0$$

$$\frac{\partial S}{\partial a_1} = 2\sum_{i=1}^{6} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) * (-x_i) = 0$$

$$\frac{\partial S}{\partial a_2} = 2\sum_{i=1}^{6} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right) * \left(-x_i^2\right) = 0$$

$$\begin{cases} \sum_{i=1}^{6} -1 * y_i + \sum_{i=1}^{6} a_0 + \sum_{i=1}^{6} a_1 * x_i + \sum_{i=1}^{6} a_2 * x_i^2 = 0 \\ \sum_{i=1}^{6} -x_i y_i + \sum_{i=1}^{6} a_0 x_i + \sum_{i=1}^{6} a_1 * x_i^2 + \sum_{i=1}^{6} a_2 * x_i^3 = 0 \\ \sum_{i=1}^{6} -x_i^2 y_i + \sum_{i=1}^{6} a_0 x_i^2 + \sum_{i=1}^{6} a_1 * x_i^3 + \sum_{i=1}^{6} a_2 * x_i^4 = 0 \end{cases} =>$$

$$\begin{cases} 6a_0 + a_1 \sum_{i=1}^{6} x_i + a_2 \sum_{i=1}^{6} x_i^2 = \sum_{i=1}^{6} y_i \\ a_0 \sum_{i=1}^{6} x_i + a_1 \sum_{i=1}^{6} x_i^2 + a_2 \sum_{i=1}^{6} x_i^3 = \sum_{i=1}^{6} x_i y_i \\ a_0 \sum_{i=1}^{6} x_i^2 + a_1 \sum_{i=1}^{6} x_i^3 + a_2 \sum_{i=1}^{6} x_i^4 = \sum_{i=1}^{6} x_i^2 y_i \end{cases} =>$$

$$\begin{cases} 6a_0 + 160a_1 + 5454a_2 = 22.1 \\ 160a_0 + 5454a_1 + 210646a_2 = -1040{,}3 \\ 5454a_0 + 210646a_1 + 8706930a_2 = -70724.3 \end{cases}$$

$$a_0 = 35.93602; a_1 = -0.94267; a_2 = -0.00783$$

Значение критерия близости данной функции и данных эксперимента:

$$S_{min} \sum_{i=1}^{6} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)^2 = 14.06373$$

3) Проверка гипотезы об адекватности выбранной модели данным эксперимента

Гипотеза об адекватности модели эксп. данных (критерий Фишера)

Формулировки основной и альтернативной гипотез:

H_0: модель (1) считается адекватной: $a_{m+1} = a_{m+2} = \dots = a_k = 0$

H_1: модель (1) не адекватна: $\exists a_i$: $i > m : a_i \neq 0$

Уровень значимости:

$$\alpha = 0.05$$

Статистический критерий:

$$F = \frac{\frac{1}{k-m}\left(S_{min}^{(1)} - S_{min}^{(2)}\right)}{\frac{1}{n-k-1} S_{min}^{(2)}}$$

$$F = \frac{\frac{1}{3-2}(24{,}37142 - 14{,}06373)}{\frac{1}{6-3-1} 14{,}06373} = 1{,}46585$$

Если гипотеза верна, то F имеет распределения Фишера с числами $k_1 = 1$ и $k_2 = 2$. Находим критическую точку

$$F_{\text{кр}} = 18.51$$

Видим, что значение критерия попадает в допустимую область ($F < F_{\text{кр}}$), значит, гипотеза принимается на уровне значимости 0.05.

4) Построение доверительных интервалов для коэффициентов функции регрессии и самой функции