МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«Национальный исследовательский ядерный университет «МИФИ»
(НИЯУ МИФИ)
Институт Финансовых Технологий и Экономической Безопасности
Кафедра Финансового мониторинга

Лабораторная работа №1
по дисциплине «Эконометрика»

Выполнил студент группы С20-702: Нуритдинходжаева А.А.
Проверил: Домашова Д.В.

Москва 2023

Содержание

1. Постановка задачи

Провести регрессионный анализ на основе следующих данных, собранных по каждому субъекту Российской Федерации на 2021 год:

1. Результаты голосования по поправкам в Конституцию, % за – результативный признак (y);
2. Объясняющие признаки (x_1, x_2, … x_9):

X_1	Число организаций, на которых проходили забастовки
X_2	Численность населения на одну больничную койку
X_3	Численность инвалидов по группе инвалидности
X_4	Задолженность по кредитам в рублях, предоставленным кредитными организациями физическим лицам
X_5	Средства (вклады) физических лиц на рублевых счетах в Сбербанке России
X_6	Средний размер назначенных пенсий, тыс.
X_7	Среднедушевые доходы населения, тыс.
X_8	Уровень бедности, %
X_9	Средний прожиточный минимум

Для этого:

1. Оценить функцию регрессии
2. Исследовать уравнение регрессии на значимость.
3. Для значимой модели регрессии исследовать значимость коэффициентов.
4. Построить доверительные интервалы для значимых параметров связи.
5. Провести экономический анализ результатов.

Исходные данные приведены в приложении А.

2 Оценки линейного уравнения множественной регрессии

Не требуя нормальный закон распределения вектора $(Y, X_1,..., X_k)$, мы должны оценивать функцию $M_{Y/X_1...X_k} = f(x_1, x_2,..., x_k)$, вообще говоря, не обязательно линейную. Неизвестная функция, $f(x_1, x_2,..., x_k)$ возможно задается сложным аналитическим выражением, но на практике подбирают “хорошую” аппроксимацию для $f(x_1, x_2,..., x_k)$ из некоторого параметрического класса функций $\tilde{f}(x_1, x_2,..., x_k, \bar{\beta})$ ($\bar{\beta}$ - вектор неизвестных параметров), удобных в вычислительном плане при оценке параметров и т.д. Возьмем в качестве аппроксимирующей функции $\tilde{f}(x_1, x_2,..., x_k, \bar{\beta})$ линейную функцию: $M_{Y/X} \approx \tilde{f}(x_1, x_2,..., x_k, \bar{\beta}) = \tilde{Y} = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$

или в векторной форме: $\tilde{y} = x^T \beta^T$

где $x = (1, x_1,..., x_k)^T$, $\bar{\beta} = (\beta_0,..., \beta_k)$

Для оценки линейной функции (уравнения) множественной регрессии построим математическую модель, получим оценки коэффициентов $\bar{\beta}$, изучим свойства оценок уравнения и отдельных коэффициентов.

Классическая линейная модель множественной регрессии.

Предположим, что для оценки уравнения регрессии взята выборка объемом *n*. Результаты наблюдений над результативным признаком представлены вектором $Y = (y_1,..., y_n)^T$ и матрицей *X* типа объект-свойство наблюденных значений признаков $x_1, ...x_к$:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & ... & x_{1k} \\ 1 & x_{21} & x_{22} & ... & x_{2k} \\ . & . & . & ... & . \\ 1 & x_{n1} & x_{n2} & ... & x_{nk} \end{pmatrix}$$

здесь x_{ij} – значение *j*-го признака на *i*-м объекте наблюдения; столбец из "1" можно считать столбцом "наблюденных" значений для признака $x_0^0 = 1$.

Очевидно, что, подставляя в правую часть наблюденные значение объясняющих переменных для объекта наблюдение O_i, а в левую часть наблюденное значение результативного признака мы сможем их уравновесить величиной ε_i - регрессионным остатком, характеризующей расхождение между наблюденным значением y_i и "усредненным" значением $\tilde{y}_i$ - то есть значением линейной функции регрессии:

$$y_i = \beta_0 + \beta_1 x_{i_1} + \beta_2 x_{i_2} + \ldots + \beta_k x_{i_k} + \varepsilon_i \quad i = \overline{1,n}$$

или, если $\bar{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$

$$Y = X\bar{\beta} + \bar{\varepsilon}$$

Система линейных уравнений называется линейной моделью множественной регрессии (ЛММР).

Отметим, что наличие регрессионных остатков в (ЛММР) объясняется вариацией результативного признака относительно "среднего значения" при данных значениях объясняющих переменных, что в свою очередь можно объяснить влиянием на результативный признак ряда неучтенных факторов-признаков.

Будем предполагать, что

1) $x_1, \ldots, x_k$ – детерминированные переменные;
2) ранг матрицы X равен "$k+1$" – среди признаков нет линейно зависимых;
3) $M\varepsilon_i = 0$, $i = \overline{1,n}$ - нет систематических ошибок в измерении y;
4) $D\varepsilon_i = M\varepsilon_i^2 = \sigma^2$, $i = \overline{1,n}$ - гомоскедастичность регрессионных остатков (равноточные измерения);
5) $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = M(\varepsilon_i \cdot \varepsilon_j) = 0$, $i \neq j$, $i = \overline{1,n}$ $j = \overline{1,n}$ - условие некоррелированных регрессионных остатков.

Условия 4 –5 можно заменить одним условием в векторной форме

4') $M\bar{\varepsilon}\bar{\varepsilon}^T = \sigma^2 E_n$

Условия (1–5) известны как условия Гаусса – Маркова. ЛММР, удовлетворяющая требованиям (1–5), называется классической ЛММР (КЛММР).

3. Оценка коэффициентов классической линейной модели множественной регрессии

Рассматривается линейная модель множественной регрессии:

$$Y = X\bar{\beta} + \bar{\varepsilon} .$$

Оценку коэффициентов $\bar{\beta}$ уравнения регрессии можно искать исходя из требований минимума модуля отклонения наблюденных значений y_i от "значений" функции регрессии, либо из критерия минимума суммы квадратов отклонений наблюденных значений y_i от "значений" функции регрессии (метод наименьших квадратов) более удобного с позиции технической реализации.

Перейдем к оценке коэффициентов методом наименьших квадратов (МНК). Выпишем квадратичный функционал, обозначив через $\bar{b} = (b_0, b_1, ..., b_k)^T$ оценку вектора $\bar{\beta}$

$$F = \sum_{i=1}^{n}(y_i - b_0 - b_1x_{i1} - b_2x_{i2} - ... - b_kx_{in})^2 = (Y - X\bar{b})^T(Y - X\bar{b}) =$$
$$= Y^TY - \bar{b}^TX^TY - Y^TX\bar{b} + \bar{b}^TX^TX\bar{b} = Y^TY - 2\bar{b}^TX^TY + \bar{b}^TX^TX\bar{b} \to \min$$

Воспользовавшись необходимым условием существования экстремума, найдем

$$2X^TX\bar{b} - 2X^TY = 0$$

В силу предположения о справедливости условий Гаусса-Маркова, в частности ($\mathrm{rang}X = k+1$), матрица X^TX – не выраждена и получим МНК - оценки для вектора $\bar{\beta}$:

$$\bar{b}_{\text{МНК}} \equiv \bar{b} = (X^TX)^{-1}X^TY$$

и, следовательно, оценку y уравнения регрессии

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + ... + b_kx_k$$

Оценим коэффициенты уравнения регрессии.

Найдем оценки основных характеристик объясняющих переменных. Результаты расчетов в пакете STATISTICA представлены в таблице 1.

Variable	Descriptive Statistics (исходные in Данные) Valid N	Mean	Minimum	Maximum	Std.Dev.
x1	85	0,02	0,00	1,0	0,15
x2	85	122,60	77,50	218,5	20,23
x3	85	0,28	0,00	10,9	1,19
x4	85	131,72	16,27	334,8	51,17
x5	85	79,52	6,46	245,4	44,95
x6	85	16097,96	11923,00	26381,0	2794,11
x7	85	48501,51	30942,17	130482,9	20357,36
x8	85	13,19	4,60	29,3	4,63
x9	85	12441,67	9610,00	24164,0	3236,54

Таблица 1 оценка математических ожиданий объясняющих переменных

```
independedVariables=data.drop(columns=['Y'])
independedVariables.describe()
```

	X1	X2	X3	X4	X5	X6	X7	X8	X9
count	85.000000	85.000000	85.000000	85.000000	85.000000	85.000000	85.000000	85.000000	85.000000
mean	0.023529	122.598824	0.280952	131.723328	79.521953	16097.964706	48501.506863	13.192941	12441.668941
std	0.152477	20.226377	1.186931	51.166005	44.945435	2794.106717	20357.355162	4.625751	3236.535971
min	0.000000	77.500000	0.001200	16.265871	6.455157	11923.000000	30942.166667	4.600000	9610.000000
25%	0.000000	110.400000	0.040855	107.432505	57.877933	14564.000000	36057.550000	10.500000	10635.000000
50%	0.000000	123.100000	0.087969	122.585805	70.509889	15345.000000	41298.475000	12.800000	11206.000000
75%	0.000000	132.500000	0.205361	146.787923	85.745007	16217.000000	50192.558333	15.300000	12810.000000
max	1.000000	218.500000	10.941124	334.799039	245.368812	26381.000000	130482.900000	29.300000	24164.000000

Таблица 1.1 Оценка математических ожиданий объясняющих переменных

Найдем оценки коэффициентов уравнения регрессии, используя пакет STATISTICA, результаты представлены в таблице 2.

N=85	Regression Summary for Dependent Variable: y (исходные in Данные) R= ,64611949 R?= ,41747040 Adjusted R?= ,34756685 F(9,75)=5,9721 p<,00000 Std.Error of estimate: 7,6670 b*	Std.Err. of b*	b	Std.Err. of b	t(75)	p-value
Intercept			86,98822	11,01731	7,89559	0,000000
x1	-0,118878	0,093574	-7,40034	5,82513	-1,27042	0,207864
x2	0,143831	0,099501	0,06750	0,04669	1,44552	0,152478
x3	-0,148039	0,122300	-1,18387	0,97804	-1,21046	0,229903
x4	-0,535199	0,199906	-0,09929	0,03709	-2,67725	0,009114

x5	-0,464147	0,176798	-0,09802	0,03734	-2,62529	0,010488
x6	-0,113874	0,234889	-0,00039	0,00080	-0,48480	0,629230
x7	0,609313	0,245773	0,00028	0,00011	2,47917	0,015415
x8	-0,188241	0,103728	-0,38627	0,21285	-1,81477	0,073558
x9	-0,007593	0,132039	-0,00002	0,00039	-0,05751	0,954292

Таблица 2Результаты оценивания параметров линейной модель множественной регрессии

```
import pandas
import statsmodels.api as sm
from numpy import ones
nones=pandas.DataFrame({'1':ones(independedVariables.shape[0])})
X=nones.join(independedVariables)
Y=data['Y']
model=sm.OLS(Y, X).fit()
model.summary()
```

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.417
Model:	OLS	Adj. R-squared:	0.348
Method:	Least Squares	F-statistic:	5.972
Date:	Thu, 18 May 2023	Prob (F-statistic):	2.88e-06
Time:	13:23:21	Log-Likelihood:	-288.43
No. Observations:	85	AIC:	596.9
Df Residuals:	75	BIC:	621.3
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
1	86.9882	11.017	7.896	0.000	65.041	108.936
X1	-7.4003	5.825	-1.270	0.208	-19.005	4.204
X2	0.0675	0.047	1.446	0.152	-0.026	0.161
X3	-1.1839	0.978	-1.210	0.230	-3.132	0.764
X4	-0.0993	0.037	-2.677	0.009	-0.173	-0.025
X5	-0.0980	0.037	-2.625	0.010	-0.172	-0.024
X6	-0.0004	0.001	-0.485	0.629	-0.002	0.001
X7	0.0003	0.000	2.479	0.015	5.58e-05	0.001
X8	-0.3863	0.213	-1.815	0.074	-0.810	0.038
X9	-2.227e-05	0.000	-0.058	0.954	-0.001	0.001

Omnibus:	1.806	Durbin-Watson:	2.349
Prob(Omnibus):	0.405	Jarque-Bera (JB):	1.207
Skew:	0.248	Prob(JB):	0.547
Kurtosis:	3.308	Cond. No.	7.46e+05

Оценка уравнения регрессии выглядит следующим образом:

$Y^\wedge = 86{,}99 – 7{,}40034x_1 + 0{,}0675x_2 – 1{,}18387x_3 – 0.09929x_4 – 0.09802x_5 –$

(11,01731) (5,82513) (0,04669) (0,97804) (0,03709) (0,03734)

- $0.0039x_6 + 0,00028x_7 - 0.38627x_8 - 0,00002x_9$

(0,0008) (0,00011) (0,21285) (0,00039)

Далее необходимо вычислить оценку вектора регрессионных остатков.

$e = y - \hat{y}$

Оценка вектора регрессионных остатков представлена в приложении Б.

4. Анализ вариации результативного признака Y. Выборочный коэффициент детерминации

В случае линейной множественной регрессии, для оценки вариации результативного признака используется коэффициент детерминации $R^2_{y/x_1,...,x_n} \equiv R^2$. Определим выборочную вариацию результативной переменной *Y*.

$$Q_{\hat{i}\acute{a}\grave{u}} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = (Y - \bar{Y})^T(Y - \bar{Y})$$

где $\bar{y} = \sum_{i=1}^{n} y_i / n$ –выборочное среднее; $Y = (y_1,..., y_n)^T$, $\bar{Y}_{n*1} = (\bar{y}, \bar{y},..., \bar{y})^T$.

Следовательно, оценка коэффициента детерминации:

$$\hat{R}^2_{y/x_1,...,x_n} \equiv \frac{Q_{факт}}{Q_{ост}} \equiv 1 - \frac{Q_{ост}}{Q_{общ}}$$

Выборочный коэффициент детерминации характеризует долю общей вариации результативного признака *y*, объясняемую вариацией выборочной функции регрессии.

Рассчитав с помощью STATISTICA (рис.1), получим:

$R^2_{y/x1,x2,x3}$=0,41747040 $R_{y/x1,x2,x3}$=0,64611949.

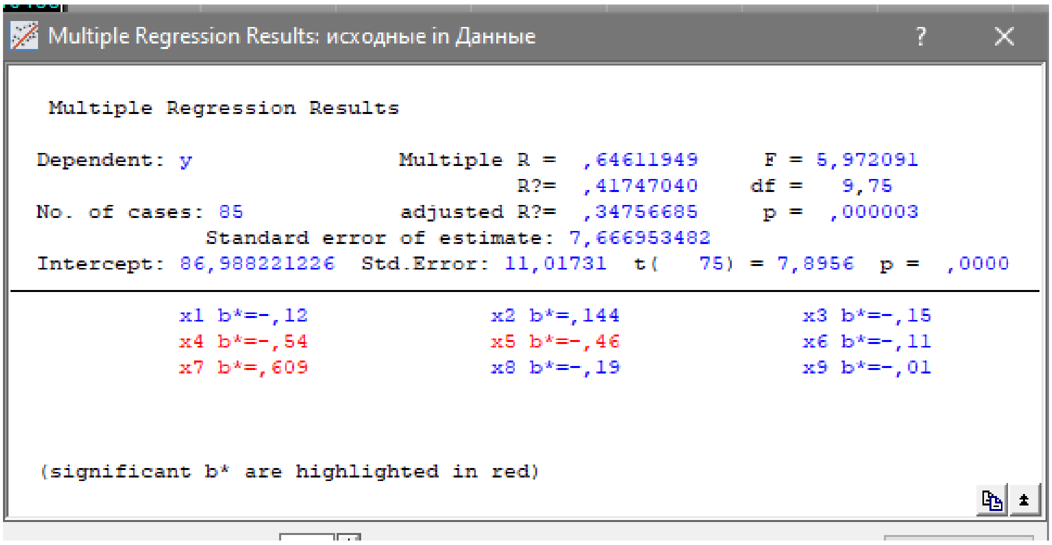


Рисунок 1оценка коэффициента детерминации

Из полученных результатов можно сделать следующий вывод:

– доля вариации результатов голосования о поправках в Конституцию, обусловленные выбранными объясняющими признаками, составляет 41,75%, т.е. на 58,25% вариация результатов голосования о поправках в Конституцию зависит от неучтенных признаков.

5. Проверка гипотезы о нормальном характере распределения регрессионных остатков

Дальнейшее изучение свойств оценок КЛММР проводится при дополнительном предположении и нормальном характере распределения регрессионных остатков:

$$\varepsilon_i \in N(0,\sigma^2), \quad i=1,n$$
$$\varepsilon \in N(0,\sigma^2 E_n)$$

Это предположение необходимо проверить.

Выдвинем гипотезы:

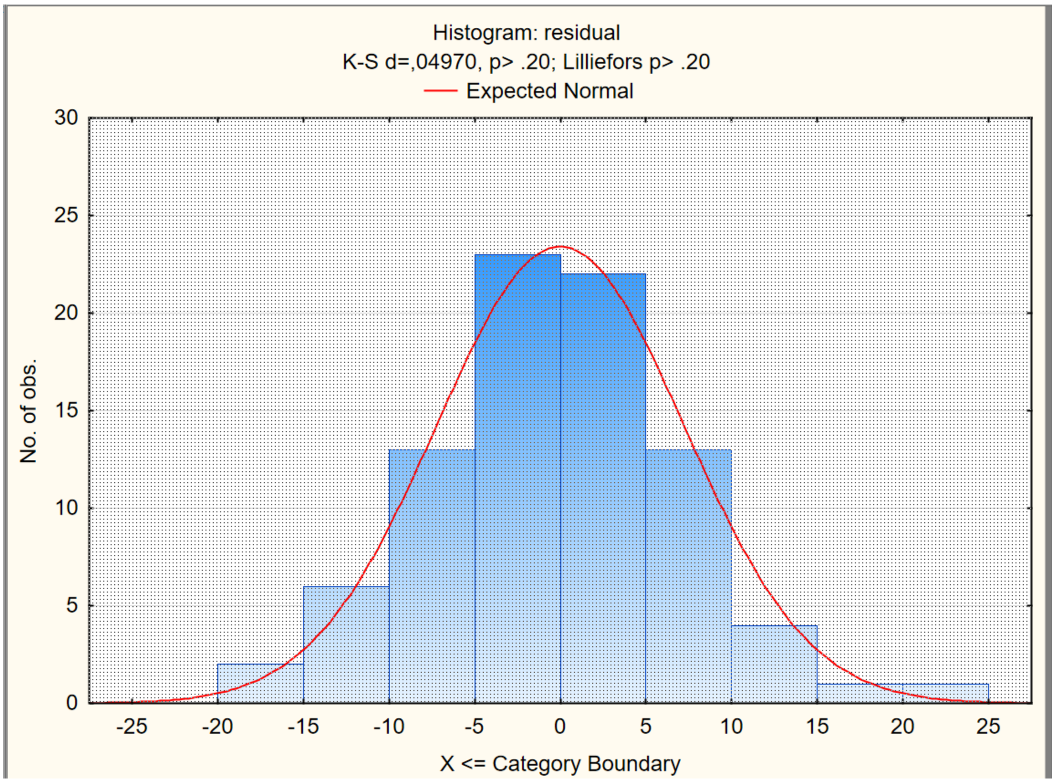
Гипотеза Н0: Распределение регрессионных остатков не отличается от нормального.

Гипотеза Н1: Распределение регрессионных остатков отличается от нормального.

Для проверки гипотезы воспользуемся критерием Колмогорова-Смирнова. В пакете программ «Statistica» получим следующий результат:

Category	Frequency table: residual (исходные in Данные) K-S d=,04970, p> .20; Lilliefors p> .20					
	Count	Cumulative Count	Percent of Valid	Cumul % of Valid	% of all Cases	Cumulative % of All
-25,0000<x<=-20,0000	0	0	0,0000	0,0000	0,00000	0,0000
-20,0000<x<=-15,0000	2	2	2,3529	2,3529	0,21906	0,2191
-15,0000<x<=-10,0000	6	8	7,0588	9,4118	0,65717	0,8762
-10,0000<x<=-5,00000	13	21	15,2941	24,7059	1,42388	2,3001
-5,00000<x<=0,000000	23	44	27,0588	51,7647	2,51917	4,8193
0,000000<x<=5,000000	22	66	25,8824	77,6471	2,40964	7,2289
5,000000<x<=10,00000	13	79	15,2941	92,9412	1,42388	8,6528
10,00000<x<=15,00000	4	83	4,7059	97,6471	0,43812	9,0909
15,00000<x<=20,00000	1	84	1,1765	98,8235	0,10953	9,2004
20,00000<x<=25,00000	1	85	1,1765	100,0000	0,10953	9,3100
Missing	828	913	974,1176		90,69003	100,0000

Таблица 3 Проверка гипотезы о нормальном законе распределения регрессионных остатков




```
import matplotlib.pyplot as plt
from scipy.stats import norm, kstest
```

```
from numpy import arange
plt.figure(figsize=(8,5))
histData=plt.hist(model.resid)
range_=arange(min(model.resid),max(model.resid),0.05)
coefY=len(model.resid)*(histData[1][1]-histData[1][0])
plt.plot(range_,
    [norm(model.resid.mean(),model.resid.std()).pdf(x)*coefY for x in range_],
    color='r')
plt.xticks(histData[1])
KS_maxD,
KS_PValue=kstest(model.resid,cdf='norm',args=(model.resid.mean(),model.resid.std()))
plt.title("Histogram of the distributionof regression residues\n"+
    "Distribution: Normal\n"+
    "Kolmogorov-Smirnov test = {:.5}, p-value = {:.5}".format(KS_maxD,
KS_PValue),fontsize=18)
plt.ylabel("No. Of observations",fontsize=15)
plt.xlabel("Category (upper limits)",fontsize=15)
plt.grid()
plt.show()
```

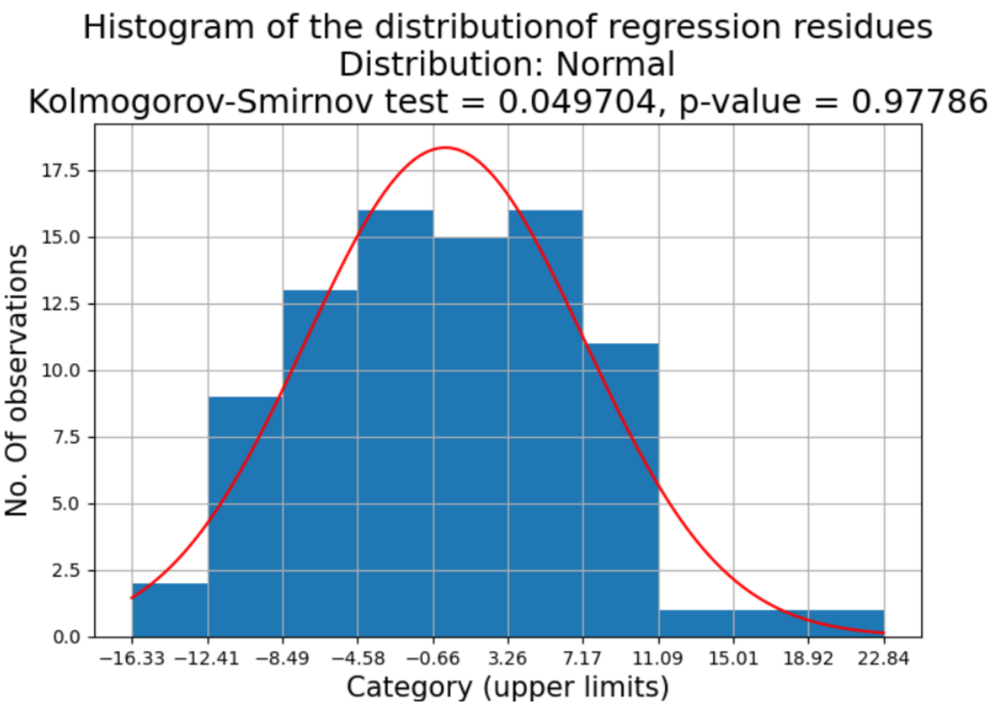


Рисунок 2 Гистограмма распределения регрессионных остатков.

Согласно тесту, величина $\lambda = \frac{\max D}{\sqrt{n}} = \frac{0,4970}{\sqrt{85}}$ = 0,05391 меньше критического уровня $\lambda_{\text{крит}}$=0,148 (квантиль распределения Колмогорова с числом степеней свободы 85 и уровнем доверия 0,05), а значение p-value > 0,05; следовательно, анализируемые остатки распределены нормально.

6. Проверка значимости уравнения регрессии и значимости коэффициентов

Для проверки значимости построенного уравнения регрессии выдвигается гипотеза H_0: линейная модель множественной регрессии не значима, что формально можно сформулировать так

H_0: $\beta_1=\beta_2=\ldots=\beta_к=0$

Альтернативная гипотеза H_1: ЛММР значима или формально

H_1: $\exists j \in [\overline{1,n}]: \quad \beta_j \neq 0$.

Для проверки гипотезы H_0 используем статистику:

$$F = \frac{Q_{факт}/k}{Q_{ост}/(n-k-1)} = \frac{\hat{R}^2_{y/x_1,\ldots,x_n}/k}{(1-\hat{R}^2_{y/x_1,\ldots,x_n})/(n-k-1)},$$

которая в случае справедливости H_0 имеет распределение Фишера – Снедекорра с числом степеней свободы $\nu_1 = k$ и $\nu_2 = n-k-1$.

Проверим гипотезу о незначимости ЛММР:

H_0: $\beta_1=\beta_2=\beta_3= \beta_4= \beta_5 = \beta_6= \beta_7 = \beta_8 =0$

Альтернативная гипотеза H_1:

H_1: $\exists j \in [\overline{1,8}]: \quad \beta_j \neq 0$.

По итогам проверки в пакете STATISTICA(см. рис.1) получили, что уровень значимости p<0.05. Таким образом, построенная ЛММР значима. Теперь проверим значимость коэффициентов ЛММР.

7. Проверка гипотез о значимости коэффициента ЛММР

В случае если нулевая гипотеза о незначимости уравнения регрессии отвергнута, проверяем гипотезы о значимости коэффициентов уравнения регрессии. Выдвигаются гипотезы вида:

H_0: коэффициент β_j незначимо отличен от нуля (или формально H_0: $\beta_j=0$); альтернативная гипотеза H_1: коэффициент β_j – значимо отличен от нуля (формально H_1: $\beta_j \neq 0$).

Для проверки таких гипотез H_0 строятся статистики:

$$t = \frac{b_j}{S_{b_j}}, \quad j = 1,2,...,k, \quad S_{b_j} = S\sqrt{[(X^T X)^{-1}]_{jj}},$$

которые в случае справедливости H_0, имеют распределение Стьюдента с $\nu = n - k - 1$ степенями свободы. Далее, либо сравниваем $t_{набл}$ с $t_{кр}(\alpha)$, либо значимость нулевой гипотезы с заданным уровнем.

Проверим гипотезы о значимости коэффициентов ЛММР

H_0: $\beta_0=0$;

H_1: $\beta_0 \neq 0$.

Нулевая гипотеза принимается (р=0,00<0,05), коэффициент β_0 значимо отличен от нуля.

Выдвинем следующую гипотезу:

H_0: $\beta_1=0$;

H_1: $\beta_1 \neq 0$.

Нулевая гипотеза принимается (р=0,21>0,05), коэффициент β_1 не значим.

Выдвинем следующую гипотезу:

H_0: $\beta_2=0$;

H_1: $\beta_2 \neq 0$.

Нулевая гипотеза принимается (р=0,15>0,05), коэффициент β_2 не значим.

Выдвинем следующую гипотезу:

H_0: $\beta_3=0$;

H_1: $\beta_3 \neq 0$.

Нулевая гипотеза принимается (р=0, 23>0,05), коэффициент β_3 не значим.

Выдвинем следующую гипотезу:

H_0: $\beta_4=0$;

H_1: $\beta_4 \neq 0$.

Нулевая гипотеза отклоняется (p=0,0091<0,05), коэффициент β_4 значимо отличен от нуля.

Выдвинем следующую гипотезу:

H$_0$: $\beta_5=0$;

H$_1$: $\beta_5\neq 0$.

Нулевая гипотеза отклоняется (p=0,015>0,05), коэффициент β_5 значимо отличен от нуля.

Выдвинем следующую гипотезу:

H$_0$: $\beta_6=0$;

H$_1$: $\beta_6\neq 0$.

Нулевая гипотеза принимается (p=0,63>0,05), коэффициент β_6 не значим.

Выдвинем следующую гипотезу:

H$_0$: $\beta_7=0$;

H$_1$: $\beta_7\neq 0$.

Нулевая гипотеза отклоняется (p=0,015<0,05), коэффициент β_6 значимо отличен от нуля.

Выдвинем следующую гипотезу:

H$_0$: $\beta_8=0$;

H$_1$: $\beta_8\neq 0$.

Нулевая гипотеза принимается (p=0,07>0,05), коэффициент β_6 не значим.

Выдвинем следующую гипотезу:

H$_0$: $\beta_9=0$;

H$_1$: $\beta_9\neq 0$.

Нулевая гипотеза принимается (p=0,95>0,05), коэффициент β_6 не значим.

8. Построение доверительных интервалов для значимых коэффициентов КЛМНР

Для коэффициентов уравнения регрессии значимо отличных от нуля находим доверительные интервалы, используя статистику:

$$t=\frac{b_j-\beta_j}{S_{b_j}}$$

имеющую распределение Стьюдента с $\nu = n - k - 1$ степенями свободы.

$b_j - \gamma S_{b_{j1}} < \beta_j < b_j + \gamma S_{b_j}$

При уровне доверия α доверительный интервал будет задаваться следующим образом:

$$\mathrm{P}\left\{-t_{\frac{a}{2},n-k-1} < \frac{b_j - B_j}{S_{b_j}} < t_{\frac{a}{2},n-k-1}\right\} = 1 - a$$

$$\mathrm{P}\left\{b_j - S_{b_j} t_{\frac{a}{2},n-k-1} < B_j < b_j + S_{b_j} t_{\frac{a}{2},n-k-1}\right\} = 1 - a$$

Построим доверительный интервал для всех коэффициентов с уровнем значимости 0,05:

			left	right
86,98822	11,01731		65,04061	108,93584
-7,40034	5,82513		-19,0046	4,20392
0,06750	0,04669		-0,02552	0,16052
-1,18387	0,97804		-3,13222	0,76447
-0,09929	0,03709		-0,17316	-0,02541
-0,09802	0,03734		-0,1724	-0,02364
-0,00039	0,00080		-0,00198	0,00120
0,00028	0,00011		5,58E-05	0,00051
-0,38627	0,21285		-0,81028	0,03775
-0,00002	0,00039		-0,00079	0,00075

Рисунок3 Доверительные интервалы

Выводы.

1. Модель регрессии значима. Оценка уравнения регрессии имеет вид:

$Y = 86{,}99 - 7{,}40034x_1 + 0{,}0675x_2 - 1{,}18387x_3 - 0.09929x_4 - 0.09802x_5 -$

$- 0.0039x_6 + 0{,}00028x_7 - 0.38627x_8 - 0{,}00002x_9$

($p=0.000<0.05$ $F_{набл} = 5{,}97 > F_{крит} = 2{,}04$).

2. При увеличении задолженностей по кредитам в рублях, предоставленными кредитными организациями физическим лицам на 1, результаты голосования по поправкам в Конституцию уменьшатся на 0,09929.
3. При увеличении средств (вкладов) физических лиц на рублевых счетах в Сбербанке России на 1 результаты голосования по поправкам в Конституцию уменьшатся на 0,09802.

4. При увеличении среднедушевого дохода населения на 1 результаты голосования по поправкам в Конституцию увеличатся на 0,0028.

9. Проверка мультиколлинеарности

9.1. Внешние признаки

1. *Небольшие изменения исходных статистических данных приводит к сильному изменению оценок коэффициентов, вплоть до изменения знака.*

Исходная таблица была изменена – данные по признаку x_4 были увеличены на 10%, а по признаку x_9 уменьшены на 10% (признаки были выбраны случайным образом) – и выгружена в Statistica для повторного проведения регрессионного анализа.

Сравним коэффициенты, полученные изначально с коэффициентами, полученными после изменения исходных данных.

N=85	Regression Summary for Dependent Variable: y (исходные in Данные) R= ,64611949 R?= ,41747040 Adjusted R?= ,34756685 F(9,75)=5,9721 p					
	b*	Std.Err. of b*	b	Std.Err. of b	t(75)	p-value
Intercept			86,98822	11,01731	7,89559	0,000000
x1	-0,118878	0,093574	-7,40034	5,82513	-1,27042	0,207864
x2	0,143831	0,099501	0,06750	0,04669	1,44552	0,152478
x3	-0,148039	0,122300	-1,18387	0,97804	-1,21046	0,229903
x4	-0,535199	0,199906	-0,09929	0,03709	-2,67725	0,009114
x5	-0,464147	0,176798	-0,09802	0,03734	-2,62529	0,010488
x6	-0,113874	0,234889	-0,00039	0,00080	-0,48480	0,629230
x7	0,609313	0,245773	0,00028	0,00011	2,47917	0,015415
x8	-0,188241	0,103728	-0,38627	0,21285	-1,81477	0,073558
x9	-0,007593	0,132039	-0,00002	0,00039	-0,05751	0,954292

Таблица 4. Regression summary для исходных значений

N=85	Regression Summary for Dependent Variable: y (с изменениями in Данные) R= ,64611949 R?= ,41747040 Adjusted R?= ,34756685 F(9,75)=5,9721 p					
	b*	Std.Err. of b*	b	Std.Err. of b	t(75)	p-value
Intercept			86,98822	11,01731	7,89559	0,000000
x1	-0,118878	0,093574	-7,40034	5,82513	-1,27042	0,207864
x2	0,143831	0,099501	0,06750	0,04669	1,44552	0,152478
x3	-0,148039	0,122300	-1,18387	0,97804	-1,21046	0,229903
x4	-0,535199	0,199906	-0,09026	0,03371	-2,67725	0,009114
x5	-0,464147	0,176798	-0,09802	0,03734	-2,62529	0,010488
x6	-0,113874	0,234889	-0,00039	0,00080	-0,48480	0,629230
x7	0,609313	0,245773	0,00028	0,00011	2,47917	0,015415
x8	-0,188241	0,103728	-0,38627	0,21285	-1,81477	0,073558

x9	-0,007593	0,132039	-0,00002	0,00035	-0,05751	0,954292

Таблица 5. Regression summary для изменённых значений

Значения коэффициентов при объясняющих переменных поменялись слабо или не поменялись вовсе. Сохранился в исходном виде и набор значимых переменных.

2. *Среди коэффициентов много незначимых коэффициентов, хотя в целом модель значима.*

Из 9 коэффициентов 3 являются значимыми – X_4, X_5, X_7. При этом модель в целом является значимой.

3. *Стандартные отклонения оценок велики настолько, что сравнимы с ними или даже превосходят их.*

Из табл. 2 видно, что у 2 из 9 объясняющих переменных значение стандартного отклонения оценки по модулю превышает величину самой оценки коэффициента – X_6, X_9.

4. *Доверительные интервалы для коэффициентов уравнения содержат точку 0.*

Из рисунка 3 видно, что у коэффициентов при всех незначимых объясняющих переменных – X_1, X_2, X_3, X_6 X_8, X_9 – доверительные интервалы оценки содержат точку 0.

9.2. Формальные признаки

1. Достаточно высокие значения коэффициентов парной или частной корреляции между объясняющими признаками.

Рассмотрим корреляционную матрицу оценок коэффициентов (табл. 6):

Variable	Correlations (с изменениями in Данные)									
	x1	x2	x3	x4	x5	x6	x7	x8	x9	y
x1	1,0000	-0,1934	-0,0154	0,107	0,1129	0,1621	0,1759	-0,1264	0,2539	-0,1435
x2	-0,1934	1,0000	0,0579	-0,341	-0,2679	-0,3184	-0,2766	0,3005	-0,2221	0,278
x3	-0,0154	0,0579	1,0000	0,1736	0,2454	0,3034	-0,0197	-0,0534	0,3435	-0,3838
x4	0,107	-0,341	0,1736	1,0000	0,6394	**0,8521**	**0,8121**	-0,3564	0,4871	-0,4583
x5	0,1129	-0,2679	0,2454	0,6394	1,0000	**0,7691**	**0,7847**	-0,4492	0,6231	-0,4242
x6	0,1621	-0,3184	0,3034	**0,8520**	**0,7691**	1,0000	**0,8325**	-0,3292	0,6604	-0,4726
x7	0,1759	-0,2766	-0,0197	**0,8121**	**0,7847**	**0,8325**	1,0000	-0,3309	0,5946	-0,2843
x8	-0,1264	0,3005	-0,0534	-0,3564	-0,4492	-0,3292	-0,3309	1,0000	-0,2731	0,1151
x9	0,2539	-0,2221	0,3435	0,4871	0,6231	0,6604	0,5946	-0,2731	1,0000	-0,332
y	-0,1435	0,278	-0,3838	-0,4583	-0,4242	-0,4726	-0,2843	0,1151	-0,3312	1,000

Таблица 6. Корреляционная матрица оценок коэффициентов

```
data.corr()
```

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
Y	1.000000	-0.143489	0.277960	-0.383814	-0.458254	-0.424248	-0.472596	-0.284304	0.115087	-0.331981
X1	-0.143489	1.000000	-0.193382	-0.015433	0.106963	0.112885	0.162071	0.175871	-0.126350	0.253945
X2	0.277960	-0.193382	1.000000	0.057859	-0.340968	-0.267945	-0.318367	-0.276564	0.300528	-0.222053
X3	-0.383814	-0.015433	0.057859	1.000000	0.173605	0.245387	0.303385	-0.019712	-0.053363	0.343512
X4	-0.458254	0.106963	-0.340968	0.173605	1.000000	0.639371	0.852005	0.812053	-0.356414	0.487143
X5	-0.424248	0.112885	-0.267945	0.245387	0.639371	1.000000	0.769099	0.784695	-0.449224	0.623081
X6	-0.472596	0.162071	-0.318367	0.303385	0.852005	0.769099	1.000000	0.832521	-0.329182	0.660357
X7	-0.284304	0.175871	-0.276564	-0.019712	0.812053	0.784695	0.832521	1.000000	-0.330922	0.594596
X8	0.115087	-0.126350	0.300528	-0.053363	-0.356414	-0.449224	-0.329182	-0.330922	1.000000	-0.273132
X9	-0.331981	0.253945	-0.222053	0.343512	0.487143	0.623081	0.660357	0.594596	-0.273132	1.000000

Таблица 6.1 Корреляционная матрица оценок коэффициентов

Считается, что наличие значимых коэффициентов корреляции, по абсолютной величине превосходящих 0,7-0,8, свидетельствуют о присутствии мультиколлинеарности.

На основе вычисленной матрицы есть основания подозревать тесную связь между X_4 и X_6 ($r(x^{(4)}, x^{(6)})=0,85$), X_4 и X_6 ($r(x^{(4)}, x^{(6)})=0,81$), X_5 и X_6 ($r(x^{(5)}, x^{(6)})=0,77$), X_5 и X_7 ($r(x^{(5)}, x^{(7)})=0,78$) и X_6 и X_7 ($r(x^{(6)}, x^{(7)})=0,83$).

2. Высокое значение коэффициента детерминации или множественного коэффициента корреляции при построении регрессионной модели зависимости одной из объясняющих переменных от всех других.

Рассмотрим оценки коэффициентов детерминации регрессионных моделей зависимости объясняющей переменной xj от всех остальных переменных x1,x2,...,xk\xj, j= 1,k.

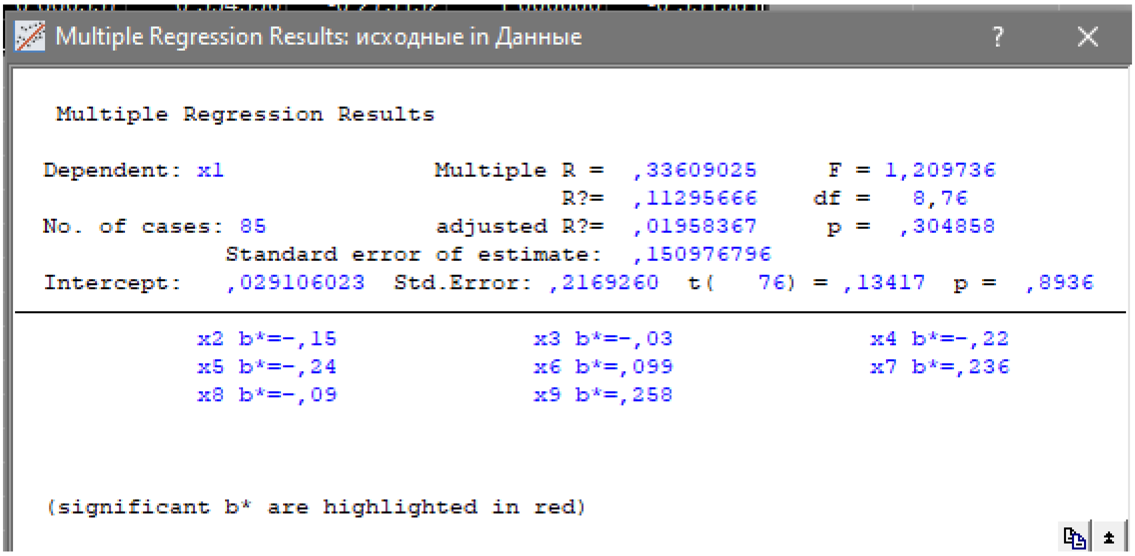


Рисунок 4 Расчет коэффициента детерминации

Все расчеты остальных коэффициенты детерминации производятся аналогичным образом. В результате получили:

${R^2}_{x_1x_2x_3x_4x_5x_6x_7x_8x_9}$=0,11295666

${R^2}_{x_2x_1x_3x_4x_5x_6x_7x_8x_9}$=0,2154847

${R^2}_{x_3x_2x_1x_4x_5x_6x_7x_8x_9}$=0,48071533

${R^2}_{x_4x_2x_3x_1x_5x_6x_7x_8x_9}$=0, 80564107

${R^2}_{x_5x_3x_4x_1x_6x_7x_8x_9}$=0,75151527

${R^2}_{x_6x_2x_3x_4x_5x_1x_7x_8x_9}$=0,85922260

${R^2}_{x_7x_2x_3x_4x_5x_6x_1x_8x_9}$=0,87141519

${R^2}_{x_8x_2x_3x_4x_5x_6x_7x_1x_9}$=0,27811396

${R^2}_{x_9x_2x_3x_4x_5x_6x_7x_8x_1}$=0,55449441

```
from statsmodels.formula.api import ols

variables = sorted(independedVariables.columns.tolist())
for response in variables:
   print("R*2({}): {}".format(response, ols(formula="{} ~ {}".format(response, " + 
".join(set(variables) - set([response]))), data=data).fit().rsquared))
```

```
R*2(X1): 0.11295665595269455
R*2(X2): 0.21548247047490054
R*2(X3): 0.480715334013578
R*2(X4): 0.8056410721350242
R*2(X5): 0.7515152744047486
R*2(X6): 0.8592226024304545
R*2(X7): 0.8714151945179595
R*2(X8): 0.2781139575849768
R*2(X9): 0.5544944079342208
```

Видим, что некоторые значения являются достаточно большими (0,6 и более) и свидетельствуют о значимой зависимости некоторых объясняющих переменных (x_4, x_5, x_6, x_7) от всех других.

10. Устранение мультиколлинеарности. Построение моделей, не подверженных мультиколлинеарности

10.1. Метод включения переменных

В результате проведения пошаговой регрессии получили следующее:

N=85	Regression Summary for Dependent Variable: y (исходные in Данные) R= ,53573911 R?= ,28701639 Adjusted R?= ,26962655 F(2,82)=16,505 p<,00000 Std.Error of estimate: 8,1120					
	b*	Std.Err. of b*	b	Std.Err. of b	t(82)	p-value
Intercept			98,10851	5,361723	18,29794	0,000000
x6	-0,392257	0,097859	-0,00133	0,000332	-4,00840	0,000134
x3	-0,264809	0,097859	-2,11769	0,782581	-2,70603	0,008282

Таблица 7 Результаты оценивания параметров ЛММР методом пошаговой регрессии

```
from numpy import ones
def YX(data, response, vars_x):
    nones = pandas.DataFrame({'Intercept': ones(data.shape[0])})
    X=nones.join(data.loc[:, vars_x])
    Y=data[response]
    return (Y, X)
from statsmodels.formula.api import ols
def forward_selected(data, response):
    remaining = set(data.columns)
    remaining.remove(response) #удаляем У
    selected = []  #присваем переменной selected пустой список
    currentScore, bestNewScore = 0.0, 0.0
    while remaining and currentScore == bestNewScore:
        scoreR = 0.0
        for candidate in remaining:
            Y, X = YX(data, response, selected + [candidate])
            tmpRes = sm.OLS(Y, X).fit()
            if scoreR < tmpRes.rsquared:
                bestNewScore, bestCandidate = tmpRes.rsquared_adj, candidate
                scoreR = tmpRes.rsquared
        if currentScore < bestNewScore:
            remaining.remove(bestCandidate)
            selected.append(bestCandidate)
            currentScore = bestNewScore
    Y, X = YX(data, response, selected)
    return sm.OLS(Y, X).fit()
```

```
forwardModel = forward_selected(data, 'Y')
forwardModel.summary()
```

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.417
Model:	OLS	Adj. R-squared:	0.356
Method:	Least Squares	F-statistic:	6.807
Date:	Thu, 18 May 2023	Prob (F-statistic):	1.05e-06
Time:	19:23:34	Log-Likelihood:	-288.43
No. Observations:	85	AIC:	594.9
Df Residuals:	76	BIC:	616.8
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	86.8721	10.759	8.074	0.000	65.443	108.301
X6	-0.0004	0.001	-0.514	0.609	-0.002	0.001
X3	-1.2007	0.927	-1.295	0.199	-3.047	0.646

Оценка уравнения регрессии значима т.к. нулевая гипотеза отклонена; коэффициенты при переменных также значимы. Коэффициент детерминации составил 0,287.

10.2. Метод исключения переменных

В результате проведения пошаговой регрессии получили следующее:

N=85	Regression Summary for Dependent Variable: y (исходные in Данные) R= ,58004301 R?= ,33644989 Adjusted R?= ,31187396 F(3,81)=13,690 p<,00000 Std/Error of estimate: 7,8739					
	b*	Std.Err. of b*	b	Std.Err. of b	t(81)	p-value
Intercept			86,03118	2,413853	35,64061	0,000000
x4	-0,664366	0,155096	-0,12325	0,028772	-4,28359	0,000050
x5	-0,519770	0,146013	-0,10977	0,030836	-3,55974	0,000624
x7	0,663058	0,192380	0,00031	0,000090	3,44661	0,000903

Таблица 8 Результаты оценивания параметров ЛММР методом пошаговой регрессии

```
def backward_elimination(data, response):
    selected = set(data.columns)
    selected.remove(response)
    currentPValues = ols(formula="{} ~ {}".format(response, " + ".join(selected)),
data=data).fit().pvalues
    for _ in range(len(data.columns)):
        currentAdjR2 = -1.0
        if (max(currentPValues) >= 0.05):
            for i in range(1, len(currentPValues)):
                candidateToRemove = currentPValues.axes[0][i]
```

```
            newModel = ols(formula="{} ~ {}".format(response, " + ".join(selected -
set([candidateToRemove]))), data=data).fit()
            newAdjR2 = newModel.rsquared_adj
            if (currentAdjR2 < newAdjR2):
              currentAdjR2, deletedVar, improvedPValues = newAdjR2, candidateToRemove,
newModel.pvalues
        if (currentAdjR2 == -1.0):
          break
        selected.remove(deletedVar)
        currentPValues = improvedPValues
  return ols(formula="{} ~ {}".format(response, " + ".join(sorted(selected))), data=data). fit()
backwardModel = backward_elimination(data, "Y")
backwardModel.summary()
```

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.336
Model:	OLS	Adj. R-squared:	0.312
Method:	Least Squares	F-statistic:	13.69
Date:	Thu, 18 May 2023	Prob (F-statistic):	2.63e-07
Time:	19:27:48	Log-Likelihood:	-293.96
No. Observations:	85	AIC:	595.9
Df Residuals:	81	BIC:	605.7
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	86.0312	2.414	35.641	0.000	81.228	90.834
X4	-0.1232	0.029	-4.284	0.000	-0.180	-0.066
X5	-0.1098	0.031	-3.560	0.001	-0.171	-0.048
X7	0.0003	8.97e-05	3.447	0.001	0.000	0.000

Omnibus:	0.223	Durbin-Watson:	2.292
Prob(Omnibus):	0.895	Jarque-Bera (JB):	0.213
Skew:	0.112	Prob(JB):	0.899
Kurtosis:	2.902	Cond. No.	1.49e+05

Оценка уравнения регрессии значима т.к. нулевая гипотеза отклонена; коэффициенты при переменных также значимы. Коэффициент детерминации составил 0,336.

10.3 Метод ридж-регрессии

В результате проведения пошаговой регрессии получили следующее:

	Ridge Regression Summary for Dependent Variable: y (исходные in Данные) l=,10000 R= ,53043830 R?= ,28136479 Adjusted R?= ,26383710 F(2,82)=16,053 p<,00000 Std.Error of estimate: 8,1441

N=85	b*	Std.Err. of b*	b	Std.Err. of b	t(82)	p-value
Intercept			85,77476	2,357738	36,38011	0,000000
x3	-0,290407	0,090392	-2,32239	0,722865	-3,21276	0,001880
x4	-0,370761	0,090392	-0,06878	0,016769	-4,10172	0,000096

Таблица 9 Результаты оценивания параметров ЛММР методом ридж-регрессии

Коэффициент детерминации составил 0,281.

11. Выводы

Модель регрессии значима.

При использовании метода включения и метода исключения переменных были получены разные уравнения.

Метод включения переменных (R^2=0,287):

y^=98,11-0,00133x_6-2,11769x_3

Метод исключения переменных (R^2=0,336):

y^=86,03118 – 0,12325x_4-0,10977x_5+0,00031x_7

Метод ридж-регрессии (R^2=0,281):

y^=85,77476 – 2,32239x_3-0,06878x_4

Таким образом, наиболее сильная зависимость результатов голосования по поправкам в Конституцию наблюдается от:

1. Задолженности по кредитам предоставленным кредитным организациями физическим лицам,
2. Средства (вклады) физических лиц на рублевых счетах в Сбербанке России
3. Среднедушевые доходы населения, тыс.

Приложение А (исходные данные)

Наименование	**Результаты голосования по поправкам в Конституцию, % за**	Число организаций, на которых проходили забастовки	Численность населения на одну больничную койку	Численность инвалидов по группе инвалидности	Задолженность по кредитам в рублях, предоставленным кредитными организациями физическим лицам	Средства (вклады) физических лиц на рублевых счетах в Сбербанке	Средний размер назначенных пенсий, тыс.	Среднедушевые доходы населения, тыс.	Уровень бедности, %	Средний прожиточный минимум
Алтайский край	**71,91**	0	110,6	0,092	114	59	14493	32729	9,3	10665
Амурская область	**70,42**	0	101,8	0,142	153	74	15902	59187	16,5	14017
Архангельская область без автономного ок	**65,78**	0	120,5	0,226	172	121	19760	57076	14,2	22219
Астраханская область	**86,7**	0	121	0,083	120	56	13970	41615	12,3	11069
Белгородская область	**81,76**	0	141	0,052	109	84	15549	41349	11,7	9720
Брянская область	**87,55**	0	135,8	0,051	98	60	14713	35578	15,5	11280
Владимирская область	**71,18**	0	114,6	0,092	105	73	15555	37708	7,0	11093
Волгоградская область	**83,39**	0	114,1	0,047	102	58	14902	38073	13,0	10158
Вологодская область	**71,16**	0	129,8	0,380	129	77	16377	45453	11,8	11767
Воронежская область	**80,07**	0	127,5	0,027	109	93	14670	40730	11,3	9747
г. Москва	**65,29**	0	157,5	0,010	178	245	16935	111395	12,4	18029
Еврейская автономная область	**77,30**	0	77,8	0,502	122	67	15345	50039	7,9	16386
Забайкальский край	**72,92**	1	102,6	0,109	132	61	14679	50193	5,5	13613,32
Ивановская область	**76,23**	0	120,4	0,107	94	71	15024	31881	22,4	10761
Иркутская область	**64,28**	0	110,8	0,062	146	83	16217	54318	19,3	12243
Кабардино-Балкарская Республика	**90,48**	0	137,9	0,111	59	29	12216	31172	12,9	12576
Калининградская область	**72,5**	0	128,7	0,988	132	81	14974	41298	16,4	12067
Калужская область	**70,17**	0	130,4	0,057	139	86	15642	47854	18,1	11618
Камчатский край	**61,76**	1	92,5	0,218	201	163	23317	92667	13,4	21797
Карачаево-Черкесская Республика	**84,8**	0	134,8	0,195	77	30	13138	32128	8,9	10451
Кемеровская область — Кузбасс	**87,29**	0	106,7	0,001	123	63	15760	48411	13,9	10727
Кировская область	**70,89**	0	111	0,077	124	60	15572	35390	22,1	10710
Костромская область	**67,92**	0	106,8	0,167	107	67	14998	34989	12,5	11241
Краснодарский край	**88,92**	0	133,8	0,012	117	66	14627	40955	13,3	11397
Красноярский край	**70,10**	0	124,5	0,046	152	72	16815	60026	11,8	13409
Курганская область	**69,58**	0	102	0,040	121	51	14688	35356	9,7	10696
Курская область	**73,92**	0	113,5	0,053	107	59	14488	40206	15,9	10459
Ленинградская область	**79,12**	0	163	0,025	159	67	16210	52513	18,5	11289
Липецкая область	**78,12**	0	119,9	0,501	105	71	15060	40185	9,1	9945
Магаданская область	**62,03**	0	96,9	0,251	242	219	23455	108460	7,9	21102
Московская область	**78,96**	0	138,8	0,003	190	106	16805	63414	8,1	13580
Мурманская область	**62,54**	0	99,6	0,196	196	172	21097	75495	7,9	18625
Ненецкий автономный округ	**44,42**	0	134,6	10,941	216	185	23854	44854	9,3	22219
Нижегородская область	**79,31**	0	110,4	0,015	116	91	15589	41162	8,4	13857
Новгородская область	**71,44**	0	124,8	0,303	118	71	15583	39473	12,9	10833
Новосибирская область	**67,59**	0	107,3	0,128	155	66	15455	46444	12,9	11408
Омская область	**62,09**	0	122,5	0,012	124	56	14824	40655	12,8	12284
Оренбургская область	**73,60**	0	131,2	0,189	136	50	14564	38146	13,7	10635
Орловская область	**79,74**	0	127,6	0,105	115	77	15240	35174	12,1	9938
Пензенская область	**85,25**	0	123,1	0,049	108	71	14520	36058	12,0	10722
Пермский край	**70,75**	0	132,6	0,022	141	85	15806	45178	12,8	9610
Приморский край	**78,86**	0	103,7	0,032	144	93	16034	55076	12,3	10844
Псковская область	**78,74**	0	112,3	0,401	108	69	14961	34727	14,6	13963
Республика Адыгея (Адыгея)	**84,55**	0	148,1	0,442	95	36	13838	35633	11,5	11807
Республика Алтай	**66,16**	0	137,4	1,230	105	36	13837	38702	22,4	9779
Республика Башкортостан	**88,68**	0	128,2	0,015	134	50	15256	41647	11,4	10838
Республика Бурятия	**71,95**	0	126,4	0,070	115	42	14703	44709	19,5	10015
Республика Дагестан	**90,34**	0	153,8	0,092	26	21	11923	32180	14,7	12810
Республика Ингушетия	**85,89**	0	218,5	0,210	16	12	13282	31230	29,3	10628
Республика Калмыкия	**84,00**	0	125,9	0,300	147	40	13269	33167	22,6	10964
Республика Карелия	**70,46**	0	131,2	0,381	169	107	19336	49303	14,4	10631
Республика Коми	**65,08**	0	110,2	0,166	195	118	19879	59721	15,3	14467
Республика Крым	**90,07**	0	117,8	0,149	31	46	13981	37993	16,2	14567
Республика Марий Эл	**75,76**	0	121,8	0,282	104	50	14134	35202	17,9	10998
Республика Мордовия	**85,60**	0	125,9	0,126	101	46	14385	34441	16,5	10184
Республика Саха (Якутия)	**58,34**	0	117,4	0,217	219	73	20793	82555	16,3	9716
Республика Северная Осетия — Алания	**82,83**	0	127,2	0,205	87	48	13841	32603	13,4	18368
Республика Татарстан (Татарстан)	**82,81**	0	156,3	0,028	136	60	15414	44854	6,2	10596
Республика Тыва	**96,79**	0	89	0,222	129	23	14781	46092	28,2	9955
Республика Хакасия	**69,52**	0	140,1	0,538	118	47	15497	47148	17,9	11494
Ростовская область	**83,54**	0	127,2	0,042	113	65	14397	39045	12,0	11784
Рязанская область	**79,1**	0	129,1	0,054	122	80	15098	40282	12,4	11053
Самарская область	**80,55**	0	134,5	0,005	121	86	15553	42729	11,7	10785
Санкт-Петербург	**77,66**	0	118,1	0,018	195	136	17279	75920	5,0	10979
Саратовская область	**82,24**	0	102	0,095	104	60	14440	37089	14,0	11910,4
Сахалинская область	**74,84**	0	87,4	0,047	212	173	20660	94587	7,4	9844
Свердловская область	**65,97**	0	111,6	0,005	146	75	16366	48347	8,6	15906
Севастополь	**84,6**	0	146,7	0,053	34	172	15203	39127	10,5	11206
Смоленская область	**71,97**	0	106	0,202	116	75	14839	36097	14,3	11380
Ставропольский край	**84,85**	0	121,4	0,067	96	62	14119	37090	13,0	11201
Тамбовская область	**87,41**	0	133,4	0,222	101	71	14137	33689	10,5	10081
Тверская область	**71,87**	0	106,7	0,184	127	77	15353	39692	10,8	10221
Томская область	**64,86**	0	99,7	0,181	131	61	17107	50481	13,5	11234,14
Тульская область	**83,06**	0	114,6	0,088	124	71	15558	43892	9,7	12218
Тюменская область без автономных округо	**85,57**	0	132,5	0,050	176	74	20292	55770	11,4	11134
Удмуртская Республика	**68,92**	0	128,2	0,041	142	58	15551	38564	13,2	16281
Ульяновская область	**71,16**	0	123,2	0,068	115	69	14534	35991	11,3	10465
Хабаровский край	**62,28**	0	108,9	0,011	158	104	18101	59247	13,8	10642
Ханты-Мансийский автономный округ — К	**69**	0	134,8	0,066	271	92	23055	85211	12,0	16222
Челябинская область	**69,54**	0	125,9	0,020	123	56	15618	43477	12,0	11430
Чеченская Республика	**97,92**	0	162,8	0,044	25	6	13485	30942	19,8	11240
Чувашская Республика — Чувашия	**72,57**	0	122,8	0,005	126	65	14639	35531	15,7	9804
Чукотский автономный округ	**80,3**	0	77,5	0,493	231	228	26381	130483	7,3	24164
Ямало-Ненецкий автономный округ	**89,16**	0	130,2	0,025	335	136	22966	115987	4,6	17029
Ярославская область	**70,69**	0	103,8	0,002	108	90	16064	41389	8,9	10742

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9
0	71.91	0	110.6	0.091998	113.818441	59.074254	14493	32729.066667	9.3	10665.0
1	70.42	0	101.8	0.142170	152.949959	74.094058	15902	59187.475000	16.5	14017.0
2	65.78	0	120.5	0.226177	171.649203	120.779311	19760	57076.416667	14.2	22219.0
3	86.70	0	121.0	0.082606	119.958612	55.908922	13970	41614.916667	12.3	11069.0
4	81.76	0	141.0	0.051596	109.315794	83.862578	15549	41348.641667	11.7	9720.0
...	...	...	...	...	...	...	...	...	...	...
80	97.92	0	162.8	0.044138	25.406658	6.455157	13485	30942.166667	19.8	11240.0
81	72.57	0	122.8	0.004701	126.209976	65.010964	14639	35530.791667	15.7	9804.0
82	80.30	0	77.5	0.492757	230.764414	227.842024	26381	130482.900000	7.3	24164.0
83	89.16	0	130.2	0.025424	334.799039	136.267614	22966	115986.616667	4.6	17029.0
84	70.69	0	103.8	0.001554	107.548126	90.056128	16064	41388.708333	8.9	10742.0

Приложение Б

Оценка вектора регрессионных остатков:

	Predicted & Residual Values y		
	Observed Value	Predicted Value	Residual
Алтайский край	71,910004	77,115379	-5,205376
Амурская область	70,419998	75,220558	-4,800560
Архангельская областъ	65,779999	68,564087	-2,784088
Астраханская область	86,699997	79,088120	7,611877
Белгородская область	81,760002	78,366714	3,393288
Брянская область	87,550003	78,694267	8,855736
Владимирская область	71,180000	78,789879	-7,609879
Волгоградская област	83,389999	78,570740	4,819260
Вологодская область	71,160004	76,708595	-5,548592
Воронежская область	80,070000	76,955391	3,114609
г. Москва	65,290001	75,818726	-10,528725
Еврейская автономная	77,300003	77,818024	-0,518021
Забайкальский край	72,919998	73,442314	-0,522316
Ивановская область	76,230003	73,084396	3,145607
Иркутская область	64,279999	73,221298	-8,941299
Кабардино-Балкарская	90,480003	86,256775	4,223228
Калининградская обла	72,500000	72,789749	-0,289749
Калужская область	70,169998	73,796486	-3,626488
Камчатский край	61,759998	61,237682	0,522316
Карачаево-Черкесская	84,800003	85,705223	-0,905220
Кемеровская область	87,290001	77,886955	9,403046
Кировская область	70,889999	71,417877	-0,527878
Костромская область	67,919998	75,864792	-7,944794
Краснодарский край	88,919998	78,413628	10,506371
Красноярский край	70,099998	78,878838	-8,778839
Курганская область	69,580002	77,207672	-7,627670
Курская область	73,919998	77,618118	-3,698120
Ленинградская област	79,120003	76,789146	2,330856
Липецкая область	78,120003	78,965622	-0,845619
Магаданская область	62,029999	65,912186	-3,882187
Московская область	78,959999	75,211060	3,748940
Мурманская область	62,540001	66,925911	-4,385910
Ненецкий автономный	44,419998	43,008446	1,411552
Нижегородская област	79,309998	76,020241	3,289757
Новгородская область	71,440002	76,407677	-4,967674
Новосибирская област	67,589996	74,164818	-6,574821
Омская область	62,090000	78,030731	-15,940731
Оренбургская область	73,599998	76,887688	-3,287689
Орловская область	79,739998	75,719589	4,020409
Пензенская область	85,250000	77,334511	7,915489
Пермский край	70,750000	75,153603	-4,403603
Приморский край	78,860001	74,956192	3,903809
Псковская область	78,739998	74,676392	4,063606

Республика Адыгея (А	84,550003	83,482155	1,067848
Республика Алтай	66,160004	77,630211	-11,470207
Республика Башкортос	88,680000	78,714325	9,965675
Республика Бурятия	71,949997	79,092300	-7,142303
Республика Дагестан	90,339996	91,197220	-0,857224
Республика Ингушетия	85,889999	90,877960	-4,987961
Республика Калмыкия	84,000000	71,915215	12,084785
Республика Карелия	70,459999	68,879494	1,580505
Республика Коми	65,080002	66,317886	-1,237885
Республика Крым	90,070000	85,947922	4,122078
Республика Марий Эл	75,760002	76,991623	-1,231621
Республика Мордовия	85,599998	78,347061	7,252937
Республика Саха (Яку	58,340000	74,666397	-16,326397
Республика Северная	82,830002	80,297516	2,532486
Республика Татарстан	82,809998	82,249268	0,560730
Республика Тыва	96,790001	73,952049	22,837952
Республика Хакасия	69,519997	79,678749	-10,158752
Ростовская область	83,540001	78,517921	5,022079
Рязанская область	79,099998	76,323593	2,776405
Самарская область	80,550003	76,962997	3,587006
Санкт-Петербург	77,660004	74,947052	2,712952
Саратовская область	82,239998	76,821373	5,418625
Сахалинская область	74,839996	70,672615	4,167381
Свердловская область	65,970001	76,403282	-10,433281
Севастополь	84,599998	77,457924	7,142075
Смоленская область	71,970001	73,769127	-1,799126
Ставропольский край	84,849998	79,249329	5,600670
Тамбовская область	87,410004	78,534973	8,875031
Тверская область	71,870003	74,802589	-2,932587
Томская область	64,860001	76,768036	-11,908035
Тульская область	83,059998	77,802238	5,257759
Тюменская область бе	85,570000	74,496475	11,073524
Удмуртская Республик	68,919998	75,272331	-6,352333
Ульяновская область	71,160004	76,997459	-5,837456
Хабаровский край	62,279999	72,704819	-10,424820
Ханты-Мансийский авт	69,000000	70,435898	-1,435898
Челябинская область	69,540001	79,143105	-9,603104
Чеченская Республика	97,919998	90,445007	7,474991
Чувашская Республика	72,570000	74,516525	-1,946526
Чукотский автономный	80,300003	69,897758	10,402245
Ямало-Ненецкий автон	89,160004	71,059753	18,100250
Ярославская область	70,690002	76,354385	-5,664383

```
0   -5.205382
1   -4.800556
2   -2.784084
3    7.611877
4    3.393288
5    8.855732
6   -7.609877
7    4.819264
dtype: float64
```