Галанов Ю.И.

Лабораторный практикум по мат. статистике

Нелинейная регрессия

Исследуется зависимость расходов на питание Y от величины совокупного дохода X.

1. Рассмотреть возможность аппроксимировать функцию регресии зависимостью

$$y = a \bullet x^b$$

2. Найти оценки параметров модели (a и b) методом наименьших квадратов и построить для них доверительные интервалы при значении коэффициента доверия 0.95;

3. Найти коэффициент детерминации (множественной корреляции) R^2

Р Е Ш Е Н И Е

Исходные данные представим в виде матрицы *M*.
Где 1-й столбец есть величина *x*, а 2-й столбец - величина *y*.
Отсортируем данные в порядке возрастания величины *x* и поместим их в матрицу *M'*

$$M := \begin{pmatrix} 646.8 & 120.6 \\ 673.5 & 122.5 \\ 701.3 & 127.4 \\ 722.5 & 129.9 \\ 751.6 & 133.4 \\ 779.2 & 134 \\ 810.2 & 136.4 \\ 865.3 & 133.4 \\ 858.4 & 130.1 \\ 875.8 & 136.3 \end{pmatrix}$$

$$i := 0..9$$

$$M' := csort(M, 0)$$

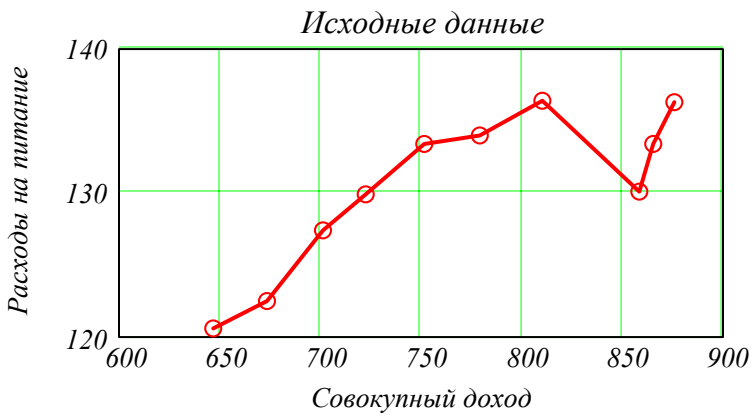
$M' =$

	0	1
0	646.8	120.6
1	673.5	122.5
2	701.3	127.4
3	722.5	129.9
4	751.6	133.4
5	779.2	134
6	810.2	136.4
7	858.4	130.1
8	865.3	133.4
9	875.8	136.3

Выделяем из матрицы переменные $x := M^{\langle 0 \rangle}$ $y := M^{\langle 1 \rangle}$

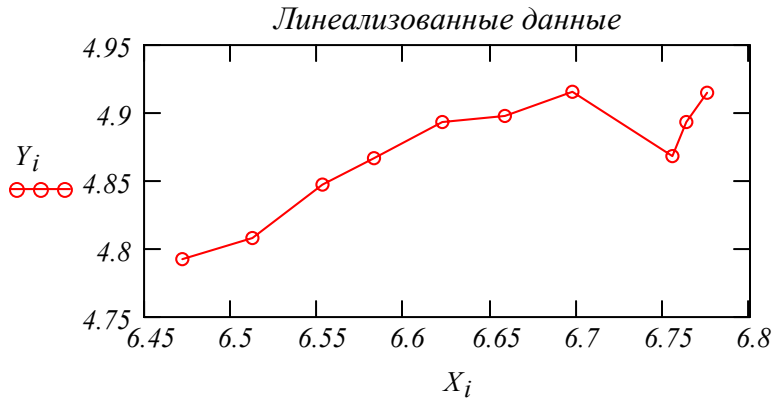
Число пар данных N: $N := length(x)$ $N = 10$

-1-



Перейдем к новым переменным $X := ln(x)\, Y := ln(y)$, тогда функция регрессии станет линейной: $Y := m \cdot X + c$, где $m := b$, $c := ln(a)$

Преобразуем исходные данные: $Y := \overrightarrow{ln(y)}$ $X := \overrightarrow{ln(x)}$



Расчет параметров линейной модели $f(x) := m \cdot x + c$
методом наименьших квадратов

Средние значения:

$$My := \frac{1}{N} \cdot \sum_{i=0}^{N-1} Y_i \qquad My = 4.87 \qquad\qquad Mx := \frac{1}{N} \cdot \sum_{i=0}^{N-1} X_i \qquad Mx = 6.639$$

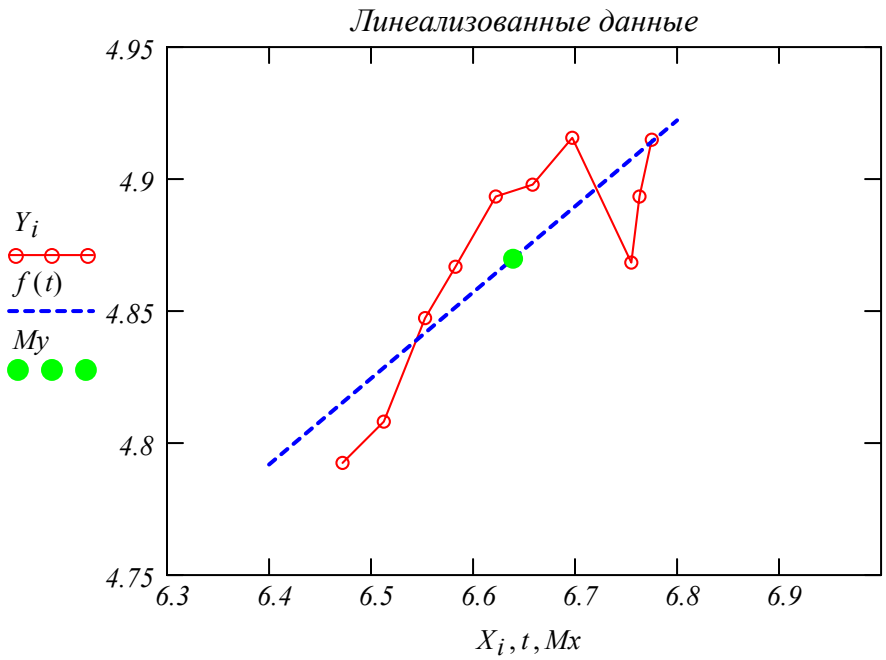
Сводка формул для нахождения параметров прямой линии методом наименьших квадратов

$$D := \sum_i \left(X_i - Mx\right)^2 \qquad D = 0.106 \qquad m := \frac{\sum_i \left(X_i - Mx\right) \cdot Y_i}{D} \qquad m = 0.326$$

$$c := My - m \cdot Mx \qquad\qquad c = 2.706$$

Предсказанные значения: $f(x) := m \cdot x + c$

$t := 6.4, 6.6 .. 6.8$



Найдем оценки дисперсий параметров линейной модели

Дисперсия углового коэффициента $Dm := \frac{1}{D} \cdot \frac{\sum_i \left(Y_i - f\left(X_i\right)\right)^2}{N-2} \qquad Dm = 5.99 \times 10^{-3}$

Дисперсия свободного члена: $Dc := \left(\frac{1}{N} + \frac{Mx^2}{D}\right) \cdot \frac{\sum_i \left(Y_i - f\left(X_i\right)\right)^2}{N-2} \qquad Dc = 0.264$

Доверительные интервалы для коэффициентов модели.
Считаем, что коэффициенты модели имеют асимптотически нормальное распределение, а их центральные статистики - распределение Стьюдента. Тогда доверительные интервалы находим из неравенств:

$$\frac{|C-c|}{\sqrt{Dc}}\cdot\sqrt{N-1}\le q_t \qquad \frac{|M-m|}{\sqrt{Dm}}\cdot\sqrt{N-1}\le q_t,$$

где $q_t := qt\left(\frac{0.95+1}{2}, N-1\right)$ - квантиль распределения Стьюдента.

$q_t = 2.262$

Доверительные границы:

$cL := c - q_t\cdot\frac{\sqrt{Dc}}{\sqrt{N-1}}$ $\quad cL = 2.319$ $\quad cR := c + q_t\cdot\frac{\sqrt{Dc}}{\sqrt{N-1}}$ $\quad cR = 3.094$

$mL := m - q_t\cdot\sqrt{\frac{Dm}{N-1}}$ $\quad mL = 0.268$ $\quad mR := m + q_t\cdot\sqrt{\frac{Dm}{N-1}}$ $\quad mR = 0.384$

Доверительные границы для параметров исходной модели получим с помощью обратных преобразований: $a := e^c$, $\quad b := m$

Границы для a:

$aL := e^{cL}$ $\quad aR := e^{cR}$ $\quad aL = 10.162$ $\quad aR = 22.057$

Для b границы будут такими же, что и для m.

Коэффициент множественной корреляции ${R_{corr}}^2$ выразим через коэффициент корреляции R_{corr} между исходными данными и предсказанными моделью

$Y1_i := f(X_i)$ $\quad M1 := \frac{1}{N}\cdot\sum_i Y1_i$ $\quad M1 = 4.87$ $\quad DY := \frac{1}{N}\cdot\sum_i (Y_i - My)^2$

$D1 := \frac{1}{N}\cdot\sum_i (Y1_i - M1)^2$ $\quad D1 = 1.125\times 10^{-3}$ $\quad DY = 1.632\times 10^{-3}$

$$R_{corr} := \frac{\sum_i (Y1_i - M1)\cdot(Y_i - My)}{N}\cdot\sqrt{D1\cdot DY} \qquad {R_{corr}}^2 = 2.323\times 10^{-12}$$

Проверим гипотезу о значимости коэффициента детерминации. Подсчитаем статистику критерия при **N=10,** $k := 2$; здесь k - число коэффициентов модели:

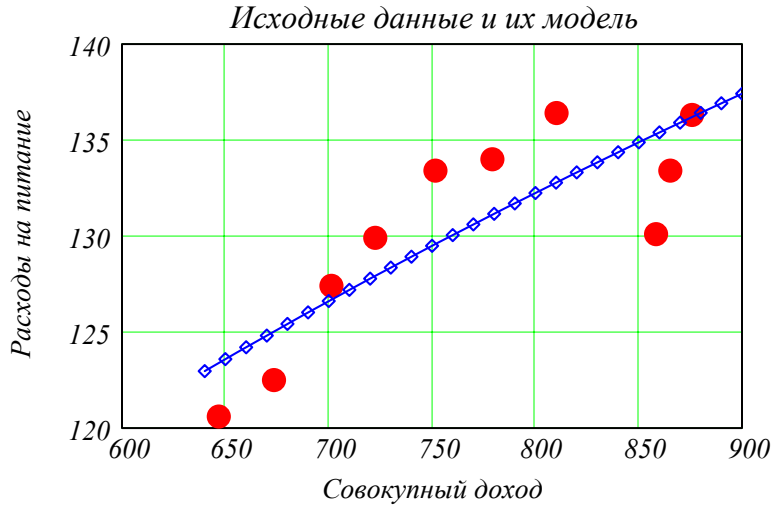
$$F := \frac{R_{corr}^{2} \cdot (N-k)}{\left(1 - R_{corr}^{2}\right) \cdot (k-1)} \qquad F = 1.858 \times 10^{-11}$$

Критическую точку F_T находим как квантиль распределения Фишера со степенями свободы $n1 := N - k$ и $n2 := k - 1$ при уровне значимости $\alpha := 0.05$

$$F_T := qF(1-\alpha, n2, n1) \qquad F_T = 5.318$$

Так как $F > F_T$, то коэффициент $R_{corr} = 1.524 \times 10^{-6}$ значим

$$t := 640, 650 .. 900$$



$a = 14.971$

$b = 0.326$