

Миргасимов Дамир Рустамович, студент,
Финансовый Университет при Правительстве РФ, г. Москва, Россия
e - mail: dammirg97@gmail.com

АНАЛИЗ ЗАВИСИМОСТИ РОЖДАЕМОСТИ ОТ УРОВНЯ БЕДНОСТИ И СРЕДНЕДУШЕВОГО ДОХОДА

Аннотация: Проведен анализ зависимости рождаемости от уровня бедности и среднедушевого дохода. Построена регрессионная модель и проведена оценка модели на наличие автокорреляции и гетероскедастичности.

Ключевые слова: линейная множественная регрессия, матрица парных коэффициентов корреляции, коэффициент детерминации, тест Гольфреда-Кванта, тест Дарбина-Уотсона.

Annotation: Analysis of the dependence of the birth rate from the level of poverty and per capita income. The author builds a regression model and the estimation models for the presence of autocorrelation and heteroscedasticity.

Keywords: multiple linear regressions, matrix of correlation coefficients, the coefficient of determination, Halfred-Quantum test, Durbin-Watson test.

Проведем анализ зависимости рождаемости от уровня бедности и среднедушевого дохода. В качестве показателей будем использовать коэффициент рождаемости (Y), численность населения с доходами ниже прожиточного минимума (X1) и среднедушевой доход (X2).

Таблица 1. Исходные данные [3].

| год | Крожд | Численность населения с доходами ниже прожиточного минимума, % | Среднедушевой доход, тыс. руб |
|------|-------|----------------------------------------------------------------|-------------------------------|
| | Y | X1 | X2 |
| 1999 | 1,157 | 28,4 | 1,66 |
| 2000 | 1,195 | 29 | 2,28 |
| 2001 | 1,223 | 27,5 | 3,06 |
| 2002 | 1,286 | 24,6 | 3,95 |
| 2003 | 1,319 | 20,3 | 5,17 |
| 2004 | 1,344 | 17,6 | 6,40 |
| 2005 | 1,294 | 17,8 | 8,09 |
| 2006 | 1,305 | 15,2 | 10,15 |
| 2007 | 1,416 | 13,3 | 12,54 |
| 2008 | 1,502 | 13,4 | 14,86 |
| 2009 | 1,542 | 13 | 16,90 |
| 2010 | 1,567 | 12,5 | 18,96 |
| 2011 | 1,582 | 12,7 | 20,78 |
| 2012 | 1,691 | 10,7 | 23,22 |
| 2013 | 1,707 | 10,8 | 25,93 |
| 2014 | 1,750 | 11,2 | 27,77 |
| 2015 | 1,777 | 13,3 | 30,47 |
| 2016 | 1,762 | 13,5 | 30,7384 |

Построим графики зависимостей Y от X1, Y от X2.

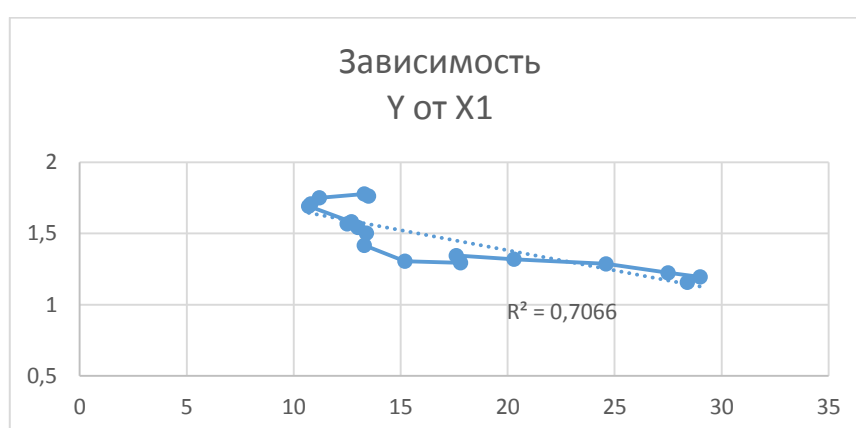


Рисунок 1. График зависимости Y от X1

По графику можно сделать вывод что связь между численностью населения с доходами ниже прожиточного минимума (X1) и коэффициентом рождаемости (Y) линейная. Коэффициент детерминации, изображенный на графике

показывает, что на 70,7% вариация рождаемости зависит от вариации численности населения с доходами ниже прожиточного минимума.

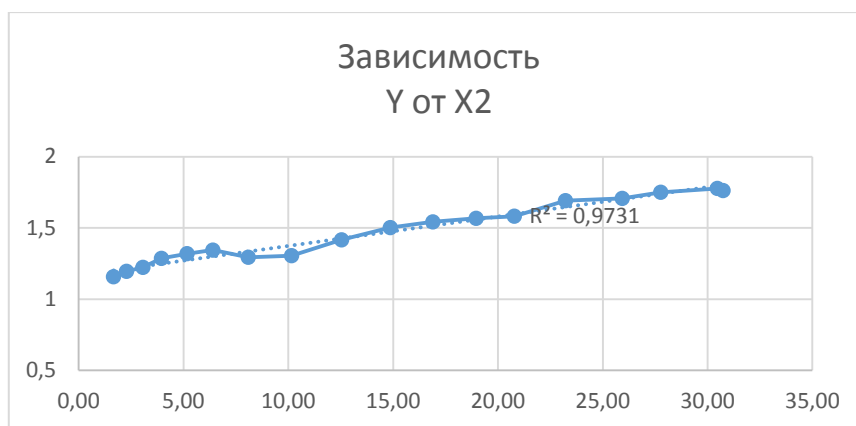


Рисунок 2. График зависимости Y от X2

По графику можно сделать вывод что связь между среднедушевым доходом (X2) и коэффициентом рождаемости (Y) линейная. Коэффициент детерминации, изображенный на графике показывает, что на 97,3% вариация рождаемости зависит от вариации среднедушевого дохода.

Чтобы проверить связь между всеми факторами, построим матрицу парных коэффициентов корреляции. Будем использовать команду КОРРЕЛЯЦИЯ из встроенного в Excel пакета АНАЛИЗ ДАННЫХ.

Таблица 2 – Матрица парных коэффициентов

| | y | x1 | x2 |
|----|--------|--------|----|
| y | 1 | | |
| x1 | -0,841 | 1 | |
| x2 | 0,986 | -0,826 | 1 |

Связь сильнее между коэффициентом рождаемости и среднедушевым доходом, корреляция равна 0,986. К тому же между факторными признаками присутствует мультиколлинеарность. Корреляционная связь между факторными переменными X1 и X2 сильная (-0,826). Это говорит о сильной взаимосвязи между численностью населения с доходами ниже прожиточного минимума и среднедушевым доходом, что может сказаться на качестве постоянной модели.

Для расчета регрессии будем использовать команду РЕГРЕССИЯ из встроенного в Excel пакета АНАЛИЗ ДАННЫХ. Получим уравнение линейной множественной регрессии (см. Приложение1 таб. 1).

$$\hat{Y} = 1,23 - 0,003 * X_1 + 0,019 * X_2 \quad (1)$$

По данному уравнению можно сделать следующий вывод: при изменении числа населения с доходами ниже прожиточного минимума на 1% и неизменном уровне второго фактора, коэффициент рождаемости изменится на 0,003 единицы. При изменении среднедушевого дохода на 1 тыс. рублей и неизменном уровне второго фактора, коэффициент рождаемости изменится на 0,019 единицы.

Верификация уравнения регрессии проводится на основе результатов автоматизированного регрессионного анализа. Экспресс-верификация проводится по следующим показателям: « R^2 », «Значимость F», «P-значение» (по каждому параметру регрессии), а также по графикам подбора и остатков [1].

« R^2 » - показатель детерминации в нашем случае он равняется $R^2=0,975$ (см. Приложение 1 таб. 1). Это означает что на 97,5 % изменения уровня рождаемости зависит от изменения исследуемых факторов. Т.к. показатель близок к единице, можно говорить о высоком уровне описания уравнения в целом.

«Значимость F» - Расчетный коэффициент Фишера равен $F_{расч} = 294,79$ (см. Приложение1 таб. 1). Табличное значения критерия Фишера равно $F_{табл} = 3,55$.

$F_{расч} > F_{табл}$, следовательно, модель в целом признается значимой.

Достоверность по уровню значимости критерия Фишера (см. Приложение1 таб. 1 Значимость F) значительно меньше 0,05, значит модель значима.

По графикам остаткам формируются предварительные суждения о выполнимости предпосылок МНК.



Рисунок 3. График остатков для переменной X1

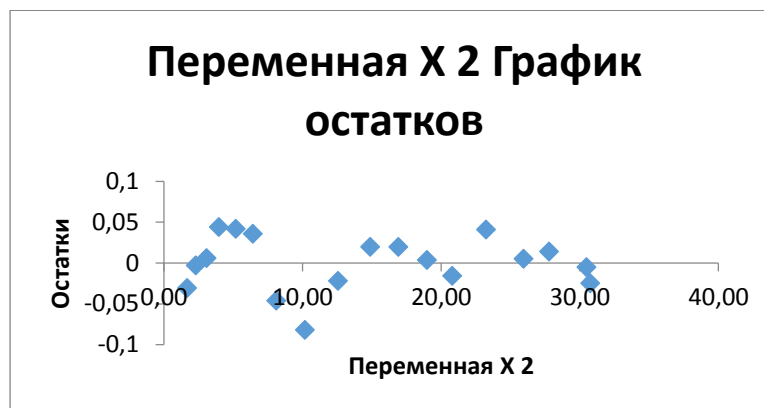


Рисунок 4. График остатков для переменной X2

Определенная связь между отклонениями отсутствует, это будет свидетельствует об отсутствии гетероскедастичности.

В целом можно сделать вывод, что построенная модель статистически значима и уровень описания уравнения в целом достаточно высок.

Рассчитаем критерий Дарбина-Уотсона для нашей модели.

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} = \frac{0,0215}{0,0207} = 1,039 \quad (2)$$

Автокорреляция отсутствует, если выполняется следующее условие: $d1 < DW$ и $d2 < DW < 4 - d2$. В нашем случае $d1=1,05$ и $d2=1,53$. В нашем случае условия отсутствия автокорреляции не выполняются: $1,05 > DW$ и $1,53 > DW < 2,47$. Можно сделать вывод, что присутствует положительная автокорреляция остатков.

В модели множественной регрессии выбирается переменная, от которой наиболее зависима результативная переменная. В нашем случае наибольшая взаимосвязь с коэффициентом рождаемости наблюдается у фактора X_2 – среднедушевой доход, т.к. коэффициент корреляции между ними равен 0,986 [2].

В данном случае предполагается, что стандартное отклонение $\sigma_i = \sigma(\varepsilon_i)$ пропорционально значению x_i переменной X в этом наблюдении, т.е. $\sigma^2_i = \sigma^2 x^2_i$, $i = 1, 2, \dots, n$.

Тест Голдфелда-Квандта состоит в следующем:

1. Все n наблюдений упорядочиваются по величине X .
2. Вся упорядоченная выборка после этого разбивается на три подвыборки размерностей k , $(n-2k)$, k .
3. Оцениваются отдельные регрессии для первой подвыборки (k первых наблюдений) и для третьей подвыборки (k последних наблюдений).
4. Для сравнения соответствующих дисперсий строится соответствующая F -статистика:

5. $F = S_3/S_1$

6. Построенная F -статистика имеет распределение Фишера с числом степеней свободы $\nu_1 = \nu_2 = (n - c - 2m)/2$.

7. Если $F > F_{кр}$, то гипотеза об отсутствии гетероскедастичности отклоняется.

Этот же тест может использоваться при предположении об обратной пропорциональности между σ_i и значениями объясняющей переменной. При этом статистика Фишера имеет вид: $F = S_1/S_3$.

Упорядочим все значения по величине X_2 . Находим размер подвыборки $k=(18-4)/2=7$. Где $c = 4n/15 = 4*18/15 = 4$.

Затем оценим регрессию для первой подвыборки. Регрессия для первой подвыборки будет иметь следующий вид:

$$Y = 1,15 + 0,025X \quad (3)$$

Найдем дисперсию уравнения регрессии первой подвыборки.

$$S_1 = \frac{\sum_{i=1}^n (Y - Y_{расч})^2}{n - m} = \frac{0,0089}{7 - 1} = 0,00148 \quad (4)$$

Оценим регрессию для третьей подвыборки. Регрессия третьей подвыборки имеет вид.

$$Y = 1,24 + 0,0022X \quad (5)$$

Найдем дисперсию уравнения регрессии третьей подвыборки.

$$S_3 = \frac{\sum_{i=1}^n (Y - Y_{\text{расч}})^2}{n - m} = \frac{0,0032}{7 - 1} = 0,00054 \quad (6)$$

Число степеней свободы $v_1 = v_2 = (n - c - 2m)/2 = (18 - 4 - 2 \cdot 1)/2 = 6$.
 $F_{\text{кр}}(6,6) = 5,99$. Строим обратную F-статистику:

$$F = 0,00148/0,00054 = 2,74 \quad (7)$$

Поскольку $F < F_{\text{кр}}$, то гипотеза об отсутствии гетероскедастичности принимается.

Подведем итог. При изучении зависимости рождаемости от уровня бедности и среднедушевого дохода мы сделали вывод о сильной взаимосвязи рождаемости с данными факторами. По построенному уравнению регрессии можно сделать вывод что, при изменении числа населения с доходами ниже прожиточного минимума на 1% и неизменном уровне второго фактора, коэффициент рождаемости измениться на 0,003 единицы. При изменении среднедушевого дохода на 1 тыс. рублей и неизменном уровне второго фактора, коэффициент рождаемости измениться на 0,019 единицы.

При этом построенная модель статистически значима и уровень описания уравнения в целом достаточно высок.

Однако, критерий Дарбина-Уотсона указывает на присутствие положительная автокорреляция остатков. Критерий Гольдфельда-Квандта указывает на отсутствие гетероскедастичности. Это означает, что наблюдения однородны, то есть дисперсии случайных ошибок модели постоянны.

Библиографический список:

1. Гладилин А.В. Эконометрика [Электронный ресурс]: учебное пособие для вузов [Гриф УМО] / А. В. Гладилин, А. Н. Герасимов, Е. И. Громов. - 3-е изд., стер. - Электрон. текстовые дан. - Москва : КноРус, 2014. - 227 с. (дата обращения 02.12.2017).

2. Новиков А.И. Эконометрика [Электронный ресурс]: учебное пособие для вузов / А. И. Новиков. - Электрон. текстовые дан. - Москва: Дашков и К°, 2013. - 223 с. - Режим доступа: <http://e.lanbook.com/view/book/5670/> (дата обращения 03.12.2017).

3. Федеральная служба государственной статистики [Электронный ресурс] URL: <http://www.gks.ru/> (дата обращения 01.12.2017).

Приложение 1.

Таблица 1. Расчетная таблица

| ВЫВОД ИТОГОВ | | | | | | | | |
|---------------------------------|---------------------|---------------------------|---------------------|-------------------|---------------------|--------------------|---------------------|----------------------|
| <i>Регрессионная статистика</i> | | | | | | | | |
| Множественный R | 0,987517 | | | | | | | |
| R-квадрат | 0,97519 | | | | | | | |
| Нормированный R-квадрат | 0,971881 | | | | | | | |
| Стандартная ошибка | 0,035437 | | | | | | | |
| Наблюдения | 18 | | | | | | | |
| Дисперсионный анализ | | | | | | | | |
| | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | <i>Значимость F</i> | | | |
| Регрессия | 2 | 0,740406 | 0,370203 | 294,7914 | 9,12E-13 | | | |
| Остаток | 15 | 0,018837 | 0,001256 | | | | | |
| Итого | 17 | 0,759244 | | | | | | |
| | <i>Коэффициенты</i> | <i>Стандартная ошибка</i> | <i>t-статистика</i> | <i>P-Значение</i> | <i>Нижние 95%</i> | <i>Верхние 95%</i> | <i>Нижние 95,0%</i> | <i>Верхние 95,0%</i> |
| Y-пересечение | 1,233632 | 0,060963 | 20,23587 | 2,66E-12 | 1,103693 | 1,363571 | 1,103693 | 1,363571 |
| Переменная X 1 | -0,00274 | 0,002417 | -1,13295 | 0,275021 | -0,00789 | 0,002413 | 0,007899 | 0,002413 |
| Переменная X 2 | 0,019201 | 0,001507 | 12,74255 | 1,89E-09 | 0,015989 | 0,022412 | 0,015989 | 0,022412 |