

Индивидуальное задание
по дисциплине «Эконометрика»
**Эконометрическое моделирование
рождаемости в Оренбургской
области**

Руководители работы:

канд. технических
наук, доцент

А.Г. Реннер

канд. экономических
наук, доцент

О.С. Чудинова

Исполнитель:

студент гр. 11ПМ(б)ОП

Л. В. Зверева

Цель, предмет и объект исследования

Цель исследования - выявление социально-демографических факторов и количественная оценка их влияния на рождаемость в регионе.

Предмет исследования - эконометрические модели, описывающие влияние социально-демографических факторов на рождаемость в регионе.

Объект исследования - муниципальные образования Оренбургской области.

Задачи исследования

- исходя из экономических соображений, отобрать социально-демографические факторы, влияющие на рождаемость;
- провести регрессионный анализ рождаемости на основе линейной модели множественной регрессии;
- проверить соответствие линейной модели множественной регрессии условиям Гаусса-Маркова;
- исследовать выборочную совокупность на регрессионную однородность;
- провести регрессионный анализ рождаемости на основе нелинейных моделей;
- осуществить сценарное прогнозирование рождаемости.

Набор показателей

К выбранным показателям относятся:

u - число родившихся (чел)

x_1 - число браков

x_2 - заболеваемость на 1000 человек населения (чел)

x_3 - численность безработных, зарегистрированных в государственных учреждениях службы занятости населения (чел)

x_4 - численность женщин (чел)

x_5 - средний возраст (лет)

Фрагмент информационной базы исследования

Таблица 1 – Фрагмент таблицы с исходными данными

Название объекта	Число родившихся	Число браков	Заболеваемость на 1000 чел	Число безработных	Численность женщин	Средний возраст
Абдулинский	139	97	769,3	499	5470	45,9
Адамовский	513	284	817,5	220	13770	36,9
Акбулакский	451	246	816,4	363	12876	37
Александровский	268	169	546,2	155	8284	40,2
Асекеевский	262	220	766,3	253	10974	41,4
Беляевский	272	191	906,1	172	8847	39,2
Бугурусланский	289	196	632,7	679	10254	40,4
Бузулукский	522	310	737,4	600	16472	40,2
Гайский	188	124	892,9	430	5380	39
Грачевский	183	152	835,4	68	7171	41,5
Домбаровский	332	185	782,2	224	8423	36,5
Илекский	417	239	991,3	330	13035	39,9
Кваркенский	310	208	887,7	180	9783	39,1
Красногвардейский	319	221	777,9	194	10940	38,9
Кувандыкский	303	194	872,9	367	10052	40,1
Курманаевский	240	214	1027,6	217	9368	42,1
Матвеевский	180	110	789,5	198	6590	42
Новоорский	455	261	811,2	285	15724	38,4

Классическая линейная модель множественной регрессии

$$\tilde{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k \quad (1)$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_k x_{ik} + z_i, i = 1..n \quad (2)$$

$$Y = X\beta + z \quad (3)$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix} \quad z_i = y_i - \tilde{y}_i, i = 1..n \quad \begin{matrix} n = 41 \\ k = 5 \end{matrix} \quad (4)$$

Условия Гаусса-Маркова

1) x_1, \dots, x_k – детерминированные (неслучайные) величины

2) ранг матрицы X равен $k+1$

3) $M\varepsilon_i = 0 \quad i = 1, n$

4) $D\varepsilon_i = M\varepsilon_i^2 = \sigma^2$

5) $\text{cov}(\varepsilon_i, \varepsilon_j) = M(\varepsilon_i \cdot \varepsilon_j) = 0 \quad i \neq j \quad i = \overline{1, n} \quad j = \overline{1, n}$

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

Оценка коэффициентов ЛММР методом наименьших квадратов

$$\sum_{i=1}^n z_i^2 \rightarrow \min \quad (6)$$

$$G(\beta_0, \dots, \beta_k) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik})^2 = (Y - X\beta)^T (Y - X\beta) = \quad (7)$$
$$= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta = Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta \rightarrow \min$$

$$X^T X\beta = X^T Y \quad (8)$$

$$\hat{\beta}_{\text{МНК}} \equiv \hat{\beta} = (X^T X)^{-1} X^T Y \quad (9)$$

$$\hat{y} = 696.76 + 0.88x_1 - 0.13x_2 - 0.01x_3 + 0.009x_4 - 13.63x_5 \quad (10)$$

(0.15) (0.05) (0.03) (0.003) (2.94)

$$\hat{R}^2 = 0.99$$

Исследование закона распределения регрессионных ОСТАТКОВ

$$H_0 : P_\varepsilon(x) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{(x-m_\varepsilon)^2}{2\sigma_\varepsilon^2}}$$

$$H_1 : P_\varepsilon(x) \neq \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{(x-m_\varepsilon)^2}{2\sigma_\varepsilon^2}}$$

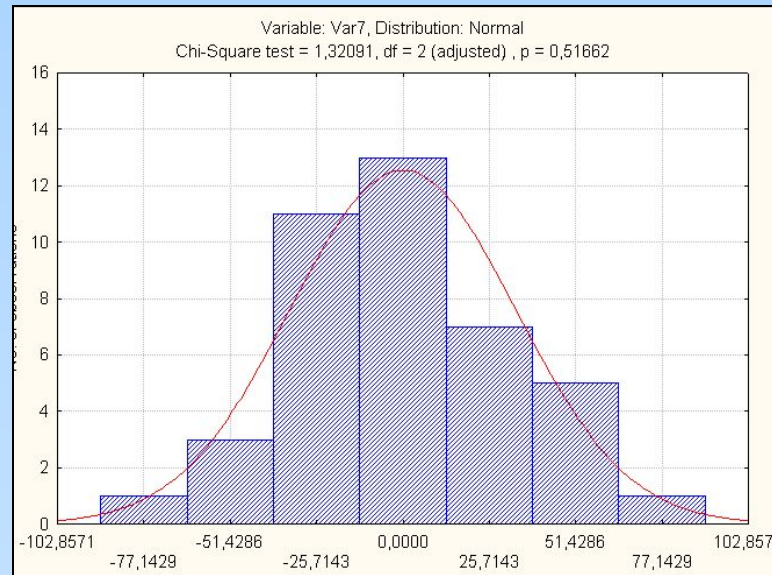


Рисунок 1 – Гистограмма частот и проверка гипотезы о характере закона распределения регрессионных остатков модели

$$p > \alpha \quad (0.52 > 0.05)$$

Проверка выполнения третьего условия Гаусса-Маркова

$$\begin{aligned}H_0 : M\varepsilon &= 0 \\H_1 : M\varepsilon &\neq 0\end{aligned}\tag{11}$$

$$t = \frac{\bar{e} - M\varepsilon}{\hat{\sigma}_{ост}} \sqrt{n} \in t(n-1)\tag{12}$$

$$t_{набл} = \frac{\bar{e} - M\varepsilon}{\hat{\sigma}_{ост}} \sqrt{n} = \frac{0.000006 - 0}{35.83} \cdot \sqrt{41} = 0,000001$$

$$t_{крлев} = u_{\frac{\alpha}{2}} = -2,02 \qquad t_{крпр} = u_{1-\frac{\alpha}{2}} = 2,02$$

Проверка значимости ЛММР

$$\begin{aligned}H_0 : \beta_1 = \beta_2 = \dots = \beta_5 = 0 \\ H_1 : \exists j \in (1..5) : \beta_j \neq 0\end{aligned}\tag{13}$$

$$F = \frac{\hat{R}^2 / k}{(1 - \hat{R}^2) / n - k - 1} = \frac{Q_{\text{факт}} / k}{Q_{\text{ост}} / n - k - 1} \in F(k, n - k - 1)\tag{14}$$

$$F_{\text{набл}} = 8291,5$$

$$F_{\text{кр}} = u_{0,95} = 2.49$$

$$F_{\text{набл}} > F_{\text{кр}} (8291,5 > 2.49)$$

Проверка значимости и построение доверительных интервалов для коэффициентов модели

$$H_0 : \beta_j = 0, j = 0..k \quad (15)$$

$$t = \frac{\hat{\beta}_j}{S_{\hat{\beta}_j}} \in t(n-k-1) \quad (16)$$

$$t_{0набл} = 5,16; t_{1набл} = 5,81; t_{2набл} = -2,7; t_{3набл} = -0,33; t_{4набл} = 3,41; t_{5набл} = -4,63$$

$$t_{крлев} = u_{\frac{\alpha}{2}} = -2,03 \quad t_{крпр} = u_{1-\frac{\alpha}{2}} = 2,03$$

$$t = \frac{\hat{\beta}_j(\eta_{1,n}) - \beta_j}{S_{\hat{\beta}_j}(\eta_{1,n})} \in t(n-k-1) \quad (17)$$

$$\hat{\beta}_j - \delta_{\gamma} S_{\hat{\beta}_j} < \beta_j < \hat{\beta}_j + \delta_{\gamma} S_{\hat{\beta}_j} \quad (18)$$

$$0.53 < \beta_1 < 1.23 \quad -0.17 < \beta_2 < -0.09 \quad -0.081 < \beta_3 < 0.061$$

$$0.002 < \beta_4 < 0.02 \quad -20.5 < \beta_5 < -6.7$$

Исследование линейной модели множественной регрессии на наличие мультиколлинеарности объясняющих переменных

Variable	Correlations (Spreadsheet13P)				
	x1	x2	x3	x4	x5
x1	1,000000	-0,102784	0,800381	0,998832	-0,231068
x2	-0,102784	1,000000	-0,149014	-0,105023	-0,175310
x3	0,800381	-0,149014	1,000000	0,807668	-0,255295
x4	0,998832	-0,105023	0,807668	1,000000	-0,222149
x5	-0,231068	-0,175310	-0,255295	-0,222149	1,000000

Рисунок 2 - Оценка матрицы парных коэффициентов корреляции

$$\hat{R}_{x_1/x_2, x_3, x_4, x_5}^2 = 0.99$$

$$\hat{R}_{x_3/x_1, x_2, x_4, x_5}^2 = 0.69$$

$$\hat{R}_{x_4/x_1, x_2, x_3, x_5}^2 = 0.99$$

$$\text{cond}(X^T X) = \frac{\max |\lambda_j|}{\min |\lambda_j|} = 1,686 \times 10^{12} \quad (19)$$

Устранение мультиколлинеарности между объясняющими переменными линейной модели множественной регрессии

$$\hat{y} = 683.57 + 0.9x_1 - 0.13x_2 + 0.009x_4 - 13.43x_5 \quad \hat{R}^2 = 0.99 \quad (20)$$

(127.58) (0.15) (0.05) (0.002) (2.84)

$$(X^T X + \tau E)\beta = X^T Y \quad (21)$$

$$\hat{\beta}_\tau = (X^T X + \tau E)^{-1} X^T Y$$

$$\tau = 0.3 \quad \hat{y} = 802.17 + 0.55x_1 - 0.13x_2 + 0.54x_3 + 0.01x_4 - 17.77x_5 \quad \hat{R}^2 = 0.88 \quad (22)$$

(1281.92) (0.12) (0.48) (0.26) (0.002) (25.69)

$$\hat{y} = 686.99 + 0.89x_1 - 0.13x_2 - 0.01x_3 + 0.009x_4 - 13.43x_5 \quad (23)$$

(134.12) (0.15) (0.05) (0.03) (0.009) (2.93)

Устранение мультиколлинеарности путем перехода к ортогональным переменным

Eigenvalues (Spreadsheet13P) Extraction: Principal components				
Value	Eigenvalue	% Total variance	Cumulative Eigenvalue	Cumulative %
1	2,846511	56,93021	2,846511	56,9302
2	1,165697	23,31395	4,012208	80,2442
3	0,736061	14,72122	4,748269	94,9654
4	0,250699	5,01397	4,998967	99,9793
5	0,001033	0,02065	5,000000	100,0000

Рисунок 3 - Вклады главных компонент в суммарную дисперсию исходных признаков

Factor Loadings (Varim) Extraction: Principal co (Marked loadings are >		
Variable	Factor 1	Factor 2
x1	0,970103	-0,015575
x2	-0,191824	-0,816096
x3	0,906583	0,001527
x4	0,971817	-0,008459
x5	-0,312640	0,709872

Рисунок 4 – Матрица нагрузок

$$z_1 = 0.58x_1^* - 0.15x_2^* + 0.54x_3^* + 0.58x_4^* - 0.15x_5^* \quad (24)$$

$$z_2 = 0.01x_1^* - 0.761x_2^* + 0.027x_3^* + 0.02x_4^* + 0.65x_5^*$$

$$\hat{y} = 622.1 + 1123.4f_1 \quad \hat{R}^2 = 0.95 \quad (25)$$

(41.78) (42.3)

$$\hat{y} = 2380.38 + 0.479x_1 - 0.819x_2 + 1.227x_3 + 0.008x_4 - 48.049x_5 \quad (26)$$

Обобщенная линейная модель множественной регрессии

$$Y = X\beta + z \quad (27)$$

1) x_1, \dots, x_k – детерминированные переменные

2) ранг матрицы X равен $k+1$

$$3) M\varepsilon_i = 0 \quad i = \overline{1, n}$$

$$4) D\varepsilon_i = M\varepsilon_i^2 = \sigma_i^2 \quad i = \overline{1, n} \quad (28)$$

$$5) \text{cov}(\varepsilon_i, \varepsilon_j) = M(\varepsilon_i \cdot \varepsilon_j) = 0 \quad i \neq j \quad i = \overline{1, n} \quad j = \overline{1, n}$$

$$4') \Sigma_\varepsilon = M\varepsilon\varepsilon^T = \sigma_0^2 \Sigma_0 \quad (\Sigma_0 \neq E_n)$$

Обобщенный метод наименьших квадратов

$$\Sigma_0 = CC^T \quad (29)$$

(30)

(31)

(32)

(33)

$$\hat{S}_0^2 = \frac{1}{n-k-1} (Y - X\beta)^T \Sigma_0^{-1} (Y - X\beta) \quad (34)$$

$$\Sigma_{\hat{\beta}_{\text{ОМНК}}} = \hat{S}_0^2 (X^T \Sigma_0^{-1} X)^{-1}$$

Исследование гетероскедастичности регрессионных остатков ЛММР по переменной x_1

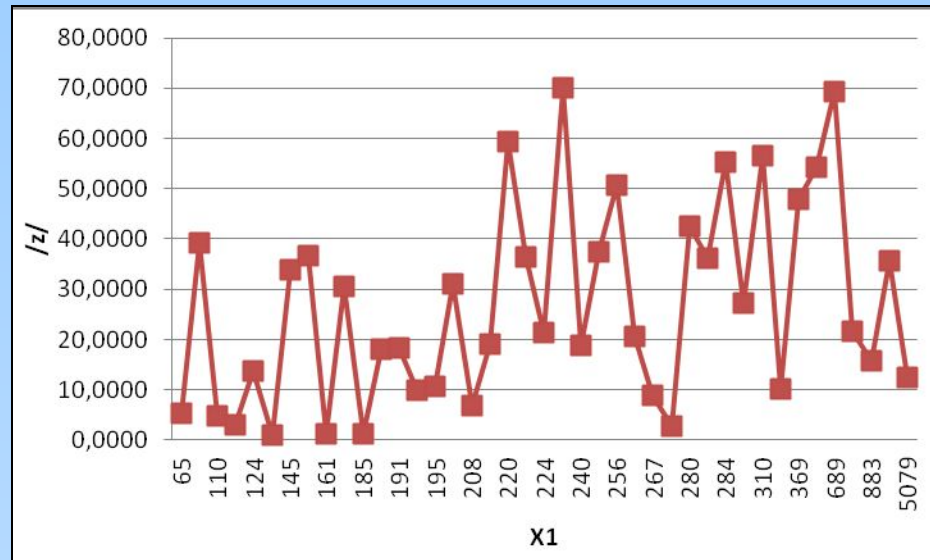


Рисунок 5 - График зависимости абсолютных значений остатков $|\hat{z}_i|$ и значений объясняющей переменной x_1

Тест ранговой корреляции Спирмена

$$H_0 : \rho_{x_i|\varepsilon} = 0 \quad H_1 : \rho_{x_i|\varepsilon} \neq 0 \quad (35)$$

$$t = r_{x_i|\varepsilon} \sqrt{n-1} \in N(0,1) \quad (36)$$

$$r_{x_i|\varepsilon} = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n^3 - n} \quad (37)$$

Исследование гетероскедастичности регрессионных остатков ЛММР по переменной x_1

Pair of Variables	Valid N	Spearman R	t(N-2)	p-level
x1 & Регр остатки	41	0,407491	2,786637	0,008184

Рисунок 6 - Результат расчета рангового коэффициента корреляции Спирмена

$$r_{x_1|z} = 0,41$$

$$p < \alpha \quad (0.008 < 0.05)$$

Тест Голдфелда-Квандта

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 \quad (38)$$

$$H_1 : \exists i, j : \sigma_i^2 \neq \sigma_j^2$$

$$F = \frac{\max\{Q_1; Q_2\} / (n' - k - 1)}{\min\{Q_1; Q_2\} / (n'' - k - 1)} \in F(n' - k - 1, n'' - k - 1) \quad (39)$$

$$n' = n'' = \frac{41 - 0.25 \cdot 41}{2} \approx 15 \quad F_{набл} = \frac{Q''}{Q'} = \frac{17340}{3852,91} = 4,5 \quad F_{крит} = 2.4$$

Тест Глейзера

$$|\hat{z}_i| = \alpha + \beta |x_{i1}|^\gamma + \delta_i \quad (40)$$

Таблица 2 - Результаты оценивания регрессионной модели

γ	α	S_α	β	S_β	\hat{R}^2	F
-2	32	3,7	-150600	73838,37	0,1	4,2
-1,5	34,2	4,3	-19538,3	8742,2	0,11	5
-1,2	36,5	5,1	-5620,4	2443,9	0,1194	5,3
-1,1	37,47	5,4	-3694,9	1603,35	0,1199	5,3
-1	38,53	5,817	-2424,33	1055,1	0,1192	5,3
-0,5	46,6	10,1	-296,9	146,2	0,0956	4,1

$$|\hat{z}_i| = 37,47 + \frac{3694,9}{(1603,4)} |x_{i1}|^{-1,1} \quad (41)$$

Исследование гетероскедастичности регрессионных остатков ЛММР по переменной x_2

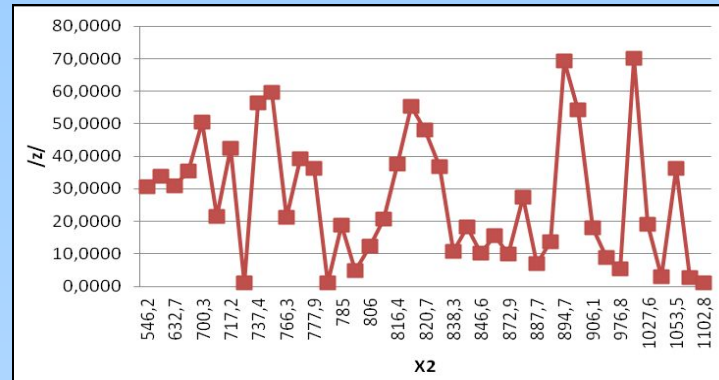


Рисунок 7 - График зависимости абсолютных значений остатков и значений объясняющей переменной x_2

Тест ранговой корреляции Спирмена

$$r_{x_2|z} = -0.29 \quad p > \alpha \quad (0.07 > 0.05)$$

Тест Голдфелда-Квандта

$$F_{набл} = \frac{Q'}{Q''} = \frac{15526}{11538} = 1.35 \quad F_{крит} = 2.4$$

Тест Глейзера

Таблица 3 - Результаты оценивания регрессионной модели

γ	α	S_α	β	S_β	\hat{R}^2	F
-1	1,63	20	20648,32	16127,7	0,0403	1,64
-0,5	-26,1	40,3	1519,4	1151,2	0,0428	1,7
0,5	83,84	41,2	-1,98	1,43	0,0468	1,92
1	56,1	20,9	-0,03	0,02	0,0485	1,99
1,5	46,74	14,2	-0,0008	0,00057	0,0499	2,05
2	42	10,82	-0,00002	0,0001	0,051	2,1

Исследование гетероскедастичности регрессионных остатков ЛММР по переменной x_4

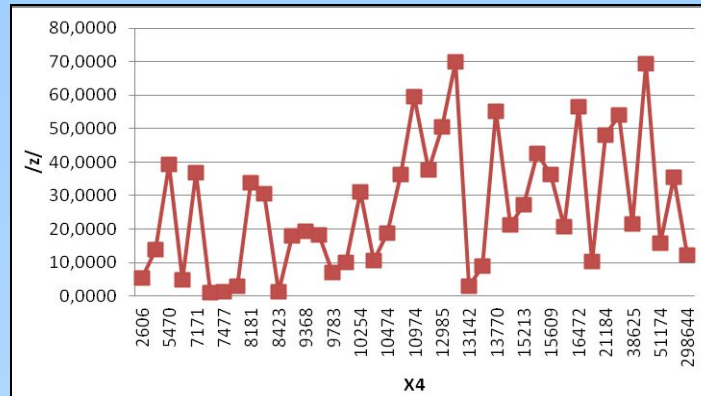


Рисунок 8 - График зависимости абсолютных значений остатков и значений объясняющей переменной x_4

Тест ранговой корреляции Спирмена

$$r_{x_2|z} = 0.42 \quad p < \alpha \quad (0.006 < 0.05)$$

Тест Голдфелда-Квандта

$$F_{набл} = \frac{Q''}{Q'} = \frac{16237}{3875.51} = 4,2 \quad F_{крит} = 2.4$$

Тест Глейзера

Таблица 4 - Результаты оценивания регрессионной модели

γ	α	S_α	β	S_β	\hat{R}^2	F
-2	30	3	-231731510	131010976	0,07	3,1
-1,5	32	4	-5155177	2485480	0,099	4,3
-1	37	5,26	-108096	46630,3	0,1211	5,4
-0,9	38,5	5,72	-49319,3	21121,7	0,12265	5,5
-0,8	40	6,3	-22419,7	9605,5	0,12257	5,45
-0,5	46,3	9,3	-2104,01	957,71	0,11	4,83

$$|\hat{z}_i| = 38.5 - 49319.3 |x_{i1}|^{-0.9} \quad (21121,67) \quad (42)$$

Исследование гетероскедастичности регрессионных остатков ЛММР по переменной x_5

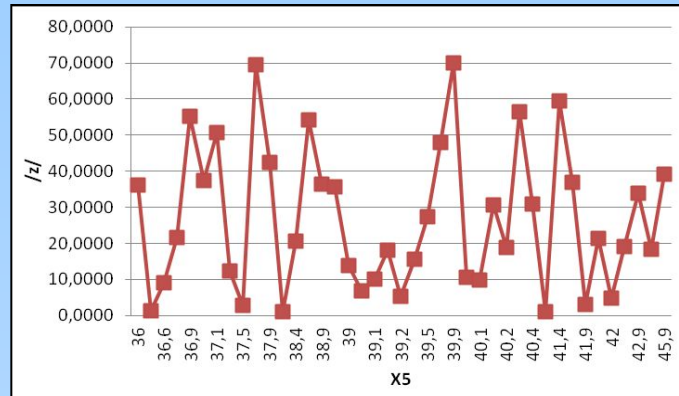


Рисунок 9 - График зависимости абсолютных значений остатков и значений объясняющей переменной x_5

Тест ранговой корреляции Спирмена

$$r_{x_2|z} = -0.02 \quad p > \alpha \quad (0.88 > 0.05)$$

Тест Голдфелда-Квандта

$$F_{набл} = \frac{Q'}{Q''} = \frac{11877}{9059,2} = 1,3 \quad F_{крит} = 2.4$$

Тест Глейзера

Таблица 5 - Результаты оценивания регрессионной модели

γ	α	S_α	β	S_β	\hat{R}^2	F
-2	32	3,7	-150600	73838,37	0,1	4,2
-1,5	34,2	4,3	-19538,3	8742,2	0,11	5
-1,2	36,5	5,1	-5620,4	2443,9	0,1194	5,3
-1,1	37,47	5,4	-3694,9	1603,35	0,1199	5,3
-1	38,53	5,817	-2424,33	1055,1	0,1192	5,3
-0,5	46,6	10,1	-296,9	146,2	0,0956	4,1

$$|\hat{z}_i| = 37,47 + \frac{3694,9}{(1603,4)} |x_{i1}|^{-1.1} \quad (43)$$

Коррекция стандартных ошибок в форме Уайта

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	683.5742	128.6024	5.315408	0.0000
X1	0.896402	0.172931	5.183581	0.0000
X2	-0.126769	0.043304	-2.927455	0.0059
X4	0.008534	0.002912	2.930430	0.0058
X5	-13.42705	2.915468	-4.605453	0.0000

R-squared	0.999154	Mean dependent var	632.1220
Adjusted R-squared	0.999060	S.D. dependent var	1154.019
S.E. of regression	35.38628	Akaike info criterion	10.08437
Sum squared resid	45078.80	Schwarz criterion	10.29335
Log likelihood	-201.7297	Hannan-Quinn criter.	10.16047
F-statistic	10626.46	Durbin-Watson stat	1.717461
Prob(F-statistic)	0.000000		

Рисунок 10 - Результаты коррекции стандартных ошибок коэффициентов регрессионной модели в форме Уайта

$$\hat{y} = 683.57 + 0.9x_1 - 0.13x_2 + 0.009x_4 - 13.43x_5 \quad \hat{R}^2 = 0.99 \quad (44)$$

$$(128.6) \quad (0.17) \quad (0.04) \quad (0.003) \quad (2.92)$$

Исследование линейной модели множественной регрессии на наличие автокорреляции регрессионных остатков

$$\begin{aligned} H_0 : \rho(\varepsilon_i, \varepsilon_{i-1}) &= 0 \\ H_1 : \rho(\varepsilon_i, \varepsilon_{i-1}) &\neq 0 \end{aligned} \quad (45)$$

Статистика Дарбина-Уотсона

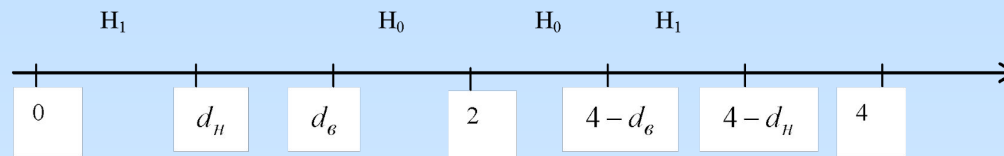
$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^{n-1} e_i^2} \quad (46)$$

$DW_{набл} \in (0, d_H)$ - принимается гипотеза о наличии положительной автокорреляции

$DW_{набл} \in (4 - d_H, 4)$ - принимается гипотеза о наличии отрицательной автокорреляции

$DW_{набл} \in (d_B, 4 - d_B)$ - принимается гипотеза об отсутствии автокорреляции

$DW_{набл} \in [d_H, d_B] \cup [4 - d_B, 4 - d_H]$ - область неопределенности критерия



Исследование ЛММР на наличие автокорреляции регрессионных остатков по переменным x_1 и x_2

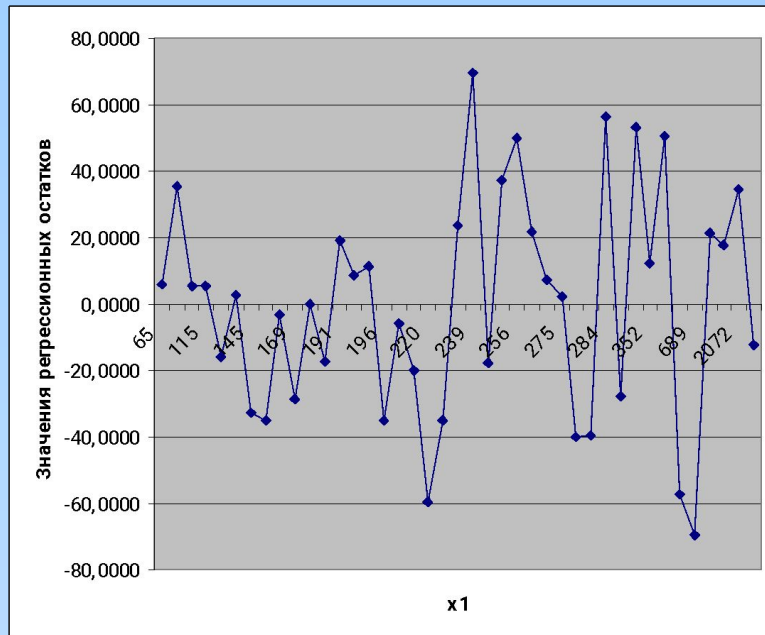


Рисунок 10 - График зависимости значений остатков и значений объясняющей переменной x_1

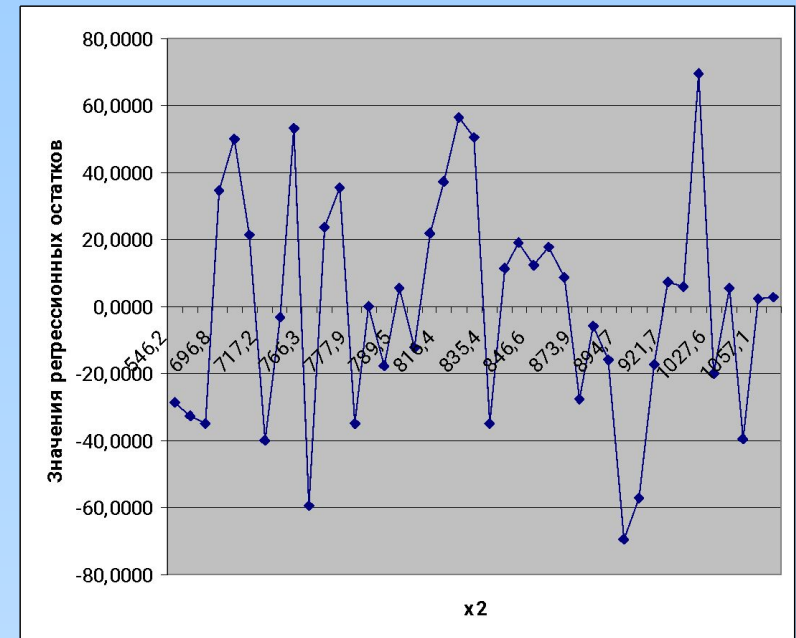


Рисунок 11 - График зависимости значений остатков и значений объясняющей переменной x_2

$$DW_{набл} = 1,743$$

$$d_H = 1,29$$

$$d_B = 1,72$$

$$DW_{набл} = 1,66$$

$$d_H = 1,29$$

$$d_B = 1,72$$

Исследование ЛММР на наличие автокорреляции регрессионных остатков по переменным x_4 и x_5

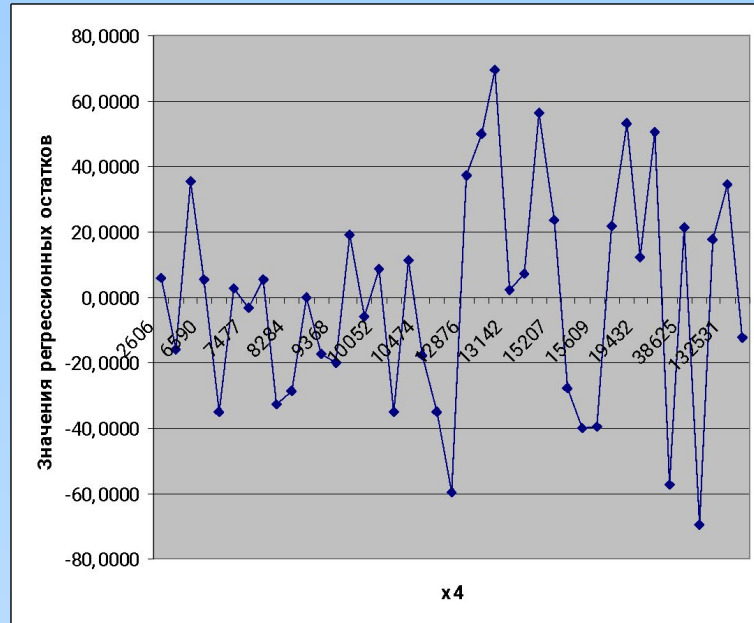


Рисунок 12 - График зависимости значений остатков и значений объясняющей переменной x_4

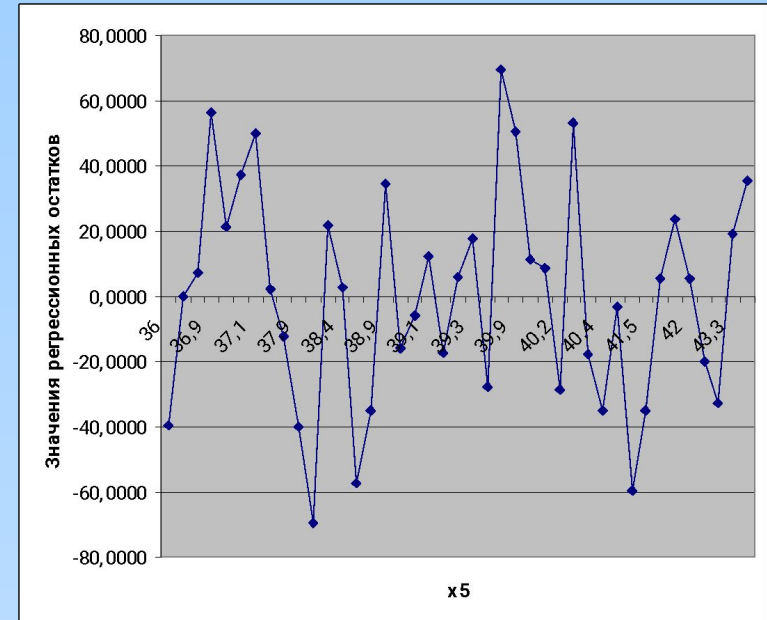


Рисунок 13 - График зависимости значений остатков и значений объясняющей переменной x_5

$$DW_{набл} = 1,73$$

$$d_H = 1,29$$

$$d_B = 1,72$$

$$DW_{набл} = 1,57$$

$$d_H = 1,29$$

$$d_B = 1,72$$

Построение точечной оценки прогнозируемого значения функции регрессии в точке

$$\hat{y}(X^{(n+1)}) = X^{(n+1)} \hat{\beta}_{МНК} \quad \hat{\beta}_{МНК} = \begin{pmatrix} 683.57 \\ 0.9 \\ -0.13 \\ 0.009 \\ -13.43 \end{pmatrix} \quad (47)$$

Таблица 6 - Результаты оценивания прогнозируемого значения регрессии в точке $X^{(n+1)}$

Номер сценария	$X^{(n+1)T}$	$\hat{y}(X^{(n+1)})$
1	$\begin{pmatrix} 1 \\ 65 \\ 1103 \\ 2606 \\ 45.9 \end{pmatrix}$	6
2	$\begin{pmatrix} 1 \\ 424 \\ 834 \\ 23964 \\ 39.5 \end{pmatrix}$	642
3	$\begin{pmatrix} 1 \\ 5079 \\ 546 \\ 298644 \\ 36 \end{pmatrix}$	7388

Построение доверительных интервалов для прогнозируемого значения и значения функции регрессии в точке

$$u(\eta_{1,n}) = \frac{\hat{y}(\eta_{1,n}) - y_{n+1}}{\hat{\sigma}_{np}} \in t(n - k - 1) \quad (48)$$

$$\hat{y}(X^{(n+1)}) - u_{\frac{1+\gamma}{2}} \cdot \hat{\sigma}_{np}(X^{(n+1)}) < y_{n+1} < \hat{y}(X^{(n+1)}) + u_{\frac{1+\gamma}{2}} \cdot \hat{\sigma}_{np}(X^{(n+1)}) \quad (49)$$

$$\hat{\sigma}_{np}(X^{(n+1)}) = \sqrt{\hat{\sigma}^2 (X^{(n+1)} (X^T X)^{-1} X^{(n+1)T} + 1)} \quad (50)$$

Таблица 7 – Доверительные интервалы для прогнозируемого значения $y_{n+1} = y(X^{(n+1)})$

Номер сценария	$\hat{\sigma}_{np}(X^{(n+1)})$	Доверительный интервал
1	42,83	$-81.166 < y_{n+1} < 92.560$
2	35,81	$569.308 < y_{n+1} < 714.574$
3	49,98	$7386.640 < y_{n+1} < 7489.357$

$$X^{(n+1)} \hat{\beta}_{МНК} - u_{\frac{1+\gamma}{2}} \cdot \hat{\sigma}_{\hat{f}}(X^{(n+1)}) < f(X^{(n+1)}) < X^{(n+1)} \hat{\beta}_{МНК} + u_{\frac{1+\gamma}{2}} \cdot \hat{\sigma}_{\hat{f}}(X^{(n+1)}) \quad (51)$$

$$\hat{\sigma}_{\hat{f}}(X^{(n+1)}) = \sqrt{\hat{\sigma}^2 \cdot X^{(n+1)} \cdot (X^T X)^{-1} \cdot X^{(n+1)T}} \quad (52)$$

Таблица 8 – Доверительные интервалы для прогнозируемого значения функции регрессии $f(X^{(n+1)})$

Номер сценария	$\hat{\sigma}_{\hat{f}}(X^{(n+1)})$	Доверительный интервал
1	24,13	$-43.25 < f(X^{(n+1)}) < 54.64$
2	5,53	$630.73 < f(X^{(n+1)}) < 653.15$
3	35,3	$7316.42 < f(X^{(n+1)}) < 7459.58$

Анализ дифференциации муниципальных образований Оренбургской области

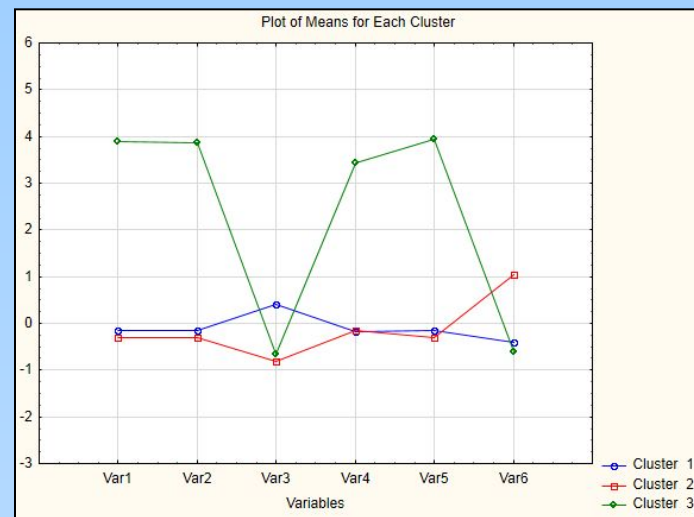
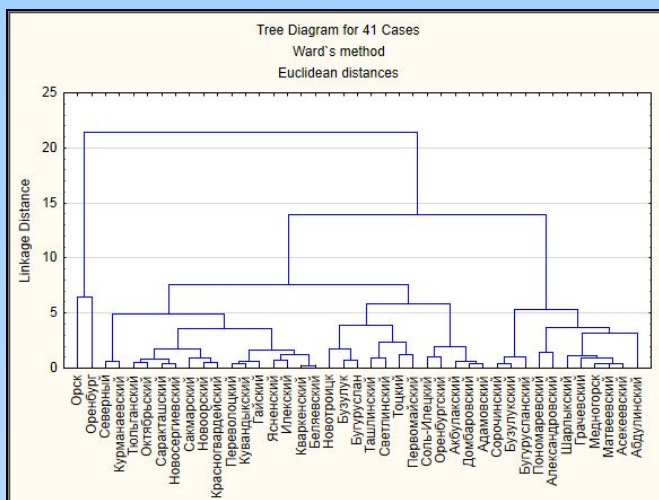


Рисунок 14 - Дендрограмма объединения классов методом Уорда

Рисунок 15 - График средних значений признаков в каждом кластере, полученном методом К-средних

Таблица 7 – Результаты классификации муниципальных образований методом К-средних

Номер кластера	Количество объектов в кластере	Состав класса
1	2	3
$\{S_{31}\}$	27	Города: г.Бузулук, г.Бугуруслан, г.Новотроицк. Районы: Адамовский, Акбулакский, Беляевский, Гайский, Домбаровский, Илекский, Кваркенский, Красногвардейский, Курманаевский, Новоорский, Новосергиевский, Переволоцкий, Сакмарский, Саракташский, Северный, Кувандыкский, Первомайский, Оренбургский, Тюльганский, Светлинский, Соль-Илецкий, Ташлинский, Тоцкий, Ясненский.
$\{S_{32}\}$	12	Города: г.Медногорск. Районы: Абдулинский, Александровский, Ассеевский, Бугурусланский, Бузулукский, Грачевский, Матвеевский, Октябрьский, Пономаревский, Сорочинский, Шарлыкский.
$\{S_{33}\}$	2	Города: г.Оренбург, г.Орск.

Регрессионный анализ рождаемости на основе модели с переменной структурой

Критерий Чоу

$$H_0 : \beta^{(1)} = \beta^{(2)}, \sigma_{\varepsilon(1)}^2 = \sigma_{\varepsilon(2)}^2 \quad (53)$$

$$\gamma = \frac{(e^T e - e^{(1)T} e^{(1)} - e^{(2)T} e^{(2)}) / (k+1)}{(e^{(1)T} e^{(1)} + e^{(2)T} e^{(2)}) / (n_1 + n_2 - 2k - 2)} \stackrel{H_0}{\in} F(k+1, n_1 + n_2 - 2k - 2) \quad (54)$$

$$\gamma_{n_1 n_2} = \frac{(e^T e - e^{(1)T} e^{(1)}) / n_2}{e^{(1)T} e^{(1)} / (n_1 - k - 1)} \stackrel{H_0}{\in} F(n_2, n_1 - k - 1) \quad (55)$$

$$\gamma_{набл} = \frac{(41988 - 25830 - 7750.2) / (4+1)}{\frac{25830 + 7750.2}{27 + 12 - 2 * 4 - 2}} = 1.452$$

$$\gamma_{крит} = 2,46$$

$$\gamma_{набл} = \frac{(45079 - 41988) / 2}{41988 / (39 - 4 - 1)} = \frac{1545.5}{1234.94} = 1.25$$

$$\gamma_{крит} = 3,28$$

$$d1 = \begin{cases} 1, & \text{объект принадлежит первому классу;} \\ 0, & \text{иначе.} \end{cases}$$

$$\hat{y} = 534 + 1,14x_1 - 0.15x_2 + 0.004x_4 - 9,71x_5 + 22,41d1_{i1}$$

(165.8) (0.22) (0.06) (0.0036) (3.98) (22.93)

Подход Бокса-Кокса

$$y_i'(\lambda) = \frac{y_i^\lambda - 1}{\lambda}, i = 1..n \quad (56)$$

$$x_{ij}'(\lambda) = \frac{x_{ij}^\lambda - 1}{\lambda}, i = 1..n; j = 1..k$$

$$y_i'(\lambda^*) = \beta_0 + \beta_1 x_{i1}'(\lambda^*) + \dots + \beta_k x_{ik}'(\lambda^*) + z_i, i = 1..n \quad (57)$$

$$y_i(\lambda^*) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + z_i, i = 1..n$$

$$\hat{\beta}_{MMIT} = (X^T X)^{-1} X^T Y'(\lambda)$$

$$\hat{\sigma}_{MMIT}^2 = \frac{1}{n} (Y'(\lambda) - X\beta)^T (Y'(\lambda) - X\beta)$$

$$l_Y(y_1, \dots, y_n) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (Y'(\lambda) - X\beta)^T (Y'(\lambda) - X\beta) + (\lambda - 1) \sum_{i=1}^n \ln y_i \quad (58)$$

$$Y'(\lambda) = \left(\frac{y_1^\lambda - 1}{\lambda}, \frac{y_2^\lambda - 1}{\lambda}, \dots, \frac{y_n^\lambda - 1}{\lambda} \right)^T$$

Моделирование числа родившихся с использованием подхода Бокса-Кокса

Таблица 9 – Значения функции максимального правдоподобия

λ	$l_T(y_1, \dots, y_n)$
-1	-161.427
-0,8	-157.222
-0.6	-153.497
-0.4	-149.912
-0.2	-146.642
0.2	-142.681
0.4	-141.733
0.5	-141.498
0.6	-141.425
0.7	-141.533
0.8	-141.875
0.9	-142.507
1	-143.488
1.2	-146.634
1.4	-151.337

$$\tilde{y}_i(0.6) = 92.04 + 0.78\tilde{x}_1(0.6) - 0.15\tilde{x}_2(0.6) + 0.04\tilde{x}_4(0.6) - 5.46\tilde{x}_5(0.6) + z_i \quad (59)$$

(17.69)
(0.17)
(0.06)
(0.01)
(1.13)

$$\hat{y} = (61.014 + 0.78x_1^{0.6} - 0.15x_2^{0.6} + 0.04x_4^{0.6} - 5.46x_5^{0.6})^{1.667} \quad (60)$$

$$\hat{R}^2 = 0.99$$

Выводы

1. Для описания зависимости уровня рождаемости населения от выбранных факторов рассмотрена линейная регрессионная модель. Получено значимое уравнение регрессии, значимое влияние на число родившихся оказывают число браков, заболеваемость на 1000 человек населения, численность женщин, средний возраст. Причем при увеличении числа браков и численности женщин число родившихся будет увеличиваться, а при увеличении заболеваемости на 1000 человек населения и среднего возраста – уменьшаться, что не противоречит здравому смыслу.

2. Анализ внешних и формальных признаков показал наличие мультиколлинеарности между объясняющими переменными. Вследствие этого реализовано несколько методов устранения мультиколлинеарности: метод пошагового включения, метод главных компонент, метод «ридж-регрессии» и рекуррентный метод наименьших квадратов. По экономическим и статистическим соображениям наилучшим было признано уравнение регрессии, построенное методом пошаговой регрессии.

3. Проверка наличия/отсутствия гетероскедастичности по каждой из объясняющих переменных показала, что гетероскедастичность наблюдается более, чем по одной переменной, поэтому стандартные ошибки коэффициентов модели были уточнены с помощью оценок в форме Уайта.

4. С использованием критерия Дарбина-Уотсона была проведена проверка наличия/отсутствия автокорреляции по каждой из объясняющих переменных, которая показала ее отсутствие.

Выводы

3. Осуществлено сценарного прогнозирования числа родившихся. Были рассмотрены 3 сценария, характеризующиеся «наихудшими», средними и «наилучшими» значениями объясняющих переменных, для каждого из которых были построены точечные оценки прогнозируемого значения функции регрессии в точке, а также доверительные интервалы для прогнозируемого значения и значения функции регрессии в точке.

4. Для выявления «схожих» по социально-экономическим показателям муниципальных образований методами кластерного анализа была проведена их классификация по количественным признакам. В результате все города и районы Оренбургской области были разбиты на три класса. Наиболее благополучная ситуация, по сравнению с остальными, наблюдается в том классе, куда входят города Оренбург и Орск. Для ранжирования муниципальных образований области был построен интегральный показатель. Результаты ранжирования городов и районов Оренбургской области в целом согласуются с интерпретацией классов.

5. Результаты проверки гипотезы о регрессионной однородности с помощью критерия Чоу показали, что выборочная совокупность является однородной в регрессионном смысле.

6. С использованием преобразования Бокса-Кокса была построена нелинейная зависимость. Для сравнения полученных оценок уравнений регрессии были найдены модельные значения, различия в которых оказались незначительными. Принимая во внимание значение остаточной дисперсии, характеризующей качество модели, было решено отдать предпочтение линейной модели множественной регрессии, на основании которой и осуществлялся ранее сценарный прогноз.