

ЭКОНОМИЧЕСКОГО
НАУЧНОГО ОБЩЕСТВА
СТУДЕНТОВ И АСПИРАНТОВ



№ 27

Эконометрические исследования

*OF STUDENTS' ECONOMIC
SCIENTIFIC SOCIETY*

Econometrics research

САНКТ-ПЕТЕРБУРГ
2010

ББК 65

В 38

Межвузовский студенческий научный журнал

Учредитель – Международный банковский институт (МБИ)

Учрежден 7 мая 2002 г.



Редакционная
коллегия:

Деревянко Ю. Д.	– <i>главный редактор журнала, ректор МБИ, член-корреспондент МАН ВШ, к.э.н.</i>
Изранцев В. В.	– <i>зам. главного редактора, д.т.н., профессор, уч. секретарь</i>
Кузютин Д.В.	– <i>ответственный редактор выпуска, к.ф.-м.н., доцент</i>
Захаров И. Н.	– <i>к.т.н., доцент</i>
Фаттахов В. В.	– <i>к.т.н., доцент</i>
Бургонова Г. Н.	– <i>к.э.н., профессор</i>
Высоцкий Ю. В.	– <i>д.филос.н., доцент</i>
Кузютин Д. В.	– <i>к.ф.-м.н., доцент</i>
Макарова Н. В.	– <i>д.пед.н., профессор</i>
Марков Я. Г.	– <i>к.т.н.</i>
Павлова И. П.	– <i>д.э.н., профессор</i>
Пивоваров С. Э.	– <i>д.э.н., профессор</i>
Погостинская Н. Н.	– <i>д.э.н., профессор</i>
Попова Е. М.	– <i>д.э.н., профессор</i>
Сутырин С. Ф.	– <i>д.э.н., профессор</i>
Утевский А. С.	– <i>д.э.н.</i>
Ястребова М.А.	– <i>к.ф.-м.н.</i>
Журавлев А.Е.	– <i>председатель Совета ЭНОС, студент МБИ</i>

Вестник экономического научного общества студентов и аспирантов. № 27 // Межвузовский студенческий научный журнал. Bulletin of students' economic scientific society. Эконометрические исследования. Econometrics research / Под редакцией Д.В. Кузютина и Я.Б. Панкратовой. – СПб.: Изд-во МБИ, 2010.– 142 с.

ISBN 978-5-903028-90-0

В настоящем выпуске Вестника ЭНОС представлены 9 научных работ студентов Международного банковского института, выполненных в рамках курсового исследования по эконометрике под научным руководством преподавателей кафедры математических методов исследования экономики МБИ.

Пять авторов награждены дипломами и два автора – грамотами VIII Санкт-Петербургского открытого конкурса им. профессора В.Н. Вениаминова на лучшую студенческую научную работу по экономике, управлению и информатике в экономической сфере, проведенного Комитетом по науке и высшей школе Санкт-Петербурга и Международной академией наук высшей школы в 2009 г. (Ю.В. Мережинская – диплом 2 степени, К.Я. Моисеев и Л.Б. Немчинова – диплом 3 степени, М.А. Сизова и А.А. Фокина – грамота).

СОДЕРЖАНИЕ

Голубкова Д.А. , научные руководители – к.ф.-м.н., доцент Евсеев Е.А., к.ф.-м.н., доцент Кузютин Д.В. Международный банковский институт. ВЛИЯНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ФАКТОРОВ НА ПРОЦЕНТ ИММИГРАНТОВ В НАСЕЛЕНИИ СТРАНЫ	5
Калитина Е.С. , научные руководители – к.ф.-м.н., доцент Кузютин Д.В., ст. преподаватель Панкратова Я.Б. Международный банковский институт. ФАКТОРЫ, ВЛИЯЮЩИЕ НА УРОВЕНЬ ЗАРАБОТНОЙ ПЛАТЫ НАСЕЛЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ	16
Ласыгина А.В. , научные руководители – к.ф.-м.н., доцент Кузютин Д.В., ст. преподаватель Панкратова Я.Б. Международный банковский институт. ФАКТОРЫ, ВЛИЯЮЩИЕ НА УРОВЕНЬ РОЖДАЕМОСТИ В РОССИЙСКОЙ ФЕДЕРАЦИИ	37
Мережинская Ю.В. , научные руководители – к.ф.-м.н., доцент Кузютин Д.В., ст. преподаватель Панкратова Я.Б. Международный банковский институт. ИССЛЕДОВАНИЕ ВЛИЯНИЯ РЯДА ЭКОНОМИЧЕСКИХ ФАКТОРОВ НА ВЕЛИЧИНУ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ НАСЕЛЕНИЯ В СТРАНАХ ЕС	49
Моисеев К.Я. , научные руководители – к.ф.-м.н., доцент Евсеев Е.А., к.ф.-м.н., доцент Кузютин Д.В. Международный банковский институт. ВЗАИМОСВЯЗЬ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ НАСЕЛЕНИЯ И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ СТРАН МИРА	67
Немчинова Л.П. , научные руководители – к.ф.-м.н., доцент Кузютин Д.В., ст. преподаватель Панкратова Я.Б. Международный банковский институт. ФАКТОРЫ, ВЛИЯЮЩИЕ НА УРОВЕНЬ ПРЕСТУПНОСТИ В РОССИЙСКОЙ ФЕДЕРАЦИИ	81
Сизова М.А. , научные руководители – к.ф.-м.н., доцент Евсеев Е.А., к.ф.-м.н., доцент Кузютин Д.В. Международный банковский институт. ВЛИЯНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ НА КОЛИЧЕСТВО БРАКОВ В РЕГИОНАХ РОССИИ	100

Тюляев О.Н. , научные руководители – к.ф.-м.н., доцент Кузютин Д.В., ст. преподаватель Панкратова Я.Б. Международный банковский институт. ИССЛЕДОВАНИЕ ФАКТОРОВ, ВЛИЯЮЩИХ НА ПЛОТНОСТЬ ЗАСЕЛЕНИЯ РЕГИОНОВ РОССИЙСКОЙ ФЕДЕРАЦИИ	111
Фокина А.А. , научные руководители – к.ф.-м.н., доцент Евсеев Е.А., к.ф.-м.н., доцент Кузютин Д.В. Международный банковский институт. ИССЛЕДОВАНИЕ ФАКТОРОВ, ВЛИЯЮЩИХ НА УРОВЕНЬ АБОРТОВ В РАЗЛИЧНЫХ СТРАНАХ МИРА	127

Голубкова Дарья Андреевна

Международный банковский институт

Кузютин Д.В., к.ф.-м.н., доцент; **Евсеев Е.А.**, к.ф.-м.н., доцент –
научные руководители

ВЛИЯНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ФАКТОРОВ НА ПРОЦЕНТ ИММИГРАНТОВ В НАСЕЛЕНИИ СТРАНЫ

Иммиграция населения – въезд населения одной страны в другую на временное или постоянное проживание, рассматриваемый по отношению к стране, куда въезжают мигранты.

В настоящее время количество людей, живущих за пределами своей исторической родины, достигло 190 млн человек. Что же побуждает людей покинуть свою родину и отправляться в чужие, незнакомые им страны? Данный вопрос я поставила ключевым в своей работе и попыталась выяснить, какие факторы влияют на миграцию населения.

1. Сбор данных и отбор факторов

Для проведения исследования были отобраны следующие факторы, влияющие на количество иммигрантов в различных странах:

- X_1 – ВВП на душу населения (GDP per capita, %);
- X_2 – инфраструктура (infrastructure, балл);
- X_3 – личный подоходный налог (Personal income tax, %);
- X_4 – уровень жизни (standard of living, балл).

В качестве результирующего признака были использованы данные по проценту иммигрантов в населении страны. В исследовании использовались страны-представители пяти частей света (Северная Америка, Южная Америка, Австралия и Океания, Европа, Азия), имеющие совершенно разные культурные и экономические положения.

Далее будет изучено влияние каждого из приведенных выше факторов на количество иммигрантов в странах по отдельности и влияние этих факторов в совокупности.

2. Исследование влияния отдельных факторов

2.1. Исследование влияния величины ВВП на душу населения на процент иммигрантов в населении страны

В статистической таблице (см. приложение) представлены исходные данные о проценте иммигрантов в населении страны и величине ВВП на душу населения в странах. Предположим, что с увеличением величины ВВП на душу населения количество иммигрантов в стране будет уменьшаться.

Для изучения влияния фактора X_1 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 1).

Зависимость процента иммигрантов от ВВП на душу населения

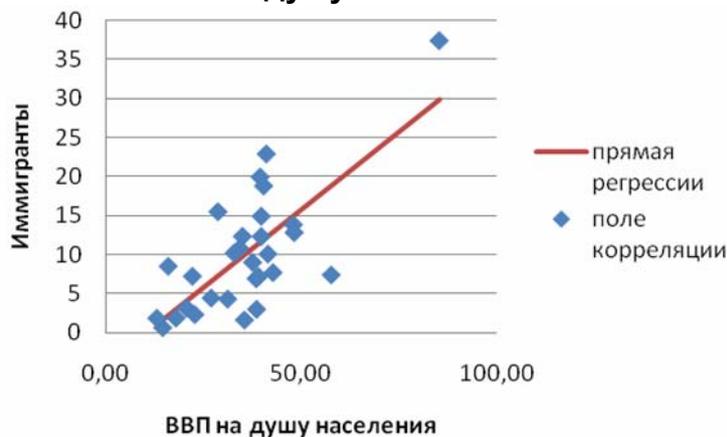


Рис. 1. Зависимость процента иммигрантов в населении страны от величины ВВП на душу населения

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную:

$$\hat{Y}_x = -4,16 + 0,39X_1.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации.

Коэффициент корреляции $r_{yx_1} = 0,733$. Согласно шкале Чеддока, такая связь изучаемых признаков является заметной.

Коэффициент детерминации $R^2 = 0,53$ говорит о том, что около 53 % вариации процента иммигрантов в населении страны объясняется вариацией величины ВВП на душу населения. Остальные 47 % вариации объясняются неучтенными в данной модели факторами. Скорректированный коэффициент детерминации $\hat{R}^2 = 0,52$. Средняя ошибка аппроксимации составила 74,05 %.

Данная модель является значимой по критерию Фишера, так как $F_{набл} > F_{табл}$, где $F_{набл} = 31,45$, а $F_{табл} = 4,21$.

Таким образом, данная математическая модель, выражающая зависимость объясняющей переменной, подходит для описания зависимой переменной. Поэтому включение данного фактора в модель множественной регрессии является целесообразным.

Для данной модели подходит следующая интерпретация: при увеличении размера ВВП на душу населения на 1 % процент иммигрантов в населении страны увеличится на 0,39 %.

2.2. Исследование влияния состояния инфраструктуры на процент иммигрантов в населении страны

В статистической таблице (см. приложение) представлены исходные данные о проценте иммигрантов в населении страны и состоянии инфраструктуры в различных странах. Предположим, что с улучшением состояния инфраструктуры количество иммигрантов в стране будет расти.

Для изучения влияния фактора X_2 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 2).

Зависимость процента иммигрантов от инфраструктуры

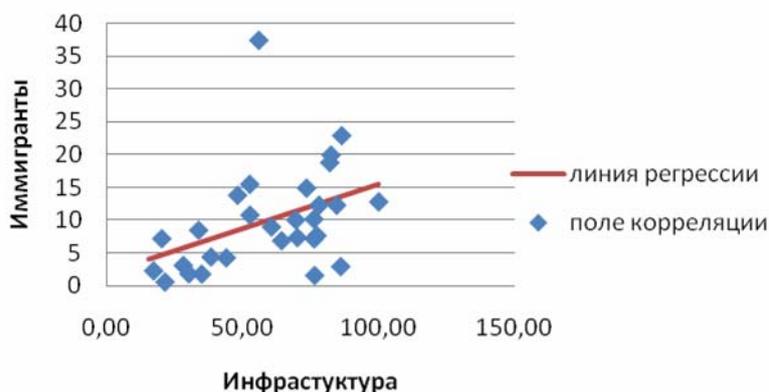


Рис. 2. Зависимость процента иммигрантов в населении страны от инфраструктуры

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную

$$\hat{y}_x = 1,87 + 0,13x_2.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации.

Коэффициент корреляции $r_{yx_2} = 0,40$, согласно шкале Чеддока, такая связь изучаемых признаков является заметной.

Коэффициент детерминации $R^2 = 0,16$ говорит о том, что около 16 % вариации процента иммигрантов в населении страны объясняется вариацией состояния инфраструктуры. Остальные 84 % вариации объясняются неучтенными в данной модели факторами. Скорректированный коэффициент детерминации $\bar{R}^2 = 0,13$. Средняя ошибка аппроксимации составила 110,37 %.

Данная модель является значимой по критерию Фишера, так как $F_{набл} > F_{табл}$, где $F_{набл} = 5,34$, а $F_{табл} = 4,21$.

Таким образом, данная математическая модель, выражающая зависимость объясняющей переменной, подходит для описания зависимой переменной. Поэтому включение данного фактора в модель множественной регрессии является целесообразным.

Для данной модели подходит следующая интерпретация: при улучшении состояния инфраструктуры на 1 балл процент иммигрантов в населении страны увеличится на 0,13 %.

2.3. Исследование влияния величины личного подоходного налога на процент иммигрантов в населении страны

В статистической таблице (см. приложение) представлены исходные данные о проценте иммигрантов в населении страны и величине личного подоходного налога в различных странах. Для изучения влияния фактора X_3 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 3). По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную:

$$\hat{y}_x = 6,33 + 0,13X_3.$$

Оценим тесноту связи с помощью показателя детерминации. Коэффициент детерминации $R^2 = 0,029$ говорит о том, что около 2,9 % вариации процента иммигрантов в населении страны объясняется вариацией величины личного подоходного налога. Остальные 97,1 % вариации объясняются неучтенными в данной модели факторами.

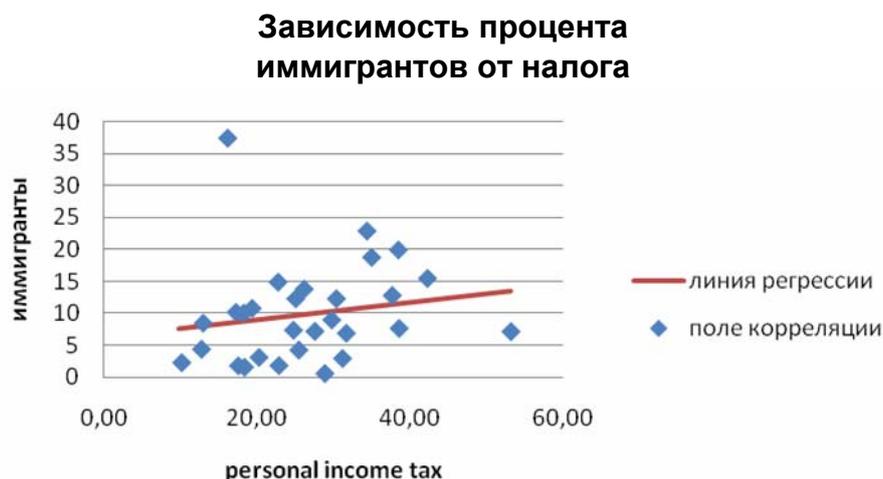


Рис. 3. Зависимость процента иммигрантов в населении страны от личного подоходного налога

Данная модель является незначимой по критерию Фишера, так как $F_{набл} < F_{табл}$, где $F_{набл} = 0,81$, а $F_{табл} = 4,21$.

Таким образом, данная математическая модель не подходит для описания зависимой переменной. Дальнейшее включение фактора X_3 в модель множественной регрессии является нецелесообразным.

2.4. Исследование влияния уровня жизни на процент иммигрантов в населении страны

В статистической таблице (см. приложение) представлены исходные данные о проценте иммигрантов в населении страны и уровне жизни в различных странах.

Для изучения влияния фактора X_4 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 4):

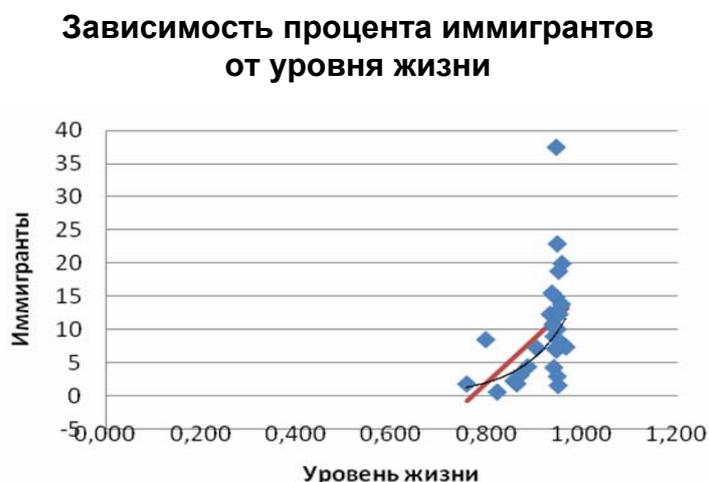


Рис. 4. Зависимость процента иммигрантов в населении страны от уровня жизни

При рассмотрении поля корреляции трудно точно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных зависимостей: линейной, квадратичной, степенной и экспоненциальной. Была изучена возможность существования каждой из этих видов зависимостей, получены следующие уравнения парных регрессий:

- линейная – $\hat{y}_x = -51,3 + 66,6 X_4$;
- квадратичная – $\hat{y}_x = 452,8 X_4^2 - 725,6 X_4 + 293,1$;
- степенная – $\hat{y}_x = 15,73 X_4^{9,209}$.

Для того чтобы осуществить выбор в пользу какой-либо из них, необходимо использовать следующие критерии:

- метод абсолютных отклонений. Лучшей из нескольких моделей является та, у которой показатель MAD наименьший;
 - средняя ошибка аппроксимации. Чем меньше эта ошибка, тем лучше построенная модель аппроксимирует наблюдаемые данные.
- Результаты представлены в табл. 1.

Таблица 1

Значение критериев отбора модели

Значения критериев отбора модели		
тип модели	MAD	ошибка аппр.
линейная	4,941515108	104,33 %
квадратичная	4,559576003	88,07 %
степенная	4,680166138	73,59 %

К сожалению, однозначно выбрать лучшую модель нельзя. По методу абсолютных отклонений наилучшей моделью является квадратичная, а по средней ошибке аппроксимации – степенная.

3. Исследование влияния совокупности факторов

В качестве факторов, оказывающих влияние на процент иммигрантов в населении страны, после предварительного исследования были отобраны три нижеследующих:

- ✓ x_1 – ВВП на душу населения (GDP per capita);
- ✓ x_2 – инфраструктура (infrastructure);
- ✓ x_4 – уровень жизни (standard of living).

Рассчитанные парные коэффициенты корреляции представим в виде корреляционной матрицы (табл. 2).

Корреляционная матрица

Q`	Y	X ₁	X ₂	X ₄
Y	1	0,733536	0,406650432	0,450805
X ₁	0,733535903	1	0,570720005	0,715053
X ₂	0,406650432	0,57072	1	0,728984
X ₄	0,450805128	0,715053	0,728984067	1

После анализа матрицы Q` становится очевидно, что наибольшее влияние на процент иммигрантов в населении страны оказывает величина ВВП на душу населения, обладающая самым большим коэффициентом корреляции.

Построим модель множественной линейной регрессии:

$$\hat{Y}_x = 19,85 + 0,45 \times X_1 + 0,02 \times X_2 - 29,94 \times X_4.$$

Коэффициент детерминации $R^2 = 0,552$ говорит о том, что 55,2 % вариации процента иммигрантов в населении страны объясняется вариацией совокупности признаков. Остальные 45,8 % вариации объясняются неучтенными в данной модели факторами. Скорректированный коэффициент детерминации $R^2_{adj} = 0,498$.

Данная модель является незначимой по критерию Фишера, т. к. $F_{набл} < F_{табл}$, где $F_{набл} = 0,78$, а $F_{табл} = 10,27$.

Так как фактор «коэффициент ВВП на душу населения» оказался наиболее значимым, было решено добавлять факторы после него. Для оценки оправданности включения фактора мы будем использовать скорректированный коэффициент детерминации множественных регрессий, который будем сравнивать со скорректированным коэффициентом регрессии Y на X₁, равным 0,520966585.

При включении второго фактора после первого скорректированный коэффициент детерминации по первому и второму факторам оказался меньше скорректированного коэффициента детерминации по одному фактору ($0,50272 < 0,520966585$). Включение второго фактора (инфраструктура) после первого оказалось неоправданно.

Третий фактор (личный подоходный налог) также был включен после первого фактора. Однако его включение оказалось неоправданным ($R^2_{adj} = 0,504588183 < R^2 = 0,520966585$).

Включение четвертого фактора (уровень жизни) после первого также оказалось неоправданным ($R^2_{adj} = 0,514515571 < R^2 = 0,520966585$).

Проведя проверку, было выявлено, что скорректированный коэффициент детерминации при добавлении факторов уменьшался, следовательно, включение этих факторов во множественную модель после первого фактора нецелесообразно. В трех моделях парной регрессии каждый фактор по отдельности был значим, а в множественной регрессии только первый фактор оказался значимым.

4. Проверка предпосылок регрессионного анализа

4.1. Проверка гомоскедастичности модели

Проверим условие гомоскедастичности в модели YX_1 , которая наиболее полно и четко описывает выборку.

По графику остатков (рис. 5) достаточно сложно определить наличие гетероскедастичности, необходимо произвести тесты Спирмена и Голдфелда-Квандта.

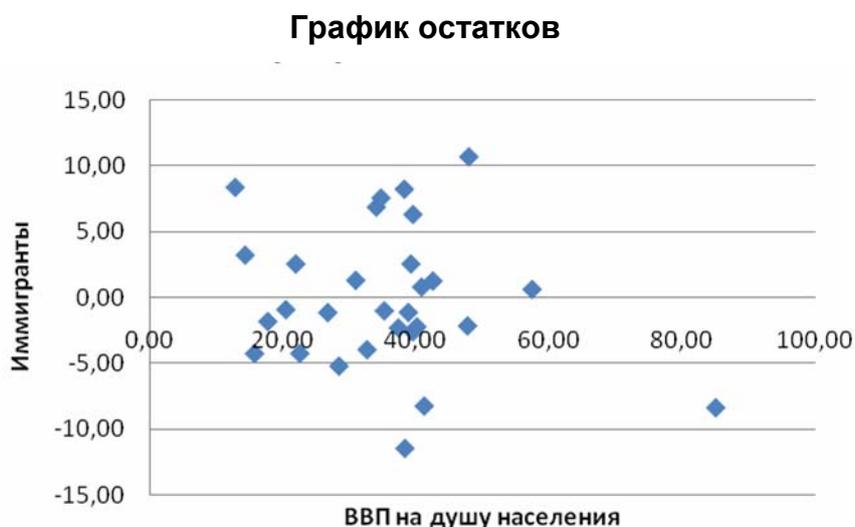


Рис. 5. График остатков

Проведем тест Голдфелда-Квандта и тем самым проверим, постоянен ли относительный разброс возмущений. Для этого упорядочим нашу выборку по возрастанию величины ВВП на душу населения, после чего разобьем ее на три части по 11, 7 и 11 элементов и построим регрессии по первой и последней группе. Проверим значение F-статистики: $F > F_{max}$, следовательно гипотеза об отсутствии гетероскедастичности отклоняется. По тесту Голдфелда-Квандта в данной модели гетероскедастичность присутствует.

Далее проведем тест Спирмена. Найдем разницу рангов объясняющей переменной и соответствующих остатков, после чего построим коэффициент ранговой корреляции. В нашем случае он равен 0,41601. Далее строим для него соответствующую статистику, равную в данном случае 2,377112, что превосходит критическое значение распределения Стьюдента для вероятности 0,05 и с 27 степенями свободы, равное 2,05183. Тем самым мы убеждаемся, что гетероскедастичность присутствует.

4.2. Автокорреляция

Как правило, если автокорреляция присутствует, то наибольшее влияние на наблюдение оказывает результат предыдущего наблюдения – так называемая автокорреляция первого порядка. Используем критерий Дарбина-Уотсона для обнаружения автокорреляции первого порядка.

$$d = 2,107411.$$

По таблице критических значений статистики Дарбина-Уотсона, при $n=29$, $k=1$: $d_l = 1,34$, $d_u = 1,48$, $4 - d_u = 2,52$.

Так как $d_u < d < 4 - d_u$: $1,48 < 2,107 < 2,52$, то гипотеза об отсутствии автокорреляции не отвергается, автокорреляции нет. Модель удовлетворяет данной предпосылке регрессионного анализа.

ЗАКЛЮЧЕНИЕ

В ходе подробного изучения и анализа факторов было выявлено, что наибольшее влияние на процент иммигрантов в населении страны оказывает величина ВВП на душу населения, что вполне естественно. Также сильно влияющими на процент иммигрантов в населении страны оказались индекс уровня жизни и развитость инфраструктуры. Совсем небольшое влияние имеет такой фактор, как личный подоходный налог.

При рассмотрении трех наиболее значимых факторов в совокупности выяснилось, что только влияние первого оказалось значимым. Коэффициент детерминации равен 0,561.

Подводя итог, надо отметить, что выборка в основном состояла из европейских стран, так как именно в этих странах тема миграции особо актуальна. Из 190 миллионов человек, иммигрировавших в чужие страны, 11,05 % (21 миллион) человек проживают на территории Евросоюза. На фоне резкого старения населения европейских стран миграция экономически активного контингента становится важной задачей союза. И правительства стран Евросоюза должны рассматривать иммигрантов как «обогащение» нации, а не как угрозу.

ПРИЛОЖЕНИЕ

Таблица

Данные

Страны	Immigrants as percentage of state population	GDP per capita	infrastructure	Personal income tax, %	standard of living
	Y	X ₁	X ₂	X ₃	X ₄
Австралия	19,93	39,30	82,47	38,50	0,957
Австрия	14,90	39,60	73,42	22,80	0,944
Бельгия	6,90	38,30	64,24	31,70	0,945
Великобритания	8,98	37,40	60,54	29,80	0,940
Венгрия	3,13	20,50	28,05	20,30	0,869
Германия	12,31	34,80	78,09	25,10	0,932
Дания	7,16	38,90	76,24	53,20	0,943
Ирландия	13,81	47,80	48,07	26,20	0,956
Исландия	7,67	42,60	77,25	38,60	0,960
Испания	10,79	34,10	52,61	19,40	0,938
Италия	4,28	31,00	43,90	25,50	0,940
Канада	18,78	40,20	81,99	35,00	0,950
Люксембург	37,42	85,10	55,88	16,20	0,945
Мексика	0,60	14,40	21,32	28,90	0,821
Нидерланды	10,05	41,30	69,47	18,30	0,947
Новая Зеландия	15,48	28,50	52,53	42,30	0,936
Норвегия	7,37	57,50	70,09	24,80	0,965
Польша	1,84	17,80	30,13	22,90	0,862
Португалия	7,20	22,00	20,14	27,60	0,904
Россия	8,48	15,80	33,73	13,00	0,797
Словакия	2,29	22,60	17,13	10,20	0,856
США	12,81	48,00	100,00	37,70	0,948
Турция	1,81	12,90	34,85	17,60	0,757
Финляндия	2,96	38,40	86,02	31,20	0,947
Франция	10,18	32,70	76,19	17,30	0,942
Чехия	4,41	26,80	38,28	12,80	0,885
Швейцария	22,89	40,90	86,40	34,40	0,947
Швеция	12,30	39,60	84,55	30,40	0,951
Япония	1,59	35,30	76,42	18,40	0,949

ЛИТЕРАТУРА

1. *Евсеев Е. А., Буре В. М.* Эконометрика. СПб.: Изд-во МБИ, 2007. – 139 с.
2. *Тарашина С. И., Панкратова Я. Б.* Выполнение курсовой работы по эконометрике: Учеб.-метод. пособие. СПб.: Изд-во МБИ, 2007. – 97 с.
3. *Кремер Н.Ш., Путко Б.А.* Эконометрика: Учебник для вузов / Под ред. проф. Н.Ш. Кремера. – М.: ЮНИТИ-ДАНА, 2003. – 311 с.
4. Практикум по эконометрике: Учеб. пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др. Под ред. И.И. Елисеевой. М.: Финансы и статистика, 2003. – 192 с.: ил.
5. Сборники государственной статистики Петростат [Электронный ресурс] <http://www.gks.ru>

Калитина Елизавета Сергеевна

Международный банковский институт

Кузютин Д.В., к.ф.-м.н., доцент, **Панкратова Я.Б.**, ст. преподаватель –
научные руководители

ФАКТОРЫ, ВЛИЯЮЩИЕ НА УРОВЕНЬ ЗАРАБОТНОЙ ПЛАТЫ НАСЕЛЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Изучение зависимости уровня доходов населения от различных факторов чрезвычайно важно в современном мире. Это необходимо как для людей, заинтересованных в увеличении своих доходов, так и для органов государственной власти для принятия эффективных управленческих решений.

Для студентов данная тема также представляет большой интерес, так как все мы после скорого окончания нашей учебы в вузе столкнемся с выбором места работы и нам всем необходимо знать, какие факторы в будущем могут повлиять на наш доход.

Таким образом, цель данной работы – выявление факторов, влияющих на уровень доходов населения Российской Федерации, отбор наиболее значимых из них и установление вида зависимости уровня доходов от этих факторов.

Сбор данных и отбор факторов

Для проведения исследования были отобраны следующие факторы, влияющие на уровень доходов населения:

- X_1 – валовой региональный продукт на душу населения за 2007 год в рублях. Данный показатель характеризует уровень экономического развития по каждому региону;

- X_2 – уровень экономически активного населения РФ в возрасте 15–72 лет за 2007 год в процентах. Экономически активное население включает в себя всех мужчин и женщин, которые обеспечивают или имеются в наличии для обеспечения предложения рабочей силы для производства товаров и услуг (ООН, 1996). В экономической литературе такое население определяется как часть населения, обеспечивающая предложение рабочей силы для производства товаров и услуг, включая занятых и безработных, или часть населения, занятая общественно полезной деятельностью, приносящей доход;

- X_3 – объем инвестиций в основной капитал за 2007 год в миллионах рублей. В зависимости от этого показателя меняется уровень заработ-

ной платы населения, потому как он влияет на финансовое состояние организации в целом, а значит, и на уровень заработной платы;

- X_4 – численность студентов высших учебных заведений по регионам РФ за 2007 год в тысячах человек. Этот показатель показывает примерный уровень грамотности каждого региона;

- X_5 – численность граждан, имеющих статус безработных, за 2007 год в тысячах человек. Этот показатель также может влиять на уровень доходов населения каждого региона.

В качестве результирующего признака (уровня доходов населения) были использованы данные по среднемесячной начисленной заработной плате (без выплат социального характера) по регионам РФ за 2007 год в рублях. Для исследования использовались данные по регионам Российской Федерации, отражающие вариацию факторов и результирующего признака. Были задействованы данные по всем регионам и городам РФ.

В исследовании использовались данные Федеральной службы государственной статистики (Росстат) за 2007 г. по регионам Российской Федерации, опубликованные на сайте Федеральной службы государственной статистики. Далее будет изучено влияние каждого из приведенных выше факторов на уровень доходов в отдельности и влияние всех этих факторов в совокупности.

Исследование влияния отдельных факторов

Исследование влияния ВРП на душу населения на уровень среднемесячной начисленной заработной платы

Представим исходные данные о среднемесячной начисленной заработной плате и объеме ВРП на душу населения по регионам за 2007 год в виде статистической таблицы, удобной для анализа (см. приложение).

Для изучения влияния фактора X_1 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 1). При его рассмотрении можно предположить, что наиболее подходящей будет линейная модель, так как она проста и визуально довольно точно отражает зависимость Y от X_1 . Построим линейное уравнение парной регрессии:

$$\hat{Y} = 6799,80 + 0,033X_1.$$

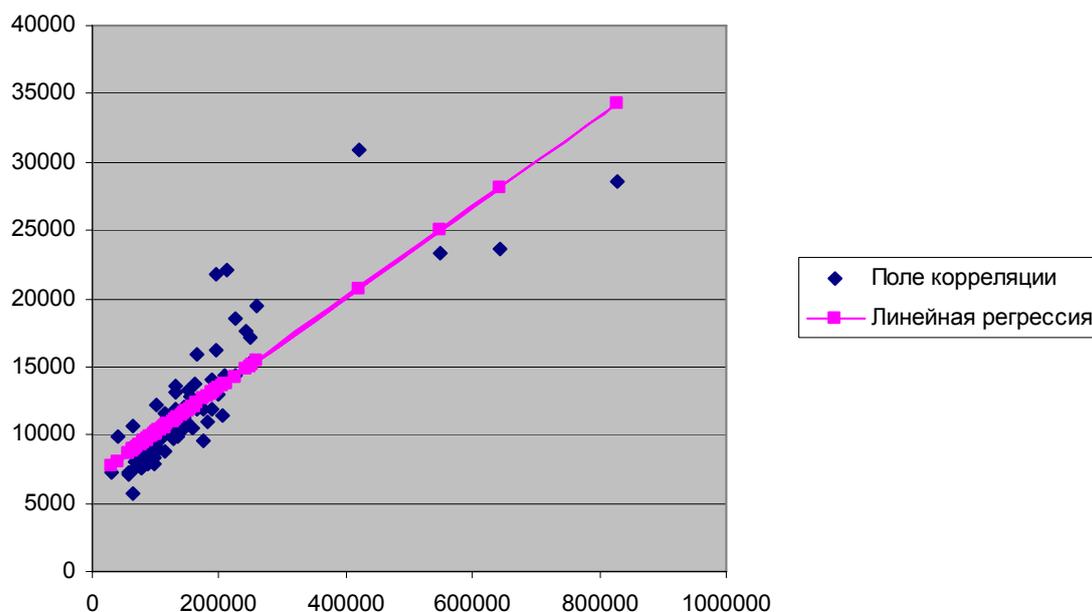


Рис. 1. Корреляционное поле $X_1 - Y$

Далее проведем проверку значимости построенного уравнения линейной регрессии по критерию Фишера: $F_{\text{набл}} = 210,8$, а $F_{\text{табл.}} = 3,96$, т. е. $F_{\text{набл}} > F_{\text{табл.}}$, а значит, основная гипотеза о том, что все коэффициенты уравнения равны нулю, отклоняется.

Полученной модели можно дать следующую экономическую интерпретацию: при увеличении валового регионального продукта на душу населения на 100 рублей среднемесячная заработная плата увеличится на 3,3 рубля.

Коэффициент $\alpha = 6799,80$ в данной модели интерпретации не имеет, т. к. существование региона с нулевым региональным продуктом на душу населения невозможно.

Оценим показатели связи с помощью показателей детерминации. Коэффициент корреляции $r_{xy} = 0,85$ показывает достаточно высокую связь между показателями.

Коэффициент детерминации $R^2 = 0,73$ говорит о том, что 73 % вариации уровня номинальной заработной платы объясняется вариацией уровня ВРП на душу населения. Остальные 27 % вариации объясняются неучтенными в данной модели факторами.

Рассчитаем средний коэффициент эластичности, который показывает, на сколько процентов изменяется в среднем результативный признак при изменении фактора на 1 %: $\varepsilon = 0,43$. Значит, при повышении ВРП на душу населения на 1 % повышается и уровень доходов на 0,43 %.

Проверим также коэффициенты уравнения линейной регрессии на значимость: $t_{\text{табл}}(0,05;78) = 1,99$, $t_a = 15,14$, $t_b = 14,52$. Оба коэффициента t_a и t_b по модулю больше $t_{\text{табл}}$, следовательно, оба коэффициента значимы.

Проблема оценки качества модели заслуживает серьезного внимания, т. к. использование некачественной модели часто приводит к неверным выводам и в дальнейшем – к принятию неверных решений.

С помощью средней ошибки аппроксимации оценим точность прогноза: $A = 13\%$. По данной модели средняя ошибка аппроксимации превышает рекомендованное значение в 8–10%. Но ее превышение не значительно. И, следовательно, приближение построенной модели к наблюдаемым статистическим значениям считается хорошим.

Оценим также точность нашего прогноза с помощью критерия MAD: $MAD = 1662,12$.

Затем была проведена проверка гипотезы равенства дисперсий наблюдений, потому как нарушение этой гипотезы существенно снижает качество оценок неизвестных параметров. Проверка выявила отсутствие гетероскедастичности, потому как: $F_{\text{табл.}} > F_{\text{стат.}}$

Последним этапом была проведена проверка гипотезы о наличии корреляционной связи соседних остатков (автокорреляции первого порядка) при помощи критерия Дарбина–Уотсона. Наличие автокорреляции означает неадекватность построенной парной регрессии истинной зависимости и недостаточность построенной парной регрессии для прогнозирования.

По таблице критических значений критерия Дарбина–Уотсона были найдены значения d_{Low} и d_{Upper} . Их значения составили: $d_{\text{Low}}=1,61$, а $d_{\text{Upper}}=1,66$. Статистика Дарбина–Уотсона составила: $d = 0,037$, а $4-d = 3,96$, а это не позволяет сделать никаких точных выводов, появилась необходимость вычислить коэффициент автокорреляции первого порядка. Его значение составило: $r_1 = 0,96$.

А значит, теперь можно сказать, что автокорреляция присутствует. Таким образом, данная модель парной линейной регрессии неадекватна. Необходимо подобрать другую модель для данного фактора. Попробуем логарифмическую модель, т. к. можно сказать, что визуально она подходит довольно хорошо (рис. 2).

Таким образом, уравнение регрессии будет выглядеть:

$$\hat{Y} = -73613,27 + 7265,6 \ln X_1.$$

Далее проведем проверку значимости построенного уравнения регрессии по критерию Фишера: $F_{\text{набл}} = 211,21$, а $F_{\text{табл.}} = 3,96$, т. е.

$F_{\text{набл}} > F_{\text{табл.}}$, а значит, основная гипотеза о том, что все коэффициенты уравнения равны нулю, отклоняется.

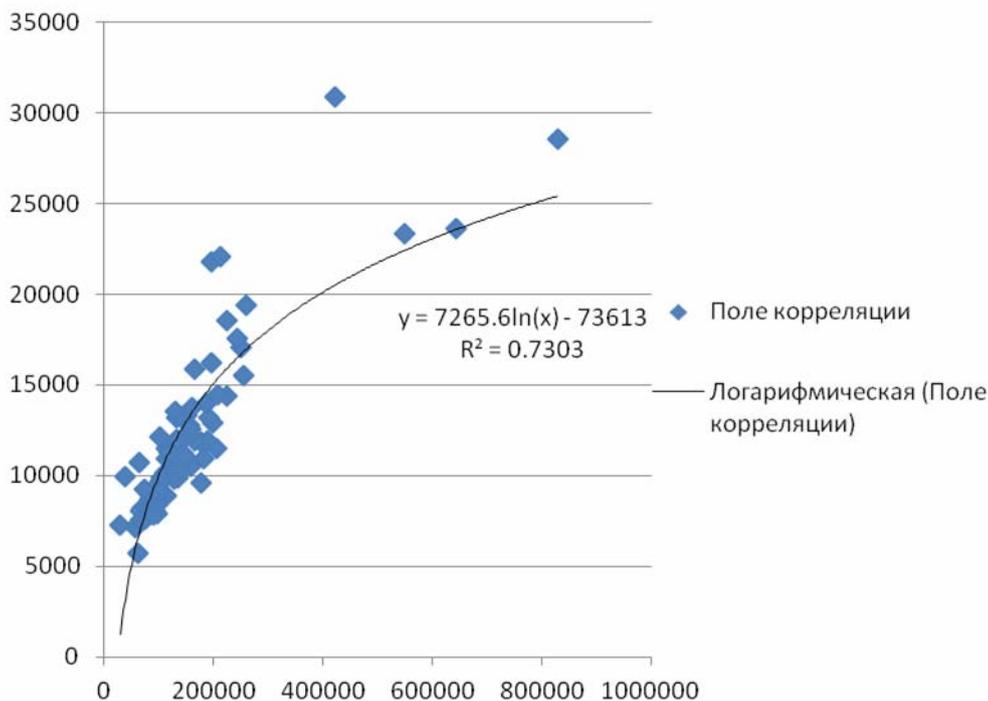


Рис. 2. Логарифмическая регрессия

Далее снова оценим точность прогноза с помощью ошибки аппроксимации: $A = 14\%$. Средняя ошибка аппроксимации превышает рекомендованное значение в 8–10%. Но ее превышение не значительно. И, значит, приближение построенной модели к наблюдаемым статистическим значениям считается хорошим.

Оценим также точность нашего прогноза с помощью критерия MAD: $MAD = 1667,88$. Этот фактор включаем во множественную модель.

Исследование влияния уровня экономически активного населения на уровень доходов

Для изучения фактора X_2 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 3). По данному полю корреляции можно смело предположить наличие линейной связи между факторами.

В этом случае линейное уравнение парной регрессии будет иметь вид:

$$\hat{Y} = -34263,438 + 687,46 X_2.$$

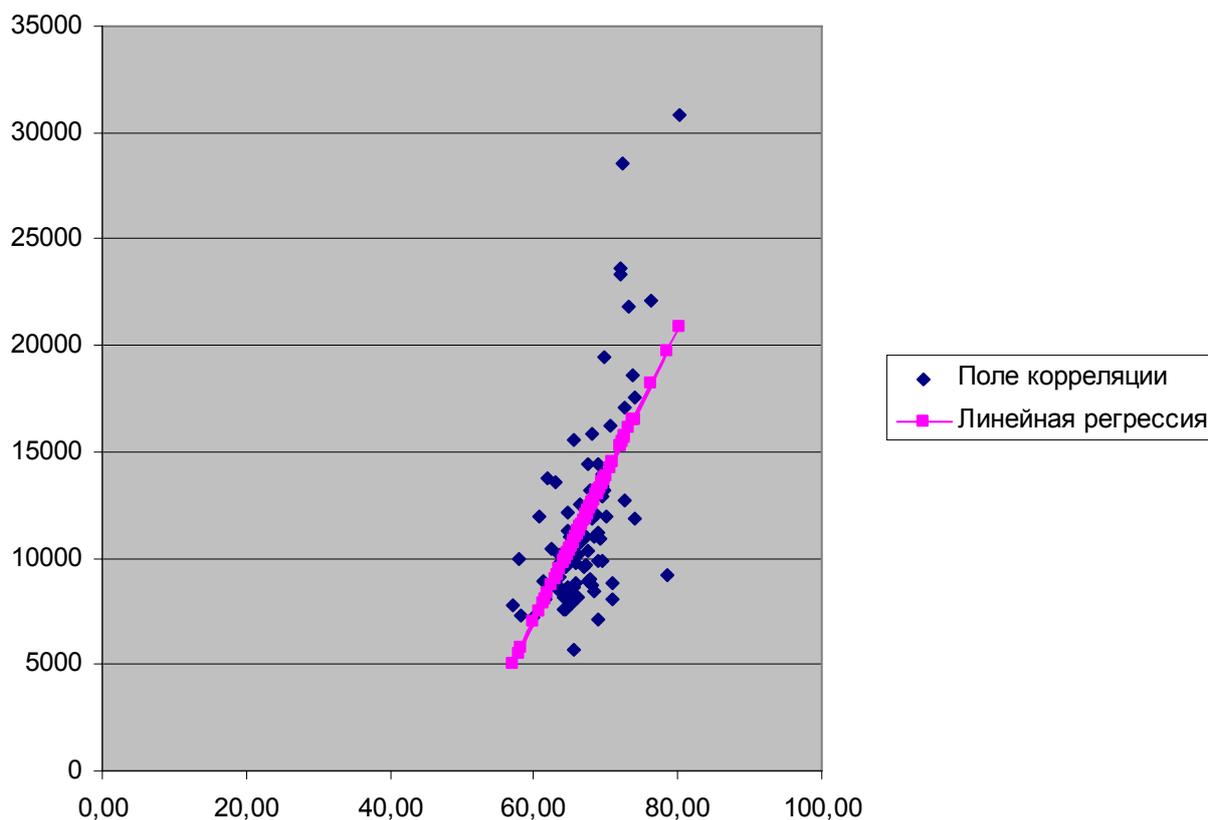


Рис. 3. Корреляционное поле $X_2 - Y$

Затем необходимо провести обязательную проверку построенного уравнения регрессии по критерию Фишера: $F_{\text{набл}} = 49,38$, а $F_{\text{табл.}} = 3.96$, т. е. $F_{\text{набл}} > F_{\text{табл.}}$, а значит, основная гипотеза о том, что все коэффициенты уравнения равны нулю, отклоняется.

Таким образом, данная модель в целом значима и подходит для описания зависимой переменной. Экономическая интерпретация для полученной модели будет выглядеть таким образом: при увеличении уровня экономически активного населения на 1 % среднемесячный доход населения увеличится почти на 700 рублей.

Так как фактически невозможно, чтобы уровень экономически активного населения в регионе был равен нулю, то коэффициент $\alpha = -34263,438$ в данной модели интерпретации не имеет.

Затем необходимо оценить показатели связи в модели. Для этого используются коэффициенты корреляции и детерминации.

- Коэффициент корреляции $r_{xy} = 0,62$ показывает достаточно высокую связь между факторами.

- Коэффициент детерминации $R^2 = 0,39$ говорит о том, что всего лишь 39 % вариации уровня доходов населения объясняется вариацией

уровня экономически активного населения. Остальные 61 % вариации объясняются неучтенными в данной модели факторами.

Далее, рассчитаем средний коэффициент эластичности. Он покажет, на сколько процентов изменяется в среднем результативный признак при изменении фактора на 1 %: $\Theta = 3,87$. Значит, при повышении экономически активного населения в регионе на 1% повышается и уровень доходов на 3,87 %.

Проверим также коэффициенты уравнения линейной регрессии на значимость: $t_{\text{табл}}(0,05;78) = 1,99$, $t_a = -5,201$, $t_b = 7,03$. Оба коэффициента t_a и t_b по модулю больше $t_{\text{табл}}$, это означает, что оба коэффициента значимы.

Помня о том, что использование некачественной модели неизбежно приводит к неверным выводам и принятию неверных решений, с помощью средней ошибки аппроксимации оценим точность прогноза: $A = 23$ %.

По данной модели средняя ошибка аппроксимации превышает рекомендованное значение в 8–10 %. И ее превышение достаточно значительно. Таким образом, приближение построенной модели к наблюдаемым статистическим значениям нельзя считать хорошим.

Оценим также точность нашего прогноза с помощью критерия MAD: $MAD = 2661,08$.

Была проведена проверка гипотезы равенства дисперсий наблюдений. Проверка выявила отсутствие гетероскедастичности. Снова была проведена проверка гипотезы о наличии корреляционной связи соседних остатков (автокорреляции первого порядка) при помощи критерия Дарбина–Уотсона.

По таблице критических значений критерия Дарбина–Уотсона были найдены значения d_{Low} и d_{Upper} . Их значения составили: $d_{\text{Low}}=1,61$, а $d_{\text{Upper}}=1,66$.

Статистика Дарбина–Уотсона составила: $d = 0,35$, а $4-d = 3,65$, а это означает, что автокорреляция положительная.

Можно сказать, что автокорреляция присутствует. Значит, данная модель парной линейной регрессии неадекватна. Необходимо подобрать другую модель для данного фактора.

Предположительно лучшей будет степенная модель (рис. 4). Уравнение парной регрессии примет вид:

$$\hat{Y} = 0,02 \times X^{3,16}.$$

Проверка значимости построенного уравнения регрессии по критерию Фишера: $F_{\text{набл}} = 45,14$, а $F_{\text{табл.}} = 3,96$, т. е. $F_{\text{набл}} > F_{\text{табл.}}$, а значит, модель значима в целом.

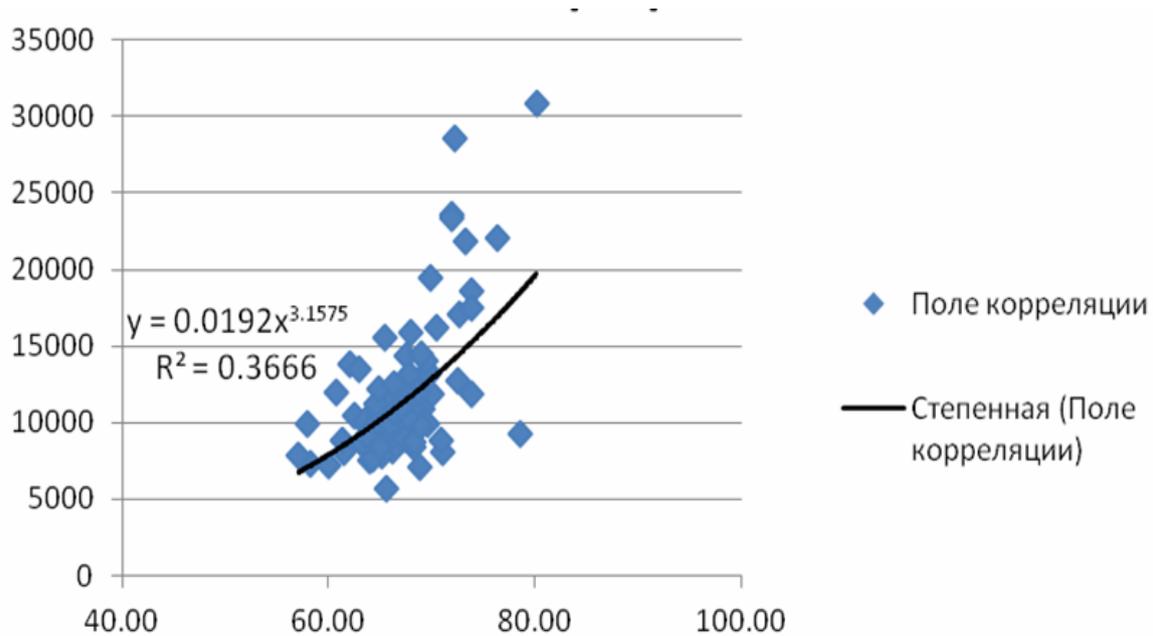


Рис. 4. Степенная регрессия

Оценим точность прогноза с помощью ошибки аппроксимации: $A = 21\%$. Ее превышение не значительно, и, значит, приближение построенной модели к наблюдаемым статистическим значениям считается хорошим.

Оценим также точность нашего прогноза с помощью критерия MAD: $MAD = 2540,66$.

Проверка на автокорреляцию показала ее отсутствие, данный фактор включаем во множественную модель.

Исследование влияния объема инвестиций в основной капитал на уровень доходов населения

Для изучения фактора X_3 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 5).

По получившемуся полю корреляции предположим и проверим наличие линейной связи между факторами Y и X_3 .

Линейное уравнение парной регрессии будет выглядеть:

$$\hat{Y} = 10322,91 + 0,02 X_3.$$

Проведем проверку построенного уравнения регрессии по критерию Фишера: $F_{\text{набл}} = 30,62$, а $F_{\text{табл.}} = 3,96$, т.е. $F_{\text{набл}} > F_{\text{табл.}}$, а значит, модель в целом значима, подходит для описания зависимой переменной, а основная гипотеза о том, что все коэффициенты уравнения равны нулю, отклоняется.

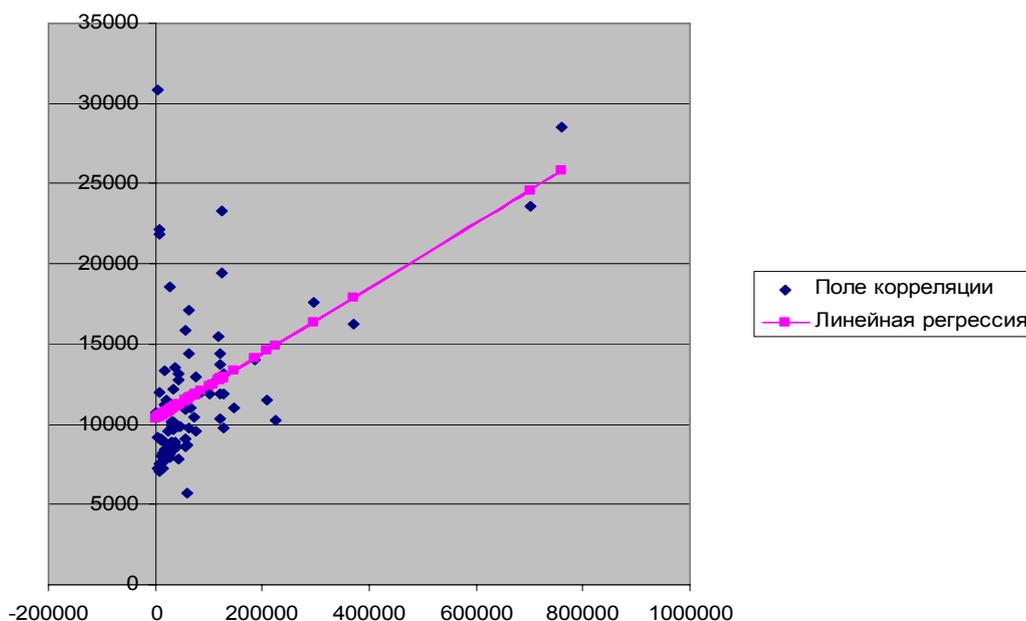


Рис. 5. Корреляционное поле $X_3 - Y$

Далее, приведем экономическую интерпретацию для полученной модели: при увеличении объема инвестиций в основной капитал на 1 млн рублей доход населения увеличится на 0,02 рубля.

В случае, если объем инвестиций в основной капитал составит 0 рублей, то среднемесячный доход населения составит приблизительно 10 323 рубля.

Далее, оценим показатели связи в модели. Для этого используем коэффициенты корреляции и детерминации:

- коэффициент корреляции $r_{xy} = 0,53$, значит, связь между показателями достаточно высокая;
- коэффициент детерминации $R^2 = 0,28$ говорит о том, что лишь 28 % вариации уровня доходов населения объясняется вариацией уровня объема инвестиций в основной капитал. Остальные 72 % вариации объясняются неучтенными в данной модели факторами.

Средний коэффициент эластичности покажет, на сколько процентов изменяется в среднем результативный признак при изменении фактора на 1 %: $\varepsilon = 0,13$. Значит, при повышении объема инвестиций в основной капитал на 1 % повышается и уровень среднемесячной заработной платы на 0,13 %.

Проверим также коэффициенты уравнения линейной регрессии на значимость: $t_{\text{табл}}(0,05;78) = 1,99$, $t_a = 19,22$, $t_b = 5,53$. Оба коэффициента t_a и t_b по модулю больше $t_{\text{табл}}$, это означает, что оба коэффициента значимы.

С помощью средней ошибки аппроксимации оценим точность прогноза: $A = 22\%$. По данной модели средняя ошибка аппроксимации превышает рекомендованное значение в 8–10%. И ее превышение достаточно значительно. Таким образом, приближение построенной модели к наблюдаемым статистическим значениям нельзя считать хорошим.

Оценим также точность нашего прогноза с помощью критерия MAD: $MAD = 2625,4$.

Проведем проверку гипотезы равенства дисперсий наблюдений. Проверка выявила отсутствие гетероскедастичности.

Проведем также проверку гипотезы о наличии корреляционной связи соседних остатков (автокорреляции первого порядка) при помощи критерия Дарбина–Уотсона. По таблице критических значений критерия Дарбина–Уотсона были найдены значения d_{Low} и d_{Upper} . Их значения составили: $d_{Low} = 1,61$, а $d_{Upper} = 1,66$.

Статистика Дарбина–Уотсона составила: $d = 0,09$, а $4-d = 3,91$, а это не позволяет сделать никаких точных выводов, появилась необходимость вычислить коэффициент автокорреляции первого порядка. Его значение составило: $r_1 = 0,9$. Следовательно, автокорреляция присутствует. Из этого следует, что данная модель парной линейной регрессии неадекватна. Таким образом, фактор во множественную регрессию не включаем.

Исследование влияния количества студентов высших учебных заведений на уровень доходов населения

Для изучения фактора X_4 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 6).

По получившемуся полю корреляции можно смело сказать, что между факторами Y и X_4 имеется линейная зависимость. Линейное уравнение парной регрессии:

$$\hat{Y} = 11102,688 + 8,767 X_4.$$

Проверка данного уравнения по критерию Фишера: $F_{набл} = 7,03$, а $F_{табл.} = 3,96$, т. е. $F_{набл} > F_{табл.}$, а значит, модель в целом значима, подходит для описания зависимой переменной, а гипотеза о том, что все коэффициенты уравнения равны нулю, отклоняется.

Экономическая интерпретация модели: при увеличении количества студентов в вузах на 1000 человек доход населения увеличится на 8,767 рубля.

В случае если в регионе не имеется ни одного высшего учебного заведения, а значит, и ни одного студента, в нем учащегося, то среднемесячный доход населения составит 11 102,7 рубля.

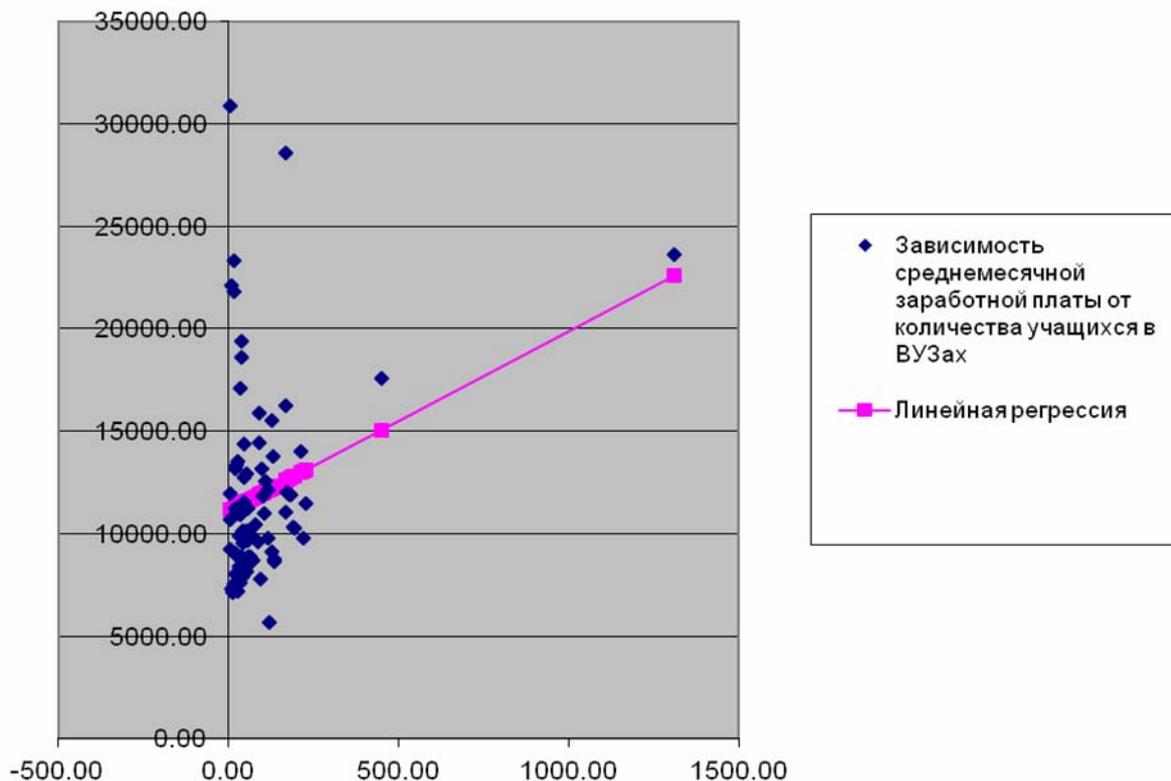


Рис. 6. Корреляционное поле $X_4 - Y$

Оценка показателей связи в модели:

- коэффициент корреляции $r_{xy} = 0,29$. Значит, связь между показателями достаточно высокая;
- коэффициент детерминации $R^2 = 0,08$ говорит о том, что только 8 % вариации уровня доходов населения объясняется вариацией уровня количества учащихся в вузах. Остальные 92 % вариации объясняются не учтенными в данной модели факторами.

Средний коэффициент эластичности: $\varepsilon = 0,07$. Значит, при повышении количества студентов в вузах на 1 %, уровень доходов повышается на 7 %.

Проверка коэффициентов уравнения на значимость: $t_{\text{табл}}(0,05;78) = 1,99$, $t_a = 18,55$, $t_b = 2,65$. Оба коэффициента t_a и t_b по модулю больше $t_{\text{табл}}$, что означает, что оба коэффициента значимы.

Средняя ошибка аппроксимации: $A = 26$ %. По данной модели средняя ошибка аппроксимации превышает рекомендованное значение в 8–10 %. И ее превышение достаточно значительно. Таким образом, приближение построенной модели к наблюдаемым статистическим значениям считается нехорошим. Критерий MAD: $\text{MAD} = 3090,4$.

Исследование влияния численности граждан, имеющих статус безработных, на уровень доходов населения

Для изучения фактора X_5 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 7). При рассмотрении данного поля корреляции трудно точно понять вид зависимости, но можно предположить существование нескольких возможных видов: линейная зависимость; квадратичная; гиперболическая; степенная.

Были получены следующие уравнения парных регрессий:

- линейная регрессия:

$$\hat{Y} = 12066,83 - 6,32 X_5;$$

- квадратичная регрессия:

$$\hat{Y} = 11993,98 - 1,7 X_5 - 0,01 X_5^2;$$

- гиперболическая регрессия:

$$\hat{Y} = 12692,74 - 14805,28/X_5;$$

- степенная регрессия:

$$\hat{Y} = 11993,68 \times X_5^{-0,025}.$$

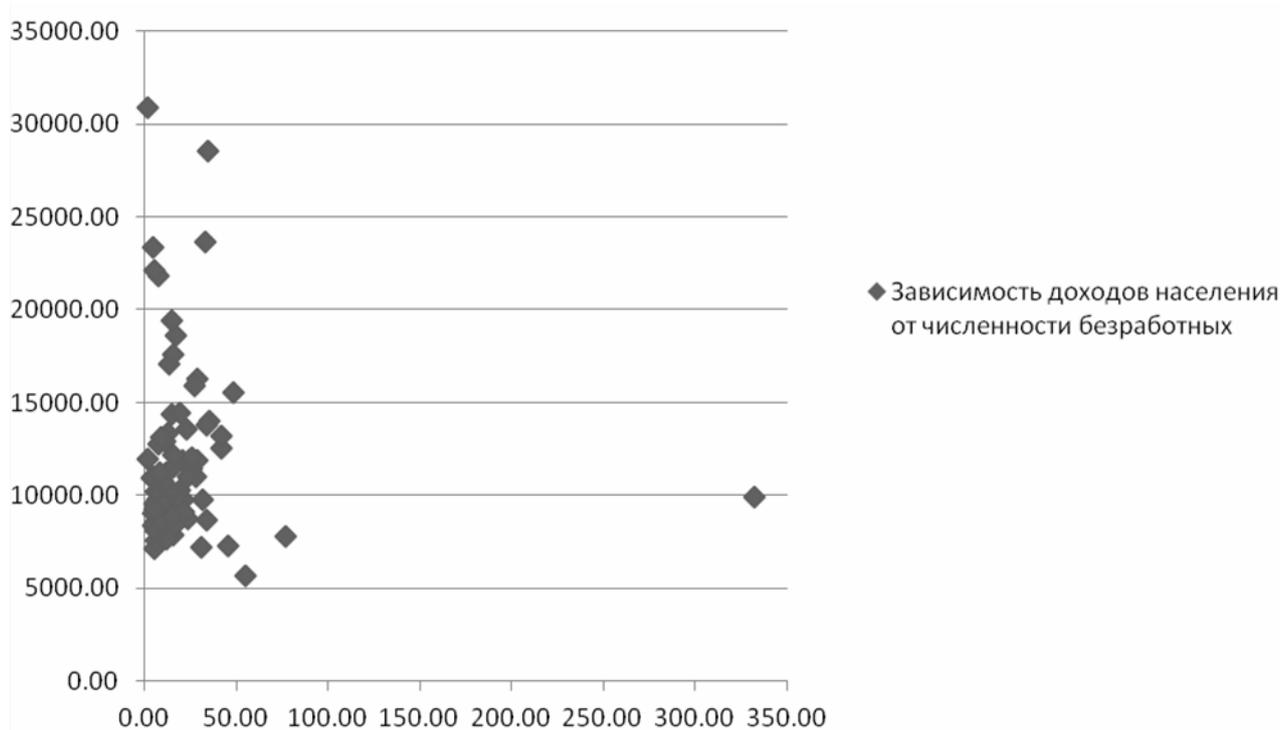


Рис. 7. Корреляционное поле $X_5 - Y$:

Для того чтобы выбрать лучшую из построенных моделей, необходимо использовать критерии (табл. 1):

- метод абсолютных отклонений. Лучшей из моделей будет та, у которой коэффициент MAD наименьший;
- средняя ошибка аппроксимации. Чем она меньше, тем лучше построенная модель аппроксимирует наблюдаемые данные.

Таблица 1

<i>Тип модели</i>	<i>MAD</i>	<i>R²</i>	<i>S_{ост}</i>	<i>A</i>
Линейная	3311,6	0,003	4765,3	0,28
Квадратичная	3304,4	0,003	4795,9	0,28
Гиперболическая	3415,8	0,07	4598,3	0,29
Степенная	3178,5	0,004	0,33	0,25

Лучше всего аппроксимирует наблюдаемые данные степенная модель. Поэтому сделаем выбор в ее пользу.

Затем проверим уравнение регрессии по критерию Фишера: $F_{\text{набл}} = 0,2$, а $F_{\text{табл.}} = 3,96$, т. е. $F_{\text{набл}} < F_{\text{табл.}}$, а значит, модель в целом незначима и не подходит для описания зависимой переменной.

Оценим показатели связи в модели:

- коэффициент корреляции $r_{xy} = -0,05$. Значит, связь между показателями очень слабая;
- коэффициент детерминации $R^2 = 0,0025$ говорит о том, что всего лишь 0,25 % вариации уровня доходов населения объясняется вариацией уровня количества безработных. Остальные 99,75 % вариации объясняются неучтенными в данной модели факторами.

В итоге мы можем сделать вывод о том, что математическая модель не подходит для описания зависимой переменной. Поэтому включение данного фактора в модель множественной регрессии нецелесообразно.

Исследование влияния совокупности факторов

В качестве факторов, влияющих на уровень средней заработной платы, после предварительного исследования были отобраны четыре:

- X_1 – валовой региональный продукт на душу населения в рублях;
- X_2 – уровень экономически активного населения РФ в возрасте 15–72 лет в процентах;
- X_3 – объем инвестиций в основной капитал в миллионах рублей;
- X_4 – численность студентов высших учебных заведений по регионам РФ в тысячах человек.

Парные коэффициенты корреляции представим в виде корреляционной матрицы:

Таблица 2

Q	X ₁	X ₂	X ₃	X ₄	Y
X ₁	1	0,4668	0,7871	0,4801	0,8544
X ₂	0,4668	1	0,23367	0,1448	0,5953
X ₃	0,7871	0,2337	1	0,7324	0,5310
X ₄	0,4801	0,1448	0,7324	1	0,2876
Y	0,8544	0,5953	0,5310	0,2876	1

Проанализировав табл. 2, можно сделать вывод, что наибольшее влияние на доходы населения оказывает первый фактор – валовой региональный продукт на душу населения, так как значение линейного коэффициента корреляции составляет 0,85. Связь между факторами, судя по парным коэффициентам корреляции, достаточно хорошая. Мультиколлинеарность присутствует, так как определитель матрицы Q близок к нулю.

Используя метод пошагового включения, получаем, что в модель множественной регрессии следует включить только два фактора X₁ и X₂:

$$\hat{Y} = -12856,7 + 0,028 X_1 + 304,507 X_2.$$

Экономическая интерпретация коэффициентов:

- при увеличении ВРП на душу населения на 100 рублей среднемесячная заработная плата увеличится на 2,8 рубля при неизменности остальных факторов;
- при увеличении уровня экономически активного населения на 1% среднемесячный доход населения увеличится на 304,507 рубля при неизменности остальных факторов.

Коэффициент $\alpha = 12856,7$ в данной модели интерпретации по-прежнему не имеет.

Оценим тесноту связи с помощью коэффициентов детерминации:

- коэффициент детерминации $R^2 = 0,788$ говорит о том, что 79 % вариации уровня доходов населения объясняется вариацией четырех рассмотренных факторов. Остальные 21 % вариации объясняются неучтенными в данной модели факторами;
- скорректированный коэффициент детерминации равен: $R^2_{adj} = 0,78$.

Данная модель является значимой по критерию Фишера: $F_{набл} = 286,92$, $F_{табл.} = 3,97$, т. е. $F_{набл} > F_{табл.}$

Проверим снова коэффициенты уравнения линейной регрессии на значимость: $t_{\text{табл}}(0,05;78) = 1,99$, $t_a = -3,005$, $t_{b_1} = 12,077$, $t_{b_2} = 4,61$. Все t -статистики по модулю больше $t_{\text{табл}}$, что означает, что все они значимы и подлежат включению в модель.

Теперь произведем сравнение построенной модели множественной регрессии с моделями парной регрессии зависимостей среднего уровня доходов от ВРП на душу населения и от уровня экономической активности населения для того, чтобы сделать выбор между лучшей из этих моделей (табл. 3).

Таблица 3

<i>Тип модели</i>	R²	S_{ост}	MAD	A
Логарифмическая (Y-X ₁)	0,73	2477,9	1667,9	14%
Степенная (Y-X ₂)	0,37	0,27	2540,66	21%
Множественная (Y-X ₁ -X ₂)	0,79	2209	1605,906	14%

Модель множественной регрессии значительно лучше по каждому из рассмотренных показателей. Значит, в целях избегания возможных ошибок при принятии решений следует принять множественную модель зависимости среднего уровня доходов населения от двух факторов X₁ (валовой региональный продукт на душу населения в рублях) и X₂ (уровень экономической активности населения РФ в возрасте 15–72 лет в процентах) как рекомендуемую модель к дальнейшему использованию.

ЗАКЛЮЧЕНИЕ

В ходе проведенного исследования была изучена зависимость пяти факторов на результирующий признак: и каждого в отдельности в парной модели, и некоторых из них в совокупности во множественной модели. С этой целью были построены четыре парные модели линейной регрессии и модель множественной линейной регрессии. На основе построенных корреляционных полей влияние каждого из факторов на результирующий фактор было выдвинуто предположение о линейной зависимости признаков.

После проведенного исследования стало ясно, что значимыми можно признать не все модели. Модель зависимости доходов от численности

безработных в регионе оказалась незначимой. Также в процессе построения множественной модели были исключены еще два фактора: количество студентов вузов и объем инвестиций в основной капитал, так как они оказались незначимыми во множественной модели.

Таким образом, в ходе проведенного исследования было установлено, что наиболее подходящей является множественная модель, отражающая зависимость среднего уровня начисленной заработной платы от ВРП на душу населения и уровня экономической активности.

Получившаяся модель приняла следующий вид:

$$\hat{Y} = - 12856,7 + 0,028 X_1 + 304,507 X_2.$$

Проведенное исследование показало, что средний уровень заработной платы населения зависит в основном от экономического развития каждого региона, а также от уровня экономически активного населения. Изменение уровня дохода на 0,79 % объясняется изменением этих двух факторов.

ЛИТЕРАТУРА

1. *Евсеев Е. А., Буре В. М.* Эконометрика. СПб.: Изд-во МБИ, 2007. –139 с.
2. *Тарашина С. И., Панкратова Я. Б.* Выполнение курсовой работы по эконометрике: Учеб.-метод. пособие. СПб.: Изд-во МБИ, 2007. – 97 с.
3. *Кремер Н.Ш., Путко Б.А.* Эконометрика: Учебник для вузов / Под ред. проф. Н.Ш. Кремера. М.: ЮНИТИ-ДАНА, 2003. – 311 с.
4. Практикум по эконометрике: Учеб. пособие / И.И. Елисеева, С.В. Курьшева, Н.М. Гордеенко и др. Под ред. И.И. Елисеевой. М.: Финансы и статистика, 2003. – 192 с.: ил.
5. Сборники государственной статистики Росстат [Электронный ресурс]: <http://www.gks.ru>

ПРИЛОЖЕНИЕ

Таблица

Данные

	Среднемесячная начисленная з/п (в руб.)	ВРП на душу населения (руб.)	Уровень экономически активного населения (%)	Объем инвестиций в основной капитал (млн руб.)	Студенты вузов (тыс. чел.)	Численность безработных (тыс. чел.)
	У	Х1	Х2	Х3	Х4	Х5
Белгородская область	10479,5	159384,9	62,50	73562,2	80,10	9,90
Брянская область	8189,6	80839	64,2	21313,8	52,20	11,80
Владимирская область	9688,1	101953,8	67,3	32824,6	57,30	13,70
Воронежская область	8730,9	99963	63,2	58245,2	134,10	23,40
Ивановская область	8171,6	69928,8	66,2	14419,4	54,50	11,60
Калужская область	10926,8	113618,9	69,1	26070	37,20	4,20
Костромская область	9058,1	98021,7	67,9	11331,8	21,80	5,00
Курская область	8856,8	113240,5	65,8	29292,8	65,40	8,30
Липецкая область	10907,1	182503,4	66,7	56339,3	36,80	4,90
Московская область	16234,5	196130,6	70,5	372018,3	170,50	28,90
Орловская область	8610,7	96802,5	65,5	23902,9	46,40	6,30
Рязанская область	9796,6	107925,7	63,7	29096,4	54,30	7,00
Смоленская область	9552,1	100307,2	66,9	22396,5	44,20	5,40
Тамбовская область	7903	96795,5	65,2	27836,7	42,80	15,30
Тверская область	10177	114717,5	66	33027,7	44,70	6,30
Тульская область	10137,1	112409,1	63,6	30445,1	55,00	7,40
Ярославская область	11214,8	144189,4	69	37463,3	59,30	11,60

Таблица (продолжение)

	Средне- мечная на- численная з/п (в руб.)	ВРП на душу на- селения (руб.)	Уровень эко- номически активного населения (%)	Объем ин- вестиций в основной капитал (млн руб.)	Студен- ты ву- зов (тыс. чел.)	Числен- ность без- работных (тыс. чел.)
	У	Х1	Х2	Х3	Х4	Х5
г. Москва	23623,3	643733,1	72	700572,4	1312,60	33,20
Республика Карелия	13342,1	151210,8	69,6	18556,7	24,50	12,60
Республика Коми	17077,3	249570,9	72,7	62333,9	35,90	13,50
Архангель- ская область	14400,3	224807,8	67,5	121729,3	47,30	15,20
Вологодская область	12913,9	199086,8	69,4	74706,8	52,90	11,40
Калинин- градская область	12750,3	155668,9	72,6	42302,4	45,30	7,20
Ленинград- ская область	13154,8	191009,9	69,7	128259,5	20,10	8,80
Мурманская область	18581	225044,6	73,8	25928	37,80	17,40
Новгород- ская область	11004,9	133676,2	68,3	21749,6	22,70	4,90
Псковская область	8950,5	88966,6	67,4	11831,2	24,60	11,40
Санкт-Пе- тербург	17552	242755	73,9	296698,6	450,10	15,60
Республика Адыгея	8056,4	66089,8	61,5	10472,5	20,40	9,50
Республика Дагестан	5696,4	62366,9	65,6	58219,8	119,50	55,10
Республика Ингушетия	7285,4	29903,7	58,2	5773,2	10,10	45,50
Кабардино- Балкарская Республика	7213,9	56528,8	60	12754,4	29,70	30,60
Республика Калмыкия	7101,2	58145,9	68,8	6081,3	11,50	5,50
Карачаево- Черкесская Республика	7558,4	65969,6	64	9057,5	17,20	5,80

Таблица (продолжение)

	Средне- мечная на- численная з/п (в руб.)	ВРП на душу на- селения (руб.)	Уровень эко- номически активного населения (%)	Объем ин- вестиций в основной капитал (млн руб.)	Студен- ты ву- зов (тыс. чел.)	Числен- ность без- работных (тыс. чел.)
	У	Х1	Х2	Х3	Х4	Х5
Республика Северная Осетия – Алания	7625,8	76455,4	64,3	12610,3	35,50	12,20
Чеченская Республика	9916,7	39102,7	57,9	31641	31,00	332,00
Краснодар- ский край	10260	128202,3	64,9	224893	194,90	19,50
Ставрополь- ский край	8647,7	81800,7	64,7	55372	137,30	33,70
Астрахан- ская область	9866,9	105326,1	69,6	45565,5	43,90	14,30
Волгоград- ская область	9770,2	127700,8	64,5	61425,3	115,80	21,70
Ростовская область	9779,6	106222,8	65,9	126259,3	220,70	31,90
Республика Башкорто- стан	11027,1	148403,5	65	148001,9	169,50	28,20
Республика Марий Эл	8404,3	79993,2	68,3	17117,4	30,70	5,00
Республика Мордовия	8103	91234,2	71	26361,7	41,90	6,20
Республика Татарстан	11468,6	204890,2	66,5	210064	227,80	25,50
Удмуртская Республика	9838,6	134061,6	69	44463	72,30	18,10
Чувашская Республика	8703,2	94911,3	68,2	37209,4	72,40	12,20
Пермский край	11856	176183,5	68	101812,8	101,10	20,80
Кировская область	8861,5	84670	70,9	37450,9	59,60	15,30
Нижегород- ская область	10302	140602	67,6	121695,3	191,40	13,90
Оренбург- ская область	9619,6	176733,3	64,3	76577,9	86,30	10,80

Таблица (продолжение)

	Средне- саячная на- численная з/п (в руб.)	ВРП на душу на- селения (руб.)	Уровень эко- номически активного населения (%)	Объем ин- вестиций в основной капитал (млн руб.)	Студен- ты ву- зов (тыс. чел.)	Числен- ность без- работных (тыс. чел.)
	У	Х1	Х2	Х3	Х4	Х5
Пензенская область	8566,4	86138,8	63,2	39929	59,60	9,20
Самарская область	11920,7	189051,6	70	126734,9	180,40	24,10
Саратовская область	9108,3	101053	63,7	56793,8	126,60	21,50
Ульяновская область	8412,7	96217,1	65	32505	55,30	16,00
Курганская область	8883,1	85744	61,3	18412	37,50	14,80
Свердловская область	13986,9	187607,3	69,4	187159,6	214,40	35,10
Тюменская область	28565	829155	72,3	761163,4	169,40	34,90
Челябинская область	11897,5	165907,5	73,9	122668,6	185,10	28,90
Республика Алтай	9228,2	74266,3	78,6	5756,2	5,30	5,20
Республика Бурятия	11528,6	114126,1	66,3	20634,2	47,50	15,70
Республика Тыва	10701,6	63686,1	66,4	2116	6,50	10,50
Республика Хакасия	11251,2	119252,8	64,6	16662,8	21,30	8,40
Алтайский край	7804,7	88935,2	57,1	42680,5	95,70	77,10
Забайкальский край	12161,5	101051,6	64,8	34024,2	117,90	15,60
Красноярский край	15510	253942,4	65,5	118002	129,50	48,20
Иркутская область	13770	160522,8	62	122445,2	131,40	33,50
Кемеровская область	12554,9	157297,5	66,4	107587	108,80	42,00
Новосибирская область	12017	144869	68,7	85229,1	170,90	25,70
Омская область	11003,6	149273,4	67,1	65440,2	106,80	23,60

Таблица (окончание)

	Среднемесячная численность з/п (в руб.)	ВРП на душу населения (руб.)	Уровень экономически активного населения (%)	Объем инвестиций в основной капитал (млн руб.)	Студенты вузов (тыс. чел.)	Численность безработных (тыс. чел.)
	У	Х1	Х2	Х3	Х4	Х5
Томская область	14429	208946,1	69	64082,3	89,70	19,40
Республика Саха (Якутия)	19409,2	259249	69,9	124002	37,60	14,60
Камчатский край	21814,7	196069,9	73,3	8476,4	16,30	7,80
Приморский край	13174,1	131578,8	67,8	43158,6	97,40	42,20
Хабаровский край	15883,5	165629,2	68	54840,3	89,50	27,00
Амурская область	13534,4	131039,6	62,9	37253,5	29,80	22,80
Магаданская область	22101,6	211896,2	76,4	7638,8	10,40	5,30
Сахалинская область	23346,3	550227,8	72	125904,6	16,40	4,70
Еврейская автономная область	11968,9	132585,7	60,8	7979,3	7,50	1,70
Чукотский авт.округ	30859,1	421287,6	80,2	5335,4	7,10	1,40

Ласыгина Анастасия Валерьевна

Международный банковский институт

Кузютин Д.В., к.ф.-м.н., доцент; **Панкратова Я.Б.**, ст. преподаватель –
научные руководители

ФАКТОРЫ, ВЛИЯЮЩИЕ НА УРОВЕНЬ РОЖДАЕМОСТИ В РОССИЙСКОЙ ФЕДЕРАЦИИ

На сегодняшний день одной из актуальных и острых проблем современной России выступает демографическая ситуация, а именно: проблема сокращения численности населения. Практически во всех средствах массовой информации идет речь о демографическом состоянии Российской Федерации и об основных направлениях деятельности государства в этой сфере. На федеральном уровне разрабатываются программы стимулирования рождаемости, а именно: меры поддержки молодых семей, поддержки женщин, принимающих решение родить и поднять на ноги ребенка. Но несмотря на некоторое увеличение коэффициента рождаемости, все равно она находится на катастрофически низком уровне. Рождаемость в России не достигает уровня, необходимого для простого воспроизводства населения. Суммарный коэффициент рождаемости составляет 1,4, тогда как для простого воспроизводства населения без прироста численности необходим суммарный коэффициент рождаемости 2,11–2,15. Важность разрешения проблемы рождаемости очевидна, и для ее решения важно знать, какие факторы могут влиять на рождаемость.

Цель данного исследования – выявить факторы, которые могут влиять на уровень рождаемости Российской Федерации, отобрать из них наиболее значимые, а также установить вид зависимости уровня рождаемости от этих факторов.

Выбор факторов, влияющих на уровень рождаемости

Для проведения исследования были отобраны следующие факторы:

X_1 – количество браков в расчете на 1000 человек в промилле;

X_2 – количество разводов в расчете на 1000 человек в промилле;

X_3 – уровень дохода от предпринимательской деятельности, %;

X_4 – ВВП. Данный показатель характеризует уровень экономического развития страны, млрд руб.;

X_5 – уровень безработицы, %.

В качестве результирующего признака было выбрано число родившихся в расчете на 1000 человек за год (промилле). В исследовании использовались данные Федеральной службы государственной статистики (Росстат) за период с 1992 по 2007 годы, опубликованные на сайте Федеральной службы государственной статистики. Далее будет изучено влияние каждого из приведенных выше факторов на уровень рождаемости в отдельности и влияние всех этих факторов в совокупности.

Исследование влияния количества браков на уровень рождаемости

Представим исходные данные об уровне рождаемости и количестве браков за период с 1992 по 2007 годы в виде статистической таблицы, удобной для анализа (табл. 1).

Таблица 1

Го- ды	Число ро- дившихся в расчете на 1000 чело- век за год, промилле, Y	Число браков в расчете на 1000 че- ловек за год, про- милле, X_1	Число разводов в расчете на 1000 че- ловек за год, про- милле, X_2	Уровень дохо- да от предпри- нимательской деятельности, от общего чи- сла доходов населения, %, X_3	Динамика ВВП (в со- поставимых ценах), млрд руб., X_4	Уро- вень безра- боти- цы, %, X_5
1992	10,7	7,1	4,3	8,4	25777,34	5,2
1993	9,4	7,5	4,5	18,6	23507,64	5,9
1994	9,6	7,4	4,6	16	19995	8,1
1995	9,3	7,3	4,5	16,4	19473,7	9,4
1996	8,9	5,9	3,8	13,1	18772,7	9,7
1997	8,6	6,3	3,8	12,5	19035,5	11,8
1998	8,8	5,8	3,4	14,4	18026,6	13,3
1999	8,3	6,3	3,7	12,4	19180,3	13
2000	8,7	6,2	4,3	15,4	21098,3	10,6
2001	9	6,9	5,3	12,6	22174,4	9
2002	9,7	7,1	6	11,9	23216,6	7,9
2003	10,2	7,6	5,5	12	24911,4	8,2
2004	10,4	6,8	4,4	11,7	26705	7,8
2005	10,2	7,5	4,2	11,4	28414,1	7,2
2006	10,4	7,8	4,5	11,1	30516,7	7,2
2007	11,3	8,9	4,8	10	32988,6	6,1

Для изучения влияния фактора X_1 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 1).

Зависимость уровня рождаемости от количества браков

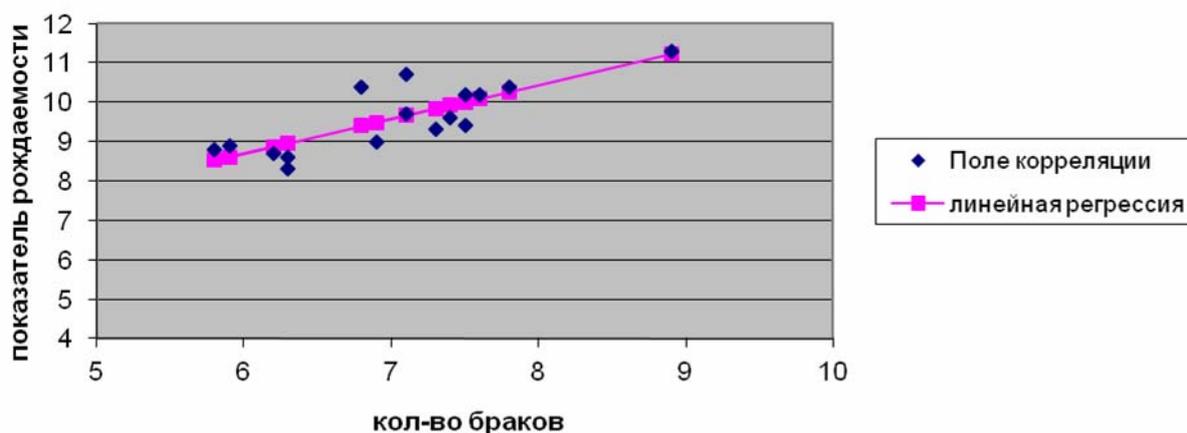


Рис. 1. Корреляционное поле YX_1

При его рассмотрении можно заявить, что наиболее подходящей будет линейная модель, т. к. она наиболее проста и визуально довольно точно отражает зависимость Y от X_1 .

Но тем не менее была изучена возможность существования нескольких возможных видов зависимостей:

- линейная зависимость;
- гиперболическая зависимость;
- квадратичная зависимость.

Для того чтобы осуществить выбор в пользу какой-либо зависимости, сравним полученные результаты по построенным моделям (табл. 2).

Таблица 2

Значения критериев отбора модели

Тип модели	R^2	\bar{A}	MAD	Сост
линейная	0,65	0,04	0,4	0,53
квадратичная	0,65	0,04	0,28	0,55
гиперболическая	0,63	0,04	0,43	0,54

На основе полученных результатов выбор был сделан в пользу линейной модели, так как она проще, и значения этой модели почти не отличаются от значений квадратичной и гиперболической моделей.

Также анализ квадратичной зависимости показал, что эта модель незначима по критерию Фишера, т. к. $F_{\text{набл}} < F_{\text{табл}}$.

Полученное линейное уравнение имеет следующий вид:

$$Y = 3,49 + 0,87 X_1.$$

Полученная модель является значимой по критерию Фишера, т. к. $F_{\text{набл.}} > F_{\text{табл.}}$, где $F_{\text{набл.}} = 26,22$, а $F_{\text{табл.}} = 4,6$. Также коэффициенты уравнения линейной регрессии по модулю превышают $t_{\text{табл.}}$ по таблице распределения Стьюдента: $t_{\text{табл.}}(0,05;14)=2,14$, а $t_{\alpha}=2,9$, $t_{\beta}=5,1$, т. е. являются значимыми. Полученная модель, выражающая зависимость между объясняющей и зависимыми переменными, подходит для описания последней. Модели можно дать следующую экономическую интерпретацию: при увеличении числа браков на 100 промилле в расчете на 1000 человек уровень рождаемости увеличится на 87 промилле в расчете на 1000 человек.

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{yx_1} = 0,81$, это означает, что связь между фактором X_1 (количество браков) и уровнем рождаемости высокая и прямая.

Коэффициент детерминации $R^2 = 0,65$ говорит о том, что 65 % вариации показателя рождаемости объясняется вариацией числа браков. Остальные 35 % вариации объясняются неучтенными в данной модели факторами.

Рассмотрим коэффициент эластичности, который показывает, на сколько процентов изменяется в среднем результирующий признак при изменении факторного на 1 %:

$$\varepsilon = 0,637.$$

Следовательно, при повышении числа браков на 1 % уровень доходов повышается на 0,637 %.

Средняя ошибка аппроксимации $\bar{A} = 4$ %, а среднее абсолютное отклонение $MAD = 0,4$. Исходя из полученных данных, можно сделать вывод, что расчетные и фактические значения достаточно приближены друг к другу, а выбранная модель адекватна.

Исследование влияния количества разводов на показатель рождаемости

Далее изучим зависимость между уровнем рождаемости и количеством разводов за период с 1992 по 2007 гг. Для изучения влияния фактора X_2 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 2).

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную:

$$Y = 7,3 + 0,5 \cdot X_2.$$

Влияние количества разводов на показатель рождаемости

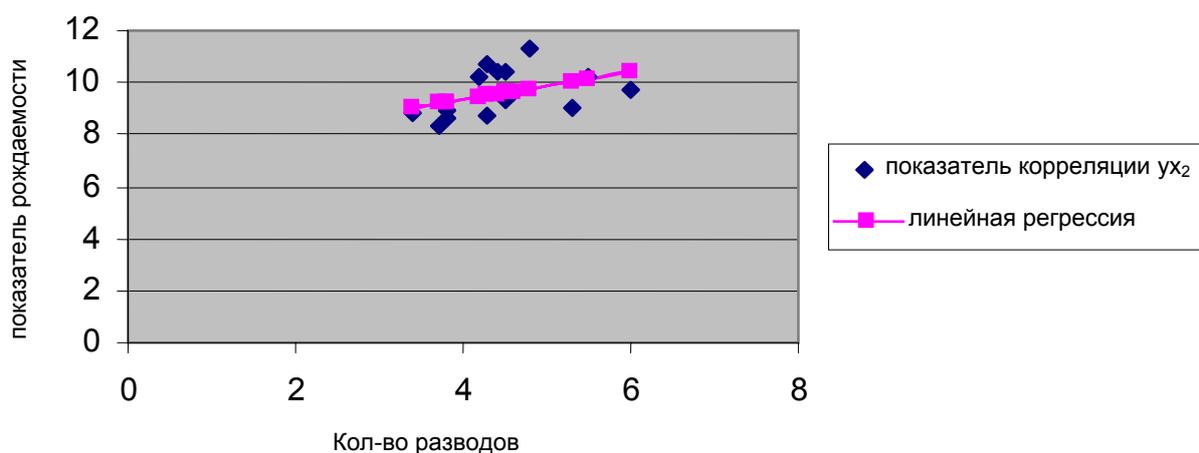


Рис. 2. Корреляционное поле YX_2

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{yx_2} = 0,4$, это означает, что связь не высокая. Коэффициент детерминации $R^2 = 0,16$ говорит о том, что лишь 16 % вариации показателя рождаемости объясняется вариацией числа разводов. Остальные 84 % вариации объясняются неучтенными в данной модели факторами.

Данная модель является незначимой по критерию Фишера, т. к. $F_{\text{набл.}} < F_{\text{табл.}}$, где $F_{\text{набл.}} = 2,76$, а $F_{\text{табл.}} = 4,6$.

Отсюда можно сделать вывод, что математическая модель, выражающая данную зависимость, не подходит для описания зависимой переменной. Поэтому включение данного фактора в модель множественной регрессии нецелесообразно.

Исследование зависимости уровня рождаемости от уровня дохода от предпринимательской деятельности

Теперь исследуем связь между факторами «уровень рождаемости» и «уровень доходов» от предпринимательской деятельности за период с 1992 по 2007 гг. Для изучения влияния фактора X_3 на результирующий признак Y необходимо построить поле корреляции (рис. 3).

При анализе зависимости коэффициента рождаемости от уровня доходов от предпринимательской деятельности была изучена возможность существования нескольких возможных видов зависимостей:

- линейная зависимость;

- гиперболическая зависимость;
- квадратичная зависимость;
- логарифмическая зависимость.

**Зависимость коэффициента рождаемости
от доходов от предпринимательской деятельности**

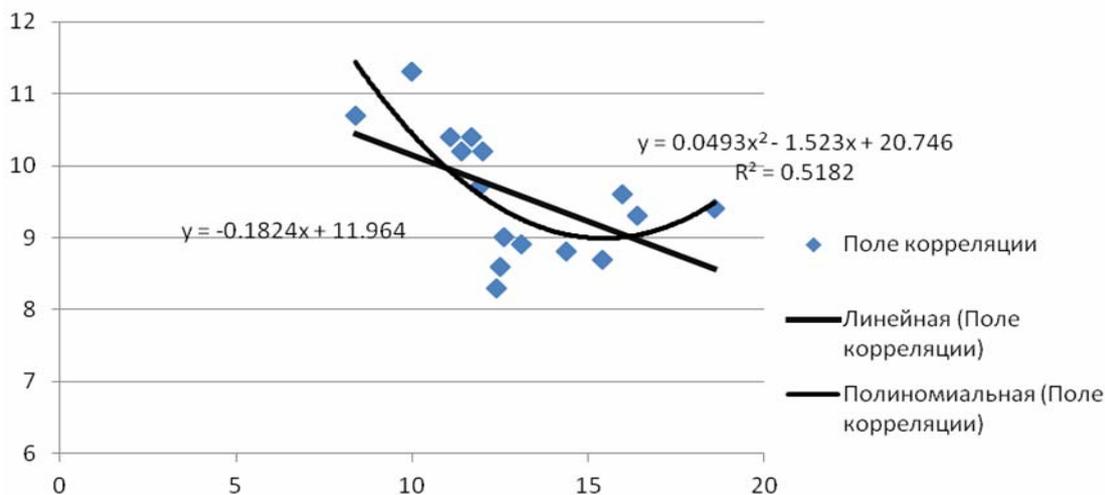


Рис. 3. Корреляционное поле YX_3

Для того чтобы осуществить выбор в пользу какой-либо зависимости, были использованы следующие критерии, представленные в табл. 3.

Таблица 3

Значения критериев отбора модели

Тип модели	R^2	\bar{A}	MAD	Сост
линейная	0,3	0,048	0,61	0,75
квадратичная	0,52	0,053	0,5	0,65
гиперболическая	0,4	0,06	0,57	0,7
логарифм	0,36	0,06	0,58	0,7

Исходя из полученных результатов, выбор был сделан в пользу квадратичной модели, т. к. эта модель более точно описывает зависимость между показателем Y и фактором X_3 .

Уравнение квадратичной модели имеет следующий вид:

$$Y = 20,75 - 1,5 X_3 + 0,05 X_3^2.$$

Квадратичная модель является значимой по критерию Фишера, т. к. $F_{\text{набл}} > F_{\text{табл}}$, где $F_{\text{набл}} = 6,99$, а $F_{\text{табл}} = 3,8$ при $\alpha=0,05$. Также коэффициенты уравнения линейной регрессии по модулю превышают $t_{\text{табл}}$ по таблице распределения Стьюдента: $t_{\text{табл}}(0,05;13)=2,16$, а $t_\alpha=5,6$, $t_\beta=-2,75$, $t_\gamma=2,44$. Таким образом, рассматриваемая модель подходит для анализа.

Также фактор X_3 был исследован на гетероскедастичность с помощью метода Гольдфельда–Квандта. Полученные результаты показали отсутствие гетероскедастичности, это означает, что наблюдения проводились с одинаковой точностью.

Средняя ошибка аппроксимации: $\bar{A}=5,3\%$ и среднее абсолютное отклонение $MAD=0,5$. Исходя из полученных данных, можно сделать вывод, что расчетные и фактические значения достаточно приближены друг к другу, а выбранная модель адекватна.

Коэффициент детерминации $R^2=0,52$, следовательно около 52 % уровня рождаемости определяется уровнем дохода от предпринимательской деятельности.

Полученной модели можно дать следующую экономическую интерпретацию: при увеличении дохода от предпринимательской деятельности на 1 % уровень рождаемости снижается на 1,45 промилле в расчете на 1000 человек, но при достижении дохода от предпринимательской деятельности 15 % уровень рождаемости повышается на 0,05 промилле в расчете на 1000 человек.

Влияние ВВП на уровень рождаемости

Далее исследуем зависимость уровня рождаемости от ВВП. Для этого построим поле корреляции (рис. 4).

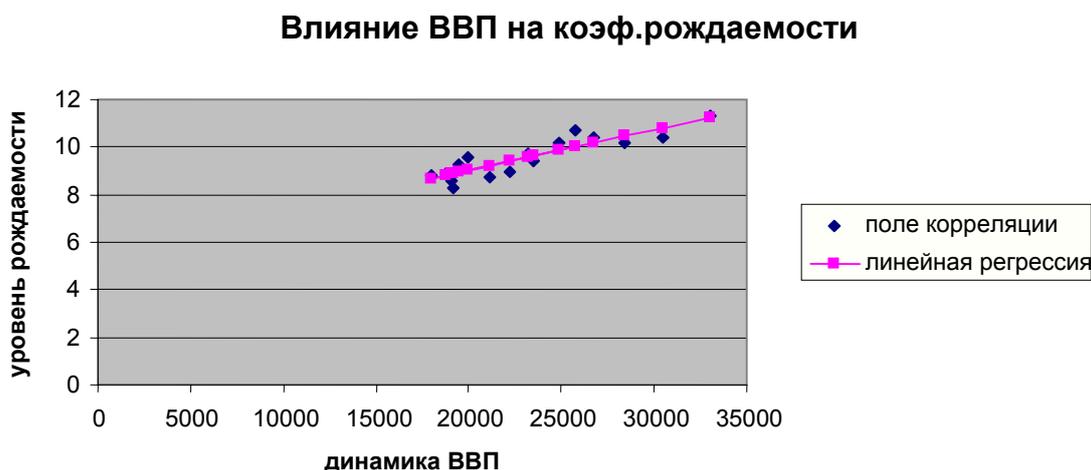


Рис. 4. Корреляционное поле YX_4

По виду поля корреляции можно предположить, что подходящей моделью для анализа влияния фактора X_4 на показатель Y будет линейная модель.

Линейное уравнение парной регрессии выглядит следующим образом:

$$Y = 5,6 + 0,00017 X_4.$$

По критерию Фишера модель значима, т. к. $F_{\text{набл}} > F_{\text{табл}}$, где $F_{\text{набл}} = 31,12$, а $F_{\text{табл}} = 4,6$, при $\alpha=0,05$. Также все коэффициенты линейной регрессии значимы, т. к. они по модулю превышают $t_{\text{табл}}$ по таблице распределения Стьюдента.

Коэффициент корреляции $r_{yx_4} = 0,89$, это означает, что связь между показателем Y и фактором X_4 сильная. Коэффициент детерминации $R^2 = 0,8$, говорит о том, что 80 % вариаций уровня рождаемости объясняется вариацией ВВП. Остальные 20 % объясняются неучтенными в данной модели факторами.

Рассмотрим коэффициент эластичности:

$$\varepsilon = 0,49.$$

Следовательно, при повышении ВВП на 1 % уровень рождаемости повышается на 0,49 %.

Для точности прогноза была рассчитана средняя ошибка аппроксимации $\bar{A} = 3,5$ % и среднее абсолютное отклонение $MAD = 0,33$. Исходя из полученных данных, можно сделать вывод, что приближение построенной модели к наблюдаемым статистическим значениям считается хорошим.

Полученной модели можно дать следующую экономическую интерпретацию: при увеличении ВВП на 10 000 млрд руб. уровень рождаемости увеличится на 1,7 промилле в расчете на 1000 человек.

Исследование влияния уровня безработицы на уровень рождаемости

Теперь исследуем зависимость уровня рождаемости от уровня безработицы. Для этого необходимо построить поле корреляции (рис. 5) фактора X_5 на результирующий признак Y . По данному полю корреляции сложно выявить вид зависимости, поэтому были построены и исследованы три модели. Результаты приведены в табл. 4.

На основе полученных результатов выбор был сделан в пользу линейной модели, т. к. она проще, и все показатели почти не отличаются от показателей в других моделях.

Полученная линейная модель имеет следующий вид:

$$Y = 12,2 - 0,3 X_5.$$

Значения критериев отбора модели

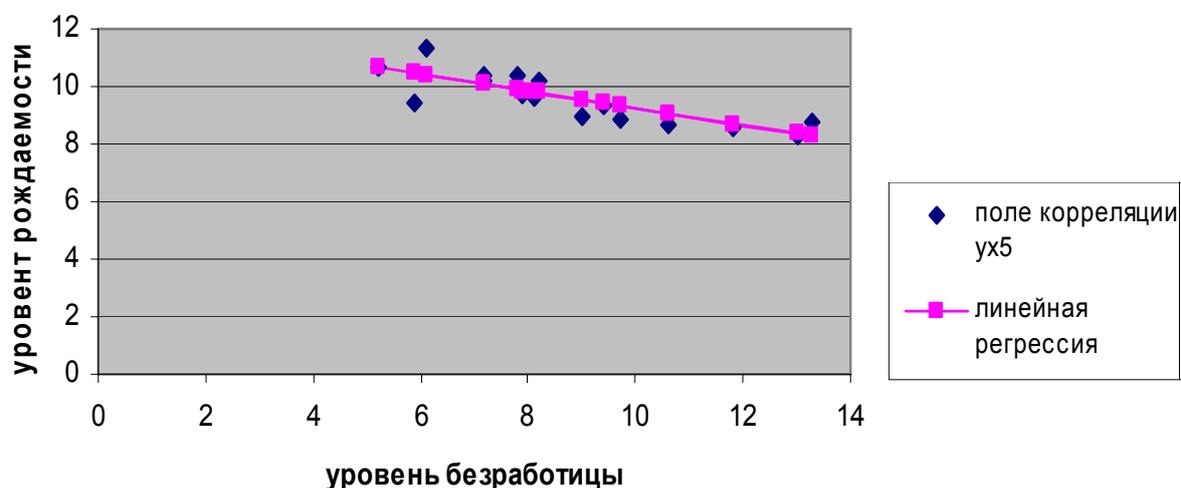
Тип модели	R^2	\bar{A}	MAD	Соств
линейная	0,7	0,04	0,37	0,5
квадратичная	0,7	0,04	0,35	0,5
гиперболическая	0,6	0,04	0,43	0,57

Проверка модели на значимость показала, что модель значима по критерию Фишера, т. к. $F_{\text{набл}} = 31,12$, а $F_{\text{табл}} = 4,6$, т. е. $F_{\text{набл}} > F_{\text{табл}}$. Также коэффициенты уравнения линейной регрессии значимы. Анализ точности прогноза показал, что средняя ошибка аппроксимации $\bar{A} = 4\%$, а среднее абсолютное отклонение $MAD = 0,37$.

Коэффициент детерминации $R^2=0,7$ говорит о том, что 70 % уровня рождаемости определяется уровнем безработицы.

Коэффициент эластичности $\Xi = -0,27$, следовательно при повышении уровня безработицы на 1 % уровень рождаемости снижается на 0,27 %.

зависимость коэф. рождаемости от уровня безработицы

Рис. 5. Корреляционное поле YX_5

Рассмотрим тесноту связи между коэффициентом рождаемости (Y) и уровнем безработицы (X_5): $r_{yx_5} = -0,85$. Приведенный коэффициент корреляции показывает, что связь между фактором X_5 и уровнем рождаемости сильная и обратная.

Также данная модель была исследована на гетероскедастичность с помощью метода Гольдфельда–Квандта. Полученные результаты показали отсутствие гетероскедастичности, это означает, что наблюдения проводились с одинаковой точностью, т. е. $F_{\text{стат}} < F_{\text{табл}}$, где $F_{\text{стат}} = 0,14$, а $F_{\text{табл}} = 6,39$.

Полученной модели можно дать следующую экономическую интерпретацию: при увеличении уровня безработицы на 1 % уровень рождаемости уменьшается на 0,3 промилле в расчете на 1000 человек.

Исследование влияния совокупности факторов

В качестве факторов, оказывающих влияние на уровень рождаемости, после предварительного исследования были отобраны 4:

- X_1 – количество браков в расчете на 1000 человек в промилле;
- X_3 – уровень дохода от предпринимательской деятельности, %;
- X_4 – ВВП, млрд долл.;
- X_5 – уровень безработицы, %.

На первоначальном этапе построения многофакторной модели была проведена проверка на наличие мультиколлинеарности (табл. 5). Как показал анализ, мультиколлинеарность в модели есть, т. к. определитель корреляционной матрицы близок к нулю.

Таблица 5

Корреляционная матрица

	у	X_1	X_3	X_4	X_5
Y	1	0,807423413	-0,546140177	0,893294112	-0,83048508
X_1	0,807423413	1	-0,217981793	0,804590096	-0,762948244
X_3	-0,546140177	-0,217981793	1	-0,557696942	0,249576181
X_4	0,893294112	0,804590096	-0,557696942	1	-0,741036255
X_5	-0,83048508	-0,762948244	0,249576181	-0,741036255	1

Полученное уравнение множественной регрессии имеет вид:

$$Y = 9,53 + 0,42 X_1 - 0,12 X_3 - 0,16 X_5.$$

Проверка модели на адекватность показала следующие результаты:

- Множественный коэффициент детерминации $R^2=0,93$ говорит о том, что около 93 % уровня рождаемости определяется факторами, введенными в модель, а именно: количеством браков, уровнем дохода от

предпринимательской деятельности и уровнем безработицы в стране. Поскольку R^2 близок к единице, можно сделать вывод, что выбранная модель удачна.

- Для точности прогноза была рассчитана средняя ошибка аппроксимации $\bar{A} = 2,3 \%$ и среднее абсолютное отклонение $MAD = 0,21$. Исходя из полученных данных, можно сделать вывод, что расчетные и фактические значения достаточно приближены друг к другу, а выбранная модель адекватна.

- Критерий Фишера $F = 28,4$, что в несколько раз превосходит его табличное значение, при $\alpha=0,05$ $F_{\text{табл}} = 3,49$. Также коэффициенты уравнения линейной регрессии по модулю превышают $t_{\text{табл}}$ по таблице распределения Стьюдента: $t_{\text{табл}}(0,05;12) = 2,18$, а $t_{\alpha} = 5,75$, $t_{\beta} = 2,5$, $t_{\gamma} = -3,34$, $t_{\delta} = -2,8$. Таким образом, рассматриваемая модель подходит для анализа.

Также модель была проверена на наличие автокорреляции. Полученные результаты показали отсутствие автокорреляции в модели.

Сравним построенную модель множественной регрессии с парными моделями (табл. 6).

Таблица 6

Сравнение моделей

Тип модели	R^2	\bar{A}	MAD	Соств
Парная $Y X_1$	0,65	0,04	0,4	0,53
Парная $Y X_3$	0,52	0,053	0,5	0,65
Парная $Y X_4$	0,8	0,035	0,33	0,4
Парная $Y X_5$	0,7	0,04	0,4	0,5
Множественная	0,93	0,023	0,21	0,34

Модель множественной регрессии несколько лучше по каждому фактору из рассмотренных показателей. Потому целесообразно рассматривать влияние данных факторов в совокупности.

ЗАКЛЮЧЕНИЕ

В ходе исследования было изучено влияние каждого из рассмотренных факторов на уровень рождаемости в отдельности и влияние всех этих факторов в совокупности. Для этого были построены пять парных регрессий и две модели множественной регрессии.

Проведенное исследование показало, что значимыми можно признать не все построенные модели. Модель, отражающая зависимость рождаемости от уровня разводов, является незначимой.

В ходе исследования было установлено, что подходящей является модель множественной регрессии:

$$Y = 9,53 + 0,42 x_1 - 0,12 x_3 - 0,16 x_5.$$

Полученной модели можно дать следующую экономическую интерпретацию:

- при увеличении числа браков в расчете на 1000 человек на 100 промилле число родившихся увеличится на 42 промилле в расчете на 1000 человек при неизменности остальных факторов;
- при увеличении дохода от предпринимательской деятельности на 1 % уровень рождаемости снижается на 0,12 промилле в расчете на 1000 человек при неизменности остальных факторов;
- при увеличении уровня безработицы на 1 % число родившихся уменьшается на 0,16 промилле при неизменности остальных факторов.

Исследование показало, что около 93 % уровня рождаемости определяется факторами, введенными в модель, а именно количеством браков, уровнем дохода от предпринимательской деятельности и уровнем безработицы в стране.

Изучать факторы снижения рождаемости необходимо, т. к. это может помочь понять и объяснить наблюдаемые тенденции, а также позволит увидеть пути выхода из сложившегося демографического кризиса и разработать эффективную демографическую политику.

ЛИТЕРАТУРА

1. *Евсеев Е. А., Буре В. М.* Эконометрика. СПб.: Изд-во МБИ, 2007. – 139 с.
2. *Тарашнина С. И., Панкратова Я. Б.* Выполнение курсовой работы по эконометрике: Учеб.-метод. пособие. СПб.: Изд-во МБИ, 2007. – 97 с.
3. Сборники государственной статистики Росстат [Электронный ресурс]: <http://www.gks.ru>

Мережинская Юлия Владимировна

Международный банковский институт

Кузютин Д.В., к.ф.-м.н., доцент; **Панкратова Я.Б.**, ст. преподаватель –
научные руководители

ИССЛЕДОВАНИЕ ВЛИЯНИЯ РЯДА ЭКОНОМИЧЕСКИХ ФАКТОРОВ НА ВЕЛИЧИНУ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ НАСЕЛЕНИЯ В СТРАНАХ ЕС

На протяжении последних десятилетий целый ряд стран столкнулся с проблемами демографии. Представители Европейского Союза переживают период стагнации, когда достаточно высокая продолжительность жизни сопровождается снижением рождаемости.

Изучение такого показателя, как ожидаемая продолжительность жизни (при рождении) чрезвычайно важно как с демографической, так и с экономической точки зрения. Показатель характеризует среднее количество лет, которое прожил бы новорожденный при условии, что в каждом возрасте условия для сохранения его жизни оставались бы такими, какими они были для соответствующей возрастной группы в год его рождения. Он отражает качество жизни в стране и может служить мерой эффективности капиталовложений в человеческий ресурс. Поскольку данный показатель рассчитывается уже на протяжении более чем 70 лет, можно сделать выводы о его объективности и значимости, ведь в большинстве случаев прогнозы оправдались. Статистическая информация регулярно обновляется, отражает реальную ситуацию и, зачастую, используется для проведения различных расчетов, сравнительных анализов и т. д. Также по данным, ежегодно публикуемым World Health Organization, можно судить об общих тенденциях развития по странам. Таким образом, изучение данного вопроса представляется еще более значимым, поскольку позволяет выявить зависимость между демографической ситуацией и социально-экономической политикой государства.

Цель данного исследования – выявить существование или отсутствие зависимости между уровнем ожидаемой продолжительности жизни населения и различными социально-экономическими факторами, отобрать из них наиболее значимые и установить вид зависимости продолжительности жизни от этих факторов.

1. Сбор данных и отбор факторов

Для проведения исследования были отобраны следующие факторы, влияющие на ожидаемую продолжительность жизни населения:

- X_1 – ВВП на душу населения по странам Европейского Союза. Величина ВВП в долларах США, приходящаяся на душу населения (на основе расчетов ВВП с учетом паритета покупательной способности) характеризует уровень экономического развития стран;

- X_2 – уровень инфляции (потребительские цены). Данный показатель характеризует изменение уровня потребительских цен (в %) по сравнению с предыдущим годом;

- X_3 – уровень безработицы. Подразумевает процент трудоспособного населения, которое активно ищет работу, но не может ее найти в данное время. Рост безработицы приводит к обострению социальной напряженности, ухудшению социально-психологического климата в обществе, также увеличение показателя приводит к ухудшению благосостояния общества;

- X_4 – социальная обеспеченность от ВВП. Данный фактор отражает, какой процент от доли ВВП страны выделяется на различные социальные нужды;

- X_5 – индекс восприятия коррупции. Данный индекс отражает данные об уровне восприятия коррупции в государственных секторах стран мира. Он представляет собой составной индекс, основанный на данных 17 различных опросов, проводимых среди экспертов по соответствующим странам. Индекс ранжирует страны по шкале от 0 до 10 баллов, причем ноль обозначает самый высокий уровень восприятия коррупции, а десять – наименьший;

- X_6 – обеспеченность кроватями в больницах на 100 000 человек. Данный фактор показывает, сколько кроватей в больницах каждой страны приходится на 100 000 человек;

- X_7 – количество занятых людей в процентах. Данный фактор показывает, сколько имеется работающего населения по сравнению со всем населением страны в совокупности;

- X_8 – количество инфарктов миокарда на 100 000 человек. Данный фактор показывает, какое количество людей в стране страдает одним из самых опасных заболеваний сердца;

- X_9 – обеспеченность врачами на 1000 человек. Данный фактор показывает, сколько врачей разных специальностей имеется в каждой стране.

В качестве результирующего признака были использованы данные по средней ожидаемой продолжительности жизни (при рождении) за 2007–2008 гг., опубликованные World Health Organization. Исследования проводились на основе данных, отражающих вариацию факторов и результирующего признака, по различным странам ЕС.

В процессе написания данной работы помимо информации, опубликованной на официальном сайте World Health Organization, также использовался официальный сайт Euro stat. В дальнейшем мы изучим влияние каждого из вышеперечисленных факторов на ожидаемый уровень продолжительности жизни в отдельности и всех факторов в совокупности, проанализируем их значимость и установим вид зависимости результирующего признака от этих факторов.

2. Исследование влияния отдельных факторов

2.1. Исследование влияния ВВП на душу населения на ожидаемую продолжительность жизни

Уровень ВВП влияет на уровень жизни людей и их продолжительность жизни. В основном по показателям уровень ВВП в странах ЕС высокий по сравнению с другими странами, т. к. он зависит от уровня развития экономики. Поскольку ожидаемая продолжительность жизни связана с понятием «качество жизни» и «экономически развитое государство», было бы логичным предположить, что с увеличением уровня ВВП в стране благосостояние всей нации в целом увеличивается, а значит, увеличивается и показатель продолжительности жизни.

Для того чтобы оценить тесноту связи фактора и результирующего признака, определяем $r_{xy} = 0,35$.

При рассмотрении поля корреляции (рис. 1) трудно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных видов зависимостей:

1. Линейная $Y = 68,7 + 0,0002 \cdot X_1$.
2. Квадратическая $Y = 59,3 + 0,0007X_1 + 0,004X_1^2$.
3. Степенная $Y = 25,58 \cdot X^{0,1}$.

Таблица 1

Модель	\bar{A}	MAD	R^2	Соств
Линейная	3,62 %	11,36541994	0,351395	3,50352016
Квадратическая	2,75 %	6,210860028	0,645557	2,64333455
Степенная	2,99 %	8,378106018	0,533997	0,04104671

Из табл. 1 можно сделать вывод, что лучшей моделью является степенная модель. Коэффициент детерминации $R^2 = 0,534$ свидетельствует о том, что 53,39 % вариации уровня ожидаемой продолжительности жизни населения по странам объясняется вариацией уровня безработицы. Остальные же 46,61 % вариации объясняются неучтенными в данной модели факторами.

Зависимость продолжительности жизни от ВВП

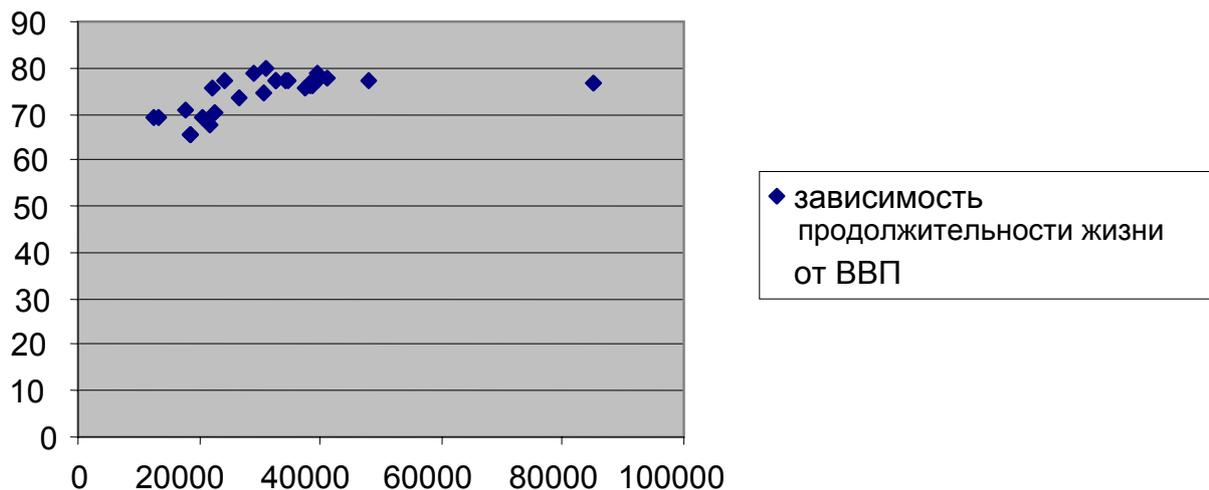


Рис. 1. Поле корреляции X_1 - Y

Кроме того, проверим статистическую значимость полученных уравнений:

- F-статистика (линейная) = 13,54.
- F-статистика (квадратическая) = 21,85.
- F-статистика (степенная) = 28,64.

Все показатели превышают по модулю критические значения распределения Фишера для вероятности 0,05 и степенями свободы 1 и 25, равное 4,24, для линейной и степенной регрессии и для той же вероятности, но со степенями свободы 1 и 24, равное 4,259, для квадратичной регрессии. Таким образом, все модели являются значимыми по критерию Фишера.

Проверив модель по тесту Спирмена, гетероскедастичности не обнаружили.

2.2. Исследование зависимости уровня инфляции потребительских цен на уровень жизни

Предположим, что с увеличением инфляции в стране благосостояние населения снижается. Естественно, что незначительный рост инфляции

может оказывать как положительный, так и отрицательный эффект на экономику, одни категории населения будут выигрывать на фоне растущей инфляции, другие – проиграть. Построим поле корреляции для фактора X_2 и результирующего признака Y .

При рассмотрении поля корреляции (рис. 2) можно предположить, что наиболее подходящей будет линейная модель, поскольку она не только проста в интерпретации, но и визуально довольно точно отражает зависимость Y от фактора X_2 .

Зависимость продолжительности жизни от уровня инфляции

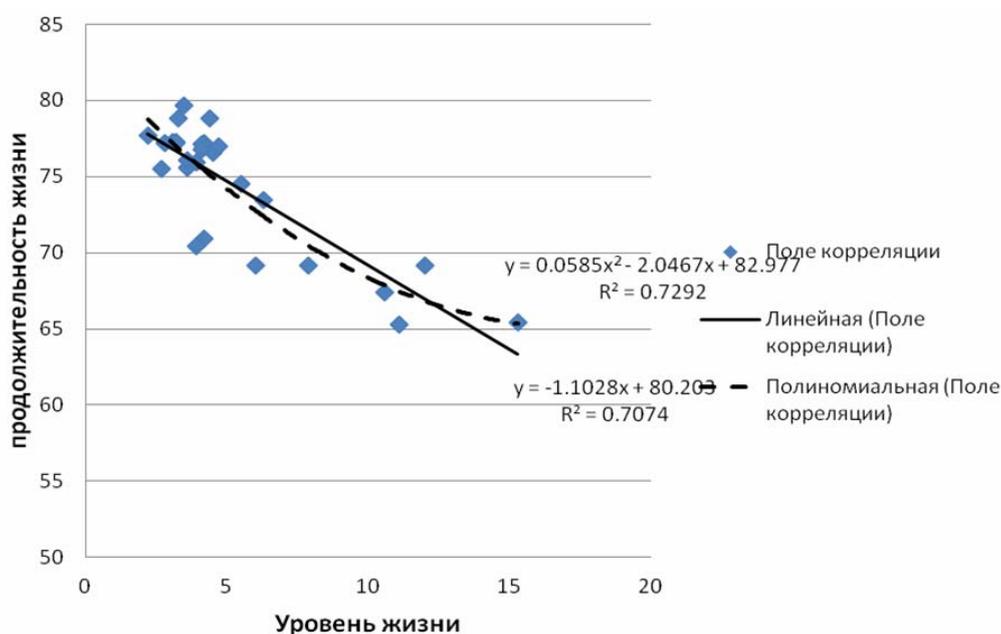


Рис. 2. Поле корреляции X_2 – Y

Для того чтобы оценить тесноту связи фактора и результирующего признака, определяем $r_{xy} = 0,7$. Построим линейное уравнение парной регрессии:

$$Y = 80,2 - 1,1 \cdot X_2 .$$

Оно показывает, что с увеличением уровня инфляции на 1 % продолжительность жизни уменьшается на 1,1 года.

Далее проведем проверку значимости построенного уравнения регрессии по критерию Фишера. F-статистика = 60,45, данный показатель превышает критическое значение распределения Фишера для вероятности 0,05 со степенями свободы 1 и 25, равное 4,24. Таким образом, математическая модель, выражающая связь между уровнем инфляции и ожидаемой продолжительностью жизни, является значимой по критерию Фишера и подходит для описания зависимости.

Согласно найденной регрессии можно сказать, что в случае, если бы в стране уровень инфляции стремился к 0, ожидаемая продолжительность жизни составила бы в среднем 80,2 лет. Дальнейшее увеличение уровня инфляции на 1 % повлечет за собой уменьшение показателя ожидаемой продолжительности жизни в среднем на год.

Коэффициент детерминации $R^2=0,7$ свидетельствует о том, что 70 % вариации уровня ожидаемой продолжительности жизни населения по странам объясняется вариацией уровня инфляции (потребительских цен) в соответствии с построенной линейной регрессией. Остальные 30 % объясняются невключенными в данную модель параметрами.

Для того чтобы оценить точность прогноза, рассчитаем среднюю ошибку аппроксимации: $\tilde{A}=2,36$ %. Полученная средняя ошибка аппроксимации не превышает рекомендуемое значение 8–10 %, а значит, приближение построенной модели к наблюдаемым статистическим значениям является хорошим.

Реальная практика показывает, что, действительно, в большинстве случаев, страны с низким уровнем инфляции имеют более высокие показатели продолжительности жизни, нежели страны с высоким уровнем. По графику мы можем наблюдать ситуацию, когда одна из стран, а именно Латвия имеет значительное отклонение от общей тенденции.

Результаты тестов свидетельствуют о том, что гетероскедастичность отсутствует, полученные выводы и интервальные оценки будут надежными.

Далее проверим гипотезу о наличии автокорреляции. Для этого проведем тест Дарбина-Уотсона. В результате исследования мы получили, что гипотеза об отсутствии автокорреляции не отвергается (автокорреляции нет).

2.3. Исследование влияния уровня безработицы на ожидаемую продолжительность жизни

Поскольку ожидаемая продолжительность жизни напрямую связана с понятием «качество жизни», было бы логичным предположить, что с увеличением уровня безработицы в стране благосостояние всей нации в целом снижается, а значит, снижается и показатель продолжительности жизни. Построим диаграмму для фактора X_3 и результирующего признака Y (см. рис. 3).

Зависимость продолжительности жизни от уровня безработицы

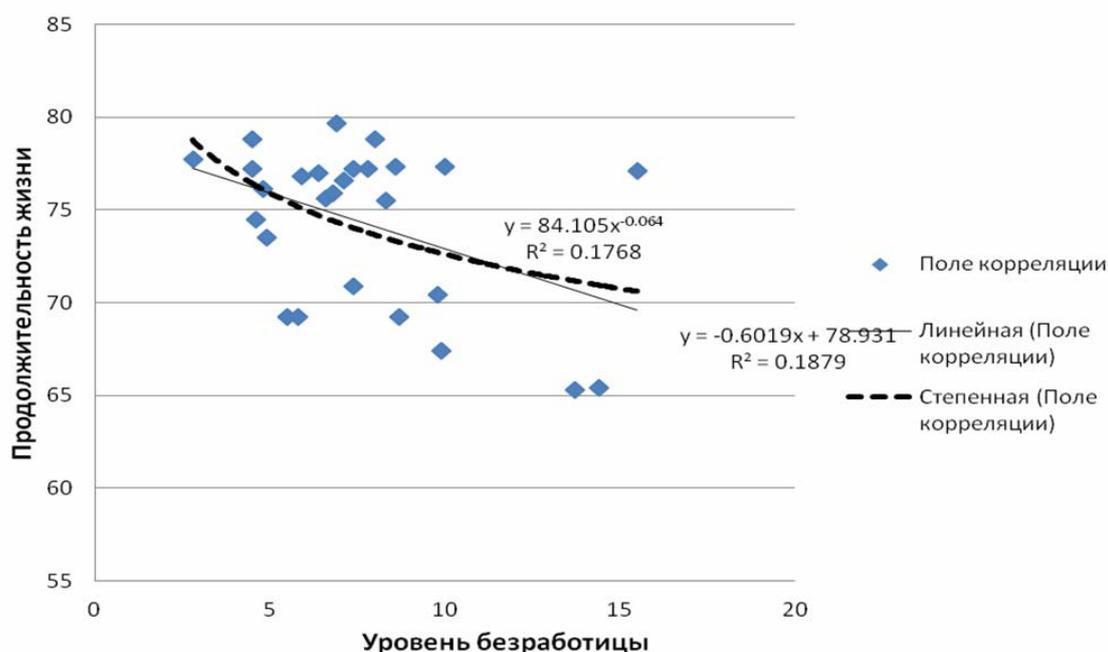


Рис. 3. Поле корреляции X_3 – Y

Для того чтобы оценить тесноту связи фактора и результирующего признака, определяем $r_{xy} = 0,17$. При рассмотрении поля корреляции трудно выявить вид зависимости, но можно выдвинуть предположение о существовании двух возможных видов зависимостей:

1. Линейная: $Y = 78,9 - 0,6 \cdot X_3$.
2. Степенная: $Y = 84,1 \cdot X_3^{-0,06}$.

Коэффициент детерминации (табл. 2) в случае линейной и степенной зависимостей отличается незначительно, однако линейная модель не только проще, она также лучше аппроксимирует наблюдаемые данные, поскольку ошибка аппроксимации в случае линейной зависимости меньше. Таким образом, на основе сравнения полученных результатов мы делаем выбор в пользу линейной модели.

Таблица 2

Модель	\bar{A}	MAD	R^2	Соцт
Линейная	4,41 %	14,22998	0,187919	3,920253
Степенная	4,57 %	14,58587	0,176808	0,054555

Эконометрическая интерпретация выбранной модели: с увеличением безработицы в стране на 1 % продолжительность жизни уменьшается на 7 месяцев.

Коэффициент детерминации $R^2 = 0,187$ свидетельствует о том, что 18,7 % вариации уровня ожидаемой продолжительности жизни населения по странам объясняется вариацией уровня безработицы. Остальные же 81,3 % вариации объясняются неучтенными в данной модели факторами.

Поскольку в ходе исследования мы сделали выбор в пользу линейной регрессии, то на следующем этапе необходимо сделать проверку коэффициентов на значимость:

$t_b = -2,40522$	$t_a = 38,35083$	$t_{\text{табл}} = 2,059539$
------------------	------------------	------------------------------

Обе статистики превышают по модулю критическое значение распределения Стьюдента для вероятности 0,05 и 25 степенью свободы, равное 2,06, т. е. коэффициенты a и b являются значимыми.

Теоретически можно представить себе идеальную ситуацию, когда в стране нулевой уровень безработицы и все население занято, в таком случае ожидаемая продолжительность жизни составила бы в среднем 78,9 лет. С другой стороны, увеличение уровня безработицы на 1 % повлечет за собой уменьшение показателя ожидаемой продолжительности жизни на 0,6 года.

Реальная практика показывает, что, действительно, в большинстве случаев, высокоразвитые страны с низким уровнем безработицы имеют более высокие показатели продолжительности жизни, нежели страны с высоким уровнем безработицы.

2.4. Исследование влияния социальной обеспеченности от ВВП на продолжительность жизни

Уровень социальной обеспеченности от ВВП влияет на уровень жизни людей и их продолжительность жизни. В основном по показателям социальная обеспеченность в странах ЕС высокая по сравнению с другими странами, т. к. уровень социальной обеспеченности зависит от уровня развития экономики и социальной политики государства. Построим диаграмму для фактора X_4 и результирующего признака Y .

Для того чтобы оценить тесноту связи фактора и результирующего признака, определяем $r_{xy} = 0,55$.

При рассмотрении поля корреляции (рис. 4) трудно точно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных видов зависимостей:

1. Линейная $Y = 62,95 + 0,52 \cdot X_4$.
2. Квадратическая $Y = 42,84 + 2,5 \cdot X_4 - 0,05 \cdot X_4^2$.

Зависимость продолжительности жизни от социальной обеспеченности

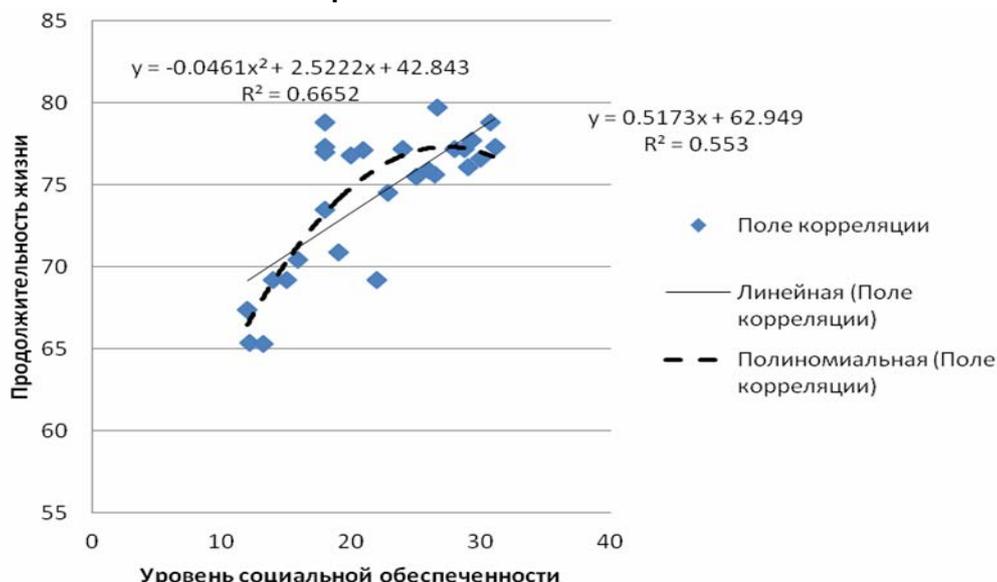


Рис. 4. Поле корреляции X_4-Y

Лучшей по всем показателям (табл. 3) оказалась линейная модель. Она показывает, что с увеличением социальной обеспеченности от ВВП продолжительность жизни увеличивается на полгода.

Таблица 3

Модель	\bar{A}	MAD	R^2	Соств
Линейная	2,94%	7,833221	0,552971	2,908587
Квадратическая	2,42%	5,866992	0,665181	2,569118

Коэффициент детерминации $R^2=0,55$ свидетельствует о том, что 55 % вариации уровня ожидаемой продолжительности жизни населения по странам объясняется вариацией уровня социальной обеспеченности населения. Остальные же 45 % вариации объясняются неучтенными в данной модели факторами.

Кроме того, проверим статистическую значимость полученных уравнений:

- F-статистика (линейная) = 30,92.
- F-статистика (квадратическая) = 23,84.

Все показатели превышают по модулю критические значения распределения Фишера для вероятности 0,05 и степенями свободы 1 и 25, равное 4,24, для линейной и степенной регрессии и для той же вероятности, но со степенями свободы 1 и 24, равное 4,259, для квадратической. Таким образом, все модели являются значимыми по критерию Фишера.

Проверив модель по тесту Спирмена, гетероскедастичности не обнаружилось и автокорреляция отсутствует.

2.5. Исследование влияния уровня восприятия коррупции на продолжительность жизни

Индекс восприятия коррупции отражает данные об уровне восприятия коррупции в государственных секторах стран мира. Он ранжирует страны по шкале от 0 до 10 баллов, причем ноль обозначает самый высокий уровень восприятия коррупции, а десять – наименьший. Известно, что чем менее коррумпирован государственный сектор, тем более эффективно он осуществляет свою деятельность. Население стран «коррумпированных», согласно данному индексу, относится с недоверием к любому виду государственных структур и, зачастую, предпочитает справляться со своими проблемами собственными силами. В данных странах наблюдается рост уровня преступности, хронической заболеваемости населения, увеличивается потребление алкоголя и табачных изделий, а количество аборт в среднем составляет 37 % от общего числа новорожденных.

Логично было бы предположить, что чем больше данный показатель (в расчете на то, что 10 означает наименьший уровень коррупции), тем выше уровень ожидаемой продолжительности жизни.

Построим диаграмму для фактора X_5 и результирующего признака Y (рис. 5).

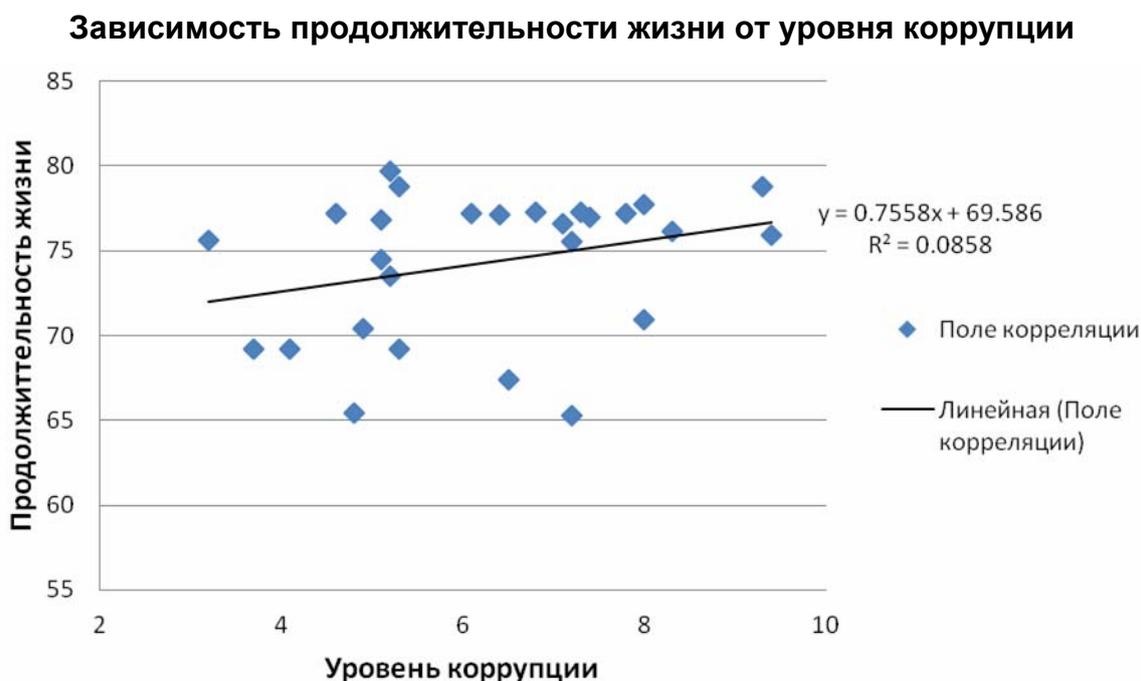


Рис. 5. Поле корреляции X_5 – Y

Для того чтобы оценить тесноту связи фактора и результирующего признака, определяем $r_{xy} = 0,08$. При рассмотрении поля корреляции трудно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных видов зависимостей:

1. Линейная $Y = 69,58 + 0,75 \cdot X_5$.
2. Квадратическая $Y = 70,83 + 0,33 \cdot X_5 + 0,033 \cdot X_5^2$.
3. Степенная $Y = 66,6 \cdot 0,75 \cdot X_5^{0,06}$.

Так как построив все 3 модели, мы увидели, что все они незначимы, мы не будем включать фактор X_5 в модель множественной регрессии.

2.6. Исследование влияния обеспеченности кроватями в больницах на 100 000 человек на продолжительность жизни

Уровень обеспеченности кроватями в больницах напрямую связан с продолжительностью жизни человека. В основном по показателям уровень обеспеченности кроватями в больницах в странах ЕС высокий по сравнению с другими странами, т. к. он зависит от уровня развития здравоохранения и социальной политики государства. Построим диаграмму для фактора X_6 и результирующего признака Y .

Для того чтобы оценить тесноту связи фактора и результирующего признака, определяем $r_{xy} = 0,83$. При рассмотрении поля корреляции (рис. 6) трудно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных видов зависимостей:

1. Линейная $Y = 56,86 + 0,03 \cdot X_6$.
2. Квадратическая $Y = 57,68 + 0,02 \cdot X_6 + 0,00247 X_6^2$.
3. Степенная $Y = 20,1 \cdot X_6^{0,2}$.

Таблица 4

Модель	\bar{A}	MAD	R^2	Сост
Линейная	2,01 %	2,924274	0,833117	1,777137
Квадратическая	2,02 %	2,921306	0,833286	1,812862
Степенная	2,02 %	3,264226	0,806672	0,026438

По результатам табл. 4 лучшей моделью была выбрана модель линейной регрессии, которая показывает, что с увеличением количества кроватей в больницах на 1 % продолжительность жизни увеличивается на 0,03 года.

Зависимость продолжительности жизни от обеспеченности кроватями на 100 000 человек

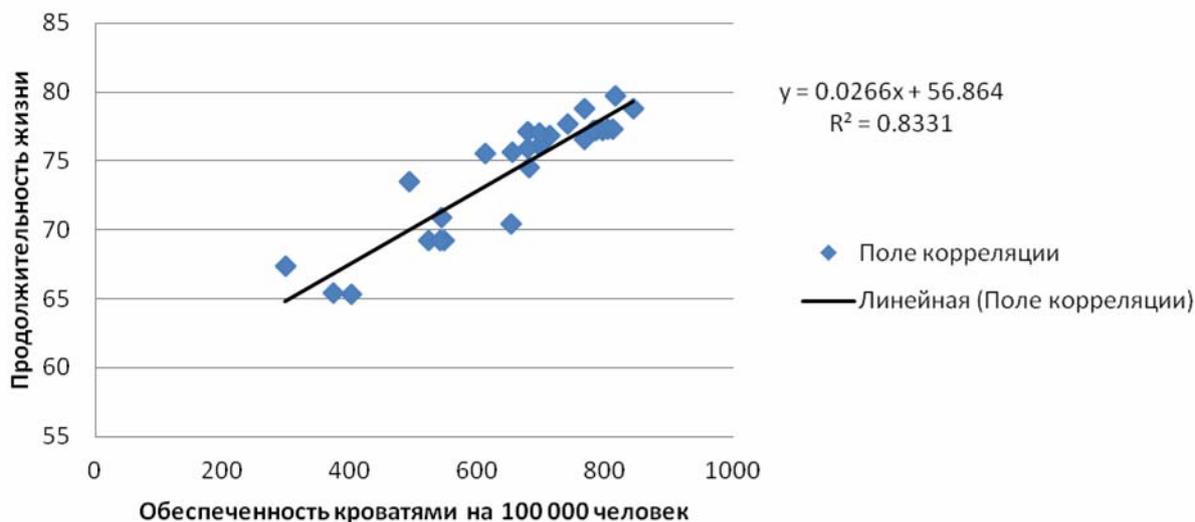


Рис. 6. Поле корреляции X_6 – Y

Коэффициент детерминации $R^2=0,83$ свидетельствует о том, что 83 % вариации уровня ожидаемой продолжительности жизни населения по странам объясняется вариацией уровня обеспеченностью кроватями. Остальные же 17 % вариации объясняются неучтенными в данной модели факторами.

Кроме того, проверим статистическую значимость полученных уравнений:

- F-статистика (линейная) = 124,8.
- F-статистика (квадратическая) = 59,97.
- F-статистика (степенная) = 104,31.

Все показатели превышают по модулю критические значения распределения Фишера при уровне значимости 0,05, равное 4,24, для линейной и степенной регрессии и 4,259 – для квадратической регрессии. Таким образом, все модели являются значимыми по критерию Фишера.

Проверив модель по тесту Спирмена, гетероскедастичности не обнаружилось и автокорреляция отсутствует.

2.7. Исследование влияния количества занятого населения на продолжительность жизни

На сегодняшний день показатель занятости и платежеспособности населения сильно связан со здоровьем самого человека, т. к. некоторые жизненно важные медицинские услуги дорого стоят.

Для того чтобы оценить тесноту связи фактора и результирующего признака, определяем $r_{xy}=0,1$. При рассмотрении поля корреляции (рис. 7)

трудно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных видов зависимостей:

1. Линейная $Y = 66,98 + 0,15 \times X_7$.
2. Квадратическая $Y = 51,31 + 0,82 \times X_7^2$.
3. Степенная $Y = 48,57 \times X_7^{0,11}$.

Зависимость продолжительности жизни от количества работающего населения

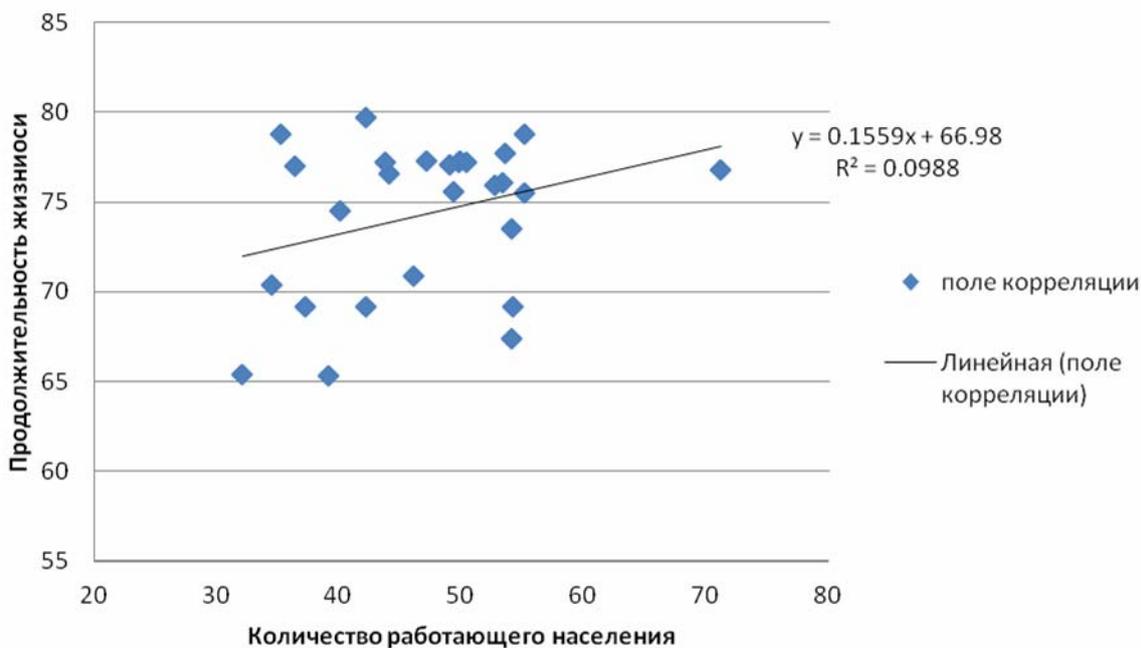


Рис. 7. Поле корреляции $X_7 - Y$

При исследовании каждой модели получилось, что они все статистически незначимы.

2.8. Исследование влияния количества инфарктов миокарда на продолжительность жизни населения

На сегодняшний день из-за заболевания сердца умирает много людей различного возраста, и исследование самого частого заболевания показалось нам влиятельным фактором.

Для того чтобы оценить тесноту связи фактора и результирующего признака, определяем $r_{xy} = 0,04$.

При рассмотрении поля корреляции (рис. 8) трудно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных видов зависимостей:

**Зависимость продолжительности жизни
от количества инфарктов миокарда на 100 000 человек**

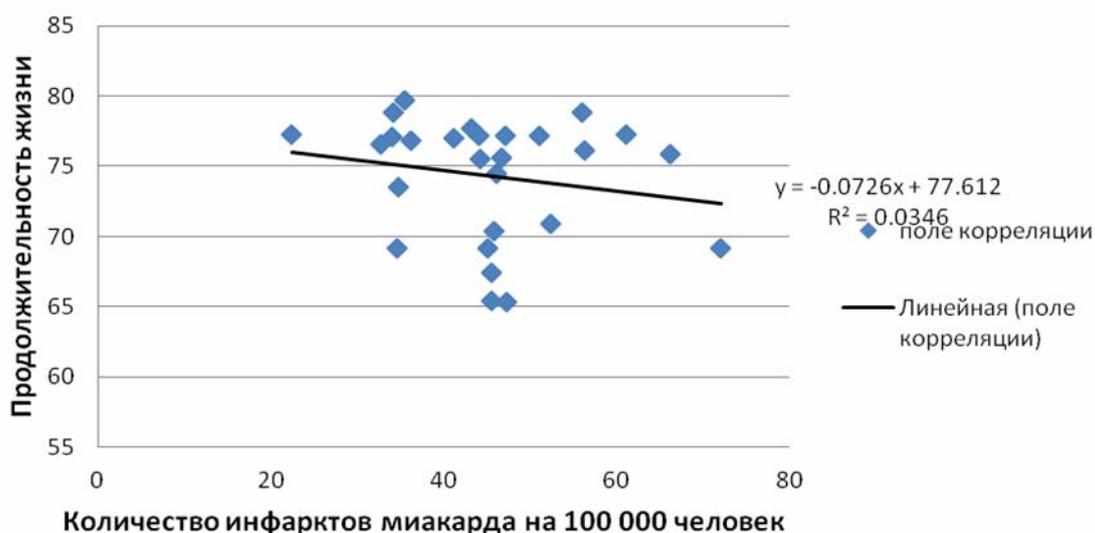


Рис. 8. Поле корреляции X_8 – Y

1. Линейная $Y = 77,6 - 0,07 \cdot X_8$.
2. Квадратическая $Y = 84,5 - 0,37 \cdot X_8 + 0,003 \cdot X_8^2$.
3. Степенная $Y = 89,3 \cdot X_8^{-0,05}$.

При исследовании каждой модели получилось, что они все статистически незначимы.

**2.9. Исследование влияния количества врачей на 1000 человек
на продолжительность жизни**

Количество врачей напрямую связано с продолжительностью жизни человека. На сегодняшний день в ЕС самый высокий уровень обеспеченности врачами в Швеции и, как следствие, продолжительность жизни там тоже высокая.

Для того чтобы оценить тесноту связи фактора и результирующего признака, определяем $r_{xy} = 0,55$. При рассмотрении поля корреляции (рис. 9) трудно точно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных видов зависимостей:

1. Линейная $Y = 59,47 + 4,06 \cdot X_9$.
2. Квадратическая $Y = 35,02 + 17,84 \cdot X_9 - 0,003 \cdot X_9^2$.
3. Степенная $Y = 56,97 \cdot X_9^{0,2}$.

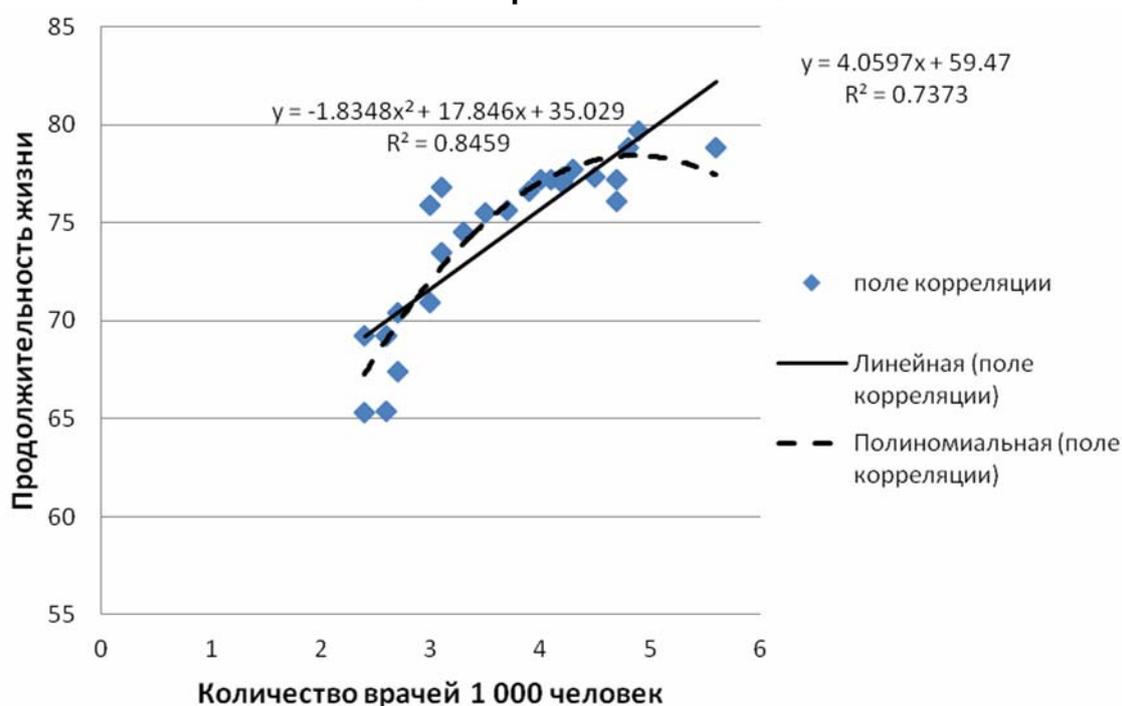
При исследовании всех моделей наиболее значимой по всем показателям (табл. 5) оказалась квадратическая модель.

Таблица 5

Модель	\bar{A}	MAD	R^2	Соств
Линейная	2,21%	4,603528	0,737284	2,229756
Квадратическая	1,60%	2,699868	0,845923	1,7428
Степенная	1,96%	3,794368	0,783669	0,027967

Коэффициент детерминации $R^2=0,84$ свидетельствует о том, что 84 % вариации уровня ожидаемой продолжительности жизни населения по странам объясняется вариацией уровня обеспеченности врачами. Остальные же 16 % вариации объясняются неучтенными в данной модели факторами.

Зависимость продолжительности жизни от количества врачей на 1000 человек

Рис. 9. Поле корреляции X_9 – Y

Кроме того, проверим статистическую значимость полученных уравнений:

- F-статистика (линейная) = 70,15.
- F-статистика (квадратическая) = 65,88.
- F-статистика (степенная) = 90,56.

Все показатели превышают по модулю критические значения распределения Фишера для уровня значимости 0,05, равное 4,24, для линейной и степенной регрессии и 4,259 – для квадратической. Таким образом, все модели являются значимыми по критерию Фишера.

Поскольку в ходе исследования мы сделали выбор в пользу квадратической регрессии, то на следующем этапе необходимо сделать проверку коэффициентов на значимость:

$t_a = 5,732893$	$t_{b1} = 5,291324$	$t_{b2} = -4,11366$	$T_{\text{табл}} = 2,063899$
------------------	---------------------	---------------------	------------------------------

Все статистики превышают по модулю критическое значение распределения Стьюдента для вероятности 0,05 и 25 степенью свободы, равное 2,05, следовательно коэффициенты являются значимыми.

Проверив модель по тесту Спирмена, гетероскедастичности не обнаружилось и автокорреляция отсутствует.

3. Построение многофакторной модели

В процессе исследования мы выяснили, что только 3 фактора из рассматриваемых статистически значимы и оказывают влияние на уровень ожидаемой продолжительности жизни.

- X_2 – уровень инфляции (потребительские цены).
- X_6 – обеспеченность кроватями в больницах на 100 000 человек.
- X_9 – обеспеченность врачами на 1000 человек.

Рассчитанные парные коэффициенты корреляции представим в виде корреляционной матрицы (табл. 6).

Таблица 6

Q	Y	X ₂	X ₆	X ₉
Y	1	-0,8410913	0,9127522	0,85865272
X ₂	-0,8410913	1	-0,8062562	-0,645784869
X ₆	0,9127522	-0,8062562	1	0,813520307
X ₉	0,8586527	-0,6457849	0,8135203	1

Из анализируемых факторов наибольшее влияние на уровень ожидаемой продолжительности жизни оказывает шестой фактор – обеспеченность кроватями в больницах на 100 000 человек. Коэффициент корреляции между объясняющими переменными X_2X_6 , X_6X_9 больше 0,7, что говорит о мультиколлинеарности.

Построим модель множественной регрессии:

$$Y_{X_2X_9} = 63,6 - 0,39 \cdot X_2 + 0,02 \cdot X_9 .$$

Экономическая интерпретация:

- Каждый дополнительный процент инфляции в стране снижает среднюю продолжительность жизни на 0,4 года.

- Увеличение количества медработников на единицу увеличивает продолжительность жизни на 0,02 года.

Проверив модель, гетероскедастичность и автокорреляция не обнаружилась.

Теперь проведем сравнение построенной модели множественной регрессии с моделями парных регрессий. Это сравнение производится для того, чтобы осуществить выбор между моделями (табл. 7).

Таблица 7

Модель	\bar{A}	MAD	R^2	Soct
Множественная	16,97 %	12,7536	0,84709	1,63
yx_2	2,36 %	5,12658	0,70744	2,35302
yx_6	2,01 %	2,92427	0,83312	1,77714
yx_9	2,21 %	4,60353	0,73728	2,22976

Модель множественной регрессии незначительно лучше парных линейных регрессий, поэтому выбор был сделан в пользу парной линейной модели YX_6 .

ЗАКЛЮЧЕНИЕ

В процессе исследования было изучено влияние каждого из приведенных выше факторов на результирующий признак – ожидаемый уровень продолжительности жизни как в отдельности, так и всех факторов в совокупности, была проанализирована их значимость и был установлен вид зависимости результирующего признака от этих факторов. Для этого были построены четыре парные регрессионные модели и одна множественная регрессия. При этом в ходе исследования выяснилось, что в ряде случаев нелинейная регрессия лучше отображает имеющуюся зависимость, нежели линейная. Однако несмотря на то, что полученные линейные и степенные модели и модель множественной регрессии были несколько лучше по каждому из рассмотренных показателей, в конечном итоге выбор был сделан в пользу модели линейной регрессии как наиболее подходящей с точки зрения проведения на ее основе дальнейших исследований и принятия решений.

В процессе исследования выяснилось, что при пошаговом отборе переменных для построения множественной регрессии включение факторов

ВВП на душу населения, уровень безработицы, социальную обеспеченность, коррупцию, количество занятого населения, количество инфарктов миокарда является неоправданным.

Итак, в ходе исследования было установлено, что наиболее подходящей является модель, отражающая зависимость уровня ожидаемой продолжительности жизни от уровня обеспеченности кроватями.

ЛИТЕРАТУРА

1. *Айвазян С.А., Мхитарян В.С.* Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998.
2. *Евсеев Е.А., Буре В.М.* Эконометрика. СПб.: Изд-во МБИ, 2007. – 139 с.
3. *Макконел Р., Брю С.Л.* Экономикс. М.: Дрофа, 1991.
4. *Тарашина С.И., Панкратова Я.Б.* Выполнение курсовой работы по эконометрике: Учеб.-метод. пособие. СПб.: Изд-во МБИ, 2007. – 97 с.
5. Официальный сайт Европейской статистики [Электронный ресурс]: www.eurostat/eu.com

Моисеев Кирилл Яковлевич

Международный банковский институт

Кузютин Д.В., к.ф.-м.н., доцент; **Евсеев Е.А.**, к.ф.-м.н., доцент –
научные руководители

ВЗАИМОСВЯЗЬ ПРОДОЛЖИТЕЛЬНОСТИ ЖИЗНИ НАСЕЛЕНИЯ И СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ СТРАН МИРА

Одним из индикаторов уровня жизни населения той или иной страны является продолжительность жизни населения. Часто ее связывают с уровнем ВВП страны. Во многих случаях это оправдано, но, например, размер ВВП США в 2007 году был в 17 раз больше, чем у Австралии, а ожидаемая продолжительность жизни в США оказалась на три года меньше. Становится ясно, что за длительностью жизни стоит гораздо больше факторов. Во многих случаях показатель продолжительности жизни зависит не только от уровня экономического развития стран. Кроме экономических факторов воздействуют также различные виды заболеваний, физическое состояние населения, пристрастия к алкоголю и сигаретам и многие другие показатели.

Целью данной работы является выявление и анализ экономических и социальных факторов, влияющих на ожидаемую продолжительность жизни в странах мира. С помощью эконометрических методов анализа была сделана попытка понять и продемонстрировать другим, на что нужно воздействовать для повышения продолжительности жизни.

1. Сбор данных и отбор факторов

Изначально рассматривалось влияние 15 факторов (таких как потребление алкоголя, курение сигарет, уровень ВВП стран, ожирение населения, раковые заболевания, диабет) на ожидаемую продолжительность жизни, но из-за слишком маленького коэффициента корреляции и их незначимости по тестам Фишера и Стьюдента они были отброшены.

Для проведения исследования были отобраны следующие факторы, влияющие на ожидаемую продолжительность жизни в странах мира:

- X_1 – уровень расходов на здравоохранение в стране на душу населения (ППС – паритет покупательной способности, \$);
- X_2 – процент работоспособного населения;
- X_3 – количество смертельных случаев на 100 000 человек в результате инфаркта;

- X_4 – количество практикующих врачей на 1000 человек;
- X_5 – уровень затрат на фармацевтику (ППС, \$).

В качестве результирующего признака (ожидаемая продолжительность жизни в стране) были использованы данные по ожидаемой продолжительности жизни по каждой отдельно взятой стране. В исследовании использовались страны-представители всех континентов Земли, имеющие совершенно разные культурные характеристики и экономические положения. В связи с отсутствием данных по многим странам выборка состоит из 30 стран. В исследовании использовались данные ОЭСР (Организация экономического сотрудничества и развития) за 2007 год.

Далее будет изучено влияние каждого из приведенных выше факторов на продолжительность жизни в странах по отдельности и влияние этих факторов в совокупности.

2. Исследование влияния отдельных факторов

2.1. Исследование влияния уровня расходов на здравоохранение на ожидаемую продолжительность жизни

Предположим, что с увеличением затрат на здравоохранение продолжительность жизни будет расти. Для изучения влияния фактора X_1 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 1).

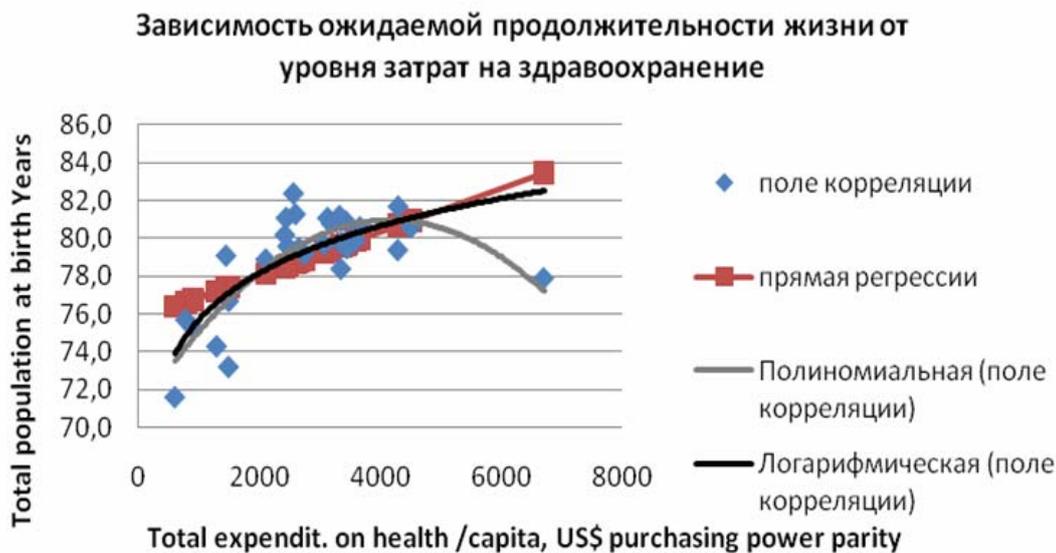


Рис. 1

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому для начала построим наиболее простую модель – линейную:

$$Y = 75,7 + 0,0011 X_1.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{yx_1} = 0,568$. Согласно шкале Чеддока, такая связь изучаемых признаков является заметной.

Коэффициент детерминации $R^2 = 0,32$ говорит о том, что около 32 % вариации продолжительности жизни объясняется вариацией уровня затрат на здравоохранение. Остальные 68 % вариации объясняются неучтенными в данной модели факторами. Скорректированный коэффициент детерминации $\bar{R}^2 = 0,29$.

Средняя ошибка аппроксимации составила 2,02 %, что свидетельствует о хорошем соответствии модели эмпирическим данным.

Данная модель является значимой по критерию Фишера, так как $F_{набл} > F_{табл}$, где $F_{набл} = 13,3$, а $F_{табл} = 4,19$.

Также можно рассмотреть и нелинейные виды зависимости.

- Полиномиальная зависимость:

$$\hat{Y} = 70,7 + 0,04 \cdot X_1 - 6 \cdot 10^{-7} X_1^2;$$

$R^2 = 0,7$ – коэффициент детерминации;

$A = 1,4$ % – средняя ошибка аппроксимации.

- Линейно-логарифмическая зависимость. Тогда:

$$\hat{Y} = 50,8 + 3,6 \ln X_1;$$

$R^2 = 0,55$ – коэффициент детерминации;

$A = 1,41$ % – средняя ошибка аппроксимации.

Хотя квадратичная модель обладает самым большим коэффициентом детерминации, ее применение нелогично, так как от роста уровня затрат на здравоохранение не может быть снижения продолжительности жизни. В данном случае больше подходит линейно-логарифмическая зависимость, которая показывает замедление скорости роста продолжительности жизни от увеличения уровня затрат, что является логичным.

Таким образом, исследовав три модели, можно сделать вывод, что наиболее подходящей для исследования является линейно-логарифмическая зависимость. Включение данного фактора в модель множественной регрессии является целесообразным.

Для данной модели подходит следующая интерпретация: при увеличении уровня затрат на здравоохранение на 1 % ожидаемая продолжительность жизни вырастет на 0,036 года.

2.2. Исследование влияния числа работоспособных граждан на ожидаемую продолжительность жизни

Для изучения влияния фактора X_2 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 2).

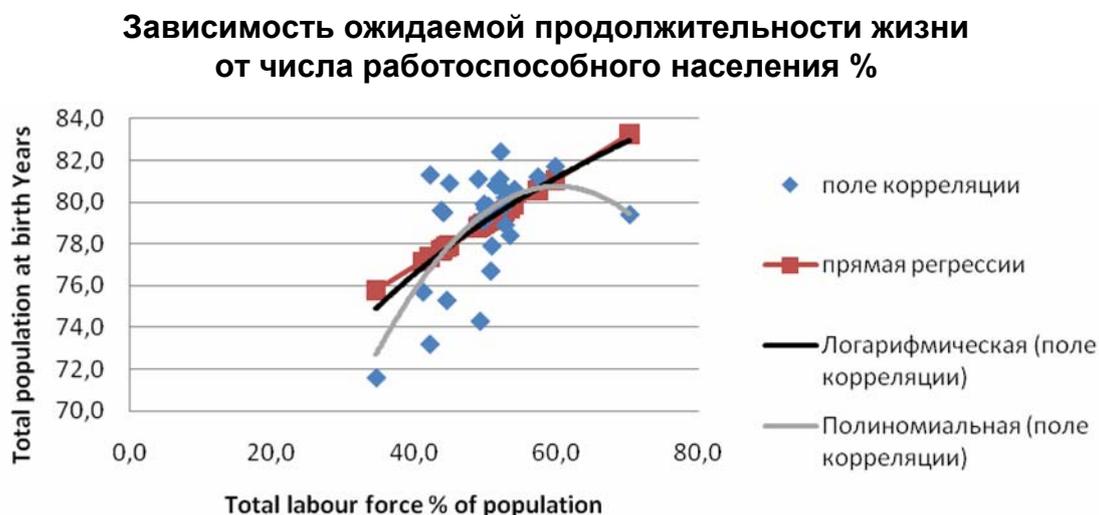


Рис. 2

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому сначала построим наиболее простую модель – линейную

$$\hat{Y} = 68,5 + 0,2 X_2.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{yx_2} = 0,51$, согласно шкале Чеддока, такая связь изучаемых признаков является заметной. Коэффициент детерминации $R^2 = 0,27$ говорит о том, что около 27 % вариации продолжительности жизни объясняется вариацией числа работоспособных граждан. Остальные 73 % вариации объясняются неучтенными в данной модели факторами.

Данная модель является значимой по критерию Фишера, т. к. $F_{набл} > F_{табл}$, где $F_{набл} = 10,2$, а $F_{табл} = 4,19$, следовательно основная гипотеза о том, что все коэффициенты уравнения равны нулю, отклоняется. Следовательно, данная модель является значимой.

Средняя ошибка аппроксимации составила 2,26 %, что свидетельствует о хорошем соответствии модели эмпирическим данным.

Также можно рассмотреть и нелинейные виды зависимости.

- Полиномиальная зависимость. Тогда:

$$\hat{Y} = 35,7 + 1,5X_2 - 0,012 \cdot X_2^2;$$

$$R^2 = 0,42 \text{ – коэффициент детерминации;}$$

$A = 1,8\%$ – средняя ошибка аппроксимации.

- Линейно-логарифмическая зависимость. Тогда:

$$\hat{Y} = 34,48 + 11,39 \ln X_2;$$

$R^2 = 0,316$ – коэффициент детерминации;

$A = 17,24\%$ – средняя ошибка аппроксимации.

Таким образом, исследовав три модели, можно сделать вывод, что наиболее подходящей для исследования является линейно-логарифмическая зависимость. Включение данного фактора в модель множественной регрессии является целесообразным.

Для данной модели подходит следующая интерпретация: при увеличении числа рабочей силы на 1 % ожидаемая продолжительность жизни вырастет на 0,11 года.

2.3. Исследование влияния количества смертельных случаев в результате инфаркта на ожидаемую продолжительность жизни

Для изучения влияния фактора X_3 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 3).

Зависимость ожидаемой продолжительности жизни от количества смертельных случаев от инфаркта

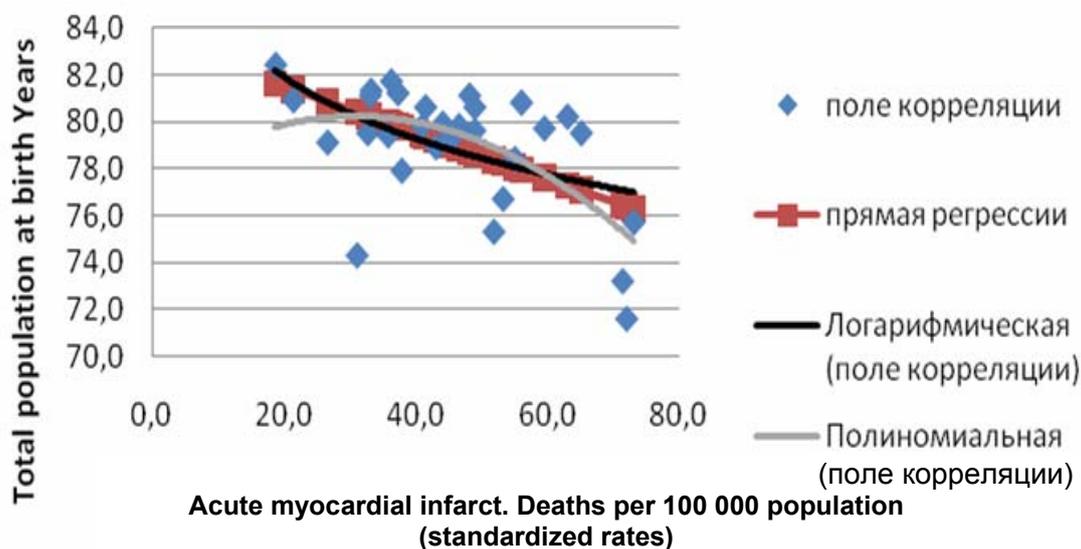


Рис. 3

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную:

$$\hat{Y} = 83,5 - 0,1 X_3.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации.

Коэффициент корреляции $r_{yx3} = 0,54$, согласно шкале Чеддока, такая связь изучаемых признаков является заметной. Коэффициент детерминации $R^2 = 0,295$ говорит о том, что около 29,5 % вариации продолжительности жизни объясняется вариацией числа смертей от инфаркта. Остальные 70,5 % вариации объясняются неучтенными в данной модели факторами.

Данная модель является значимой по критерию Фишера, т. к. $F_{набл} > F_{табл}$, где $F_{набл} = 11,7$, а $F_{табл} = 4,19$, следовательно основная гипотеза о том, что все коэффициенты уравнения равны нулю, отклоняется. Следовательно, данная модель является значимой.

Средняя ошибка аппроксимации составила 2,16 %, что свидетельствует о хорошем соответствии модели эмпирическим данным.

Также можно рассмотреть и нелинейные виды зависимости.

- Полиномиальная зависимость. Тогда:

$$\hat{Y} = 77,4 + 0,183X_3 - 0,003 \cdot X_3^2;$$

$$R^2 = 0,37 - \text{коэффициент детерминации};$$

$$A = 2 \% - \text{средняя ошибка аппроксимации}.$$

Проверим статистическую значимость коэффициентов полиномиальной модели:

$$t_a = 21,2;$$

$$t_{b1} = 1,14;$$

$$t_{b2} = -1,78.$$

Мы можем наблюдать, что t-статистики коэффициентов b_1 и b_2 не превышают критическое значение распределения Стьюдента с вероятностью 0,05 и 27 степенями свободы, равное 2,05. Отсюда можно сделать вывод, что коэффициенты незначимы.

- Линейно-логарифмическая зависимость. Тогда:

$$\hat{Y} = 93,27 - 3,79 \ln X_3 - \text{уравнение};$$

$$R^2 = 0,243 - \text{коэффициент детерминации};$$

$$A = 17,76\% - \text{средняя ошибка аппроксимации}.$$

Таким образом, исследовав три модели, можно сделать вывод, что наиболее подходящей для исследования является линейно-логарифмическая зависимость. Включение данного фактора в модель множественной регрессии является целесообразным.

Для данной модели подходит следующая интерпретация: при увеличении количества смертельных случаев от инфаркта на 1 % ожидаемая продолжительность жизни уменьшается на 0,0379 года.

2.4. Исследование влияния количества практикующих врачей на ожидаемую продолжительность жизни

Для изучения влияния фактора X_4 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 4).

Зависимость ожидаемой продолжительности жизни от количества практикующих врачей на 1000 человек

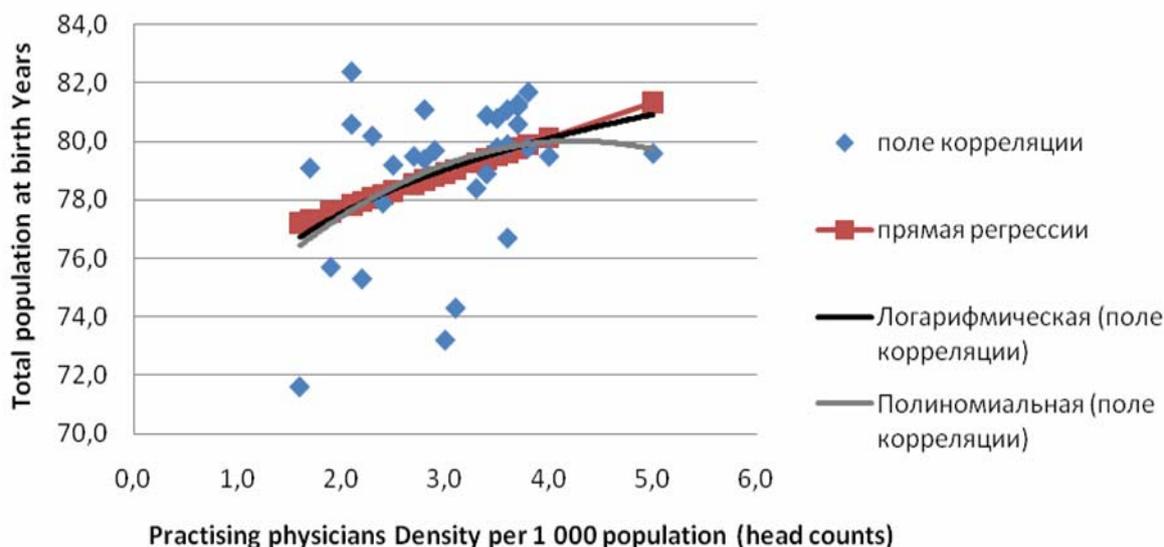


Рис. 4

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную

$$\hat{Y} = 75,2 + 1,2X_4.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{YX_4} = 0,36$, согласно шкале Чеддока, такая связь изучаемых признаков является заметной. Коэффициент детерминации $R^2 = 0,13$ говорит о том, что около 13 % вариации продолжительности жизни объясняется вариацией числа практикующих врачей. Остальные 87 % вариации объясняются неучтенными в данной модели факторами.

Данная модель является значимой по критерию Фишера, т. к. $F_{\text{набл}} > F_{\text{табл}}$, где $F_{\text{набл}} = 4,37$, а $F_{\text{табл}} = 4,19$, следовательно основная гипотеза о том, что все коэффициенты уравнения равны нулю, отклоняется. Следовательно, данная модель является значимой.

Средняя ошибка аппроксимации составила 2,37 %, что свидетельствует о хорошем соответствии модели эмпирическим данным.

Также можно рассмотреть и нелинейные виды зависимости.

- Полиномиальная зависимость. Тогда:

$$\hat{Y} = 70,83 + 4,313X_4 - 0,506 \cdot X_4^2;$$

$$R^2 = 0,16 - \text{коэффициент детерминации};$$

$$A = 2,2 \% - \text{средняя ошибка аппроксимации}.$$

Проверим статистическую значимость коэффициентов полиномиальной модели:

$$t_a = 13,118;$$

$$t_{b1} = 1,19;$$

$$t_{b2} = -0,87.$$

Мы можем наблюдать, что t-статистики коэффициентов b_1 и b_2 не превышают критическое значение распределения Стьюдента с вероятностью 0,05 и 27 степенями свободы, равное 2,05. Отсюда можно сделать вывод, что коэффициенты незначимы.

- Линейно-логарифмическая зависимость. Тогда:

$$\hat{Y}_X = 74,98 + 3,691 \ln X_4;$$

$$R^2 = 0,153 - \text{коэффициент детерминации};$$

$$A = 29,63\% - \text{средняя ошибка аппроксимации}.$$

Таким образом, исследовав три модели, можно сделать вывод, что наиболее подходящей для исследования является линейная зависимость. Включение данного фактора в модель множественной регрессии является целесообразным.

Для данной модели подходит следующая интерпретация: при увеличении количества практикующих врачей на 1 человека ожидаемая продолжительность жизни увеличится на 1,2 года.

2.5. Исследование влияния уровня фармацевтических расходов на ожидаемую продолжительность жизни

Для изучения влияния фактора X_5 на результирующий признак Y необходимо сначала построить поле корреляции (см. рис. 5).

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную

$$\hat{y}_x = 75,8 + 0,07x_5.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{yx_5} = 0,4$. Согласно шкале Чеддока, такая связь изучаемых признаков является слабой.

Коэффициент детерминации $R^2 = 0,17$ говорит о том, что около 17 % вариации продолжительности жизни объясняется вариацией затрат на фармацевтику. Остальные 83 % вариации объясняются неучтенными в данной модели факторами.

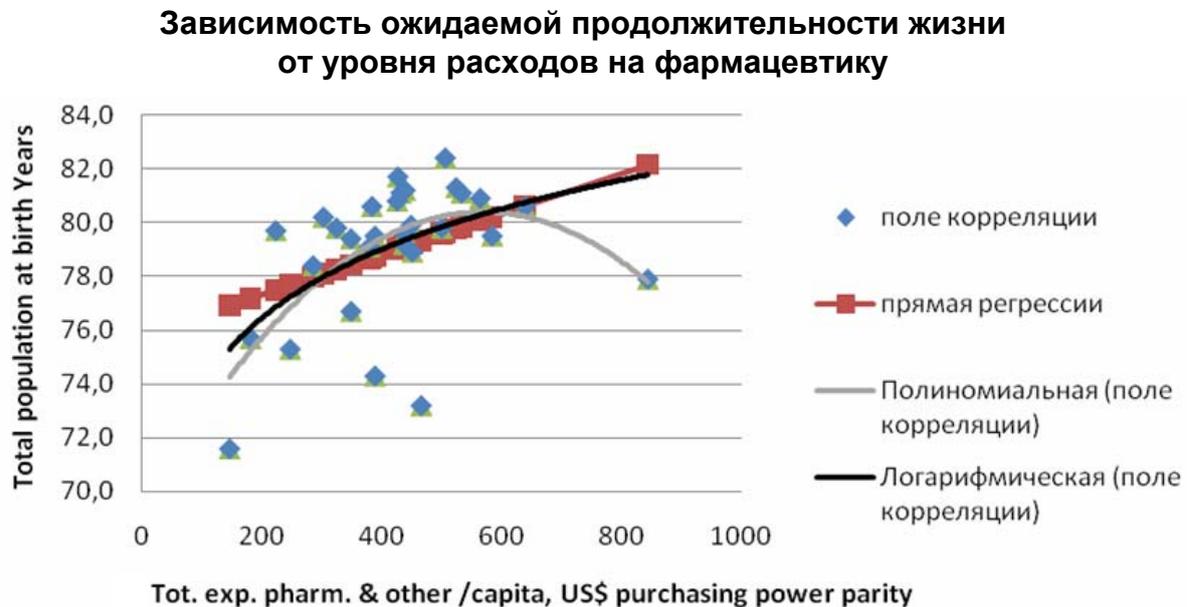


Рис. 5

Данная модель является значимой по критерию Фишера, так как $F_{набл} > F_{табл}$, где $F_{набл} = 5,54$, а $F_{табл} = 4,19$, следовательно основная гипотеза о том, что все коэффициенты уравнения равны нулю, отклоняется. Следовательно, данная модель является значимой.

Средняя ошибка аппроксимации составила 2,31 %, что свидетельствует о хорошем соответствии модели эмпирическим данным.

Также можно рассмотреть и нелинейные виды зависимости.

- Полиномиальная зависимость. Тогда:
 $\hat{Y} = 69,24 + 0,039X_5 - 3 \cdot 10^{-5} \cdot X_5^2$;
 $R^2 = 0,365$ – коэффициент детерминации;
 $A = 1,8\%$ – средняя ошибка аппроксимации.
- Линейно-логарифмическая зависимость. Тогда:
 $\hat{Y} = 56,77 + 3,71 \ln X_5$;
 $R^2 = 0,27$ – коэффициент детерминации;
 $A = 8,22\%$ – средняя ошибка аппроксимации.

Хотя квадратичная функция обладает самым большим коэффициентом детерминации, ее применение нелогично, так как от роста уровня расходов на фармацевтику не может быть снижения продолжительности жизни. В данном случае больше подходит линейно-логарифмическая за-

висимость, которая показывает замедление скорости роста продолжительности жизни от увеличения уровня расходов, что является логичным.

Таким образом, исследовав три модели, можно сделать вывод, что наиболее подходящей для исследования является линейно-логарифмическая зависимость. Включение данного фактора в модель множественной регрессии является целесообразным.

Для данной модели подходит следующая интерпретация: при увеличении уровня расходов на фармацевтику на 1 % ожидаемая продолжительность жизни вырастет на 0,0371 года.

3. Исследование влияния совокупности факторов

В качестве факторов, оказывающих влияние на продолжительность жизни, после предварительного исследования были отобраны следующие пять:

- X_1 – уровень затрат на здравоохранение (\$ ППС);
- X_2 – процент работоспособного населения;
- X_3 – количество смертельных случаев на 100 000 человек в результате инфаркта;
- X_4 – количество практикующих врачей на 1000 человек;
- X_5 – уровень расходов на фармацевтику (\$ ППС).

При оценке влияния этих 5 факторов на признак была получена следующая модель множественной линейной регрессии:

$$\hat{Y} = 74,1 + 0,0006X_1 + 0,089X_2 - 0,063X_3 + 0,632X_4 - 0,00097X_5.$$

Коэффициент детерминации $R^2 = 0,52$ говорит о том, что 52 % вариации продолжительности жизни объясняется вариацией совокупности признаков X_1 – X_5 . Остальные 48 % вариации объясняются неучтенными в данной модели факторами.

Данная модель является значимой по критерию Фишера, так как $F_{\text{набл}} > F_{\text{табл}}$, где $F_{\text{набл}} = 5,28$, а $F_{\text{табл}} = 2,62$.

Построим корреляционную матрицу, рассчитав парные коэффициенты корреляции (см. табл. 1).

После анализа матрицы становится очевидно, что наибольшее влияние на продолжительность жизни оказывает общая сумма затрат на здравоохранение, обладающая самым большим коэффициентом корреляции с признаком. При этом не существует коэффициента корреляции между факторами, превышающего 0,7, что позволяет удостовериться в отсутствии проблемы мультиколлинеарности.

Таблица 1

	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Y	1	0,568	0,517	-0,543	0,367	0,406
X ₁	0,568	1	0,549	-0,358	0,293	0,627
X ₂	0,517	0,549	1	-0,334	0,117	0,127
X ₃	-0,543	-0,358	-0,334	1	-0,211	-0,570
X ₄	0,367	0,293	0,117	-0,211	1	0,224
X ₅	0,406	0,627	0,127	-0,570	0,224	1

При рассмотрении значимости каждого из коэффициентов с помощью р-значения, полученного в пакете анализа, оказалось, что все факторы незначимы, так как превышают значение 0,05 (табл. 2).

Таблица 2

	<i>P-Значение</i>
Y	1,25276E-12
X ₁	0,229047011
X ₂	0,273401519
X ₃	0,075293682
X ₄	0,21093529
X ₅	0,826086559

Для того чтобы понять, почему факторы в совокупности являются значимыми, а по отдельности незначимыми, была выполнена оценка мультиколлинеарности факторов при помощи определителя матрицы парных коэффициентов корреляции между факторами. Определитель матрицы оказался равен 0,19, что подтверждает подозрения о ненадежности, так как чем ближе к нулю определитель межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. Проведем оценку значимости мультиколлинеарности факторов. Если фактическое значение χ^2 превосходит табличное (критическое) $\chi_{\text{факт.}}^2 > \chi_{\text{табл.}}^2$, то мультиколлинеарность считается доказанной. В нашем случае мультиколлинеарность оказалась доказанной, так как $\chi_{\text{факт.}}^2 = 45,77$, $\chi_{\text{табл.}}^2 = 18,31$, т. е. $\chi_{\text{факт.}}^2 > \chi_{\text{табл.}}^2$.

Далее из матрицы были поочередно исключены факторы X_1 и X_5 , так как они имели самые большие показатели корреляции между собой. Была получена следующая матрица (табл. 3):

Таблица 3

	X_2	X_3	X_4
X_2	1	-0,33478	0,1178537
X_3	-0,33478	1	-0,211808
X_4	0,1178537	-0,211808	1

Ее определитель оказался равен 0,845. Чем ближе к единице определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов. Проведем оценку значимости мультиколлинеарности факторов: $\chi_{\text{факт.}}^2 = 4,71$, $\chi_{\text{табл.}}^2 = 18,31$, т. е. $\chi_{\text{факт.}}^2 < \chi_{\text{табл.}}^2$, значит, мультиколлинеарность факторов является недоказанной.

Если построить регрессии факторов X_1 и X_5 на остальные, то можно увидеть, что полученные коэффициенты детерминации являются достаточно большими: 0,66 и 0,64 соответственно. Также данные регрессии являются значимыми по F-тесту. Следовательно, это является одним из подтверждений того, что X_1 и X_5 зависят от остальных факторов, и их исключение из модели множественной регрессии было оправданным шагом.

Таким образом, после более тщательного исследования в качестве факторов, оказывающих влияние на продолжительность жизни, были отобраны:

X_2 – процент работоспособного населения;

X_3 – количество смертельных случаев на 100 000 человек в результате инфаркта;

X_4 – количество практикующих врачей на 1000 человек.

Построим корреляционную матрицу, рассчитав парные коэффициенты корреляции (табл. 4).

Таблица 4

	Y	X_2	X_3	X_4
Y	1	0,517722	-0,54319	0,367654
X_2	0,517722	1	-0,33478	0,117854
X_3	-0,54319	-0,33478	1	-0,21181
X_4	0,367654	0,117854	-0,21181	1

Среди анализируемых факторов по показателю корреляции наибольшее влияние на продолжительность жизни оказывает первый фактор – уровень расходов на здравоохранение в стране на душу населения.

Построим модель множественной линейной регрессии:

$$\hat{Y}_x = 72,13 + 0,148X_2 - 0,066X_3 + 0,81X_4.$$

Коэффициент детерминации $R^2 = 0,48$ говорит о том, что 48 % вариации рождаемости объясняется вариацией этих трех факторов.

Скорректированный коэффициент детерминации $\bar{R}^2 = 0,42$, что не намного меньше коэффициента детерминации.

Данная модель является значимой по критерию Фишера, так как $F_{набл} > F_{табл}$, где $F_{набл} = 8,11$, а $F_{табл} = 2,975$.

Теперь произведем сравнение построенной модели множественной регрессии с моделью парной регрессии зависимости продолжительности жизни от уровня расходов на здравоохранение X_1 (табл. 5).

Таблица 5

Тип модели	R^2	\bar{R}^2	\bar{A}
Парная (X_1)	0,55	0,54	2,02%
Множественная	0,48	0,42	1,83%

Модель парной регрессии несколько лучше по каждому из рассмотренных показателей. Следовательно, для дальнейшего рассмотрения больше подходит модель парной регрессии зависимости ожидаемой продолжительности жизни от уровня расходов на здравоохранение.

ЗАКЛЮЧЕНИЕ

В ходе подробного изучения и анализа факторов было выявлено, что наибольшее влияние на продолжительность жизни оказывает уровень расходов на здравоохранение в стране. Следовательно, каждое государство, которое заботится о своих гражданах и задумывается над своим будущим, просто обязано увеличивать расходы на здравоохранение и следить за состоянием здоровья своего населения.

Кроме расходов на здравоохранение, достаточно сильно влияющими на продолжительность жизни, оказались факторы: уровень смертности населения от инфаркта и количество практикующих врачей на 1000 человек.

Совершенно не оказывающим никакого влияния на продолжительность жизни населения оказался такой фактор, как размер ВВП страны.

Подводя итог, надо отметить, что выборка в основном состояла из европейских стран. Опыт Европы в области здравоохранения может послужить хорошим примером для других стран мира. Несмотря на все сложности, продолжительность и качество жизни в Европе неуклонно растут. В России, по данным за 2008 год, ожидаемая продолжительность жизни составляет около 67 лет, что является достаточно печальным фактом. Если правительство нашей страны начнет уделять внимание этой проблеме, то есть твердая уверенность в том, что в будущем показатели продолжительности жизни в России достигнут европейского уровня.

ЛИТЕРАТУРА

1. Буре В.М., Евсеев Е.А. Эконометрика: Учеб. пособие. СПб., МБИ, 2007. – 168 с.
2. Тарашнина С.И., Панкратова Я.Б. Выполнение курсовой работы по эконометрике: Учеб.-метод. пособие. СПб., МБИ, 2007. – 97 с.
3. Эконометрика: Учебник под редакцией И.И. Елисеевой. М.: Финансы и статистика, 2003. – 344 с.
4. Официальный сайт Организации экономического сотрудничества и развития (ОЭСР) [Электронный ресурс]: <http://www.oecd.org>.

Немчинова Любовь Павловна

Международный банковский институт

Кузютин Д.В., к.ф.-м.н., доцент; **Панкратова Я.Б.** ст. преподаватель –
научные руководители

ФАКТОРЫ, ВЛИЯЮЩИЕ НА УРОВЕНЬ ПРЕСТУПНОСТИ В РОССИЙСКОЙ ФЕДЕРАЦИИ

Актуальность темы. В настоящее время чрезвычайно остро стоит проблема борьбы с преступностью, так как статистические данные за последние годы показывают резкий ее рост¹. Отсюда актуальным становится отбор факторов, влияющих на уровень преступности. Изучение зависимости преступности от различных факторов имеет большое значение как для населения Российской Федерации в целом, так и для правоохранительных органов и органов власти, поскольку они будут знать, на какие сферы жизни населения необходимо направить основные силы.

Цель исследования. Выявить факторы, которые могут влиять на уровень преступности в Российской Федерации, отобрать из них наиболее значимые, затем на основе выбранных факторов установить вид зависимости уровня преступности от этих факторов.

Проведенное исследование может носить рекомендательный характер правоохранительным органам и органам власти, так как оно может применяться в их работе в ближайшей перспективе, но будет необходимо проведение нового исследования в связи с экономической нестабильностью, а тем более в послекризисный период.

1. Сбор данных и отбор факторов

Для проведения исследования были отобраны следующие факторы, влияющие на уровень преступности в Российской Федерации:

- X_1 – численность граждан, имеющих статус безработных (тыс. человек);

¹ По данным Федеральной службы государственной статистики, число зарегистрированных преступлений, тысяч единиц

2002 г.	2003 г.	2004 г.	2005 г.	2006 г.	2007 г.
2526	2756	2894	3555	3855	3583

- X_2 – доля людей, имеющих высшее профессиональное образование в общей численности экономически активного населения (тыс. человек). Этот фактор можно расценивать как уровень образования или грамотности;
- X_3 – средние потребительские цены на алкогольную продукцию крепостью 40 % об. спирта и выше обыкновенного качества (рублей);
- X_4 – среднемесячная номинальная начисленная заработная плата (рублей);
- X_5 – величина прожиточного минимума для трудоспособного населения (руб.).

В качестве результирующего признака (количества преступлений) были использованы данные об общем количестве преступлений по регионам Российской Федерации без выделения отдельных видов преступлений (такие как причинение вреда здоровью, нарушение правил дорожного движения или преступления в экономической сфере). Для исследования были применены данные, отражающие вариацию факторов и результирующего признака в региональном разрезе, причем были задействованы данные не по всем регионам Российской Федерации. В связи с отсутствием данных из исследования были исключены следующие регионы:

- Коми-Пермяцкий автономный округ;
- Корякский автономный округ;
- Таймырский (Долгано-Ненецкий) автономный округ;
- Эвенкийский автономный округ;
- Чеченская республика.

Читинская область. В работе использовались данные Федеральной службы государственной статистики (Росстат) за 2007 год по регионам Российской Федерации, опубликованные в различных статистических сборниках.

В последующих главах будет проанализировано влияние каждого фактора на уровень преступности в отдельности и влияние всех этих факторов в совокупности. С помощью анализа влияния каждого фактора на результирующий признак в отдельности будет исследована возможность включения этого фактора в общую модель или исключение этого фактора.

2. Исследование влияния отдельных факторов

2.1. Исследование влияния уровня безработицы на уровень преступности

Представим исходные данные о количестве совершенных преступлений за 2007 г. и уровне безработицы по регионам в виде статистической таблицы, удобной для анализа (см. таблицу приложения).

Для изучения влияния фактора X_1 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 1). На основе данного поля корреляции трудно определить тип зависимости, но можно предположить наличие линейной и квадратичной зависимости. Была изучена возможность существования каждой из этих видов зависимостей и построены парные регрессии:

- 1) линейная $\hat{Y} = 18147,6 + 1570X_1$;
- 2) квадратичная $\hat{Y} = -9678 + 5200,2X_1 - 71,5X_1^2$;

Линейной модели можно дать следующую интерпретацию: при увеличении числа безработных на тысячу человек количество преступлений возрастет на 1570. Данная интерпретация вполне адекватно отражает реальность.

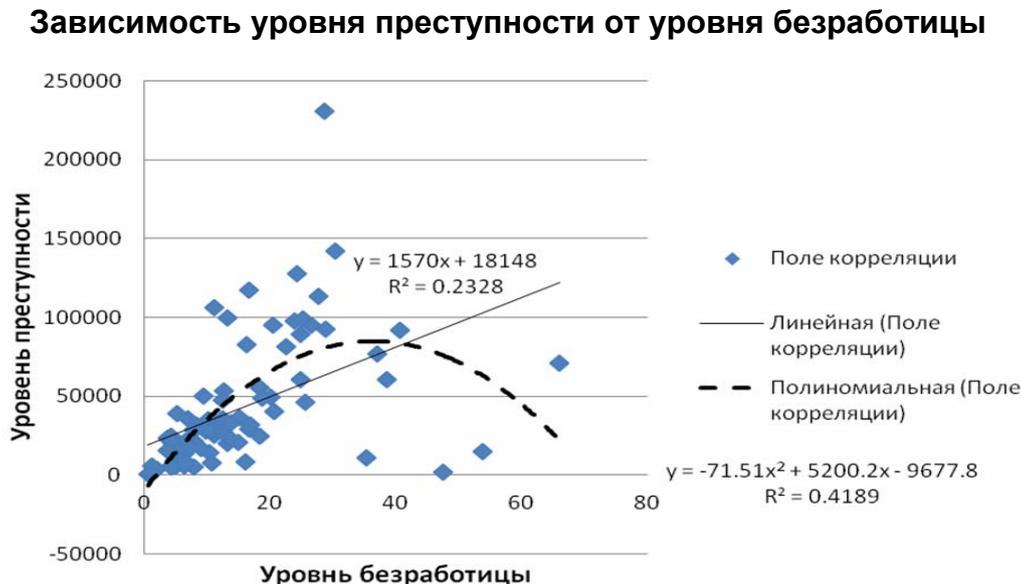


Рис. 1. Поле корреляции X_1 – Y

Для того чтобы определить целесообразность включения этого фактора в модель множественной регрессии, необходимо проанализировать следующие критерии:

- метод абсолютных отклонений (MAD); чем меньше этот показатель, тем лучше;

- средняя ошибка аппроксимации (\bar{A}); чем она меньше, тем лучше построенная модель аппроксимирует данные;
 - коэффициент детерминации.
- Результаты приведены в табл. 1.

Таблица 1

Значение критериев отбора модели

Тип модели	\bar{A}	MAD	R^2	$S_{ост}$
Линейная модель	2,07%	22532,13	0,23	35127,43
Квадратичная модель	1,38%	20269,60	0,42	30761,77
Гиперболическая модель	2,33%	27380,70	0,13	37455,71

Все модели являются значимыми по критерию Фишера, так как $F_{набл} > F_{табл}$, для линейной модели $F_{набл} = 24,887$, для квадратичной модели $F_{набл} = 29,189$ и для гиперболической $F_{набл} = 12011$, при этом $F_{табл} = 3,96$.

Далее оценим тесноту связи с помощью коэффициентов корреляции и детерминации. Коэффициент корреляции $r = 0,48$, что классифицируется как средняя связь между признаками.

Коэффициент детерминации у квадратичной модели самый высокий и равен 0,42, т. е. 42 % вариации уровня преступности объясняется вариацией уровня безработицы.

Отсюда можно сделать вывод, что данный фактор (X_1) подходит для описания зависимой переменной и целесообразно его включение в модель.

2.2. Исследование влияния уровня образования населения на уровень преступности

За уровень образования принимаем долю людей, имеющих высшее профессиональное образование в общей численности активного населения, для того чтобы оценить зависимость между уровнем образования и количеством совершаемых в стране преступлений рентабельности. Данные представлены в приложении.

Для изучения влияния фактора X_2 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 2).

При его рассмотрении можно заявить, что наиболее подходящими будут линейная и квадратичная модели, так как они смогут более точно отразить зависимость Y от фактора X_2 . Построим парные регрессии, для

того чтобы определить, включать данный фактор в модель множественной регрессии или нет:

- линейная $\hat{Y} = 23343,63 + 206,05X_2$;
- квадратичная $\hat{Y} = 6277,6 + 480,05X_2 - 0,24X_2^2$.

Зависимость уровня преступности от уровня образования населения

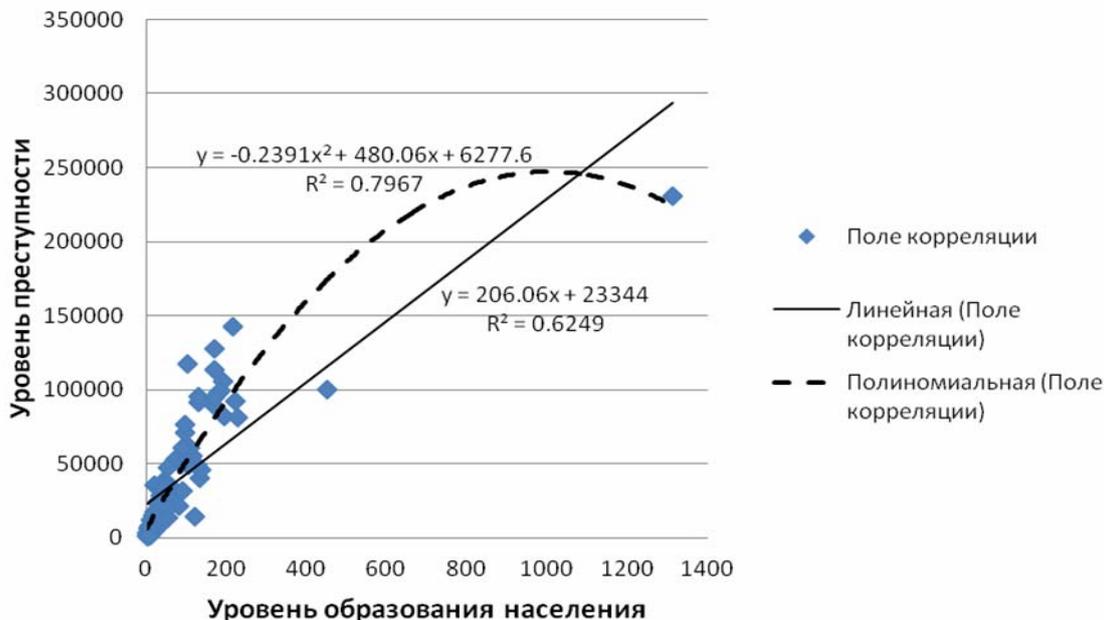


Рис. 2. Поле корреляции X_2 – Y

Линейной модели можно дать следующую интерпретацию: с увеличением студентов высших учебных заведений на тысячу человек количество зарегистрированных преступлений увеличится на 206 единиц. Такая зависимость может быть объяснена тем, что люди, имеющие высшее образование, имеют больше знаний, как, например, уйти от налогов или, совершив незаконные действия, остаться безнаказанным.

Для того чтобы определить целесообразность включения этого фактора в модель множественной регрессии, необходимо проанализировать ряд критериев: (результаты приведены в табл. 2).

Таблица 2

Значение критериев отбора модели

Тип модели	\bar{A}	MAD	R^2	$S_{\text{ост}}$
Линейная модель	2,48%	41440,86	0,62	24562,63
Квадратичная модель	0,66%	11854,23	0,8	18193,22

Видно, что по данным моделям средняя ошибка аппроксимации не превышает 3 %, следовательно приближение построенных моделей к наблюдаемым статистическим значениям считается очень хорошим.

Обе построенных модели являются значимыми по критерию Фишера, так как $F_{\text{набл}} > F_{\text{табл}}$, для линейной модели $F_{\text{набл}}=136,609$, $F_{\text{табл}}=3,96$, для квадратичной модели $F_{\text{набл}}=158,7$ и $F_{\text{табл}}=3,1$. Проверка коэффициентов моделей на значимость показала, что все коэффициенты значимы (для квадратичной модели $t_a=2,03$, $t_{b_1}=13,5$, $t_{b_2}=-8,27$ и $t_{\text{табл}}=1,99$, t_a , t_{b_1} , t_{b_2} по модулю больше $t_{\text{табл}}$.)

Далее оценим тесноту связи с помощью коэффициентов корреляции и детерминации. Коэффициент корреляции $r = 0,78$, что можно классифицировать как высокую связь между признаками.

Коэффициент детерминации у линейной модели равен 0,62, у квадратичной модели самый высокий и равен 0,8, т. е. 80 % вариации уровня преступности объясняется вариацией уровня образования.

Отсюда можно сделать вывод, что данный фактор (X_2) подходит для описания зависимой переменной и целесообразно его включение в модель.

2.3. Исследование влияния цен на алкогольную продукцию на уровень преступности

Интересно было проследить влияние цен на алкогольную продукцию и как этот фактор влияет на количество совершенных преступлений. Данные, взятые из исследований Федеральной службы государственной статистики, представлены в таблице приложения.

Для изучения влияния фактора X_3 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 3). По построенному полю корреляции можно предположить, что данный фактор не влияет на зависимую переменную. Попробуем построить парные регрессии, на основе чего решим, стоит ли включать данный фактор в модель множественной регрессии:

- линейная $\hat{Y} = 57731,14 - 90,73X_3$;
- квадратичная $\hat{Y} = -19206,7 + 614,4X_3 - 1,44X_3^2$.

Линейная зависимость может быть объяснена следующим образом: с ростом средней цены на алкогольную продукцию на рубль количество преступлений уменьшится на 90, т. е. если цена на алкогольную продукцию вырастет, то люди меньше будут употреблять алкоголь и, как следствие, будет меньше совершаться преступлений.

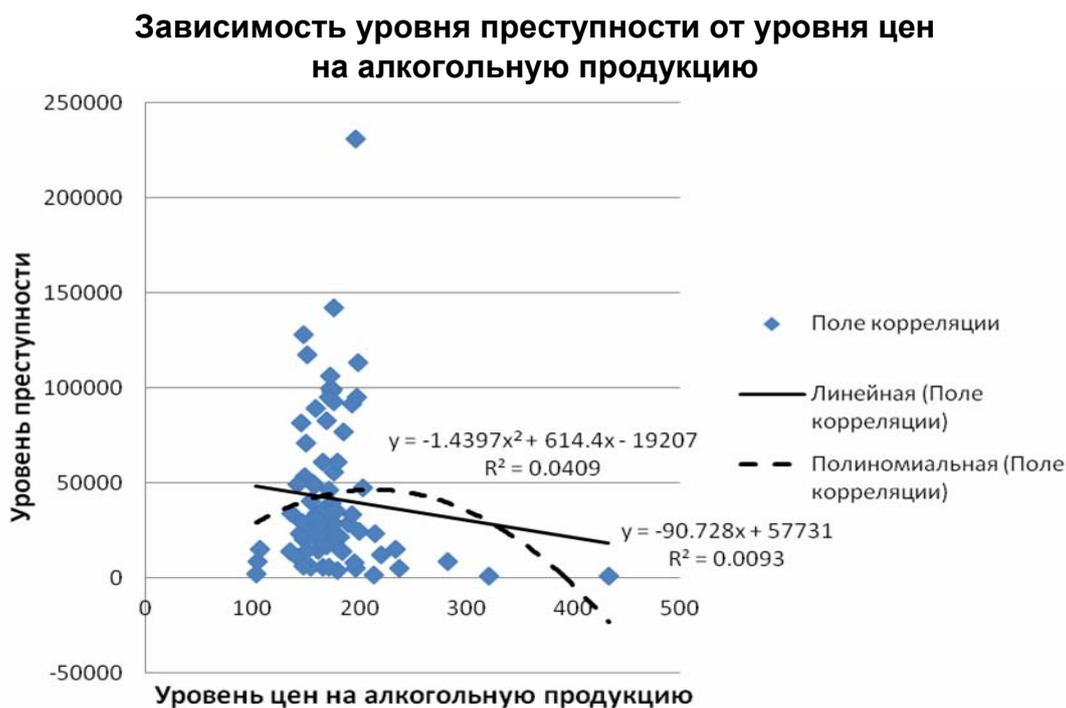


Рис. 3. Поле корреляции X_3 – Y

Все построенные модели являются незначимыми по критерию Фишера, так как $F_{\text{набл}} < F_{\text{табл}}$, для линейной модели $F_{\text{набл}}=0,77$, $F_{\text{табл}}=3,96$, для квадратичной модели $F_{\text{набл}}=1,72$ и $F_{\text{табл}}=3,1$.

Далее, оценим тесноту связи с помощью коэффициентов корреляции и детерминации. Коэффициент корреляции $r = -0,09$, что свидетельствует о полном отсутствии связи между признаками.

Коэффициент детерминации у линейной модели равен 0,009, у гиперболической модели – 0,0007, т. е. 0,9 % вариации уровня преступности объясняется вариацией средних цен на алкогольную продукцию, а 99,1 % объясняются неучтенными факторами.

Нет смысла проверять остальные показатели, так как очевидно, что данный фактор (X_3) не влияет на зависимую переменную и нецелесообразно включать его в модель множественной регрессии.

2.4. Исследование влияния уровня доходов (средней заработной платы) на уровень преступности

В качестве фактора (уровень доходов населения) были использованы данные по среднемесячной номинальной заработной плате населения по регионам. Совокупность данных сведена в таблицу приложения.

Для изучения влияния фактора X_4 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 4). На основе данного поля корреляции трудно определить тип зависимости, но можно предположить наличие линейной и квадратичной зависимости. Была изучена

возможность существования каждой из этих видов зависимостей и построены парные регрессии:

- линейная $\hat{Y} = 300350 + 0,9X_4$;
- квадратичная $\hat{Y} = -52692,8 + 11,9X_4 - 0,00028X_4^2$.

Интерпретация линейной модели: при увеличении средней начисленной заработной платы на 1 рубль количество совершенных преступлений увеличится на 1 зарегистрированную единицу. Данная интерпретация расходится с экономической логикой: ведь если население в стране в среднем начинает получать большую заработную плату, то и преступлений должно совершаться меньше.

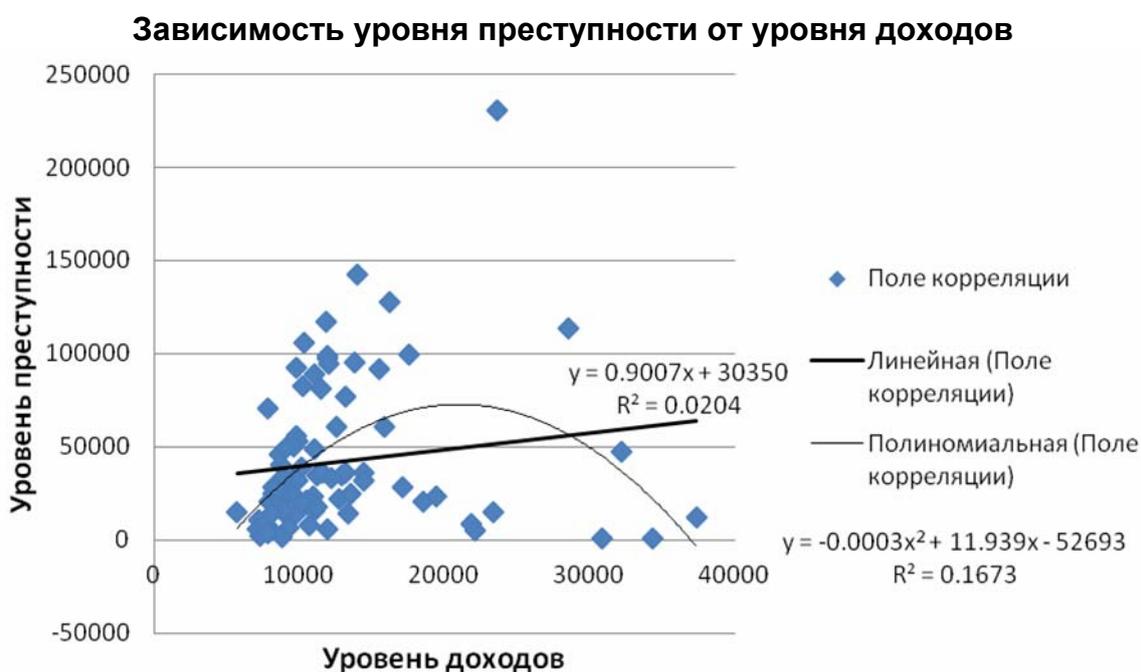


Рис. 4. Поле корреляции X_4 – Y

Для того чтобы определить целесообразность включения этого фактора в модель множественной регрессии, рассмотрим посчитанные критерии (результаты приведены в табл. 3).

Таблица 3

Значение критериев отбора модели

Тип модели	\bar{A}	MAD	R^2	$S_{\text{ост}}$
Линейная модель	3,17 %	30846,97	0,02	39694,45
Квадратичная модель	2,36 %	26798,31	0,17	36821,86

Средняя ошибка аппроксимации не превышает 8–10 %, значит, приближение считается хорошим.

Линейная модель является незначимой по критерию Фишера, так как $F_{\text{набл}} < F_{\text{табл}}$, $F_{\text{набл}}=1,7$, $F_{\text{табл}}=3,96$. Квадратичная модель является значимой по критерию Фишера, так как $F_{\text{набл}}=8,13 > F_{\text{табл}}=3,1$. Проверка коэффициентов моделей на значимость показала, что все коэффициенты значимы (для квадратичной модели $t_a=-2,21$, $t_{b1}=3,99$, $t_{b2}=-3,78$ и $t_{\text{табл}}=1,99$, t_a , t_{b1} , t_{b2} по модулю больше $t_{\text{табл}}$).

Далее оценим тесноту связи с помощью коэффициентов корреляции и детерминации. Коэффициент корреляции $r = 0,14$, что характеризует связь между признаками как очень слабую. Коэффициент детерминации у квадратичной модели самый высокий и равен $0,16$, у линейной модели еще ниже – $0,02$.

Исходя из проведенного анализа, сложно сказать, целесообразно ли включать фактор в модель множественной регрессии или нет, поскольку коэффициент детерминации очень низкий. Попробуем включить данный фактор X_4 в модель и посмотрим, будет ли этот фактор коррелировать с другими факторами в множественной модели.

2.5. Исследование влияния уровня доходов (среднего прожиточного минимума) на уровень преступности

В качестве фактора (уровень доходов населения) были использованы данные по средней величине прожиточного минимума по регионам. Совокупность данных сведена в таблицу приложения.

Для изучения влияния фактора X_5 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 5).

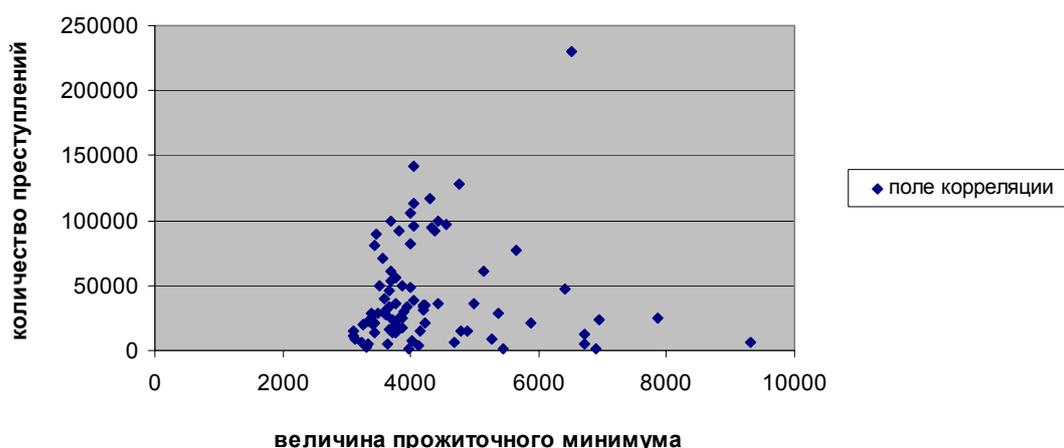


Рис. 5. Поле корреляции X_5 – Y

По построенному полю корреляции сложно определить какую-либо зависимость. Можно предположить, что данный фактор не влияет на зави-

симую переменную. Попробуем построить парные регрессии, на основе чего решим, стоит ли включать данный фактор в модель множественной регрессии.

Были построены три парные регрессии: линейная, квадратичная и гиперболическая, и ни одна модель не является значимой по критерию Фишера, так как $F_{набл} < F_{табл}$, для линейной модели $F_{набл}=0,09 < F_{табл}=3,96$, для квадратичной модели $F_{набл}=2,79 < F_{табл}=3,1$, для гиперболической модели $F_{набл}=1,1 < F_{табл}=3,96$. Коэффициент детерминации у линейной модели равен 0,001, у гиперболической модели – 0,01, у квадратичной – 0,06, что доказывает нецелесообразность включения этого фактора во множественную модель.

Проверка гипотезы гомоскедастичности с помощью метода Гольдфелда показала, что во всех парных моделях присутствует гетероскедастичность, так как $F > F_c$, то гипотеза о гетероскедастичности принимается. Значит, оценки метода наименьших квадратов могут не быть наименьшими (табл. 4).

Таблица 4

Тест Гольдфелда-Куандта

	Модель Y-X ₁	Модель Y-X ₂	Модель Y-X ₃	Модель Y-X ₄	Модель Y-X ₅
F	27,98	18,7	207640,2	11,37	5,99
F _c	1,93	1,93	1,93	1,93	1,93
Вывод	F>F _c				

Так как $F > F_c$, то гипотеза гетероскедастичности принимается.

Проверка автокорреляции моделей с помощью критерия Дарбина-Уотсона выявила, что во всех моделях отсутствует автокорреляция (табл. 5), что означает адекватность построенных парных регрессий истинной зависимости, следовательно построенные регрессии достаточны для прогнозирования. На их основе будем строить модель множественной регрессии из отобранных факторов.

Таблица 5

Тест Дарбина-Уотсона

	Модель Y-X ₁	Модель Y-X ₂	Модель Y-X ₃	Модель Y-X ₄	Модель Y-X ₅
d	1,7	1,7	1,95	1,89	1,91
dl	1,62	1,62	1,62	1,62	1,62
du	1,67	1,67	1,67	1,67	1,67
Вывод	d>du и 4-d>du				

Так выполняется тождество $d > du$ и $4 - d > du$, то гипотеза H_0 (отсутствие автокорреляции) принимается.

3. Исследование влияния совокупности факторов

В качестве факторов, оказывающих влияние на уровень преступности, после предварительного исследования были отобраны следующие:

- X_1 – уровень безработицы;
- X_2 – средняя месячная зарплата, попробуем включить этот фактор в множественную модель, несмотря на то, что по ряду критериев он имеет меньшее влияние на результирующий фактор;
- X_3 – доля людей, имеющих высшее профессиональное образование в общей численности экономически активного населения. Этот фактор можно расценивать как уровень образования или грамотности.

Рассчитанные парные коэффициенты корреляции представим в виде корреляционной матрицы (табл. 6).

Таблица 6

Расширенная матрица коэффициентов корреляции

Q	Y	X_1	X_2	X_3
Y	1	0,48	0,14	0,79
X_1	0,48	1	-0,12	0,31
X_2	0,14	-0,12	1	0,16
X_3	0,79	0,31	0,16	1

Из анализируемых факторов наибольшее влияние на количество совершенных преступлений оказывает фактор X_3 – количество лиц, обучающихся в вузах, так как значение линейного коэффициента корреляции выше 0,7. Парные коэффициенты корреляции говорят об отсутствии между ними связи. Определитель матрицы коэффициентов корреляции равен 0,85, следовательно – нет мультиколлинеарности.

Построим модель множественной регрессии:

$$\hat{Y} = 6778,7 + 893,9X_1 + 0,42X_2 + 181,06X_3.$$

Экономическая интерпретация коэффициентов множественной линейной регрессии:

- 1) при увеличении числа безработных на 1 тысячу человек количество совершенных преступлений увеличится на 893 при неизменности остальных факторов;
- 2) при увеличении средней номинальной начисленной заработной платы на 1 рубль количество совершенных преступлений увели-

чится на 0,42 при неизменности остальных факторов. Данная интерпретация не соответствует действительности;

- 3) при увеличении числа студентов на 1 тысячу человек количество преступлений увеличится на 181 единицу при неизменности остальных факторов. Эта зависимость может говорить о том, что чем больше людей имеют высшее образование (в частности экономическое), тем больше существует методов и способов уйти от налогов;
- 4) без учета влияния данных факторов количество совершенных преступлений будет равно 6779 единиц.

Проверим модель на значимость. Данная модель является значимой по критерию Фишера, так как $F_{набл} > F_{табл}$ ($F_{набл} = 29,84$, $F_{табл} = 2,72$). Проверим также коэффициенты множественной регрессии на значимость: $t_a = 1,015$, $t_{b1} = 4,13$, $t_{b2} = 1,04$, $t_{b3} = 10,4$ и $t_{табл} = 1,99$. Таким образом, коэффициенты α и β_2 незначимы, так как по модулю меньше $t_{табл}$.

Оценим тесноту связи с помощью коэффициентов детерминации. Коэффициент детерминации $R^2 = 0,69$, это говорит о том, что 69 % вариации уровня преступности объясняется вариацией трех рассмотренных факторов. Скорректированный коэффициент детерминации $R_{adj}^2 = 0,68$. Так как $R_{adj}^2 < R^2$, то качество регрессии считается хорошим.

Оценим точность прогноза, для этого рассчитаем среднюю ошибку аппроксимации $\bar{A} = 1,53$ %, значит, приближение построенной модели к наблюдаемым статистическим данным очень хорошее. Оценим точность прогноза с помощью критерия $MAD = 15153$.

По построенной модели можно сделать следующий вывод: модель является хорошей, но так как коэффициенты α и β_2 незначимы и фактору X_2 нельзя дать экономическую интерпретацию, придется исключить этот фактор из модели и построить новую модель множественной регрессии, оставив только два фактора X_1 и X_3 , после чего сравнить, какая модель лучше описывает влияние факторов на результирующий признак. Назовем модель, построенную выше, модель № 1, а новую модель с двумя факторами – модель № 2.

Построение модели № 2

Строим новую модель, оставляя в качестве факторов, оказывающих влияние на уровень преступности, только два фактора:

- X_1 – уровень безработицы;
- X_3 – доля людей, имеющих высшее профессиональное образование в общей численности экономически активного населения. Этот фактор можно расценивать как уровень образования или грамотности.

Рассчитанные парные коэффициенты корреляции представим в виде корреляционной матрицы (табл. 7).

Таблица 7

Расширенная матрица коэффициентов корреляции

	Y	X ₁	X ₃
Y	1	0,48	0,79
X ₁	0,48	1	0,31
X ₃	0,79	0,31	1

Из анализируемых факторов наибольшее влияние на количество совершенных преступлений, также как и модели № 1, оказывает фактор X₃ – количество лиц, обучающихся в вузах, так как значение линейного коэффициента корреляции выше 0,7. Парные коэффициенты корреляции говорят об отсутствии между ними связи. Определитель матрицы коэффициентов корреляции равен 1,107, следовательно – нет мультиколлинеарности.

Построим модель множественной регрессии:

$$\hat{Y} = 12408,58 + 852,9X_1 + 184,8X_3.$$

Экономическая интерпретация коэффициентов множественной линейной регрессии:

- 1) при увеличении числа безработных на 1 тысячу человек количество совершенных преступлений увеличится на 853 при неизменности остальных факторов;
- 2) при увеличении числа студентов на 1 тысячу человек количество преступлений увеличится на 185 единиц при неизменности остальных факторов;
- 3) без учета влияния данных факторов количество совершенных преступлений будет равно 12 408 единиц.

Данная экономическая интерпретация соответствует действительности.

Проверим модель на значимость. Данная модель является значимой по критерию Фишера, так как $F_{набл} > F_{табл}$ ($F_{набл} = 88,89$, $F_{табл} = 2,72$). Проверим также коэффициенты множественной регрессии на значимость: $t_a = 3,14$, $t_{b1} = 4$, $t_{b3} = 10,84$ и $t_{табл} = 1,99$. Таким образом, все коэффициенты значимы, так как по модулю больше $t_{табл}$.

Оценим тесноту связи с помощью коэффициентов детерминации. Коэффициент детерминации $R^2 = 0,69$, это говорит о том, что 69 % вариации уровня преступности объясняется вариацией трех рассмотренных

факторов. Скорректированный коэффициент детерминации $R_{adj}^2=0,68$. Так как $R_{adj}^2 < R^2$, то качество регрессии считается хорошим.

Посмотрим, какой из факторов оказывает большее влияние на зависимую переменную, для этого посчитаем эластичность. Фактор X_3 оказывает большее влияние на зависимую переменную, так как эластичность $X_1=0,3$, а эластичность $X_3=0,39$.

Оценим точность прогноза, для этого рассчитаем среднюю ошибку аппроксимации $\bar{A}=1,34$ %, значит, приближение построенной модели к наблюдаемым статистическим данным очень хорошее. Оценим точность прогноза с помощью критерия $MAD=15\ 202$.

Теперь произведем сравнение построенных моделей множественной регрессии. Это сравнение производится, чтобы осуществить выбор в пользу одной из моделей (табл. 8).

Таблица 8

Сравнение моделей

	R^2	R_{adj}^2	\bar{A}	MAD	$S_{ост}$
Модель № 1	0,69	0,68	1,53	15153	22562,7
Модель № 2	0,69	0,68	1,34	15202	22575,8

Согласно сравнительной таблице R^2 и R_{adj}^2 равны в обеих моделях, тогда смотрим на ошибку аппроксимации: в модели № 2 она несколько меньше, приближение считается отличным, так как \hat{Y} расходится с истинными значениями Y_i только на 1,34 %. В модели № 2 все коэффициенты значимы и имеют экономическую интерпретацию, поэтому, несмотря на то, что в модели № 1 MAD меньше, но незначительно, необходимо принять модель № 2 – модель зависимости уровня преступности от числа безработных и числа учащихся в вузах.

ЗАКЛЮЧЕНИЕ

В ходе исследования была изучена зависимость каждого из приведенных выше факторов на результирующий признак как в отдельности, так и в совокупности. Для этого были построены пять парных моделей и две модели множественной регрессии. Предположение о линейной зависимости признака от факторов было выдвинуто на основе построенных полей корреляции.

Приведенное исследование показало, что значимыми можно признать не все построенные модели. Модель, отражающая зависимость уровня преступности от цен на алкогольную продукцию, а также модель, отражающая зависимость уровня преступности от среднего прожиточного минимума, были признаны незначимыми. После предварительного анализа было решено включить в модель множественной регрессии три фактора, но фактор, отражающий зависимость уровня преступности от средней заработной платы, оказался незначим, и его экономическая интерпретация шла вразрез с экономической действительностью, поэтому этот фактор был исключен из модели множественной регрессии.

В ходе подсчетов и оценки модели с помощью различных критериев было установлено, что наиболее подходящей является модель, отражающая зависимость уровня преступности от числа лиц, имеющих статус безработных, и числа лиц, получающих высшее образование. Даная модель имеет следующий вид:

$$\hat{Y} = 12408,58 + 852,9X_1 + 184,8X_3 .$$

Полученной модели можно дать экономическую интерпретацию: при увеличении числа лиц, имеющих статус безработных, на 1 тыс. человек количество совершенных преступлений увеличится на 853. Эта интерпретация соответствует действительности, ведь когда люди остаются без работы, они теряют средства к существованию и могут пойти на преступление. Второй коэффициент показывает, что при увеличении числа лиц, получающих высшее образование, на 1 тыс. человек количество совершенных преступлений увеличится на 185. Эта зависимость может быть объяснена тем, что чем больше людей имеют высшее образование, в частности экономическое, тем больше они имеют знаний и методов, как, например, уйти от налогов или совершить незаконные спекулятивные махинации.

$F_{\text{набл.}}$	$F_{\text{табл.}}$	R^2	R^2_{adj}	S_e	MAD	A
88,9	2,72	0,69	0,68	22575,8	15202,8	1,34

Коэффициент детерминации R^2 равен 0,69, т. е. 69 % вариации количества преступлений объясняется вариацией двух рассмотренных факторов. Скорректированный коэффициент детерминации R^2_{adj} равен 0,68, т. е. меньше чем R^2 , что говорит о хорошем качестве модели. Ошибка аппроксимации показывает, что \hat{Y} расходится с истинными значениями Y_i только

на 1,34 %. Также модель является значимой по тесту Фишера и все коэффициенты значимы.

Подводя итог, можно утверждать, что количество преступлений по регионам РФ в большей степени зависит от уровня безработицы, что особенно актуально в период кризиса. Поэтому данный анализ может иметь рекомендательный характер Правительству РФ с целью концентрирования большего внимания на проблеме безработицы и недопущению резкого роста преступности, возможно, путем создания дополнительных рабочих мест.

ЛИТЕРАТУРА

1. *Евсеев Е. А., Буре В. М.* Эконометрика. СПб.: Изд-во МБИ, 2007. – 139 с.
2. *Тарашина С. И., Панкратова Я. Б.* Выполнение курсовой работы по эконометрике: Учеб.-метод. пособие. СПб.: Изд-во МБИ, 2007. – 97 с.
3. Практикум по эконометрике: Учеб. пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др. Под ред. И.И. Елисеевой. М.: Финансы и статистика, 2003. – 192 с.: ил.
4. Сборники государственной статистики Росстат [Электронный ресурс]: <http://www.gks.ru>

ПРИЛОЖЕНИЕ

Таблица

Регион	Число преступлений, единиц, X_1	Численность безработных, тысяча человек, X_2	Число лиц, получающих ВО, тысяча человек, X_3	Средняя цена водки обыкновенного качества, X_4	Среднемесячная заработная плата, рубль, X_5	Величина прожиточного минимума, рубль, X_6
1	2	3	4	5	6	7
Российская Федерация	3582541	15625,4	7461.3	170,712	13593.4	4159
Центральный федеральный округ	764203	1866	2369.5	172,934	15876.7	3444,75
Белгородская обл.	21184	7,900	80,1	181,955	10479,5	3444,75
Брянская обл.	28097	9,867	52,2	163,24	8189,6	3489,5
Владимирская обл.	33779	12,950	57,3	160,8	9688,1	3949,5
Воронежская обл.	40322	20,692	134,1	155,09	8730,9	3593
Ивановская обл.	24990	11,092	54,5	168,54	8171,6	3826,75
Калужская обл.	23058	3,708	37,2	168,68	10926,8	3712
Костромская обл.	13779	4,575	21,8	135,28	9058,1	3717,75
Курская обл.	28612	7,850	65,4	176,8	8856,8	3393,5
Липецкая обл.	20020	4,450	36,8	148	10907,1	3398,5
Московская обл.	127714	24,267	170,5	147,875	16234,5	4768,75
Орловская обл.	20032	5,483	46,4	160,76	8610,7	3247,5
Рязанская обл.	13839	6,458	54,3	184,405	9796,6	3778,25
Смоленская обл.	24555	4,208	44,2	158,13	9552,1	3866,75
Тамбовская обл.	20316	13,308	42,8	165,645	7903	3244,5
Тверская обл.	38927	5,308	44,7	174,195	10177	4038,75
Тульская обл.	19592	6,833	55	175,025	10137,1	3765,75
Ярославская обл.	34756	10,158	59,3	178,03	11214,8	4186,5
Москва	230631	28,742	1312,6	196,225	23623,3	6520
Республика Карелия	14505	10,317	24,5	161,33	13342,1	4779,75
Республика Коми	28358	11,708	35,9	189,6	17077,3	5360,75
Архангельская обл.	36087	15,125	47,3	167,73	14400,3	4997,25
Ненецкий авт. округ	804	0,583	0,4	321,45	34380,7	6895,5
Вологодская обл.	34299	10,425	52,9	171,33	12913,9	4225,75
Калининградская обл.	21682	6,567	45,3	164,785	12750,3	4229,5

Таблица (продолжение)

1	2	3	4	5	6	7
Ленинградская обл.	35816	6,992	20,1	175,945	13154,8	3772,25
Мурманская обл.	20813	14,875	37,8	176,865	18581	5871,75
Новгородская обл.	15325	4,442	22,7	180,57	11004,9	4146,25
Псковская обл.	16597	9,100	24,6	169,415	8950,5	3659,5
Санкт-Петербург	99776	13,158	450,1	173,445	17552	4426,75
Республика Адыгея	5165	7,858	20,4	196,68	8056,4	3647,5
Республика Дагестан	14730	53,858	119,5	106,655	5696,4	3092,25
Республика Ингушетия	2104	47,625	10,1	103,37	7285,4	3315,5
Кабардино-Балкарская Республика	10790	35,367	29,7	142,7	7213,9	3104
Республика Калмыкия	5435	4,967	11,5	166,12	7101,2	3346
Карачаево-Черкесская Республика	5602	6,350	17,2	154,515	7558,4	3225,5
Республика Северная Осетия	8468	16,067	35,5	105,095	7625,8	3126,25
Краснодарский край	82360	16,308	194,9	169,665	10260	4007,5
Ставропольский край	45968	25,575	137,3	171,985	8647,7	3665
Астраханская обл.	31707	8,242	43,9	156,73	9866,9	3608,75
Волгоградская обл.	55453	18,325	115,8	175,765	9770,2	3766,25
Ростовская обл.	92516	28,883	220,7	176,19	9779,6	3813,5
Республика Башкортостан	89109	24,867	169,5	159,475	11027,1	3448
Республика Марий Эл	22958	4,242	30,5	144,14	8404,3	3332,75
Республика Мордовия	14201	5,525	41,9	151,265	8103	3430,75
Республика Татарстан	81251	22,558	227,8	145,25	11468,6	3431,5
Удмуртская Республика	53035	12,600	72,3	148,53	9838,6	3692
Чувашская Республика	27690	10,517	72,4	157,185	8703,2	3421,5
Пермский край	117198	16,758	101,1	151,59	11856	4308,75
Кировская обл.	29257	16,633	59,6	146,555	8861,5	3884,75
Нижегородская обл.	106007	11,192	191,4	173,23	10302	3994
Оренбургская обл.	49824	9,475	86,3	156,74	9619,6	3501,25
Пензенская обл.	28367	7,958	59,6	165,725	8566,4	3596,25
Самарская обл.	97633	23,867	180,4	173,835	11920,7	4550
Саратовская обл.	49132	20,158	126,6	142,47	9108,3	3859,75

Таблица (окончание)

1	2	3	4	5	6	7
Ульяновская обл.	27726	12,417	55,3	162,35	8412,7	3612
Курганская обл.	33820	12,542	37,5	136,955	8883,1	3672,75
Свердловская обл.	142269	30,325	214,4	176,625	13986,9	4051
Тюменская обл.	113456	27,742	169,4	199,435	28565	4053
Ханты-Мансийский авт. округ	47393	12,442	53,8	203,53	32227,9	6419,5
Ямало-Ненецкий авт. округ	11823	5,175	12,1	220,16	37363,5	6719,5
Челябинская обл.	99097	25,283	185,1	175,175	11897,5	3697
Республика Алтай	6326	4,475	5,3	147,77	9228,2	4687
Республика Бурятия	35828	12,475	47,5	177,215	11528,6	4428,25
Республика Тыва	7781	10,783	6,5	195,765	10701,6	4020,5
Республика Хакасия	17540	7,400	21,3	167,49	11251,2	3874,25
Алтайский край	70860	66,108	95,7	150,62	7804,7	3561,5
Забайкальский край	32988	13,758	45,6	193,84	12161,5	4210
Агинский Бурятский авт. округ	1392	1,158	1,3	213,28	8763,3	3980
Красноярский край	91518	40,608	129,5	192,95	15510	4375,25
Иркутская обл.	95221	26,650	131,4	197,32	13770	4042,5
Усть-Орд Бурятский авт. округ	3655	1,950	0,2	180,08	7816,9	4133
Кемеровская обл.	60665	38,633	108,8	179,43	12554,9	3691,75
Новосибирская обл.	94858	20,525	170,9	171,195	12017	4336
Омская обл.	48637	18,850	106,8	157,94	11003,6	4005,25
Томская обл.	31655	16,908	89,7	159,395	14429	4195,75
Республика Саха (Якутия)	23075	13,675	43,6	214,475	19409,2	6939,75
Камчатский край	8305	6,642	16,3	282,925	21814,7	5276,75
Приморский край	76636	36,992	97,4	185,915	13174,1	5650
Хабаровский край	60593	24,783	89,5	166,28	15883,5	5145
Амурская обл.	24470	18,308	29,8	199,715	13534,4	7864,333
Магаданская обл.	4835	4,325	10,4	237,56	22101,6	6709,25
Сахалинская обл.	15214	3,708	16,4	234,31	23346,3	4879,5
Еврейская автоном- ная обл.	5777	1,192	7,5	171,875	11968,9	9311,25
Чукотский авт. округ	1033	1,127	5,8	433,53	30859,1	5457,5

Сизова Мария Александровна

Международный банковский институт

Кузютин Д.В., к.ф.-м.н., доцент; **Евсеев Е.А.**, к.ф.-м.н., доцент –
научные руководители

ВЛИЯНИЕ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ НА КОЛИЧЕСТВО БРАКОВ В РЕГИОНАХ РОССИИ

Сегодня политика государства нацелена на укрепление статуса семьи, повышение престижа материнства и отцовства. Брак – в семейном праве – добровольный, равноправный союз женщины и мужчины, заключаемый для создания семьи и порождающий взаимные права и обязанности супругов.

Но заключение брака зависит не только от желания двух любящих людей, а также от многих социально-экономических факторов.

Была предпринята попытка выявить те факторы, которые оказывают наибольшее влияние на количество заключаемых браков в России.

1. Сбор данных и отбор факторов

Для проведения исследования были отобраны следующие факторы, влияющие на количество браков по регионам России:

X_1 – общая площадь жилых помещений, приходящаяся в среднем на одного жителя (m^2);

X_2 – численность постоянного населения;

X_3 – среднемесячная номинальная начисленная заработная плата;

X_4 – объем платных услуг населению (млн руб.);

X_5 – число высших учебных учреждений по регионам РФ.

В качестве результирующего признака были использованы данные по количеству браков по регионам России. В исследовании использовались данные Росстата за 2007 год.

Далее будет изучено влияние каждого из приведенных выше факторов на количество браков по регионам по отдельности и влияние этих факторов в совокупности.

2. Исследование влияния отдельных факторов

2.1. Исследование влияния общей площади жилых помещений, приходящейся в среднем на одного жителя, на количество браков

В статистической таблице приложения представлены исходные данные о количестве браков по регионам России.

Для изучения влияния фактора X_1 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 1).



По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную:

$$\hat{Y} = 18341,66 + 199,31 \cdot X_1.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{yx_1} = -0,0662$. Согласно шкале Чеддока, такая связь изучаемых признаков является незаметной.

Коэффициент детерминации 0,0044 говорит о том, что только около 0,44 % вариации количества браков объясняется вариацией общей площади жилых помещений. Остальные 99,56 % вариации объясняются неучтенными в данной модели факторами. Средняя ошибка аппроксимации составила 106,2 %.

Данная модель является незначимой по критерию Фишера, так как $F_{\text{набл}} < F_{\text{табл}}$, где $F_{\text{набл}} = 0,3$, а $F_{\text{табл}} = 3,97$.

Таким образом, данная математическая модель не подходит для описания зависимой переменной. Дальнейшее включение фактора X_1 в модель множественной регрессии является нецелесообразным.

2.2. Исследование влияния численности постоянного населения на количество браков

Для изучения влияния фактора X_2 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 2).

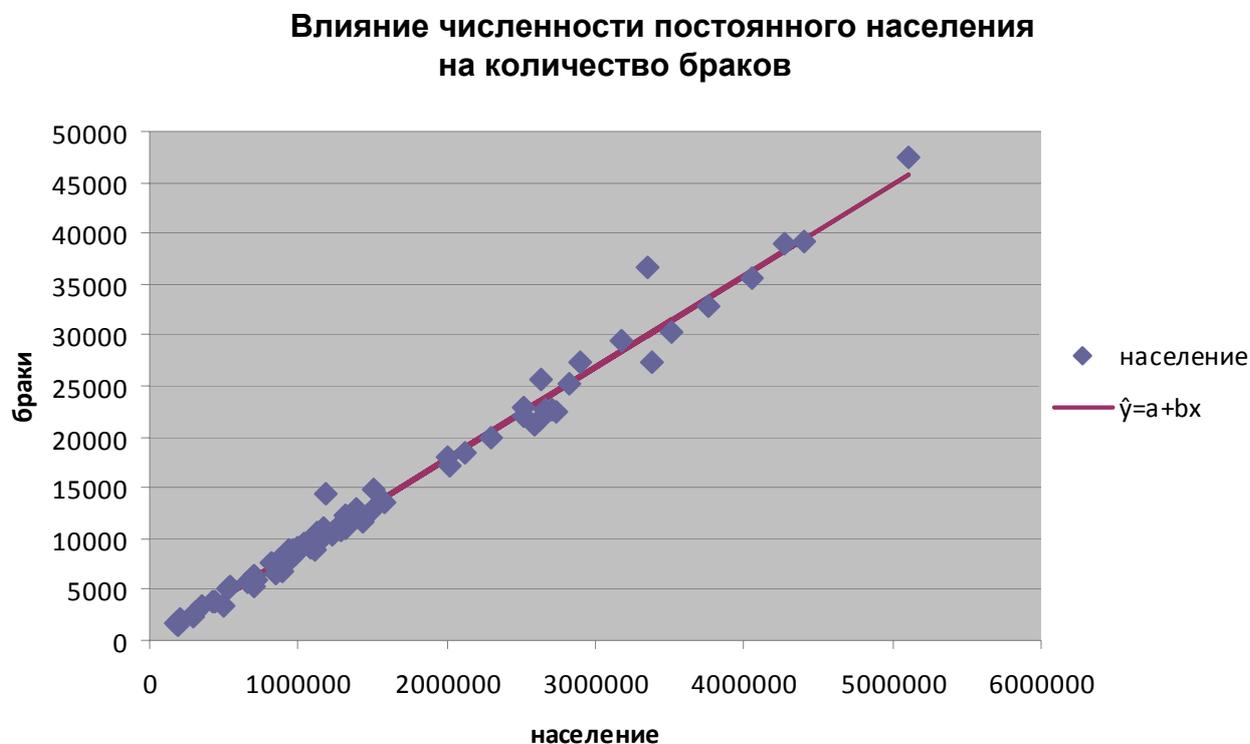


Рис. 2

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную:

$$Y = -209,71 + 0,01 \cdot X_2 .$$

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{yx_2} = 0,993$. Согласно шкале Чеддока такая связь изучаемых признаков является очень заметной.

Коэффициент детерминации 0,98 говорит о том, что около 98 % вариации количества браков объясняется вариацией количества населения. Остальные 2 % вариации объясняются неучтенными в данной модели факторами. Средняя ошибка аппроксимации составила 6,15 %, что свидетельствует о хорошем соответствии модели эмпирическим данным.

Данная модель является значимой по критерию Фишера, так как $F_{\text{набл}} > F_{\text{табл}}$, где $F_{\text{набл}} = 4851,390647$, а $F_{\text{табл}} = 3,97$.

Таким образом, данная математическая модель подходит для описания зависимой переменной. Включение данного фактора в модель множественной регрессии является целесообразным.

Для данной модели подходит следующая интерпретация: при увеличении количества населения на 100 человек количество браков увеличивается на 1.

2.3. Исследование влияния среднемесячной номинальной начисленной заработной платы на количество браков

Для изучения влияния фактора X_3 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 3).

Влияние среднемесячной номинальной начисленной заработной платы на количество браков

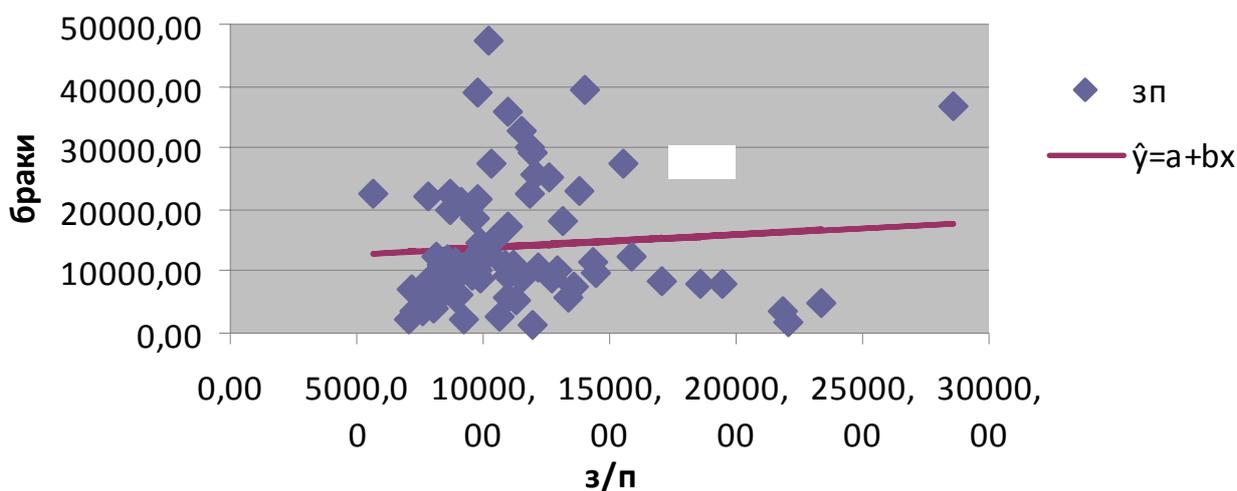


Рис. 3

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную:

$$\hat{Y} = 11638,12 + 0,22 \cdot X_3.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{yx_3} = 0,085$. Согласно шкале Чеддока такая связь изучаемых признаков является незаметной.

Коэффициент детерминации 0,0072 говорит о том, что только около 0,72 % вариации количества браков объясняется вариацией заработной

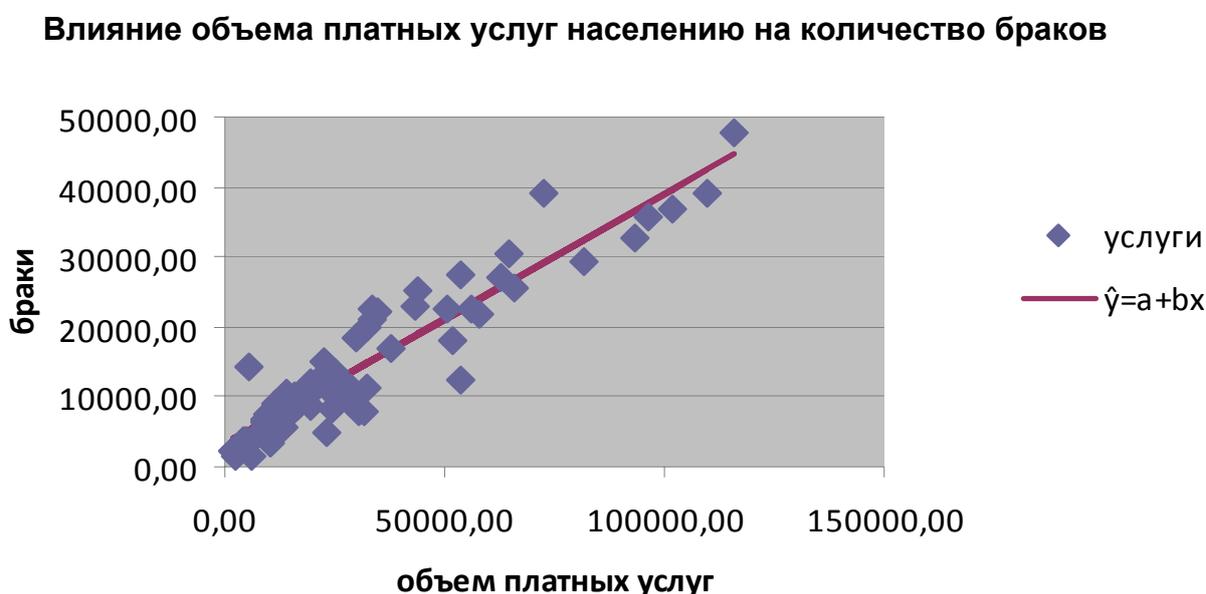
платы. Остальные 99,28 % вариации объясняются неучтенными в данной модели факторами. Средняя ошибка аппроксимации составила 107,03.

Данная модель является незначимой по критерию Фишера, так как $F_{\text{набл}} < F_{\text{табл}}$, где $F_{\text{набл}} = 0,53$, а $F_{\text{табл}} = 3,97$.

Таким образом, данная математическая модель не подходит для описания зависимой переменной. Ее дальнейшее включение в модель множественной регрессии является нецелесообразным.

2.4. Исследование влияния объема платных услуг населению на количество браков

Для изучения влияния фактора X_4 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 4).



По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную:

$$\hat{Y} = 3280,12 + 0,36 \cdot X_4.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{yx_4} = 0,94$. Согласно шкале Чеддока такая связь изучаемых признаков является очень заметной.

Коэффициент детерминации 0,882 говорит о том, что около 88,2 % вариации количества браков объясняется вариацией объема платных услуг. Остальные 11,8 % вариации объясняются неучтенными в данной модели факторами. Средняя ошибка аппроксимации составила 28,79 %.

Данная модель является значимой по критерию Фишера, так как $F_{\text{набл}} > F_{\text{табл}}$, где $F_{\text{набл}} = 546,5091367$, а $F_{\text{табл}} = 3,97$.

Таким образом, данная математическая модель подходит для описания зависимой переменной. Включение данного фактора в модель множественной регрессии является целесообразным.

Данная модель имеет следующую интерпретацию: при увеличении объема платных услуг на 1 млн происходит увеличение количества браков на 0,36.

2.5. Исследование влияния числа высших учебных учреждений на количество браков

Для изучения влияния фактора X_5 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 5).

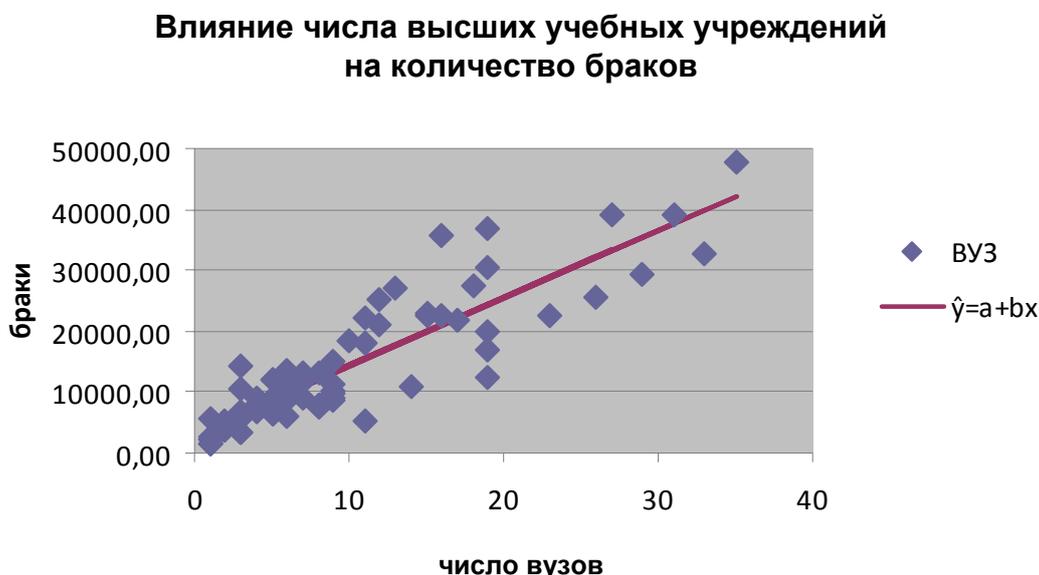


Рис. 5

По данному полю корреляции трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную:

$$\hat{Y} = 3257,54 + 1112,67 \cdot X_5.$$

Оценим тесноту связи с помощью показателей корреляции и детерминации. Коэффициент корреляции $r_{yx_5} = 0,89$. Согласно шкале Чеддока такая связь изучаемых признаков является очень заметной.

Коэффициент детерминации 0,789122827 говорит о том, что около 78,9 % вариации количества браков объясняется вариацией числа высших учебных учреждений. Остальные 21,1 % вариации объясняются неучтен-

ными в данной модели факторами. Средняя ошибка аппроксимации составила 34,18 %.

Данная модель является значимой по критерию Фишера, так как $F_{\text{набл}} > F_{\text{табл}}$, где $F_{\text{набл}} = 273,17$, а $F_{\text{табл}} = 3,97$.

Таким образом, данная математическая модель подходит для описания зависимой переменной. Включение данного фактора в модель множественной регрессии является целесообразным.

Для данной модели подходит следующая интерпретация: при увеличении числа высших учебных учреждений на 1 ед. происходит увеличение количества браков на 1112.

Включение фиктивной переменной

В работе была предпринята попытка исследовать влияние включения фиктивной переменной. Все регионы были разделены на 4 группы: обработка, добыча, торговля и сельское хозяйство.

Исследование проводилось на основе трех значимых факторов: X_2 , X_4 , X_5 .

Коэффициенты детерминации при множественной регрессии: добыча – 0,19; обработка – 0,23; сельское хозяйство – 0,35; торговля – 0,12. Можно увидеть, что коэффициенты достаточно небольшие.

Модель влияния выбранных факторов на результирующий признак следующая:

$$Y = -0,00043 \cdot X_1 + 0,03 \cdot X_2 - 7,1 \cdot X_3 + 5081,9 \cdot d_1 - 76,9 \cdot d_2 - 1521,9 \cdot d_3.$$

Кроме того, построим множественную регрессию для всей совокупности в целом. Коэффициент детерминации $R^2 = 0,37$ говорит о том, что только около 37 % вариации количества браков объясняется данной совокупностью факторов. Остальные 63 % вариации объясняются неучтенными в данной модели факторами. Таким образом, включение фиктивной переменной считается нецелесообразным.

Исследование влияния совокупности факторов

В качестве факторов, оказывающих влияние на количество, после предварительного исследования были отобраны следующие три:

X_1 – общая площадь жилых помещений;

X_2 – численность постоянного населения;

X_3 – среднемесячная номинальная начисленная заработная плата.

Построим модель множественной линейной регрессии:

$$\hat{Y} = 274,2 + 100,3 \cdot X_1 + 0,009 \cdot X_2 - 0,14 \cdot X_3 .$$

Коэффициент детерминации $R^2 = 0,998$ говорит о том, что 99,8 % вариации количества браков объясняется вариацией совокупности признаков X_1, X_2, X_3 . Остальные 0,2 % вариации объясняются неучтенными в данной модели факторами.

Данная модель является значимой по критерию Фишера, т. к. $F_{набл} > F_{табл}$, где $F_{набл} = 22257,17$, а $F_{табл} = 3,97$.

Проверим коэффициенты уравнения линейной регрессии на значимость по Стьюденту: $t_{табл} = 1,99$, а $t_{\beta_1} = -2,72$, $t_{\beta_2} = -81,1$, $t_{\beta_3} = 4,8$, следовательно все коэффициенты значимы.

Построим корреляционную матрицу, рассчитав парные коэффициенты корреляции (таблица).

Таблица

Матрица Q'	Y	X ₁	X ₂	X ₃
Y	1,000	-0,066	0,993	0,085
X ₁	-0,066	1,000	-0,038	0,075
X ₂	0,993	-0,038	1,000	0,029
X ₃	0,085	0,075	0,029	1,000

После анализа матрицы становится очевидно, что наибольшее влияние на количество браков оказывает численность населения, обладающая самым большим коэффициентом корреляции с признаком. При этом не существуют коэффициентов корреляции между факторами, превышающего 0,7, что говорит об отсутствии проблемы мультиколлинеарности.

Однако очевидно, что при рассмотрении фактора 2 отдельно по определенным показателям видно, что он оказывает наибольшее воздействие на количество браков, нежели совокупность факторов. Таким образом, дальнейшее исследование будет происходить на основе данного фактора.

ЗАКЛЮЧЕНИЕ

В ходе подробного изучения и анализа факторов было выявлено, что наибольшее влияние на количество браков оказывает численность населе-

ния в данном регионе, что вполне естественно. Отношения между людьми в браке являются личными и сугубо индивидуальными. Вместе с тем они имеют и общественное значение. Так как браки являются показателем социальной устойчивости в стране, государству необходимо уделять повышенное внимание демографической политике, которая в данный момент имеет отрицательный характер.

Также сильно влияющими на количество браков оказались объем платных услуг населению и число высших учебных учреждений по регионам РФ, что в период кризиса оказывает особое влияние на данную тенденцию.

Совсем небольшое влияние имеют такие факторы, как общая площадь жилых помещений, приходящаяся в среднем на одного жителя, и среднемесячная номинальная начисленная заработная плата, что может отражать особенности российского общества.

Подводя итог, надо отметить, что в прошлом году в России стали регистрировать меньше браков, зато больше – разводов. Эксперты считают это следствием финансового кризиса. Огромную роль играет желание бывших и будущих супругов с наименьшими потерями пережить эпоху финансовых потрясений. Что будет в России с браками в дальнейшем? Есть пессимистическая и оптимистическая точки зрения. Если экономический кризис окажется действительно глубоким, то браков станет меньше, ведь они сопряжены с расходами, а денег во время кризиса у людей становится тоже меньше.

Однако брак, как отмечалось, является важной составляющей развития страны в целом. Таким образом, государству следует уделять этому фактору большое внимание и делать все возможное, чтобы люди могли вступать в узы брака спокойно и максимально независимо от социально-экономических показателей.

ЛИТЕРАТУРА

1. *Евсеев Е.А., Буре В.М.* Эконометрика. СПб.: Изд-во МБИ, 2007. –139 с.
2. *Тарашина С.И., Панкратова Я.Б.* Выполнение курсовой работы по эконометрике: Учеб.-метод. пособие. СПб.: Изд-во МБИ, 2007. – 97 с.
3. *Кремер Н.Ш., Путко Б.А.* Эконометрика: Учебник для вузов / Под ред. проф. Н.Ш. Кремера. М.: ЮНИТИ-ДАНА, 2003. – 311 с.
4. Практикум по эконометрике: Учеб. пособие / И.И. Елисеева, С.В. Курьшева, Н.М. Гордеенко и др. Под ред. И.И. Елисеевой. М.: Финансы и статистика, 2003. – 192 с.: ил.
5. Сборники государственной статистики Росстат [Электронный ресурс]: <http://www.gks.ru>

ПРИЛОЖЕНИЕ

Таблица

	Число браков, У	X ₁	X ₂	X ₃	X ₄	X ₅
1	2	3	4	5	6	7
Белгородская область	14899	23,9	1513620	10479,5	22574,6	9
Брянская область	12215	23,8	1317589	8189,6	20629,7	5
Владимирская область	12330	24	1459574	9688,1	22616,9	6
Воронежская область	19826	25	2294616	8730,9	32512	19
Ивановская область	9148	23	1087886	8171,6	14349,8	9
Калужская область	9105	23,7	1008968	10926,8	17149,3	5
Костромская область	6291	21,5	702209	9058,1	8930,2	3
Курская область	11017	24,3	1170730	8856,8	18643,9	14
Липецкая область	10918	24,5	1173870	10907,1	18142,8	6
Орловская область	7574	22,6	826588	8610,7	10023,7	8
Рязанская область	10164	24,7	1172325	9796,6	16136,3	9
Смоленская область	9087	25,2	993514	9552,1	13936,7	7
Тамбовская область	8803	23,2	1117119	7903	16256,9	6
Тверская область	13020	26,5	1390444	10177	24958,1	8
Тульская область	13600	24,2	1580531	10137,1	25272,3	6
Ярославская область	11172	22,9	1320140	11214,8	28489,5	9
Республика Карелия	5823	23,3	693150	13342,1	13232,7	3
Республика Коми	8427	23,2	974617	17077,3	24379,4	5
Архангельская область	11294	23,8	1280187	14400,3	32227,2	6
Вологодская область	10349	24,6	1227778	12913,9	24580,8	6
Калининградская область	8794	21,2	937353	12750,3	19684,5	9
Мурманская область	7930	22,8	856969	18581	31579,6	4
Новгородская область	5750	25,6	657595	11004,9	11759,5	1
Псковская область	5901	26,4	713392	8950,5	10685,5	6
Республика Адыгея	3826	23,2	441199	8056,4	4023,7	2
Республика Дагестан	22647	16,2	2658630	5696,4	33773,7	16
Республика Ингушетия	3319	10,6	492669	7285,4	5509,7	3
Кабардино-Балкарская Республика	6875	16,6	891299	7213,9	8984,5	4
Республика Калмыкия	2430	20,9	287199	7101,2	2120,5	1
Карачаево-Черкесская Республика	3717	19,7	428706	7558,4	4704,8	2
Республика Северная Осетия – Алания	5293	25,8	701444	7625,8	10672,5	11
Чеченская Республика	14386	4,9	1183745	9916,7	5260,2	3
Краснодарский край	47558	20,5	5101081	10260	115688,9	35
Ставропольский край	22616	20,6	2701215	8647,7	50361,9	23

Таблица (окончание)

1	2	3	4	5	6	7
Астраханская область	8836	19,8	994127	9866,9	15361,1	5
Волгоградская область	21697	20,8	2619955	9770,2	57988	17
Ростовская область	38940	20,6	4275978	9779,6	72829,8	27
Республика Башкортостан	35635	20	4050989	11027,1	96393,2	16
Республика Марий Эл	6298	21,2	706680	8404,3	8920,3	5
Республика Мордовия	6607	22,9	847645	8103	8973,9	3
Республика Татарстан	32795	21,3	3760534	11468,6	93325,8	33
Удмуртская Республика	13344	19,4	1537858	9838,6	22827,9	7
Чувашская Республика	10906	21,5	1286239	8703,2	19586,4	6
Пермский край	22481	20,3	2730892	11856	56287	15
Кировская область	11560	21,3	1426917	8861,5	21370,4	7
Нижегородская область	27313	22,7	3381328	10302	53392,5	18
Оренбургская область	18466	20,3	2125503	9619,6	30122,4	10
Пензенская область	12030	22,9	1395981	8566,4	19517,8	5
Самарская область	29410	21,4	3178577	11920,7	81728,1	29
Саратовская область	21177	23,7	2595315	9108,3	33749,7	12
Ульяновская область	11104	22,1	1321710	8412,7	20244,1	6
Курганская область	8913	20,1	969304	8883,1	11825,5	4
Свердловская область	39256	21,6	4399738	13986,9	109753,6	31
Тюменская область	36749	19,3	3345127	28565	101898,3	19
Челябинская область	30281	21,5	3516355	11897,5	64683,7	19
Республика Алтай	2088	17,5	205387	9228,2	1531,4	1
Республика Бурятия	8761	18,4	959985	11528,6	13642,5	5
Республика Тыва	2697	12,3	309439	10701,6	2207,3	1
Республика Хакасия	5289	20	536609	11251,2	8496	2
Алтайский край	22059	20,4	2523308	7804,7	34530,2	11
Забайкальский край	10631	19	1122104	12161,5	14064,9	3
Красноярский край	27232	21,4	2893748	15510	63015,5	13
Иркутская область	22814	20,1	2513808	13770	43338,1	15
Кемеровская область	25231	20,7	2826295	12554,9	44092	12
Новосибирская область	25608	20,1	2640656	12017	65727,8	26
Омская область	17084	21,1	2025626	11003,6	37733,8	19
Томская область	9630	20,3	1033102	14429	27800,6	9
Республика Саха (Якутия)	7954	19,8	949972	19409,2	30621	8
Камчатский край	3323	21,9	347123	21814,7	10658,8	3
Приморский край	18073	19,9	2005917	13174,1	52126	11
Хабаровский край	12327	20,1	1405452	15883,5	53844,4	19
Амурская область	7629	20,7	874613	13534,4	14333,1	5
Магаданская область	1672	27,3	168530	22101,6	6178,5	1
Сахалинская область	5040	21,7	521206	23346,3	23358,1	2
Еврейская автономная об- ласть	1518	20,5	185645	11968,9	2648,2	1

Тюляев Олег Николаевич

Международный банковский институт

Кузютин Д.В., к.ф.-м.н., доцент; **Панкратова Я.Б.** ст. преподаватель –
научные руководители

ИССЛЕДОВАНИЕ ФАКТОРОВ, ВЛИЯЮЩИХ НА ПЛОТНОСТЬ ЗАСЕЛЕНИЯ РЕГИОНОВ РОССИЙСКОЙ ФЕДЕРАЦИИ

С 1992 года, за 16 лет, население России сократилось на три процента: с 148,7 миллиона человек до 143,8 миллионов. Россия находится в состоянии глубокого прогрессирующего демографического кризиса, угрожающего как ее территориальной целостности, так и самому существованию государственности.

Как известно, 2008 год прошел под девизом «года семьи», что указывает на озабоченность российского государства данной проблемой. Сейчас в стране пока не чувствуется дефицит кадров, однако в ближайшей перспективе при неизменных демографических трендах ситуация будет ухудшаться.

В сложившейся ситуации, по моему мнению, необходима мощная государственная поддержка, направленная прежде всего на решение социальных вопросов, которые тесно взаимосвязаны с экономической ситуацией. Однако ввиду «специфики» России (ее многонациональности, различиях климатических, экономических и т. д.) выделение определяющего фактора, влияющего на плотность населения, становится весьма сложной задачей.

На сегодняшний день перед нашим правительством стоит задача выбора приоритетного направления в налоговой политике, которое помогло бы изменить в лучшую сторону демографическую ситуацию. В качестве возможных представляются следующие направления: ужесточение налоговой политики с целью перераспределения доходов в пользу малоимущих либо, наоборот, налоговые послабления, чтобы открывалось больше новых предприятий, росли зарплаты работников.

Цель данного исследования – выявить факторы, которые могут влиять на региональную плотность населения Российской Федерации, отобрать из них наиболее значимые, а также установить вид зависимости между численностью населения и этими факторами.

Сбор данных и отбор факторов

Чтобы дать ответ на поставленный вопрос, было проведено математическое исследование, которое должно помочь в разрешении данной дилеммы. В качестве факторов, способных объяснить региональные различия плотности населения, были предложены следующие:

- X_1 – среднедушевые денежные доходы в месяц, руб.;
- X_2 – количество предприятий. Данный фактор дает представление о наличии рабочих мест в каждом регионе;
- X_3 – величина прожиточного минимума, руб.;
- X_4 – оборот розничной торговли, млн руб.;
- X_5 – количество больничных коек на 10 000 человек;
- X_6 – обеспеченность общественным транспортом.

В качестве результирующего признака были использованы данные численности населения в различных регионах Российской Федерации. Для исследования использовались данные, отражающие вариацию факторов и результирующего признака в региональном разрезе, причем были задействованы данные по всем регионам нашей страны. С целью комплексного анализа поставленной задачи в выборку были включены все данные, даже те, которые имеют серьезные отличия в экономическом и социальном развитии (Москва, Санкт-Петербург, Чеченская республика, газово-нефтяные округа).

Информация, которая подверглась статистической обработке, является статичной и взята по данным за 2006 год. Основным источником данных – сайт федеральной службы государственной статистики (gks.ru), некоторые недостающие данные взяты с сайта statistika.ru.

Исследование влияния отдельных факторов

Исследование влияния уровня среднедушевых денежных доходов на численность населения

Для изучения фактора X_1 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 1). При его рассмотрении трудно точно выявить вид зависимости, но можно выдвинуть предположение о существовании двух видов зависимостей:

- Линейная
 $\hat{Y}_X = 0,179X_1 + 276,07.$
- Полулогарифмическая
 $\hat{Y}_X = 1811,7\ln(X_1) - 14439.$

Для того чтобы осуществить выбор в пользу какой-либо из них, необходимо использовать следующие критерии:

- метод абсолютных отклонений. Лучшей из нескольких моделей является та, у которой этот показатель наименьший;
- средняя ошибка аппроксимации. Чем меньше эта ошибка, тем лучше построенная модель аппроксимирует наблюдаемые данные;
- установление характера скедастичности. При установлении гетероскедастичности в модели рекомендуется изменить ее спецификацию, например заменой форм зависимостей.

В данном случае выбор определили результаты проверок значимости коэффициентов:

Линейная регрессия

$Tb=b/Mb$	$Ta=a/Ma$	$t_{табл}=\bar{t}$
4,599674	0,755699	1,990847

Полулогарифмическая регрессия

$Tb=b/Mb$	$Ta=a/Ma$	$t_{табл}=\bar{t}$
4,277933	-3,80388	1,990847

Поле корреляции

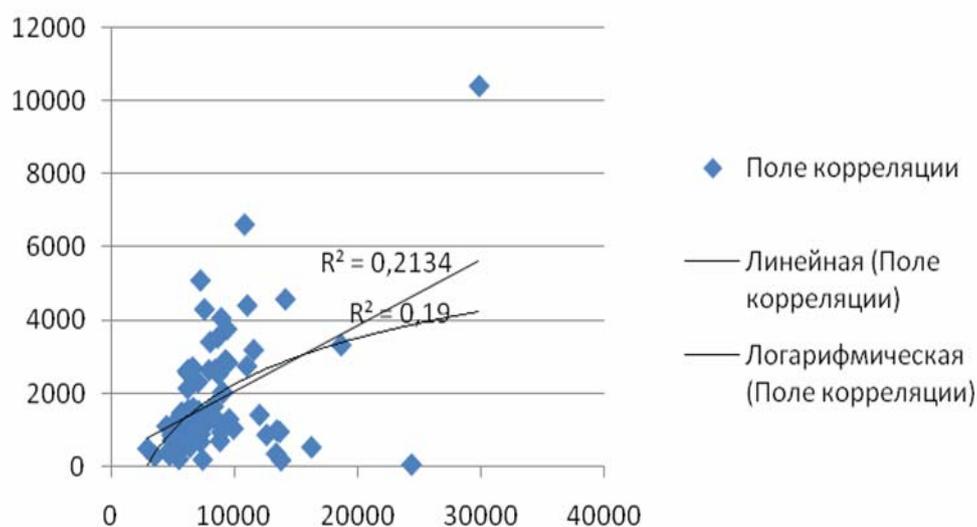


Рис. 1. Корреляционное поле X_1-U

Однако тест ранговой корреляции Спирмена установил наличие гетероскедастичности в полулогарифмической модели

$$(tr_{xe}=6,09 > t_{крит}=1,99).$$

Отсюда можно сделать вывод, что математическая модель, выражающая данную зависимость объясняющей переменной, не подходит для описания зависимой переменной. Поэтому включение данного фактора в модель множественной регрессии нецелесообразно.

Исследование влияния количества предприятий на численность населения

Для изучения фактора X_2 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 2).

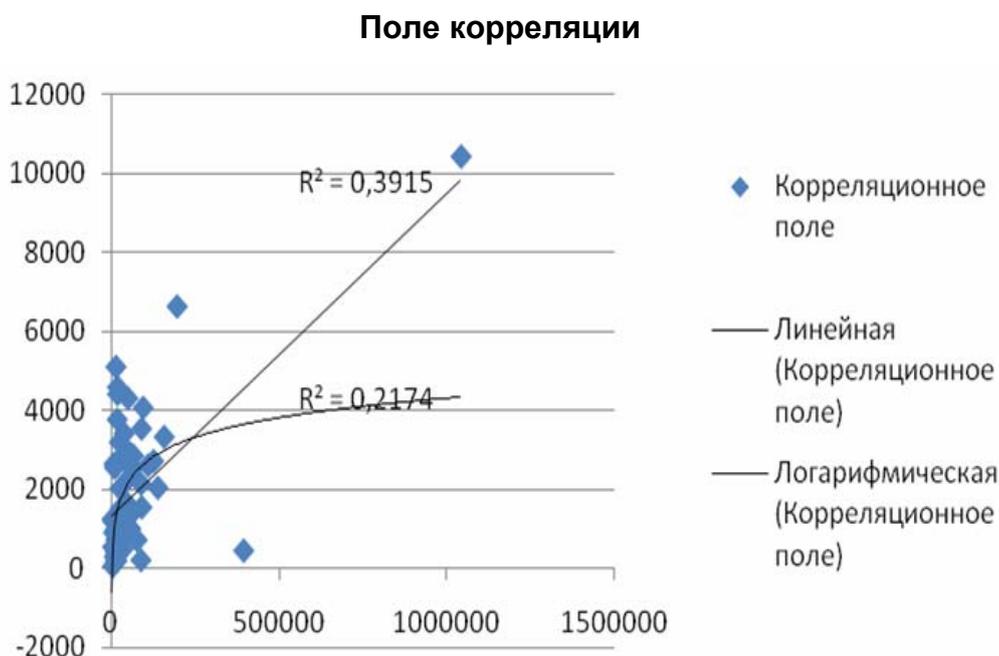


Рис. 2. Корреляционное поле X_2 – Y

При его рассмотрении трудно точно выявить вид зависимости, но можно выдвинуть предположение о существовании двух видов зависимостей:

- Линейная
 $\hat{Y}_x = 0,0082 X_2 + 1324,9$.
- Полулогарифмическая
 $\hat{Y}_x = 710,65 \ln(X_2) - 5517$.

Тип модели	MAD	A
Линейная	924,33	134,25 %
Полулогарифмическая	1019,21	107,21 %

Линейная регрессия

$Tb=b/Mb$	$Ta=a/Ma$	$t_{табл}=\bar{t}$
7,083991	8,5063433	1,990847

$$F_{табл} = 3,96 < F_{набл} = 50,18.$$

Полулогарифмическая регрессия

$Tb=b/Mb$	$Ta=a/Ma$	$t_{табл}=\bar{t}$
4,655041	-3,499084	1,990847

$$F_{табл} = 3,96 < F_{набл} = 21,67.$$

В двух моделях получилось достаточно «странное» значение средней ошибки аппроксимации, что связано с включением ряда аномальных наблюдений (отношение остатков к наблюдаемым значениям зачастую целые числа, так как в линейной модели эта величина для Чукотки более 25). Ввиду чего А будет нести второстепенный характер в определении качества построенных моделей.

Так как R^2 в линейной модели больше, чем в полулогарифмической (0,39 и 0,21), абсолютные отклонения (MAD) меньше в линейной, чем в полулогарифмической, а также, принимая во внимание простоту первой, модель целесообразно строить по линейной регрессии.

Проверка по тесту Голдфелда-Квандта не установила наличия гетероскедастичности в выбранной регрессии ($F_{набл} = 1,09 < F_{табл} = 1,96$).

Оценим тесноту связи с помощью показателей корреляции и детерминации.

Коэффициент корреляции $r=0,63$, что по шкале Чеддока классифицирует такую связь изучаемых признаков, как заметную.

Коэффициент детерминации $R^2 = 0,39$ говорит о том, что 39 % вариации плотности населения объясняется вариацией количества предприятий. Остальные 61% вариации объясняются неучтенными в данной модели факторами.

Отсюда можно сделать вывод, что данная математическая модель вполне способна отражать экономические реалии (коэффициент 0,0082 при X_2 значит, что с открытием 10 000 предприятий население возрастает на 820 тысяч человек). С вероятностью 95 % коэффициент при x будет варьироваться в интервале от 0,0058 до 0,0105. Рассмотренный фактор будет включен в модель множественной регрессии.

Исследование влияния величины прожиточного минимума на численность населения

Для изучения фактора X_3 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 3).

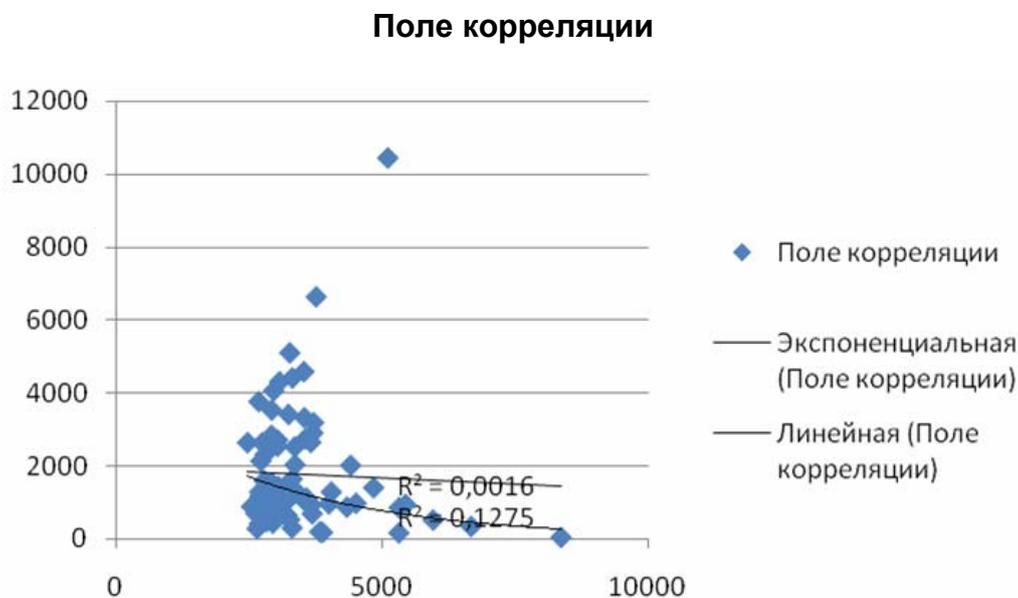


Рис. 3. Корреляционное поле X_3 – Y

При его рассмотрении трудно точно выявить вид зависимости, но можно выдвинуть предположение о существовании трех видов зависимостей:

- Линейная
 $\hat{Y}_X = -0,0662 X_3 + 2013,4$.
- Гиперболическая
 $\hat{Y}_X = 2063 - 913363 / X_3$.
- Экспоненциальная
 $\hat{Y}_X = 3855,2e^{-3E-04X_3}$.

В данном случае выбор определили результаты проверки значимости коэффициентов и моделей в целом:

Линейная регрессия

$T_b = b/M_b$	$T_a = a/M_a$	$t_{\text{табл}} =$
-0,35764	3,025789	1,990847

$$F_{\text{табл}} = 3,96 > F_{\text{набл}} = 0,13.$$

Гиперболическая регрессия

$Tb=b/Mb$	$Ta=a/Ma$	$t_{\text{табл}}=$
-0,29836835	2,168977	1,990847

$$F_{\text{табл}} = 3,96 > F_{\text{набл}} = 0,09.$$

Экспоненциальная регрессия

$Tb=b/Mb$	$Ta=a/Ma$	$t_{\text{табл}}=$
-3,37669	11244,72	1,990847

$$F_{\text{табл}} = 3,96 < F_{\text{набл}} = 11,4.$$

Тест ранговой корреляции Спирмена установил наличие гетероскедастичности в экспоненциальной модели ($tr_{\text{хе}}=2,62 > t_{\text{крит}}=1,99$).

Отсюда можно сделать вывод, что математическая модель, выражающая данную зависимость объясняющей переменной, не подходит для описания зависимой переменной. Поэтому включение данного фактора в модель множественной регрессии нецелесообразно.

Исследование влияния объема розничной торговли на численность населения

Для изучения фактора X_4 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 4).

Корреляционное поле

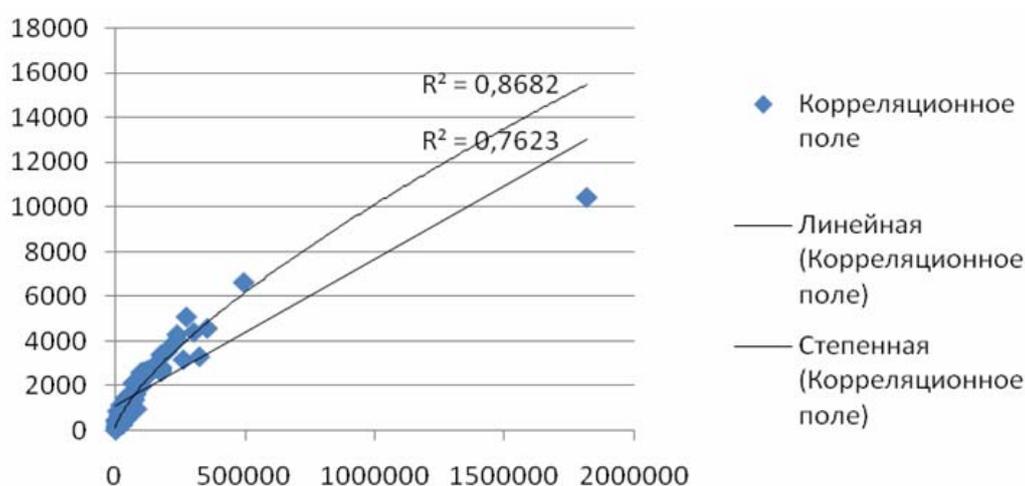


Рис. 4. Корреляционное поле X_4 – Y

При анализе корреляционного поля и подборе лучших линий тренда я остановился на двух зависимостях:

- Линейная
 $\hat{Y}_X = 0,0066 X_4 + 1071,2$.
- Степенная
 $\hat{Y}_X = 0,526 X_4^{0,7141}$.

Тип модели	MAD	A
Линейная	595,02	88,65 %
Степенная	328,3	24,42 %

Линейная регрессия

Tb=b/Mb	Ta=a/Ma	t _{табл} =
15,816	10,78496	1,990847

$$F_{\text{табл.}} = 3,96 < F_{\text{набл.}} = 250,14.$$

Степенная регрессия

Tb=b/Mb	Ta=a/Ma	t _{табл} =
22,66339	1,522709	1,990847

$$F_{\text{табл}} = 3,96 < F_{\text{набл}} = 513,63.$$

Несмотря на статистическую незначимость коэффициента a ($a = 1071,2$) в степенной регрессии, предпочтение в выборе модели будет отдано ей.

Тест ранговой корреляции Спирмена установил наличие гетероскедастичности в степенной модели ($tr_{xe}=4,3 > t_{\text{крит}}=1,99$). Однако все равно эта модель имеет хорошую объяснительную способность и хорошо отражает экономические реалии.

Коэффициент корреляции $r=0,93$, что по шкале Чеддока классифицирует такую связь изучаемых признаков как высокую.

Коэффициент детерминации $R^2 = 0,86$ говорит о том, что 86 % вариации плотности населения объясняется вариацией количества преступлений. Остальные 14 % вариации объясняются неучтенными в данной модели факторами.

Принимая во внимание вышесказанное, математическая модель, построенная на основе степенной регрессии, вполне способна отражать экономические реалии.

Уравнение $\hat{Y}_X = 0,526 X_4^{0,7141}$ показывает нам постоянную эластичность $\Xi=0,71$, что свидетельствует о прямой привлекательности региона от оборота розничной торговли (увеличение последней на 1 %, увеличива-

ет население на 0,7 %). Однако рассмотренный фактор будет включен в модель множественной регрессии, ввиду того, что факторы – показатель плотности населения и объем розничной торговли – являются следствием друг друга. Поэтому фактор X_4 не способен помочь в разрешении поставленной задачи.

Исследование влияния количества больничных коек на 10 000 человек на численность населения

Для изучения фактора X_5 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 5). При его рассмотрении трудно точно выявить вид зависимости, но можно выдвинуть предположение о существовании трех видов зависимостей:

- Линейная
 $\hat{Y}_X = -19,517 X_5 + 4056.$
- Степенная
 $\hat{Y}_X = 3E+06 X_5^{1,66}.$
- Экспоненциальная
 $\hat{Y}_X = 10932e^{-0,019X_5}.$

Тип модели	MAD	A
Линейная	1041	122 %
Степенная	1009	88 %
Экспоненциальная	975	77 %

Линейная регрессия

$T_b=b/M_b$	$T_a=a/M_a$	$t_{\text{табл}}=$
-2,86223	4,993843	1,990847

$$F_{\text{табл.}} = 3,96 < F_{\text{набл.}} = 8,19.$$

Степенная регрессия

$T_b=b/M_b$	$T_a=a/M_a$	$t_{\text{табл}}=$
-3,96047	7,5516895	1,990847

$$F_{\text{табл}} = 3,96 < F_{\text{набл}} = 15,68.$$

Экспоненциальная регрессия

$Tb=b/Mb$	$Ta=a/Ma$	$t_{табл} =$
-5,53344	23,320532	1,990847

$$F_{табл} = 3,96 < F_{набл} = 30,61.$$

Поле корреляции

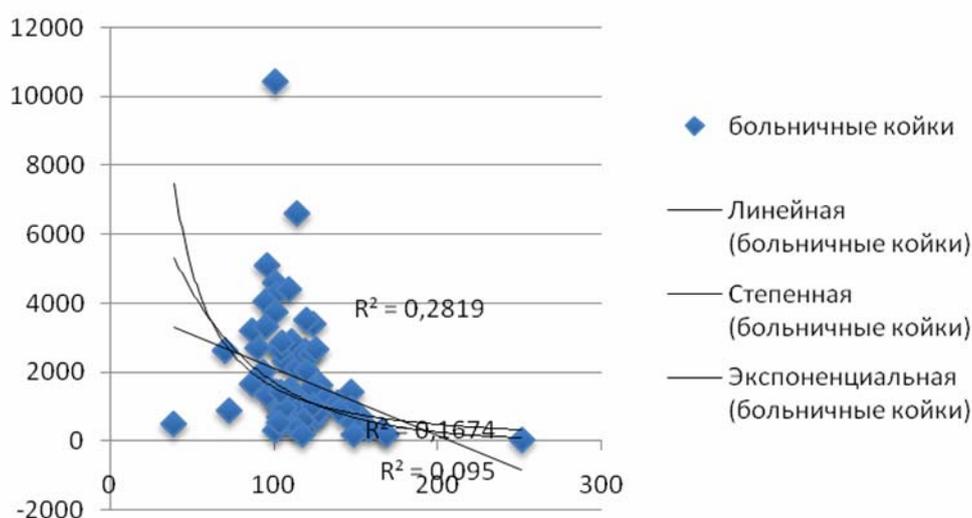


Рис. 5. Корреляционное поле X_5 - Y

Так как R^2 в экспоненциальной модели больше, чем в степенной и линейной (0,28, 0,17 и 0,095), абсолютные отклонения (MAD) наименьшие в экспоненциальной, модель целесообразно строить по экспоненциальной регрессии.

Тест ранговой корреляции Спирмена не установил наличие гетероскедастичности в степенной модели ($tr_{хе}=0,35 < t_{крит}=1,99$).

Коэффициент корреляции $r=0,53$, что по шкале Чеддока классифицирует такую связь изучаемых признаков, как заметную.

Коэффициент детерминации $R^2 = 0,28$ говорит о том, что 28 % вариации плотности населения объясняется вариацией количества преступлений. Остальные 62 % вариации объясняются неучтенными в данной модели факторами.

Рассмотренный фактор будет включен в модель множественной регрессии.

Исследование влияния числа автобусов общего пользования на 100 000 человек на численность населения

Для изучения фактора X_6 на результирующий признак Y необходимо сначала построить поле корреляции (рис. 6).

Поле корреляции

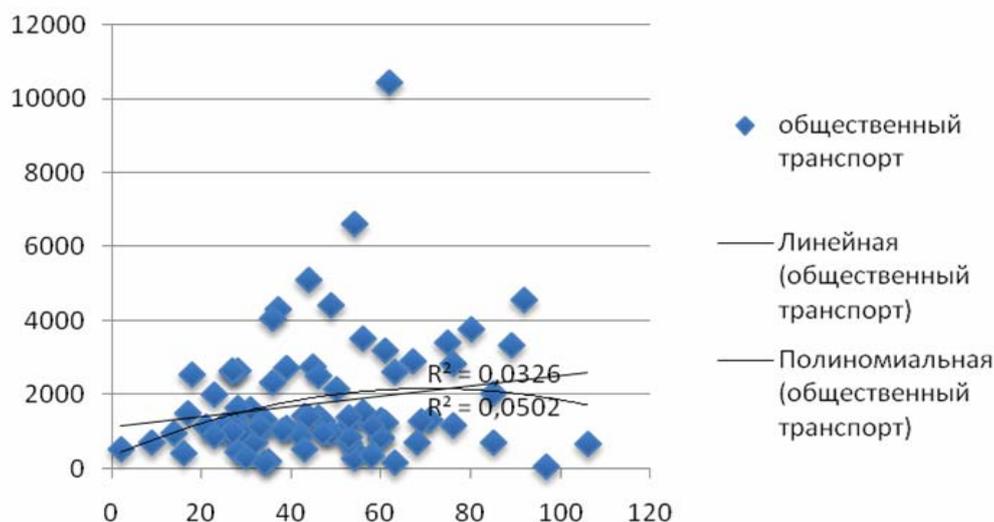


Рис. 6. Корреляционное поле X_6 – Y

При его рассмотрении трудно точно выявить вид зависимости, но можно выдвинуть предположение о невозможности построения каких-либо моделей. Для «лучших» регрессионных моделей (линейной и полиномиальной) коэффициенты R^2 чрезвычайно малы 0,05 и 0,03 соответственно.

Дальнейшее изучение влияния фактора на показатель Y нецелесообразно.

Исследование влияния совокупности факторов

В качестве факторов, оказывающих влияние на уровень доходов, после предварительного исследования были отобраны только два:

- X_2 – количество предприятий;
- X_5 – количество больничных коек.

Была спроектирована следующая многофакторная модель:

$$Y_{X_2X_5} = 3099,8 - 15,07 \times X_2 + 0,008 X_5,$$

которую можно интерпретировать следующим образом: 3099,8 тыс. людей – базовый уровень населения; с открытием 10 000 предприятий население возрастает на 800 тысяч человек, увеличение количества больничных коек на одну уменьшает численность населения на 15,07 тысяч чело-

век. Последнее справедливо лишь для российской практики и свидетельствует о недостаточном уровне медицинского обеспечения.

1. Доверительные интервалы: для коэффициента при уровне значимости в 95 %:

- для **a** от 1799 до 4400;
- для **b1** от -25,82 до -4,32;
- для **b2** от 0,006 до 0,01.

Следовательно, все коэффициенты статистически значимы.

2. Корреляционная матрица

	Y	X ₅	X ₂
Y	1	0,461919	0,625695
X ₅	0,461919	1	0,5109
X ₂	0,625695	0,5109	1

Наблюдалась средней силы межфакторная связь в 0,51, т. е. не приходится говорить о наличии мультиколлинеарности.

3.

\bar{A}	MAD	Скорр. R ²	F	F _{табл}
106%	850,83	0,43	31,17	3,96

4. Гетероскедастичность отмечена по фактору «больничные койки» ($F_{набл}=5,11$, $F_{табл}=2,01$), по фактору «количество предприятий» не отмечена ($F_{набл}=0,87$, $F_{табл}=2,01$) согласно тесту Голдфелда-Квандта.

Теперь произведем сравнение построенной модели множественной регрессии с моделями парных регрессий. Это сравнение производится для того, чтобы осуществить выбор между четырьмя моделями:

Модель	R ²	\bar{A}	MAD
Экспоненциальная (больничные койки)	0,282	76,5 %	975,73
Экспоненциальная (прожиточный минимум)	0,128	79,6 %	1005,76
Логарифмическая (денежные доходы)	0,19	196 %	991,7
Степенная (оборот розничной торговли)	0,87	24,42 %	328,3
Линейная (кол-во предприятий)	0,392	134 %	924,3
Множественная (кол-во больничных коек и кол-во предприятий)	0,45 (скорр. 0,43)	106 %	850,83

Как видно из таблицы, лучшей объяснительной способностью обладает парная регрессия с участием фактора «оборот розничной торговли»,

однако для решения поставленной задачи более интересна множественная модель с участием факторов количество предприятий и количество больничных коек.

ЗАКЛЮЧЕНИЕ

В процессе исследования было изучено влияние каждого из приведенных выше факторов на результирующий признак – региональная плотность населения – как в отдельности, так и всех факторов в совокупности, была проанализирована их значимость и был установлен вид зависимости результирующего признака от этих факторов. Для этого были построены шесть моделей парной регрессии и одна модель множественной регрессии. При этом в ходе исследования выяснилось, что в ряде случаев нелинейная регрессия лучше отображает имеющуюся зависимость, нежели линейная.

Из проведенных исследований можно сделать следующий вывод: социальные составляющие менее важны для населения, чем возможность работать и зарабатывать. Это связано со следующими причинами: человек выбирает место жительства по комплексным параметрам, поэтому улучшение каждого в отдельности не будет приносить эффекта постоянной отдачи, наблюдается эффект насыщения фактором.

Принимая во внимание вышесказанное, хочется еще раз отметить сложность предстоящего выбора для государства приоритетных направлений, способных улучшить качество жизни населения, а в свою очередь, и демографическую ситуацию. Поэтому к вопросу надо подходить комплексно, принимая в расчет разнородности регионов, однако ключевым фактором в решении соответствующих задач должны быть меры по улучшению их экономик. Так, стимулирование бизнеса, уменьшение налогового бремени, создание свободных экономических зон может принести большие плоды в решении демографической проблемы, чем увеличение социальной обеспеченности населения.

ЛИТЕРАТУРА

1. Буре В.М., Евсеев Е.А. Эконометрика: Учеб. пособие. СПб.: Изд-во МБИ, 2007. – 168 с.
2. Тарашнина С.И., Панкратова Я.Б. Выполнение курсовой работы по эконометрике: Учеб.-метод. пособие. СПб.: Изд-во МБИ, 2007. – 97 с.
3. Доугерти К. Введение в эконометрику. М.: ИНФРА-М, 2004.
4. Официальный сайт Федеральной службы государственной статистики: <http://www.gks.ru>
5. Сайт статистики РФ <http://www.statistika.ru>

Данные

Регион	Численность населения, Y	Среднедушевые денежные доходы (в месяц), руб., X ₁	Кол-во предприятий, X ₂	Величина прожиточного минимума, X ₃	Оборот розничной торговли, млн руб., X ₄	Число больничных коек на 10 000 человек населения, X ₅	Число автобусов общего пользования на 100 000, X ₆
1	2	3	4	5	6	7	8
Белгородская область	1511	7083,6	26538	2901	62616	110	17
Брянская область	1331	6111,6	19669	2873	48205	127	60
Владимирская область	1473	5684,7	30673	3222	40721	99	28
Воронежская область	2314	7020,2	54984	2813	93569	107	36
Ивановская область	1100	4464,8	27316	3090	29381	114	39
Калужская область	1014	6994,1	25114	3197	49220	114	33
Костромская область	709	6227,7	16471	2952	22793	153	68
Курская область	1184	6706,9	24379	2849	48814	114	53
Липецкая область	1181	7706	21005	2721	53124	133	76
Московская область	6628	10797,7	194881	3775	495548	114	54
Орловская область	834	5875,4	14943	2629	29415	119	60
Рязанская область	1182	6175,4	31450	2966	44548	120	25
Смоленская область	1006	6870,6	22720	3187	46881	125	48
Тамбовская область	1130	6869,9	17663	2736	48269	122	21
Тверская область	1407	7311,1	50027	2990	70216	127	46
Тульская область	1600	6636,2	36081	2795	61144	128	31
Ярославская область	1328	8153,6	43131	2995	52671	130	71
Москва	10425	29802,6	1041326	5121	1817771	101	62
Республика Карелия	698	8802,5	19687	3699	34827	118	9
Республика Коми	985	13406,8	21181	4525	82144	116	29
Архангельская область	1291	9539,4	23068	4060	64708	114	47
Вологодская область	1235	8673,2	1024	3450	44860	118	61
Калининградская область	940	8887,6	32549	3696	44555	104	26

Таблица (продолжение)

1	2	3	4	5	6	7	8
Ленинградская область	1644	8288,3	45944	3321	79166	87	28
Мурманская область	864	12580	44652	5333	53115	110	60
Новгородская область	665	7207	20528	3225	27391	126	106
Псковская область	725	6382,3	13473	2951	33161	124	85
Санкт-Петербург	4581	14097,7	17246	3544	354857	101	92
Республика Адыгея	443	4707,4	393399	2962	13089	111	28
Республика Дагестан	2641	6260,7	7099	2484	108645	70	28
Республика Ингуш- тия	487	2923,8	34522	2839	3809	39	54
Кабардино- Балкарская Респу- блика	894	5080,1	5984	2558	10398	73	42
Республика Калмы- кия	289	3521,1	7886	2664	28401	101	54
Карачаево-Черкес- ская Республика	431	5534,2	11511	2709	5175	115	16
Республика Северная Осетия – Алания	702	6027,3	14885	2638	15712	103	32
Чеченская Республи- ка	1163	5973,6	5934	3132	23094	102	58
Краснодарский край	5096	7220	13011	3278	274693	96	44
Ставропольский край	2710	6587,3	124861	3050	123896	90	39
Астраханская область	994	7089,6	56899	2883	42088	117	27
Волгоградская область	2636	7911,5	18123	2762	117868	119	63
Ростовская область	4304	7541,4	48169	3086	239407	102	37
Республика Башкор- тостан	4063	8909,4	93834	2973	239124	95	36
Республика Марий Эл	712	4909,5	75785	2789	20305	121	53
Республика Мордовия	857	4878,5	14872	2816	22466	127	53
Республика Татарстан	3762	9368,9	16598	2695	213946	101	80
Удмуртская Респу- блика	1544	6227,8	89732	2947	49737	123	56
Чувашская Республи- ка	1292	5383,5	37955	2712	37213	108	69
Пермский край	2748	10982,3	23527	3571	178031	104	45
Кировская область	1443	5772,5	61395	3066	43930	147	43
Нижегородская об- ласть	3411	8016,1	37410	3255	177636	124	75

Таблица (окончание)

1	2	3	4	5	6	7	8
Оренбургская область	2138	6174,9	87317	2734	69487	114	50
Пензенская область	1408	5658,9	40368	2999	47514	102	44
Самарская область	3189	11529,5	23015	3727	262912	87	61
Саратовская область	2608	6134,5	108806	2919	99902	104	27
Ульяновская область	1336	6157,1	55065	2936	53983	96	34
Курганская область	980	6535,3	28818	2977	39879	105	49
Свердловская область	4410	11012,2	17300	3324	303376	109	49
Тюменская область	3323	18621,9	156414	3555	324694	95	89
Челябинская область	3531	8613,5	89221	2939	198686	120	56
Республика Алтай	204	5470,4	87223	3841	5165	117	35
Республика Бурятия	964	7020	9993	4014	42270	107	14
Республика Тыва	309	4720,1	15863	3321	6742	166	30
Республика Хакасия	538	6366,7	3203	3275	16182	103	43
Алтайский край	2543	6255,9	9041	3049	105612	117	18
Красноярский край	2906	9242,4	55846	3716	148700	111	67
Иркутская область	2527	8684	64889	3373	128017	122	46
Кемеровская область	2839	9442,8	65351	2933	178979	105	76
Новосибирская область	2650	8399,4	47048	3674	176559	126	27
Омская область	2035	8993,3	137828	3375	101514	120	85
Томская область	1034	9896,5	52564	3572	56651	124	38
Читинская область	1128	6999,3	33864	3588	47346	139	33
Республика Саха (Якутия)	950	13629,7	15314	5456	63641	147	47
Приморский край	2019	9040,6	24518	4423	101578	93	23
Хабаровский край	1412	12011,2	57423	4854	74310	104	53
Амурская область	881	7232,9	39071	4353	35042	140	23
Камчатская область	349	13346,9	12170	6688	17621	158	58
Магаданская область	172	13748,2	11648	5329	8338	169	63
Сахалинская область	526	16211,5	7398	5974	42580	148	2
Еврейская автономная область	187	7395,6	15057	3894	8126	149	34
Чукотский автономный округ	51	24320,1	3242	8379	2376	251	97

Фокина Анастасия Алексеевна

Международный банковский институт

Кузютин Д.В., к.ф.-м.н., доцент; **Евсеев Е.А.**, к.ф.-м.н., доцент –
научные руководители

ИССЛЕДОВАНИЕ ФАКТОРОВ, ВЛИЯЮЩИХ НА УРОВЕНЬ АБОРТОВ В РАЗЛИЧНЫХ СТРАНАХ МИРА

Введение

Аборт – это слово режет слух, об этом не говорят, об этом не пишут, в средствах массовой информации это слово практически не встречается, но ежедневно в специальных заведениях только в России ежедневно убивают 21 000 беззащитных детей. Аборты совершаются во всех странах мира, вне зависимости от того, разрешены ли они законом и противоречат ли это религии или нет.

Когда заходит речь об абортах, почему-то принято называть какие-нибудь причины: тяжелое материальное положение, болезнь, социальные причины, недостаток времени на детей и пр. Даже возникла довольно циничная поговорка: «Зачем плодить нищету?». И сегодня, в условиях экономического кризиса, все больше женщин склоняются к тому, чтобы прервать свою беременность. Так как на тему абортотв люди стараются не распространяться, а, наоборот, поскорее забыть об этом, как о страшном сне, то мне стало интересно, что же толкает женщин на такой ответственный шаг.

Целью данной работы является анализ существования или отсутствия зависимости между количеством абортотв, совершенных в разных странах, и группой факторов, выбранных в результате исследования.

1. Сбор данных и отбор факторов

Для проведения исследования были отобраны следующие факторы:

- X_1 – количество заключенных по странам, рассчитанное на 100 тысяч человек (Prisoners > Per capita (most recent) by country);
- X_2 – количество мужчин, покончивших жизнь самоубийством, рассчитанное на 100 тысяч человек (Suicide rate > Young males (most recent) by country);
- X_3 – доход на человека в долларах США (Income (most recent) by country);

- X_4 – число не очень счастливых людей в стране (%);
- X_5 – количество людей, принимающих наркотики, рассчитанное на 100 тысяч человек.

В качестве результирующего признака (количества абортс на 100 тысяч человек) были использованы данные по суммарной величине абортс за год по каждой отдельно взятой стране. В исследовании использовались страны-представители всех континентов Земли, имеющие совершенно разные культурные характеристики и экономические положения. В связи с отсутствием данных из выборки были исключены такие страны, как Андорра, Бахрейн, Гренландия, Гонконг, Кыргызстан, Македония, Сингапур, Таджикистан, Тунис, Узбекистан, Венесуэла, что сузило объем выборки до 50 стран. В исследовании использовались данные Johnston's Archive за 2008 год [5].

Далее будет изучено влияние каждого из приведенных выше факторов на количество абортс в странах по отдельности и влияние этих факторов в совокупности.

2. Исследование влияния отдельных факторов

2.1. Влияние числа заключенных на количество абортс

Предположим, что с увеличением числа заключенных количество абортс также увеличится, т. к. женщины не хотят быть матерями-одиночками или рожать в тюрьме.

Проверим коррелированность фактора и результирующего признака: $r_{yx_1} = 0,68793269$. Согласно шкале Чеддока, такая связь изучаемых признаков является заметной.

При рассмотрении поля корреляции (рис. 1) трудно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных видов зависимостей:

- Линейная зависимость. Тогда:
 $Y = 26,96 + 1,30296 \cdot X_1$ – уравнение;
 $R^2 = 0,473251386$ – коэффициент детерминации;
 $\bar{A} = 312,977 \%$ – средняя ошибка аппроксимации.
- Полиномиальная зависимость. Тогда:
 $Y = 130,77 + 0,1247 \cdot X_1 + 0,0022 \cdot X_1^2$ – уравнение;
 $R^2 = 0,5079$ – коэффициент детерминации;
 $\bar{A} = 305,292 \%$ – средняя ошибка аппроксимации.

Графические изображения этих моделей показаны на рис. 1.

- Кусочно-линейная зависимость.

Зависимость количества абортот от числа заключенных

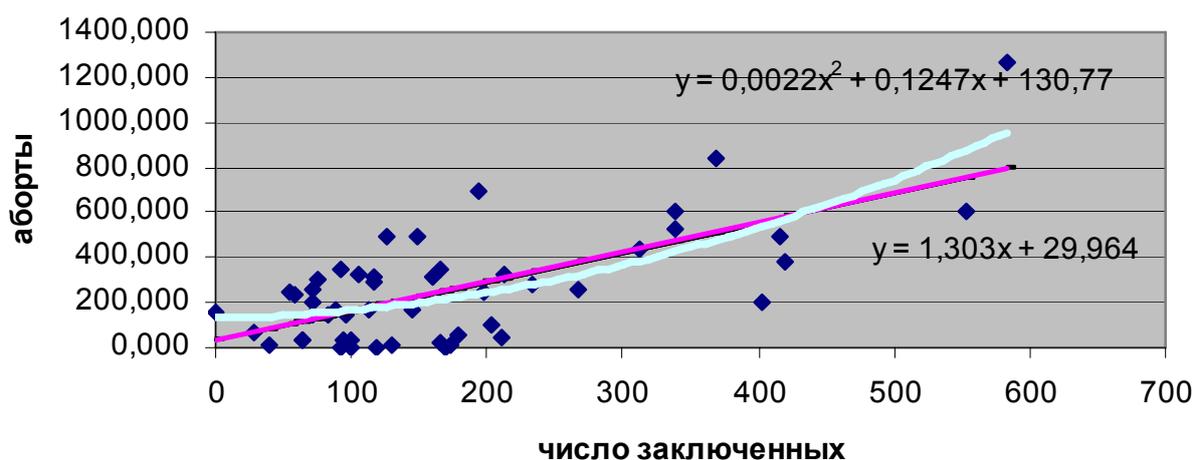


Рис. 1

В процессе анализа поля корреляции мы можем убедиться, что существуют достаточно сильные различия между двумя группами стран в зависимости от числа заключенных. Если заключенных меньше 200, то наблюдается достаточно низкий уровень абортов в стране, который с увеличением числа заключенных повышается. Если же количество заключенных превышает 200, то число абортов в стране находится на схожем высоком уровне в большинстве стран. Поэтому целесообразно будет разделить нашу выборку на две группы. Проверим оправданность такого шага с помощью теста Чоу. Для этого построим линейные регрессии для каждой группы:

- Заключенных < 200 :

$$Y = 65,69 - 0,095 \cdot X_1 - \text{уравнение};$$

$$R^2 = 0,011114798 - \text{коэффициент детерминации};$$

$$S_{\text{ост}} = 72,8090401 - \text{стандартное отклонение остатков}.$$

- Заключенных > 200 :

$$Y = 186,94 - 1,094 \cdot X_1 - \text{уравнение};$$

$$R^2 = 0,506 - \text{коэффициент детерминации};$$

$$S_{\text{ост}} = 167,32 - \text{стандартное отклонение остатков}.$$

Проведя необходимые в тесте Чоу вычисления статистики, получаем $F = -5,471976831$, что по модулю больше критического значения распределения Фишера для вероятности 0,05 и степеней свободы 2 и 46, которое равно 3,19958171.

Таким образом, мы убедились, что деление выборки на две части по данному признаку является оправданным. Введем фиктивную переменную d , принимающую значение 0 в случае, если заключенных меньше 200 человек, и значение 1 в случае, если количество заключенных превы-

шает 200 человек. Построим регрессию по трем факторам (X_1 , d и X_1d). Тогда:

$$Y = 65,69 + 0,095 \cdot X_1 + 121,26 \cdot X_1 \cdot d + 0,999 \cdot d - \text{уравнение};$$

$$R^2 = 0,74904 - \text{коэффициент детерминации};$$

$$\bar{A} = 132\% - \text{средняя ошибка аппроксимации}.$$

График кусочно-линейной модели изображен на рис. 2.

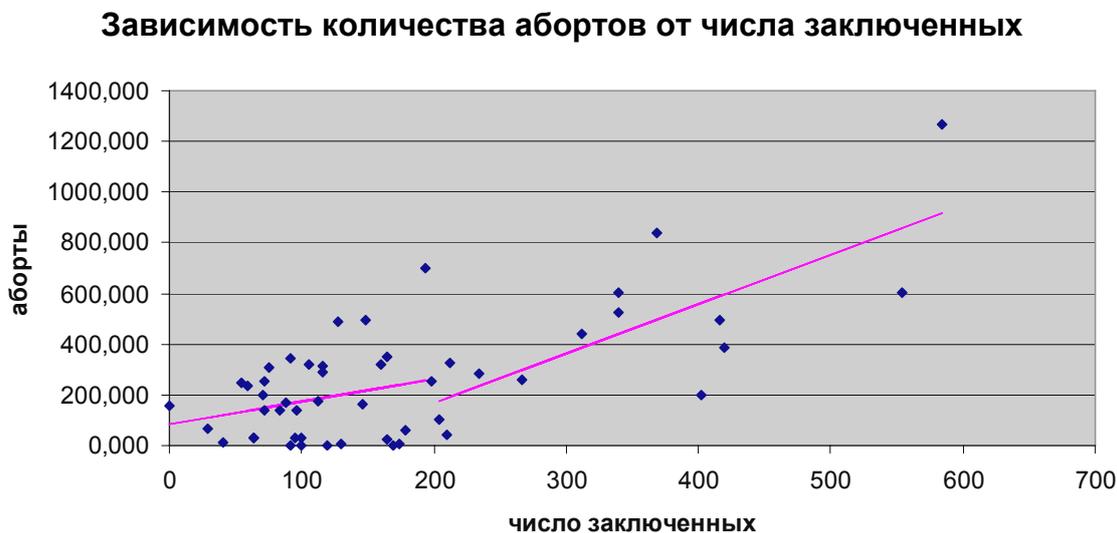


Рис. 2

Сравним полученные результаты по кусочно-линейной модели с предыдущими зависимостями. Кусочно-линейная модель обладает наибольшим коэффициентом детерминации среди всех моделей. Это говорит о том, что 74,9 % вариации уровней абортотв в стране объясняется вариацией числа заключенных, построенной кусочно-линейной регрессией, 50,8 % у полиномиальной и 47,3 % у линейной. Также кусочно-линейная модель обладает наименьшей средней ошибкой аппроксимации в 132 % при 305 % у полиномиальной и 313 % у линейной.

Проверим статистическую значимость полученных уравнений:

- F-статистика в линейной модели = 43,125;
- F-статистика в полиномиальной модели = 24,25;
- F-статистика в кусочно-линейной модели = 45,76.

Все три полученные F-статистики значительно превышают критические значения распределения Фишера для вероятности 0,05 и со степенями свободы: для линейной – 1 и 48, равное 4,043, для полиномиальной – 2 и 47, равное 3,195, для кусочно-линейной – 3 и 46, равное 2,807.

Проверим статистическую значимость коэффициентов кусочно-линейной модели:

- $t_a = 1,384553494$;

- $t_b = 0,286903583$;
- $t_{d1} = 1,807302938$;
- $t_{j1} = 2,669026412$.

Не все t-статистики превышают критическое значение распределения Стьюдента с вероятностью 0,05 и 46 степенями свободы, равное 2,012895567. Отсюда можно сделать вывод, что значим только последний коэффициент кусочно-линейной модели.

Следовательно, кусочно-линейную модель не рационально использовать для исследования зависимости между количеством аборт в стране и количеством заключенных.

Проверим статистическую значимость коэффициентов полиномиальной модели:

- $t_a = 1,877$;
- $t_{b1} = 0,184$;
- $t_{b2} = 1,819$.

Мы можем наблюдать, что t-статистики не превышают критическое значение распределения Стьюдента с вероятностью 0,05 и 47 степенями свободы, равное 2,0117. Отсюда можно сделать вывод, что коэффициенты незначимы.

Таким образом, исследовав три модели, можно сделать вывод, что наиболее подходящей для исследования является линейная зависимость. Уравнение такой модели – $Y = 26,96 + 1,30296 \cdot X_1$. Коэффициент детерминации $R^2 = 0,473251386$ говорит о том, что 47,3 % вариации уровня аборт объясняется вариацией уровня заключенных. Можно сделать вывод, что если число заключенных будет равно нулю, в стране ежегодно будут совершаться 29,96 аборт на сто тысяч населения. С увеличением числа заключенных на 1 % количество аборт будет увеличиваться в среднем на 1,30269. Поэтому включение данного фактора в модель множественной регрессии целесообразно.

2.2. Влияние числа мужчин, покончивших жизнь самоубийством, на количество аборт

Предположим, что многие аборт совершаются по причине отсутствия мужа, который покончил жизнь самоубийством.

Проверим коррелированность фактора и результирующего признака: $r_{yx2} = 0,352255597$. Согласно шкале Чеддока, такая связь изучаемых признаков является умеренной.

По данному полю корреляции (рис. 3) трудно выдвинуть предположение о наличии какой-либо связи, поэтому построим наиболее простую модель – линейную. Тогда:

- $Y = 124,03 + 5,9074 \cdot X_2$ – уравнение;
- $R^2 = 0,12408$ – коэффициент детерминации;
- $\bar{A} = 306,8 \%$ – средняя ошибка аппроксимации.

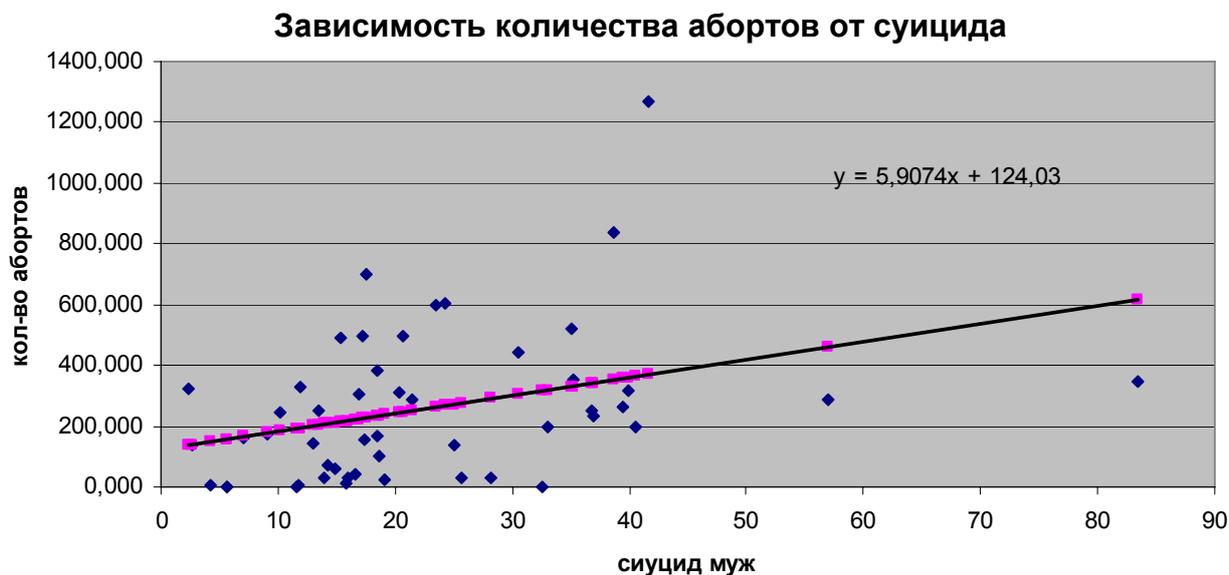


Рис. 3

Проверим статистическую значимость полученного уравнения.

F-статистика = 6,799775651. Полученная F-статистика превышает критическое значение распределения Фишера для вероятности 0,05 и со степенями свободы 1 и 48, равное 4,042652.

Проверив статистическую значимость коэффициентов линейной модели, мы увидели, что коэффициенты значимы, т. к. превышают критическое значение распределения Стьюдента с вероятностью 0,05 и 49 степенями свободы, равное 2,0106.

Таким образом, мы принимаем линейную модель для исследования зависимости между количеством абортотв в стране и числом суицидов. При этом можно сделать вывод о том, что если бы количество суицидов стремилось к нулю, в стране ежегодно убивали бы 124,03 нерожденных детей на сто тысяч населения. При приросте числа суицидов на 0,01 уровень абортотв будет увеличиваться на 5,907 единиц.

2.3. Влияние дохода (на человека) на количество абортотв

Предположим, что большинство абортотв совершается из-за отсутствия средств для воспитания и для жизни ребенка.

Проверим коррелированность фактора и результирующего признака: $r_{yx3} = -0,312$. Согласно шкале Чеддока, такая связь изучаемых признаков является умеренной.

Зависимость количества абортотв от дохода

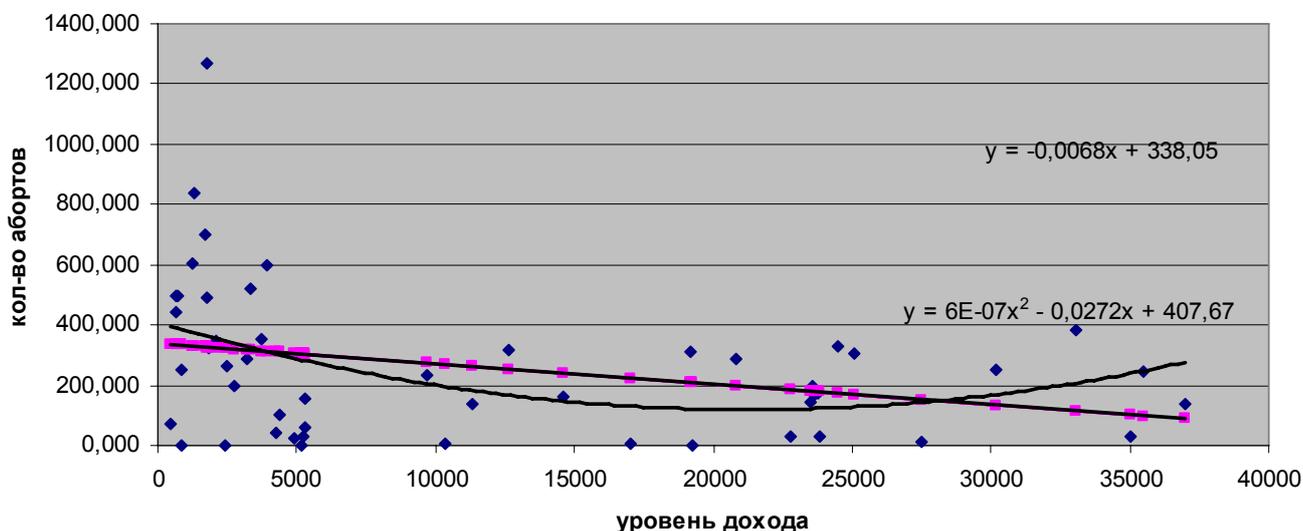


Рис. 4

При рассмотрении поля корреляции (рис. 4) трудно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных видов зависимостей:

- Линейная зависимость. Тогда:
 $Y = 338,052 - 0,0068 \cdot X_3$ – уравнение;
 $R^2 = 0,0973$ – коэффициент детерминации;
 $\bar{A} = 444 \%$ – средняя ошибка аппроксимации.
- Полиномиальная зависимость. Тогда:
 $Y = 407,666 - 0,0271 \cdot X_3 + 0,000000638 \cdot X_3^2$ – уравнение;
 $R^2 = 0,164$ – коэффициент детерминации;
 $\bar{A} = 380 \%$ – средняя ошибка аппроксимации.
- Кусочно-линейная зависимость.

В процессе анализа поля корреляции мы можем убедиться, что существуют достаточно сильные различия между двумя группами стран в зависимости от доходов их жителей. Если душевой доход меньше 10 000 долларов США, то наблюдается достаточно высокий уровень абортотв в стране, который с увеличением дохода снижается. Если же душевой доход в стране превышает 10 000 долларов США, то уровень абортотв в стране находится на схожем низком уровне в большинстве стран. Поэтому целесообразно будет разделить нашу выборку на две группы. Про-

верим оправданность такого шага с помощью теста Чоу. Для этого построим линейные регрессии для каждой группы:

- Доход меньше 10 000 \$:

$$Y = 480,421 - 0,0525 \cdot X_3 - \text{уравнение};$$

$$R^2 = 0,132 - \text{коэффициент детерминации};$$

$$S_{\text{ост}} = 287,11 - \text{стандартное отклонение остатков}.$$

- Доход больше 10 000 \$:

$$Y = 235,012 - 0,000239 \cdot X_3 - \text{уравнение};$$

$$R^2 = 0,0226 - \text{коэффициент детерминации};$$

$$S_{\text{ост}} = 141,69 - \text{стандартное отклонение остатков}.$$

Проведя необходимые в тесте Чоу вычисления статистики, получаем $F = -10,14999625$, что по модулю больше критического значения распределения Фишера для вероятности 0,05 и степеней свободы 2 и 46, которое равно 3,199581706.

Таким образом, мы убедились, что разделение выборки на две части по данному признаку является оправданным. Введем фиктивную переменную d , принимающую значение 0 в случае, если доход меньше 10 000 долларов США, и значение 1 в случае, если доход превышает 10 000 долларов США.

Зависимость абортот от дохода

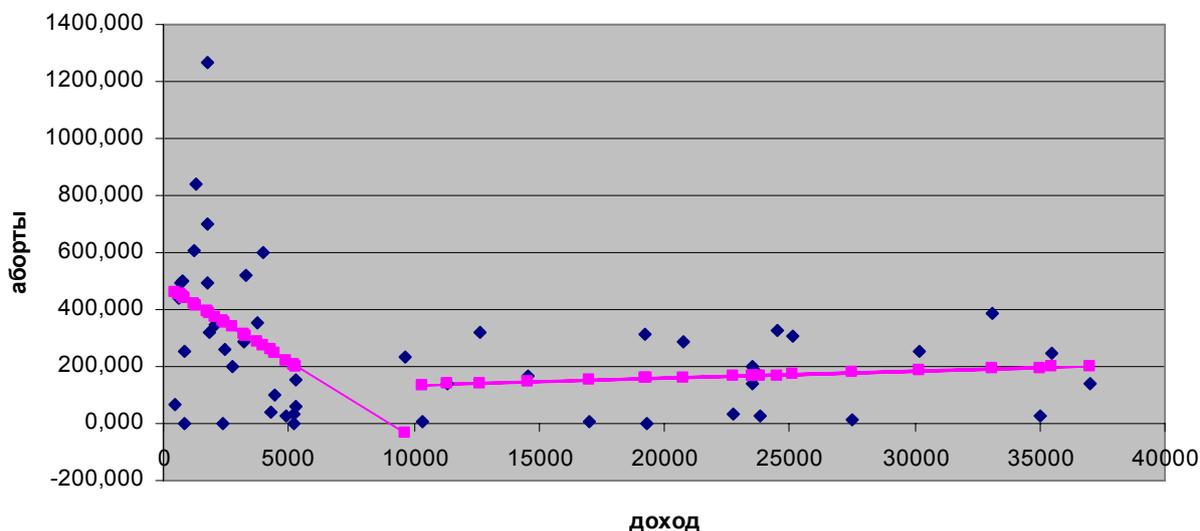


Рис. 5

Построим регрессию по трем факторам (X_3 , d и X_3d). Тогда:

$$Y = 485,491 - 0,0535 \cdot X_3 - 377,119 \cdot X_3 \cdot d + 0,0561 \cdot d - \text{уравнение};$$

$$R^2 = 0,2191 - \text{коэффициент детерминации};$$

$$\bar{A} = 410 \% - \text{средняя ошибка аппроксимации}.$$

График кусочно-линейной модели изображен на рис. 5.

Сравним полученные результаты по кусочно-линейной модели с предыдущими зависимостями. Кусочно-линейная модель обладает наибольшим коэффициентом детерминации среди всех моделей. Это говорит о том, что 21,91 % вариации уровня аборт в стране объясняется вариацией доходов населения построенной кусочно-линейной регрессии при 16,4 % у полиномиальной и 9,7 % у линейной.

Проверим статистическую значимость полученных уравнений:

- F-статистика в линейной модели = 5,1714;
- F-статистика в полиномиальной модели = 4,617;
- F-статистика в кусочно-линейной модели = 4,303.

Все три полученные F-статистики значительно превышают критические значения распределения Фишера для вероятности 0,05 и со степенями свободы: для линейной – 1 и 48, равное 4,0426, для полиномиальной – 2 и 47, равное 3,1950, для кусочно-линейной – 3 и 46, равное 2,8068.

Проверим статистическую значимость коэффициентов кусочно-линейной модели, т. к. как мы убедились ранее, именно она наиболее полно описывает нашу выборку:

- $t_a = 6,4856$;
- $t_b = -2,5666$;
- $t_{d1} = -2,1364$;
- $t_{j1} = 2,5649$.

Все t-статистики превышают критическое значение распределения Стьюдента с вероятностью 0,05 и 46 степенями свободы, равное 2,0129. Отсюда можно сделать вывод, что все коэффициенты кусочно-линейной модели значимы.

Таким образом, мы принимаем кусочно-линейную модель для исследования зависимости между количеством абортов в стране и доходом на душу населения. Уравнение такой модели:

$$Y = 485,491 - 0,0535 \cdot X_3 - 377,119 \cdot X_3 \cdot d + 0,0561 \cdot d .$$

При этом можно сделать вывод, что если душевой доход будет равен нулю, в стране ежегодно будет совершаться 485,491 абортов на сто тысяч населения. В значениях душевого дохода до 10 000 долларов США каждая дополнительная 1000 долларов дохода будет снижать количество абортов на величину в 53,5 человека на сто тысяч населения, а каждая 1000 долларов увеличения дохода поспособствует снижению абортов на 56,1 человек на сто тысяч. То есть после преодоления величины в 10 000 долларов душевого дохода наблюдается резкое падение темпов снижения количества абортов в зависимости от доходов населения.

2.4. Влияние количества не очень счастливых людей в стране на количество аборт

Предположим, что многие аборты совершаются женщинами из-за того, что они несчастливы в личной жизни, на работе, возможно несчастливы их мужья или родители, что также может влиять на решение о совершении аборта.

Проверим коррелированность фактора и результирующего признака: $r_{yx4} = 0,4568$. Согласно шкале Чеддока, такая связь изучаемых признаков является умеренной.

При рассмотрении поля корреляции (рис. 6) также трудно выявить вид зависимости, но можно выдвинуть предположение о существовании нескольких возможных видов зависимостей:

- Линейная зависимость. Тогда:

$$Y = 107,165 - 6,642 \cdot X_4 - \text{уравнение};$$

$$R^2 = 0,209 - \text{коэффициент детерминации};$$

$$\bar{A} = 421 \% - \text{средняя ошибка аппроксимации}.$$

- Полиномиальная зависимость. Тогда:

$$Y = 187,221 - 2,618 \cdot X_4 + 0,16 \cdot X_4^2 - \text{уравнение};$$

$$R^2 = 0,241 - \text{коэффициент детерминации};$$

$$\bar{A} = 355 \% - \text{средняя ошибка аппроксимации}.$$

Зависимость абортов от количества не очень счастливых людей

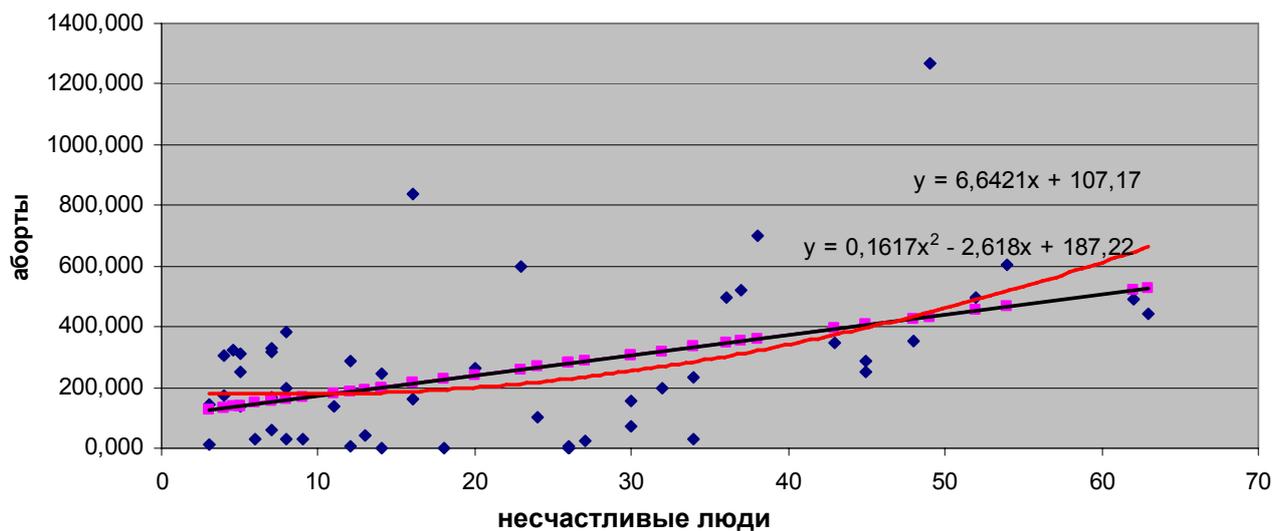


Рис. 6

Проверим статистическую значимость полученных уравнений:

- F-статистика для линейной зависимости = 12,65957909;
- F-статистика для полиномиальной зависимости = 7,455824353.

Полученные F-статистики значительно превышают критическое значение распределения Фишера для линейной модели с вероятностью 0,05 и со степенями свободы 1 и 48, равное 4,0426, и для полиномиальной модели с вероятностью 0,05 и со степенями свободы 2 и 47, равное 1,195.

Проверив статистическую значимость коэффициентов линейной модели, мы увидели, что обе t-статистики превышают критическое значение распределения Стьюдента с вероятностью 0,05 и 48 степенями свободы, равное 2,0106. Отсюда можно сделать вывод, что все коэффициенты линейной модели значимы.

Линейная модель обладает коэффициентом детерминации в 0,2068. Это говорит о том, что 20,68 % вариации уровней аборт в стране объясняется вариацией не очень счастливых людей, построенной линейной регрессией.

Таким образом, мы принимаем линейную модель для исследования зависимости между количеством аборт в стране и количеством не очень счастливых людей.

2.5. Влияние количества людей в стране, принимающих наркотики, на количество аборт

Можно предположить, что аборт совершаются из-за того, что один из родителей употребляет наркотики.

Проверим коррелированность фактора и результирующего признака: $r_{yx5} = 0,1334$. Согласно шкале Чеддока, такая связь изучаемых признаков является очень слабой.

Анализируя поле корреляции, крайне трудно принять решение о выборе какого-либо типа зависимости. Попробуем построить линейную модель, тогда:

- $Y = 234,153 + 0,0264 \cdot X_5$ – уравнение;
- $R^2 = 0,0178$ – коэффициент детерминации;
- $A' = 409 \%$ – средняя ошибка аппроксимации.

Проверим статистическую значимость линейной модели:

F-статистика в линейной модели = 0,869292966, что меньше критического значения распределения Фишера для вероятности 0,05 и со степенями свободы 1 и 48, равное 4,042651985. Следовательно, уравнение незначимо в целом.

Как мы убедились, зависимость между количеством наркоманов в стране и абортами крайне слаба. Так, только 1,78 % вариации уровней аборт в странах объясняется вариацией наркоманов построенной ли-

нейной регрессии. Возможно, это объясняется тем, что люди, употребляющие наркотики, надеются с появлением ребенка начать «новую жизнь». Исходя из этого, крайне сложно оценивать данный фактор как значимый. Но при построении множественной регрессии будет любопытным добавление этого фактора в совокупность.

3. Исследование влияния совокупности факторов

В качестве факторов, оказывающих влияние на количество абортов в стране, после предварительного исследования были отобраны четыре нижеследующих:

➤ X_1 – количество заключенных по странам, рассчитанное на 100 тысяч человек. Здесь:

- $Y = 26,96 + 1,30296 \cdot X_1$ – уравнение.

➤ X_2 – количество мужчин, покончивших жизнь самоубийством, рассчитанное также на 100 тысяч человек. Здесь:

- $Y = 124,03 + 5,9074 \cdot X_2$ – уравнение.

➤ X_3 – доход на человека в долларах США. Здесь:

- $Y = 485,491 + 0,0535 \cdot X_3 - 377,119 \cdot X_3 \cdot d + 0,0561 \cdot d$ – уравнение.

➤ X_4 – число не очень счастливых людей в стране. Здесь:

- $Y = 107,165 - 6,642 \cdot X_4$ – уравнение.

Хотя влияние фактора X_5 – количество людей, принимающих наркотики в стране, – было признано весьма небольшим, мы добавим его в многофакторную регрессию и понаблюдаем за результатами. Показатели для X_5 :

- $Y = 234,153 + 0,0264 \cdot X_5$ – уравнение.

Построим матрицу выборочных коэффициентов корреляции Q' (таблица):

Таблица

Q'	Y	X_1	X_2	X_3	X_4	X_5
Y	1	0,68793269	0,3522556	-0,311866	0,456836	0,13337
X_1	0,68793269	1	0,27274061	-0,3856655	0,4576051	0,24749
X_2	0,3522556	0,27274061	1	-0,279175	0,382428	0,251583
X_3	-0,311866	-0,385665	-0,279175	1	-0,705318	-0,220283
X_4	0,456836	0,4576051	0,382428	-0,705318	1	0,288086
X_5	0,13337	0,24749	0,251583	-0,220283	0,288086	1

После анализа матрицы Q' становится очевидно, что наибольшее влияние на величину аборт в стране оказывает X_1 (количество заключенных), обладающий самым большим коэффициентом детерминации. При этом очень велика корреляция между факторами X_3 и X_4 , что ставит под сомнение возможность построения модели множественной регрессии при данной выборке. Однако попробуем ее построить.

Для оценки оправданности включения фактора мы используем коэффициент F , который будем сравнивать со значениями распределения Фишера для вероятности 0,05. Скорректированный коэффициент детерминации множественных регрессий мы будем сравнивать со скорректированным коэффициентом регрессии Y на X_1 , равным 0,473251386.

Построим регрессию Y через X_1 и X_2 , тогда:

- $F = 2,766387977$;
- $R^2_{adj} = 0,470778737$;
- F -статистика = 23,73922064.

Заметим, что значение коэффициента F меньше значения распределения Фишера для вероятности 0,05 и степеней свободы 3 и 47, равного 2,80235. Следовательно, включение фактора X_1 во множественную регрессию неоправданно, хотя в то же время полученное уравнение признается значимым.

Построим регрессию Y через X_1 и X_3 , тогда:

- $F = 0,988848322$;
- $R^2_{adj} = 0,45092083$;
- F -статистика = 11,51538737.

Значение коэффициента F меньше значения распределения Фишера для вероятности 0,05 и степеней свободы 4 и 45, равного 2,578739184. Следовательно, включение фактора X_3 во множественную регрессию неоправданно, хотя в то же время полученное уравнение признается значимым.

Построим регрессию Y через X_1 и X_4 , тогда:

- $F = 2,39270586$;
- $R^2_{adj} = 0,466774897$;
- F -статистика = 23,38451433.

Заметим, что значение коэффициента F меньше значения распределения Фишера для вероятности 0,05 и степеней свободы 2 и 47, равного 3,195056281. Следовательно, включение фактора X_4 во множественную регрессию неоправданно, хотя в то же время полученное уравнение признается значимым.

Построим регрессию Y через X_1 и X_5 , тогда:

- $F = 0,12967256$;
- $R^2_{adj} = 0,441170938$;
- F -статистика = 21,23639923.

Заметим, что значение коэффициента F меньше значения распределения Фишера для вероятности 0,05 и степеней свободы 2 и 47, равного 3,195056281. Следовательно, включение фактора X_5 во множественную регрессию неоправданно, хотя в то же время полученное уравнение признается значимым.

Построим регрессию Y через X_3 , X_5 и X_1 , тогда:

- $F = 1,195$;
- $R^2_{adj} = 0,7349$;
- F -статистика = 24,398.

Значение коэффициента F меньше значения распределения Фишера для вероятности 0,05 и степеней свободы 3 и 46, равного 2,8068. Следовательно, включение фактора X_1 в эту регрессию является неоправданным, хотя в то же время полученное уравнение признается значимым.

Построим регрессию Y через X_3 , X_5 и X_2 , тогда:

- $F = 0,1173$;
- $R^2_{adj} = 0,7268$;
- F -статистика = 34,673.

Значение коэффициента F меньше значения распределения Фишера для вероятности 0,05 и степеней свободы 1 и 48, равного 4,0426. Следовательно, включение фактора X_2 в эту регрессию является неоправданным, хотя в то же время полученное уравнение признается значимым.

Стоит отметить, что в конечной множественной регрессии присутствует только один фактор из пяти, что, скорее всего, говорит о наличии в данной модели эффекта мультиколлинеарности. Из матрицы выборочных коэффициентов корреляции Q' следует, что эффект мультиколлинеарности в данной модели присутствует. Очевидно, что факторы X_1 , X_3 и X_4 связаны друг с другом. Построив поля корреляции и регрессии для зависимостей между указанными факторами, можно сделать выводы о том, что количество заключенных на 19,65 % определяется уровнем доходов в стране и на 21,26 % – уровнем не очень счастливых людей. Тогда как количество несчастливых людей на 57,6 % определяется уровнем доходов в стране. Это говорит о том, что мультиколлинеарность между тремя факторами существует.

ЗАКЛЮЧЕНИЕ

Во время подробного анализа было выявлено, что наибольшее влияние на количество аборт в стране оказывают число заключенных и уровень доходов в стране. Таким образом, женщинам стоит задумываться о том, с кем они связывают свою жизнь, а государству, чтобы улучшить демографическую проблему, необходимо сконцентрировать внимание на уровне жизни и доходов в стране. Также достаточно значимыми для количества аборт являются такие факторы, как суицид мужчин и количество несчастливых людей в стране. Абсолютно не влияющим на уровень аборт был признан такой социальный фактор, как потребление наркотиков жителями страны.

Как показала данная работа, экономические и социальные факторы имеют заметное влияние на уровень аборт. Но женщинам не стоит обвинять экономику и правительство, в первую очередь необходимо разобраться в самом себе и не позволить себе убить кого-то. Однако государствам также необходимо задуматься об этом, хотя они предпринимать какие-то шаги не спешат. В департаменте здравоохранения Москвы отвечают: «Нет статистики по абортам. А пока нет данных, нет и проблемы».

ЛИТЕРАТУРА

1. *Евсеев Е.А., Буре В.М.* Эконометрика. СПб.: Изд-во МБИ, 2007. –139 с.
2. *Тарашина С.И., Панкратова Я.Б.* Выполнение курсовой работы по эконометрике: Учеб.-метод. пособие. СПб.: Изд-во МБИ, 2007. – 97 с.
3. Практикум по эконометрике: Учеб. пособие / И.И. Елисеева, С.В. Курышева, Н.М. Гордеенко и др. Под ред. И.И. Елисеевой. М.: Финансы и статистика, 2003. – 192 с.: ил.
4. Официальный сайт Европейской статистики [Электронный ресурс]: www.eurostat/eu.com
5. Johnston's Archive [электронный ресурс]: <http://www.johnstonsarchive.net/>

Научное издание

ВЕСТНИК № 27

Межвузовский студенческий журнал

Корректор *И.С. Ловкис*

Технический редактор *Л.В. Соловьева*

Директор РИЦ МБИ А.И. Стригун

Подписано в печать 18.10.10.

Усл. печ. л. 8,9. Тираж 100 экз. Заказ 499.

РИЦ МБИ

191011, Санкт-Петербург, Невский пр., д. 60

тел. (812) 570-55-72

ISBN 978-5-903028-90-0



9 785903 028900