МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова» (БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова»)

Факультет	О	Естественнонаучный
	шифр	наименование
Кафедра	О6	Высшая математика
	шифр	наименование
Дисциплина	Математическая статистика и случайные процессы	

ЛАБОРАТОРНАЯ РАБОТА №9

на тему «Регрессионный анализ в пакетах STATGRAPHICS и MATHCAD»

Вариант №4

Выполнил студент группы И967
Васильев Н.А.
Фамилия И.О.

ПРЕПОДАВАТЕЛЬ
Мартынова Т.Е.
Фамилия И.О. Подпись

«_____» ____________________ 2019 г.

САНКТ-ПЕТЕРБУРГ
2019 г.

КРАТКИЕ СВЕДЕНИЯ ИЗ ТЕОРИИ

Оценка параметров линейной регрессии методом наименьших квадратов

Перепишем уравнение регрессии в несколько ином виде

$$y = \alpha + \beta\left(x - \bar{x}\right), \tag{7.2.1}$$

где $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Эта прямая называется теоретической линией регрессии или прямой отклика. Уравнение

$$\hat{y} = a + b\left(x - \bar{x}\right) \tag{7.2.2}$$

определяет кривую, которая является оценкой для прямой регрессии.

Суть метода наименьших квадратов состоит в выборе таких оценок a и b, которые бы минимизировали сумму квадратов отклонений наблюденных значений y_i от прогнозируемых величин $\hat{y}_i$, полученных подстановкой значений x_i в уравнение (7.2.2), т.е. $R = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\left[y_i - a - b\left(x_i - \bar{x}\right)\right]^2 \Rightarrow \min$. Чтобы найти значения a и b, минимизирующие R, продифференцируем это уравнение по a и b и приравняем производные нулю:

$$\begin{cases} \dfrac{\partial R}{\partial a} = -2\sum_{i=1}^{n}\left[y_i - a - b\left(x_i - \bar{x}\right)\right] = 0, \\ \dfrac{\partial R}{\partial b} = -2\sum_{i=1}^{n}\left[y_i - a - b\left(x_i - \bar{x}\right)\right]\left(x_i - \bar{x}\right) = 0. \end{cases}$$

Раскроем здесь члены под знаком суммы: $\sum_{i=1}^{n} y_i - an - b\sum_{i=1}^{n}\left(x_i - \bar{x}\right) = 0$, $\sum_{i=1}^{n} y_i\left(x_i - \bar{x}\right) - a\sum_{i=1}^{n}\left(x_i - \bar{x}\right) - b\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = 0$. Но $\sum_{i=1}^{n}\left(x_i - \bar{x}\right) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\bar{x} =$ $= n\bar{x} - n\bar{x} = 0$. Тогда $na = \sum_{i=1}^{n} y_i$, $b\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)y_i$. Отсюда легко получить оценки параметров a и b:

$$a = \hat{\alpha} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y}, \quad b = \hat{\beta} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)y_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}. \tag{7.2.3}$$

Вторую оценку часто видоизменяют и переписывают в следующем виде $\sum_{i=1}^{n}\left(x_i - \bar{x}\right)y_i = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)y_i + \bar{y}\sum_{i=1}^{n}\left(x_i - \bar{x}\right) = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)y_i + \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\bar{y} =$ $= \sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)$. Тогда

$$b = \frac{\sum_{i=1}^{n}\left[\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)\right]}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}. \tag{7.2.4}$$

Рассмотрим теперь свойства полученных оценок. Они являются несмещенными, состоятельными и эффективными в классе линейных (относительно наблюдений) оценок. Действительно,

$$M(a) = M\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(y_i) = \frac{1}{n}\sum_{i=1}^{n} M\left[\alpha + \beta\left(x_i - \bar{x}\right) + \varepsilon_i\right] =$$

$$= \frac{1}{n}\left[\alpha n + \beta\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\right] = \alpha,$$

$$M(b) = M\left(\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)y_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right) = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)M(y_i)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{\sum_{i=1}^{n}\left[\left(x_i - \bar{x}\right)\left(\alpha + \beta\left(x_i - \bar{x}\right)\right)\right]}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} =$$

$= \alpha\dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} + \beta\dfrac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \beta$. Здесь учтено, что переменные x_i - неслучайные, а y_i - случайные величины. Кроме того, математическое ожидание y_i есть теоретическая линия регрессии (7.2.1).

Найдем теперь дисперсии оценок a и b в предположении, что наблюдения y_i независимы и нормально распределены, причем $D(y_i) = D = \sigma^2$ (предположения 3, 4 и 5 предыдущего подраздела). Имеем: $D(a) = D\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(y_i) = \frac{D}{n^2}n = \frac{D}{n}$,

$$D(b) = D\left(\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)y_i}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right) = \frac{1}{\left[\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]^2}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 D(y_i) = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 D}{\left[\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2\right]^2} =$$

$$= \frac{D}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}.$$

Состоятельность оценок a и b немедленно следует после применения к ним неравенства Чебышева. Например, для оценки a получим

$P\left(\left|a - \alpha\right| \ge \varepsilon\right) \le \dfrac{D(a)}{\varepsilon^2} = \dfrac{D}{n\varepsilon^2}$. Отсюда $\lim\limits_{n\to\infty} P\left(\left|a - \alpha\right| \ge \varepsilon\right) = 0$.

В общем случае доказательство того, что метод наименьших квадратов дает оценки с наименьшей дисперсией в классе всех несмещенных оценок, довольно сложно. Приведем его для оценки b параметра β. Предположим, что существует еще одна линейная оценка b' параметра β, отличная от оценки b и пусть, например, $b' = \sum_{i=1}^{n} c_i y_i$. Очевидно, что $M(b') = \sum_{i=1}^{n} c_i M(y_i) = \sum_{i=1}^{n} c_i\left[\alpha + \beta\left(x_i - \bar{x}\right)\right] = \alpha\sum_{i=1}^{n} c_i + \beta\sum_{i=1}^{n}\left(x_i - \bar{x}\right)c_i$. Оценка b' будет несмещенной, если $M(b') = \beta$, т.е.

$$\begin{cases} \sum_{i=1}^{n} c_i = 0, \\ \sum_{i=1}^{n}\left(x_i - \bar{x}\right)c_i = 1. \end{cases} \tag{7.2.5}$$

В этих условиях $D(b')=\sum_{i=1}^{n}c_i^2 D(y_i)=D\sum_{i=1}^{n}c_i^2=$

$$=D\sum_{i=1}^{n}\left[c_i-\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}+\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^2=D\left[\sum_{i=1}^{n}c_i^2+\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2}-\right.$$

$$\left.-2\sum_{i=1}^{n}c_i\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}+\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2}+2\sum_{i=1}^{n}c_i\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}-2\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2}\right]=$$

$$=\sum_{i=1}^{n}\left[c_i-\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^2+\sum_{i=1}^{n}\left[c_i-\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}+\sum_{i=1}^{n}c_i\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

Но $\sum_{i=1}^{n}(x_i-\bar{x})^2$ - это константа, т.е. выражение под этой суммой уже не зависит от индекса внешнего суммирования. Тогда $\sum_{i=1}^{n}c_i\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sum_{i=1}^{n}c_i(x_i-\bar{x})=\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ с учетом условий (7.2.5). Аналогично

$$\sum_{i=1}^{n}\left[c_i-\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{1}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2}\sum_{i=1}^{n}\left[c_i(x_i-\bar{x})\sum_{i=1}^{n}(x_i-\bar{x})^2-(x_i-\bar{x})^2\right]=$$

$$=\frac{1}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}c_i(x_i-\bar{x})-\sum_{i=1}^{n}(x_i-\bar{x})^2\right]=$$

$$=\frac{1}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^2}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2-\sum_{i=1}^{n}(x_i-\bar{x})^2\right]=0.$$

Поэтому

$$D(b')=D\sum_{i=1}^{n}\left[c_i-\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^2+\frac{D}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

Последний член в полученном выражении является константой. Следовательно, минимизировать $D(b')$ можно только за счет уменьшения первого члена. Полагая $c_i=\frac{x_i-\bar{x}}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$, мы обратим первый член в нуль (меньше он не может быть) и тем самым минимизируем $D(b')$. Но если в формулу $b'=\sum_{i=1}^{n}c_i y_i$

подставить значения c_i, при которых $D(b')$ минимальна, то альтернативная оценка b' примет вид $b' = \sum_{i=1}^{n} c_i y_i = \dfrac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$, что совпадает с оценкой наименьших квадратов. Поэтому b - линейная несмещенная оценка параметра β с минимальной дисперсией.

Интервальные оценки параметров линейной регрессии и кривой регрессии

Построим теперь доверительные границы для параметров α и β и кривой регрессии. Так как $\hat{y} = a + b(x - \overline{x})$ и $D(a) = \dfrac{D}{n}$, $D(b) = \dfrac{D}{\sum_{i=1}^{n}(x_i - \overline{x})}$, то

$$M(\hat{y}) = M[a + b(x - \overline{x})] = M(a) + (x - \overline{x})M(b) = \alpha + \beta(x - \overline{x}) = y,$$

$$D(\hat{y}) = D[a + b(x - \overline{x})] = D(a) + (x - \overline{x})^2 D(b) = \frac{D}{n} + \frac{(x - \overline{x})^2 D}{\sum_{i=1}^{n}(x_i - \overline{x})^2} =$$

$$= D\left[\frac{1}{n} + \frac{(x - \overline{x})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right]$$ - выражение для дисперсии $D(\hat{y})$ в текущей точке x.

Очевидно, что $\hat{y}$ - кроме того линейная функция от оценок a и b, которые в свою очередь являются линейными оценками от нормально распределенных наблюдений y_i. Следовательно, $\hat{y}$ - нормально распределенная случайная величина, и для нее может быть построен доверительный интервал стандартным образом. То же можно сказать и об оценках коэффициентов регрессии.

Заметим, что a и b независимы друг от друга, так же как независима от них оценка $\hat{D}$ дисперсии D. Это можно доказать, рассмотрев, например, $M(a \cdot b)$. После непродолжительных вычислений будет видно, что $M(a \cdot b) = K(a, b) = 0$. Следовательно a и b - некоррелированы, а поскольку мы остаемся в рамках гауссовской модели, то и независимы.

В предыдущих разделах было показано, что дробь $n\hat{D}/D \in \chi^2_{n-1}$, $\hat{D} = D^*$. В нашем случае $\hat{D} = D^* = \dfrac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 =$
$= \dfrac{1}{n}\sum_{i=1}^{n}(y_i - a - b(x - \overline{x}))^2$. Так как на случайные величины y_i, входящие в эту формулу, наложены два условия связи вида $\dfrac{\partial R}{\partial a} = 0$ и $\dfrac{\partial R}{\partial b} = 0$, то число степеней свободы уменьшается на число связей и $n\hat{D}/D \in \chi^2_{n-2}$.

Составим дроби Стьюдента для a и b. В нашем случае $a \in N\left(\alpha, \dfrac{D}{n}\right)$, $b \in N\left(\beta, \dfrac{D}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right)$, а по теории $t = \dfrac{z\sqrt{n}}{\sqrt{v}}$, где $z \in N(0,1)$, $v \in \chi^2_n$, причем в этой дроби под корнем в числителе стоит число степеней свободы случайной величины v. Выберем в качестве стандартной нормальной случайной величины z сначала выражение $\dfrac{a - \alpha}{\sqrt{D/n}} = \dfrac{(a - \alpha)\sqrt{n}}{\sigma} \in N(0,1)$, затем $\dfrac{b - \beta}{\sigma}\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2} \in N(0,1)$. Подставляя эти результаты в дробь Стьюдента, будем иметь $t_a = \dfrac{(a - \alpha)\sqrt{n}\sqrt{n-2}}{\sqrt{D}\,(\sqrt{n\hat{D}}/\sqrt{D})} = \dfrac{(a - \alpha)\sqrt{n-2}}{\sqrt{\hat{D}}} \in t_{n-2}$. Аналогично

$$t_b=\frac{(b-\beta)\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\sqrt{n-2}}{\sqrt{D}\sqrt{n\hat{D}/D}}=\frac{(b-\beta)\sqrt{(n-2)\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}}{\sqrt{n\hat{D}}}\in t_{n-2}.$$

Наконец, получим в явном виде доверительные интервалы для коэффициентов линейной регрессии. $P(|a-\alpha|<\varepsilon)=\beta'$ по определению, где β' - доверительная вероятность. $P\left(\frac{|a-\alpha|\sqrt{n-2}}{\sqrt{\hat{D}}}<\frac{\varepsilon\sqrt{n-2}}{\sqrt{\hat{D}}}\right)=P\left(\left|\frac{(a-\alpha)\sqrt{n-2}}{\sqrt{\hat{D}}}\right|<t_{\beta'}\right)=$

$=P(|t|<t_{\beta'})=\beta'$, величина $t_{\beta'}$ может быть найдена из уравнения $2\int_0^{t_{\beta'}}s_{n-2}(t)dt=\beta'$. Тогда $\varepsilon=t_{\beta',n-2}\sqrt{\frac{\hat{D}}{n-2}}$ и $I_\alpha=\left(a-t_{\beta',n-2}\sqrt{\frac{\hat{D}}{n-2}},\ a+t_{\beta',n-2}\sqrt{\frac{\hat{D}}{n-2}}\right)$.

Точно такие же преобразования дают интервал для второго коэффициента.

$P(|b-\beta|<\varepsilon)=\beta'$, тогда $P\left(\left|\frac{(b-\beta)\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}}{\sqrt{n\hat{D}/(n-2)}}\right|<\frac{\varepsilon\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}}{\sqrt{n\hat{D}/(n-2)}}\right)=$

$=P(|t|<t_{\beta'})=\beta'$. Отсюда $\varepsilon=t_{\beta',n-2}\sqrt{\frac{n}{n-2}\frac{\hat{D}}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}}$ и

$$I_\beta=\left(\varepsilon-t_{\beta',n-2}\sqrt{\frac{n}{n-2}\frac{\hat{D}}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}},\ \varepsilon+t_{\beta',n-2}\sqrt{\frac{n}{n-2}\frac{\hat{D}}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}}\right).$$

На практике часто возникает вопрос об оценке отклонения истинной прямой $y=\alpha+\beta\left(x-\bar{x}\right)$ от ее оценки $\hat{y}=a+b\left(x-\bar{x}\right)$ при некотором заданном значении x. Особенно важен этот вопрос при построении прогноза. Оценкой точности здесь также может служить интервальная оценка y.

Используя обычные рассуждения, приводящие к t - статистикам, получаем:

$$M(\hat{y})=\alpha+\beta\left(x-\bar{x}\right)=y,\ D(\hat{y})=D\left[\frac{1}{n}+\frac{\left(x-\bar{x}\right)^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\right],\quad \hat{y}\in N(M(\hat{y}),D(\hat{y})).$$

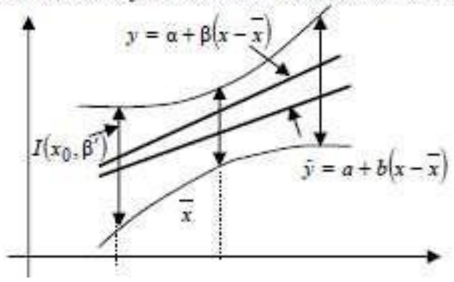
Тогда $z=\frac{\hat{y}-y}{\sqrt{D(\hat{y})}}\in N(0{,}1)$, а $t=\frac{z\sqrt{n-2}}{\sqrt{n\hat{D}/D}}\in t_{n-2}$. В нашем случае дробь

Стьюдента равна

$$t = \frac{(\hat{y} - y)}{\sqrt{D\left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\right)}\sqrt{\frac{n\hat{D}}{D}}}\sqrt{n-2} = \frac{(\hat{y} - y)\sqrt{n-2}}{\sqrt{\left(1 + \frac{n(x - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\right)\hat{D}}} = (\hat{y} - y)d \in t_{n-2}.$$

$$P\left(\left|\hat{y} - y\right| < \varepsilon\right) = \beta' \quad \text{и} \quad P\left(\left|(\hat{y} - y)d\right| < \varepsilon d\right) = P\left(\left|(\hat{y} - y)d\right| < t_{\beta'}\right) = P\left(|t| < t_{\beta'}\right) = \beta'.$$

Тогда $\varepsilon = \frac{t_{\beta'}}{d} = t_{\beta', n-2}\sqrt{\frac{\hat{D}}{n-2}}\sqrt{\left(1 + \frac{n(x - \bar{x})^2}{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\right)}$, $I_y = (\hat{y} - \varepsilon,\ \hat{y} + \varepsilon)$ для любого конкретного x, так как $\varepsilon = \varepsilon(x)$. Очевидно, что длина доверительного интервала минимальна в точке $x = \bar{x}$. По мере удаления от $\bar{x}$ точность оценки будет заметно снижаться. Наименее надежная оценка по МНК будет получаться для ординат, отвечающим очкам, наиболее удаленным от $\bar{x}$ (рис. 7.1). Вертикальные отрезки на рисунке представляют собой доверительные интервалы в соответствующих точках.



ФОРМУЛИРОВКА ЗАДАНИЯ

Найти в пакетах STATGRAPHICS и MATCHAD оценки параметров линейной регрессии y на x, доверительные интервалы для параметров и линии регрессии и проверить согласие линейной регрессии с результатами наблюдений. Принять уровень доверительной вероятности равным 0.90.

$$x := \begin{pmatrix} 0 \\ 0.05 \\ 0.1 \\ 0.15 \\ 0.2 \\ 0.25 \\ 0.3 \\ 0.35 \\ 0.4 \\ 0.45 \\ 0.5 \\ 0.55 \end{pmatrix} \qquad y := \begin{pmatrix} 2.43 \\ 2.67 \\ 2.71 \\ 3.15 \\ 3.47 \\ 3.76 \\ 3.91 \\ 4.46 \\ 4.76 \\ 5.15 \\ 5.54 \\ 5.61 \end{pmatrix}$$

СКРИНШОТЫ

Решение в пакете STATGRAPHICS

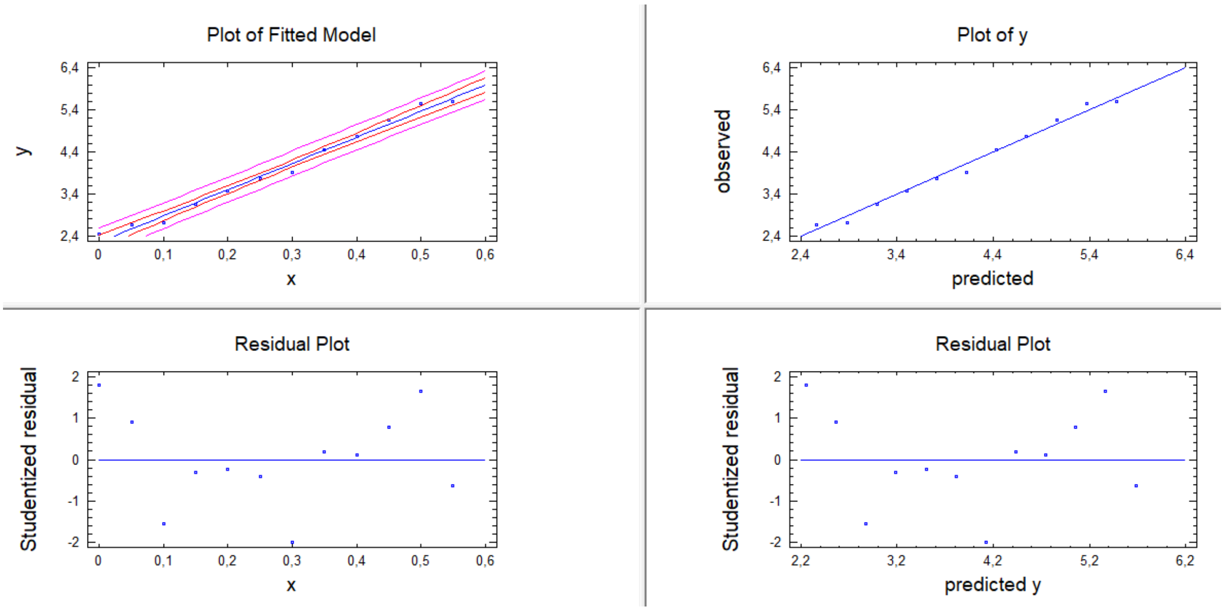


Рисунок 1 – Графики линейной регрессии и предсказанных наблюдений; стьюдентизированных остатков модели линейной регрессии

```
Regression Analysis - Linear model: Y = a + b*X
-----------------------------------------------------------------------------
Dependent variable: y
Independent variable: x
-----------------------------------------------------------------------------
                               Standard          T
Parameter       Estimate         Error       Statistic        P-Value
-----------------------------------------------------------------------------
Intercept        2,25564      0,0694954       32,4574         0,0000
Slope            6,22797       0,214043       29,0968         0,0000
-----------------------------------------------------------------------------


                           Analysis of Variance
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------------
Model                     13,8666      1      13,8666     846,62       0,0000
Residual                 0,163787     10    0,0163787
-----------------------------------------------------------------------------
Total (Corr.)             14,0304     11

Correlation Coefficient = 0,994146
R-squared = 98,8326 percent
R-squared (adjusted for d.f.) = 98,7159 percent
Standard Error of Est. = 0,127979
Mean absolute error = 0,0962005
Durbin-Watson statistic = 1,52372 (P=0,1043)
Lag 1 residual autocorrelation = 0,129935
```

Рисунок 2 – Результаты расчета модели простой линейной регрессии

```
Analysis of Variance with Lack-of-Fit
-----------------------------------------------------------------------------
Source             Sum of Squares     Df  Mean Square    F-Ratio      P-Value
-----------------------------------------------------------------------------
Model                     13,8666      1      13,8666     846,62       0,0000
Residual                 0,163787     10    0,0163787
-----------------------------------------------------------------------------
   Lack-of-Fit           0,163787     10    0,0163787
   Pure Error                 0,0      0
-----------------------------------------------------------------------------
Total (Corr.)             14,0304     11
```

Рисунок 3 – Результаты анализа адекватности линейной модели

```
Predicted Values
----------------------------------------------------------------------------
                                   95,00%                  95,00%
                  Predicted     Prediction Limits      Confidence Limits
           X              Y     Lower        Upper      Lower        Upper
----------------------------------------------------------------------------
         0,0        2,25564   1,93115      2,58013     2,1008      2,41049
        0,55        5,68103   5,35654      6,00551    5,52618      5,83587
----------------------------------------------------------------------------
```

Рисунок 4 – Предсказанные по умолчанию наблюдения и их доверительные границы

Comparison of Alternative Models

Model	Correlation	R-Squared
Square root-Y	0,9957	99,15%
Exponential	0,9943	98,86%
Linear	0,9941	98,83%
Reciprocal-Y	-0,9822	96,48%
Square root-X	0,9350	87,41%
Reciprocal-X	<no fit>	
Double reciprocal	<no fit>	
Logarithmic-X	<no fit>	
Multiplicative	<no fit>	
S-curve	<no fit>	
Logistic	<no fit>	
Log probit	<no fit>	

Рисунок 5 – Результаты анализа альтернативных моделей

Unusual Residuals

Row	X	Y	Predicted Y	Residual	Studentized Residual

Рисунок 6 – Результаты анализа резко выделяющихся наблюдений

Решение в пакете MATHCAD

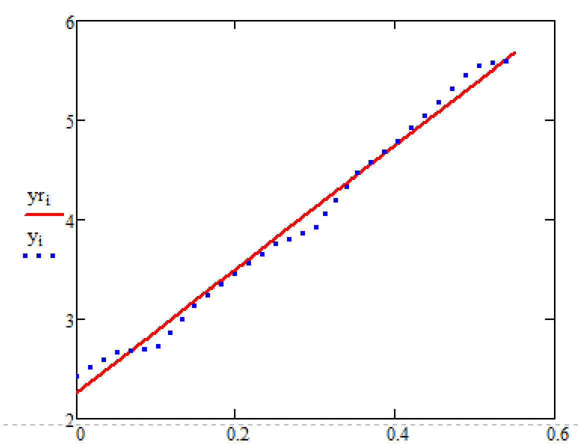
$n := \mathrm{rows}(x) = 12$

$a := \mathrm{intercept}(x, y) = 2.256$

$b := \mathrm{slope}(x, y) = 6.228$

$i := 1..n$

$yr_i := a + b \cdot x_i$



$xmean := \mathrm{mean}(x) = 0.275$

$ymean := \mathrm{mean}(y) = 3.968$

$\beta 1 := 0.9$

$\alpha 1 := 1 - \beta 1 = 0.1$

$t := \mathrm{qt}\left(1 - \frac{\alpha 1}{2}, n - 2\right) = 1.812$

$D1 := \frac{1}{n-2} \cdot \sum_{k=1}^{n} \left(y_k - yr_k\right)^2 = 0.016$

$aleft := a - t \cdot \sqrt{\frac{D1}{n-2}} = 2.182$

$aright := a + t \cdot \sqrt{\frac{D1}{n-2}} = 2.329$

$\varepsilon b := \sqrt{\frac{n}{n-2} \cdot \frac{D1}{\sum_{k=1}^{n} \left(x_k - xmean\right)^2}} = 0.234$

$bleft := b - t \cdot \varepsilon b = 5.803$

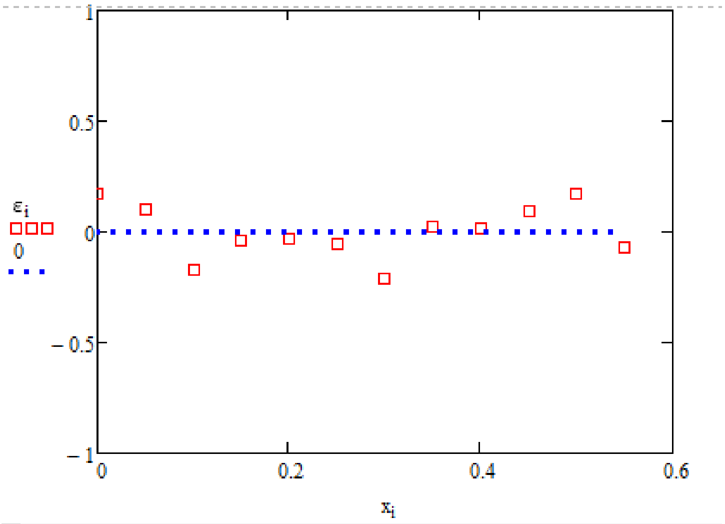
$bright := b + t \cdot \varepsilon b = 6.653$

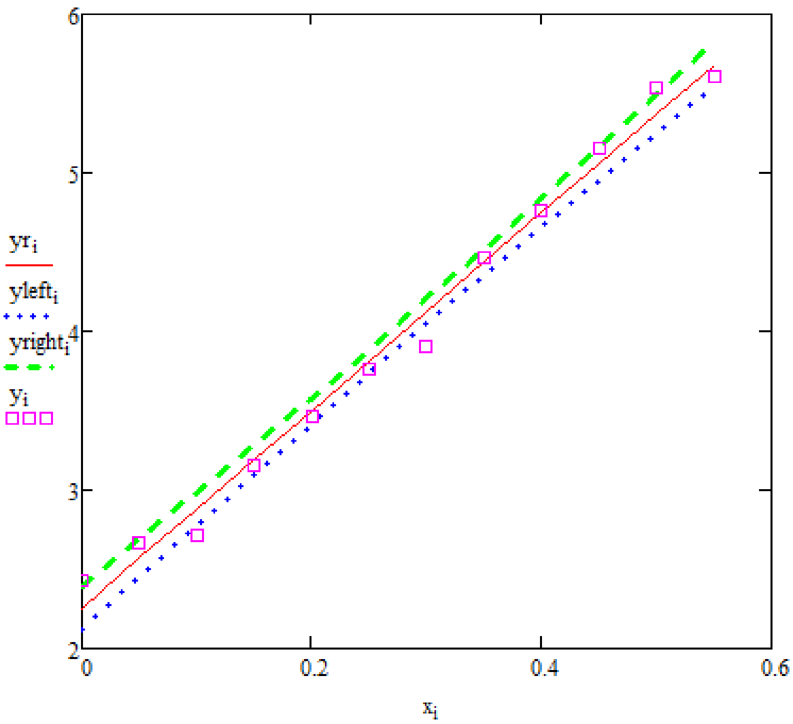
$i := 1..n$

$\varepsilon_i := y_i - yr_i$

$$yleft_i := yr_i - t\cdot\sqrt{\frac{D1}{n-2}\cdot\left[1+\frac{n\cdot(x_i - xmean)^2}{\sum_{k=1}^{n}(x_k - xmean)^2}\right]}$$

$$yright_i := yr_i + t\cdot\sqrt{\frac{D1}{n-2}\cdot\left[1+\frac{n\cdot(x_i - xmean)^2}{\sum_{k=1}^{n}(x_k - xmean)^2}\right]}$$





Вывод: В ходе выполнения данной лабораторной работы была проведена проверка согласия линейной регрессии с результатом наблюдений. По данным пакета STATGRAPHICS после сравнения альтернативных моделей можно сказать, что линейная модель лучше всего согласуется с результатами наблюдений. По данным пакета MATCHAD при доверительной вероятности равной 0.90 в доверительный интервал из 12 наблюдений не попало только 3, то есть 25%. Это значит, что линейная модель удовлетворительно аппроксимирует исходные данные.